```
# Imports for all the models needed
import sklearn.tree as tree
import sklearn.ensemble as ensemble
import sklearn.linear_model as lm
from sklearn.model_selection import train_test_split, cross_validate, mean_absolute_error
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.preprocessing import StandardScaler, LabelEncoder, OneHotEncoder
from sklearn.pipeline import Pipeline, FeatureUnion
from sklearn.base import BaseEstimator, TransformerMixin
from sklearn.model_selection import GridSearchCV

# Imports for plotting
import matplotlib.pyplot as plt
import matplotlib.ticker as mticker, cycler, numpy as np
```

Reading Data into Pandas Dataframe

```
# Read in the data into a pandas dataframe
housing = pd.read_csv('data/housing.csv')

# Display summary of the data frame and get a feel for it
print("Dimensions: {0}".format(housing.shape))
housing.info(verbose=True, show_counts=True)
```

```
Dimensions: (20640, 10)
<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
```

```
# Generate histograms of all data
housing.hist(bins=50, figsize=(20,15))
```

```
array([[<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>,
        
```