NAME: NERELLA VENKATA RADHAKRISHNA

ID: 190031187

PRACTICAL-3

In [2]:
```
import pandas as pd
import numpy as np
```

In [3]:
```
#1
#Importing Data From a CSV file
csvfile=pd.read_csv('housing.csv')
csvfile
```

Out[3]:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_inco
0	-122.23	37.88	41	880	129.0	322	126	8.32
1	-122.22	37.86	21	7099	1106.0	2401	1138	8.30
2	-122.24	37.85	52	1467	190.0	496	177	7.25
3	-122.25	37.85	52	1274	235.0	558	219	5.64
4	-122.25	37.85	52	1627	280.0	565	259	3.84
...	...	...	...	...	...	...	...	
20635	-121.09	39.48	25	1665	374.0	845	330	1.56
20636	-121.21	39.49	18	697	150.0	356	114	2.55
20637	-121.22	39.43	17	2254	485.0	1007	433	1.70
20638	-121.32	39.43	18	1860	409.0	741	349	1.86
20639	-121.24	39.37	16	2785	616.0	1387	530	2.38

20640 rows × 10 columns

In [4]:
```
#1
#Importing Data From a delimited text file (like TSV)
tsvfile=pd.read_table('housing.csv')
tsvfile
```

Out[4]:

	longitude,latitude,housing_median_age,total_rooms,total_bedrooms,population,households,median_income,median_
0	-122.23,37.88,41,8
1	-122.22,37.86,21,7
2	-122.24,37.85,52,1
3	-122.25,37.85,52,1
4	-122.25,37.85,52,1
...	
20635	-121.09,39.48,25,1
20636	-121.21,39.49,18,(
20637	-121.22,39.43,17,
20638	-121.32,39.43,18,1
20639	-121.24,39.37,16,2

20640 rows × 1 columns

In [5]:
```
#1
#Importing Data From an Excel file
exfile=pd.read_excel(r'houseexcel.xlsx')
exfile
```

Out[5]:

	total_rooms	total_bedrooms	population	households	median_income	median_house_value	ocean_proximity
0	880	129	322	126	8.3252	452600	NEAR BAY
1	7099	1106	2401	1138	8.3014	358500	NEAR BAY
2	1467	190	496	177	7.2574	352100	NEAR BAY
3	1274	235	558	219	5.6431	341300	NEAR BAY
4	1627	280	565	259	3.8462	342200	NEAR BAY
5	919	213	413	193	4.0368	269700	NEAR BAY
6	2535	489	1094	514	3.6591	299200	NEAR BAY
7	3104	687	1157	647	3.1200	241400	NEAR BAY
8	2555	665	1206	595	2.0804	226700	NEAR BAY
9	3549	707	1551	714	3.6912	261100	NEAR BAY
10	2202	434	910	402	3.2031	281500	NEAR BAY
11	3503	752	1504	734	3.2705	241800	NEAR BAY
12	2491	474	1098	468	3.0750	213500	NEAR BAY
13	696	191	345	174	2.6736	191300	NEAR BAY
14	2643	626	1212	620	1.9167	159200	NEAR BAY
15	1120	283	697	264	2.1250	140000	NEAR BAY
16	1966	347	793	331	2.7750	152500	NEAR BAY
17	1228	293	648	303	2.1202	155500	NEAR BAY
18	2239	455	990	419	1.9911	158700	NEAR BAY

In [8]:
```
#1
#Importing Data  from a JSON formatted string
jsonfile=pd.read_json('housejson.json')
jsonfile
```

Out[8]:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_inco
0	-122.23	37.88	41	880	129.0	322	126	8.32
1	-122.22	37.86	21	7099	1106.0	2401	1138	8.30
2	-122.24	37.85	52	1467	190.0	496	177	7.25
3	-122.25	37.85	52	1274	235.0	558	219	5.64
4	-122.25	37.85	52	1627	280.0	565	259	3.84
...	...	...	...	...	...	...	...	
20635	-121.09	39.48	25	1665	374.0	845	330	1.56
20636	-121.21	39.49	18	697	150.0	356	114	2.55
20637	-121.22	39.43	17	2254	485.0	1007	433	1.70
20638	-121.32	39.43	18	1860	409.0	741	349	1.86
20639	-121.24	39.37	16	2785	616.0	1387	530	2.38

20640 rows × 10 columns

In [9]:
```
#2
#Viewing/Inspecting Data
csvfile.describe()
```

Out[9]:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households
count	20640.000000	20640.000000	20640.000000	20640.000000	20433.000000	20640.000000	20640.000000
mean	-119.569704	35.631861	28.639486	2635.763081	537.870553	1425.476744	499.539680
std	2.003532	2.135952	12.585558	2181.615252	421.385070	1132.462122	382.329753
min	-124.350000	32.540000	1.000000	2.000000	1.000000	3.000000	1.000000
25%	-121.800000	33.930000	18.000000	1447.750000	296.000000	787.000000	280.000000
50%	-118.490000	34.260000	29.000000	2127.000000	435.000000	1166.000000	409.000000
75%	-118.010000	37.710000	37.000000	3148.000000	647.000000	1725.000000	605.000000
max	-114.310000	41.950000	52.000000	39320.000000	6445.000000	35682.000000	6082.000000

In [10]:
```
#2
#Viewing/Inspecting Data
csvfile.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20640 entries, 0 to 20639
Data columns (total 10 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   longitude           20640 non-null  float64
 1   latitude            20640 non-null  float64
 2   housing_median_age  20640 non-null  int64  
 3   total_rooms         20640 non-null  int64  
 4   total_bedrooms      20433 non-null  float64
 5   population          20640 non-null  int64  
 6   households          20640 non-null  int64  
 7   median_income       20640 non-null  float64
 8   median_house_value  20640 non-null  int64  
 9   ocean_proximity     20640 non-null  object 
dtypes: float64(4), int64(5), object(1)
memory usage: 1.6+ MB
```

In [11]:
```
#2
#Viewing/Inspecting Data
csvfile.head()
```

Out[11]:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income
0	-122.23	37.88	41	880	129.0	322	126	8.3252
1	-122.22	37.86	21	7099	1106.0	2401	1138	8.3014
2	-122.24	37.85	52	1467	190.0	496	177	7.2574
3	-122.25	37.85	52	1274	235.0	558	219	5.6431
4	-122.25	37.85	52	1627	280.0	565	259	3.8462

In [12]:
```
#2
#Viewing/Inspecting Data
csvfile.tail()
```

Out[12]:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_inco
20635	-121.09	39.48	25	1665	374.0	845	330	1.56
20636	-121.21	39.49	18	697	150.0	356	114	2.55
20637	-121.22	39.43	17	2254	485.0	1007	433	1.70
20638	-121.32	39.43	18	1860	409.0	741	349	1.86
20639	-121.24	39.37	16	2785	616.0	1387	530	2.38

In [13]:
```
#2
#Viewing/Inspecting Data
#Apply
csvfile.apply(pd.Series.value_counts)
```

Out[13]:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_inc
-124.35	1.0	NaN	NaN	NaN	NaN	NaN	NaN	
-124.3	2.0	NaN	NaN	NaN	NaN	NaN	NaN	
-124.27	1.0	NaN	NaN	NaN	NaN	NaN	NaN	
-124.26	1.0	NaN	NaN	NaN	NaN	NaN	NaN	
-124.25	1.0	NaN	NaN	NaN	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	
<1H OCEAN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
INLAND	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
ISLAND	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
NEAR BAY	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
NEAR OCEAN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

24864 rows × 10 columns

In [31]:
```
#3
#Glimpse
#slicing
csvfile.loc[1:8,'Longitude':'total_rooms']
```

Out[31]:

	Longitude	Latitude	housing_median_age	total_rooms
1	-122.22	37.86	21	7099
2	-122.24	37.85	52	1467
3	-122.25	37.85	52	1274
4	-122.25	37.85	52	1627
5	-122.25	37.85	52	919
6	-122.25	37.84	52	2535
7	-122.25	37.84	52	3104
8	-122.26	37.84	42	2555

In [14]:
```
#3
#Data Cleaning
#unique
csvfile.households.unique()
```

Out[14]:
```
array([ 126, 1138,  177, ..., 1767, 1832, 1818])
```

In [15]:
```
#3
#Data Cleaning
#select (First Row)
csvfile.iloc[0,:]
```

Out[15]:
```
longitude              -122.23
latitude                 37.88
housing_median_age          41
total_rooms                880
total_bedrooms             129
population                 322
households                 126
median_income           8.3252
median_house_value      452600
ocean_proximity       NEAR BAY
Name: 0, dtype: object
```

In [16]:
```
#3
#Data Cleaning
#select (By position)
csvfile.iloc[0,:5]
```

Out[16]:
```
longitude            -122.23
latitude               37.88
housing_median_age        41
total_rooms              880
total_bedrooms           129
Name: 0, dtype: object
```

In [17]:
```
#3
#Data Cleaning
#select (By Index)
csvfile.loc[0]
```

Out[17]:
```
longitude              -122.23
latitude                 37.88
housing_median_age          41
total_rooms                880
total_bedrooms             129
population                 322
households                 126
median_income           8.3252
median_house_value      452600
ocean_proximity       NEAR BAY
Name: 0, dtype: object
```

In [18]:
```
#3
#Data Cleaning
#list
csvfile.values.tolist()
```

Out[18]:
```
[[-122.23, 37.88, 41, 880, 129.0, 322, 126, 8.3252, 452600, 'NEAR BAY'],
 [-122.22, 37.86, 21, 7099, 1106.0, 2401, 1138, 8.3014, 358500, 'NEAR BAY'],
 [-122.24, 37.85, 52, 1467, 190.0, 496, 177, 7.2574, 352100, 'NEAR BAY'],
 [-122.25, 37.85, 52, 1274, 235.0, 558, 219, 5.6431, 341300, 'NEAR BAY'],
 [-122.25, 37.85, 52, 1627, 280.0, 565, 259, 3.8462, 342200, 'NEAR BAY'],
 [-122.25, 37.85, 52, 919, 213.0, 413, 193, 4.0368, 269700, 'NEAR BAY'],
 [-122.25, 37.84, 52, 2535, 489.0, 1094, 514, 3.6591, 299200, 'NEAR BAY'],
 [-122.25, 37.84, 52, 3104, 687.0, 1157, 647, 3.12, 241400, 'NEAR BAY'],
 [-122.26, 37.84, 42, 2555, 665.0, 1206, 595, 2.0804, 226700, 'NEAR BAY'],
 [-122.25, 37.84, 52, 3549, 707.0, 1551, 714, 3.6912, 261100, 'NEAR BAY'],
 [-122.26, 37.85, 52, 2202, 434.0, 910, 402, 3.2031, 281500, 'NEAR BAY'],
 [-122.26, 37.85, 52, 3503, 752.0, 1504, 734, 3.2705, 241800, 'NEAR BAY'],
 [-122.26, 37.85, 52, 2491, 474.0, 1098, 468, 3.075, 213500, 'NEAR BAY'],
 [-122.26, 37.84, 52, 696, 191.0, 345, 174, 2.6736, 191300, 'NEAR BAY'],
 [-122.26, 37.85, 52, 2643, 626.0, 1212, 620, 1.9167, 159200, 'NEAR BAY'],
 [-122.26, 37.85, 50, 1120, 283.0, 697, 264, 2.125, 140000, 'NEAR BAY'],
 [-122.27, 37.85, 52, 1966, 347.0, 793, 331, 2.775, 152500, 'NEAR BAY'],
 [-122.27, 37.85, 52, 1228, 293.0, 648, 303, 2.1202, 155500, 'NEAR BAY'],
 [-122.26, 37.84, 50, 2239, 455.0, 990, 419, 1.9911, 158700, 'NEAR BAY'],
```

In [ ]:
```
#3
#Data Cleaning
#Selective renaming
csvfile.rename(columns={'housing_median_age':'Housing_Median_Age'},inplace=True)
csvfile
```

Out[29]:

	longitude	latitude	Housing_Median_Age	total_rooms	total_bedrooms	population	households	median_inco
0	-122.23	37.88	41	880	129.0	322	126	8.3
1	-122.22	37.86	21	7099	1106.0	2401	1138	8.3
2	-122.24	37.85	52	1467	190.0	496	177	7.2
3	-122.25	37.85	52	1274	235.0	558	219	5.6
4	-122.25	37.85	52	1627	280.0	565	259	3.8
...	...	...	...	...	...	...	...	
20635	-121.09	39.48	25	1665	374.0	845	330	1.5
20636	-121.21	39.49	18	697	150.0	356	114	2.5
20637	-121.22	39.43	17	2254	485.0	1007	433	1.7
20638	-121.32	39.43	18	1860	409.0	741	349	1.8
20639	-121.24	39.37	16	2785	616.0	1387	530	2.3

20640 rows × 10 columns

In [20]:
```
#3
#Data Cleaning
#Mass renaming of columns
csvfile.rename(columns={'latitude':'Latitude','longitude':'Longitude'},inplace=True)
csvfile
```

Out[20]:

	Longitude	Latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_inc
0	-122.23	37.88	41	880	129.0	322	126	8.3
1	-122.22	37.86	21	7099	1106.0	2401	1138	8.3
2	-122.24	37.85	52	1467	190.0	496	177	7.2
3	-122.25	37.85	52	1274	235.0	558	219	5.6
4	-122.25	37.85	52	1627	280.0	565	259	3.8
...	...	...	...	...	...	...	...	
20635	-121.09	39.48	25	1665	374.0	845	330	1.5
20636	-121.21	39.49	18	697	150.0	356	114	2.5
20637	-121.22	39.43	17	2254	485.0	1007	433	1.7
20638	-121.32	39.43	18	1860	409.0	741	349	1.8
20639	-121.24	39.37	16	2785	616.0	1387	530	2.3

20640 rows × 10 columns

In [ ]:
```
#3
#Mass renaming of index
csvfile.rename(index=lambda x:x+1)
```

Out[30]:

	longitude	latitude	Housing_Median_Age	total_rooms	total_bedrooms	population	households	median_inco
1	-122.23	37.88	41	880	129.0	322	126	8.3
2	-122.22	37.86	21	7099	1106.0	2401	1138	8.3
3	-122.24	37.85	52	1467	190.0	496	177	7.2
4	-122.25	37.85	52	1274	235.0	558	219	5.6
5	-122.25	37.85	52	1627	280.0	565	259	3.8
...	...	...	...	...	...	...	...	
20636	-121.09	39.48	25	1665	374.0	845	330	1.5
20637	-121.21	39.49	18	697	150.0	356	114	2.5
20638	-121.22	39.43	17	2254	485.0	1007	433	1.7
20639	-121.32	39.43	18	1860	409.0	741	349	1.8
20640	-121.24	39.37	16	2785	616.0	1387	530	2.3

20640 rows × 10 columns

In [ ]:
```
#3
#Change the index
csvfile.rename(index={0:"x",1:"y",2:"z"})
```

Out[32]:

	longitude	latitude	Housing_Median_Age	total_rooms	total_bedrooms	population	households	median_inco
x	-122.23	37.88	41	880	129.0	322	126	8.3
y	-122.22	37.86	21	7099	1106.0	2401	1138	8.3
z	-122.24	37.85	52	1467	190.0	496	177	7.2
3	-122.25	37.85	52	1274	235.0	558	219	5.6
4	-122.25	37.85	52	1627	280.0	565	259	3.8
...	...	...	...	...	...	...	...	
20635	-121.09	39.48	25	1665	374.0	845	330	1.5
20636	-121.21	39.49	18	697	150.0	356	114	2.5
20637	-121.22	39.43	17	2254	485.0	1007	433	1.7
20638	-121.32	39.43	18	1860	409.0	741	349	1.8
20639	-121.24	39.37	16	2785	616.0	1387	530	2.3

20640 rows × 10 columns

In [ ]:
```
#4
#Filter
csvfile[csvfile['households']>150]
```

Out[34]:

	longitude	latitude	Housing_Median_Age	total_rooms	total_bedrooms	population	households	median_inco
1	-122.22	37.86	21	7099	1106.0	2401	1138	8.3
2	-122.24	37.85	52	1467	190.0	496	177	7.2
3	-122.25	37.85	52	1274	235.0	558	219	5.6
4	-122.25	37.85	52	1627	280.0	565	259	3.8
5	-122.25	37.85	52	919	213.0	413	193	4.0
...	...	...	...	...	...	...	...	
20634	-121.56	39.27	28	2332	395.0	1041	344	3.7
20635	-121.09	39.48	25	1665	374.0	845	330	1.5
20637	-121.22	39.43	17	2254	485.0	1007	433	1.7
20638	-121.32	39.43	18	1860	409.0	741	349	1.8
20639	-121.24	39.37	16	2785	616.0	1387	530	2.3

19218 rows × 10 columns

In [22]:
```
#4
#Sort (By a Column)
csvfile.sort_values('Longitude')
```

Out[22]:

	Longitude	Latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_inc
2655	-124.35	40.54	52	1820	300.0	806	270	3.(
1861	-124.30	41.84	17	2677	531.0	1244	456	3.(
1851	-124.30	41.80	19	2672	552.0	1298	478	1.9
2631	-124.27	40.69	36	2349	528.0	1194	465	2.5
2653	-124.26	40.58	52	2217	394.0	907	369	2.3
...	...	...	...	...	...	...	...	
12450	-114.56	33.69	17	720	174.0	333	117	1.(
2780	-114.55	32.80	19	2570	820.0	1431	608	1.2
12447	-114.49	33.97	17	2809	635.0	83	45	1.(
13923	-114.47	34.40	19	7650	1901.0	1129	463	1.8
13924	-114.31	34.19	15	5612	1283.0	1015	472	1.4

20640 rows × 10 columns

In [23]:
```
#4
#Sort (By a Column) Descending Order
csvfile.sort_values('Longitude',ascending=False)
```

Out[23]:

	Longitude	Latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_inc
13924	-114.31	34.19	15	5612	1283.0	1015	472	1.4
13923	-114.47	34.40	19	7650	1901.0	1129	463	1.8
12447	-114.49	33.97	17	2809	635.0	83	45	1.(
2780	-114.55	32.80	19	2570	820.0	1431	608	1.2
12450	-114.56	33.69	17	720	174.0	333	117	1.(
...	...	...	...	...	...	...	...	
2653	-124.26	40.58	52	2217	394.0	907	369	2.3
2631	-124.27	40.69	36	2349	528.0	1194	465	2.5
1851	-124.30	41.80	19	2672	552.0	1298	478	1.9
1861	-124.30	41.84	17	2677	531.0	1244	456	3.(
2655	-124.35	40.54	52	1820	300.0	806	270	3.(

20640 rows × 10 columns

In [25]:
```
#4
#GroupBy
csvfile.groupby("total_rooms")
csvfile
```

Out[25]:

	Longitude	Latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_inc
0	-122.23	37.88	41	880	129.0	322	126	8.3
1	-122.22	37.86	21	7099	1106.0	2401	1138	8.3
2	-122.24	37.85	52	1467	190.0	496	177	7.2
3	-122.25	37.85	52	1274	235.0	558	219	5.6
4	-122.25	37.85	52	1627	280.0	565	259	3.8
...	...	...	...	...	...	...	...	
20635	-121.09	39.48	25	1665	374.0	845	330	1.5
20636	-121.21	39.49	18	697	150.0	356	114	2.5
20637	-121.22	39.43	17	2254	485.0	1007	433	1.7
20638	-121.32	39.43	18	1860	409.0	741	349	1.8
20639	-121.24	39.37	16	2785	616.0	1387	530	2.3

20640 rows × 10 columns

In [26]:
```
#4
#GroupBy
csvfile.groupby(["total_rooms","total_bedrooms"])
csvfile
```

Out[26]:

	Longitude	Latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_inc
0	-122.23	37.88	41	880	129.0	322	126	8.3
1	-122.22	37.86	21	7099	1106.0	2401	1138	8.3
2	-122.24	37.85	52	1467	190.0	496	177	7.2
3	-122.25	37.85	52	1274	235.0	558	219	5.6
4	-122.25	37.85	52	1627	280.0	565	259	3.8
...	...	...	...	...	...	...	...	
20635	-121.09	39.48	25	1665	374.0	845	330	1.5
20636	-121.21	39.49	18	697	150.0	356	114	2.5
20637	-121.22	39.43	17	2254	485.0	1007	433	1.7
20638	-121.32	39.43	18	1860	409.0	741	349	1.8
20639	-121.24	39.37	16	2785	616.0	1387	530	2.3

20640 rows × 10 columns

In [27]:
```
#4
#pivot_table
csvfile.pivot_table(index="total_rooms",values=["total_bedrooms","population"],aggfunc=np.sum)
```

Out[27]:

total_rooms	population	total_bedrooms
2	6	2.0
6	8	2.0
8	13	1.0
11	24	11.0
12	18	4.0
...	...	...
30450	9419	5033.0
32054	15507	5290.0
32627	28566	6445.0
37937	16122	5471.0
39320	16305	6210.0

5926 rows × 2 columns

In [28]:
```
#5
#Perform Exploratory Data Analysis
csvfile.sample(8)
```

Out[28]:

	Longitude	Latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_inc
14324	-117.15	32.72	52	344	177.0	460	147	1.2
6227	-117.93	34.06	28	3342	688.0	2210	647	3.4
12209	-117.02	33.60	7	1972	352.0	964	317	3.2
1158	-121.53	39.53	35	1806	293.0	683	295	4.5
4012	-118.60	34.16	37	3441	584.0	1283	544	4.1
14842	-117.08	32.69	31	2558	487.0	1938	492	3.4
9911	-122.28	38.31	52	58	18.0	48	22	1.7
15162	-117.03	32.96	16	3424	698.0	1940	645	4.1