```
from sklearn.model_selection import train_test_split
X = data.drop(['median_house_value'], axis =1)
y = data['median_house_value']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25)
train_data = X_train.join(y_train)
train_data
train_data.hist(figsize=(15,8))
housing_median_age
                                                                                        2500
                                             6000
  4000
                                                                                        2000
  3000
                                             4000
                                                                                        1500
  2000
                                                                                        1000
                                             2000
  1000
                                                                                         500
                  -120 -118
       -124 -122
                               -116
                                    -114
                                                       34
                                                                                                                    40
                                                                   38
                                                                          40
                                                                                                             30
                                                                                                         population
                                                           total bedrooms
                  total rooms
                                                                                       15000
 12500 -
                                            10000
 10000
                                                                                       10000
                                             7500
  7500
                                             5000
  5000
                                                                                        5000
                                             2500
  2500
    0 -
                                               0
                                                                                           0
              10000
                    20000
                                                           2000
                                                                             6000
                                                                                                    10000
                                                                                                            20000
                                                                                                    median_house_value
                  households
                                                           median_income
                                                                                        3000
                                             5000
 10000
                                             4000
  7500
                                                                                        2000
                                             3000
  5000
                                             2000
                                                                                        1000
  2500
                                             1000
           1000 2000 3000 4000 5000 6000
                                                                                                 100000 200000 300000 400000 500000
                                                 0.0
                                                                     10.0
plt.figure(figsize=(15,8))
sns.scatterplot(x="latitude", y="longitude", data = train_data, hue="median_house_value", palette="coolwarm")
<Axes: xlabel='latitude', ylabel='longitude'>
                                                                                               median_house_value
                                                                                                     200000
                                                                                                     300000
  -116
  -118
   -120
  -122
  -124
```

```
#Feauture Engineering
  train_data['bedroom_ratio'] = train_data['total_bedrooms']/ train_data['total_rooms']
train_data['household_rooms'] = train_data['total_rooms']/train_data['households']
  plt.figure(figsize=(15.8))
  sns.heatmap(train_data.corr(), annot=True, cmap='YlGnBu')
                                                                                                                                                                       1.00
                            1 -0.92 -0.12
                                                                                   -0.016 -0.043 0.32 -0.057 0.0083 -0.48
               longitude -
                                                                                                                                                    -0.07
                                                                                                   -0.45
                                                                   -0.14
                latitude
                           -0.92
                                                  -0.037
                                                          -0.073
                                                                           -0.092
                                                                                   -0.081
                                                                                           -0.15
                                                                                                                   -0.015
                                                                                                                                   -0.16
                                                                                                                                           -0.12
                                                                                                                                                                       0.75
                                                    -0.32
                                                           -0.27
                                                                   -0.25
                                                                           -0.24
                                                                                                           -0.24
                                                                                                                                                   -0.039
    housing_median_age
             total_rooms
                                           -0.32
                                                           0.95
                                                                            0.93
                                                                                                           -0.014
                                                                                                                   -0.01
                                                                                                                          -0.015
                                                                                                                                                   -0.38
         total_bedrooms
                                   -0.073
                                           -0.27
                                                   0.95
                                                                    0.9
                                                                            0.98
                                                                                   0.025
                                                                                                           -0.046
                                                                                                                   0.0054 -0.015
                                                                                                                                           0.68
                                                                                                                                                    -0.55
                                                   0.87
                                                            0.9
                                                                    1
                                                                            0.94
                                                                                                           -0.074 -0.014 -0.057
                                                                                                                                           0.58
                                                                                                                                                   -0.62
                                   -0.14
                                           -0.25
                                                                                           -0.02
                                                                                                                                  -0.016
              population
                                           -0.24
                                                                                                                   -0.011 -0.0037
                                                                                                                                                    -0.65
             households
                                  -0.092
                                                                                                           -0.087
         median_income -
                                           -0.12
                                                                                                           -0.24
                                                                                                                                           -0.51
    median_house_value
                           -0.043
                                   -0.15
                                                                   -0.02
                                                                                                           -0.49
                                                                                                                                            -0.19
                                   -0.45
                                                                                                                   -0.012 -0.31
                                                                                                                                   -0.34
            <1H OCEAN
                                                                                                           -0.61
                                                                                                                                                   -0.13
                                                  -0.014 -0.046 -0.074 -0.087 -0.24
                                                                                           -0.49
                                                                                                   -0.61
                           -0.057
                                           -0.24
                                                                                                                   0.0096 -0.24
                                                                                                                                   -0.26
                                                                                                                                            -0.1
                 INLAND
                                   -0.015
                                                   -0.01
                                                           0.0054 -0.014
                                                                           -0.011 -0.0084
                                                                                                   0.012
                 ISLAND
                                                  -0.015 -0.015 -0.057
                                                                                                   -0.31
                                                                                                           -0.24
                                                                                                                                   -0.13
                                                                                                                                                                       -0.50
           NEAR OCEAN
                                   -0.16
                                                                  -0.016
                                                                                                   -0.34
                                                                                                           -0.26
                                                                                                                          -0.13
                                                                                                                                                   -0.039
                                                                                   -0.51 -0.19
          bedroom_ratio
                                   -0.12
                                          -0.032
                                                                   0.58
                                                                                                            -0.1
                                                                                                                           -0.012
                                                                                                                                                   -0.74
                                                                                                                                                                       -0.75
                           -0.07
                                           -0.039
                                                   -0.38
                                                                   -0.62
                                                                                                   -0.13
                                                                                                                           -0.017
                                                           -0.55
                                                                            -0.65
                                                                                                                                   -0.039
       household rooms
                                                                                                    OCEAN
                                                                                                                    ISLAND
                                                                                                                            BAY
                                                                                                                                     OCEAN
                                                                                     income
                                                                                                                            VEAR B
                                             median
                                                                                                                                     NEAR
from sklearn.linear_model import LinearRegression
  X_train, y_train = train_data.drop(['median_house_value'], axis =1), train_data['median_house_value']
  reg = LinearRegression()
  reg.fit(X_train, y_train)

    tinearRegression
   LinearRegression()
test_data = X_test.join(y_test)
   #Data Preprocessing
   test_data['total_rooms'] = np.log(test_data['total_rooms']+1)
  test_data['total_bedrooms'] = np.log(test_data['total_bedrooms']+1)
test_data['population'] = np.log(test_data['population']+1)
test_data['households'] = np.log(test_data['households']+1)
```

```
test_data['population'] = np.log(test_data['population']+1)
test_data['households'] = np.log(test_data['households']+1)

test_data = test_data.join(pd.get_dummies(test_data.ocean_proximity)).drop(['ocean_proximity'], axis =1) #Made the different strut
#Feauture Engineering
test_data['bedroom_ratio'] = test_data['total_bedrooms']/ test_data['total_rooms']
test_data['household_rooms'] = test_data['total_rooms']/test_data['households']

**

**X_test, y_test = test_data.drop(['median_house_value'], axis =1), test_data['median_house_value']
: reg.score(X_test, y_test)
```

0.673389329293721

```
#Random Forest & Hyper Parameter Tuning
from sklearn.ensemble import RandomForestRegressor
forest = RandomForestRegressor()
forest.fit(X_train, y_train)
▼ RandomForestRegressor
RandomForestRegressor()
forest.score(X_test, y_test)
0.8222769738969176
from sklearn.model_selection import GridSearchCV
forest = RandomForestRegressor()
param_grid = {
   "n_estimators": [3,10,30],
    "max_features": [2,4,6,8]
grid_search = GridSearchCV(forest, param_grid, cv=5,
                         scoring="neg_mean_squared_error",
                          return_train_score=True)
grid_search.fit(X_train, y_train)
           GridSearchCV
 ► estimator: RandomForestRegressor
      ► RandomForestRegressor
best_forest = grid_search.best_estimator_
best_forest
                 RandomForestRegressor
RandomForestRegressor(max_features=8, n_estimators=30)
```

```
best_forest.score(X_test, y_test)
```

0.8183072898131833