LAB – 0

1) to load .csv file into the data

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

```
Choose files   housing.csv
housing.csv(text/csv) - 1423529 bytes, last modified: 04/02/2026 - 100% done
Saving housing.csv to housing.csv
{'housing.csv': b'longitude,latitude,housing_median_age,total_rooms,total_bedrooms,population,households,median_income,median_house_value,ocean_proximity\n-
122.23,37.88,41.0,880.0,129.0,322.0,126.0,8.3252,452600.0,NEAR BAY\n-122.22,37.86,21.0,7099.0,1106.0,2401.0,1138.0,8.3014,358500.0,NEAR BAY\n-
122.24,37.85,52.0,1467.0,190.0,496.0,177.0,7.2574,352100.0,NEAR BAY\n-122.25,37.85,52.0,1274.0,235.0,558.0,219.0,5.6431,341300.0,NEAR BAY\n-
122.25,37.85,52.0,1627.0,280.0,565.0,259.0,3.8462,342200.0,NEAR BAY\n-122.25,37.85,52.0,919.0,213.0,413.0,193.0,4.0368,269700.0,NEAR BAY\n-
122.25,37.84,52.0,2535.0,489.0,1094.0,514.0,3.6591,299200.0,NEAR BAY\n-122.25,37.84,52.0,3104.0,687.0,1157.0,647.0,3.12,241400.0,NEAR BAY\n-
122.26,37.84,42.0,2555.0,665.0,1206.0,595.0,2.0804,226700.0,NEAR BAY\n-122.25,37.84,52.0,3549.0,707.0,1551.0,714.0,3.6912,261100.0,NEAR BAY\n-
122.26,37.85,52.0,2202.0,434.0,910.0,402.0,3.2031,281500.0,NEAR BAY\n-122.26,37.85,52.0,3503.0,752.0,1504.0,734.0,3.2705,241800.0,NEAR BAY\n-
122.26,37.85,52.0,2491.0,474.0,1098.0,468.0,3.075,213500.0,NEAR BAY\n-122.26,37.84,52.0,696.0,191.0,345.0,174.0,2.6736,191300.0,NEAR BAY\n-
122.26,37.85,52.0,2643.0,626.0,1212.0,620.0,1.9167,159200.0,NEAR BAY\n-122.26,37.85,50.0,1120.0,283.0,697.0,264.0,2.125,140000.0,NEAR BAY\n-
122.27,37.85,52.0,1966.0,347.0,793.0,331.0,2.775,152500.0,NEAR BAY\n-122.27,37.85,52.0,1228.0,293.0,648.0,303.0,2.1202,155500.0,NEAR BAY\n-
122.26,37.84,50.0,2239.0,455.0,990.0,419.0,1.9911,158700.0,NEAR BAY\n-122.27,37.84,52.0,1503.0,298.0,690.0,275.0,2.6033,162900.0,NEAR BAY\n-
122.27,37.85,40.0,751.0,184.0,409.0,166.0,1.3578,147500.0,NEAR BAY\n-122.27,37.85,42.0,1639.0,367.0,929.0,366.0,1.7135,159800.0,NEAR BAY\n-
122.27.37.84.52.0.2426.0.541.0.1015.0.478.0.1.725.112000.0.NEAR BAY\n-122.27.37.84.52.0.1688.0.337.0.853.0.325.0.2.1806.99700.0.NEAR BAY\n-
```

```
122.21.37.8.39.0.2003.0.500.0.1100.0.464.0.3.0682.156500.0.NEAR BAY\n-122.21.37.8.39.0.2018.0.447.0.1221.0.446.0.3.0757.151000.0.NEAR BAY\n-
```

```
# 1) Load the CSV file into a DataFrame
df = pd.read_csv("housing.csv")
```

```
df.head()
```

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value	ocean_proximity
0	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252	452600.0	NEAR BAY
1	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	358500.0	NEAR BAY
2	-122.24	37.85	52.0	1467.0	190.0	496.0	177.0	7.2574	352100.0	NEAR BAY
3	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431	341300.0	NEAR BAY
4	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462	342200.0	NEAR BAY

2)to display information of all columns

```
print("Column Information:")
print(df.info())
```

```
Column Information:
<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  float64
 3   total_rooms         20640 non-null  float64
 4   total_bedrooms      20433 non-null  float64
 5   population          20640 non-null  float64
 6   households          20640 non-null  float64
 7   median_income       20640 non-null  float64
 8   median_house_value  20640 non-null  float64
 9   ocean_proximity     20640 non-null  object
dtypes: float64(9), object(1)
memory usage: 1.6+ MB
None
```

3) to display statistical information of all numericals

```
print("\nStatistical Summary of Numerical Columns:")
print(df.describe())
```

```
Statistical Summary of Numerical Columns:
          longitude    latitude  housing_median_age   total_rooms \
count  20640.000000  20640.000000       20640.000000  20640.000000
mean    -119.569704     35.631861          28.639486   2635.763081
std        2.003532      2.135952          12.585558   2181.615252
min     -124.350000     32.540000           1.000000      2.000000
25%     -121.800000     33.930000          18.000000   1447.750000
50%     -118.490000     34.260000          29.000000   2127.000000
75%     -118.010000     37.710000          37.000000   3148.000000
max     -114.310000     41.950000          52.000000  39320.000000

       total_bedrooms     population     households  median_income \
count    20433.000000   20640.000000   20640.000000   20640.000000
mean       537.870553    1425.476744     499.539680       3.870671
std        421.385070    1132.462122     382.329753       1.899822
min          1.000000       3.000000       1.000000       0.499900
25%        296.000000     787.000000     280.000000       2.563400
50%        435.000000    1166.000000     409.000000       3.534800
75%        647.000000    1725.000000     605.000000       4.743250
max       6445.000000   35682.000000    6082.000000      15.000100

       median_house_value
count         20640.000000
mean         206855.816909
std          115395.615874
min           14999.000000
25%          119600.000000
50%          179700.000000
75%          264725.000000
max          500001.000000
```

4)to display the count of unique lables for ocean proximity column

5) to display which attributes(column)in a dataset have missing values count greater than zero

```python
df['ocean_proximity'].value_counts()
```

	count
ocean_proximity	
<1H OCEAN	9136
INLAND	6551
NEAR OCEAN	2658
NEAR BAY	2290
ISLAND	5

dtype: int64

```python
print("\nColumns with missing values:")
missing_values = df.isnull().sum()
print(missing_values[missing_values > 0])
```

```
Columns with missing values:
total_bedrooms    207
dtype: int64
```