```
In [1]: import numpy as np #numerical analysis
        import pandas as pd
```
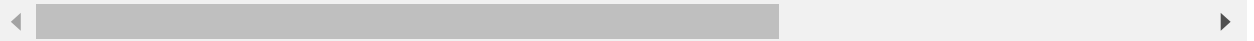
```
In [2]: df=pd.read_csv('bank.csv')
```

```
In [52]: df
```

Out[52]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month
0	59	admin.	1	secondary	no	2343	yes	no	unknown	5	may
1	56	admin.	1	secondary	no	45	no	no	unknown	5	may
2	41	technician	1	secondary	no	1270	yes	no	unknown	5	may
3	55	services	1	secondary	no	2476	yes	no	unknown	5	may
4	54	admin.	1	tertiary	no	184	no	no	unknown	5	may
...	...	...	...	...	...	...	...	...	...	...	...
11157	33	blue-collar	2	primary	no	1	yes	no	cellular	20	apr
11158	39	services	1	secondary	no	733	no	no	unknown	16	jun
11159	32	technician	2	secondary	no	29	no	no	cellular	19	aug
11160	43	technician	1	secondary	no	0	no	yes	cellular	8	may
11161	34	technician	1	secondary	no	0	no	no	cellular	9	jul

11162 rows × 17 columns

```
In [4]: df.head()
```

Out[4]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month	dura
0	59	admin.	married	secondary	no	2343	yes	no	unknown	5	may	
1	56	admin.	married	secondary	no	45	no	no	unknown	5	may	
2	41	technician	married	secondary	no	1270	yes	no	unknown	5	may	
3	55	services	married	secondary	no	2476	yes	no	unknown	5	may	
4	54	admin.	married	tertiary	no	184	no	no	unknown	5	may	

In [5]: `df.tail()`

Out[5]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month
11157	33	blue-collar	single	primary	no	1	yes	no	cellular	20	apr
11158	39	services	married	secondary	no	733	no	no	unknown	16	jun
11159	32	technician	single	secondary	no	29	no	no	cellular	19	aug
11160	43	technician	married	secondary	no	0	no	yes	cellular	8	may
11161	34	technician	married	secondary	no	0	no	no	cellular	9	jul

In [6]: `df.describe()`

Out[6]:

	age	balance	day	duration	campaign	pdays	prev
count	11162.000000	11162.000000	11162.000000	11162.000000	11162.000000	11162.000000	11162.00
mean	41.231948	1528.538524	15.658036	371.993818	2.508421	51.330407	0.83
std	11.913369	3225.413326	8.420740	347.128386	2.722077	108.758282	2.29
min	18.000000	-6847.000000	1.000000	2.000000	1.000000	-1.000000	0.00
25%	32.000000	122.000000	8.000000	138.000000	1.000000	-1.000000	0.00
50%	39.000000	550.000000	15.000000	255.000000	2.000000	-1.000000	0.00
75%	49.000000	1708.000000	22.000000	496.000000	3.000000	20.750000	1.00
max	95.000000	81204.000000	31.000000	3881.000000	63.000000	854.000000	58.00

In [7]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11162 entries, 0 to 11161
Data columns (total 17 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   age        11162 non-null  int64
 1   job        11162 non-null  object
 2   marital    11162 non-null  object
 3   education  11162 non-null  object
 4   default    11162 non-null  object
 5   balance    11162 non-null  int64
 6   housing    11162 non-null  object
 7   loan       11162 non-null  object
 8   contact    11162 non-null  object
 9   day        11162 non-null  int64
 10  month      11162 non-null  object
 11  duration   11162 non-null  int64
 12  campaign   11162 non-null  int64
 13  pdays      11162 non-null  int64
 14  previous   11162 non-null  int64
 15  poutcome   11162 non-null  object
 16  deposit    11162 non-null  object
dtypes: int64(7), object(10)
memory usage: 1.4+ MB
```

finding null values in your data

In [8]: `df.isnull()`

Out[8]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month	d
0	False	False	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	
11157	False	False	False	False	False	False	False	False	False	False	False	
11158	False	False	False	False	False	False	False	False	False	False	False	
11159	False	False	False	False	False	False	False	False	False	False	False	
11160	False	False	False	False	False	False	False	False	False	False	False	
11161	False	False	False	False	False	False	False	False	False	False	False	

11162 rows × 17 columns

In [9]: `df.isnull().sum()`

Out[9]:
```
age          0
job          0
marital      0
education    0
default      0
balance      0
housing      0
loan         0
contact      0
day          0
month        0
duration     0
campaign     0
pdays        0
previous     0
poutcome     0
deposit      0
dtype: int64
```

In [10]: `df.columns`

Out[10]:
```
Index(['age', 'job', 'marital', 'education', 'default', 'balance', 'housing',
       'loan', 'contact', 'day', 'month', 'duration', 'campaign', 'pdays',
       'previous', 'poutcome', 'deposit'],
      dtype='object')
```

In [13]: `df.loc[:]` *#prints all rows and columns*

Out[13]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month
0	59	admin.	married	secondary	no	2343	yes	no	unknown	5	may
1	56	admin.	married	secondary	no	45	no	no	unknown	5	may
2	41	technician	married	secondary	no	1270	yes	no	unknown	5	may
3	55	services	married	secondary	no	2476	yes	no	unknown	5	may
4	54	admin.	married	tertiary	no	184	no	no	unknown	5	may
...	...	...	...	...	...	...	...	...	...	...	...
11157	33	blue-collar	single	primary	no	1	yes	no	cellular	20	apr
11158	39	services	married	secondary	no	733	no	no	unknown	16	jun
11159	32	technician	single	secondary	no	29	no	no	cellular	19	aug
11160	43	technician	married	secondary	no	0	no	yes	cellular	8	may
11161	34	technician	married	secondary	no	0	no	no	cellular	9	jul

11162 rows × 17 columns

In [14]: `df.loc[:10] #prints first 10 rows and all columns`

Out[14]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month
0	59	admin.	married	secondary	no	2343	yes	no	unknown	5	may
1	56	admin.	married	secondary	no	45	no	no	unknown	5	may
2	41	technician	married	secondary	no	1270	yes	no	unknown	5	may
3	55	services	married	secondary	no	2476	yes	no	unknown	5	may
4	54	admin.	married	tertiary	no	184	no	no	unknown	5	may
5	42	management	single	tertiary	no	0	yes	yes	unknown	5	may
6	56	management	married	tertiary	no	830	yes	yes	unknown	6	may
7	60	retired	divorced	secondary	no	545	yes	no	unknown	6	may
8	37	technician	married	secondary	no	1	yes	no	unknown	6	may
9	28	services	single	secondary	no	5090	yes	no	unknown	6	may
10	38	admin.	single	secondary	no	100	yes	no	unknown	7	may

In [16]: `df.age<60`

Out[16]:
```
0        True
1        True
2        True
3        True
4        True
         ...
11157    True
11158    True
11159    True
11160    True
11161    True
Name: age, Length: 11162, dtype: bool
```

In [17]: `df.balance>1000`

Out[17]:
```
0         True
1        False
2         True
3         True
4        False
         ...
11157    False
11158    False
11159    False
11160    False
11161    False
Name: balance, Length: 11162, dtype: bool
```

In [18]: `df.iloc[:,:] #prints all rows and column`

Out[18]:

	age	job	marital	education	default	balance	housing	loan	contact	day	mon
0	59	admin.	married	secondary	no	2343	yes	no	unknown	5	ma
1	56	admin.	married	secondary	no	45	no	no	unknown	5	ma
2	41	technician	married	secondary	no	1270	yes	no	unknown	5	ma
3	55	services	married	secondary	no	2476	yes	no	unknown	5	ma
4	54	admin.	married	tertiary	no	184	no	no	unknown	5	ma
...	...	...	...	...	...	...	...	...	...	...	
11157	33	blue-collar	single	primary	no	1	yes	no	cellular	20	a
11158	39	services	married	secondary	no	733	no	no	unknown	16	ju
11159	32	technician	single	secondary	no	29	no	no	cellular	19	au
11160	43	technician	married	secondary	no	0	no	yes	cellular	8	ma
11161	34	technician	married	secondary	no	0	no	no	cellular	9	j

11162 rows × 17 columns

In [20]: `df.iloc[0:10,:6] #prints first 10 rows and first 6 columns`

Out[20]:

	age	job	marital	education	default	balance
0	59	admin.	married	secondary	no	2343
1	56	admin.	married	secondary	no	45
2	41	technician	married	secondary	no	1270
3	55	services	married	secondary	no	2476
4	54	admin.	married	tertiary	no	184
5	42	management	single	tertiary	no	0
6	56	management	married	tertiary	no	830
7	60	retired	divorced	secondary	no	545
8	37	technician	married	secondary	no	1
9	28	services	single	secondary	no	5090

In [22]: `df.iloc[50:75,0:4] #prints 50-75 rows with 4 columns`

Out[22]:

	age	job	marital	education
50	41	blue-collar	single	primary
51	39	management	divorced	tertiary
52	59	retired	married	secondary
53	41	blue-collar	married	secondary
54	48	blue-collar	married	secondary
55	40	admin.	married	secondary
56	48	blue-collar	married	secondary
57	60	retired	married	primary
58	40	technician	single	unknown
59	57	technician	married	tertiary
60	51	blue-collar	married	secondary
61	41	blue-collar	divorced	secondary
62	41	blue-collar	married	secondary
63	52	blue-collar	divorced	primary
64	59	blue-collar	married	primary
65	44	blue-collar	married	unknown
66	49	unknown	married	primary
67	40	services	married	primary
68	41	admin.	married	primary
69	44	technician	married	tertiary
70	60	services	single	primary
71	29	management	married	tertiary
72	41	retired	divorced	primary
73	41	blue-collar	married	primary
74	42	technician	single	secondary

In [23]: `df.iloc[100:150,:]` *#prints 100-150 rows with all columns*

Out[23]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month
100	29	services	single	secondary	no	271	yes	no	unknown	20	may
101	35	blue-collar	married	primary	no	102	yes	no	unknown	20	may
102	31	blue-collar	married	secondary	no	2	yes	no	unknown	20	may
103	35	services	married	secondary	no	4170	yes	no	unknown	20	may
104	34	technician	single	secondary	no	85	yes	no	unknown	20	may
105	31	admin.	single	secondary	no	431	yes	yes	unknown	20	may
106	44	technician	divorced	secondary	no	982	yes	no	unknown	20	may
107	36	blue-collar	married	primary	no	408	yes	yes	unknown	20	may
108	35	blue-collar	married	unknown	no	4822	yes	no	unknown	20	may
109	41	blue-collar	married	primary	no	1250	yes	no	unknown	20	may
110	31	blue-collar	single	primary	no	216	yes	no	unknown	21	may
111	34	blue-collar	married	secondary	no	1207	yes	no	unknown	21	may
112	32	technician	married	secondary	no	791	yes	no	unknown	21	may
113	36	management	married	tertiary	no	849	yes	yes	unknown	21	may
114	30	blue-collar	single	secondary	yes	239	yes	no	unknown	21	may
115	37	technician	single	secondary	no	1211	yes	no	unknown	21	may
116	36	blue-collar	married	primary	no	599	yes	no	unknown	21	may
117	31	management	single	tertiary	no	825	yes	no	unknown	21	may
118	27	technician	married	secondary	no	2183	yes	yes	unknown	21	may
119	34	retired	married	primary	no	4499	no	no	unknown	21	may
120	30	technician	single	tertiary	no	1289	yes	no	unknown	21	may
121	32	technician	single	tertiary	no	4665	yes	no	unknown	21	may
122	37	technician	single	secondary	no	3326	yes	no	unknown	21	may
123	32	blue-collar	married	secondary	no	783	yes	no	unknown	21	may
124	33	blue-collar	married	secondary	no	0	yes	no	unknown	21	may
125	35	blue-collar	married	primary	no	994	yes	no	unknown	23	may
126	35	self-employed	divorced	tertiary	no	1354	yes	no	unknown	23	may
127	30	blue-collar	single	secondary	no	239	yes	no	unknown	23	may
128	45	blue-collar	divorced	primary	no	-311	yes	no	unknown	23	may
129	45	admin.	married	secondary	no	149	yes	no	unknown	23	may
130	33	blue-collar	married	secondary	no	1464	yes	no	unknown	23	may
131	44	management	divorced	tertiary	no	5773	no	no	unknown	23	may
132	40	blue-collar	married	primary	no	278	yes	no	unknown	23	may

	age	job	marital	education	default	balance	housing	loan	contact	day	month
133	31	blue-collar	married	secondary	no	2910	yes	no	unknown	23	may
134	30	technician	married	tertiary	no	541	yes	no	unknown	26	may
135	44	technician	divorced	secondary	no	1262	yes	no	unknown	26	may
136	34	services	married	secondary	no	-538	yes	no	unknown	26	may
137	37	blue-collar	married	secondary	no	125	no	no	unknown	26	may
138	28	self-employed	single	tertiary	no	0	yes	no	unknown	26	may
139	32	admin.	single	secondary	no	620	yes	no	unknown	26	may
140	37	blue-collar	divorced	secondary	no	316	yes	no	unknown	26	may
141	34	management	married	tertiary	no	2287	yes	no	unknown	26	may
142	38	admin.	divorced	secondary	no	198	yes	no	unknown	26	may
143	28	admin.	single	secondary	no	460	yes	no	unknown	26	may
144	31	management	divorced	tertiary	no	1145	yes	no	unknown	26	may
145	59	services	married	secondary	no	-22	yes	no	unknown	26	may
146	24	blue-collar	married	secondary	no	685	yes	no	unknown	26	may
147	26	services	single	secondary	no	901	yes	no	unknown	26	may
148	59	retired	single	secondary	no	351	yes	no	unknown	27	may
149	30	entrepreneur	single	primary	no	0	yes	yes	unknown	27	may

In [26]: df

Out[26]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month
0	59	admin.	married	secondary	no	2343	yes	no	unknown	5	may
1	56	admin.	married	secondary	no	45	no	no	unknown	5	may
2	41	technician	married	secondary	no	1270	yes	no	unknown	5	may
3	55	services	married	secondary	no	2476	yes	no	unknown	5	may
4	54	admin.	married	tertiary	no	184	no	no	unknown	5	may
...	...	...	...	...	...	...	...	...	...	...	...
11157	33	blue-collar	single	primary	no	1	yes	no	cellular	20	apr
11158	39	services	married	secondary	no	733	no	no	unknown	16	jun
11159	32	technician	single	secondary	no	29	no	no	cellular	19	aug
11160	43	technician	married	secondary	no	0	no	yes	cellular	8	may
11161	34	technician	married	secondary	no	0	no	no	cellular	9	jul

11162 rows × 17 columns

In [27]: df.columns

Out[27]: Index(['age', 'job', 'marital', 'education', 'default', 'balance', 'housing',
        'loan', 'contact', 'day', 'month', 'duration', 'campaign', 'pdays',
        'previous', 'poutcome', 'deposit'],
       dtype='object')

In [29]: df.age.value_counts()

Out[29]: 31   496
32   477
34   466
33   464
35   461
      ...
92   2
93   2
90   2
89   1
95   1
Name: age, Length: 76, dtype: int64

In [30]: `df.poutcome.value_counts()`

Out[30]:
```
unknown    8326
failure    1228
success    1071
other       537
Name: poutcome, dtype: int64
```

In [32]: `df.balance.value_counts()`

Out[32]:
```
0        774
1         39
3         35
2         34
4         29
        ...
4641       1
2975       1
918        1
2959       1
7561       1
Name: balance, Length: 3805, dtype: int64
```

```
In [33]: df.day.value_counts()
```

```
Out[33]: 20    570
         18    548
         30    478
         5     477
         15    466
         14    463
         13    453
         21    452
         6     447
         12    445
         8     419
         17    411
         28    410
         4     402
         29    388
         19    384
         7     382
         11    373
         16    369
         9     364
         2     334
         3     306
         27    284
         22    269
         26    252
         23    245
         25    224
         10    163
         31    140
         24    122
         1     122
         Name: day, dtype: int64
```

```
In [34]: df.deposit.value_counts()
```

```
Out[34]: no     5873
         yes    5289
         Name: deposit, dtype: int64
```

```
In [35]: df.housing.value_counts()
```

```
Out[35]: no     5881
         yes    5281
         Name: housing, dtype: int64
```

```
In [36]:   df.pdays.value_counts()
```

```
Out[36]:   -1      8324
           92       106
           182       89
           91        84
           181       81
                    ...
           587        1
           579        1
           515        1
           491        1
           683        1
           Name: pdays, Length: 472, dtype: int64
```

```
In [37]:   df.job.value_counts()
```

```
Out[37]:   management      2566
           blue-collar     1944
           technician      1823
           admin.          1334
           services         923
           retired          778
           self-employed    405
           student          360
           unemployed       357
           entrepreneur     328
           housemaid        274
           unknown           70
           Name: job, dtype: int64
```

```
In [38]:   df.education.value_counts()
```

```
Out[38]:   secondary    5476
           tertiary     3689
           primary      1500
           unknown       497
           Name: education, dtype: int64
```

```
In [48]:   from sklearn.preprocessing import LabelEncoder
           le=LabelEncoder()
           print("Before Label Encoding:",df.marital.value_counts('marital'))
```

```
           Before Label Encoding: married     0.568984
           single      0.315176
           divorced    0.115839
           Name: marital, dtype: float64
```

In [49]:
```python
df['marital']=le.fit_transform(df['marital'])
print("After Label Encoding:",df.marital.value_counts('marital'))
```

```
After Label Encoding: 1     0.568984
2     0.315176
0     0.115839
Name: marital, dtype: float64
```

In [50]:
```python
df
```

Out[50]:

	age	job	marital	education	default	balance	housing	loan	contact	day	month
0	59	admin.	1	secondary	no	2343	yes	no	unknown	5	may
1	56	admin.	1	secondary	no	45	no	no	unknown	5	may
2	41	technician	1	secondary	no	1270	yes	no	unknown	5	may
3	55	services	1	secondary	no	2476	yes	no	unknown	5	may
4	54	admin.	1	tertiary	no	184	no	no	unknown	5	may
...	...	...	...	...	...	...	...	...	...	...	...
11157	33	blue-collar	2	primary	no	1	yes	no	cellular	20	apr
11158	39	services	1	secondary	no	733	no	no	unknown	16	jun
11159	32	technician	2	secondary	no	29	no	no	cellular	19	aug
11160	43	technician	1	secondary	no	0	no	yes	cellular	8	may
11161	34	technician	1	secondary	no	0	no	no	cellular	9	jul

11162 rows × 17 columns

In [51]:
```python
df.marital.value_counts() # 1 for married,2 for single and 0 for divorced
```

Out[51]:
```
1    6351
2    3518
0    1293
Name: marital, dtype: int64
```

In [ ]: