10/3/15   Lab 7

```
import pandas as pd
df = pd. read_csv ('housing.csv')
print ("sample data !!!")
print (df.head())
```

→ Display statistical information of all numerical

```
print (df. describe())
```

→ Display the count of unique labels for
"Ocean Proximity" column

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

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

→ To display which attributes in a dataset
have missing values count greater than zero.
missing_values = df.isnull().sum()

```
print (missing_values [missing_values > 0])
```

total_bedrooms   207

1. Which columns in the dataset had missing
   values? How did you handle them.
   Age, TG
   Handled using simpleImputer with strategy

median, mean.

2. which Categorical columns did you identify
in the dataset? How did you encode them.
Gender
Using Ordinal Encoding

3. what is the difference between Min-Max Scaling
and Standardization? when would you
use one over the other?

In Min-Max Scaling, it transforms data
to fit within a specific range, usually
between 0 and 1

Standardization is used to transform
data to have a mean of 0 and std deviation
of 1. No specific range to fit the data.

Shehad B
10/3/25