Date-07-12-2023

In [ ]:
```
############### Packages ##########################

************* EDA ******************************

1) pandas : Data frame analysis

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
```

Step-1 :

Import Requierd Packages

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plot
import seaborn as sns
```

Step-2:

Read the Dataset

Visa-dataset

In [2]:
```
file_path="C:\\Users\\kurre\\OneDrive\\Documents\\Naresh IT\\datafiles\\Vis
pd.read_csv(file_path)
```

Out[2]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_traini
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

bank-Dataset

In [5]:
```
file_path1="C:\\Users\\kurre\\OneDrive\\Documents\\Naresh IT\\datafiles\\ba
pd.read_csv(file_path1,sep=',')
```

Out[5]:

	age	job	marital	education	default	balance	housing	loan	contact	day
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

In [ ]:
```
# pd.read_csv(file_path1,sep=';')  (if all dataset is mix than use seperati
```

Step-3:

Create Data Frames

A) Using list

In [11]:
```
list1=['Ram','Raheem','Robert']
list2=[30,31,32]
pd.DataFrame(zip(list1,list2))
```

Out[11]:

	0	1
0	Ram	30
1	Raheem	31
2	Robert	32

In [9]:
```
list1=['Ram','Raheem','Robert']
list2=[30,31,32]

cols=['Name','Age']
names=pd.DataFrame(zip(list1,list2),columns=cols)
```

In [10]:
```
names
```

Out[10]:

	Name	Age
0	Ram	30
1	Raheem	31
2	Robert	32

$Step$-4 :

How to Save the dataframe in local

In [ ]:
```
# What is your data frame name: names

# Where you want to save: C:\Users\kurre\OneDrive\Documents\Naresh IT\dataf

# What new names you want to provide : ex- NAMES
```

In [18]:
```
names.to_csv('C:\\Users\\kurre\\OneDrive\\Documents\\Naresh IT\\datafiles\\
```

In [19]:
```
names.to_csv('NAMESS.csv')

# this will save where our python file located

# it will create extra coloumn as index
```

In [15]:
```
names.to_csv('NAMES1.csv',index=False)

# it will not create extra coloumn as index
```

$Task-1$

Read NAMESS and NAMES1

In [20]:
```
filepath1="NAMESS.csv"
pd.read_csv(filepath1)
```

Out[20]:

	Unnamed: 0	Name	Age
0	0	Ram	30
1	1	Raheem	31
2	2	Robert	32

In [ ]:
```
filepath1="NAMESS.csv"
pd.read_csv(filepath1)

# how can remove unnamed coloumn?
```

In [21]:
```
filepath2="NAMES1.csv"
pd.read_csv(filepath2)
```

Out[21]:

	Name	Age
0	Ram	30
1	Raheem	31
2	Robert	32

$Step - 5$:

Add a new coloumn to exiting dataframe

In [22]:
```
names

# What is your data frame name : names
# what is the name of the coloumn you want to add : city
# how many rows in original data : 3
# What is your new data= ['Hyd','chennai','blr']

# if there is 3 rows then put 3  new data, than means new data is depend on
```

Out[22]:

	Name	Age
0	Ram	30
1	Raheem	31
2	Robert	32

In [24]:
```
names['city']=['Hyd','chennai','blr']
```

In [25]:
```
names
```

Out[25]:

	Name	Age	city
0	Ram	30	Hyd
1	Raheem	31	chennai
2	Robert	32	blr

In [26]:
```
names['id']=[1,2,3,4]    # this will give error becouse row=3 and new data=
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call las
t)
Cell In[26], line 1
----> 1 names['id']=[1,2,3,4]

File ~\anaconda3\Lib\site-packages\pandas\core\frame.py:3950, in DataFram
e.__setitem__(self, key, value)
   3947     self._setitem_array([key], value)
   3948 else:
   3949     # set column
-> 3950     self._set_item(key, value)

File ~\anaconda3\Lib\site-packages\pandas\core\frame.py:4143, in DataFram
e._set_item(self, key, value)
   4133 def _set_item(self, key, value) -> None:
   4134     """
   4135     Add series to DataFrame in specified column.
   4136
   (...)
   4141     ensure homogeneity.
   4142     """
-> 4143     value = self._sanitize_column(value)
   4145     if (
   4146         key in self.columns
   4147         and value.ndim == 1
   4148         and not is_extension_array_dtype(value)
   4149     ):
   4150         # broadcast across multiple columns if necessary
   4151         if not self.columns.is_unique or isinstance(self.columns,
MultiIndex):

File ~\anaconda3\Lib\site-packages\pandas\core\frame.py:4870, in DataFram
e._sanitize_column(self, value)
   4867     return _reindex_for_setitem(Series(value), self.index)
   4869 if is_list_like(value):
-> 4870     com.require_length_match(value, self.index)
   4871 return sanitize_array(value, self.index, copy=True, allow_2d=True)

File ~\anaconda3\Lib\site-packages\pandas\core\common.py:576, in require_l
ength_match(data, index)
    572 """
    573 Check the length of data matches the length of the index.
    574 """
    575 if len(data) != len(index):
--> 576     raise ValueError(
    577         "Length of values "
    578         f"({len(data)}) "
    579         "does not match length of index "
    580         f"({len(index)})"
    581     )

ValueError: Length of values (4) does not match length of index (3)
```

Step – 6 :

How to Remove colomun

In [ ]:
```
# What is your data frame name : names

************** axis ********************
# i want to remove coloumn
axis=1 ===> coloumn
axis=0 ===> rows

*****************************************************
# What is the name of the coloumn you want remove : city

**************************** inplace ****************************
# When you drop the coloumn means= you are modifing the tabel
# The data frame overwrite or you want to create new dataframe

# inplace=true (data will be overwrite)
```

In [27]:
```
names.drop('city',axis=1,inplace=True)
```

In [28]:
```
names
```

Out[28]:

	Name	Age
0	Ram	30
1	Raheem	31
2	Robert	32

How to create dataframe using dictionary

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

dict1
```

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

In [31]:
```
pd.DataFrame(dict1)

# keys automatically consider as columns
# values automatically consider as rows
```

Out[31]:

	Names	Age
0	Ram	30
1	Raheem	31
2	Robert	32

In [34]:
```
pd.DataFrame(dict1,index=['A','B','C'])   # change the rows names
```

Out[34]:

	Names	Age
A	Ram	30
B	Raheem	31
C	Robert	32

In [36]:
```
dict1={'Names':['Ram','Raheem','Robert'],
       'Age':[30,31,32]}

pd.DataFrame(dict1,columns=['A','B'])
```

Out[36]:

	A	B

In [37]:
```
dict1={'Names':['Ram','Raheem','Robert'],
       'Age':[30,31,32]}

pd.DataFrame(dict1,columns=['Names','Age'])
```

Out[37]:

	Names	Age
0	Ram	30
1	Raheem	31
2	Robert	32

In [ ]:
```

```