Name : Mohit Manish Bhavsar

Roll No : 20U437

Div : 4

```
pip install pandas
```

```
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (1.1
Requirement already satisfied: numpy>=1.15.4 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/di
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (f
```

```
import csv
file=open("names.csv","r")
csv_reader=csv.reader(file)
list_from_csv=[]
for row in csv_reader:
  list_from_csv.append(row)
print(list_from_csv)
```

```
[['Name', 'Marks', 'Age', 'Place'], ['Maithili', '90', '20', 'Jalgoan'], ['Viahsnavi
```

```
import csv
file=open("Placement_Data_Full_Class.csv","r")
csv_reader=csv.reader(file)
list_from_csv=[]
for row in csv_reader:
  list_from_csv.append(row)
  print(list_from_csv)
```

```
IOPub data rate exceeded.
The notebook server will temporarily stop sending output
to the client in order to avoid crashing it.
To change this limit, set the config variable
`--NotebookApp.iopub_data_rate_limit`.

Current values:
NotebookApp.iopub_data_rate_limit=1000000.0 (bytes/sec)
NotebookApp.rate_limit_window=3.0 (secs)
```

```
import pandas as pd
from google.colab import files
uploaded= files.upload()
```

Choose Files No file chosen Upload widget is only available when the cell has been execute

```
Saving Placement_Data_Full_Class.csv to Placement_Data_Full_Class (1).csv
```

```
df=pd.read_csv('Placement_Data_Full_Class.csv')
```

```
print(df)
```

```
     sl_no gender  ssc_p    ssc_b  ...  specialisation  mba_p      status    salary
0        1      M  67.00   Others  ...          Mkt&HR  58.80      Placed  270000.0
1        2      M  79.33  Central  ...         Mkt&Fin  66.28      Placed  200000.0
2        3      M  65.00  Central  ...         Mkt&Fin  57.80      Placed  250000.0
3        4      M  56.00  Central  ...          Mkt&HR  59.43  Not Placed       NaN
4        5      M  85.80  Central  ...         Mkt&Fin  55.50      Placed  425000.0
..     ...    ...    ...      ...  ...             ...    ...         ...       ...
210    211      M  80.60   Others  ...         Mkt&Fin  74.49      Placed  400000.0
211    212      M  58.00   Others  ...         Mkt&Fin  53.62      Placed  275000.0
212    213      M  67.00   Others  ...         Mkt&Fin  69.72      Placed  295000.0
213    214      F  74.00   Others  ...          Mkt&HR  60.23      Placed  204000.0
214    215      M  62.00  Central  ...          Mkt&HR  60.22  Not Placed       NaN

[215 rows x 15 columns]
```

```
df.head()
```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	wor
0	1	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	
1	2	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	
2	3	M	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	
3	4	M	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	
4	5	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	

```
df.tail(10)
```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	w
205	206	M	61.00	Others	62.0	Others	Commerce	65.0	Comm&Mgmt	
206	207	M	41.00	Central	42.0	Central	Science	60.0	Comm&Mgmt	
207	208	M	83.33	Central	78.0	Others	Commerce	61.0	Comm&Mgmt	
208	209	F	43.00	Central	60.0	Others	Science	65.0	Comm&Mgmt	
209	210	M	62.00	Central	72.0	Central	Commerce	65.0	Comm&Mgmt	
210	211	M	80.60	Others	82.0	Others	Commerce	77.6	Comm&Mgmt	

Double-click (or enter) to edit

```
df.columns
```

```
Index(['sl_no', 'gender', 'ssc_p', 'ssc_b', 'hsc_p', 'hsc_b', 'hsc_s',
       'degree_p', 'degree_t', 'workex', 'etest_p', 'specialisation', 'mba_p',
       'status', 'salary'],
      dtype='object')
```

```
df[['gender', 'ssc_p']]
```

	gender	ssc_p
0	M	67.00
1	M	79.33
2	M	65.00
3	M	56.00
4	M	85.80
...	...	...
210	M	80.60
211	M	58.00
212	M	67.00
213	F	74.00
214	M	62.00

215 rows × 2 columns

```
df.describe()
```

	sl_no	ssc_p	hsc_p	degree_p	etest_p	mba_p	s
count	215.000000	215.000000	215.000000	215.000000	215.000000	215.000000	148.0
mean	108.000000	67.303395	66.333163	66.370186	72.100558	62.278186	288655.4
std	62.209324	10.827205	10.897509	7.358743	13.275956	5.833385	93457.4
min	1.000000	40.890000	37.000000	50.000000	50.000000	51.210000	200000.0
25%	54.500000	60.600000	60.900000	61.000000	60.000000	57.945000	240000.0
50%	108.000000	67.000000	65.000000	66.000000	71.000000	62.000000	265000.0
75%	161.500000	75.700000	73.000000	72.000000	83.500000	66.255000	300000.0
max	215.000000	89.400000	97.700000	91.000000	98.000000	77.890000	940000.0

Double-click (or enter) to edit

```
df.count()
```

```
sl_no             215
gender            215
ssc_p             215
ssc_b             215
hsc_p             215
hsc_b             215
hsc_s             215
degree_p          215
degree_t          215
workex            215
etest_p           215
specialisation    215
mba_p             215
status            215
salary            148
dtype: int64
```

```
df.isnull()
```

	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	et
0	False	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...

```
df.isnull().count()
```

```
sl_no             215
gender            215
ssc_p             215
ssc_b             215
hsc_p             215
hsc_b             215
hsc_s             215
degree_p          215
degree_t          215
workex            215
etest_p           215
specialisation    215
mba_p             215
status            215
salary            215
dtype: int64
```

```
di={'Roll':[1,2,3,4,5],
    'Name':['Maithili','Vaishnavi','Siddhant','Diksha','Mohit'],
    'Marks':['First','First','Second','Distinction','Second'],}
import pandas as pd
df=pd.DataFrame(di)
df
```

	Roll	Name	Marks
0	1	Maithili	First
1	2	Vaishnavi	First
2	3	Siddhant	Second
3	4	Diksha	Distinction
4	5	Mohit	Second

```
df.shape
```

```
(5, 3)
```

```
df.columns
```

```
Index(['Roll', 'Name', 'Marks'], dtype='object')
```

```
df.size
```

```
15
```

```
df.dtypes
```

```
Roll      int64
Name     object
Marks    object
dtype: object
```

```
df.astype({'Roll':'int32'}).dtypes
```

```
Roll      int32
Name     object
Marks    object
dtype: object
```

```
df.dtypes
```

```
Roll      int64
Name     object
Marks    object
dtype: object
```

```
df.astype({'Name':'string'}).dtypes
```

```
Roll      int64
Name     string
Marks    object
dtype: object
```

```
df.dtypes
```

```
Roll      int64
Name     object
Marks    object
dtype: object
```

```
df=df.astype({'Name':'string','Roll':'int32'}).dtypes
```

```
df
```

```
Roll      int32
Name     string
Marks    object
dtype: object
```

```
df=pd.DataFrame(di)
```

```
df
```

	Roll	Name	Marks
0	1	Maithili	First
1	2	Vaishnavi	First
2	3	Siddhant	Second
3	4	Diksha	Distinction
4	5	Mohit	Second

```
df['Marks'].replace(['Distinction','First','Second'],[92,82,72],inplace=True)
```

```
df
```

	Roll	Name	Marks
0	1	Maithili	82
1	2	Vaishnavi	82
2	3	Siddhant	72
3	4	Diksha	92
4	5	Mohit	72

```
df['Marks']=df['Marks'].astype('category')
```

```
df.dtypes
```

```
Roll        int64
Name       object
Marks    category
dtype: object
```

```
df['Marks']=df['Marks'].cat.codes
```

```
df
```

	Roll	Name	Marks
0	1	Maithili	1
1	2	Vaishnavi	1
2	3	Siddhant	0
3	4	Diksha	2
4	5	Mohit	0