KENDRIYA VIDYALAYA SANGTHAN, RANCHI REGION

Support Material

Informatics Practices (065)

2023-24

Sh. D. P Patel
Deputy Commissioner
Kendriya Vidyalaya Sangathan Ranchi Region
Chief Patron

SHRI SURESH KUMAR SINGH
Assistant Commissioner
KVS RO Ranchi
Patron

SMT. SUJATA MISHRA
Assistant Commissioner
KVS RO Ranchi
Patron

SHRI BALENDER KUMAR
Assistant Commissioner
KVS RO Ranchi
Patron

SHRI SILAS PURTI
Principal
KV No. 1, HEC Ranchi
Subject Convenor

SHRI RAVI PRAKASH
PGT (CS)
KV Namkum, Ranchi
Assistant Convenor

मेरे प्रिय विद्यार्थियों के लिए दो शब्द

केन्द्रीय विद्यालय संगठन, राँची संभाग के 12वीं कक्षा के विद्यार्थियों हेतु छात्र सहायता सामग्री प्रस्तुत करते हुए मुझे अपार हर्ष हो रहा है। 12वीं कक्षा के छात्रों, यह सामग्री आपकी आवश्यकताओं को ध्यान में रखते हुए तैयार की गई है। जब आप परीक्षा की तैयारी के अंतिम चरण में होते हैं तब आप एक स्थान पर सभी संभावित प्रश्नों को देख अपने श्रम को एकाग्र कर पाते हैं। इस सहायता सामग्री से प्रश्नों को दोहराना और अभ्यास करना सहज होगा। जब आप आवंटित समय में प्रश्न पत्र पूरा करने की अपनी क्षमता का परीक्षण करना चाहते हों या जब आप अध्ययन करते समय कोई प्रश्न देखते हों तब यह सहायक सामग्री आपकी सहायता हेतु होगा। तैयारी करते समय कभी-कभी तत्काल उत्तर की आवश्यकता है, लेकिन पाठ्य-पुस्तक से खोजने और पढ़ने में समय लगेगा। वैसी स्थिति में जब आप कम समय में पूरी अवधारणा या विचार को समझना और जानना चाहते हैं तब यह सामग्री आपको तुरंत परेशानी से बचा लेगी। यह पिछले सीबीएसई बोर्ड परीक्षा के पेपर और प्रतियोगी परीक्षा के किसी प्रश्न को जानने और समझने में मदद करेगा।

अपने विषयों में विशेषज्ञता रखने वाले समर्पित और अनुभवी शिक्षकों की एक टीम ने कड़ी मेहनत के बाद इस सामग्री को तैयार किया है। केवल उन्हीं वस्तुओं को शामिल करने का ध्यान रखा गया है जो प्रासंगिक हैं और पाठ्य-पुस्तक के अनुरूप हैं। इस सामग्री को एनसीईआरटी पाठ्य पुस्तक के विकल्प के रूप में नहीं लिया जाना चाहिए बल्कि इसे इसके पूरक के रूप में डिज़ाइन किया गया है। छात्रों की सहायता सामग्री में आपके लिए आवश्यक सभी महत्वपूर्ण पहलू हैं: प्रश्न पत्र का डिज़ाइन, पाठ्यक्रम, सभी इकाइयों/अध्यायों या बिंदुओं में अवधारणाएं, प्रत्येक अध्याय से नमूना परीक्षण आदि । मुझे यकीन है कि सहायक सामग्री का उपयोग छात्रों और शिक्षकों दोनों द्वारा किया जाएगा और मुझे विश्वास है कि यह सामग्री आपको अपनी परीक्षाओं में अच्छा प्रदर्शन करने में मदद करेगी । आनेवाली परीक्षा के लिए शुभकामनाओं के साथ आप यह अवश्य याद रखें मेहनत का कोई विकल्प नहीं है।

डी.पी.पटेल
उपायुक्त
केन्द्रीय विद्यालय संगठन
क्षेत्रीय कार्यालय राँची

INDEX

Syllabus	4
Marks Division	5
Introduction to Python libraries- Pandas, Matplotlib	7-11
Pandas Series	12-23
Pandas DataFrame	24-46
Data Visualization	51-69
Database Query using SQL	70-110
Database Concept	
Structured Query Language	
Introduction to MySQL	
Simple Query	
Aggregate Function and Grouping Data	
Joins	
Computer Networks	111-133
Societal Impact	134-148
Bibliography -	149

Syllabus

Unit 1: Data Handling using Pandas and Data Visualization

- Introduction to Python libraries- Pandas, Matplotlib.
- Data structures in Pandas - Series and data frames.
- Series: Creation of series from ndarray, dictionary, scalar value; mathematical operations; head and tail functions; selection, indexing and slicing.
- Data Frames: creation of data frames from dictionary of series, list of dictionaries, text/CSV files, display, iteration. Operations on rows and columns: add, select, delete, rename, Head and Tail functions, indexing using labels, Boolean indexing.
- Importing/Exporting Data between CSV files and Data Frames.
- Purpose of plotting; drawing and saving following types of plots using Matplotlib – line plot, bar graph, histogram
- Customizing plots: adding label, title, and legend in plots.

Unit 2: Database Query using SQL

Revision of Database concepts and SQL commands covered in Class XI [Addition]

- Math functions: POWER (), ROUND (), MOD ().
- Text functions: UCASE ()/UPPER (), LCASE ()/LOWER (), MID ()/SUBSTRING ()/SUBSTR(), LENGTH (), LEFT (), RIGHT (), INSTR (), LTRIM (), RTRIM (), TRIM ().
- Date Functions: NOW (), DATE (), MONTH (), MONTHNAME (), YEAR (), DAY (), DAYNAME ().
- Aggregate Functions: MAX (), MIN (), AVG (), SUM (), COUNT (); using COUNT (*).
- Querying and manipulating data using Group by, Having, Order by.
- **Working with two tables using equi-join. [Addition]**

Unit 3: Introduction to Computer Networks

- Introduction to networks, Types of networks: LAN, MAN, WAN.
- Network Devices: modem, hub, switch, repeater, router, gateway.
- Network Topologies: Star, Bus, Tree, Mesh.
- Introduction to Internet, URL, WWW and its applications- Web, email, Chat, VoIP.
- Website: Introduction, difference between a website and webpage, static vs dynamic web page, web server and hosting of a website.
- Web Browsers: Introduction, commonly used browsers, browser settings, add-ons and plug-ins, cookies.

Unit 4: Societal Impacts

- Digital footprint, net and communication etiquettes,
- Data protection, intellectual property rights (IPR), plagiarism, licensing and copyright,
- Free and open-source software (FOSS),
- Cybercrime and cyber laws, hacking, phishing, cyber bullying, overview of Indian IT Act.
- E-waste: hazards and management.

● Awareness about health concerns related to the usage of technology.

● E-waste: hazards and management. Awareness about health concerns related to the usage of technology.

UNIT	UNIT NAME	MARKS	PERIODS	
			THEORY	PRACTICAL
1	Data Handling using Pandas and Data Visualization	25	25	25
2	Database Query using SQL	25	20	17
3	Introduction to Computer Networks	10	12	0
4	Societal Impacts	10	14	--
	Project	-	--	7
5	Practical	30	-	-
	TOTAL	100	71	49

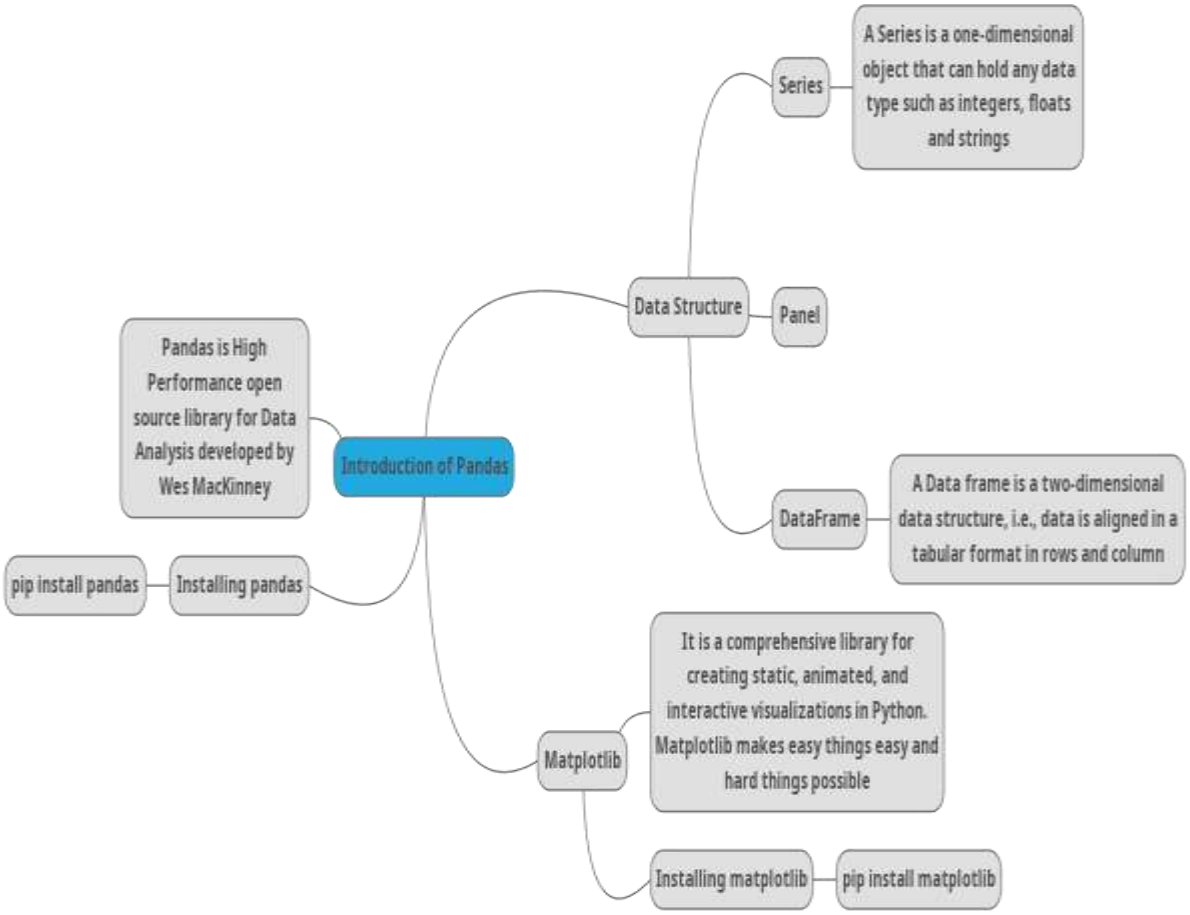
Contributions

SL	Name of Teachers	Topic
01	Sh Mangal Rai, PGT (Comp) KV Maithon Dam	Introduction to Python Library Pandas and Matplotlib Data Structure in Pandas Series Operation
02	Ms. Uzma Ghazali, PGT (Comp) KV Chakradarpur	Dataframe operations Data Transfer CSV/Dataframe Data Visualization
03	Sh. Mithilesh Kumar, PGT (Comp) KV No. 1, HEC Ranchi	Database Query using SQL
04	Smt. G. Sadangi, PGT (Comp) KV Tatanagar	Introduction to Computer Networks
05	Sh. Chandan Lohani, PGT (Comp) KV Surda	Societal Impact

Moderation and Compilation

Sh. Ravi Prakash, PGT (Comp) KV Namkum, Ranchi

Introduction to Python libraries- Pandas, Matplotlib



Introduction to Python libraries- Pandas, Matplotlib.

- Python libraries contains a collection of built-in modules that allow us to perform many actions without writing detailed programs for it.
- Each library in Python contains a large number of modules that one can import and use.
- **NumPy, Pandas and Matplotlib** are three well established python libraries for scientific and analytical use.
- These libraries allow us to manipulate, transform and visualize data easily and efficiently.

NumPy

- It stands for 'Numerical Python'.
- It is a package that can be used for numerical data analysis and scientific computing.
- NumPy uses a multidimensional array object and has functions and tools for working with these arrays.
- Elements of an array stay together in memory; hence, they can be quickly accessed.

Pandas

- PANDAS (PANelDAta) is a high-level data manipulation tool used for analysing data.

- It is very easy to import and export data using Pandas library which has a very rich set of functions.
- It is built on packages like NumPy and Matplotlib and gives us a single, convenient place to do most of our data analysis and visualization work.
- Pandas has three important data structures, namely – **Series**, **DataFrame** and **Panel** to make the process of analysing data organised, effective and efficient.
- It is installed using command **pip install pandas**

Matplotlib

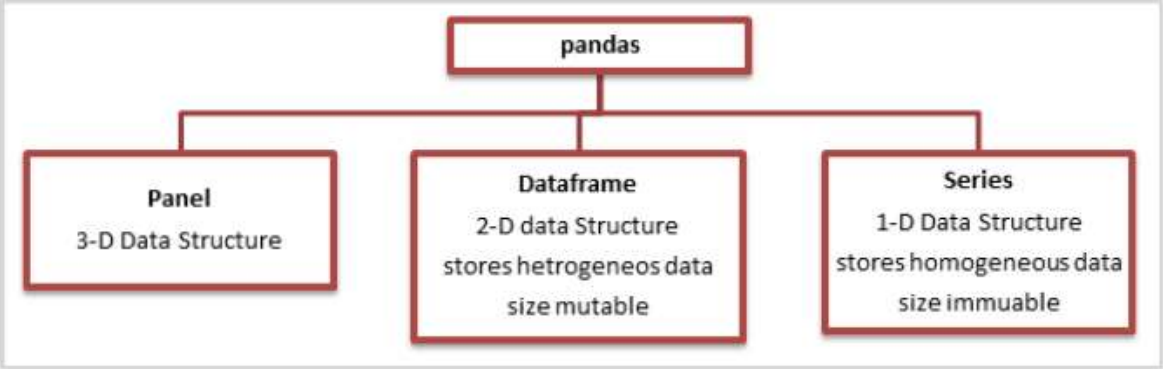
- The Matplotlib library in Python is used for plotting graphs and visualisation.
- Using Matplotlib, with just a few lines of code we can generate publication quality plots, histograms, bar charts, scatterplots, etc.
- It is also built on Numpy, and is designed to work well with Numpy and Pandas.
- It is used to installed using command **pip install matplotlib**

Data structures in Pandas - Series and Data Frames.

- Out of many data structures of Pandas, two basic data structures – Series and Dataframe- are universally popular for their dependability.
- For working in Pandas, we have to import both pandas and numpy module/Libraries by giving the following import statements:

  import pandas as pd

  import numpy as np



Series:

- Series is an important data structure of pandas.
- It represents s a **one-dimensional** array of indexed data.
- A series type object has two main components:
  - any array of actual data.
  - an associated array of index s of data labels.
- A series type object can be created in many ways using pandas library's function **Series**().
- It is value mutable i.e., value stored in Series can be modified

DataFrame :

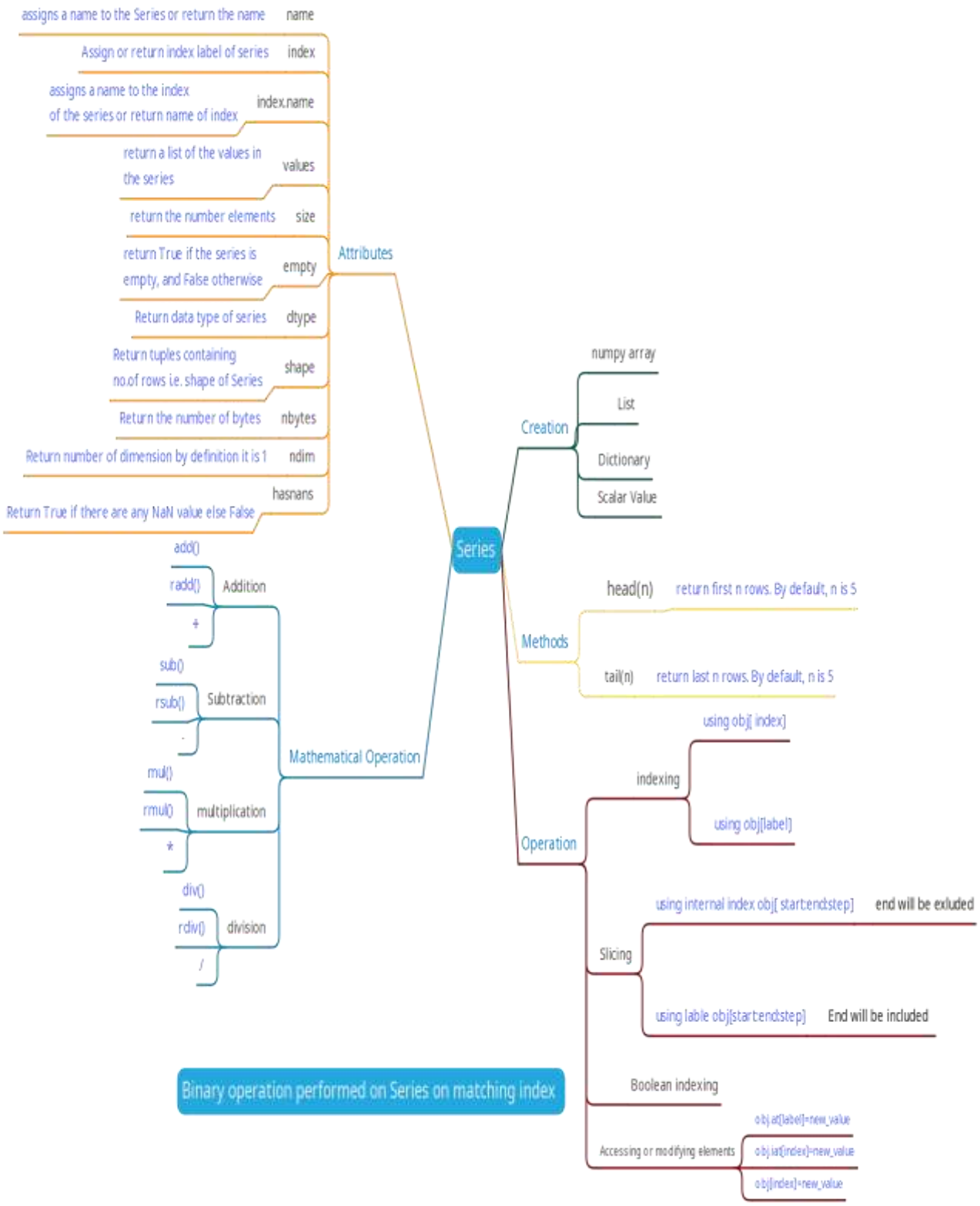
- A DataFrame is another pandas data structure, which store data in two-dimensional way.
- It is actually a **two dimensional** (tabular and spreadsheet like) labeled array, which is an

ordered collection of columns where columns may store different types of data, e.g., numeric or string or floating point or Boolean type etc.

Major characteristics of a DataFrame data structure is as follows:

- It has two indices or two axes- a row index(axis=0) and a column index (axis=1).
- It is like a spreadsheet where each value is identifiable with the combination of **row index and column index**. The **row index is known as index** and the **column index is called column-name.**
- The indices can be of numbers or letters or strings.
- There is no condition of having all the data of same type across columns; its columns can have data of different types.
- You can easily change its value, i.e. it is value mutable.
- You can add or delete rows/columns in a DataFrame, i.e. it is size –mutable.

Pandas Series



- A series is a one-dimensional array containing a sequence of values of any data type (int ,

float, list, string etc.

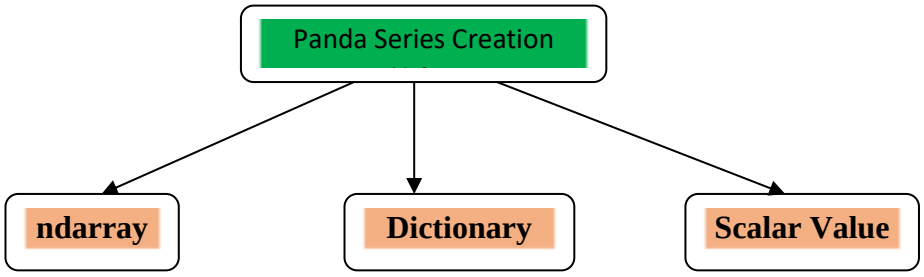
- It has by default numeric data labels (index) starting from zero.
- We can also assign values of other data types as index. It is like a column in a spreadsheet.

For Example

Index	Value
0	Aman
1	Smriti
2	Ronit
3	Divya
4	Kirty

Creation of Series from –list, ndarray, dictionary, scalar value.

- There are different ways in which a series can be created in Pandas.
- To create or use series, we first need to the Pandas import library (import pandas as **pd**)



- **Series creation from scalar values**

```
import pandas as pd

ser1 = pd.Series([10,20,30])

print(ser1)
```

Output

```
>>>
= RESTART: C:/Users/KV
py
0    10
1    20
2    30
dtype: int64
>>>
```

- **Assigning user defined labels to the index and use them to access elements of a series.**

```
import pandas as pd

ser2 = pd.Series(["Ram","Shyam","Sita"],index =[3,5,2])

print(ser2)
```

Output:

```
>>>
= RESTART: C:/Us
3     Ram
5   Shyam
2    Sita
dtype: object
>>>
```

Question: Create a series having names of any five famous Dams of India and assign their state as index value.

- **Creation of series from NumPy arrays**

We can create series from 1-Dimensional NumPy array as follows: -

```
>>>import numpy as np  # import NumPy with alias np
>>>import pandas as pd
>>>ar1 = np.array([1,2,3,4])
>>>ser3 = pd.Series(ar1)
>>>print(ser3)
```

```
Output:
0    1
1    2
2    3
3    4
dtype: int32
```

- **Creation of series and using letters or strings as indices:**

```
import pandas as pd
import numpy as np
array1 = np.array([1,2,3])
ser4 = pd.Series(array1,index =["JAN","FEB","MAR"])
print(ser4)
Output:
```

```
>>>
= RESTART: C:/
py
JAN    1
FEB    2
MAR    3
dtype: int32
>>>
```

Note: When index labels are passed with the array, then the length of the index and array must be of the same size, else it will result in a ValueError.

- **Creation of Series from Dictionary**
  - Python Dictionary is a key-value pair and value can be quickly retrieved when its key is known.
  - Dictionary keys can be used to construct an index for a Series. e.g.

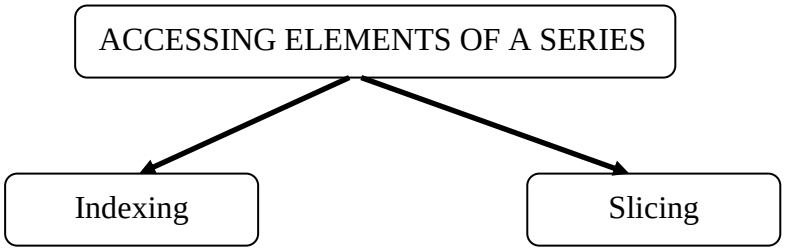
```
>>>import pandas as pd
>>>dict1 = {'India': 'NewDelhi', 'UK':'London', 'Japan': 'Tokyo'}
>>>print(dict1)                #Display the dictionary
>>> series8 = pd.Series(dict1)
>>>print(series8)              #Display the series
```

OUTPUT:

```
>>>
= RESTART: C:/Users/KVM/AppData/Local/Programs/Python/Pyt
{'India': 'NewDelhi', 'UK': 'London', 'Japan': 'Tokyo'}
India    NewDelhi
UK         London
Japan       Tokyo
dtype: object
>>>
```

- **Accessing Elements of a series**

There are two ways for accessing the elements of a series: Indexing and slicing



- **Indexing**

Indexing in Series is similar to that for NumPy arrays, and is used to access elements in a series. Indexes are of two types: **Positional index** and **Labelled index**.

- ❖ Positional index takes an integer value that corresponds to its position in the series starting from 0.
- ❖ Labelled index takes any user-defined label as index.
- ❖ Using obj.at[label] and obj.iat(index)

Following example shows usage of the positional index for accessing a value from a Series.

```
>>>seriesNum = pd.Series([10,20,30])
>>>seriesNum[2]
Output: 30
```

Here, the value 30 is displayed for the positional index 2

In the following example, value NewDelhi is displayed for the labelled index India.

```
>>>seriesCapCntry = pd.Series(['NewDelhi', 'WashingtonDC', 'London', 'Paris'], index=['India', 'USA', 'UK', 'France'])

>>>seriesCapCntry['India']
```

Output: 'NewDelhi'

We can also access an element of the series using the positional index:

```
>>>seriesCapCntry[1]
```

Output: 'WashingtonDC'

More than one element of a series can be accessed using a list of positional integers or a list of index labels as shown in the following examples:

```
>>>seriesCapCntry[[3,2]]
France  Paris
UK      London
dtype: object
>>>seriesCapCntry[['UK','USA']]
UK      London
USA     WashingtonDC
dtype: object
```

The index values associated with the series can be altered by assigning new index values as shown in the following example:

```
>>>seriesCapCntry.index=[10,20,30,40]

>>>seriesCapCntry
```

Output:

```
10 NewDelhi
20 WashingtonDC
30 London
40 Paris
dtype: object
```

- **Slicing**
  - Slicing is process of getting part of a sequence.
  - We can define which part of the series is to be sliced by specifying the start and end parameters

    SeriesObj[start: end:step]

  - When we use positional indices for slicing, the value at the end_index position is excluded, i.e., only (end -start) number of data values of the series are extracted. Consider the following series seriesCapCntry:

    >>>**seriesCap = pd.Series(['NewDelhi', 'WashingtonDC', 'London', 'Paris'], index=['India', 'USA', 'UK', 'France'])**

    >>>**seriesCap[1:3]** #excludes the value at index position 3

    ```
    OUTPUT:
    USA    WashingtonDC
    UK     London
    dtype: object
    ```

  - If labelled indexes are used for slicing, then value at the end index label is also included in the output, for example:

    >>>**seriesCap['USA' : 'France']**

    ```
    OUTPUT:
    USA          WashingtonDC
    UK           London
    France Paris
    dtype: object
    ```

  - We can also get the series in reverse order, for example:

    >>>**seriesCap[: : -1]**

    OUTPUT:

    ```
    France Paris
    UK           London
    USA          WashingtonDC
    India  NewDelhi
    dtype: object
    ```

- We can also use slicing to modify the values of series elements as shown in the following example:

  >>> **import numpy as np**

  >>>**seriesAlpha = pd.Series(np.arange(10,16,1),index = ['a', 'b', 'c', 'd', 'e', 'f'])**

  >>>**seriesAlpha**

  a 10
  b 11
  c 12
  d 13
  e 14
  f 15
  dtype: int32

- Updating the values in a series using slicing also excludes the value at the end index position.

  But it changes the value at the end index label when slicing is done using labels.

  >>>**seriesAlph['c':'e'] = 500**

  >>>seriesAlph

  a 10

  b 50

  c 500

  d 500

  e 500

  f 15

  dtype: int32

- **Attributes of Pandas Series**

Attribute Name	Purpose	Example
name	Assigns or return a name to the Series	>>> seriesCap.name = 'Capitals' >>> print(seriesCap)
index.name	assigns a name to the index of the series	>>>seriesCap.index.name ='Countries' >>> print(seriesCap)
values	return a list of the values in the series	>>> print(seriesCap.values)
size	return the number elements	>>> print(seriesCap.size)
empty	return True if the series is empty, and False otherwise	>>>print(seriesCap.empty)
dtype	Return data type of series	>>>print(seriesCap.dtype)

shape	Return tuples (no.of rows) of the shape.	>>>print(seriesCap.shape)
nbytes	Return number of bytes	>>>print(seriesCap.nbytes)
ndim	Return number of dimension by definition it is 1	>>>print(seriesCap.ndim)
hasnans	Return True if there are any NaN value else False	>>>print(seriesCap.hasnans)

We can access certain properties called attributes of a series by using that property with the series name. using **seriesCap** as an example:

>>>**seriesCap**
India NewDelhi
USA WashingtonDC
UK London
France Paris
dtype: object

Some of the methods those are available for Pandas Series.

Let us consider the following series:

>>>**seriesTenTwenty=pd.Series(np.arange(10,20, 1))**

>>>**print(seriesTenTwenty)**

0 10
1 11
2 12
3 13
4 14
5 15
6 16
7 17
8 18
9 19
dtype: int32

Method	Description	Example
head(n)	Returns the first n members of the series. Default value for **n is 5**. i.e. if n is not gives, it return first 5 rows from the Series.	>>>seriesTenTwenty.head(2)
count()	Returns the number of non-NaN values in the Series	>>>seriesTenTwenty.count()

tail(n)	Returns the last n members of the series. Default value for **n is 5**. i.e. if n is not gives, it return last 5 rows from the Series.	>>>seriesTenTwenty.tail(3)

- **Mathematical Operations on Series**
  - On Series various mathematical operation can be performed.
  - Binary operation: While performing mathematical operations in series, the operation is performed on matching index and rest filled with NaN by default.

Consider the following series: **seriesA** and **seriesB** for understanding mathematical operations on series in Pandas.

```
seriesA = pd.Series([1,2,3,4,5], index =['a', 'b', 'c', 'd', 'e'])
seriesB = pd.Series([10,20,-10,-50,100],index = ['z', 'y', 'a', 'c', 'e'])
```

```
print(seriesA)
```

```
a    1
b    2
c    3
d    4
e    5
dtype: int64
```

```
print(seriesB)
```

```
z     10
y     20
a    -10
c    -50
e    100
dtype: int64
```

- ***Addition of two Series***

```
In [15]: seriesA+seriesB

Out[15]: a     -9.0
         b      NaN
         c    -47.0
         d      NaN
         e    105.0
         y      NaN
         z      NaN
         dtype: float64
```

- The **second method** is applied when we do not want to have NaN values in the output.
- We can use the series method add() and a parameter fill value to replace missing value with a specified value.

- That is calling seriesA.add(seriesB) is equivalent to calling seriesA+seriesB, but add() allows explicit specification of the fill value for any element in seriesA or seriesBthat might be missing,

  >>>seriesA.add(seriesB, fill_value=0)

  a -9.0

  b 2.0

  c -47.0

  d 4.0

  e 105.0

  y 20.0

  z 10.0

  dtype: float64

➢ ***Subtraction of two Series***

Again, it can be done in two different ways, as shown inthe following examples:

>>>**seriesA – seriesB #using subtraction operator**

a 11.0
b NaN
c 53.0
d NaN
e -95.0
y NaN
z NaN
dtype: float64

Now replace the missing values with 1000 before subtracting seriesB from seriesA using explicit subtraction method sub().

>>>**seriesA.sub(seriesB, fill_value=1000)**

using fill value 1000 while making explicit

call of the method"

a 11.0
b -998.0
c 53.0
d -996.0
e -95.0
y 980.0
z 990.0
dtype: float64

- ***Multiplication of two Series***

Again, it can be done in two different ways, as shown inthe following examples:

```
>>>seriesA * seriesB  #using multiplication operator
a      -10.0
b       NaN
c     -150.0
d       NaN
e      500.0
y       NaN
z       NaN
dtype: float64
```

Now replace the missing values with 0 before multiplication of seriesB with seriesA using explicit multiplication method mul().

```
>>>seriesA.mul(seriesB, fill_value=0)

# using fill value 0 while making

#explicit call of the method

a      -10.0

b       0.0

c     -150.0

d       0.0

e      500.0

y       0.0

z       0.0

dtype: float64
```

- ***Division of two Series***

Again, it can be done in two different ways, as shown inthe following examples:

```
>>>seriesA/seriesB # using division operator

a      -0.10

b       NaN

c      -0.06

d       NaN

e       0.05

y       NaN
```

z NaN

dtype: float64

Now replace the missing values with 0 beforedividing seriesA by seriesB using explicit division method div().

using fill value 0 while making explicit call of the method

a -0.10

b inf

c -0.06

d inf

e 0.05

y 0.00

z 0.00

dtype: float64

Question Answer on Series

1 Marks Questions

Q1. Which of the following command will show the last 3 rows from a Pandas Series named NP?

a. NP.Tail( ) b. NP.tail(3)

c. NP.TAIL(3) d. All of the above

Ans : b. NP.tail(3)

Q2. To display last five rows of a series object **'S'**, you may write:

a. S.head() b. S.Tail(5)

c. S.head(5) d. S.tail()

Ans: **d. S.tail()**

Q3. **Assertion (A):-** To use the Pandas library in a Python program, one must import it.

Reasoning (R): - The only alias name that can be used with the Pandas library is pd.

i. Both A and R are true and R is the correct explanation for A

ii. Both A and R are true and R is not the correct explanation for A

iii. A is True but R is False

iv. A is false but R is True

Ans: iii. A is True but R is False

Q4. Which of the following statement will import pandas library?

i. Import pandas as pd ii. import Pandas as py

iii. import pandas as pd iv. import panda as pd

Ans: iii. import pandas as pd

Q5. Fill in the blanks:

```
import pandas as pd

l1 =[11,12,13,14]
```

(a) row (b) index (c) row_index (d) All of the above
Ans: (b)index

Q6. While trying to create series from dictionary, keys of dictionary become index.

(a) **True (b) False (c) Depends on Python Version (d) None of the above**

Q7. Which index,data element will be printed by below code as output:-

```
import pandas as pd

l1 =[11,12,13,14]

series1 = pd.Series(l1)
```

(a) Data : 12,13,14 along with series-index 1,2,3
(b) Data : 13,14 along with series-index 'C','D'
(c) Data : 13,14 along with series-index 2,3
(d) Data : 12,13,14 along with series-index 'B','C','D'
Ans: (c)Data : 13,14 along with series-index 2,3

Q8. To get the number of dimensions of Series object, _____ attribute is used.

(a) **size (b) shape (c) itemsize (d) ndim**
Ans:(d) ndim

Q9. To skip not numeric or null values in series, we can use _____ attribute

(a) skip **(b) skipna (c) skipNaN (d) None of the above**

Ans:**(b) skipna**

Q10: What will be correct syntax for pandas series?

(a) pandas_Series(data, index, dtype)
(b) panda.series(data, index, dtype)
(c) pandas.Series(data, index, dtype)
(d) panda_Series(data, index, dtype)
Ans: (c) pandas.Series(data, index, dtype)

Q11. To get the number of dimensions of a Series object, _____ attribute is displayed.

(a) Index **(b) Size (c) Itemsize (d) Ndim**

Ans: **(d) Ndim**

Q12. To get the size of the datatype of the items in Series object, you can use _____ attribute.

(a) Index (b) Size (c) Itemsize (d)Ndim

Ans: (c) Itemsize

Q13. To get the number of elements in a Series object, _____ attribute may be used.

(a) Index (b) Size (c)Itemsize (d)Ndim

Ans: (b) Size

Q14. To get the number of bytes of the Series data, _____ attribute is displayed.

(a) hasnans (b) nbytes (c) ndim (d) dtype

Ans: (b) nbytes

Q15. To display third element of a Series object S, you will write_____.

```
import pandas as pd
list1=[10,20,30,40,50]
S = pd.Series(list1)
print(S[2])
```

(a) S[:3] (b) S[2] (c) S[3] (d) S[:2]

Ans: (b) S[2]

[2 Marks Questions]

Q1. Predict the output of the given Python code:

```
import pandas as pd
list1=[-10,-20,-30]
ser = pd.Series(list1*2)
print(ser)
```

Ans:
```
0   -10
1   -20
2   -30
3   -10
4   -20
5   -30
```
(2 marks for correct output)

Q2. Complete the given Python code to get the required output as: **Rajasthan**

```
import _____ as pd
di = {'Corbett': 'Uttarakhand', 'Sariska': 'Rajasthan', 'Kanha': 'Madhya Pradesh', 'Gir':'Gujarat'}
NP = _____. Series( _____ )
print(NP[ _____ ])
```

Ans:
```
import pandas as pd
di = {'Corbett': 'Uttarakhand', 'Sariska':'Rajasthan', 'Kanha': 'Madhya
Pradesh','Gir':'Gujarat'}
```

```
NP = pd.Series( di)
print(NP[ 'Sariska'])
```

Q3. Write a program to create a series object using a dictionary that stores the number of students in each house of class 12D of your school.

Note: Assume four house names are Beas, Chenab, Ravi and Satluj having 18, 2, 20, 18 students respectively and Pandas library has been imported as pd.

Ans: **St={'Beas' :18, 'Chenab' :20 , ' Ravi' :20, ' Satluj' :18}**
S1=pd.Series(St)

Q4. What will be the output of the following code:

```
>>>import pandas as pd

>>>A=pd.Series(data=[35,45,55,40])

>>>print(A>45)
```

Ans:
```
0    False
1    False
2    True
3    False
```

Q5. Which of the following code will generate the following output?

```
Jan    31
Feb    28
Mar    31
dtype: int64
```

(a)
```
import pandas
S1 = pd.Series(data = [31,28,31], index=["Jan","Feb","Mar"])
print(S1)
```

(b)
```
import pandas as pd
S1 = p1.series([31,28,31], index=["Jan","Feb","Mar"])
print(S1)
```

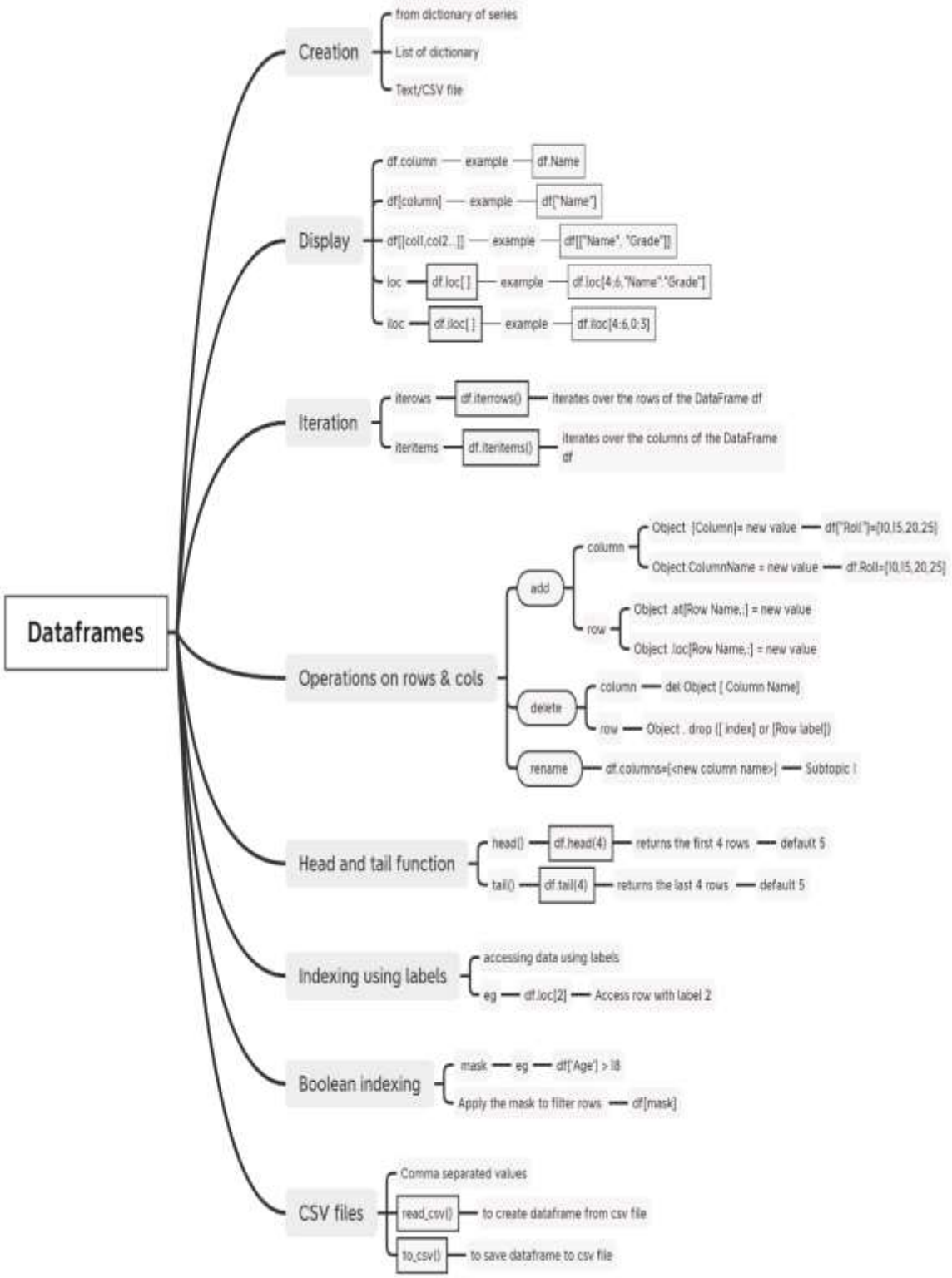
(c)
```
import pandas as pd
S1 = pd.Series([31,28,31], columns=["Jan","Feb","Mar"])
print(S1)
```

(d)
```
import pandas as pd
S1 = pd.Series([31,28,31], index=["Jan","Feb","Mar"])
print(S1)
```

Ans.(d)
```
import pandas as pd

S1 = pd.Series([31,28,31], index=["Jan","Feb","Mar"])
print(S1)
```

Pandas- DataFrame

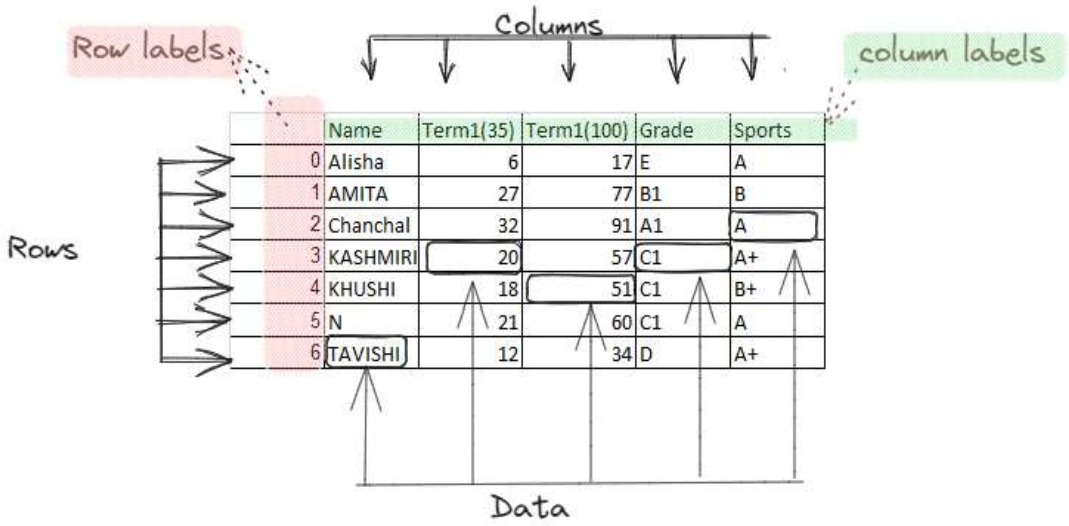


Pandas: Introduction

- Pandas has derived its name from ***"Panel Data System"***.
- ***Data Analysis*** refers to process of evaluating big data sets using analytical and statistical tools so as to discover useful information and conclusions to support business decision making.
- Python Pandas Library introduces data structures ***Series*** and ***Dataframe*** and some useful functions with these data structures.
- ***Data Structure:*** Data Structures refer to specialized way of storing data so as to apply a specific type of functionality on them.
- Pandas has two basic data structures: dataframes, series

Data Frames

- A DataFrame is a two-dimensional labelled data structure that is similar to a spreadsheet.
- It is a collection of data that is organized into rows and columns.
- Each row represents a single data point, and each column represents a single attribute of that data point.
- It is both size mutable and value mutable.



	Name	Term1(35)	Term1(100)	Grade	Sports
0	Alisha	6	17	E	A
1	AMITA	27	77	B1	B
2	Chanchal	32	91	A1	A
3	KASHMIRI	20	57	C1	A+
4	KHUSHI	18	51	C1	B+
5	N	21	60	C1	A
6	TAVISHI	12	34	D	A+

Main characteristics of DataFrame

- It has two indexes i.e. two axis- a row index (axis = 0) and a column index (axis=1).
- The row index is known as index and the column index is called column-name.
- The indexes can be of numbers, letters or strings.
- It is value mutable as well as size mutable (rows /columns can be elete/added/modified)

DataFrame Attributes

All information related to a DataFrame object is available through attributes. **<DataFrame object> . <attribute name>**

Attribute	Description
index	Returns the index (row labels) of the DataFrame
columns	Returns the column labels of the DataFrame
axes	Returns a list representing both the axes of the Data Frame (axis=0 i.e. index and axis=1 i.e. columns)
values	Returns a Numpy representation of the DataFrame
dtypes	Returns the dtypes of data in the DataFrame
shape	Returns tuple of the shape of the DataFrame
ndim	Returns number of dimensions of the dataframe
size	Returns the number of elements in the dataframe
empty	Returns True if the DataFrame object is empty, otherwise False
index	Returns the index (row labels) of the DataFrame
columns	Returns the column labels of the DataFrame
axes	Returns a list representing both the axes of the Data Frame (axis=0 i.e. index and axis=1 i.e. columns)
T	Transpose index and columns of DataFrame

Creation of DataFrames :

There are several ways to create a DataFrame. One way is to use the pd.DataFrame() function. This function takes a dictionary of series or lists of dictionaries as input and creates a DataFrame from it.

- **Creating DataFrame from dictionary of Series**

```
import pandas as pd
s1=pd.Series([11,12,13,14],index=['r1','r2','r3','r4'])
s2=pd.Series([21,22,23,24],index=['r1','r2','r3','r4'])
s3=pd.Series([31,32,33,34],index=['r1','r2','r3','r4'])
di1={"one":s1,"two":s2,"three":s3}
df2=pd.DataFrame(di1,index=['r1','r2','r3','r4'])
print(df2)
```

Output :

	one	two	three
r1	11	21	31
r2	12	22	32
r3	13	23	33
r4	14	24	34

- Dictionary key became column header
- Series index became index of DataFrame
-

- **Creating DataFrame - from list of dictionaries**

```
import pandas as pd
d1={"one":33, "two":44}
d2={"one":55, "two":66}
d3={"one":77, "two":88}
list_of_d =[d1,d2,d3]
df3=pd.DataFrame(list_of_d, index=["r1","r2","r3"])
print(df3)
```

Output :

```
    one  two
r1   33   44
r2   55   66
r3   77   88
```

- **Creating DataFrame - from list of list**

```
import pandas as pd
data=[[1,2,3], [4,5,6], [7,8,9]]
df=pd.DataFrame(data)
print(df)
```

Output :

```
   0  1  2
0  1  2  3
1  4  5  6
2  7  8  9
```

- Creating DataFrame from a **CSV file**

The read_csv()This function reads a CSV file and creates a DataFrame from it

Example :

	A	B	C	D
1	Name	Test (35)	100	Grade
2	Alisha Bajra	6	17	E
3	AMITA PRADHAN	27	77	B1
4	Chanchal Jindal	32	91	A1
5	KASHMIRI JAMUDA	20	57	C1
6	KHUSHI KUMARI	18	51	C1
7	N NANDITA	21	60	C1
8	TAVISHI SAHU	12	34	D
9				

Code:

```
import pandas as pd
df=pd.read_csv("marks.csv")
print(df)
```

Output:

```
              Name  Test (35)  100 Grade
0     Alisha Bajra          6   17     E
1    AMITA PRADHAN         27   77    B1
2  Chanchal Jindal         32   91    A1
3  KASHMIRI JAMUDA         20   57    C1
4    KHUSHI KUMARI         18   51    C1
5        N NANDITA         21   60    C1
6     TAVISHI SAHU         12   34     D
```

Accessing Data in a DataFrame

df.column

df[column]

df[[col1,col2...]]

loc and iloc

- Data in a DataFrame can be accessed using the loc and iloc accessors.
- The loc accessor allows you to access data by label, while the iloc accessor allows you to access data by index.

Object .**loc** [start row : end row , start column : end column]
Object .**iloc** [start row index: end row index, start column index : end column index]

Note : The loc and iloc method both access part of rows and columns work like slices, but **loc include end row/column name** and the **iloc excluded end index** from the result.

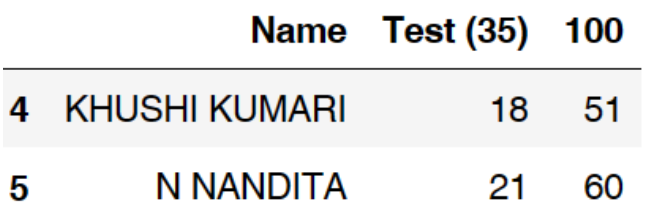
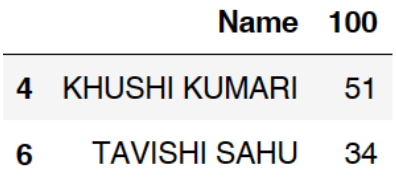
Examples

df.Name	Output : 0 Alisha Bajra 1 AMITA PRADHAN 2 Chanchal Jindal 3 KASHMIRI JAMUDA 4 KHUSHI KUMARI 5 N NANDITA 6 TAVISHI SAHU Name: Name, dtype: object
df["Name"]	Output : 0 Alisha Bajra 1 AMITA PRADHAN 2 Chanchal Jindal 3 KASHMIRI JAMUDA 4 KHUSHI KUMARI 5 N NANDITA 6 TAVISHI SAHU Name: Name, dtype: object

df[["Name", "Grade"]]	

	Name	Grade
0	Alisha Bajra	E
1	AMITA PRADHAN	B1
2	Chanchal Jindal	A1
3	KASHMIRI JAMUDA	C1
4	KHUSHI KUMARI	C1
5	N NANDITA	C1
6	TAVISHI SAHU	D

Example of using loc and iloc

df.loc[4:6,"Name":"Grade"]	Output :
df.iloc[4:6,0:3]	Output : 
df.iloc[4:8:2,0:6:2]	Output: 

	Name	Test (35)	100	Grade
4	KHUSHI KUMARI	18	51	C1
5	N NANDITA	21	60	C1
6	TAVISHI SAHU	12	34	D

	Name	Test (35)	100
4	KHUSHI KUMARI	18	51
5	N NANDITA	21	60

	Name	100
4	KHUSHI KUMARI	51
6	TAVISHI SAHU	34

- **ITERATION**

There are two main ways to iterate over a DataFrame:

- Using the **iterrows()** method: This method returns an iterator that iterates over the rows of the DataFrame. Each row is represented as a tuple of (index, row_data).
- Using the **iteritems()** method: This method returns an iterator that iterates over the items of the DataFrame. Each item is represented as a tuple of (column_name, column_data).

Using iterrows() to extract row-wise series object

Iterate over the rows of the DataFrame using iterrows()

```
for index, row in df.iterrows():
    print("Row index :" , index)
    print("containing :" )
    print(row)
```

Output :

```
Row index : 0
containing :
Name          Alisha Bajra
Test (35)                6
100                     17
Grade                    E
Name: 0, dtype: object
Row index : 1
containing :
Name          AMITA PRADHAN
Test (35)                27
100                      77
Grade                    B1
Name: 1, dtype: object
Row index : 2
containing :
Name          Chanchal Jindal
Test (35)                  32
100                        91
Grade                      A1
Name: 2, dtype: object
Row index : 3
containing :
Name          KASHMIRI JAMUDA
Test (35)                  20
100                        57
Grade                      C1
Name: 3, dtype: object
Row index : 4
containing :
Name          KHUSHI KUMARI
Test (35)                18
100                      51
Grade                    C1
Name: 4, dtype: object
Row index : 5
Name          N NANDITA
Test (35)            21
100                  60
Grade                C1
Name: 5, dtype: object
Row index : 6
containing :
Name          TAVISHI SAHU
Test (35)                12
100                      34
Grade                     D
Name: 6, dtype: object
```

Using iteritems() to extract data from a dataframe column-wise

```
for (col,colSeries) in df.iteritems():
    print("col index :", col)
    print("containing")
    print(colSeries)
```

```
col index : Name
containing
0       Alisha Bajra
1      AMITA PRADHAN
2    Chanchal Jindal
3    KASHMIRI JAMUDA
4      KHUSHI KUMARI
5          N NANDITA
6       TAVISHI SAHU
Name: Name, dtype: object
col index : Test (35)
containing
0     6
1    27
2    32
3    20
4    18
5    21
6    12
Name: Test (35), dtype: int64
col index : 100
containing
0    17
1    77
2    91
3    57
4    51
5    60
6    34
Name: 100, dtype: int64
col index : Grade
containing
0     E
1    B1
2    A1
3    C1
4    C1
5    C1
6     D
Name: Grade, dtype: object
```

OPERATIONS ON ROWS AND COLUMNS

add (insert /append), select, delete (drop column and row, rename

Arithmetic operations:

Example

```
df["Percentage"]=df["Total"]/70*100
print(df4)
```

- **Assigning a value to a column:**
  - Will modify it, if the column already exists.
  - Will add a new column if it does not exist already.

To change or add a column:

Object . Column Name = new value

OR

Object [Column]= new value

Example: Add a new column.

```
# adding a new column sports in DataFrame df
df["sports"]=["A","B","A","A+","B+","A","A+"]
print(df)
```

Output

	Name	Test (35)	100	Grade	sports
0	Alisha Bajra	6	17	E	A
1	AMITA PRADHAN	27	77	B1	B
2	Chanchal Jindal	32	91	A1	A
3	KASHMIRI JAMUDA	20	57	C1	A+
4	KHUSHI KUMARI	18	51	C1	B+
5	N NANDITA	21	60	C1	A
6	TAVISHI SAHU	12	34	D	A+

Modify an existing column

```
# modifying an existing column
df4["sports"]=["B","B","A","A","C","C","B"]
print(df4)
```

Output

	Name	Test (35)	100	Grade	sports
0	Alisha Bajra	6	17	E	B
1	AMITA PRADHAN	27	77	B1	B
2	Chanchal Jindal	32	91	A1	A
3	KASHMIRI JAMUDA	20	57	C1	A
4	KHUSHI KUMARI	18	51	C1	C
5	N NANDITA	21	60	C1	C
6	TAVISHI SAHU	12	34	D	B

There are some other ways of adding a column to a dataframe:

Object **. at** [: , Col. Name] = Value for column

Object **. loc** [: , Col. Name] = Value for column

New Object = Object **. assign** (Col. Name =[Value])

Adding/ Modifying a Row:

To change or add a row , the syntax is:

Object **.at**[Row Name,:] = new value

Object **.loc**[Row Name,:] = new value

Example:

Dataframe Df

	C1	C2	C3
0	13	23	37
1	19	20	21
2	11	12	13
3	13	14	15

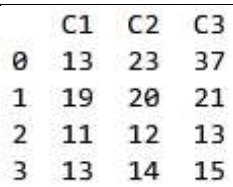
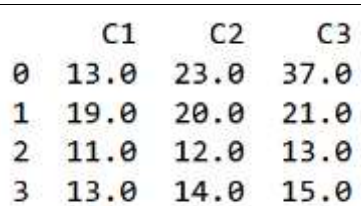
```
# code to add a row

df.loc[4,:]=[16,17,18]
print(df)
```

Output

	C1	C2	C3
0	13.0	23.0	37.0
1	19.0	20.0	21.0
2	11.0	12.0	13.0
3	13.0	14.0	15.0
4	16.0	17.0	18.0

Dataframe Df	# code to add a row	Output

	df3.**at**[5]=[19,20,21]	

Modifying a Single cell:

Syntax:

Object . Column Name [Row Name] = new value

OR

Object **. loc** [Row Index, Column index] = new value

Example:

Deleting Rows/Columns:

Python Pandas provides two ways to delete rows and columns:

del and **drop ( )** function.

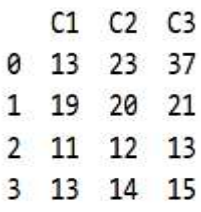
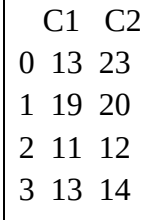
To delete a column: **del** Object [Column Name]

Drop() : to remove the columns or rows

It returns the DataFrame with dropped columns or None if inplace=True

Example:

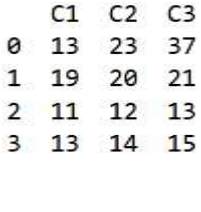
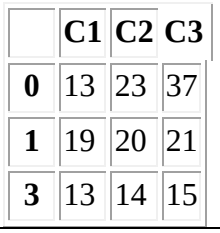
Deleting column C3 from dataframe d1

Dataframe d1	# Code to delete column C3 from dataframe	Output
	del df[“C3”] **or** df.drop(columns=["C3"],inplace=True) print(df)	

Deleting rows from a dataframe, **drop( )** function is used.

Syntax: Object **. drop** ([index or sequence of indexes] or [Row label])

Example:

Dataframe d1	# Code to delete column C3 from dataframe	Output
	df.drop(index=[2])	

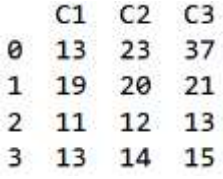
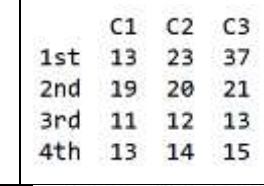
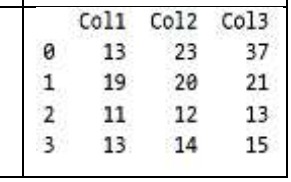
Note: **drop ( )** function can also used to delete a column by giving axis =1 with index/labels.

Renaming Rows/Columns :

To change the name of any row/column individually , **rename ( )** function is used.

Syntax:

Object **. rename** (index = { names dictionary }, columns = { names dictionary }, inplace = False)

Dataframe df	# Renaming row labels	Output
	df.**rename(**index={0:"1st",1:"2nd",2:"3rd", 3:"4th"},inplace=True**)** print(df)	
	# Renaming column labels df.**rename(**columns={0:"Col1",C1:"Col2",C2: "Col3", C3:"Col4"},inplace=True**)** print(df)	

Note:

- By default **rename ( )** function creates a new DataFrame and changes shows in the new DataFrame only, not in the original DataFrame.
- Hence to make changes in the original DataFrame , the argument **inplace = True** must be specify.

Head and Tail functions

head():

- ❖ This method returns the first few rows of a DataFrame.
- ❖ By default, it displays the first five rows, but you can specify the number of rows you want to see by passing an argument to the method.
- ❖ For example, df.head(10) would display the first ten rows of the DataFrame df.

tail() :

- ❖ This method returns the last few rows of a DataFrame.
- ❖ By default, it displays the last five rows, but you can specify the number of rows you want to see by passing an argument to the method.
- ❖ For example, df.tail(10) would display the first ten rows of the DataFrame df.

➢ **Indexing Using Labels, Boolean Indexing**

- ❖ Indexing using Labels: Indexing using labels refers to accessing or retrieving specific data from a DataFrame using the labels assigned to rows and columns.
- ❖ In pandas, you can use the .loc[] indexer to perform label-based indexing.
- ❖ The .loc[] indexer allows you to specify row labels and column labels to extract specific data from the DataFrame.
- ❖ For example, consider a DataFrame df with columns 'Name', 'Age', and 'Gender'. To access a specific row with the label 'A', you can use df.loc['A']. To access a specific cell in the 'Age' column for the row labeled 'A', you can use df.loc['A', 'Age'].

Example of indexing using labels

Dataframe df

	Name	Age	Gender
0	John	25	Male
1	Alice	18	Female
2	Bob	32	Male
3	Emma	21	Female
4	Charlie	27	Male

```
# Indexing using labels
print("Indexing using labels:")
print(df.loc[2])          # Access row with label 2
print(df.loc[1, 'Age'])   # Access cell at row label 1 and column 'Age'
```

Output

```
Indexing using labels:
Name      Bob
Age        32
Gender   Male
Name: 2, dtype: object
18
```

➢ Boolean Indexing

- Boolean indexing involves filtering data in a DataFrame based on a specific condition or set of conditions.
- You can use Boolean expressions to create a Boolean mask, which is a Series or DataFrame of Boolean values indicating which rows satisfy the given condition.
- For example, suppose you have a DataFrame `df` with a column 'Age' and you want to filter out all rows where the age is greater than 18. You can create a Boolean mask using the condition `df['Age'] > 18`.
- Then, you can apply this mask to the DataFrame using square brackets to retrieve only the rows where the condition is True: `df[df['Age'] > 18]`.
- This will return a new DataFrame containing only the rows where the condition is satisfied.

Example of Boolean indexing

Dataframe df

	Name	Age	Gender
0	John	25	Male
1	Alice	18	Female
2	Bob	32	Male
3	Emma	21	Female
4	Charlie	27	Male

```
# Boolean indexing
print("\nBoolean indexing:")
mask = df['Age'] > 20      # Create a boolean mask
print(df[mask])            # Apply the mask to filter rows with Age > 20
```

Output

```
Boolean indexing:
      Name  Age  Gender
0     John   25    Male
2      Bob   32    Male
3     Emma   21  Female
4  Charlie   27    Male
```

MCQ : Dataframes

1. Which library in Python is used for working with DataFrames?
   a) NumPy b) Pandas c) Matplotlib d) Scikit-learn
2. Which of the following is a valid method to create a DataFrame in Pandas?
   a) From a CSV file b) From dictionary of series
   b) c) From CSV File d) All of the above
3. How can you access the first five rows of a DataFrame?
   a) df.head(5) b) df.tail(5) c) df[0:5] d) df.loc[0:4]
4. Which method is used to check the dimensions of a DataFrame?
   a) df.info() b) df.shape() c) df.describe() d) df.columns()
5. Which method is used to check the summary statistics of a DataFrame?
   a) df.info() b) df.shape() c) df.describe() d) df.columns()
6. How can you select a specific column in a DataFrame?
   a) df['column_name'] b) df.column_name
   b) c) df.loc[:, 'column_name'] d) All of the above
7. How can you drop a column from a DataFrame?
   a) df.drop('column_name', axis=1) b) df.delete('column_name', axis=1)
   b) df.remove('column_name', axis=1) d) df.drop_column('column_name')
8. How can you filter rows in a DataFrame based on a condition?
   a) df[condition] b) df.filter(condition) c) df.loc[condition] d) df.select(condition)
9. How can you sort a DataFrame based on a specific column?
   a) df.sort_by('column_name') b) df.sort_values('column_name') c) df.sort('column_name') d) df.order_by('column_name')
10. How can you rename a column in a DataFrame?
    a) df.rename_column('old_name', 'new_name') b) df.rename('old_name', 'new_name')
    b) df.rename(columns={'old_name': 'new_name'}) d) df.column_name('old_name', 'new_name')
11. Which method is used to fill missing values in a DataFrame?
    a) df.fill_missing() b) df.fillna() c) df.fill_null() d) df.impute()
12. Pandas Dataframe is:
    a) 3 Dimensional b) 1 Dimensional c) 2 Dimensional d) Multidimensional
13. The axis 0 identifies a dataframe's_____.
    a) rows b) columns c) values d) datatype
14. What is the correct syntax to return the 5th row of a pandas DataFrame? Consider the name of the DataFrame is df:
    a) df[5] b) df.loc[5] c) df.loc[4] d) df.iloc[5]
15. The function to_csv used for converting a DataFrame to a CSV File is found in ____
    a) Pandas Module b) CSV Module c) Python Module d) None of these
16. When a DataFrame to a CSV File the attribute _____ is used to specify a separator other than comma(,)
    a) other b) sep c) separator d) None of these

17. Which attribute of dataframe is used to perform the transpose operation on a dataframe?
    a) T b) Ndim c) Empty d) Shape
18. What is a correct syntax to return the values of 1st row of a Pandas DataFrame dfRent.?
    a. dfRent[0] b. dfRent.loc[1] c. dfRent.loc[0] d. dfRent.iloc
19. To display the 3rd ,4th and 5th columns from the 6th to 9th rows of a dataframe DF is….
    a) DF.loc [6:9,3:5] b) DF.loc [6:10,3:6] c) DF.iloc [6:10,3:6] d) DF.iloc [6:9,3:5]
20. If a dataframe is created using a 2D dictionary, then the column labels are formed from _
    (a) dictionary's values (b) inner dictionary's keys (c) outer dictionary's key (d) none of these
21. Which among the following options can be used to create a DataFrame in Pandas?
    (a) Scalar values (b) An ndarray (c) A python dict (d) All of these

Competency based

2 Mark Questions: DataFrame

1. Consider the following DataFrame 'df' and answer following questions:

	rollno	name	UT1	UT2	UT3	UT4
0	1	Prerna Singh	24	24	20	22
1	2	Manish Arora	18	17	19	22
2	3	Tanish Goel	20	22	18	24
3	4	Falguni Jain	22	20	24	20
4	5	Kanika	15	20	18	22
5	6	Ramandeep Kaur	20	15	22	24

(a) The teacher needs to know the roll no, name and marks scored by the student with roll number 4. Help her to write the code.

(b) Ms. Sharma, the class teacher wants to add a new column, the 'Grade' with the values, ' A', 'B', 'A', 'A', 'B', 'A' ,to the DataFrame. Help her to write code.

2.Consider the following DataFrame, named 'Stud'

	Name	Rollno	English	Hindi	Maths	Ssc	Science
0	Aman	101	65	54	87	69	74
1	Suman	102	69	64	90	87	59
2	Priya	103	75	72	98	90	75
3	Tahir	104	88	80	78	45	87
4	Bharti	105	45	53	81	69	98

Write commands to:

i. Display the Name and Rollno.

ii. Add a new row with values (Rakesh ,106, 79 , 86, 91, 77, 93)

3. Consider the Dataframe DF as shown below.NOTE:Libraries has been imported.

	A	B	C
0	10	20	30
1	40	50	60

What will be the output of the following code?

a) print("I :",DF.iloc([0][0]) b) print("II :",DF.loc([0]['C'])

4. Write the correct output on execution of the following Pandas code

```
import pandas as pd
df = pd.DataFrame({"A":["p01", "p02", "p03"], "B" : ["Pen", "Pencil", "Eraser"]})
df = df.rename(columns = {"A":"PID", "B":"PNAME"})
df = df.rename(index = {0:"A", 1:"B", 2:"C"})
print(df)
```

5. Name the functions you can use to iterate over Dataframes.

6.Find the output of the following commands based on DataFrame TOY given below:

	TNAME	MATERIAL	PRICE	AGE
T1	TRAIN	STEEL	500	15
T2	DOLL	COTTON	100	6
T3	PUPPY	FUR	200	8
T4	SOLDIER	CLAY	50	12
T5	AEROPLANE	PLASTIC	150	10

a) TOY.loc[['T2','T4'],['MATERIAL','AGE']] b) TOY.iloc[1:3,1:3]

7. Answer the following questions based on DataFrame MRK given below:

	COMPANY	CITY
C1	MICROSOFT	SAN FRANCISCO
C2	MORGAN STANLEY	NEW YORK,
C3	COCA-COLA	ATLANTA
C4	INFOSYS	BANGALORE

a) What is the output of MRK.T command?

b) What is the output of the following commands based on the DataFrame (MRK)

i) MRK.size ii) MRK.shape

8.Write a python code to create a dataframe with headings (a and b) from the list given below:

[[10,30],[15,20],[25,35]

9.Write a python code to create the following dataframe Books by any method of Dataframe creation and Give index as 'B1','B2','B3','B4'

BookName	Class	Price
Let us c	BCA	270
A. I.	B.Tech	350
DBMS	MCA	450
C. A.	BCA	250

3 Mark Questions: **DataFrame**

1.Consider a DataFrame (MRK) given below:

	NAME	**PHY**	**CHEM**
S1	**ATUL**	**51**	**35**
S2	**ANIL**	**62**	**61**
S3	**ANUJ**	**45**	**46**

(A) Write commands to create the DataFrame (MRK) using a Dictionary of Lists

(B) If a DataFrame is created using a 2D dictionary, then the column labels are formed from the inner dictionary's keys or outer dictionary's keys?

2. Assume the following Data Frame D1 and D2 given below:

	X	**Z**
P3	**877**	**966**
P4	**544**	**366**
P5	**244**	**553**

	X	**Y**	**Z**
P0	**45**	**87**	**98**
P1	**24**	**53**	**86**
P2	**85**	**92**	**47**

a) Write a command append the second to the first

b) Add a new Column between Y and Z in DataFrame D1 with any dummy values using insert command

c) Rename the index of D2 : “P3” as PT2, “P4” as PT4 and “P5” as “PT5”

3. A dictionary ‘Toys’ contains the following:

Toys = {‘Name’:[‘Doll’, ‘Ludo’, ‘Chess’ , ‘Blocks’], ‘Price’:[400,250,300,150]}

Write statements for the following:

i) Create a dataframe named “Stock” using dictionary ‘Toys’.

ii) Add a column called ‘Discount’ with the following data. [30,40,15,25]

iii) Delete column discount with all values

4. Write a python code for the following Dataframe Library using python pandas. Give index as ‘B1’, ‘B2’,’B3’,’B4’

ItemNo	ItemName	Price
P99	SUGAR	100
P10	TEA	150
P50	COFFEE	200
P60	GREEN TEA	250

i) Display Item Number and name whose price is less than 150.

ii) Display details of different types of Tea available in shop.

iii) Display the dataframe.

5. Consider the following DataFrame ‘df’ and answer questions from (i)-(iii):

ID	Name	Age	Fav_Colo	Points
T01	Rahul	32	Blue	73
T02	Mohak	25	Green	82
T03	Rajeev	45	Orange	29
T04	Rohini	30	Pink	39

I. Write down the command that will add a column “eligible" with default value as ‘yes’.

II. Write the command to extract the complete row ‘T03’

III. Which command will be used to drop a row from dataframe ‘df’ labeled as ‘T04’?

4 Mark Questions:**DataFrame**

1. Write a Python code to create the following Dataframe Empdf from a Dictionary:

	EName	Salary
0	Kush	10000
1	Ruchika	12000
2	Divya	20000
3	John	25000

a) Write python code to rename the index to 'A','B','C','D' in place of 0,1,2,3.

b) Write python code to add a column 'Commission' with values as 5% of the Salary.

2. Assume a dataframe DF That contains data about climate condition of various cities with C1,C2,C3,C4,C5 as index and give the output for the questions given below.

	City	Maxtemp	Mintemp	Rainfall
C1	Delhi	40	32	24.1
C2	Bengaluru	31	25	36.2
C3	Chennai	35	27	40.8
C4	Mumbai	29	21	35.2
C5	Kolkata	39	23	41.8

i) DF.shape ii) DF[1:2] iii) DF.loc['C1': 'C3', 'City'] iv)DF.City

3. Using the below DataFrame mdf answer the questions a to d.

	C1	C2	C3
0	13	23	37
1	19	20	21
2	11	12	13
3	13	14	15

a) Write code to create a new Dataframe n1 that stores the values of the Dataframe mdf multiplied by 3.

b) Write code to drop the index 2 from the above Dataframe. The Dataframe should be modified after this statement.

c) Write the code to display the sum of rows with indexes 2 onwards from the Dataframe mdf.

d)Write code to add a column C4 in the dataframe which stores the differences of column C3 with column C2.

5 Mark Questions: DataFrame

1. Give the output of the command that follow based on the DataFrame DF below:

AUTHOR	PRICE	QTY
WILLIAM	1500	42
AGATHA	1300	85
LEO	600	48
ALEXANDER	1200	17
JIN	800	15
CAO	1700	20

a) >>> DF['PRICE']>=1000 b) >>> DF[DF['QTY']<=50] c) >>> DF['QTY']%5

d) >>> DF.iloc[[2,3],[1,2]] e)>>>DF.QTY[2]=300

2. Atul has stored some data in a DataFrame "MDF" as shown below. He wants to shift the data to CSV file. Write commands for questions that follow:

YEAR	**SALES**	**LOSS**
2001	**85**	**42**
2002	**75**	**15**
2003	**64**	**62**
2004	**93**	**78**
2005	**67**	**52**

a) What is the full form of CSV

b) A CSV file can take ; character as separator

A) True B) False

c) To convert the above DataFrame "MDF" to a CSV file "TEXT.CSV"

d) To read and display the CSV file "TEXT.CSV" thus formed without header

e) To delete the column LOSS

3. Write code for the following questions on the basis of following DataFrame 'Sales'.

INDEX

	2014	2015	2016	2017
Madhu	100.5	12000	20000	50000
Kusum	150.8	18000	50000	60000
Kinshu	200.9	22000	70000	70000

Ankit	30000	30000	100000	80000
Shruti	40000	45000	125000	90000

a)Display the transpose of DataFrame Sales.

b) Display the sales made by Madhu and Ankit in the year 2016 and 2017.

c) Delete the data for sales man Kinshuk from the DataFrame Sales.

d) Update the sale made by Shruti in 2017 to 100000.

e) Rename the index as ‘A’,’B’,’C’,’D’,’E’

Answer key

Dataframe MCQ

1. b) Pandas

2. d) All of the above

3. a) df.head(5)

4. b) df.shape()

5. c) df.describe()

6. d) All of the above

7. a) df.drop('column_name', axis=1)

8. a) df[condition]

9. b) df.sort_values('column_name')

10. c) df.rename(columns={'old_name': 'new_name'})

11. b) df.fillna()

12.c 13.a 14.c 15.b 16.b 17.a 18.a 19.c 20.b 21.d

Competency based

Answers of 2 Mark Questions of Dataframe

1.a)df.loc[3]

b)df[‘Grade’]=[‘A’,’B’,’A’,’A’,’B’,’A’]

2. i)Stud[[‘Name’,’Rollno’]]

ii)Stud.loc[5]= [‘Rakesh’,106,79,86,91,77,93]

3. a)I:10 b) II:30

4.

	PID	PNAME
A	P01	Pen
B	P02	Pencil

5. Iterrows( ) , iteritems ( )

C P03 Eraser

6. a)

	MATERIAL	AGE
T2	COTTON	6
T4	CLAY	12

b)

	MATERIAL	PRICE
T2	COTTON	100
T3	FUR	200

7. a)

MRK.T

	C1	C2	C3	C4
COMPANY	MICROSOFT	MORGAN STANLEY	COCA-COLA	INFOSYS
CITY	SAN FRANCISCO	NEW YORK	ATLANTA	BANGALORE

b) i) MRK.size 8 ii) MRK.shape (4, 2)

8. import pandas as p
import numpy as np
A=np.array([[10,20],[30,40],[50,60]])
df=p.DataFrame(A,columns=['a','b'])
print(df)

9. B={'Book_Name':['Let us C',A.I.','DBMS' , 'C.A.']
'Class' : ['BCA', 'B. Tech', 'MCA' , 'BCA']
'Price' : [270,350,450,250] }
Books= pd.DataFrame(B)
print(Books)

Answers of 3 Mark Questions:DataFrame

1.a)

```
MRK=pd.DataFrame({'NAME':['ATUL','ANIL','ANUJ'],'PHY':[51,62,45],'CHEM':[35,61,46]},
                 index=['S1','S2','S3'])
```

b) outer dictionary's keys

2. a) D1.append(D2) b) D1.insert(2,'A',[100,200,300])

	X	Y	Z
P0	45	87.0	98
P1	24	53.0	86
P2	85	92.0	47
P3	877	NaN	966
P4	544	NaN	366
P5	244	NaN	553

c) D2.rename({'P3':'PT2','P4':'PT4','P5':'PT5'},axis='index'}

3. .i) stock = pd.DataFrame(Toys)

4.i) print(Library["Price"]<150])

ii) stock [“discount”] = [30,40,15,25] ii) print(Library.Itemname)

iii) stock.drop(“discount”,axis=1) iii) print(Library)

5. I. df['eligible'] = 'yes' II)Ans: df[df.ID == 'T03'] III. df.drop(3, axis = 0, inplace = True)

4 Mark Questions: DataFrame

1.D = {'EName': ['Kush','Ruchika','Divya','John'],'Salary': [10000,12000,20000,25000]}
Empdf = pd.DataFrame(D)
a) Empdf.rename(index = {0:'A',1:'B',2:'C',3:'D'}, inplace = True)
b) Empdf['Commission'] = Empdf['Salary']*.05

2. i) (5,4) ii) C2 Bengaluru 31 25 36.2 iii) C1 Delhi
C2 Bengaluru
C3 Chennai

iv) 0 Delhi
1 Bengaluru
2 Chennai
3 Mumbai
4 Kolkata

3. a) n1 = mdf * 3 b) mdf = mdf.drop(2) c)mdf.iloc[2:].sum()
d) mdf[‘C4’] = mdf[‘C3’] – mdf[‘C2’]

Answer of 5 Mark Questions

1. a) C1 True
C2 True
C3 False
C4 True
C5 False
C6 True

b)

	AUTHOR	PRICE	QTY
C1	WILLIAM	1500	42
C3	LEO	600	48
C4	ALEXANDER	1200	17
C5	JIN	800	15
C6	CAO	1700	20

c) C1 2
C2 0
C3 3
C4 2
C5 0
C6 0

d)

	PRICE	QTY
C3	600	48
C4	1200	17

e)

AUTHOR	PRICE	QTY
LEO TOLSTOY	600	300

2. a) Comma Separated Values b) True c) MDF.to_csv(“TEXT.CSV”)

d) import pandas as pd e) del MDF[‘LOSS’]

pd.read_csv(‘TEXT.CSV”,header=None)

3. a)Sales.T b) Sales.loc[['Madhu','Ankit'],[2016,2017]]

c) Sales.drop(['Kinshuk'], axis = 0, inplace = True)

d) Sales.loc['Shruti',2017] = 100000

e)Sales.rename(index={‘MADHU’:’A’,‘KUSUM’:’B’,‘Kinshu’:’C’,‘ANKIT’:’D’, ‘SHRUTI’:’E’})

Importing/Exporting data between CSV files and Data Frames

CSV FILES

- CSV stands for Comma-separated values.
- This type of file can be opened with excel file or notepad.
- CSV file is nothing more than a simple text file. However, it is the most common,simple and easiest method to store tabular data.
- This particular format arranges tables by a specific structure divided into rows and columns.
- To import data from a CSV file into a DataFrame, you can use the read_csv() function.
- This function takes the path to the CSV file as its input, and returns a DataFrame object.
- For example, the following code imports the data from the data.csv file into a DataFrame:

```
import pandas as pd
df = pd.read_csv('data.csv')
```

- Parameter of read_csv() function
  - **sep**: str, default ‘,’ - Delimiter to use as column separator
  - **header :** int, list of int, None, default ‘infer’ - Row number(s) to use as the column names, and the start of the data. Default behavior is to infer the column names.
  - **index_col** : int, str, sequence of int / str, or False, optional, default None - Specifies the column that should be used as the index of the DataFrame
  - **skiprows**: list-like, int or callable, optional - Line numbers to skip (0-indexed) or number of lines to skip (int) at the start of the file.
  - **nrows**: int, optional -Number of rows of file to read.

- To export a DataFrame to a CSV file, you can use the to_csv() function.

```
import pandas as pd
df= {'Name': ['Pankaj', 'Meghna'], 'ID': [1, 2], 'Role': ['CEO', 'CTO']}
df.to_csv(‘employee.csv’)
```

- Parameters of to_csv() function:
  - **sep:** str, default ‘,’ - String of length 1. Field delimiter for the output file.
  - **columns**: sequence, optional - Columns to write.
  - **header**: bool or list of str, default True - Write out the column names. If a list of strings is given it is assumed to be aliases for the column names.
  - **index**: bool, default True - Write row names (index).

- This function takes the path to the CSV file as its output, and writes the DataFrame to the file. For example, the following code exports the df DataFrame to the output.csv file:

```
df.to_csv('output.csv')
```

MCQ

1. Which of the following functions is used to import data from a CSV file into a DataFrame?
   a) read_csv() b) to_csv()
   c) write_csv() d) import_csv()
   Answer: read_csv().
2. Which of the following functions is used to export a DataFrame to a CSV file?
   a) read_csv() b) to_csv()
   c) write_csv()) import_csv()
   Answer: to_csv().
3. Which of the following parameters is used to specify whether the first row of the CSV file contains column names?
   a) header b) sep c) index_col d) na_values
   Answer: header.
4. Which of the following parameters is used to specify the separator character used in the CSV file?
   a) header b) sep c) index_col d) na_values
   Answer: sep.
5. Which of the following parameters is used to specify the column that should be used as the index of the DataFrame?
   a) header b) sep c) index_col d) na_values
   Answer: index_col.

Competency based

1. You have a CSV file with data about students. The file has the following columns:
   Name, age, gender, grade
   Write a Python program to import the data from the CSV file into a DataFrame.
2. You have a DataFrame with data about students. You want to export the DataFrame to a CSV file. The file should have the same columns as the original CSV file.
   Write a Python program to export the DataFrame to a CSV file.
3. You have a CSV file with data about students. The file has the first row as the column names.
   Write a Python program to import the data from the CSV file into a DataFrame, ignoring the first row.
4. You have a CSV file with data about students. The file uses a comma as the separator character.
   Write a Python program to import the data from the CSV file into a DataFrame, using a tab as the separator character.
5. You have a CSV file with data about students. The file has some missing values.
   Write a Python program to import the data from the CSV file into a DataFrame, and replace the missing values with the value -1.

Assertion reasoning questions

1. **1. Assertion:** The read_csv() function can be used to import data from a CSV file into a DataFrame.
   Reason: The read_csv() function is a built-in function in the pandas library that can be used to import data from a CSV file into a DataFrame.
2. **2. Assertion:** The to_csv() function can be used to export a DataFrame to a CSV file.
   Reason: The to_csv() function is a built-in function in the pandas library that can be used to export a DataFrame to a CSV file.
3. **3. Assertion:** The header parameter in the read_csv() function specifies whether the first row of the CSV file contains column names.
   Reason: The header parameter in the read_csv() function specifies whether the first row of the CSV file contains column names. If the header parameter is set to True, then the first row of the CSV file will be used as the column names. If the header parameter is set to False, then the first row of the CSV file will be ignored.
4. **4. Assertion:** The sep parameter in the read_csv() function specifies the separator character used in the CSV file.
   Reason: The sep parameter in the read_csv() function specifies the separator character used in the CSV file. The default separator character is a comma (,). However, you can specify a different separator character by setting the sep parameter to the desired character.
5. **5. Assertion:** The na_values parameter in the read_csv() function specifies the values that should be interpreted as missing values.
   Reason: The na_values parameter in the read_csv() function specifies the values that should be interpreted as missing values. The default missing values are NaN and NA. However, you can specify a different set of missing values by setting the na_values parameter to a list of the desired values.

Answer key of Competency based questions –

1	import pandas as pd df = pd.read_csv('students.csv') print(df)
2	import pandas as pd df = pd.read_csv('students.csv') df.to_csv('students_out.csv')
3	import pandas as pd df = pd.read_csv('students.csv', header=None) print(df)
4.	import pandas as pd df = pd.read_csv('students.csv', sep='\t')

	print(df)
5.	import pandas as pd df = pd.read_csv('students.csv', na_values='-1') print(df)

Answers of assertion reasoning questions

1. **Answer:** The assertion is true because the read_csv() function can be used to import data from a CSV file into a DataFrame.
2. **Answer:** The assertion is true because the to_csv() function can be used to export a DataFrame to a CSV file.
3. **Answer:** The assertion is true because the header parameter in the read_csv() function specifies whether the first row of the CSV file contains column names.
4. **Answer:** The assertion is true because the sep parameter in the read_csv() function specifies the separator character used in the CSV file.
5. **Answer:** The assertion is true because the na_values parameter in the read_csv() function specifies the values that should be interpreted as missing values.

What is Data Visualization?

- Data visualization is the technique to present the data in a pictorial or graphical format. It enables stakeholders and decision makers to analyze data visually. The data in a graphical format allows them to identify new trends and patterns easily.

- The main benefits of data visualization are as follows:

  - It simplifies the complex quantitative information
  - It helps analyze and explore big data easily
  - It identifies the areas that need attention or improvement
  - It identifies the relationship between data points and variables
  - It explores new patterns and reveals hidden patterns in the data

- **matplotlib Library and pyplot Interface**
  - The matplotlib is a python library that provides many interfaces functionally for 2D graphics
  - In short, we can call mattplotlib as a high-quality plotting library of Python.
  - The matplotlib library offers many different named collections of methods, pyplot is one such interface.
  - pyplot is a collection of methods within matplotlib which allows user to construct 2D plots easily and interactively.
- Installing matplotlib

  It is done using pip command in Command Prompt: **pip install matplotlib**

- **Importing PyPlot**

  Syntax: import matplotlib.pyplot
  or
  import matplotlib.pyplot as plt

  importing matplotlib in the form of plt we can use plt for accessing any function of matplotlib

- **Steps to plot in matplotlib:**
  - Create a .py file & import matplotlib library to it using import statement

    import matplotlib.pyplot as plt
  - Set data points in plot( ) method of plt object
  - Customize plot by setting different parameters
  - Call the show() method to display the plot

- Save the plot/graph if required using function plt.savefig(file_path)

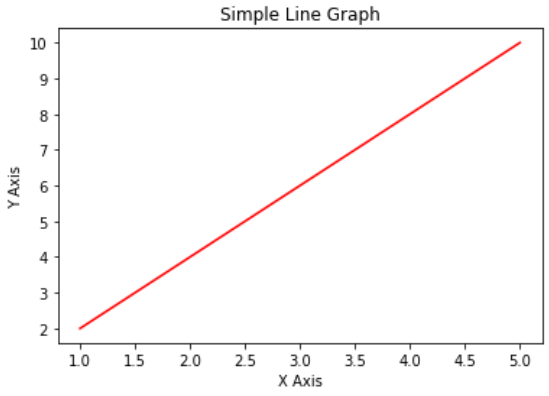
- Types of plot using matplotlib
  - LINE PLOT
  - BAR GRAPH
  - HISTOGRAM etc.

- Line Plot:
  - A line plot/chart is a graph that shows the frequency of data occurring along a number line.
  - The line plot is represented by a series of data points called markers connected with a straight line.
  - Generally, line plots are used to display trends over time. A line plot or line graph can be created using the **plot ()** function available in pyplot library.
  - We can, not only just plot a line but we can explicitly define the grid, the x and y axis scale and labels, title and display options etc.
  - Line chart: displaying data in form of lines.
  - We can create line graph with x coordinate only or with x and y coordinates.
  - Function to draw line chart – plot()
  - Default colour of line- blue
  - The default width for each bar is .0.8 units, which can be changed.

- Syntax: plt.plot(x,y)

- **Line Plot customization**
  - Custom line color
    - plt.plot(x,y,'red')
    - Change the value in color argument like 'b' for blue,'r','c',…..
  - Custom line style and line width
    - plt.plot(x,y, linestyle='solid' , linewidth=4).
  - set linestyle to solid/dashed/dotted/dashdot set linewidth as required
  - Title plt.title('DAY – TEMP Graph ') – Change it as per requirement
  - Label:
    - plt.xlabel('Time') – to set the x axis label
    - plt.ylabel('Temp') – to set the y axis label
  - Changing Marker Type, Size and Color
    - plt.plot(x,y,'blue',marker='*',markersize=10,markeredgecolor='magenta')
  - Order of methods used in plot() function:
    - Plt.plot(x,y,color,linewidth,linestyle,marker, markersize,markeredgecolor)
  - Function used to show the graph – show()

Example

import matplotlib.pyplot as plt X=[1,2,3,4,5] Y=[2,4,6,8,10] plt.title('Simple Line Graph') plt.xlabel('X Axis') plt.ylabel('Y Axis') plt.plot(X,Y,'r') plt.show()	Output 

- Bar Graph
  - A graph drawn using rectangular bars to show how large each value is.
  - The bars can be horizontal or vertical.
  - A bar graph makes it easy to compare data between different groups at a glance.
  - Bar graph represents categories on one axis and a discrete value in the other.
  - The goal of bar graph is to show the relationship between the two axes.
  - Bar graph can also show big changes in data over time.
  - Syntax : plt.bar(x,y) Bar graph customization
- Custom bar color
  - To set same color to all bars

    plt.bar(x,y, color="color code/color name")

  - To set different colors for different bars

    plt.bar(x,y, color="color code/color name sequence")

- Custom bar width
  - To set same color to all bars

    plt.bar(x,y, width=float value)
  - To set different widths for different bars

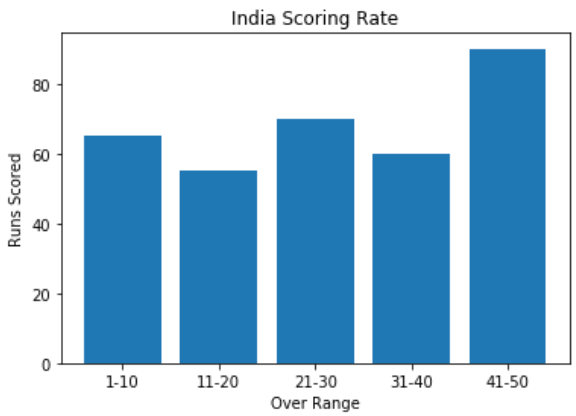
    plt.bar(x,y, width=float value sequence)
- Title

  plt.title(' Bar Graph ') – Change it as per requirement

- Label

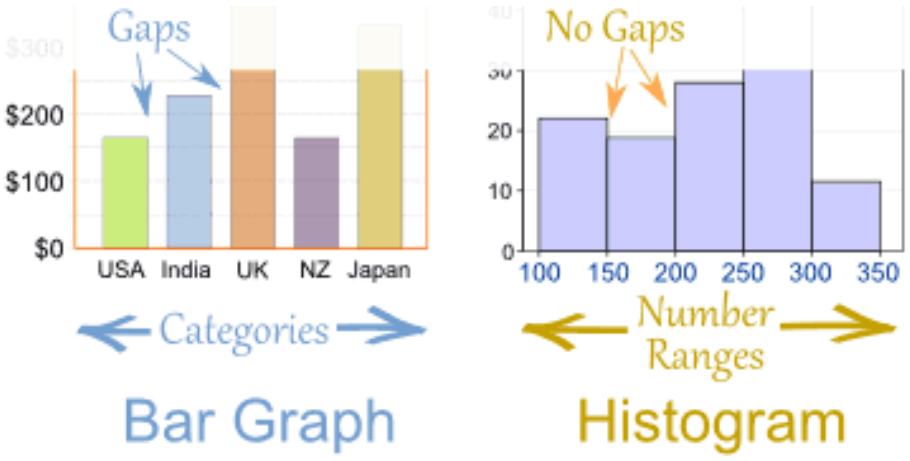
  plt.xlabel('Overs') – to set the x axis label

  plt.ylabel('Runs') – to set the y axis label

# Code to plot a bar graph import matplotlib.pyplot as plt overs=['1-10','11-20','21-30','31-40','41-50'] runs=[65,55,70,60,90] plt.xlabel('Over Range') plt.ylabel('Runs Scored') plt.title('India Scoring Rate') plt.bar(overs,runs) plt.show( )	Output 

- **Histogram**
  - A histogram is a graphical representation which organizes a group of data points into user specified ranges.
  - Histogram provides a visual interpretation of numerical data by showing the number of data points that fall within a specified range of values ("bins").
  - It is similar to a vertical bar graph but without gaps between the bars.
- **Difference between a histogram and a bar chart / graph** –

A bar chart majorly represents categorical data (data that has some labels associated with it), they are usually represented using rectangular bars with lengths proportional to the values that they represent. While histograms on the other hand, is used to describe distributions.



- **<u>Creating a Histogram :</u>**
  - It is a type of bar plot where X-axis represents the bin ranges while Y-axis gives information about frequency.

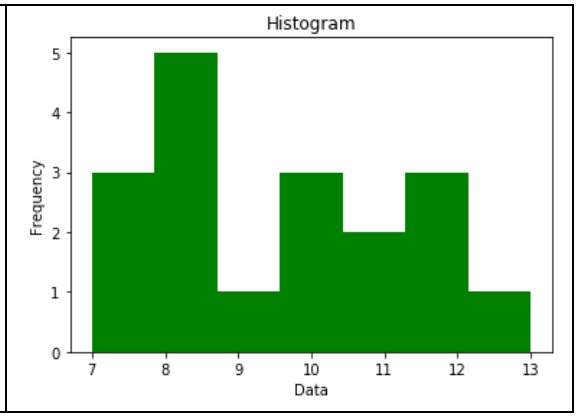
- To create a histogram the first step is to create bin of the ranges, then distribute the whole range of the values into a series of intervals, and count the values which fall into each of the intervals.
  - Bins are clearly identified as consecutive, non-overlapping intervals of variables.
  - The hist() function is used to create histogram
  - Syntax: **plt.hist(x,other parameters)**

➢ **Optional Parameters**

x	array or sequence of array
bins	optional parameter contains integer or sequence or strings default =5
histtype	optional parameter used to create type of histogram [bar, barstacked, step, stepfilled], default is "bar"
align	optional parameter controls the plotting of histogram [left, right, mid]
orientation	Optional. Possible values are 'horizontal' or 'vertical'
color	optional parameter used to set color or sequence of color specs

Program to plot histogram

```
import matplotlib.pyplot as plt
data=[7,7,7,8,8,8,8,8,9,10,10,10,11,11,12,12,12,13]
plt.xlabel('Data')
plt.ylabel('Frequency')
plt.title('Histogram')
plt.hist(data,bins=7,color='green')
plt.show()
```



➢ Title (title)
  - plt.title('Histogram ') – Change it as per requirement • Labelplt.xlabel('Data') – to set the x axis label plt.ylabel('Frequency') – to set the y axis label

➢ Legend –
  - A legend is an area describing the elements of the graph.
  - In the matplotlib library there is a function named legend() which is used to place a legend on the axes .
  - When we plot multiple ranges in a single plot, it becomes necessary that legends are specified. It is a color or mark linked to a specific data range plotted .
  - To plot a legend you need to do two things.
    - In the plotting function like bar() or plot() , give a specific label to the data range using label
    - Add legend to the plot using legend ( ) as per the sytax given below .

Syntax : - plt.legend((loc=position number or string)

- ❖ position number can be 1,2,3,4 specifying the position strings upper right/'upper left/'lower left/lower right respectively.
- ❖ Default position is upper right or 1

➢ **Saving the Plot**

- To save any plot savefig() method is used.
- Plots can be saved in various formats like pdf,png,eps etc .

```
plt.savefig('line_plot.pdf') // save plot in the current directory
plt.savefig('d:\\plot\\line_plot.pdf') // save plot in the given path
```

Case Study Based MCQs

1. Consider the following program and answer any the questions from (i) to (v):

```
1. import _______ as plt
2. plt.bar ([2,3,4,8,1],[2,4,7,3,5], label= _____ )
3. plt.legend( )
4. plt.xlabel(____ )
5. plt.ylabel('Height')
6. plt._______ ('Vertical Bar Chart')
7. _________
```

(i) Which Module will be imported in Line 1 for above code?

(a)matplotlib (b) matplotlib.pyplot
c)plotlib (d) None of these

Ans: (b)

(ii) Name the label that can be used to represent the bar chart in Line 2.

(a) Data (b) Data Values
(b) Values for X axis (d) All of these

Ans: (d)

(iii) Which message is best suited for xlabel?

(a) X values (b) Y values
(c) Legend (d) Vertical

Ans: (a)

(iv) Which method will take place at Line 6 for setting heading on the top of Chart?

(a)Title () (b) title()
(c) Head() (d) All of these.

Ans: (b)

(v) Choose the statement to be place at Line7 of the above code.

(a)Plt.print() (b) plt.show()
(c) Plt.display() (d) plot.show()

Ans: (b)

2. If you are given to plot a histogram using numpy array as per the code given below then answer any of four question from (i) to (v)

1. from matplotlib import _____ as plt #line 1
2. import numpy as np
3. fig, ax = plt._____ (1, 1)
4. a= np.array([26,59,44,39,76,16,23,11,18,78])
5. ax.hist(a, bins=[0,10,20,30,40])
6. ax._____ ('Histogram')
7. ax.set_xticks ([0,10,20,30,40,])
8. ax.set_xlabel('Percentage')
9. ax._____ ('Students')
10. _____

(i) Choose the correct option to import for given program:

(a) matplotlib (b) matplot
c) pyplot (d) plot

Ans: (c)

(ii) Fill in the blank at Line 3

(a) subplots (b) subplot
(c) plot (d) subplt

Ans: (a)

(iii) Which statement is used to set title in Line 6?

(a) title (b) set_title
(c) set (d) Title

Ans: (b) set_title

(iv) How to set Y-axis label?

(a) label () (b) set_y
(c) set_ylab () (d) set_ylabel

Ans: (d) set_ylabel

(v) To fill in blank on Line 10 for showing histogram what can bee used?

(a) plt.show() (b) plt_show()
(c) plot_show() (d) plt.show

Ans: plt.show()

3. Consider the following case and answer the from (i) to (v)

1. import ……. as pd
2. import matplotlib. _____ as plt
3. data= {'Name' : ['Karan', 'Adi', 'Abhinav', 'Kirti', 'Rahul'], 'Height' : [60,61,63,65,61], 'Weight' : [47,89,52,58,50,47]}
4. df=pd. _____ (data)
5. df._____ (Kind ='hist', edgecolor = 'Green', linewidth =2, linestyle=':' , fill= False)
6. _____

(i)Fill in the blank in Line 1.

(a) numpy (b) pandas
(c) Python (d) matplot

Ans: (b) pandas

(ii) Fill in the blank in Line 2.

(a) pyplot (b) plot
(c) pyp (d) None of these

Ans: (a) pyplot

(iii) Which of the following is used in Line 6 to represent the data?

(a) Series (b) Matplot
(c) DataFrame (d) Plot

Ans: (c) DataFrame

(iv) For blank of Line 7 command used may be:

(a) plt (b) pyplot
(c) plot (d) figure

Ans: (c)plot

(v) To show the above graph which statement is used in Line 8?

(a) plot.show (b) plot.show()
(c) plt.show (d) plt.show()

Ans: (d) plt.show()

1. What is true about Data Visualization?

A. Data Visualization is used to communicate information clearly and efficiently to users by the usage of information graphics such as tables and charts.
B. Data Visualization helps users in analyzing a large amount of data in a simpler way.
C. Data Visualization makes complex data more accessible, understandable, and usable.
D. All of the above

Ans: D

Explanation: Data Visualization is used to communicate information clearly and efficiently to users by the usage of information graphics such as tables and charts. It helps users in analyzing a large amount of data in a simpler way. It makes complex data more accessible, understandable, and usable.

2. Which method is used to save the output of pyplot in the form of image file?

A. savefig('filename')
B. save_fig('filename)
C. save_figure('filename')
D. save_img('filename')

Answer: A

3. Data visualization is also an element of the broader _____.

A. deliver presentation architecture
B. data presentation architecture
C. dataset presentation architecture
D. data process architecture

Answer: B

4. The command used to give a heading to a graph is _____

A. plt.show()
B. plt.plot()
C. plt.xlabel()
D. plt.title()

Answer: D

5. Common use cases for data visualization include?

A. Politics
B. Sales and marketing
C. Healthcare
D. All of the above

Answer: D

6. Which are pros of data visualization?

A. It can be accessed quickly by a wider audience.
B. It can misrepresent information
C. It can be distracting
D. None Of the above

Answer: A

Explanation: Pros of data visualization: it can be accessed quickly by a wider audience.

7. Which are cons of data visualization?

A. It conveys a lot of information in a small space.
B. It makes your report more visually appealing.
C. visual data is distorted or excessively used.
D. None Of the above

Answer: C

Explanation: It can be distracting: if the visual data is distorted or excessively used.

8. Which one of the following is most basic and commonly used techniques?

A. Line charts
B. Scatter plots
C. Population pyramids
D. Area charts

Answer: A

Explanation: Line charts. This is one of the most basic and common techniques used. Line charts display how variables can change over time.

9. Which is a python package used for 2D graphics?

A. matplotlib.pyplot
B. matplotlib.pip
C. matplotlib.numpy
D. matplotlib.plt

Answer: A

10. Observe the output figure. Identify the coding for obtaining this output.

A.
```
import matplotlib.pyplot as plt
plt.plot([1,2,3],[4,5,1])
plt.show()
```
B.
```
import matplotlib.pyplot as plt
plt.plot([1,2],[4,5])
plt.show()
```
C.
```
import matplotlib.pyplot as plt
plt.plot([2,3],[5,1])
plt.show()
```
D.
```
import matplotlib.pyplot as plt
plt.plot([1,3],[4,1])
plt.show()
```

Answer: A

12. Identify the right type of chart using the following hints.

Hint 1: This chart is often used to visualize a trend in data over intervals of time.

Hint 2: The line in this type of chart is often drawn chronologically.

A. Line chart
B. Bar chart
C. Pie chart
D. Scatter plot

Answer: A

13. Which of the following method will be used to display plot?

A. show()
B. display()
C. plot()
D. execute()

Answer: show()

14. Which method is used to plot horizontal bar graph in pyplot?

A. horizontal_bar()
B. barh()
C. hbar()
D. bar()

Answer: C

15. Which method is used to plot histogram in pyplot?

A. his()
B. hist()
C. Hist()
D. histogram()

Answer: B

Two Marks Questions.

1. Name the library and interface used to plot a chart in python.
2. What are the ways of importing matplotlib?
3. What are the basic elements/components of the chart?
4. What types of graphs can be plotted using matplotlib?
5. 5 How to change the thickness of line, line style, line color, and marker properties of a chart?
6. Mr. Sharma is a working in a game development industry. He saw a table of rating of various games available on the play store.

GAME	RATING
Subway Surfer	4
Temple Run	5
Candy Crush	2
Bubble Shot	5
Runner Best	8

   a. Draw a bar graph showing the above information.
   b. Mark the x axis as “Games” and y axis as “Rating”.

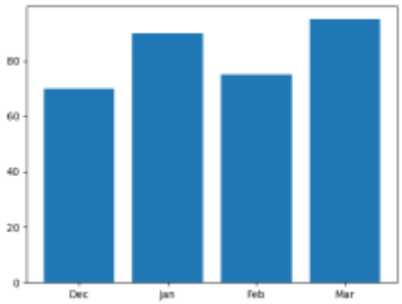
7. Write a program to create a single bar chart from two data sequences as given below:

   Means=[20,35,30,35,27]

   Stds=[2,3,4,1,2]

   Make sure to show legends.

8. Consider the plot shown below. Write the code to plot it.

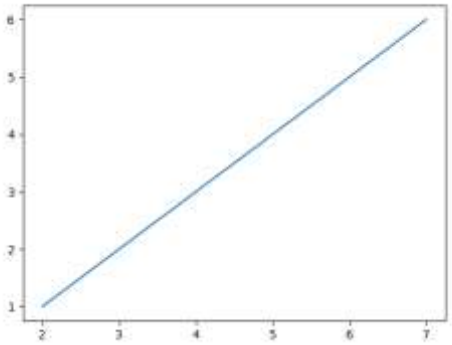


9. Write a program to create a horizontal bar chart for India's medal tally.

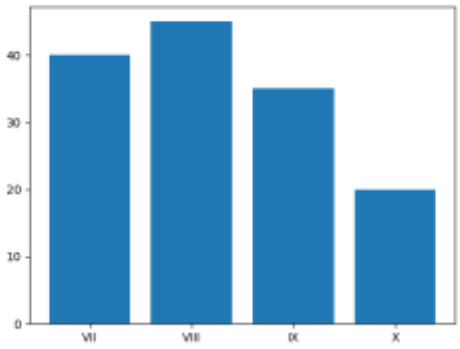
Gold	Silver	Bronze	Total
3	6	17	26

Three Marks Questions.

1. Consider the following line graph. Write the code to plot it.



2. Consider the following bar graph. Write the code to plot it.



3. Given an array Rev storing Revenue of a small business in past some years:

Rev= np.array([12, 15, 13, 15, 16, 17, 14, 19, 19, 17, 19, 16, 17, 16, 18, 18, 18, 19, 19, 18, 19, 21, 20, 18, 19, 19, 21, 19, 19, 22])

(a) Write code to create a histogram as per following specifications:

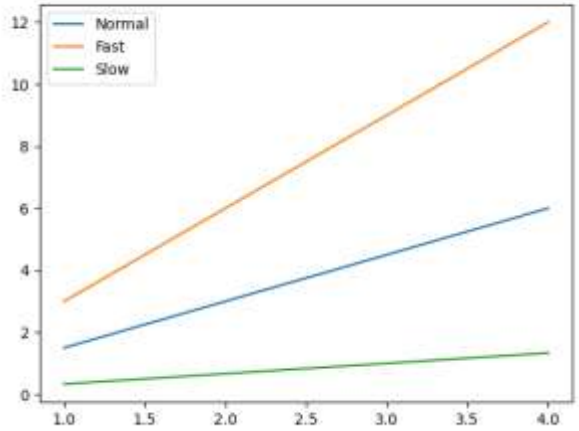
- Number of bins should be 10.
- Plot color should be blue.
- Outline/edge color should be red.

(b) Create a step type histogram.

4. (a) Create a histogram that plots two ndarrays x and y with 48 bins, in stacked horizontal histogram.

   (b) Add plot title and axes titles to above plot.

5. Write a code to plot the speed of a passenger train as shown in the figure given below:



Four Marks Questions.

1. Consider the following dataframe, and answer the questions given below:

df = pd.DataFrame({ "Quarter1":[2000, 4000, 5000, 4400, 10000], "Quarter2":[5800, 2500, 5400, 3000, 2900], "Quarter3":[20000, 16000, 7000, 3600, 8200], "Quarter4":[1400, 3700, 1700, 2000, 6000]})

Plot the line graph of data of all the quarters in one chart and also gives the legends to identify the quarters.

2. Draw a double bar graph chart showing the Cities in the X axis and happiness index in the Y axis for Male and Female. Use legends as Male and Female for identifying purpose.

City	Happiness Index Male	Happiness Index Female
Delhi	60	30
Beijing	40	60
Washington	70	70
Tokyo	65	55
Moscow	85	75

3. Given two ndarrays. Write a Program to plot them in the same line graph and show legends.

Five Marks Questions

1. Using the above data, plot the following:

App Name	App. Price	Total Downlaod
Angry Bird	75	197000
Teen Titan	120	209000
Marvel Comics	190	414000
ColorMe	245	196000
Fun Run	550	272000
Crazy Taxi	55	311000
Igram Pro	175	213000
Wapp Pro	75	455000
Maths Formulas	140	278000

(a) A line chart depicting the prices of the apps.

(b) A bar chart depicting the downloads of the apps.

(c) Convert the Est downloads sequence that has each download value divided by 1000.

Now create a bar chart that plots multiple bars for prices as well as downloads.

2. Given a dataframe namely prodf showing the production of various agricultural yields by four states UP,MP,AP and JK as shown below:

	Fruits	Pulses	Rice	Wheat
UP	44.1	23.2	102.4	120.7
MP	23.5	19.6	87.4	103.5
AP	18.9	12.6	89.4	76.4
JK	10.3	9.7	98.3	78.1

Write code to draw the multiple bar chart showing states on the x axis and production of various agricultural products on the y axis.

3. Questions given below are based on the following data:

Weight measurements for 16 small orders of French fries (in grams).

78 72 69 81 63 67 65 75

79 74 71 83 71 79 80 69

(I) Given the following set of data:

(a) Create a simple histogram from above data (b) Create a horizontal histogram from above data

(c) Create a step type of histogram from above data

(d) Create a cumulative histogram from above data [Textbook Q. Chapter 3 (Type C)]

(II) Create an ndarray containing 16 values and then plot this array along with dataset of previous question in same histogram

(a) normal histograms

(b) cumulative histograms

(c) horizontal histograms

(III) Out of above plotted histograms, which ones can be used for creating frequency polygons? Can you draw frequency polygons from all the above histogram?

(IV) Create/draw frequency polygon from the data used in above parts.

Marking scheme

One-mark questions

1. a
2. a
3. c
4. a
5. b
6. a
7. d
8. a
9. b
10. b

Two marks questions

1. Library – matplotlib

interface – pyplot

2. You can import matplotlib in following two ways:

    1. Using alias name: import matplotlib.pyplot as pp
    2. Without alias name: import matplotlib.pyplot

3 The chart has following elements/components:

1. Chart area or figure
2. Axis
3. Titles

4. Legends

4 The matplotlib provides following types of charts:

1. Line chart
2. Bar chart
3. Horizontal bar chart
4. Histogram

5 To change the thickness of line, use the linewidth parameter inside matplotlib.pyplot.plot()

To change the line style, use linestyle parameter.

6 import matplotlib.pyplot as plt

```
game=[ “Subway Surfer”, “Temple Run”, “Candy Crush”, “Bubble Shot”, “Runner Best”]
rating=[4,5,2,5,8]
plt.bar(game,rating)
plt.xlabel(“Games”)
plt.ylabel(“Rating”)
plt.show()
```

7 import matplotlib.pyplot as plt

```
import numpy as np
means = [20, 35, 30, 35, 27]
stds = [2, 3, 4, 1, 2]
indx = np.arange(len(means))
plt.bar(indx, means, color = 'cyan', label = 'Means')
plt.bar(indx+0.25, stds, color = 'olive', label = 'Stds')
plt.legend()
plt.show()
```

8 import matplotlib.pyplot as plt

```
Months = ['Dec', 'Jan', 'Feb', 'Mar']
Attendance = [70, 90, 75, 95]
plt.bar (Months, Attendance)
plt.show()
```

9 import matplotlib.pyplot as plt

medal= [‘Gold’,’Silver’,’Bronze’,’Total’]

num = [3,6,17,26]

plt.barh(num,medal)

plt.show()

Three marks questions.

1. import matplotlib.pyplot as plt

a=[2,3,4,5,6,7]

b=[1,2,3,4,5,6]

plt.plot(a,b)

plt.show()

2 import matplotlib.pyplot as plt

a=["VII","VIII","IX","X"]

b=[40,45,35,20]

plt.bar(a,b)

plt.show()

3 import matplotlib.pyplot as plt

Rev= np.array([12, 15, 13, 15, 16, 17, 14, 19, 19, 17, 19, 16, 17, 16, 18, 18, 18, 19, 19, 18, 19, 21, 20, 18, 19, 19, 21, 19, 19, 22])

(a) plt.hist (wt, bins=10, color ='b', edgecolor = 'r')
(b) plt.hist(wt,bins=10,histtype=’step’)

4 (a)import matplotlib.pyplot as plt

plt.hist([y,x], bins = 48, orientation = 'horizontal', histtype = 'barstacked')

(b) import matplotlib.pyplot as plt

plt.hist([y,x], bins = 48, orientation = 'horizontal', histtype = 'barstacked')
plt.title(“Horizontal stackedbar histogram”)
plt.x.label(“X-axix”)
plt.y.label(“Y-axix”)

5 import matplotlib.pyplot as plt

import numpy as np

x= np.arange(1, 5)

plt.plot(x, x* 1.5, label = 'Normal')

```
plt.plot(x, x* 3.0, label = 'Fast')
plt.plot(x, x/3.0, label = 'Slow')
plt.legend()
plt.show()
```

Four marks questions.

1
```
import matplotlib.pyplot as plt
X=[1,2,3,4,5]
plt.plot(X,df[“Quarter1”, label=” Quarter1”)
plt.plot(X,df[“Quarter1”, label=” Quarter2”)
plt.plot(X,df[“Quarter1”, label=” Quarter3”)
plt.plot(X,df[“Quarter1”, label=” Quarter4”)
plt.legend()
plt.show()
```

2
```
import matplotlib.pyplot as plt
HIM=[60,40,70,65,85]
HIF=[30,60,70,55,75]
X=[“Delh”,” Beijing”,” Washington”,” Tokyo”,” Moscow’]
plt.bar(X,HIM, label=”Happy Index Male”)
plt.bar(X+0.4,HIF, label=”Happy Index Female”)
plt.legend(loc=”upper left”)
plt.show()
```

3
```
import matplotlib.pyplot as plt
import numpy as np
# given series of data are y and z to be plotted on Y-axis
x= np.arange(0,10)
y = np.random.randint(10, size (10,))
z = np.random.randint(50, size (10,))
plt.figure(figsize = (8,4))
plt.plot(x, y, label = "Y data")
plt.plot(x, z, label = "Z data")
plt.legend (loc= "upper left")
```

```
plt.show()
```

Five marks questions.

1
```
plt.figure(figsize=(15, 7))
plt.plot(df2["App Price'])
plt.title("App Prices")
plt.xticks(np.arange(9), df2["App Name"])
plt.xlabel("App names")
plt.ylabel("prices")
plt.show()
```

b)
```
plt.figure(figsize = (15, 7))
plt.bar(df2['App Name'], df2[ 'Total Downloads'])
plt.title('Total Downloads of the Apps')
plt.xticks(np.arange(9), df2['App Name'])
plt.xlabel('App names')
plt.ylabel('Downloads')
plt.show()
```

(c)
```
Est = df2[ 'Total Downloads']/1000
x1 = np.arange(9)
x2=x1+0.2
plt.figure(figsize = (15, 7))
plt.bar(x1, df2[ 'App Price'], width = 0.2)
plt.bar(x2, Est, color = 'r', width = 0.2)
plt.title('Prices vs Est')
plt.xticks(np.arange(9), df2['App Name'])
plt.xlabel('App names')
plt.ylabel('Prices, Est')
plt.show()
```

2
```
import matplotlib.pyplot as plt
indx = [“UP”,”MP”,”AP”,”JK”]
plt.bar(indx, prodf[UP], label = 'UP')
plt.bar(indx, prodf[MP], label = 'MP')
```

```
plt.bar(indx, prodf[AP], label = 'AP')

plt.bar(indx, prodf[JK], label = 'JK')

plt.legend(loc="upper left")

plt.show()
```

3 (I) Given set of data:

```
wt=[78, 72, 69, 81, 63, 67, 65, 75, 79, 74, 71, 83, 71, 79, 88, 69]
```

(a) plt.hist (wt, bins=32)
(b) plt.hist (wt, bins=32,orientation="horizontal")
(c) plt.hist (wt, bins=32,histtype="step")
(d) plt.hist (wt, bins=32,cumulative="True")

(II)created another array having 16 values as:

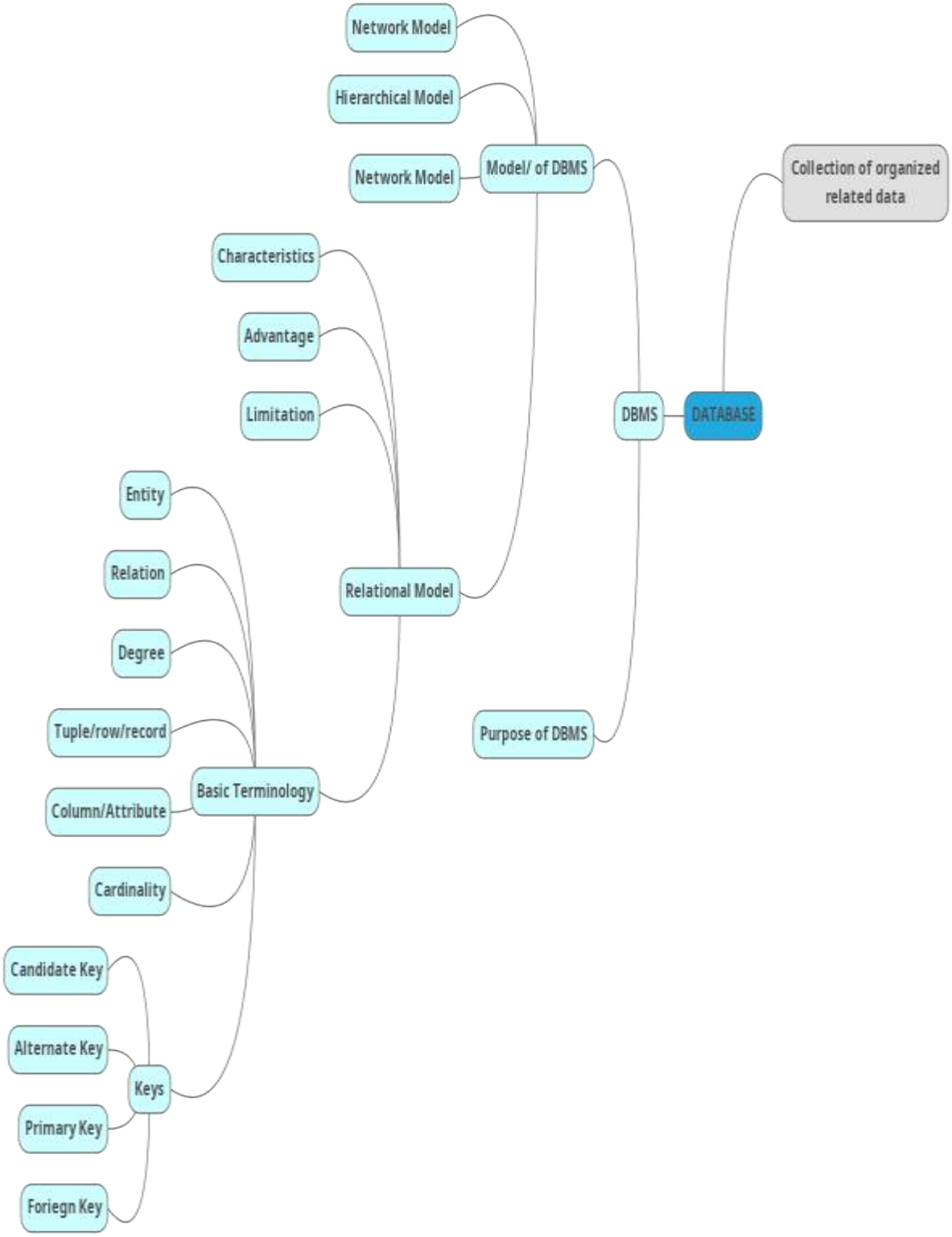
```
ar=np.linspace(40,100,16)

wt=[78, 72, 69, 81, 63, 67, 65, 75, 79, 74, 71, 83, 71, 79, 88, 69]
```

(a) plt.hist ([wt, ar])
(b) plt.hist ([wt, ar],cumulative="True")
(c) plt.hist ([wt, ar],orientation="horizontal")

(III) Since the above datasets do not represent continuous data or a pr0oper frequency distribution, we cannot create an ideal frequency polygon from any of the above datasets.

(IV) Same reason as above.


Network Model
Hierarchical Model
Network Model
Model/ of DBMS
Collection of organized related data
Characteristics
Advantage
Limitation
DBMS
DATABASE
Entity
Relation
Degree
Relational Model
Tuple/row/record
Purpose of DBMS
Basic Terminology
Column/Attribute
Cardinality
Candidate Key
Alternate Key
Keys
Primary Key
Foriegn Key

DATABASE:

- May be defined as a collection of interrelated data stored together to serve multiple application
- It is computer based record keeping system.
- It not only allows to store but also allows us modification of data as per requirements

DBMS:

- A DBMS refers to Database Management System
- It is a software that is responsible for storing, manipulating, maintaining and utilizing database.
- A database along with a DBMS is referred to as a database system.
- There are various DBMS software available in the market like :- Oracle, MS SQL Server, MySQL, Sybase, PostgreSQL, SQLite

Purpose of DBMS:

- Reduced Data redundancy –
- Control Data Inconsistency
- Sharing of data
- Ensure data integrity
- Enforce standard

Relational Database Model:

- In relational database model data is organized into table (i.e. rows and columns).
- These tables are also known as relations.
- A row in a table represent relationship among a set of values.
- A column represent the field/attributes related to relation under which information will be stored.
- For example if we want to store details of students then : Roll, Name, Class, Section, etc. will be the column/attributes and the collection of all the column information will become a Row/Record

Sample Tables:

EMPLOYEE

EMPNO	ENAME	GENDER	DEPTNO	SALARY	COMM
1	ANKITA	F	10	20000	1200
2	SUJEET	M	20	24000	
3	VIJAYA	F	10	28000	2000
4	NITIN	M	30	18000	3000
5	VIKRAM	M	30	22000	1700

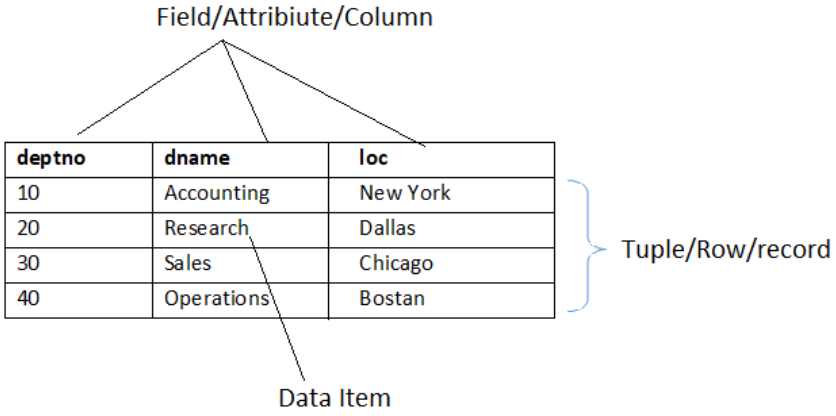
DEPT

DEPTNO	DNAME	LOC
10	Accounting	New York
20	Research	Dallas
30	Sales	Chicago
40	Operations	Bostan

Component of a table:

- Byte: group of 8 bits and is used to store a character.

- Data Item: smallest unit of named data. It represent one type of information and often referred to as a field or column information
- Record : collection of data items which represent a complete unit of information
- Table: collection of all Rows and Columns.



Common RDBMS:

- Oracle, MS SQL Server, MySQL, IBM DB2, IBM Informix, SAP Sybase, Adaptive Server Enterprise, SAP Sybase IQ, Teradata, PostgreSQL, SQLite, etc.
- Out of these MySQL, PostgreSQL and SQLite are Open source implementation.

MySQL

- Runs on virtually all platforms including Linux, Unix and Windows. Popular for web based application and online publishing. It is a part of LAMP (Linux, Apache, MySQL, PHP) stack

SQLite

- Relational DBMS but it is not client-server database engine rather, it is embedded into end program. Arguably the most widely deployed database engine as it is used by several browsers, OS and embedded systems (Mobiles).

PostgreSQL

- General purpose object-relational DBMS.
- It is the most advanced open source database system. It is free and open source i.e. source code is available under PostgreSQL license, a liberal open source license.

Common DBMS Tools for Mobile Devices:

- SQL Anywhere, DB2 Everywhere, IBM Mobile Database, SQL Server Compact, SQL Server Express, Oracle DatabaseLite, SQLite, SQLBase etc.
- Out of these SQLite is public domain open source implementation.

RELATIONAL DATA MODEL:-

Data is organized in two-dimensional tables called **relations**. The tables or relations are related to each other.

Characteristics of relational database are:-

- Data is arranged into rows and columns,
- Each relation is represented as a table.
- Every row in a table represents a single entity.
- At any given row or column position, there is one and only one value.

Advantages of a relational model are as follows:

- Changes made in the table structure do not affect the data access or other application programs.
- Tabular view also provides easier database design, use, implementation and management.
- Built-in query support .
- Mathematical operations can be successfully carried out using RDBMS.

The limitations of relational model are:

- RDBMS incurs hardware and system software overheads.
- The size of database becomes very large.

Various Terms Used in Relational Model: - A **relational database** is a type of database that stores and provides access to data points that are related to one another.

Basic Terminologies related to a Relational Database:-

- **Entity:** An entity is something that exists and an object which can be distinctly identified.

  For example, student entity, employee entity,

- **Attribute:** The term attribute is also used to represent a column.
- **Tuple:** Each row in a table is known as tuple.
- **Cardinality of Relation:** It is the number of records or tuples in the relation.
- **Degree of Relation:** Number of columns or attributes is known as degree of a relation.
- **Domain of Relation:** It defines the kind of data represented by the attribute.
- **Body of the Relation:** It consists of an unordered set of 0 or more tuples.

Concept of Keys

- In relation each record must be unique i.e. no two identical records are allowed in the Database.
- A key attribute identifies the record and must have unique values. There are various types of Keys:

Primary Key:

- A set of one or more attribute that can identify a record uniquely in the relation is called Primary Key.
- There can be only 1 primary key in a table
- Allows only distinct (no duplicate) values and also forces mandatory entry (NOT NULL) i.e. we cannot left it blank.

Candidate Key

- In a table there can be more than one attribute which contains unique values. These columns are known as candidate key as they are the candidate for primary key.

- Among these database analyst select one as a primary key based on requirement like must contain unique value, compulsory entry and where maximum searching is done etc.

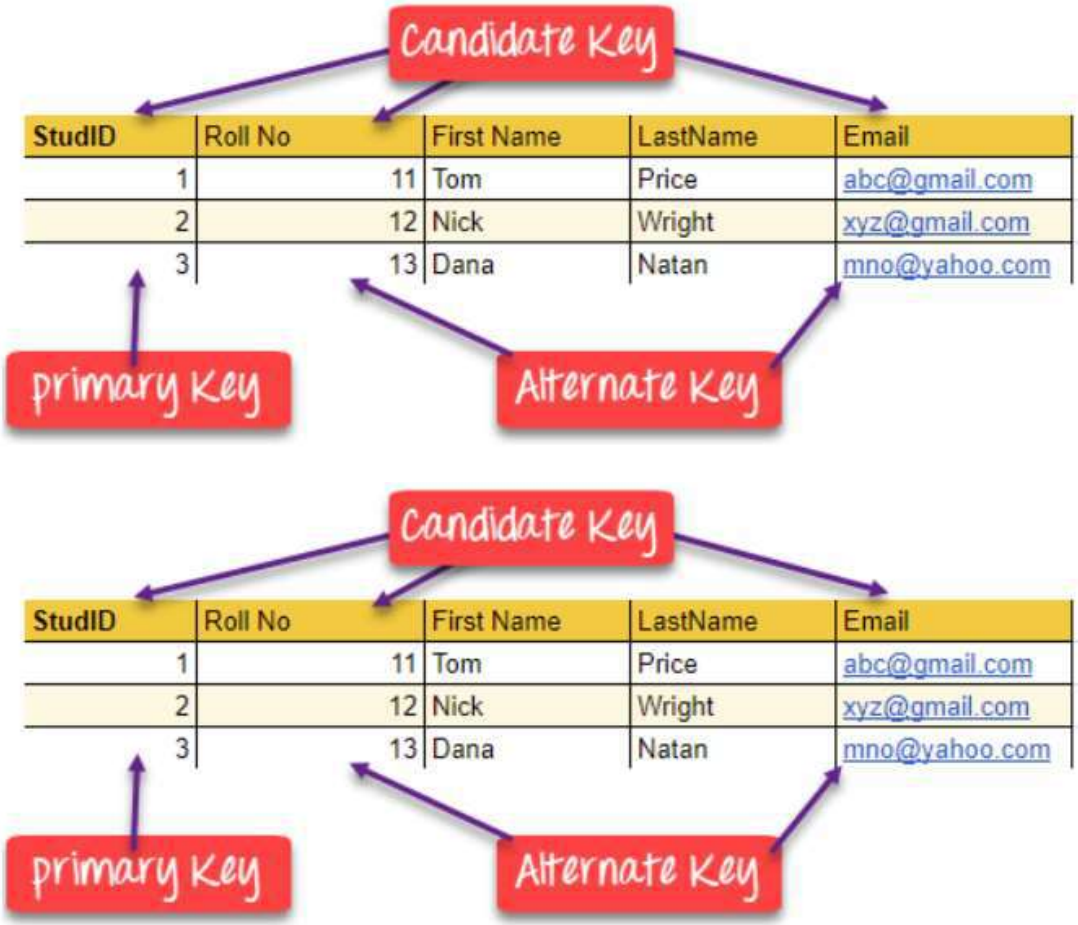
Alternate Key

- In case of multiple candidate keys, one of them will be selected as Primary Key and rest of the column will serve as Alternate Key
- A Candidate Key which is not a primary key is an Alternate Key.

Foreign key

- Used to create relationship between two tables.
- It is a non-key attribute whose value is derived from the Primary key of another table.
- Foreign key column will for the value in Primary key of another table, if present then entry will be allowed otherwise data will be rejected.
- Primary Key column table from where values will be derived is known as Primary Table or Master Table or Parent Table and Foreign key column table will be Foreign Table or Detail Table or Child table.

Example



Example 2:

EMPLOYEE

EMPNO	ENAME	GENDER	DEPTNO	SALARY	COMM
1	ANKITA	F	10	20000	1200
2	SUJEET	M	20	24000	
3	VIJAYA	F	10	28000	2000
4	NITIN	M	30	18000	3000
5	VIKRAM	M	30	22000	1700

Child Table

DEPARTMENT

DEPTNO	DNAME	LOCATION
10	HR	NEW YORK
20	ACCOUNTS	BRAZIL
30	SALES	CANADA
40	IT	INDIA

Parent Table

From the Above table definition, we can observe that the DEPTNO column of EMPLOYEE table is deriving its value from DEPTNO of table DEPARTMENT. So we can say that the DEPTNO of EMPLOYEE table is a foreign key whose value is dependent upon the Primary key column DEPTNO of table DEPARTMENT.

REFERENTIAL INTEGRITY:

- Used to ensure relationship between records in related tables are valid and user don't accidentally delete or change the related data.
- Referential integrity can be applied when:
- The master table's column is a Primary Key or has a unique index
- The related fields have the same data type
- Both tables must belong to same database.
- When referential integrity is enforced using Foreign Key you must observe the following rules:
- You cannot enter a value in Child Table which is not available in Master Table's Primary key column. However you can enter NULL values in foreign key
- You cannot delete a record from Master Table if matching record exists in related table.
- You cannot modify or change the Primary Key value in Master table if its matching record is present in related table.

STRUCTURED QUERY LANGUAGE

- It is a language that enables you to create and operate on relational databases
- It is the standard language used by almost all the database s/w vendors.
- Pronounced as SEQUEL
- It is portable i.e. it is compatible with most of the database.
- It is not a case sensitive language.
- It is very easy to learn.

SQL – features

- Allows creating/modifying a database's structure
- Changing security settings for system
- Permitting users for working on databases or tables
- Querying database
- Inserting/modifying/deleting the database contents

Classification of SQL

- DDL (Data Definition Language)
- DML (Data Manipulation Language)
- DCL (Data Control Language) or TCL (Transaction Control Language)

Data Definition Language:

- It allows to create database objects like creating a table, view or any other database objects.
- The information about created objects are stored in special file called DATA DICTIONARY
- DATA DICTIONARY contains metadata i.e. data about data.
- The commands of DDL are –
    - CREATE – To create a new database object
    - ALTER – To modify existing database object
    - DROP – To permanently remove existing database object.

Data Manipulation Language:

- It allows to perform following operation on data in the table
- Retrieval of information stored in table
- Insertion of new data in table
- Modification of existing data in table
- Deletion of existing data from table
- DML is of 2 type

- Procedural DML (in this we specify what data is needed and how to get it)
- Non-Procedural DML (in this we specify what data is needed without specifying how to get it)
- The commands of DML are
  - SELECT – To retrieve data from the table
  - INSERT – To insert new tuple/row in the table
  - UPDATE – To update existing tuple/row from the table
  - DELETE – To delete existing tuple/row from the table

Data Transaction Language:

- The commands here are used for Database Administration such as providing access or restriction to data/table to users.
- Some commands of DCL are
  - ROLLBACK
  - COMMIT
  - GRANT

EXERCISE:

1. What is Database? What are the advantages of Database System?
2. What is Primary Key? What are the restriction imposed by Primary Key? How many primary key can be applied on a Table? What is the difference between Primary key and Candidate Key?
3. What is Degree and Cardinality of table?
4. Explain the purpose of DDL and DML commands used in SQL. Also give two examples of each
5. Consider the following table and answer the questions that follow:

APPLICANTS

NO	NAME	FEE	GENDER	C_ID	DATEOFJOIN
1012	AMANDEEP	30000	M	A01	2012-01-02
1102	AVISHA	25000	F	A02	2009-03-24
1103	EKANT	30000	M	A03	2011-11-04
1049	ARUN	40000	M	A02	2009-05-06
1025	AMBER	30000	M	A01	2011-11-03

COURSES

CR_ID	CNAME
A01	DATA MINING
A02	PROGRAMMING
A03	LATHING

a) Identify the Primary Key and Foreign Key in APPLICANT and COURSES table
b) What is the Cardinality and Degree of APPLICANT table

6. Direction: In the questions given below, there are two statements marked as Assertion (A) and Reason (R). Choose the correct option out of the choices given below in each question:
Assertion (A): Primary key is a set of one or more attributes that identify tuples in a relation.
Reason (R): Every table must have one primary key.
A) Both (A) and (R) are correct and (R) is correct explanation of (A).
B) Both (A) and (R) are correct and (R) is not the correct explanation of (A).
C) (A) is True but (R) is False.
D) (A) is False but (R) is True.
Ans : (A)

7. Assertion (A): Foreign key is a non-key attribute whose value is derived from primary key of another table.
Reason (R): Each foreign key refers a candidate key in a relation.
A) Both (A) and (R) are correct and (R) is correct explanation of (A).
B) Both (A) and (R) are correct and (R) is not the correct explanation of (A).
C) (A) is True but (R) is False.
D) (A) is False but (R) is True.
Ans : (B)

8. Assertion (A): Data Integrity means that data is accurate and consistent in the database.
Reason (R): Data integrity also ensures that your data is safe from any outside forces.
A) Both (A) and (R) are correct and (R) is correct explanation of (A).
B) Both (A) and (R) are correct and (R) is not the correct explanation of (A).
C) (A) is True but (R) is False.
D) (A) is False but (R) is True.
Ans : (A)

9. Assertion (A): There can be multiple primary key in a table.
Reason (R): Duplicate values cannot be entered in primary key.
A) Both (A) and (R) are correct and (R) is correct explanation of (A).
B) Both (A) and (R) are correct and (R) is not the correct explanation of (A).
C) (A) is True but (R) is False.
D) (A) is False but (R) is True.
Ans : (D)

INTRODUCTION OF MYSQL

Brief history of MySQL:

- MySQL is freely available open source RDBMS
- It can be downloaded from www.mysql.org
- In MySQL information is stored in Tables.
- Provides features that support secure environment for storing, maintaining and accessing data.
- It is fast, reliable, scalable alternative to many of the commercial RDBMS today.
- It is developed and supported by MySQL AB, a company based in Sweden.
- This company is now subsidiary of Sun Microsystems. On April 2009 Oracle Corp. acquires Sun Microsystems.
- The chief inventor of MySQL was **Michael Widenius (a.k.a Monty)**. MySQL has been named after Monty's daughter My. The logo of MySQL is dolphin and name of that dolphin is 'Sakila'.

MYSQL DATABASE SYSTEM:

- MySQL database system refers to the combination of a MySQL server instance and MySQL database.
- It operates using Client/Server architecture in which the server runs on the machine containing the database and client connects to server over a network
- MySQL is a multiuser database system, meaning several users can access the database simultaneously.

The Server

- Listens for client requests coming in over the network and access the database as per the requirements and provide the requested information to the Client.

The Client

- Are the programs that connect to MySQL server and sends requests to the server and receives the response of Server. Client may be the MySQL prompt or it may be Front-end programming which connect to server programmatically like connecting to MySQL using Python Language or Java or any other language.

FEATURES OF MYSQL:

- Speed - MySQL runs very fast.
- Ease of Use -Can be managed from command line or GUI
- Cost Is available free of cost. It is Open Source
- Query language Support -Supports SQL
- Portability – Can be run on any platform and supported by various compilers
- Data Types - supports various data types like Numbers, Char etc.
- Security -Offers privileges and password systems that is very flexible and secure.
- Scalability and Limits -Can handle large databases. Some of real life MySQL databases contains millions of records.
- Connectivity-Clients can connect to MySQL using drivers
- Localization -The server can provide error message to client in many language
- Client and Tools -Provides several client and utility programs. Like mysqldump and mysqladmin. GUI tools like MySQL Administration and Query Browser.

STARTING MYSQL:

Click on Start →All Programs →MySQL → MySQL Server→ MySQL →Command Line Client

To exit from MySQL type **exit** or **quit** in front of MySQL prompt.

SQL and MYSQL:

- SQL stands for Structured Query Language.
- It is a language that enables you to create and operate on relational databases.
- MySQL uses SQL in order to access databases.
- It is the standard language used by almost all the database s/w vendors.

MYSQL Elements

- Literals
- Data types
- Nulls
- Comments

Literals

- It means the fixed value or constant value. It may be of character, numeric or date time type.
- Character and date/time literals are always in single quotation marks whereas numeric literals must be without single quotation marks
- For example – 'Virat', 12, 12.56, '04-20-2018'
- Date and time values are always in the format YYYY-MM-DD HH:MI:SS
- Special character like quotes are always written be preceding it back-slash(\). For example if we want to store value as Tom's Cat then it should be written as Tom\'s Cat

Data Type

- Means the type of value and type of operation we can perform on data. For example on numeric value we can store numbers and perform all arithmetic operations and so on.
- Numeric
- Date and time
- String types

Numeric Data Types

DATA TYPE	DESCRIPTION
INT	Numbers without decimal. Store up to 11 digits. -2147483648 to 2147483647
TINYINT	Small integer value between 0 – 255 (4 digits)
SMALLINT	More than TINYINT between -32768 to 32767 (5 digit)
MEDIUMINT	Integer values up to 9 digits
BIGINT	Very large integer value up to 11 digits
FLOAT(M,D)	Real numbers i.e. number with decimal. M specify length of numeric value including decimal place D and decimal symbol. For example if it is given as FLOAT(8,2) then 5 integer value 1 decimal symbol and 2 digit after decimal TOTAL – 8. it can work on 24 digits after decimal.
DOUBLE(M,D)	Real numbers with more precision up to 53 place after decimal.
DECIMAL	It is used to store exact numeric value that preserve exact precision for e.g. money data in accounting system. **DECIMAL(P,D) means P no. of significant digits (1-65), D represent no. of digit after decimal(0-30), for e.g DECIMAL(6,2) means 4 digit before decimal and 2 digit after decimal. Max will be 9999.99**

Date and Time Types

- **DATE**-A date in YYYY-MM-DD format between 1000-01-01 to 9999-12-31.
- **DATETIME**-Combination of date and time. For example to store 4th December 2018 and time is afternoon 3:30 then it should be written as – 2018-12-04 15:30:00
- **TIMESTAMP** - Similar to DATATIME but it is written without hyphen for example the above date time is stored as 20181204153000
- **TIME**-To store time in the format HH:MM:SS
- **YEAR**(M) -To store only year part of data where M may be 2 or 4 i.e. year in 2 digit like 18 or 4 digit like 2018.

String Types

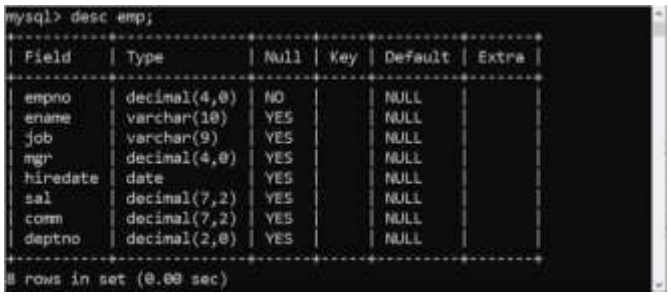
- CHAR(M)
  - Fixed length string between 1 and 255.
  - It always occupy M size, irrespective of actual number of characters entered.
  - Mostly use in the case where the data to be insert is of fixed size like Grade (A,B,C,..) or Employee code as E001, E002, etc
- VARCHAR(M)
  - Variable length string between 1 and 65535 (from MySQL 5.0.3) , earlier it was 255
  - It takes size as per the data entered for example with VARCHAR(20) if the data entered is MOBILE then it will take only 6 byte.
  - It is useful for the data like name, address where the number of character to be enter is not fixed.
- Difference between CHAR & VARCHAR

CHAR	VARCHAR
Fixed length string	Variable length string

Fast, no memory allocation every time	Slow, as it take size according to data so every time memory allocation is done
It takes more memory	It takes less space

Simple Queries in SQL

- **SHOW DATABASES**- This command is used to list all databases on MySql Server
- **USE <DATABASENAME>-** This command is used to change/select/open given database
  e.g. To open a database named 'test'
  >>> **use test**
- **SHOW TABLES** – This command will list all the tables from current database. If no database is selected it will generate error.
- **SELECT DATABASE ()** – This command will display the name of current database.
- **DESC <TABLENAME> OR DESCRIBE <TABLENAME>** - This command will be used to display the structure of the table. For Eg: The above figure display the use of **describe** command



```
mysql> desc emp;
+----------+--------------+------+-----+---------+-------+
| Field    | Type         | Null | Key | Default | Extra |
+----------+--------------+------+-----+---------+-------+
| empno    | decimal(4,0) | NO   |     | NULL    |       |
| ename    | varchar(10)  | YES  |     | NULL    |       |
| job      | varchar(9)   | YES  |     | NULL    |       |
| mgr      | decimal(4,0) | YES  |     | NULL    |       |
| hiredate | date         | YES  |     | NULL    |       |
| sal      | decimal(7,2) | YES  |     | NULL    |       |
| comm     | decimal(7,2) | YES  |     | NULL    |       |
| deptno   | decimal(2,0) | YES  |     | NULL    |       |
+----------+--------------+------+-----+---------+-------+
8 rows in set (0.00 sec)
```

- **CREATE DATABASE <DATABASENAME>** - This command is use to create a new database.
  For example – **create database mydb** will new database **mydb**
  Above command will generate error, if database already exist. To supress the error following command can be used
  create database if not exist mydb

- **DROP DATABASE <DATABASENAME>** - This command is used to remove existing database.
  For example – **drop database mydb** will permanently delete **mydb** database.
  Above command will generate error, if database already exist. To supress the error following command can be used
  drop database if exist mydb

- **CREATE TABLE <TABLENAME>** - This command / statement is used to create a new table in a database. The syntax the is
  CREATE TABLE table_name
  (
  column1 datatype[(size)] [constraint] ,
  column2 datatype[(Size)] [constraint],
  column3 datatypeI[(size)] [constraints] ,
  ....
  );
  Here size and constraints are optional.
  e. g.
  CREATE TABLE PET

(

NAME VARCHAR(20),
OWNER VARCHAR(20),
SPECIES VARCHAR(20),
SEX CHAR(1),
BIRTH DATE, DEATH DATE

);

- **DROP TABLE <TABLENAME>** - This command is used to permanently delete the table from database.
  For example, **drop table pet**; will delete **pet** table from database
- **ALTER TABLE <TABLENAME>** - This command is used to modify the structure of existing table such as adding new column, removing existing column, rename or changing data type, size and constraints.
  - **Adding new column to exiting table**
    Syntax :

    Alter table <tablename>
    Add [column] column_name datatype [(size)];

  - **Removing new existing column from the table**
    Syntax:
    Alter table <tablename>
    drop column <column_name>;

  - **Changing datatype/size of the column**
    Systax:
    alter table <tablename>modify column <colname> datatype(size);

  - **Renaming column name**
    Alter table tablename
    Change old_columnname new_column_name datatype (size)

- **UPDATE <TABLENAME>** - This command is used to update data from the table
  Syntax:

  UPDATE table_name SET column_name=new_value, column2_name=new_value
  WHERE condition;
  e.g. UPDATE emp set sal=sal+100 where ename =’Scot’

  The above query will update salary of Scot by 100.
  If where clause is not given, it will update/modify the value of column from each row.

- **INSERTING INTO TABLE_NAME** – This command is used to add new row to the table
  Syntax :
  INSERT INTO table_name VALUES (list of values)

- **SELECT DATA FROM TABLE USING SELECT STATEMENT**

  Syntax:

```
SELECT      */ COLUMN LIST
FROM        TABLE(S)
WHERE       CLAUSE/CONDITION
GROUP BY    CLAUSE/COLUMN NAME
HAVING      CLAUSE/CONDITION
ORDER BY    COLUMN NAME ASC/DESC
```

- **Selecting/displaying entire data from the table**
  Syntax:
  SELECT * From tablename;

- **Selecting/displaying data from specific column**
  Syntax:
  SELECT column1, column2, column3, …. FROM tablename;
- **Giving descripting name for column using column alias**
  Syntax:
  SELECT */col_name as 'alias1' FROM tablename;

- **Removing duplicate value from the column using distinct clause**
  Syntax:
  DISTINCT column_name FROM tablename;

- **Display all data from column using all clause**
  Syntax:
  SELECT ALL column_name FROM tablename;

- **Inserting text in the query**
  Syntax:
  SELECT 'text' FROM tablename;

```
mysql> select ename,' is working as' job from emp;
+--------+---------------+
| ename  | job           |
+--------+---------------+
| SMITH  |  is working as |
| ALLEN  |  is working as |
| WARD   |  is working as |
| JONES  |  is working as |
| MARTIN |  is working as |
```

Number of rows in result depend on total number of rows in the table

- **Performing simple calculation in the query**
  Syntax:
  Select 2+4;
  All operation can be perform such as addition (+), subtraction (-), multiplication (*), division (/) and remainder (%)

```
mysql> select 3+5;
+-----+
| 3+5 |
+-----+
|   8 |
+-----+
1 row in set (0.00 sec)
```

Number of row in result depend on total number of rows in the table

- **Performing calculation on column**

  Syntax:

  SELECT column1 operator value, column2…. FROM Tablename;

```
mysql> select ename, sal, sal*12 'Annual Salary' from emp;
+--------+---------+---------------+
| ename  | sal     | Annual Salary |
+--------+---------+---------------+
| SMITH  |  800.00 |       9600.00 |
| ALLEN  | 1600.00 |      19200.00 |
| WARD   | 1250.00 |      15000.00 |
| JONES  | 2975.00 |      35700.00 |
| MARTIN | 1250.00 |      15000.00 |
| BLAKE  | 2850.00 |      34200.00 |
```

- **Working with null value**

  Any operation on NULL will result into NULL, MySql provide **ifnull**() function to work with null value. If column contain null value, it is replaced by given value, otherwise display original value.

  Syntax:

  ifnull(column, value_to_replace')

```
mysql> select ename, sal, comm, ifnull(comm,'N/A') from emp;
+--------+---------+---------+---------------------+
| ename  | sal     | comm    | ifnull(comm,'N/A')  |
+--------+---------+---------+---------------------+
| SMITH  |  800.00 |    NULL | N/A                 |
| ALLEN  | 1600.00 |  300.00 | 300.00              |
| WARD   | 1250.00 |  500.00 | 500.00              |
| JONES  | 2975.00 |    NULL | N/A                 |
| MARTIN | 1250.00 | 1400.00 | 1400.00             |
| BLAKE  | 2850.00 |    NULL | N/A                 |
| CLARK  | 2450.00 |    NULL | N/A                 |
| SCOTT  | 3000.00 |    NULL | N/A                 |
| KING   | 5000.00 |    NULL | N/A                 |
| TURNER | 1500.00 |    0.00 | 0.00                |
| ADAMS  | 1100.00 |    NULL | N/A                 |
| JAMES  |  950.00 |    NULL | N/A                 |
| FORD   | 3000.00 |    NULL | N/A                 |
| MILLER | 1300.00 |    NULL | N/A                 |
+--------+---------+---------+---------------------+
14 rows in set (0.00 sec)
```

  Here, comm column contains null value which is replaced by 'N/A'.

- **Restricting rows using where clause**

  Where clause in the query will restrict number of rows in the output based on condition.

  Syntax:

  SELECT */column list
  FROM Tablename
  where condition ;

  Condition (column_name operator expression)

  Relational Operator

  > greater than

< less than
>= greater than equal to
<= less than equal to
= equal
! = or <> not equal to

Logical Operator

And – evaluated true if all the logical expression is true otherwise false.
Or - evaluated true if any the logical expression is true otherwise false.

Logical operator is used to combine two or more logical expression,

Membership Operator

in –
Not in
The IN operator allows you to specify multiple values in a WHERE clause.
The IN operator is a shorthand for multiple OR conditions.

- **Comparing NULL**

  is null
  is not null
  NULL (Absence of value) value cannot be compared using Relational operator. The above statement is used to check whether column contains NULL or not.

- **Range Operator**

  Between

- **Pattern Matching**

  Like
  Not Like

  Like clause is used to match pattern using two wild card characters
  _ (underscore) – single unknown character
  % (modulo) - Zero or more unknown characters
  e.g.

words staring with 't'	't%'
words ending with 't'	'%t'
words containing 't' -	'%t%'
word with 't' as second letter -	'_t%'
words with 't' as third last character –	'%t_ _'
words containing four letter and 't' as second letter –	'_ t_ _'

- **Ordering data using ORDER BY clause**
  - ORDER BY clause is used to arrange data in ascending/descending order
  - Ordering can be done on more than one column, in this case data is arranged according first column and then second column and onwards.
  - By default data is arranged in ascending order, in no ordering information (asc/desc) is given.

Syntax:

SELECT */col_list
FROM tablename
ORDER BY col1 asc/desc, col2 asc/desc;

AGGREGATE FUNCTIONS

- An aggregate function performs a calculation on multiple values and returns a single value.
- These function work on multiple rows collectively return single value.
- List of Aggregate functions are
  - max() : return maximum value in set of value

```
mysql> select max(sal) from emp;
+----------+
| max(sal) |
+----------+
|  5000.00 |
+----------+
1 row in set (0.01 sec)
```

  - min() – return minimum value from the set of values

```
mysql> select min(sal) from emp;
+----------+
| min(sal) |
+----------+
|   800.00 |
+----------+
1 row in set (0.00 sec)
```

  - avg() –return average value in set of non-null values

```
mysql> select min(sal) from emp;
+----------+
| min(sal) |
+----------+
|   800.00 |
+----------+
1 row in set (0.00 sec)
```

  - sum() - Return the summation of all non-NULL values of the set of values.

```
mysql> select sum(sal) from emp;
+----------+
| sum(sal) |
+----------+
| 29025.00 |
+----------+
1 row in set (0.01 sec)
```

  - count() - Return the number of rows in a group
    - Count(*) – return number of rows, including rows with NULL

```
mysql> select count(*) from emp;
+----------+
| count(*) |
+----------+
|       14 |
+----------+
1 row in set (0.00 sec)
```

    - Count(column_name) - return number of rows, excluding rows with NULL for the given column

```
mysql> select count(comm) from emp;
+-------------+
| count(comm) |
+-------------+
|           4 |
+-------------+
1 row in set (0.00 sec)
```

- Count (distinct column_name) – return number of rows with duplicate removed

```
mysql> select count(distinct job) from emp;
+--------------------+
| count(distinct job) |
+--------------------+
|                  5 |
+--------------------+
1 row in set (0.01 sec)
```

In above example there are 14 row in the EMP table, but distinct clause only consider unique value.

- **Group By Clause**
  - It is used in a SELECT statement to collect data across multiple records and group the results by one or more columns.
    Syntax:
    SELECT column_name, aggregate_function
    FROM table_name
    GROUP BY column_name

```
mysql> select job, max(Sal) from emp group by job;
+-----------+----------+
| job       | max(Sal) |
+-----------+----------+
| ANALYST   |  3000.00 |
| CLERK     |  1300.00 |
| MANAGER   |  2975.00 |
| PRESIDENT |  5000.00 |
| SALESMAN  |  1600.00 |
+-----------+----------+
```

In above example salary is grouped on job and maximum salary from each job is displayed.

  - Select clause involving group by clause can contain column present in group by clause, aggregate function or no column. Otherwise it will return random data from other column as given below.

```
mysql> select ename, job, max(sal) from emp group by job;
+-------+-----------+----------+
| ename | job       | max(sal) |
+-------+-----------+----------+
| SCOTT | ANALYST   |  3000.00 |
| SMITH | CLERK     |  1300.00 |
| JONES | MANAGER   |  2975.00 |
| KING  | PRESIDENT |  5000.00 |
| ALLEN | SALESMAN  |  1600.00 |
+-------+-----------+----------+
```

- **Having clause** –
  - Having clause is used to place condition on aggregate function in conjunction with group by clause.

- Having clause in placed after where clause in select statement.
  Syntax:
  SELECT columm_name, aggregate_function(col_name)
  FROM table
  WHERE condition
  GROUP BY column_name
  HAVING aggregate_function(column_name) operator expression;

```
mysql> select deptno, max(sal), count(*) from emp group by deptno;
+--------+---------+---------+
| deptno | max(sal) | count(*) |
+--------+---------+---------+
|     10 |  5000.00 |        3 |
|     20 |  3000.00 |        5 |
|     30 |  2850.00 |        6 |
+--------+---------+---------+
3 rows in set (0.00 sec)
```

The above query will display deptno, maximum salary and number of employees from each department.
The query given below display deptno, maximum salary and number of employees from those department which have maximum salary greater than equal to 3000.

```
mysql> select deptno, max(sal), count(*) from emp
    -> group by deptno having max(sal)>=3000;
+--------+---------+----------+
| deptno | max(sal) | count(*) |
+--------+---------+----------+
|     10 |  5000.00 |        3 |
|     20 |  3000.00 |        5 |
+--------+---------+----------+
```

As condition is on aggregate function max(), where clause can't be used in this case.

EXERCISE QUESTIONS:

1. What is MySQL used for? Abhay wants to start learning MySQL. From where can he obtain the MySQL software ?
2. In the table "Student", Priya wanted to increase the Marks (Column Name:Marks) of those students by 5 who have got Marks below 33. She has entered the following statement:
3. SELECT Marks+5 FROM Student WHERE Marks<33;
   Identify errors(if any) in the above statement. Rewrite the correct SQL statement.
4. Write SQL statement to add a column ''COUNTRY'' with data type and size as VARCHAR(70) to the existing table named ''PLAYER''. Is it a DDL or DML or TCL command ?
5. Table Student has the columns RNO and SCORE. It has 3 rows in it. Following two SQL statements were entered that produced the output (AVG(SCORE) as 45 and COUNT(SCORE) as 2) :
   (i) AVG(SCORE)
   (ii) Count(score)
6. 'Employee' table has a column named 'CITY' that stores city in which each employee resides. Write SQL query to display details of all rows except those rows that have CITY as 'DELHI' or 'MUMBAI' or 'CHANDIGARH'.
7. How is a database related to a table ?

8. Mr. Sen has to create a table named 'Employee' with Columns to store EmpID, Name, Designation, Age and Salary. EmpID is the Primary key and Name cannot be NULL. Some of the rows that will be inserted are shown below.

101	Smita Kumar	Secretary	28	39500.00
102	Mani Scott	Programmer	32	45300.00
103	Firdaus Ali	Programmer II	45	67500.00

Write SQL query to create the above table with appropriate data types and sizes of columns.

9. Ms. Rajshri is the Class Teacher of Class XII. She wants to create a table named 'Student' to store marks in different subjects of her class. Identify any 4 columns for the table along with their suitable data types.

10. "XYZ" Company conducts workshops for employees of organizations. The company requires data of workshops that are organized. Write SQL query to create a table 'Workshop' with the following structure:

Field	Type	Constraint
WorkshopId	integer	Primary Key
Title	Varchar(50)	
DateWorkshop	Date	
NumSpeakers	Integer	

11. Ariya wants to add another column 'Gender' in the already existing table 'CUSTOMERS'. She has written the following statement. However, it is showing error. Rewrite the correct statement.

MODIFY TABLE CUSTOMERS GENDER char(1);

12. Explain the following statement with the help of example:

13. "In a transaction either all the SQL statements be committed or all rolled back."

14. How is HAVING clause similar to WHERE clause? How is HAVING clause different from WHERE clause? Explain with the help of examples of each.

15. Consider the following table 'Transporter' that stores the order details about items to be transported. Write SQL commands for the statements (i) to (viii).

Table : TRANSPORTER

ORDERNO	DRIVERNAME	DGRD	ITEM	TRAVELDATE	DESTINATION
10012	RAM YADAV	A	TELEVISION	2019-04-19	MUMBAI
10014	SOMNATH SINGH		FURNITURE	2019-01-12	PUNE
10016	MOHAN VERMA	B	WASHING MACHINE	2019-06-06	LUCKNOW
10018	RISHI SINGH	A	REFRIGERATOR	2019-04-07	MUMBAI
10019	RADHE MOHAN		TELEVISION	2019-05-30	UDAIPUR
10020	BISHEN PRATAP	B	REFRIGERATOR	2019-05-02	MUMBAI
10021	RAM		TELEVISION	2019-05-03	PUNE

(i) To display names of drivers and destination city where TELEVISION is being transported.

(ii) To display driver names and destinations where destination is not MUMBAI.

(iii) To display the names of destination cities where items are being transported. There should be no duplicate values.

(iv) To display details of rows that have some value in DRIVERGRADE column.

(v) To display names of drivers, names of items and travel dates for those items that are being transported on or before 1st April 2019.

(vi) To display the number of drivers who have 'MOHAN' anywhere in their names.

(vii) To display the names of drivers, item names and travel dates in alphabetic (ascending) order of driver names.

(viii) To display names of drivers whose names are three characters long.

17. In CHAR(10) and VARCHAR(10), what does the number 10 indicate ?

18. 'Employee' table has a column named 'CITY' that stores city in which each employee resides. Write SQL query to display details of all rows except those rows that have CITY as 'DELHI' or 'MUMBAI' or 'CHANDIGARH'.

19. Consider the table given below. Write SQL queries for (i) to (vii).

TABLE : GYM

REGID	NAME	PREWEIGHT	CURRWEIGHT	DOJ	GENDER	BRANCH
1001	SHEELA SIA	115	98	2017-09-02	F	ADARSH VIHAR
1002	FAWAD KHAN	92	80	2018-10-11	M	MODEL TOWN
1003	PAWAN SINGH	85	80	2018-02-03	M	NIRMAN NAGAR
1004	SUPRIYA ARORA	113	100	2018-01-16	F	MODEL TOWN
1005	AJOY BASU	67	75	2017-12-09	M	NIRMAN NAGAR
1006	TANMAY JACOB	100	71	2017-11-18	M	ADARSH VIHAR
1007	LAKSHMI VENKAT	98	95	2018-06-09	F	MODEL TOWN

Columns REGID stores Registration Id, PREWEIGHT stores weight of the person before joining Gym, CURRWEIGHT stores current weight, DOJ stores Date of Joining, BRANCH stores the branch of Gym where the person has enrolled.

(i) To display names of members along with their previous and current weights who are in Model Town branch.

(ii) To display all names of members, previous weight, current weight, Change in weight (i.e. how much increase from previous weight or decrease from previous weight, Decrease will be displayed with negative sign)

(iii) To display BRANCH wise count of members in the Gym. (i.e. display the BRANCH and number of members in each BRANCH)

(iv) To display names and date of joining of all the members who joined in the year 2018.

(v) To display Names and Current weight of all the members in descending order of Current Weight.

(vi) To display the names and date of joining of male members who have joined after 27th September 2018.

(vii) To display names and date of joining of members who have their names starting with 'S' and ending with 'a'.

20. Consider the table Flight given below, write command in SQL for (i) to (iv) and output for (v) to (viii)

Table : FLIGHT

FLIGHT_NO	ORIGIN	DESTINATION	SEATS	FLIGHTDATE	RATE
1005	Varanasi	Nepal	275	12-Dec-07	3000
2785	Delhi	Kerala	290	17-Jan-08	5500
6587	Mumbai	Varanasi	435	19-Feb-08	5000
1265	Varanasi	Nepal	200	02-Jan-08	5400
4457	Delhi	Lucknow	150	22-Feb-08	4500
6856	Varanasi	Mumbai	180	03-Mar-08	6000

i) To display Flight flying between Varanasi and Nepal.

ii) To display the different Origin of Flights.

iii) To display list of flights in descending order of Rate.

iv) To display flight details of the flight whose flightdate is after Jan 2008.

v) SELECT Flight_No, Destination FROM Flight WHERE Destination LIKE '_u%';

vi) SELECT Origin, COUNT(*) FROM Flight GROUP BY Origin;
vii) SELECT Origin, Destination FROM Flight WHERE seats>400;
viii) SELECT SUM(Rate),MAX(Seats) FROM Flight;

21. Consider the following tables PARTICIPANT and ACTIVITY and answer the questions that follow :

Table : PARTICIPANT

ADMNO	NAME	HOUSE	ACTIVITYCODE
6473	Kapil Shah	Gandhi	A105
7134	Joy Mathew	Bose	A101
8786	Saba Arora	Gandhi	A102
6477	Kapil Shah	Bose	A101
7658	Faizal Ahmed	Bhagat	A104

Table : ACTIVITY

ACTIVITYCODE	ACTIVITYNAME	POINTS
A101	Running	200
A102	Hopping bag	300
A103	Skipping	200
A104	Bean bag	250
A105	Obstacle	350

(i) When the table ‘‘PARTICIPANT’’ was first created, the column ‘NAME’ was planned as the Primary key by the Programmer. Later a field ADMNO had to be set up as Primary key. Explain the reason.

OR

Identify data type and size to be used for column ACTIVITYCODE in table ACTIVITY.

(ii) To display Names of Participants along with Activity Codes and Activity Names for only those participants who are taking part in Activities that have ‘bag’ in their Activity Names and Points of activity are above 250.

22. In a database there are two tables ‘LOAN’ and ‘BORROWER’ as shown below:

Table: LOAN

LOAN_NO	BRANCH_NAME	AMOUNT
K-70	Downtown	5000
K-230	Redwood	6000
K-260	Perryridge	3700

Table : BORROWER

CUSTOMER_NAME	LOAN_NO
Jones	K-170
Smith	K-230
Hayes	K-155

(i) Identify the foreign key column in the table BORROWER.
(ii) Write sql query to list names of all customers with their Amount in ascending order?

23. Complete exercise on SQL

HOSPITAL

NO	NAME	AGE	DEPARTMENT	DATEOFADM	CHARGES	SEX	DOCTOR_ID
1	ARPIT	62	SURGERY	21-Jan-1998	3000	M	101
2	ZARINA	22	ENT	12-Dec-1997	2500	F	102
3	KAREEM	32	ORTHOPEDIC	19-Feb-1998	2000	M	103
4	ARUN		SURGERY	1-Nov-1996	3000	M	104
5	ZUBIN	30	ENT	1-Dec-1998	2500	M	103

6	KETAKI	16	RADIOLOGY	24-Feb-1998	2500	F	102
7	ANKITA	29	CARDIOLOGY	20-Feb-1999	8000	F	103
8	ZAREEN	45	GYNECOLOGY	22-Feb-1998	3000	F	104
9	KUSH		CARDIOLOGY	13-Jan-1999	8000	M	101
10	SHILPA	23	NUCLEAR MEDICINE	21-Feb-2000	4000	F	102

DOCTOR

D_ID	DOCTOR_NAME	PHONE	DATE_APPOINTMENT
101	Dr ASTHANA	9835366731	01-Feb-1996
102	Dr SANJEEV REDDY	8093626276	01-Mar-1995
103	Dr AFTAB ALAM	9816457891	01-Apr-1993
104	Dr NIBEDITA MISHRA	1234567890	01-Apr-1992

SELECT WHERE

i) To display the records of all patients who have Department as ENT
ii) To display the records of all patients with Charges greater than 3200
iii) To display the records of all patients who have Department as ENT and CARDIOLOGY
iv) To display the records of all patients in Department SURGERY and DATEOFADM before 01-01-99
v) To display the records of all patients who are MALE and pay Charges less than 2500
vi) To display the records of all patients who are admitted from 01-01-1980 and pay charges between 3200 to 4000

PATTERN

i) To display the records of all patients whose names start with S.
ii) To display the records of all patients whose Department have the letters OGY in them.
iii) To display the records of all patients whose names have E as their second last character.
iv) To display the records of all patients with names of five characters and Sex as Male

NOT

i) To display the Records of all patients who do not have their age entered
ii) To display the Records of all patients who are not admitted in ENT Department
iii) To display the Records of all patients who do not pay Charges in between 2800 to 4200
iv) To display the records of all patients not in Department SURGERY and ENT and DATEOFADM before 01-01-99

ORDER BY

i) To display the NAME, DEPARTMENT and SEX of all the patients sorted in descending order of CHARGES.
ii) To display the NO and DATEOFADM of all the patients sorted in ascending order of DOCTORID
iii) To display the NAME, AGE and SEX of all the patients sorted in ascending order of DATEODADM

DISTINCT

i) To display the unique types of Department of the patients.
ii) To display the unique types of DOCTORID of patients.

NULL

i) To display the details of the patients who do not have entered their age.
ii) To display the list all patients who have entered their age..

BETWEEN

i) To display records of patients with Department CARDIOLOGY and Charges between 2000 and 3200
ii) To display the record of all patients who have charges in the range of 2800 to 3500
iii) To display the record of all patients who were admitted in the period between 01-01-1996 to 01-01-1998
iv) To display the record of all patients who have age from 40 to 50.

LIST

i) To display the record of all patients who have Department as SURGERY, ENT and RADIOLOGY
ii) To display the record of all patients who are under Doctorid 101,102 and 103
iii) To display the record of all patients who pay charges 2500, 3000 and 4000
iv) To display patients not in Department SURGERY and ENT and DATEOFADM before 01-01-99

SELECT WITH CONCAT

i) To display the NAME, DEPARTMENT and DATEOFADM of all patients in the following format.

ARPIT IS ADMITTED TO **SURGERY** DEPARTMENT SINCE **21-Jan-1998**
ZARINA IS ADMITTED TO **ENT** DEPARTMENT SINCE **12-Dec-1997**
KAREEM IS ADMITTED TO **ORTHOPEDIC** DEPARTMENT SINCE **19-FEB-1998**

i) To display the NAME, AGE, CHARGES, CHARGES * 13% as "TOTAL" of all patients in the following format.

NAME	AGE	CHARGES	TOTAL
ARPIT	62	3000	390
ZARINA	22	2500	325

AGGREGATE FUNCTIONS

i) Find the max Charges which a patient pays
ii) Find the min Charges which a patient pays
iii) Find the sum of all Charges for all patients who are 30 years or above of age
iv) Find the average of Charges of Department ENT and CARDIOLOGY
v) Find the number of male patients

CREATE

i) Create the table HOSPITAL with datatypes and constraint as provided

COLUMN NAME	DATA TYPE	LENGTH	CONSTRAINT
NO	Integer	5	PRIMARY KEY
NAME	Character	20	NOT NULL
AGE	Integer	3	
DEPARTMENT	Character	20	NOT NULL
DATEOFADM	Date		
CHARGES	Float	7,2	Should not be less than zero
SEX	Character	1	Default value if M
DOCTOR_ID	Integer	4	References DOCTOR (D_ID)

ii) Create the table DOCTOR with datatypes and constraint as provided

COLUMN NAME	D_ID	DOCTOR_NAME	PHONE	DATE_APPOINTMENT
DATATYPE	Integer	Character	Character	Date
LENGTH	3	20	10	
CONSTRAINT	PRIMARY KEY	NOT NULL	UNIQUE	

INSERT

i) Write the commands to insert the first two records in the table HOSPITAL

UPDATE

i) Change the Department of ZUBIN to RADIOLOGY
ii) Increase the Charges of Male patients by 5 %
iii) Input the age of ARUN as 26

DELETE

i) Delete the record of SHILPA
ii) Delete the record of all patients who under DOCTORID 104

ALTER

i) Add a new column Contact_Address char(30) to the table HOSPITAL
ii) Increase the column size of name by 10
iii) Add NOT NULL Constraint to Contact_Address column

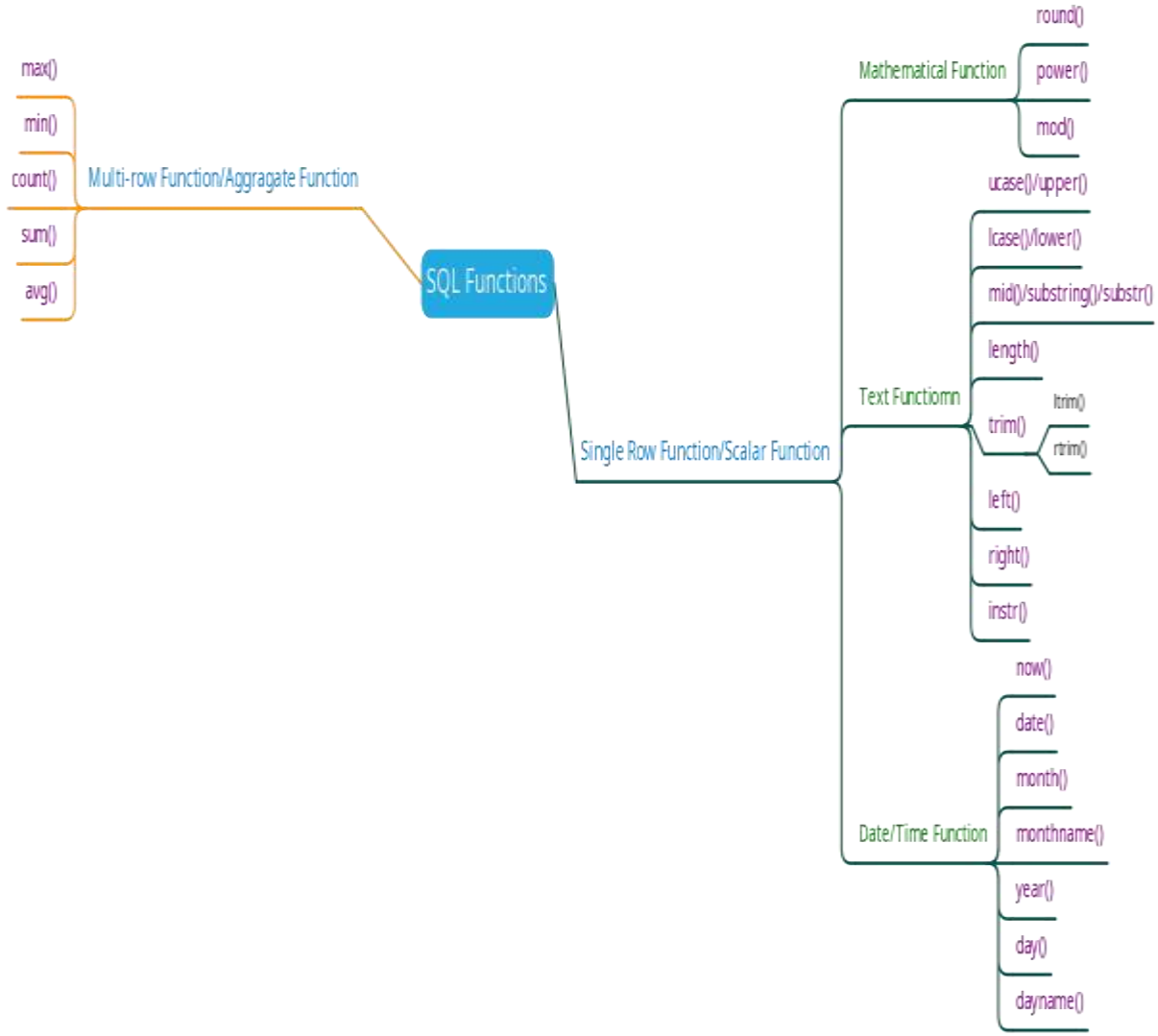
DROP

i) Drop the table HOSPITAL

GROUP BY

i) Show the number of patients in each Department
ii) Show the average Charges paid by the patients in department ENT and CARDIOLOGY
iii) Show the min DATEOFADM of each Department where minimum DATEOFADM is greater than 22-Feb-1998

SQL FUNCTIONS (STRING, MATHS and DATE/TIME)



➢ MATHS FUNCTIONS

Mathematical functions are also called number functions that accept numeric input and return numeric values.

Function	Description	Example	
		Function	Result
ROUND ( )	Round the column, expression or value to given decimal places.	ROUND (45.926,2)	45.93
TRUNC ( )	Truncates the column, expression or value to specified decimal places.	TRUNC (45.926,2)	45.92
MOD ( )	Returns remainder of division.	MOD (1600, 300)	100
POWER ( )	Returns the value of one expression raised to the power of another expression.	POWER (3, 2)	9

Solved Exercise :

Find the output of the following SQL statements.

1. SELECT MOD(11,2);
2. SELECT MOD(41,4);
3. SELECT MOD(25,3);
4. SELECT MOD(27,7);
5. SELECT POWER(2,3);
6. SELECT POW(3,4);
7. SELECT POWER(4,.5);
8. SELECT POW(16,.5);
9. SELECT POWER(2.5,3);
10. SELECT SIGN(-20);
11. SELECT SIGN(20);
12. SELECT SIGN(0);
13. SELECT SQRT(81);
14. SELECT SQRT(25);
15. SELECT ROUND(45678.2546,3);
16. SELECT ROUND(45678.2946,2);
17. SELECT ROUND(45678.6676,1);
18. SELECT ROUND(52789.2546,-1);
19. SELECT ROUND(65874.2546,-2);
20. SELECT ROUND(95687.2546,-3);
21. SELECT ROUND(4578.24,-1);
22. SELECT ROUND(25348.1,-2);
23. SELECT ROUND(68475.21,-3);
24. SELECT TRUNCATE(45678.2546,3);
25. SELECT TRUNCATE(45678.2946,2);
26. SELECT TRUNCATE(45678.6676,1);
27. SELECT TRUNCATE(52789.2546,-1);
28. SELECT TRUNCATE(65874.2546,-2);
29. SELECT TRUNCATE(95687.2546,-3);
30. SELECT TRUNCATE(4578.24,-1);
31. SELECT TRUNCATE(25348.1,-2);
32. SELECT TRUNCATE(68475.21,-3);

Answers:

1. 1
2. 1
3. 1
4. 6
5. 8
6. 81
7. 2

8. 4
9. 15.625
10. -1
11. 1
12. 0
13. 9
14. 5
15. 45678.255
16. 45678.29
17. 45678.7
18. 52790
19. 65900
20. 96000
21. 4580
22. 25300
23. 68000
24. 45678.254
25. 45678.29
26. 45678.6
27. 52780
28. 65800
29. 95000
30. 4570
31. 25300
32. 68000

➢ DATE and TIME FUNCTIONS

Date functions operate on values of the DATE data type:

Function	Description	Example	
		Function	Result
SYSDATE ( )	Returns the current database server date and time.	Select SYSDATE ( ) from Dual;	12-Jan-15
CURDATE ( )	Returns the current date.	Select CURDATE ( ) from Dual,	12-Jan-15
NOW ( )	Returns the current date and time.	Select NOW ( ) from Dual	12-Jan-15
DATE ( )	Extracts the date part of a date or datetime expression.	DATE ('2015-1-12')	12
MONTH ( )	Returns the month.	MONTH ('2015-1-12')	1
YEAR ( )	Returns the year.	YEAR ('2015-1-12')	2015
DAYNAME ( )	Returns the name of week day.	DAYNAME ('2015-1-22')	THURSDAY
DAYOFMONTH ( )	Returns the day of month.	DAYOFMONTH ('2015-1-22')	22
DAYOFWEEK ( )	Returns the week day index of the argument.	DAYOFWEEK ('2015-1-22')	5
DAYOFYEAR ( )	Returns the day of year (1-366).	DAYOFYEAR ('2015-01-07')	7

Solved Exercise :

If today"s date is 12th Apr 1995 and time is 10:50:23 AM then what will the following commands display.

1. SELECT CURDATE();
2. SELECT SYSDATE();
3. SELECT NOW();

Find the output of the following SQL statements.

4. SELECT DAYOFMONTH("1995-04-12");
5. SELECT DAY("1995-04-12");
6. SELECT MONTH("1995-04-12");
7. SELECT YEAR("1995-04-12");

8. SELECT DAYOFWEEK("1995-04-12");
9. SELECT DAYOFYEAR("1995-04-12");

10. SELECT DAYNAME("1995-04-12");
11. SELECT MONTHNAME("1995-04-12");

12. SELECT DATE("2022-02-02 12:15:46");

Answers:

1. 2022-02-09
2. 2022-02-09 14:22:17
3. 2022-02-09 14:22:17
4. 12
5. 12
6. 4
7. 1995
8. 4
9. 102
10. Wednesday
11. April
12. 2022-02-02

STRING FUNCTIONS

(a) Case-manipulation Functions

These functions convert case for character strings:

Function	Description	Examples	
		Function	Result
LOWER ( )	Returns the argument in lowercase.	LOWER ('SQL')	sql
UPPER ( )	Returns the argument in uppercase.	UPPER ('Sql')	SQL
INITCAP ( )	Returns the argument's first letter in uppercase and the remaining letters in lowercase.	INITCAP ('SQL')	Sql

(b) Character-manipulation Functions

These functions manipulate character strings:

Function	Description	Examples	
		Function	Result
CHAR ( )	Returns the character for each integer passed.	CHAR (65)	A
CONCAT ( )	Returns concatenated string.	CONCAT ('HELLO', 'WORLD')	HELLOWORLD
SUBSTR ( )	Returns the specified substring.	SUBSTR ('HELLOWORLD', 1,5)	HELLO
INSTR ( )	Finds numeric position of a named character.	INSTR ('HELLO', 'E')	2
LENGTH ( )	Shows the length of a string as a numeric value.	LENGTH ('HELLO')	5
TRIM ( )	Trims leading or trailing characters (or both) from a character string.	TRIM ('H' FROM 'HELLO')	ELLO
LTRIM ( )	Trims leading spaces.	LTRIM (' HELLO')	HELLO
RTRIM ( )	Trims trailing spaces.	RTRIM ('HELLO ')	HELLO
MID ( )	Returns a substring starting from the specified position.	MID ('HELLO', 3)	LLO
LEFT ( )	Returns the left most number of characters as specified.	LEFT ('HELLO' , 1)	H
RIGHT ( )	Returns the right most number of characters as specified.	RIGHT ('HELLO', 2)	LO
ASCII ( )	Returns the ASCII value for each character passed.	ASCII ('A')	65

Solved Exercise :

Find the output of the following SQL statements.

1. SELECT CHAR(69);
2. SELECT CHAR(73,80);
3. SELECT CHAR(80,82,65,67);
4. SELECT CONCAT("MR","ANIL");
5. SELECT CONCAT("MS","AARTI");
6. SELECT LOWER("AAkarsh");
7. SELECT LOWER("KUMAR");
8. SELECT UPPER("AAkarsh");
9. SELECT UPPER("kumar");
10. SELECT LENGTH("KUMAR");
11. SELECT LENGTH("EDUCATION");

12. SELECT LTRIM(“ EDUCATION”);
13. SELECT RTRIM(“EDUCATION ”);
14. SELECT TRIM(“ EDUCATION ”);
15. SELECT INSTR(“EDUCATION”,”CAT”);
16. SELECT INSTR(“SUBSTANTIAL”,”TAN”);
17. SELECT INSTR(“ALOGRITHM”,”LOG”);
18. SELECT INSTR(“BANANA”,”AN”);
19. SELECT INSTR(“MALAYALAM”,”LA”);
20. SELECT INSTR(“EDUCATION”,”COT”);
21. SELECT INSTR(“MALAYALAM”,”LAL”);

22. SELECT LEFT(“MALAYALAM”,4);
23. SELECT LEFT(“SUCESSFUL”,5);
24. SELECT RIGHT(“MALAYALAM”,4);
25. SELECT RIGHT(“SUCESSFUL”,5);
26. SELECT MID(“EDUCATION”,3,4);
27. SELECT MID(“SUBSTANTIAL”,2,5);
28. SELECT SUBSTR(“ALOGRITHM”,5,3);
29. SELECT SUBSTR(“BANANA”,2,2);
30. SELECT SUBSTRING(“MALAYALAM”,6,4);
31. SELECT SUBSTRING (“APPROXIMATELY”,8,5);
32. SELECT RIGHT (LEFT(“APPROXIMATELY”,8),5);
33. SELECT LEFT(RIGHT(“APPROXIMATELY”,8),5);
34. SELECT RIGHT(LEFT(“MALAYALAM”,5),3);
35. SELECT LEFT(RIGHT(“MALAYALAM”,5),3);
36. SELECT MID(LEFT(“APPROXIMATELY”,9),5,3);
37. SELECT MID(RIGHT(“APPROXIMATELY”,10),3,5);

Answers:
1. E
2. IP
3. PRAC
4. MRANIL
5. MSAARTI
6. aakarsh
7. kumar
8. AAKARSH
9. KUMAR
10. 5
11. 9
12. EDUCATION
13. EDUCATION
14. EDUCATION

15. 4
16. 5
17. 2
18. 2
19. 3
20. 0
21. 0
22. MALA
23. SUCES
24. ALAM
25. SSFUL
26. UCAT
27. UBSTA
28. RIT
29. AN
30. ALAM
31. MATEL
32. ROXIM
33. XIMAT
34. LAY
35. YAL
36. OXI
37. XIMAT

Unsolved Exercise:

Find the output of the following SQL commands based on the table "DOCTOR" given below:

DOCTOR

	NAME	LOC	AGE	DEPARTMENT	DATEOFADM	BILL	SEX	DOC_ID
1	ARPIT	RANCHI	60	SURGERY	21-Jan-1998	4562.215	M	101
2	ZARINA	BOKARO	22	ENT	12-Dec-1997	84517.625	F	102
3	KAREEM	CHATRA	32	ORTHOPEDIC	19-Feb-1998	42621.325	M	103
4	ARUN	JAMTARA	36	SURGERY	1-Nov-1996	9568.124	M	104
5	ZUBIN	SILLI	30	ENT	1-Feb-2022	61547.450	M	103
6	KETAKI	PAKUR	16	RADIOLOGY	24-Feb-1998	87421.324	F	102
7	ANKITA	SAHEBGANJ	29	CARDIOLOGY	20-Feb-1999	69658.457	F	103
8	zareen	GARHWA	45	GYNECOLOGY	22-Feb-1998	12478.650	F	104
9	KUSH	LATEHAR	37	CARDIOLOGY	13-Jan-1999	32543.389	M	101
10	SHILPA	RANCHI	23	ENT	21-Feb-2000	6547.647	F	102
11	anita	LATEHAR	51	GYNECOLOGY	15-Apr-1997	65237.487	F	101
12	SHIVAM	SAHEBGANJ	18	SURGERY	25-Sep-1999	2674.987	M	103
13	MAHI	SILLI	24	ENT	23-Oct-1998	26478.99	F	102
14	SURYA	BUNDU	60	RADIOLOGY	15-Dec-2000	6547.658	M	104
15	nitya	CHATRA	47	GYNECOLOGY	03-Mar-1998	4500.500	F	102

1. SELECT MOD(AGE,3) FROM DOCTOR WHERE DEPARTMENT = "ENT";

2. SELECT MOD(AGE,4) FROM DOCTOR WHERE AGE>=45;
3. SELECT SQRT(AGE) FROM DOCTOR WHERE NAME=”ARUN”;
4. SELECT ROUND(BILL,2) FROM DOCTOR WHERE DEPARTMENT=”ENT”;
5. SELECT ROUND(BILL,1) FROM DOCTOR WHERE NAME LIKE “K%”;
6. SELECT ROUND(BILL,-1) FROM DOCTOR WHERE DEPARTMENT=”SURGERY”;
7. SELECT ROUND(BILL,-1) FROM DOCTOR WHERE NAME LIKE “S%”;
8. SELECT ROUND(BILL,-2) FROM DOCTOR WHERE LOC LIKE “B%”;
9. SELECT ROUND(BILL,-1) FROM DOCTOR WHERE AGE BETWEEN 30 AND 40;
10. SELECT TRUNCATE(BILL,1) FROM DOCTOR WHERE AGE BETWEEN 30 AND 40;
11. SELECT TRUNCATE(BILL,-2) FROM DOCTOR WHERE DEPARTMENT=”ENT”;
12. SELECT TRUNCATE(BILL,1) FROM DOCTOR WHERE NAME LIKE “S%”;
13. SELECT TRUNCATE(BILL,-1) FROM DOCTOR WHERE DOC_ID = 101;
14. SELECT DAYOFMONTH(DATEOFADM) FROM DOCTOR WHERE DOC_ID = 101;
15. SELECT DAY(DATEOFADM) FROM DOCTOR WHERE AGE>=45;
16. SELECT MONTH(DATEOFADM) FROM DOCTOR WHERE NAME LIKE “A%”;
17. SELECT MONTH(DATEOFADM) FROM DOCTOR WHERE SEX = “M”;
18. SELECT YEAR(DATEOFADM) FROM DOCTOR WHERE DOC_ID = 102;
19. SELECT YEAR(DATEOFADM) FROM DOCTOR WHERE AGE BETWEEN 35 AND 40;
20. SELECT DAYOFWEEK(DATEOFADM) FROM DOCTOR WHERE NAME = “ZUBIN”;
21. SELECT DAYOFYEAR(DATEOFADM) FROM DOCTOR WHERE NAME = “ZUBIN”;
22. SELECT DAYNAME(DATEOFADM) FROM DOCTOR WHERE NAME = “ZUBIN”;
23. SELECT MONTHNAME(DATEOFADM) FROM DOCTOR WHERE NAME LIKE “A%”;
24. SELECT MONTHNAME(DATEOFADM) FROM DOCTOR WHERE DOC_ID = 102;
25. SELECT MONTHNAME(DATEOFADM) FROM DOCTOR WHERE AGE = 60;
26. SELECT LOWER(NAME) FROM DOCTOR WHERE AGE = 60;
27. SELECT UPPER(NAME) FROM DOCTOR WHERE DEPARTMENT= “GYNECOLOGY”;
28. SELECT LENGTH(LOC) FROM DOCTOR WHERE AGE>=45;
29. SELECT INSTR(NAME,”E”) FROM DOCTOR WHERE DEPARTMENT=”ORTHOPEDIC”;
30. SELECT INSTR(LOC,”GAN”) FROM DOCTOR WHERE NAME =”ANKITA”;
31. SELECT INSTR(DEPARTMENT,”HOP”) FROM DOCTOR WHERE NAME=”KAREEM”;
32. SELECT LEFT(NAME,3) FROM DOCTOR WHERE AGE BETWEEN 30 AND 40;
33. SELECT LEFT(LOC,3) FROM DOCTOR WHERE DEPARTMENT=”ENT”;
34. SELECT RIGHT(NAME,3) FROM DOCTOR WHERE NAME LIKE “S%”;
35. SELECT RIGHT(LOC,3) FROM DOCTOR WHERE DOC_ID = 101;
36. SELECT LEFT(DEPARTMENT,6) FROM DOCTOR WHERE SEX = “M”;
37. SELECT RIGHT(DEPARTMENT,5) FROM DOCTOR WHERE DOC_ID = 102;
38. SELECT MID(LEFT(DEPARTMENT,6),2,3) FROM DOCTOR WHERE AGE BETWEEN 30 AND 40;
39. SELECT MID(LEFT(DEPARTMENT,6),2,3) FROM DOCTOR WHERE DOC_ID = 101;
40. SELECT MID(RIGHT(DEPARTMENT,6),2,3) FROM DOCTOR WHERE DOC_ID = 103;
41. SELECT SUBSTRING(RIGHT(DEPARTMENT,7),3,2) FROM DOCTOR WHERE NAME LIKE “A%”;
42. SELECT SUBSTR(LEFT(DEPARTMENT,6),2,3) FROM DOCTOR WHERE AGE >= 40;
43. SELECT SUBSTR(LEFT(DEPARTMENT,6),2,3) FROM DOCTOR WHERE SEX = “M”;

JOINS

- A relational database consists of multiple related tables linking together using common columns, which are known as foreign key columns.
- **It is used retrieve data from multiple tables.**
- Consider the tables below EMP, DEPT & SALGARDE that stored related information, all the examples on join will be explained with help of these following three tables

EMP Table

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
7369	SMITH	CLERK	7902	1993-06-13	800.00	0.00	20
7499	ALLEN	SALESMAN	7698	1998-08-15	1600.00	300.00	30
7521	WARD	SALESMAN	7698	1996-03-26	1250.00	500.00	30
7566	JONES	MANAGER	7839	1995-10-31	2975.00		20
7698	BLAKE	MANAGER	7839	1992-06-11	2850.00		30
7782	CLARK	MANAGER	7839	1993-05-14	2450.00		10
7788	SCOTT	ANALYST	7566	1996-03-05	3000.00		20
7839	KING	PRESIDENT		1990-06-09	5000.00	0.00	10
7844	TURNER	SALESMAN	7698	1995-06-04	1500.00	0.00	30
7876	ADAMS	CLERK	7788	1999-06-04	1100.00		20
7900	JAMES	CLERK	7698	2000-06-23	950.00		30
7934	MILLER	CLERK	7782	2000-01-21	1300.00		10
7902	FORD	ANALYST	7566	1997-12-05	3000.00		20
7654	MARTIN	SALESMAN	7698	1998-12-05	1250.00	1400.00	30

DEPT TABLE

DEPTNO	DNAME	LOCATION
10	Accounting	New York
20	Research	Dallas
30	Sales	Chicago
40	Operations	Boston

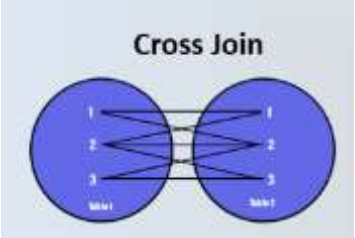
SALGRADE TABLE

GRADE	LOSAL	HISAL
1	700.00	1200.00
2	1201.00	1400.00
4	2001.00	3000.00
5	3001.00	99999.00
3	1401.00	2000.00

- Types of Join
  - Cartesian Product or Cross join
  - Equi Join
  - Natural Join
  - Non-Equi Join
  - Self Join
  - Left Outer Join
  - Right Outer Join

- **Cartesian Product or Cross join**
  - The cross join makes a Cartesian product of rows from the joined tables.
  - The cross join combines each row from the first table with every row from the right table to make the result set.

- If Table1 has degree d1 and cardinality c1 and table2 has degree d2 and cardinality c2, their Cartesian Product has degree d=d1+d2 and cardinality c=c1*c2;



Ven Diagram

Syntax:

```
SELECT * FROM table1, table2;
        OR
SELECT * FROM table1 CROSS JOIN table2;
        OR
SELECT * FROM table1 JOIN table2;
```

e.g. SELECT * FROM emp, dept;
SELECT * FROM emp CROSS JOIN dept;
SELECT * FROM emp JOIN DEPT;

Output:

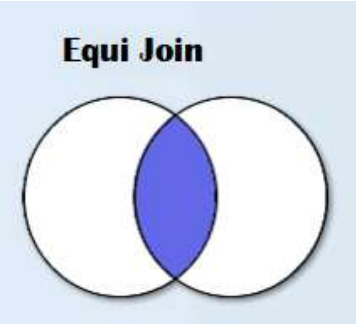
empno	ename	job	mgr	hiredate	sal	comm	deptno	deptno	dname	loc
7369	SMITH	CLERK	7902	1980-12-17	800.00	NULL	20	10	ACCOUNTING	NEW YORK
7369	SMITH	CLERK	7902	1980-12-17	800.00	NULL	20	20	RESEARCH	DALLAS
7369	SMITH	CLERK	7902	1980-12-17	800.00	NULL	20	30	SALES	CHICAGO
7369	SMITH	CLERK	7902	1980-12-17	800.00	NULL	20	40	OPERATIONS	BOSTON
7499	ALLEN	SALESMAN	7698	1981-02-20	1600.00	300.00	30	10	ACCOUNTING	NEW YORK
7499	ALLEN	SALESMAN	7698	1981-02-20	1600.00	300.00	30	20	RESEARCH	DALLAS
7499	ALLEN	SALESMAN	7698	1981-02-20	1600.00	300.00	30	30	SALES	CHICAGO
7499	ALLEN	SALESMAN	7698	1981-02-20	1600.00	300.00	30	40	OPERATIONS	BOSTON
7521	WARD	SALESMAN	7698	1981-02-22	1250.00	500.00	30	10	ACCOUNTING	NEW YORK
7521	WARD	SALESMAN	7698	1981-02-22	1250.00	500.00	30	20	RESEARCH	DALLAS
7521	WARD	SALESMAN	7698	1981-02-22	1250.00	500.00	30	30	SALES	CHICAGO
7521	WARD	SALESMAN	7698	1981-02-22	1250.00	500.00	30	40	OPERATIONS	BOSTON
7566	JONES	MANAGER	7839	1981-04-02	2975.00	NULL	20	10	ACCOUNTING	NEW YORK
7566	JONES	MANAGER	7839	1981-04-02	2975.00	NULL	20	20	RESEARCH	DALLAS
7566	JONES	MANAGER	7839	1981-04-02	2975.00	NULL	20	30	SALES	CHICAGO
7566	JONES	MANAGER	7839	1981-04-02	2975.00	NULL	20	40	OPERATIONS	BOSTON
7654	MARTIN	SALESMAN	7698	1981-09-28	1250.00	1400.00	30	10	ACCOUNTING	NEW YORK
7654	MARTIN	SALESMAN	7698	1981-09-28	1250.00	1400.00	30	20	RESEARCH	DALLAS
7654	MARTIN	SALESMAN	7698	1981-09-28	1250.00	1400.00	30	30	SALES	CHICAGO
7654	MARTIN	SALESMAN	7698	1981-09-28	1250.00	1400.00	30	40	OPERATIONS	BOSTON
7698	BLAKE	MANAGER	7839	1981-05-01	2850.00	NULL	30	10	ACCOUNTING	NEW YORK
7698	BLAKE	MANAGER	7839	1981-05-01	2850.00	NULL	30	20	RESEARCH	DALLAS
7698	BLAKE	MANAGER	7839	1981-05-01	2850.00	NULL	30	30	SALES	CHICAGO
7698	BLAKE	MANAGER	7839	1981-05-01	2850.00	NULL	30	40	OPERATIONS	BOSTON
7782	CLARK	MANAGER	7839	1981-06-09	2450.00	NULL	10	10	ACCOUNTING	NEW YORK
7782	CLARK	MANAGER	7839	1981-06-09	2450.00	NULL	10	20	RESEARCH	DALLAS
7782	CLARK	MANAGER	7839	1981-06-09	2450.00	NULL	10	30	SALES	CHICAGO
7782	CLARK	MANAGER	7839	1981-06-09	2450.00	NULL	10	40	OPERATIONS	BOSTON
7788	SCOTT	ANALYST	7566	1982-12-09	3000.00	NULL	20	10	ACCOUNTING	NEW YORK
7788	SCOTT	ANALYST	7566	1982-12-09	3000.00	NULL	20	20	RESEARCH	DALLAS
7788	SCOTT	ANALYST	7566	1982-12-09	3000.00	NULL	20	30	SALES	CHICAGO
7788	SCOTT	ANALYST	7566	1982-12-09	3000.00	NULL	20	40	OPERATIONS	BOSTON

7839	KING	PRESIDENT	NULL	1981-11-17	5000.00	NULL	10	10	ACCOUNTING	NEW YORK
7839	KING	PRESIDENT	NULL	1981-11-17	5000.00	NULL	10	20	RESEARCH	DALLAS
7839	KING	PRESIDENT	NULL	1981-11-17	5000.00	NULL	10	30	SALES	CHICAGO
7839	KING	PRESIDENT	NULL	1981-11-17	5000.00	NULL	10	40	OPERATIONS	BOSTON
7844	TURNER	SALESMAN	7698	1981-09-08	1500.00	0.00	30	10	ACCOUNTING	NEW YORK
7844	TURNER	SALESMAN	7698	1981-09-08	1500.00	0.00	30	20	RESEARCH	DALLAS
7844	TURNER	SALESMAN	7698	1981-09-08	1500.00	0.00	30	30	SALES	CHICAGO
7844	TURNER	SALESMAN	7698	1981-09-08	1500.00	0.00	30	40	OPERATIONS	BOSTON
7876	ADAMS	CLERK	7788	1983-01-12	1100.00	NULL	20	10	ACCOUNTING	NEW YORK
7876	ADAMS	CLERK	7788	1983-01-12	1100.00	NULL	20	20	RESEARCH	DALLAS
7876	ADAMS	CLERK	7788	1983-01-12	1100.00	NULL	20	30	SALES	CHICAGO
7876	ADAMS	CLERK	7788	1983-01-12	1100.00	NULL	20	40	OPERATIONS	BOSTON
7900	JAMES	CLERK	7698	1981-12-03	950.00	NULL	30	10	ACCOUNTING	NEW YORK
7900	JAMES	CLERK	7698	1981-12-03	950.00	NULL	30	20	RESEARCH	DALLAS
7900	JAMES	CLERK	7698	1981-12-03	950.00	NULL	30	30	SALES	CHICAGO
7900	JAMES	CLERK	7698	1981-12-03	950.00	NULL	30	40	OPERATIONS	BOSTON
7902	FORD	ANALYST	7566	1981-12-03	3000.00	NULL	20	10	ACCOUNTING	NEW YORK
7902	FORD	ANALYST	7566	1981-12-03	3000.00	NULL	20	20	RESEARCH	DALLAS
7902	FORD	ANALYST	7566	1981-12-03	3000.00	NULL	20	30	SALES	CHICAGO
7902	FORD	ANALYST	7566	1981-12-03	3000.00	NULL	20	40	OPERATIONS	BOSTON
7934	MILLER	CLERK	7782	1982-01-23	1300.00	NULL	10	10	ACCOUNTING	NEW YORK
7934	MILLER	CLERK	7782	1982-01-23	1300.00	NULL	10	20	RESEARCH	DALLAS
7934	MILLER	CLERK	7782	1982-01-23	1300.00	NULL	10	30	SALES	CHICAGO
7934	MILLER	CLERK	7782	1982-01-23	1300.00	NULL	10	40	OPERATIONS	BOSTON

56 rows in set (0.02 sec)

Note: **Above output has 56 row (14*4) and 11 columns (8+3)**

- **Equi Join-**
  - It performs a JOIN against equality or matching column(s) values of the associated tables.



Ven Diagram

Syntax
SELECT * / Column_list
FROM Table1, Table 2
WHERE table1.column=Table2.column;
OR
SELECT * / Column_list
FROM Table1 join Table2 on Table1.Column=Table2.Column;

Example: SELECT * FROM emp JOIN dept ON emp.deptno=dept.deptno;
OR
SELECT * FROM emp, dept
WHERE emp.deptno=dept.deptno;

Output:

empno	ename	job	mgr	hiredate	sal	comm	deptno	deptno	dname	loc
7369	SMITH	CLERK	7902	1980-12-17	800.00	NULL	20	20	RESEARCH	DALLAS
7499	ALLEN	SALESMAN	7698	1981-02-20	1600.00	300.00	30	30	SALES	CHICAGO
7521	WARD	SALESMAN	7698	1981-02-22	1250.00	500.00	30	30	SALES	CHICAGO
7566	JONES	MANAGER	7839	1981-04-02	2975.00	NULL	20	20	RESEARCH	DALLAS
7654	MARTIN	SALESMAN	7698	1981-09-28	1250.00	1400.00	30	30	SALES	CHICAGO
7698	BLAKE	MANAGER	7839	1981-05-01	2850.00	NULL	30	30	SALES	CHICAGO
7782	CLARK	MANAGER	7839	1981-06-09	2450.00	NULL	10	10	ACCOUNTING	NEW YORK
7788	SCOTT	ANALYST	7566	1982-12-09	3000.00	NULL	20	20	RESEARCH	DALLAS
7839	KING	PRESIDENT	NULL	1981-11-17	5000.00	NULL	10	10	ACCOUNTING	NEW YORK
7844	TURNER	SALESMAN	7698	1981-09-08	1500.00	0.00	30	30	SALES	CHICAGO
7876	ADAMS	CLERK	7788	1983-01-12	1100.00	NULL	20	20	RESEARCH	DALLAS
7900	JAMES	CLERK	7698	1981-12-03	950.00	NULL	30	30	SALES	CHICAGO
7902	FORD	ANALYST	7566	1981-12-03	3000.00	NULL	20	20	RESEARCH	DALLAS
7934	MILLER	CLERK	7782	1982-01-23	1300.00	NULL	10	10	ACCOUNTING	NEW YORK

Example 1: Display the employee name, sal and name of department name

Ans: In the above query ename and sal belong to emp table whereas dname belongs to DEPT table. So, to retrieve data in this we will use join

```
SELECT emp.ename, emp.sal, dept.dname
FROM emp, dept WHERE emp.deptno=dept.deptno;
```

Output:

ename	sal	dname
SMITH	800.00	RESEARCH
ALLEN	1600.00	SALES
WARD	1250.00	SALES
JONES	2975.00	RESEARCH
MARTIN	1250.00	SALES
BLAKE	2850.00	SALES
CLARK	2450.00	ACCOUNTING
SCOTT	3000.00	RESEARCH
KING	5000.00	ACCOUNTING
TURNER	1500.00	SALES
ADAMS	1100.00	RESEARCH
JAMES	950.00	SALES
FORD	3000.00	RESEARCH
MILLER	1300.00	ACCOUNTING

Note:

- In case of join **full qualified (table_name.column_name)** name is used to avoid ambiguity as both table contains common columns as PRIMARY KEY and FOREIGN KEY.
- Table Alias – Like column alias table alias can be used in case of join as given below.
  SELECT e.ename, e.sal FROM emp e, dept d WHERE emp.deptno=dept.deptno;
- Here ‘e’ & ‘d’ are table alias for EMP & DEPT table respectively.

- **Natural Join**
  - A natural join is a type of join operation that creates an implicit join by combining tables based on columns with the same name and data type.
  - It makes the SELECT query simpler with minimal use of conditions.
  - There is no need to specify the name of common column in the SELECT statement.
  - Common column is present only once in the output.

Syntax:

```
SELECT */column_list
FROM Table1 NATURAL JOIN TABLE2;
                OR
SELECT * / column_list
FROM TABLE1 join on common_column_name;
```

Example :

```
SELECT * FROM emp JOIN dept USING(deptno);
                OR
SELECT * FROM emp NATURAL JOIN DEPT;
```

Output:

```
+--------+-------+--------+-----------+------+------------+---------+---------+------------+----------+
| deptno | empno | ename  | job       | mgr  | hiredate   | sal     | comm    | dname      | loc      |
+--------+-------+--------+-----------+------+------------+---------+---------+------------+----------+
|    20 |  7369 | SMITH  | CLERK     | 7902 | 1980-12-17 |  800.00 |    NULL | RESEARCH   | DALLAS   |
|    30 |  7499 | ALLEN  | SALESMAN  | 7698 | 1981-02-20 | 1600.00 |  300.00 | SALES      | CHICAGO  |
|    30 |  7521 | WARD   | SALESMAN  | 7698 | 1981-02-22 | 1250.00 |  500.00 | SALES      | CHICAGO  |
|    20 |  7566 | JONES  | MANAGER   | 7839 | 1981-04-02 | 2975.00 |    NULL | RESEARCH   | DALLAS   |
|    30 |  7654 | MARTIN | SALESMAN  | 7698 | 1981-09-28 | 1250.00 | 1400.00 | SALES      | CHICAGO  |
|    30 |  7698 | BLAKE  | MANAGER   | 7839 | 1981-05-01 | 2850.00 |    NULL | SALES      | CHICAGO  |
|    10 |  7782 | CLARK  | MANAGER   | 7839 | 1981-06-09 | 2450.00 |    NULL | ACCOUNTING | NEW YORK |
|    20 |  7788 | SCOTT  | ANALYST   | 7566 | 1982-12-09 | 3000.00 |    NULL | RESEARCH   | DALLAS   |
|    10 |  7839 | KING   | PRESIDENT | NULL | 1981-11-17 | 5000.00 |    NULL | ACCOUNTING | NEW YORK |
|    30 |  7844 | TURNER | SALESMAN  | 7698 | 1981-09-08 | 1500.00 |    0.00 | SALES      | CHICAGO  |
|    20 |  7876 | ADAMS  | CLERK     | 7788 | 1983-01-12 | 1100.00 |    NULL | RESEARCH   | DALLAS   |
|    30 |  7900 | JAMES  | CLERK     | 7698 | 1981-12-03 |  950.00 |    NULL | SALES      | CHICAGO  |
|    20 |  7902 | FORD   | ANALYST   | 7566 | 1981-12-03 | 3000.00 |    NULL | RESEARCH   | DALLAS   |
|    10 |  7934 | MILLER | CLERK     | 7782 | 1982-01-23 | 1300.00 |    NULL | ACCOUNTING | NEW YORK |
+--------+-------+--------+-----------+------+------------+---------+---------+------------+----------+
```

Example

Consider the following tables – Bank_Account and Branch:

Table: Bank_Account

ACode	Name	Type
A01	Amrita	Savings
A02	Parthodas	Current
A03	Miraben	Current

Table: Branch

ACode	City
A01	Delhi
A02	Mumbai
A01	Nagpur

SELECT * FROM Bank_Account NATURAL JOIN Branch;

Acode	Name	Type	City
A01	Amrita	Savings	Delhi
A01	Amrita	Savings	Nagpur
A02	Parthodas	Current	Mumbai

- Difference between Equi-Join vs Natural Join

EQUI-JOIN	NATURAL JOIN
Join performed on equality of value of the columns	Join is performed on column having common name.
Where clause is used to specify the condition	There is no need to use where clause
Both columns from tables are displayed in the result.	Common column is displayed only once

SOLVED EXERCISE ON JOIN:

Consider the following two table Customer and Saleman

SALESMAN TABLE:

SALESMAN_ID	NAME	CITY	COMMISSION
5001	James Hoog	New York	0.15
5002	Nail Knite	Paris	0.13
5005	Pit Alex	London	0.11
5006	Mc Lyon	Paris	0.14
5007	Paul Adam	Rome	0.13

CUSTOMER TABLE:

CUSTOMER_ID	CUST_NAME	CITY	GRADE	SALESMAN_ID
3002	Nick Rimando	New York	100	5001
3007	Brad Davis	New York	200	5001
3005	Graham Zusi	California	200	5002
3008	Julian Green	London	300	5002
3004	Fabian Johnson	Paris	300	5006
3009	Geoff Cameron	Berlin	100	5003
3003	Jozy Altidor	Moscow	200	5007

Q1. Write a SQL query to display Salesman, cust_name and city from above table where the salesperson and customer belongs to same city.

Ans:

```
SELECT s.name AS "Salesman", c.cust_name, c.city
FROM salesman s, customer c WHERE s.city=c.city;
                    OR
SELECT salesman.name AS "Salesman", customer.cust_name, customer.city
FROM salesman,customer WHERE salesman.city=customer.city;
```

Q2. write a SQL query to display ord_no, purch_amt, cust_name, city of those orders where order amount exists between 500 and 2000.

Ans:

```
SELECT  o.ord_no,o.purch_amt, c.cust_name,c.city
FROM orders o,customer c
```

WHERE o.customer_id=c.customer_id AND o.purch_amt BETWEEN 500 AND 2000;

Q3. Write a SQL query to display Customer Name, city, Salesman, commission the all salesperson(s) and their respective the customer(s).
Ans:

SELECT c.cust_name AS "Customer Name", c.city, s.name AS "Salesman", s.commission FROM customer c, salesman s WHERE c.salesman_id=s.salesman_id;

Q4. What are Joins in SQL?
Ans. Joins in SQL are the statements or clauses using which we can combine two or more tables, based on some common fields present among the tables.

Q5.Explain the different types of Joins in SQL?
Ans: Some of the major joins in SQL are-

- Equi Join – Inner join is used to return the records which are having matching values in both the tables.
- Left Join – Left join is used to concatenate all the rows of the left table and the matching rows in the right table.
- Right Join-Right join is used to concatenate all the rows of the right table and the matching rows in the left table.
- Full Join-Full join is used to return all the records of both the tables as long as there is a matching record in either table.
- Self Join-Self join is a join that is used to join a table to itself. In a self-join, a table is considered as if it were two tables.
- Cartesian Join-Cartesian join is used to return the number of rows in the first table multiplied by the number of rows in the second table. It is also referred to as cross join.

Q6. What is Natural Join?
Ans: Natural join is used to create an implicit join clause based on the value of common attributes in the two tables. Common attributes are the attributes that have the same name in both tables. Natural join does not need any comparison operator as in the case of equi join.

Q7. What is an Equi Join?
Ans: An Equi Join is a type of join that combines tables based on matching values in the specified columns.

- The column names do not need to be the same.
- The resultant table can contain repeated columns.
- It is possible to perform an equi join on more than two tables.

Q 8. What is the difference between cross join and natural join?
Ans: A cross join produces a cross product or cartesian product of two tables whereas the natural join is based on all the columns having the same name and data types in both the tables.

UNSOLVED EXERCISE ON JOIN:

1. Consider the tables PARTICIPANT and ACTIVITY and answer questions that follow :

Table : PARTICIPANT

ADMNO	NAME	HOUSE	ACTIVITYCODE
6473	Kapil Shah	Gandhi	A105
7134	Joy Mathew	Bose	A101
8786	Saba Arora	Gandhi	A102
6477	Kapil Shah	Bose	A101
7658	Faizal Ahmed	Bhagat	A104

Table : ACTIVITY

ACTIVITYCODE	ACTIVITYNAME	POINTS
A101	Running	200
A102	Hopping bag	300
A103	Skipping	200
A104	Bean bag	250
A105	Obstacle	350

To display Names of Participants, Activity Code, Activity Name in alphabetic ascending order of names of participants. Also write your output. How many records were found when the above tables were used in an equi-join?

2. In a database there are two tables 'LOAN' and 'BORROWER' as shown below:

Table: LOAN

LOAN_NO	BRANCH_NAME	AMOUNT
K-70	Downtown	5000
K-230	Redwood	6000
K-260	Perryridge	3700

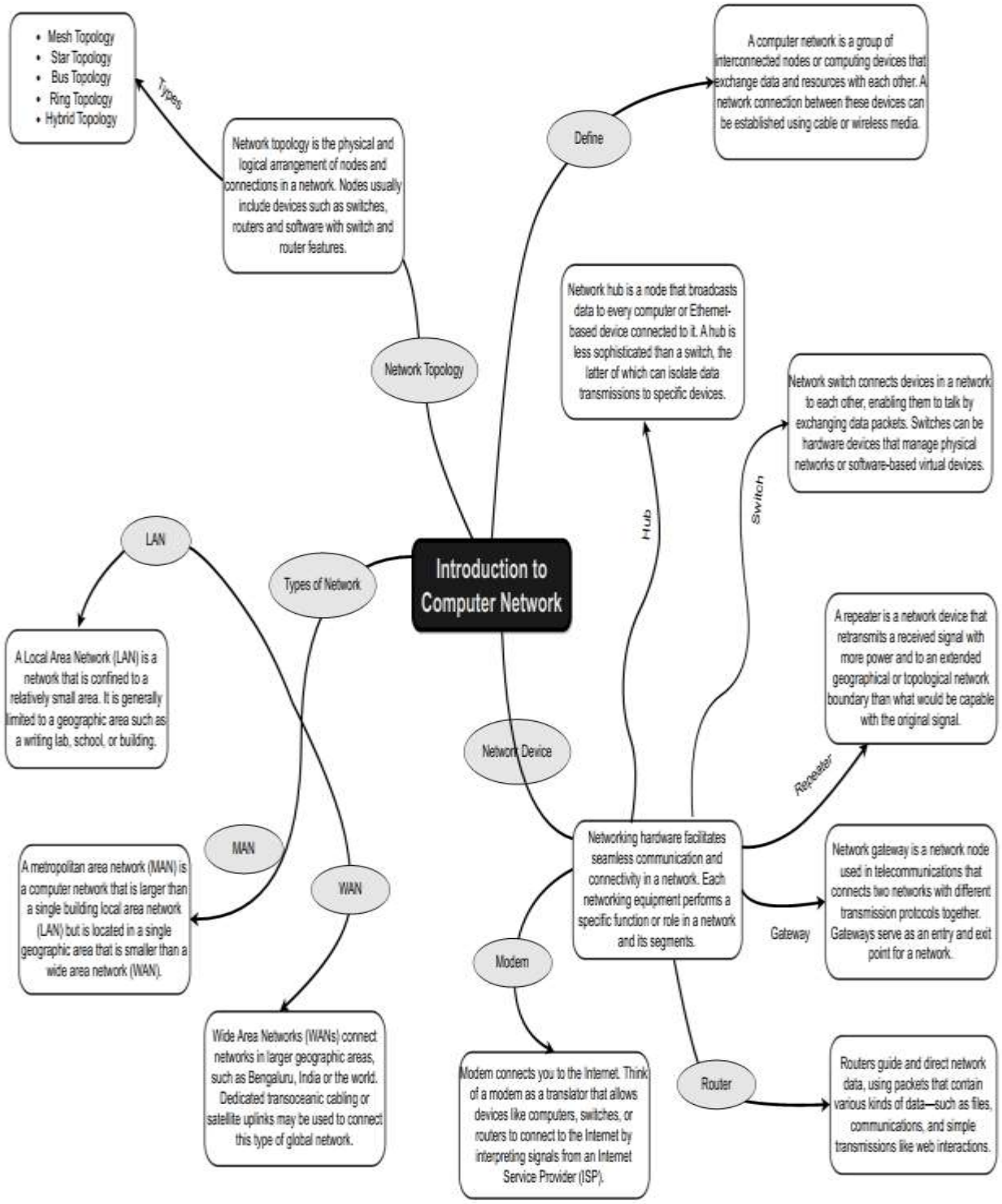
Table : BORROWER

CUSTOMER_NAME	LOAN_NO
Jones	K-170
Smith	K-230
Hayes	K-155

(iii) How many rows and columns will be there in the cross join of these two tables?

(iv) Write sql query to display customer smith's loan record with their Name, Loan _no, Branch _name and Loan amount.

Introduction to Computer Networks:



Evolution of Networking:

Network: - To connect the more than one device via a medium, is called network.

Why do we need network?

1. Fast and Secure Communication
2. Resource sharing
3. Reduce Cost

Evolution of Network:

- **ARPANET:** The Advanced Research Projects Agency Network (**ARPANET**) was an early packet switching network and the first network to implement the protocol suite TCP/IP. Both technologies became the technical foundation of the Internet.
- **NSFNET**:
  - The **N**ational **S**cience **F**oundation **Net**work (NSFNET) was a program created and funded by the National Science Foundation to coordinate and promote advanced research and education in networking in the United States.
  - NSFNET was founded in 1985. NSFNET was a general purpose research network wherein the connection is not limited to the super computer centers, it was to serve as a backbone connection for regional networks at every supercomputing site and use ARPANET's TCP/IP protocol.
  - In 1986, the super computer centers were officially connected and it became open to all academic networks.
  - The NSF decided to transfer the operations of NSFNET to the private sector in the midst of the rapid growth of the network. NSFNET was officially dissolved on October 30, 1995.
- **Internet:** A network of networks.
- **WWW**: World Wide Web started on 6th August 1991, started by ' ***Berners Lee'*** .
- **Interspace:** Interspace is a client/server software program that allows multiple users to communicate online with real-time audio, video and text chat in dynamic 3D

environments. Interspace provides the most advanced form of communication available on the Internet today.

- **Communication Terminologies:**

  - **Channel**: A channel is a separate path through which signals can flow. A channel has a certain capacity for transmitting information, often measured by its bandwidth in Hz or its data rate in bits per second.

  - **Bandwidth**: Bandwidth refers to the amount of information that can be transmitted over a network in a given amount of time, usually expressed in bits per second or bps.

  - **Data Transfer Rate:** The data transfer rate of a computer network connection is normally measured in units of bits per second (bps).
    Larger units are *Kbps*, *Mbps* and *Gbps*, KBps, MBps,GBps

    bps means bits per second.

    Bps means Byte per second

    1 kilobit per second (Kbps) = 1000 bits per second (bps).
    1 megabit per second (Mbps) = 1000 Kbps or 1000^2 bps.
    1 gigabit per second (Gbps) = 1000 Mbps
    1 Terabit per second (Tbps) = 1000 Gbps

- **Types of Networks:**

  - Local Area Network (LAN)
  - Metropolitan Area Network (MAN)
  - Wide Area Network (WAN)
  - Personal Area Network (PAN)

- **LAN**:

  - Use in small local area, like in an institute or an organization.
  - Devices are connected via physical medium.
  - Limited distance, up to 150 Meter.
  - Example – **Intranet**

- **MAN**:

  - Larger than LAN.
  - Used in Metropolitan cities.
  - Range up to 50 KM.

- **WAN**:

- It connects device across globe.
- It uses public network
- Internet
- BSNL
- VSNL

- **PAN**:

  - For very small distance
  - Private Communication
  - Example: **Bluetooth**

- **Network Devices:**

  - Modem
  - Hub
  - Switch
  - Gateway
  - Bridge
  - Router
  - Repeater

- **Modem**:

  - The full form of modem is Modulator and demodulator**.**
  - A modem is a device or program that enables a computer to transmit data over telephone or cable lines.
  - A modem converts analog signal to digital signal and vice- versa.
  - Modem connects computer to internet.
  - There are two types of modems:
    - Internal Modem
    - External Modem

- **Hub**:

  - A network device that contains multiple ports.
  - Provides multiple connections.
  - When a packet arrives at one port, it is copied to the other ports so that all segments of the LAN can see all packets.
  - There are two types of hubs:

    a. Active Hub:
       - It strengthens the signal may boost noise too.
       - It needs electricity.

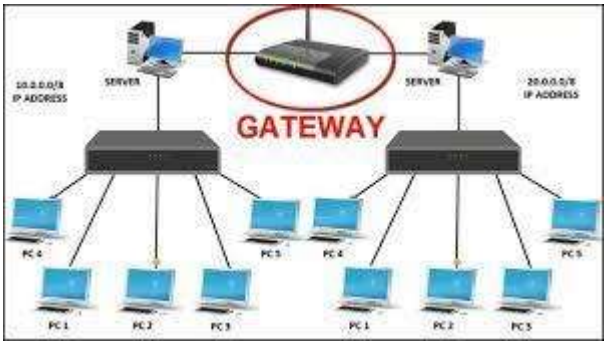
    b. Passive Hub
       - It repeat/copy signals.
       - It does not need electricity

- **Switch**:

  - A switch is called *smart hub.*
  - Provides multiple connections
  - A device that filters and forwards packets between LAN segments.

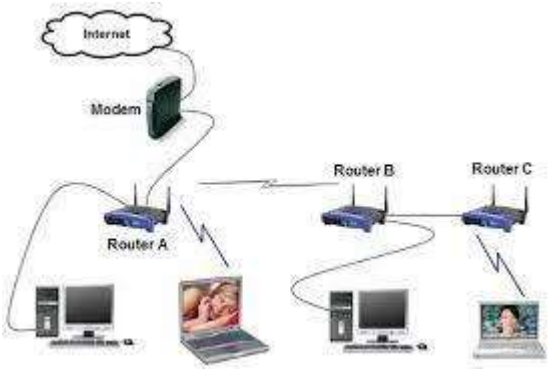
- **Gateway**:

  - A gateway is a network point that acts as an entrance to another network.
  - Used to connect two dissimilar networks



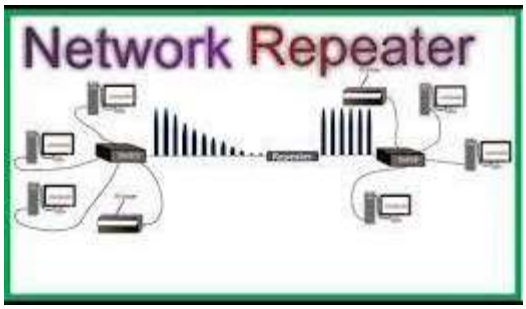
- **Router**:

  - A router is a device that forwards data packets along networks. A router is connected to at least two networks, commonly two LANs or WANs. Routers are located at gateways, the places where two or more networks connect.
  - A router acts as a dispatcher, choosing the best path for information to travel so it's received quickly.

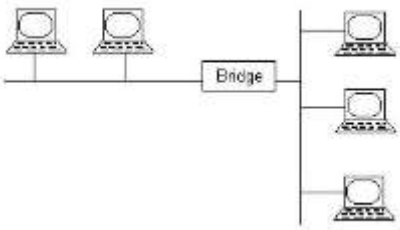


- **Repeater:**

  Network repeaters regenerate and amplify the weak signals to transmit the information for long distance.



- **Bridge**:
  - It connects multiple network segments having same protocol
  - It works at Data Link Layer (Layer 2).
  - Bridge does not simply broadcast traffic from one network.
  - Bridges use bridge table to send frames across network segments.
  - It also improves the overall network performance.



- **RJ45**

  - It stands for Registered Jack.
  - It is common interface to connect Twisted Pair Cable.
  - It is used for Ethernet and Token Ring Network.

- **Ethernet Card**
  - It also known as NIC card.
  - It enables a computer to access an Ethernet network (LAN)

- It has MAC id which gives it unique identity in the network.

- **Wi-Fi card**
  - It is also known wireless network adaptor.
  - It is a wireless network technology that allows devices to communicate over wireless signals.
  - It uses radio waves for the communication

- **Difference between Router and Switch**
  - A network switch forwards data packets between groups of devices in the same network, whereas a router forwards data between different networks**.**

- **Difference between a Router and a Modem**
  - A router forms networks and manages the flow of data within and between those networks, while a modem connects those networks to the Internet.
- **Difference between a Router and Gateway**
  - A gateway is a concept while a router is a device that implements a gateway.

Router	Gateway
It ensures that data packets are switched to the right address with the best route.	To connect two networks of different protocols as a translator
It routes the data packets via similar networks	It connects two dissimilar networks
It supports dynamic Routing.	It does not support dynamic Routing.

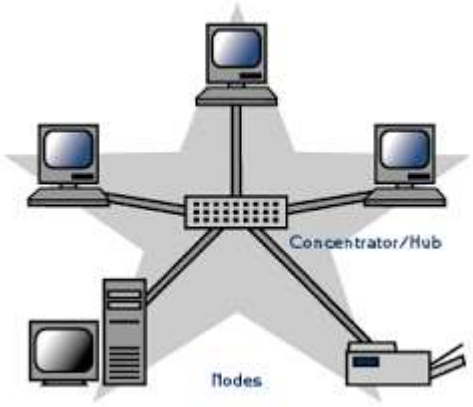
- **Network Topologies:**

The term Topology refers to the way/layout in which the various nodes or computers of a network are linked together.

The following factors are considered while selecting a topology:

- Cost
- Reliability
- Bandwidth Capacity
- Ease of installation
- Ease of Troubleshooting

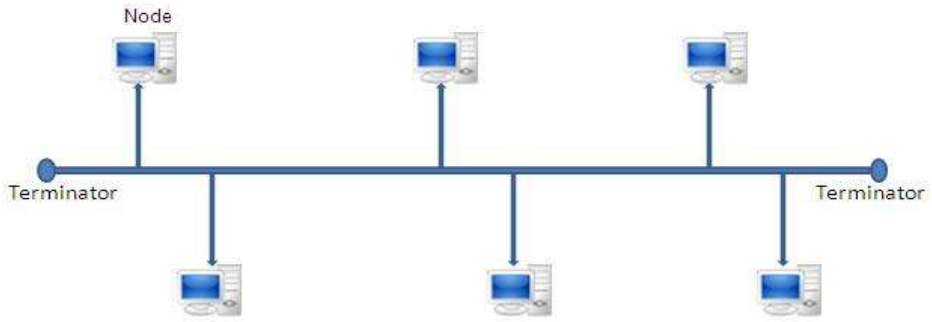
- **Star Topology**
  - The physical star Topology uses a central controlling hub with dedicated legs pointing in all directions – like points of a star.
  - Each network device has a dedicated point-to-oint link to the central hub.
  - There is no direct link between these computers and the computers can communicate via central controller only.



Advantages	Disadvantages
▪ Reliable ▪ Robust ▪ Failure of node does not affect the working of the network. ▪ Fault detection and isolation is easy. ▪ Maintenance of the network is easy. ▪ It doesn't create bottlenecks where data collisions occur.	▪ Central node dependency. If the central controller or hub fails, entire system collapses. ▪ Cabling cost is more ▪ Difficult to install

- **Bus Topology**
  - It consists of one continuous length of cable (trunk) that is shared by all the nodes in the network and a terminating resistor (terminator) at each end that absorbs the signal when it reaches the end of line.

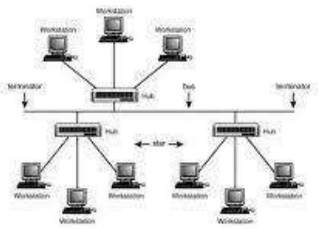
- Without a terminator the electrical signal would reach the end of copper wire and bounce back, causing errors on the network.
- Data communication message travels along the bus in both directions until it is picked up by a workstation or server NIC.
- If the message is missed or not recognized, it reaches the end of the cabling and dissipates at the terminator. Bus Network Topology requires a multipoint connection.



Advantages	Disadvantages
▪ Easy to install and the use for small networks. ▪ Requires less cable ▪ Failure of one node does not affect the network functioning. ▪ Cost is less ▪ New node can be easily added	▪ If the main cable fails the entire network collapses. ▪ Difficult to reconfigure, due to more connections. ▪ Difficult to troubleshoot ▪ Slow, due to traffic on single cable ▪ Only one device transmits at a time, other devices wait for their turn.

➢ **Tree Topology**

- This topology has Hierarchical structure.
- This topology connects the node via hubs. Hub, which is present at top level, is called **root hub** or **active hub**.
- Another hub is called secondary hub or passive hub.

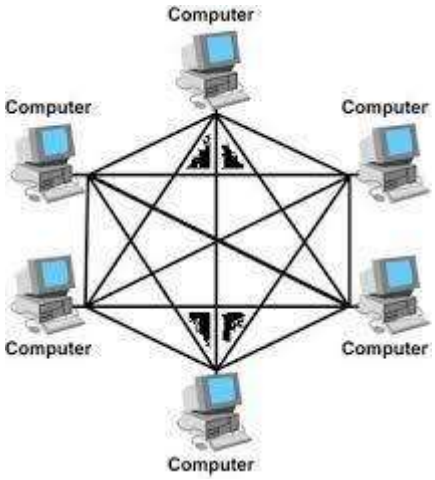


TREE TOPOLOGY

Advantages	Disadvantages
▪ New node can be added easily. ▪ Signal can travel for long distance. ▪ Isolate and prioritize communication.	▪ If the backbone line breaks, the entire segment goes down. ▪ More difficult to configure ▪ Higher cabling cost

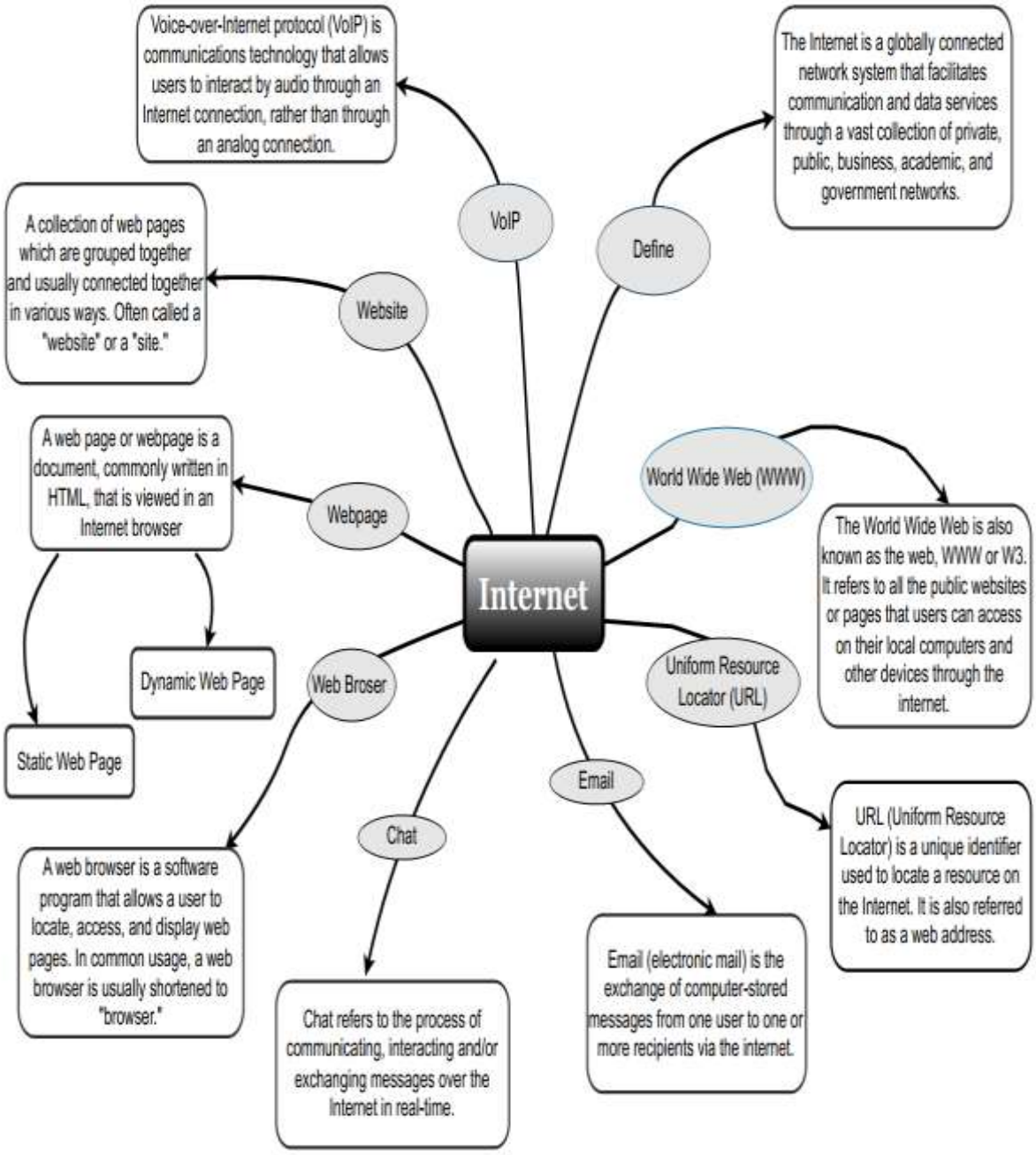
➢ **Mesh Topology:**

- ❖ In mesh each node is connected to every other node in the network i.e. each node has a dedicated point to point link to every other node as shown.
- ❖ Dedicated means that the link carries the traffic only between two devices it connects.
- ❖ In this way there exist multiple paths between two nodes of the network.
- ❖ In case of failure of one path, the other one can be used.



Advantages	Disadvantages
▪ It is robust. Failure of one node does not collapse the entire system. ▪ No traffic congestion ▪ Dedicated links ensure faster transmission ▪ Point to point links makes fault identification and isolation easier.	▪ Network installation and reconfiguration difficult. ▪ High cabling cost. If there are n nodes in the network then each node has (n-1) connections.

INTRODUCTION TO INTERNET



- **The Internet**: it is the global network of computing devices including desktop, laptop, servers, tablets, mobile phones, other handheld devices as well as peripheral devices such as printers, scanners, etc.

- **URL**:
  - URL stands for Uniform Resource Locator. It is a unique identifier used to locate a resource on the internet. It is also referred to as web address.
  - The URL contains: -
    - the name of the protocol to be used to access the file resource
    - a domain name that identifies a specific computer on the internet
    - a pathname with hierarchical description that specifies the location of a file on that computer.

- **Applications of Internet:**

  Following are some of the broad areas or services provided through Internet:

  - World Wide Web (WWW)
  - Electronic mail (Email)
  - Chat
  - Voice Over Internet Protocol (VoIP)

- **World Wide Web**:
  - It is a way of exchanging information between computers on the internet, tying them together into vast collection of interactive multimedia resources.
  - its development began in 1989 by Tim Berners Lee and his colleagues at CERN

- **Differences between internet and www**

Internet	www
It is a global network of networks	It is a way of accessing information over the medium of internet
It connects millions of computers together globally	It is information sharing model that is built on the top of internet
It does not utilize web browsers. It is the structure on which www is based	It utilizes web browsers
It does not use any protocols	The web uses http protocol to transmit data.

- **Advantages of World Wide Web are**:
  - Information can be accessed from anywhere.
  - It provides rapid interaction communication which can be used for different services.
- **E-mail (Electronic Mail)** :

- It is a paperless method of sending messages, notes, pictures and even sound files from one place to another using the internet as a medium.
- E-mail address is an individual name which is used to send and receive email on the internet.
- E-mail address is used to identify source and destination of an email message.
- the format of email address is:

  username@domainname

  where username is a unique username and domain name identifies the mail server. eg xyz@gmail.com

- **Components of email message**:
  - To field used to provide the email address of the receiver
  - from field to specify address of the sender
  - subject specifies the purpose of the email
  - Cc stands for carbon copy, this field is optional
  - Bcc stands for blind carbon copy. Here the recipients do not get to know who the other recipients are.
  - Message Body
- **Most commonly used email protocols on internet** are:
  - POP3(Post Office Protocol):
    - It is a standard mail protocol used to receive emails from remote server to a local client
  - IMAP (Internet Message Access Protocol):
    - It is a mail protocol for accessing email on a remote web server from a local client.
  - SMTP (Simple Mail Transfer Protocol):
    - Protocol for sending emails across internet

- **Chat**:
  - Chatting or instant messaging over the Internet means communicating to people at different geographic locations in real time through text message(s).
  - Applications such as WhatsApp, Snapchat, Skype, Yahoo Messenger, Google Talk, Facebook Messenger, Google Hangout, etc., are examples of instant messengers.
- **VoIP (Voice over Internet Protocol):**
  - It is a technology that allows you to make voice calls using a broadband internet connection instead of a regular phone line.
  - VoIP works on the simple principle of converting the analogue voice signals into digital and then transmitting them over the broadband line.
  - These services are either free or very economical
  -
- **Website**: Collection of web pages which are interlinked to one another. These pages are hypertext pages and link between pages is known as hyperlink.

- **Webpages**:

  - It is a document written in HTML) viewed on internet browser
  - It is used show information to viewers. It may contain text, graphics, hyperlinks to other web pages and files.
  - First main page of a website is called as Home Page.
  - Types of web pages:
    - Static Web Page
    - Dynamic Web Page

- **Difference between Static and Dynamic webpage:**

Static Web Page	Dynamic Web Page
The static web pages display the same content each time when someone visits it.	In the dynamic Web pages, the page content changes according to the user.
It takes less time to load over internet.	Dynamic web pages take more time while loading
No Database used.	A database is used in at the server end in a dynamic webpage.
Changes very rarely	Changes very frequently

- **Difference between website and webpage:**

Website	Webpage
It is a cluster of related webpages addressed to a typical URL	It is a part of website which comprises links to other web pages
There is no extension used in the URL of a website	The webpage URL has an extension.

- **Web Hosting:**

  - online service that enables user to publish website or web application on the internet.
  - When user sign up for a hosting service, basically rent some space on a server on which user can store all the files and data necessary for website to work properly.
  - A server is a physical computer that runs without any interruption so that website is available all the time for anyone who wants to see it.

- **Web Browser:**

  - it is a special software that enables the user to read/view the web page and jump from one web page to another.
  - Mosaic was the first web browser developed by the National Centre for Supercomputing Application (NCSA).
  - eg. Microsoft Edge,Mozilla Firefox, Google Chrome,Opera etc.

- **Browser Setting:**

  - Every web browser has got certain settings that define the manner in which the browser will behave.
  - These settings may be with respect to privacy, search engine preferences, download options, auto signature, autofill and autocomplete feature, theme and much more.

- **Add-ons, Plug-in:**

  - Add-ons and plug-ins are the tools that help to extend and modify the functionality of the browser.
  - Both the tools boost the performance of the browser, but are different from each other.
  - A plug-in is **a complete program** or may be a third-party software. For example, Flash and Java are plug-ins.
  - A Flash player is required to play a video in the browser.
  - A plug-in is a software that is installed on the host computer and can be used by the browser for multiple functionalities and can even be used by other applications as well.
  - An add-on is **not a complete program** and so is used to add only a particular functionality to the browser. An add-on is also referred to as extension in some browsers

- **Cookies:**
  - A cookie is a text file, containing a string of information, which is transferred by the website to the browser when we browse it.
  - This string of information gets stored in the form of a text file in the browser.
  - The information stored is retransmitted to the server to recognize the user, by identifying pages that were visited, choices that were made while browsing various menu(s) on a particular website.

QUESTIONS FROM COMPUTER NETWORKS:

Section A: Multiple Choice Questions

1. For web pages where the information is changed frequently, for example, stock prices, weather information which out of the following options would you advise.
   a. static webpage b. dynamic webpage
   c. both a and b d. none of the above

Answer: (b) dynamic webpage

2. which among the following is not an example of browser.
   a. Chrome b. Firefox
   c. Avast d. edge

Answer: (c) Avast

3. _____ is a networking device that forwards the data packets between computer network
   a. Repeater b. Hub
   c. Switch d. Router

Answer: (d) router

4. Which transmission media provides highest transmission speed in a network?
   a. Twisted Pair Cable b. Coaxial cable
   c. Optical Fiber d. All of these

Answer: (c) Optical Fiber

5. NSFNET stands for _____
   a. National Science Foundation Network
   b. National Science fundamental network
   c. neuro Science Foundation Network
   d. National Search Foundation Network

Answer: (a) National Science Foundation Network

6. Physical or logical arrangement of network refers to as _____
   a. Routing b. Looping
   c. Topology d. Networking

Answer: (c) Topology

7. Which network topology has a central device, which brings all signals together
   a. Bus b. Star

c. Ring d. Hybrid

Answer: (b) Star

8. Bluetooth is an example of _____
   a. Personal Area Network b. Local Area Network
   c. Virtual Private Network d. wide area network

Answer: (a) Personal Area Network

9. Which networking device is used to transfer communication signal to long directions?
   a. Amplifier b. Repeater
   c. Router d. all of the above

Answer(b)

10. Which is called as smart hub?
   a. Hub with high-speed ports b. Switch
   c. Router d. All of the above

Answer(b)

11. Which topology in general uses less wire length compare to other?
   a. Star topology b. Ring topology
   c. Bus topology d. All use same length of wire

Answer(c)

12. A collection of hyperlinked documents on the internet forms the
   a. World Wide Web (WWW) b. Email system
   c. Mailing list d. Hypertext Markup Language

Answer (a)

13. HTML is used to create:
   a. Machine Language program b. High level program
   c. web page d. web server

Answer(c)

14. Which term identifies a specific computer on the web and the main page of the entire site?
   a. WWW b. link
   c. SQL d.URL

Answer (d)

15. How is a domain name different from URL?

Answer: Domain names are used in URLs to identify particular web servers. for example in the URL **http ://www.cbse.nic.in/welcome.htm** , the domain name is www.cbse.nic.in

16. Rashmi has to send an email to Raees. She also wants to send the same email to Anjali but does not want Raees to know about it. Which option out of Cc and Bcc should Rashmi use to enter the email address of Anjali?

Answer: Bcc(Blind Carbon Copy)

17. When a user browses a website the web server sends a text file to the web browser. What is the name of the file.

Answer: cookies

ASSERTION AND REASONING QUESTIONS:

Directions:

In the following questions a statement of assertion (A) is followed by a statement of Reason(R)Mark the correct choice as:

a) Both assertion (A) and reason(R) are true and reason(R) is the correct explanation of assertion(A)
b) Both assertion (A) and reason(R) are true and reason(R) is not the correct explanation of assertion(A)
c) Assertion (A) is true but reason(R) is false
d) Assertion (A) is false but reason(R) is true

18. **Assertion** (A) Browser is the software to access internet-based webpages in the computer.
Reason (R) Local Area Network (LAN) is a network where two or more computers are connected within 1 km.

Answer: (b)

19. **Assertion**(A): Cookies are plain text files
Reason (R): Cookies store the profile picture on social media.

Answer: (c)

20. **Assertion**(A): Static webpage contains content that do not change

Reason(R): They may only change if the actual HTML file is manually edited.

Answer: (a)

21. **Assertion**(A): the Personal Area Network (PAN) is established within a very small area (20 to 30 sq. ft.) for sharing information among computers

    Reason (R): the Campus Area network is used to interconnect computers located with within a campus such as university campus, corporate campus, hospital campus etc.

Answer: (b)

22. **Assertion**(A): A repeater is a device that amplifies network over geographical distance
    Reason (R): A hub is a device which is used to connect more than one device in the network.

Answer: (b)

23. **Assertion**(A): VoIP stands for Voice over Internet Protocol.
    Reason (R): It is a technology that allows you to make voice calls using a broadband internet connection in stead of a regular phone line.

Answer (A)

Short Answer type questions (2 marks)

24. Aman a freelance website developer has been assigned a task to design a few web pages for a book shop. Help Aman in deciding out of static page and dynamic webpage, what kind of web pages should be designed by clearly differentiating between static and dynamic web pages on at least two points.

Answer:

Static web page	Dynamic web page
▪ content of this type of webpage cannot be changed at run time.	▪ content of this type of webpages can be changed at run time.
▪ no interaction with server in case of static web pages.	▪ interaction with server database is possible in case of dynamic web pages.

25. Priyanka a beginner in IT field has just started learning web technologies. Help her in understanding the difference between website and web pages with the help of a suitable general example of each.

Answer:

The difference between a website and a web page is that a website is a collection of different web pages containing information on a particular topic.
A web page is an individual page of a big website usually containing more specific information.

26. Sahil a class X student has just started understanding the basics of internet and web technologies. He is a bit confused in between the terms World Wide Web and Internet. Help him in understanding both the terms with the help of suitable examples of each.

Answers:

INTERNET	WWW
▪ It is a global network of networks.	▪ It stands for world wide web
▪ it is a means of connecting a computer to any other computer anywhere in the world.	▪ it is a collection of information which is accessed via internet
▪ it can be viewed as a big book store.	▪ it can be viewed as collection of books on that store.

27. Murugan wants to send a report on his trip to the North East to his mentor. The report contains images and videos. How can he accomplish his task through internet?

Answer:

He can accomplish task by uploading files to a Cloud Storage Services like Google Drive, Drop Box, OneDrive and sharing them or email them. Murugan can send his report to his mentor through internet.

28. Explain VoIP.

Answer:

VoIP allows us to have voice call over internet. i.e., the voice transmission over a computer network rather than through the regular telephone network.

29. What is the difference between Hub, Switch and Router?

Answer:

- Hub is least expensive, least intelligent and least complicated of the three and it broadcasts data to every port which may cause serious security and reliability concern.
- Switch works similarly like hub but in a more efficient manner. It creates connections dynamically and provides information only to the requesting port.
- Router is smartest and most complicated out of the three. It comes in all shapes and sizes. Routers are located at gateway and forwards packets in a network.

LONG QUESTIONS(4 MARKS)

30. You as a network expert have to suggest the best network related solutions for their problems raised in (i) to (v), keeping in mind the distances between the buildings and other given parameters.

Shortest distances between various buildings

MUMBAI CAMPUS: ADMIN TO ACCOUNTS	55 m
ADMIN TO EXAMINATION	90 m
ADMIN TO RESULT	50 m
ACCOUNTS TO EXAMINATION	55 m
ACCOUNTS TO RESULT	50 m
EXAMINATION TO RESULT	45 m
DELHI Head Office to MUMBAI campus	2150 m

Number of computers installed at various buildings are as follows:

ADMIN	110
ACCOUNTS	75
EXAMINATION	40
RESULT	12
DELHI HEAD OFFICE	20

(i) Suggest the most appropriate location of the server inside the MUMBAI campus (out of the four buildings) to get the best connectivity for maximum number of computers. Justify your answer.

(ii) Suggest and draw cable layout to efficiently connect various buildings within the MUMBAI campus for a wired connectivity.

(iii) Which networking device will you suggest to be procured by the company to interconnect all the computers of various buildings of MUMBAI campus?

(iv) Company is planning to get its website designed which will allow students to see their results after registering themselves on its server. Out of the static or dynamic, which type of website will you suggest?

(v) Which of the following will you suggest to establish the online face to face communication between the people in the ADMIN office of Mumbai campus and Delhi head office? a) Cable TV b) Email c) Video conferencing d) Text chat

31. Multipurpose Public School, Bengaluru is setting up the network between its different wings of school campus. There are 4 wings named as: - SENIOR(S), JUNIOR(J), ADMIN(A) and HOSTEL(H)

Distance between various buildings is as follows:

Hostel to Admin	55m
Admin to Senior	150m
Senior to Junior	160m
Junior to hostel	60m
admin to junior	125m
senior to hostel	180m
Number of Computers in each unit:	
Admin	25
Junior	100
Senior	15
Hostel	60

(i) Suggest and draw the cable layout to efficiently connect various wings of Multipurpose Public School, Bengaluru.

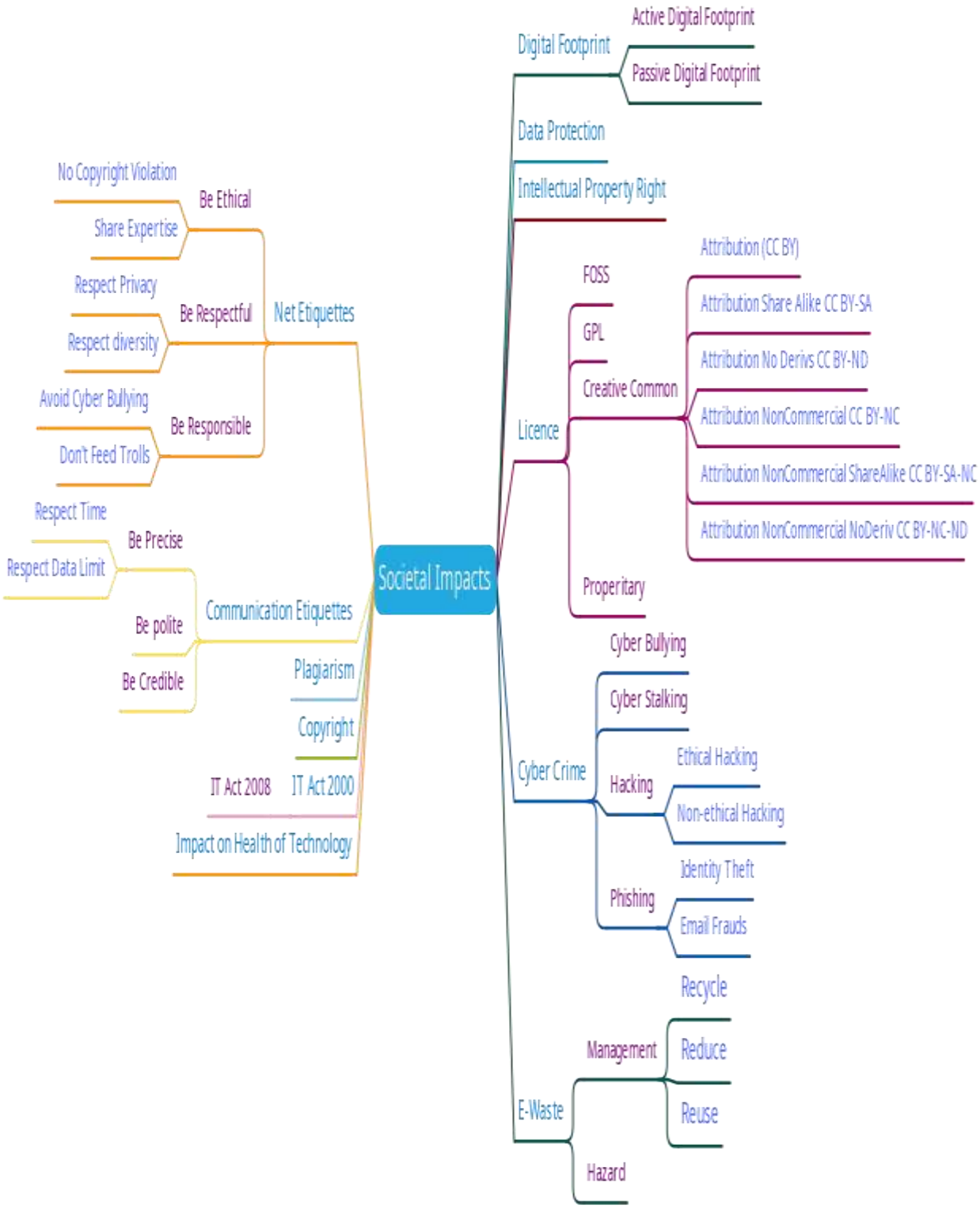
(ii) Name the most suitable wing where the servers should be installed. Justify your answer.

(iii) Suggest a device/software and its placement that would provide data security for the entire network of the school

(iv) Suggest a device and a protocol that shall be needed to provide wireless internet access to all smartphone/laptop users in the campus of Multipurpose Public School, Bengaluru.

(v) Suggest the placement of repeater with justification

SOCIETAL IMPACTS



What is a Digital Footprint?

- A digital footprint is an impact you create on the Web through your online activity, which incorporates browsing, interactions with others, and publication of content.
- In other words, it can be considered as the data trail – intentional and unintentional - you leave behind while you surf the Web or Internet.
- Digital footprint or digital shadow refers to the trail of data left behind through the utilization of the Web or on digital devices.
- The digital footprint of any person can have a positive as well as a negative impact on him.

How Digital Footprints are created?

Find some of the examples of digital footprints:

- Visiting Websites and Online Shopping
- Online Searching
- Posting on social media, blogs, etc.
- Online Image and Video Upload
- Communicating Online (Ex: - Chat, Email, etc.)
- Any activity you perform Online etc.

Types of Digital Footprints:

Active and Passive Digital Footprints.

- **Active digital footprints**: -Active digital footprints are those data trails that a person leaves intentionally on the Web.
  Ex: - Twitter, blog posts, Facebook, social network connections, image and video uploaded on Internet, phone calls, email, and chats are among the ways people create active digital footprints.
- **Passive digital footprints**: - This suggests that a passive footprint would be defined as the unintentional traces of data that an individual creates on the Web.
  Ex: -Website visits and actions, searches, and online purchases are among the activities that add passive data traces to a digital footprint.

Positive and Negative Digital Footprints.

Positive digital Footprints: -That reflects your Positive Personality.

- Increased opportunity
- Higher profits
- Less risk
- Gentler treatment

Negative digital footprints: - Things that reflect your Negative Personality sort of a drunken photo, a silly comment, logging on to an inappropriate website.

- Fewer Opportunities (like Admission, Job, etc.)
- Negative Personal Image

How Digital Footprints can affect you?

- **Privacy Issues**: Digital footprints are a privacy concern because they're a group of traceable actions, contributions, and concepts shared by users. They are often tracked and may allow internet users to find out about human actions.
- **Cyber Vetting**: where interviewers could research about the applicants before the interview based on their online activities on the Web.
- **Target advertisement**: It is used by marketers so as to seek out what products a user is curious about or to inspire ones' interest during a certain product that supported similar interests.
- Less/More Opportunities depends upon your Positive/Negative Digital Footprints.

How can you manage your Digital Footprint?

- Though it is not possible to fully hide your digital footprints, you can follow some of the given techniques to manage your digital footprints.
- You can search your name on different search engines and they provide you facilities where you can set up an alert for future notifications when your name searched online
- Have different email addresses, so professional and private accounts aren't automatically related to each other
- You can change privacy settings on social media accounts where you do not make more things public. But you should not trust them completely because your data is still with those platforms.
- Exercise caution altogether our activities, and refrain from oversharing

Net Etiquettes

- Etiquette means the general behaviour you must follow in your daily life.
- Net Etiquette are rules that you must be obeyed when you are online
- Online etiquette or 'netiquette' refers to the dos and don'ts of online communication.
- You can do Online Communication using Snapchat, Instagram, WhatsApp, etc. These are the most popular online mediums available online.

Net Etiquettes you should follow.

- **Be Ethical**
  - **No copyright violation**: we should not use copyrighted materials without the permission of the creator or owner
  - **Share the expertise**: it is good to share information and knowledge on internet so that others can access it.
- **Be Respectful**
  - **Respect privacy**: We should respect this privacy of other as we care for our privacy and should not share that such information without each other's consent.
  - **Respect diversity**: We should respect the diversity of the people in terms of knowledge, experience, culture, beliefs and other aspects.
- **Be Responsible**
  - **Avoid cyber bullying**:
    - **Cyber Bullying** implies repeatedly targeting someone with intentions to hurt or embarrass.
    - It includes insulting, degrading or intimidating online behaviour such as posting of rumours, giving threats online, posting the victim's personal information, sexual harassment or comments aimed to publicly ridicule a victim.
  - **Don't feed the troll**:
    - **Cyber Trolling** is a form of online harassment that involves intentionally posting provocative, offensive, or inflammatory comments on social media platforms, discussion forums, or blogs with the intention of causing emotional distress or anger.
    - The best way to discourage trolls is not to pay any attention to their comments.

Communication Etiquettes

- Digital communication includes email, texting, instant messaging, talking on the cell phone, audio or video conferencing, posting on forums, social networking sites.
- Being a good digital citizen (netizen), we must abide by following Communication Etiquettes.
- **Be Precise**
  - **Respect time**:
    - Don't waste stime in responding to unnecessary emails.
    - Don't expect an instant response as the recipient may have other priorities.
  - **Respect data limits**:
    - Avoid large attachment
    - Send data through storage such as Google Drive, Microsoft OneDrive, DropBox etc.
- **Be Polite**: We should be polite, non-aggressive and non-abusing in our communication even if we don't agree with their point of view.
- **Be Credible**: We should be cautious while making a comment, replying or writing an email or forum post as such acts decide our credibility over a period of time.

Data Protection

- To protect these data from substantial harm, embarrassment, inconvenience and unfairness to an individual.
- Sensitive data like biometrics information such as fingerprint, health information, financial information, or other personal data like documents, photos, audio clips, videos are required to be protected.

The data protection can be implemented by using one of the following methods:

1. **Encryption:** It is a method of representing in such a way that only authorized parties or systems can understand the data patterns. In this technique, the text data will be converted into ciphertext. Ciphertext refers to converting user-readable data into incomprehensible data.
2. **Authentication**: Authentication involves the user's identity using different credentials like username and password, security keys, different sensor locking systems, OTP and other verification methods.

Intellectual Property Right (IPR)

- Intellectual Property refers to inventions, literary and artistic expressions, designs and symbols, names and logos.
- The Intellectual Property Right gives ownership to the creator of the Intellectual Property holder.
- By this, they can get recognition and financial benefits from their property.
- These intellectual properties are legally protected by copyrights, patents, trademarks, etc.

Copyrights

- Copyrights refers to the legal rights to use a material like writing, articles, photographs, audios, videos, software or any other literacy or artistic work.
- Copyrights are **automatically** granted to the creators or the owners.
- The right includes the right to copy, reproduce, distribution of the work or content.
- If a person needs to use copyrighted materials, then the needs to obtain a license or written permission of the creators.
- There are some CC (Creative Common) license and GNU GPL license.
- They allow to use their creation as loyalty free materials

Patents

- The patents are given for the inventions.
- The creator needs to apply for the invention.
- When the patent is granted, the owner gets rights to prevent others from using, selling or distributing the protected invention.
- Patent gives full control to the patentee to decide how others can use the invention.
- A patent protects an invention for 20 years, after that public can use it freely.

Trademark

- Trademark is applicable for the visual symbol, word, name, design, slogan, label, etc for the product.
- It provides uniqueness for the other brands and commercial enterprise.
- It also gives recognition to the company.
- The trademark product denoted by ® or ™ symbols.
- There is no expiry time for the trademark.

Licensing

- A license refers to a contract or permission or agreement given to any party by a creator to use their product or service or creation.
- A license can be purchased by paying money. License is the term that gives special rights to the user to use the copyrighted material.
- Similarly, a software license is an agreement that provides legal rights to the authorised use of digital material.
- All the software, digital documents or games you are downloading from the internet provides the license agreement to use the material. If anyone is not following will be considered a criminal offence.

Violation of IPR.

Knowingly or unknowingly, people are violating IPR while doing work. So, the violation of IPR done in following ways:

1. Plagiarism
2. Copyright Infringement
3. Trademark Infringement

Plagiarism

- Plagiarism refers to copy or share the intellectual property of someone without giving any credit or any mention of the creator.
- Sometimes if you derived an idea or product which is already available, then also it is considered plagiarism. Sometimes it is also considered fraud.
- Whenever you are using any online material for your personal use or for any purpose, always cite the author and source to avoid plagiarism.

Copyright Infringement

- When you use the work of others without taking their written permission or don't pay for that using that is considered as copyright infringement.
- If you download an image from google and use in your work even after giving the credit or reference you are violating copyright.
- So before downloading any content check it for copyright violation.

Trademark Infringement

- The unauthorized use of trademark product is known as trademark infringement.
- The trademark owner can take a legal action for trademark infringement.

Public Access and Open Source

- Open source allows using the material without any special permission.
- Some software is there which are available for free of cost and allows redistribution. User can use them, copy them and redistribute them.
- They are available with modifiable source code. Free and Open-Source Software (FOSS) is a large community of users and developers who are contributing towards open-source software.
- These tools are Linux, Ubuntu, open office, Firefox are examples of open-source software.

CC and GPL

You have a creative commons license and General Public License to use such material.

- The CC is commonly used for some creative resources like websites, music, film, or literature etc.
- GPL is designed for the software, add-on, plugin or themes.
- It allows the end-user to use, run, study, share and modify the software.
- GPL license do not provide the regular updates and support.
- The CC is a non-profit organization that provides a wide range of resources like images. videos, and other libraries. Some them offers proper attribution and credit to use them.
- CC license offer a copyright license to use someone's materials to share, use and extend the work done by them.
- Types Creative Common
    - **CC BY** (Attribution) - This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation.

- **CC BY-SA (Attribution ShareAlike)** - This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms.
- **CC BY-ND (Attribution NoDerivs)** - This license lets others reuse the work for any purpose, including commercially; however, it cannot be shared with others in adapted form, and credit must be provided to you.
- **CC BY-NC (Attribution NonCommercial)** - This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial.
- **CC BY-NC-SA (Attribution NonCommercial ShareAlike)** - This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.
- **CC BY-NC-ND (Attribution NonCommercial NoDerivs)** - This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.

Cyber Crime and Cyber Laws

- Cybercrime refers to criminal activities or offences carried out in a digital world.
- In cybercrime, the computer can be targeted or used to perform such activities.
- It can be performed against an individual or organization or any group.
- Any activity related to physical harassments or financial loss or mental harassments using computers, the internet or network is considered a cybercrime.
- The cybercrime covers Phishing, Credit Card Frauds, Illegal Downloading, Cyber Bullying, Creation and Distribution of viruses, spam etc.
- These types of activities increasing day by day through hacking, ransomware like attacks, denial-of-service attack, phishing, email fraud, banking fraud and identity theft.

Hacking

- Hacking refers to entering into someone's account without the user's consent or stealing login information from someone's account and unauthorized access to data.
- The person who is doing this kind of activities known as a Hacker.
- He or she is doing this using software or hardware or expertise.
- They gain unauthorized access to data, steal some information from your computer, manipulate the digital data, capture user's personal or financial information or install malware on your systems.

Ethical Hacker or White Hacker

- If this hacking is done for positive intent, then it is known as Ethical Hacking or White Hat Hacking.
- The hacker is known as Ethical Hacker or White Hat Hacker.
- They help to protect the system from hacking and improves the security of the system. An ethical hacker exploits a website to find a security glitch on the website.

Back Hat Hacker

- A Black Hat Hacker or Unethical Hacker tries to gain untheorized access and steal the sensitive information with the aim to damage or break down the system.
- Their main focus is security cracking and stealing the sensitive information.
- They may do some illegal activities like identify theft, monetary gain for their personal benefits or leak sensitive data in the market and earn money.

Phishing:

- Phishing is an unlawful activity where fake websites or emails that look original or authentic are presented to the user to fraudulently collect sensitive and personal details, particularly usernames, passwords, banking and credit card details.

Identity Theft:

- Identity theft is a cybercrime in which personal information of a person is stolen from computers or computer networks, to commit fraud by using the data gained unlawfully.
- Financial identity theft: when the stolen identity is used for financial gain.
- Criminal identity theft: criminals use a victim's stolen identity to avoid detection of their true identity
- Medical identity theft: criminals can seek medical drugs or treatment using a stolen identity

How to reduce the risk of cybercrime?

To reduce the risk of cybercrime, follow these practices:

- Use antivirus and keep it updated.
- Take backup of data at regular intervals.
- Download and use software from authorized sources only.
- Update the software including browser as and when it is prompted.
- Avoid to download and use pirated software.
- Follow some security alerts provided by the system and browser.
- Use strong passwords for the accounts.
- Don't write or save your passwords at the common places.
- Don't allow third party cookies from the websites.
- Perform online payment transaction securely and through authorized websites only.
- Don't share your OTP or PIN number with anyone.
- Always secure wireless network with a password when connecting via hotspot.

Indian IT Act

- Government of India came up with the **Indian Information Technology Act in 2000** to deal with rapidly growing, the risk of cybercrimes, internet fraud, cyber-attacks.
- This act was amended in **2008**.
- The Indian IT Act provides guidelines and rules to the user on the processing, storage and transmission of sensitive information.
- The states of our country have their own cyber cells in the police station to report cybercrime.
- The IT act provides legal provisions and implications against cybercrime.
- If the cybercrime is proved then the accused may get the penalties also.
- It resolves issues such as hacking, tampering with digital documents, unauthorized use of passwords, using other's personal data without consent etc.

- The activities help people to use the digital environment without fear.
- On the other side, the IT act also helps the government in good governance.

E-waste Hazards and Management

- E-Waste refers to the garbage of electronic devices and gadgets like computer peripherals, laptop components, mobile phones and accessories, TVs, Tablets, Printers, Scanners.
- When these devices are not in use or damaged or not in a working condition becomes E-Waste.
- The E Waste comprises of multitude of components, some containing toxic substances that can have an adverse impact on human health and the environment if not handled properly.
- In India, e-waste management assumes greater significance not only due to the generation of its own e-waste but also because of the dumping of e-waste from developed countries.
- WEEE (Waste Electrical and Electronic Equipment) is becoming a major concern day by day globally.

What are the impacts of E-Waste on the environment?

The E-Waste impacts a lot in our environments. They are:

1. Emission of gases
2. Fumes in the atmosphere
3. Discharge of liquid waste into drains
4. Disposal of solid e-waste materials

The above listed points are contributing in the environment pollution either one or other way. The e-waste pollutes soil, water and air. When the e-waste dumped or thrown in landfills, the harmful chemicals seep into the soil and pollutes the soil.
Similarly, when it reaches to the water surface or into the water of rivers, sea water or any lake or pond, it releases some particles into the water and this water becomes useless for humans, animals or for agriculture use.
When the particles loaded with heavy metals enters into the atmosphere it causes the air pollution.

What are the impacts of E-Waste on the humans?

- The electrical or electronic devices are manufactured using certain metals and elements like lead, beryllium, cadmium, plastics.
- The above materials are difficult to recycle and contaminate water, soil, air, food.
- Lead from batteries and monitor may cause lead poisoning which affects the kidneys, brain and central nervous system.
- Children are particularly vulnerable to lead poisoning.
- Burning of electronic circuit boards for disposal generate.
- Harmful chemical called beryllium which causes skin diseases, allergies and an increased risk of lung cancer.
- Mercury content in the e-waste causes respiratory disorders and brain damage.
- The cadmium found in semiconductors and resistors can damage kidneys, liver and bones.
- At plastic, after reacting with soil, moisture produce harmful chemical causing various psychological problems like stress and anxiety

E-Waste Management

The efficient disposal of e-waste is E-Waste management. It is not possible to dispose it completely but at certain level it can be reduced through Reduce, Recycle and Reuse.

- **Reduce**: Reduce the purchase of electronic devices by using some alternates or as per the need only. They can only be discarded only after their useful life ends. Good maintenance of these devices increases its life.
- **Reuse:** The devices can be modified and reused for other purposes. The discarded devices can be supplied or sold to someone who can use them. The process of re-selling old electronic goods at lower prices is called refurbishing.
- **Recycle:** Recycling refers to the process of conversion of electronic devices into something else which can be used again and again. The e-waste which cannot be recycled can be repaired, refurbished or re-used. Many NGOs provide doorstep services to collect e-waste.

E-Waste Management in India.

- According to the Environmental Protection Act, 1986 – "Polluter pays Principle" means that anyone causing the pollution will pay for the damage caused.
- Any kind of violation will be punished according to this act.
- The Central Pollution Control Board (CPCB) has issued guidelines for the proper handling and disposal of e-waste.
- The guideline says that the manufacture of the product will be responsible for the disposal of the product when it becomes e-waste.
- The Department of Information Technology (DIT) issued a comprehensive technical guide on "Environmental Management for Information Technology Industry in India." The industries have to follow these guidelines for recycling and reuse of e-waste.
- Many companies started the process of recycling programs.

Awareness about health concerns related to the usage of technology

- Spending lots of time in front of various electronic devices console with improper poster manage both physically as well as mentally.
- The following are some side effects of spending too much time on the internet or devices:
- The Internet can be addictive
- Can have a negative impact on our physical and psychological well being
- Muscle and joint problems
- Overuse injuries like swelling, pain, joint stiffness, weakness and numbness.
- Eyestrain
- While sitting in front of a computer always keep in mind these points:
  - Always take care of viewing distance and viewing angle.
  - The right position should be straight and with a 90 degree.
  - Always take support at your lower back.
  - Knee Angle and Seat Back angle should be a 90-degree angle.
  - Adjust the height of the chair.
  - Feet on floor for shorter people.

01 Mark Questions

1. From which year did the IT Amendment Act, 2008 come into force?

i. 2009 ii. 2008 iii. 2000 iv. 20004

2.Online _____ is a theft of personal information in order to commit a fraud.

i. Phishing ii. Prank iii. Identity theft iv. All of these

3 _____ is a specific type of disorder where a person can't strike a balance between their online and offline time?

i. Phishing ii. Tech savvy iii. Internet Addiction iv. None of these

4. Any fraudulent business practice that extract money from an ignorant person is called.

i. Phishing ii. Scam iii. Both iv. None of these

5. Any information about you created by you that exists in digital form is referred to as.

i. Intellectual property ii. Digital property iii. Both iv. None of these

6. Which of the following come under cybercrime?

a. The theft of brand-new sealed laptop.

b. Access to a bank account for an unauthorized money transaction

c. Modification in a company's data with unauthorized access

d. Photocopy of a print report

i. a and b ii. a and c iii. b and c iv. D

7. Using someone else's twitter handle to post something will be termed as.

i. Fraud ii. Identity Theft iii. Online stealing iv. Violation

8. Standard security protocol that establish encrypted link between web server and browser is called.

i. Web encryption Technology ii. STD Technology

iii. Online safety Technology iv. Security socket Technology

9. Intellectual property right protect the use of information and idea that are of.

i. Ethical value ii. Moral value iii. Commercial value iv. Social value

10. Credit Card fraud may include.

i. Stealing for credit card ii. Unauthorized and illegal use of credit card

iii. Stealing of information iv. Phishing

11. Which of the following is an open-source software?

i. Microsoft Windows ii. Adobe Photoshop iii. My SQL iv. MS power point

12. The term intellectual property right covers.

i. Copyrights ii. Trademark iii. Patent iv. All of these

13. What is the term used for entering someone's account without their consent?

i. Hacking ii. Phishing

iii. Cybercrime vi. Copyright infringement

14.Online personal account, personal websites are example of.

i. Digital Wallet ii. Digital property iii Digital Certificate iv. Digital signature

15. Unsolicited commercial email is known as.

i. Spam ii. Malware iii. Virus iv. Spyware

16. Which of the following is not a type of cybercrime?

i. Data Theft ii. Forgery

iii. Damage of Data iv. Installation antivirus for protection

17. Which of the following is not covered under IPR?

i. Music ii. Insurance iii. Logo Designed iv. Invention

18. Which of the following is one of the impacts of e-waste on the environment?

i. Global Warming ii. Deforestation iii. Soil Erosion iv. Emission of Gases

19. When the e-waste dumped or thrown in landfills, the chemicals seep into the _____ and pollutes it.

i. Sea ii. Drain iii. Soil iv. All of these

20. When circuits burnt for disposals, it creates which of the following harmful chemical?

i. Beryllium ii. Lead iii. Copper iv. Mercury

21. Which of the following reduce e-waste?

i. Purchasing more and more gadgets. ii. Using them for a short time and then discarded.

iii. Good maintenance. iv. All of these.

22. The process of re-selling old electronic goods at lower prices is called.

i. Recycling ii. Refurbishing ii. Reusing iv. All of these

23. _____ refers to the process of conversion of electronic devices into something else which can be used again and again.

i. Recycling ii. Refurbishing iii. Reusing iv. All of these

24. The _____ has issued guidelines for the proper handling and disposal of e-waste.

i. State Pollution Control Board ii. Central Pollution Control Board

iii. Both iv. None of these.

25. _____ deals with designing and arranging such furniture, equipment and systems to make the work comfortable on computer.

i. Economic ii. Micro-economic iii. Ergonomics iv. None of these

02 marks Questions

1. Name the crimes for which cyber laws are enforced strictly in India.

2. What do you mean by cyber security?

3. What is phishing?

4. Explain the term 'web beacon'.

5. List any four health hazards related to excessive use of Technology.

6. Mention net etiquettes.

03 marks Questions (Case Based)

1. Ms. Sen is an IT expert and a freelancer. She recently used her skill to access the Admin password for the network server of super Dopers technology Ltd and provided confidential data of the organisation to its CEO informing him about the vulnerability of their network security. Out of the following option (i) to (iv) which one most appropriately defines Ms. Sen?

(i) Hacker (ii) Cracker (iii) Operator (iv) Network Admin

2. Munish has topped in all the subjects in his school exams. His father has gifted him a laptop. He would like to make Munish aware of the posture health hazards associated with inappropriate and excessive use of laptop. Help his father list the points which he should discuss with Munish.

3. Rishika found a crumpled paper under her desk. She picked it up and opened it. It contained some text which was struck off thrice. But she could still figure out easily that the struck-off text was the email ID and password of Garvit, her classmate. What is ethically correct for Rishika to do?

4. Jeet has to prepare a project on "Swachh Bharat Abhiyan". He decides to gather information from the Internet. He downloads three web pages (webpage1, webpage2, webpage3) containing information on Swachh Bharat Mission. Which of the following steps taken by Jeet is/are an example of plagiarism or copyright infringement? Give justification in support of your answer.

(a) He read a paragraph on "Swachh Bharat Abhiyan" from webpage1 and rephrased it. He finally pasted the rephrased paragraph in his project.

(b) He downloaded three images of “Swachh Bharat Abhiyan” from webpage2. He made a collage for his project using these images.

(c) He downloaded “Swachh Bharat Abhiyan” icon from webpage3 and pasted it on the front page of his project report.

Answer

01 Mark Answers

1 i

2 iii

3 iii

4 ii

5 ii

6 iii

7 ii

8 iv

9 iii

10 ii

11 iii

12 iv

13 i

14 ii

15 i

16 iv

17 iii

18 iv

19 iii

20 i

21 iii

22 ii

23 i

24 ii

25 iii

02 Marks Answers

1. i. Cybercrimes ii. Electronic and Digital Signature iii. Intellectual Property iv Data Protection and Privacy.

2. Cyber security is the combination of best process and practices to ensure the security of networks, computers, programs, data and information from attack, damage or unauthorised access.

3. Phishing is the attempt to acquire sensitive information such as username, password, credit card details etc.

4. Picture in email message also called 'web beacons. A Web beacon gets your computer through an email, or it can be in a webpage that you visit. Some people might call it "spyware," in that it's used to take note of your online activity, but in most cases (especially with websites) it's not there to do any harm.

5. i. Impact on bones, joints. ii. Impact on hearing. iii Impact on eyes. iv. Asleep disorder.

6. i. Keep Messages and Posts Brief ii. Use Discretion iii. Protect Personal Information iv. Obey Copyright Laws v. Help Others vi. Respect another people's privacy vii. Verify facts before reposting viii. Check messages and respond promptly ix. Thank others who help you.

03 Marks Answers (Case Based)

Hacker: hacker is a computer expert or skilled computer user who spends enormous amount of time trying to breach the security of network. Web servers and e-mail servers

Munish must be made aware of the following points regarding the pitfall of excessive use of Technology.

i. Strain Vision. ii. Muscle and Joint Pain. iii. Too much dependence.

iv. Failing Memory. v. Sleep Deprivation. vi. Distraction.

vii. Waste of time.

Inform Garvit, so that he may change his password.

All of three steps (a, b and c) taken by jeet shall be considered as plagiarism or copyright infringement as they are in violation of copyright laws, which state that one should use one's own words and ideas. Here, Jeet has copied the ideas along with the images from a website without giving reference to the author or the website.it is necessary to provide some reference or give credit to the source from where you have received your information. Since he has used information from a website, he should have cited the name of the website, its URL or name of the author along with proper acknowledgement.

Bibliography

1. NCERT Informatics Practices Book class XI and XII.
2. https://app.mindmup.com/ - for mind map.
3. https://www.geeksforgeeks.org
4. https://pandas.pydata.org/
5.