

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

1. Load the data

```
In [3]: ## Read the "housing.csv" file from the folder into the program
houseData = pd.read_excel('1553768847_housing.xlsx')
```

```
In [4]: houseData.info()
```

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

```
In [9]: ## Print first few rows of this data.
houseData.head()
```

Out[9]:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population
0	-122.23	37.88	41	880	129.0	322
1	-122.22	37.86	21	7099	1106.0	2401
2	-122.24	37.85	52	1467	190.0	496
3	-122.25	37.85	52	1274	235.0	558
4	-122.25	37.85	52	1627	280.0	565