In [ ]:

2) numpy : Numerical python/Number python [maths]

3) Matplotlib : plot

4) seaborn

5) plotly

6) bokhe

********** ML packages **********
7) Scikit-learn (Sklearn)
8) pickle : save the ML model
9) joblib : save the ML model

********** DL packages**********
10) Tensorflow
11) kears
12) PyTorch

********** NLP Packages**********
13) NLTK
14) SciPy

********** BERT MOdels(Hugging face transformers) ********** 15) Transformers

********** GEN AI **********
16) Chat GPT: OpenAI + Azure : Azure packages seperately 17) MakerSuit: Google : Google packages
18) Amazon Q : Amazon : Aws packages

****
Others**********
random
math
time
sys
os

In [2]:
###############
Packages##########################

****
EDA**********
1) pandas : Data frame analysis

********* – 1 :

Import required packages

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

������ – 2 :

In [3]: In [4]: Out[4]:

Read the data

������-������

```
file_path="C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\D

# <package_name>.<method_name>()
pd.read_csv(file_path)
```

case_id continent education_of_employee

has_job_experience requires_job_trainin 0 EZYV01 Asia

High School N

1 EZYV02 Asia Master's Y

2 EZYV03 Asia Bachelor's N

3 EZYV04 Asia Bachelor's N

4 EZYV05 Africa Master's Y

... ...

25475 EZYV25476 Asia Bachelor's Y **25476** EZYV25477

Asia High School Y **25477** EZYV25478 Asia Master's Y

25478 EZYV25479 Asia Master's Y **25479** EZYV25480 Asia

Bachelor's Y

25480 rows × 12 columns

������-������

###

In [6]: Out[6]:

In [11]:

```
file_path1="C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\
pd.read_csv(file_path1,sep=';')
```

age job marital education default balance housing loan

contact day m 0 30 unemployed married primary no 1787 no

no cellular 19 **1** 33 services married secondary no 4789 yes

yes cellular 11 **2** 35 management single tertiary no 1350 yes

no cellular 16 **3** 30 management married tertiary no 1476 yes

yes unknown 3 **4** 59 blue-collar married secondary no 0 yes

no unknown 5 **...** ... ...

4516 33 services married secondary no -333 yes no cellular

30 **4517** 57 self employed married tertiary yes -3313 yes yes

unknown 9 **4518** 57 technician married secondary no 295 no

no cellular 19 **4519** 28 blue-collar married secondary no 1137

no no cellular 6 **4520** 44 entrepreneur single tertiary no 1136

yes yes cellular 3

4521 rows × 17 columns

������ – 3

Create data frames

A) Using list

```
list1=['Hyd','Mumbai','Benguluru']
```

Out[11]:

	City
0	Hyd
1	Mumbai
2	Benguluru

```
list2=['040','022','080']

cols=['City','Code']
city=pd.DataFrame(zip(list1,list2),columns
=cols)
# I created a dataframe and I saved in a
varaiable city
city
```

	Code
0	040
1	022
2	080

B) Save the data in local

```
os.getcwd(
)
```

In [12]:

```
import os
```

Out[12]:
```
'C:\\Users\\omkar\\OneDrive\\Documents\\Data science\\Naresh IT\\Data scie
nce\\Batch-3_Sep28\\EDA-Python'
```

In [13]:

```
city.to_csv('C:\\Users\\omkar\\OneDrive\\D
ocuments\\Data science\\Naresh IT\
```

In [14]: In [ ]:

```
# In file handling we read file in two
ways

# python file and data file both are in

different location # we need to provide

the full path
```

In [15]: In [22]:

```
# python file and data file both are in
same location
# we provide only data filename


city.to_csv('naresh_it_city1.csv',index=Fa
lse)

# there is no extra index column
```

we create a dataframe using list, data frame name: **city**

we saved the dataframe in local with index: **naresh_it_city.csv**

we saved the dataframe in local with out index: **naresh_it_city1.csv**

```
city.to_csv('naresh_it_city.csv')

# extra index column created
```

********* – 1

:

Read those two dataframes

```
# both python file and data file are at same location
# no need to provide full path location
pd.read_csv('naresh_it_city.csv')
```

Out[22]:

	Unnamed: 0	City	Code
0	0	Hyd	40
1	1	Mumbai	22
2	2	Benguluru	80

```
pd.read_csv('naresh_it_city
.csv',index_col=0)
```

In [23]:

Out[23]:

	City	Code
0	Hyd	40
1	Mumbai	22
2	Benguluru	80

```
In [18]:                t_city1.csv')
pd.read_csv('naresh_i
```

Out[18]:

	City	Code
0	Hyd	40
1	Mumbai	22
2	Benguluru	80

C) Change the index

```
                    city=pd.DataFrame(zip(l
                    ist1,list2),
In [25]:            columns=cols,
list1=['Hyd','Mumbai',' index=['A','B','C'])
Benguluru']         city
list2=['040','022','080
']

cols=['City','Code']
```

Out[25]:

	City	Code
A	Hyd	040
B	Mumbai	022
C	Benguluru	080

D) How to add new column

```
                                # what is the new column want to
                                add: Product # How many rows in
In [ ]:                         your data: 3
                                # what are the new values for


                                product:['Car','Bus','Train']
In [26]: In [27]:

                                city['Product']=['Car','Bus','Trai
# what is your dataframe name:city
```

```
n']
```

```
city
```

Out[27]:

	City Code	Product
A	Hyd	040 Car
B	Mumbai	022 Bus
C	Benguluru	080 Train

E) How to drop the column

In [30]:

```
# what is the name of dataframe: City
# Which column you need to drop: Product
```

In [31]:

```
# axis =1 ============ > columns
# axis=0 ============== > rows
# You are modifying the table , do you
want save the results in same 'city'  # or
you want to create a new variable

# Suppose I want to keep in same 'City
varaible': inplace=True

city.drop('Product',
 axis=1,
 inplace=True)
```

```
city
```

Out[31]:

	City Code
A	Hyd 040
B	Mumbai 022
C	Benguluru 080

F) Create a Dataframe using Dictionary

In [32]:

```
dict1={'Names':['Ram','Ra
heem','Robert'],
                'Age':[30,31,32]}

dict1
```

Out[32]:

```
{'Names': ['Ram', 'Raheem', 'Robert'], 'Age': [30, 31, 32]}
```

In [33]:

```
# keys automatically
consider as columns #
                values automatically
                consider as rows
pd.DataFrame(dict1)
```

Out[33]:

	Names	Age
0	Ram	30
1	Raheem	31
2	Robert	32

If the values are not in list, then pass the index

In [37]:
```
dict2={'Name':'Narpd.DataFrame(dict2
esh_IT',           ,
'City':'Hyderbad', index=[0,1,2])
'Age':25}
```

Out[37]:

	Name	City	Age
0	Naresh_IT	Hyderbad	25
1	Naresh_IT	Hyderbad	25
2	Naresh_IT	Hyderbad	25

In [ ]:

In [ ]:

In [ ]: