```
# Import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import os
os.environ['KAGGLE_CONFIG_DIR'] = '/content/drive/My Drive/kaggle'
%cd /content/drive/My Drive/kaggle
     /content/drive/My Drive/kaggle
!kaggle datasets download -d camnugent/california-housing-prices
     california-housing-prices.zip: Skipping, found more recently modified local copy (use --force to force download)
!unzip \*.zip && rm *.zip

    Archive: california-housing-prices.zip

     replace housing.csv? [y]es, [n]o, [A]ll, [N]one, [r]ename: y
       inflating: housing.csv
data = pd.read_csv("/content/drive/MyDrive/kaggle/housing.csv")
```

data

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_h
0	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252	
1	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	
2	-122.24	37.85	52.0	1467.0	190.0	496.0	177.0	7.2574	
3	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431	
4	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462	
20635	-121.09	39.48	25.0	1665.0	374.0	845.0	330.0	1.5603	
20636	-121.21	39.49	18.0	697.0	150.0	356.0	114.0	2.5568	
20637	-121.22	39.43	17.0	2254.0	485.0	1007.0	433.0	1.7000	
20638	-121.32	39.43	18.0	1860.0	409.0	741.0	349.0	1.8672	
20639	-121.24	39.37	16.0	2785.0	616.0	1387.0	530.0	2.3886	

20640 rows × 10 columns

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
     RangeIndex: 20640 entries, 0 to 20639
     Data columns (total 10 columns):
                              Non-Null Count Dtype
     # Column
     ---
      0 longitude
                              20640 non-null float64
          latitude 20640 non-null float64 housing_median_age 20640 non-null float64
          latitude
          total_rooms
                              20640 non-null float64
          total_bedrooms 20433 non-null float64
         population
                              20640 non-null float64
          households
                              20640 non-null float64
          median income
         median_house_value 20640 non-null float64
     9 ocean_proximity 20640 dtypes: float64(9), object(1)
                              20640 non-null object
     memory usage: 1.6+ MB
data.dropna(inplace=True)
data.info()
     <class 'pandas.core.frame.DataFrame'>
     Int64Index: 20433 entries, 0 to 20639
     Data columns (total 10 columns):
```

dtypes: float64(9), object(1
memory usage: 1.7+ MB

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=0.2)

train_data = X_train.join(y_train)

train_data

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	ocean_pr
7415	-118.22	33.94	42.0	1046.0	287.0	1218.0	289.0	2.6538	<1F
3799	-118.36	34.16	45.0	1755.0	335.0	822.0	342.0	5.1423	<1F
6063	-117.83	33.99	14.0	17527.0	2751.0	8380.0	2676.0	6.2734	<1F
92	-122.28	37.80	52.0	96.0	31.0	191.0	34.0	0.7500	NI
15273	-117.28	33.06	8.0	4172.0	1022.0	2585.0	941.0	4.0118	NEAF
					•••				
10040	-121.05	39.20	48.0	1759.0	389.0	716.0	350.0	2.3125	
6539	-118.04	34.04	35.0	1734.0	363.0	1527.0	344.0	3.0000	<1F
6969	-118.05	33.98	41.0	1694.0	413.0	1222.0	387.0	2.8311	<1F
13853	-117.31	34.50	14.0	2443.0	447.0	883.0	465.0	2.1111	
5617	-118.26	33.79	42.0	1162.0	264.0	1044.0	241.0	3.5488	<1F

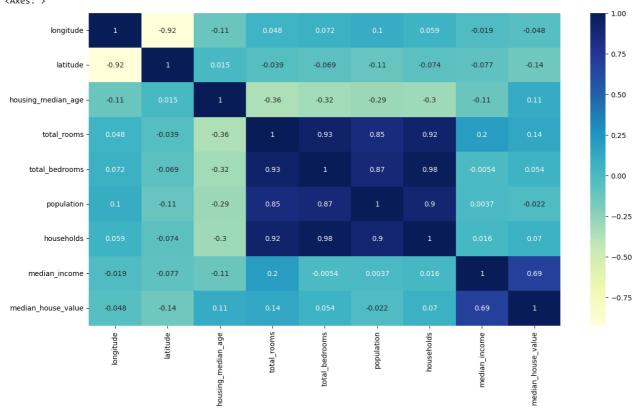
16346 rows × 10 columns

train_data.hist(figsize=(15, 8))

```
array([[<Axes: title={'center': 'longitude'}>,
          <Axes: title={'center': 'population'}>],
           [<Axes: title={'center': 'households'}>,
            <Axes: title={'center': 'median_income'}>
            <Axes: title={'center': 'median_house_value'}>]], dtype=object)
                     longitude
                                                           latitude
                                                                                           housing_median_age
      5000
                                                                                2500
                                           6000
      4000
plt.figure(figsize= (15,8) )
```

sns.heatmap(train_data.corr(), annot=True, cmap= "YlGnBu")

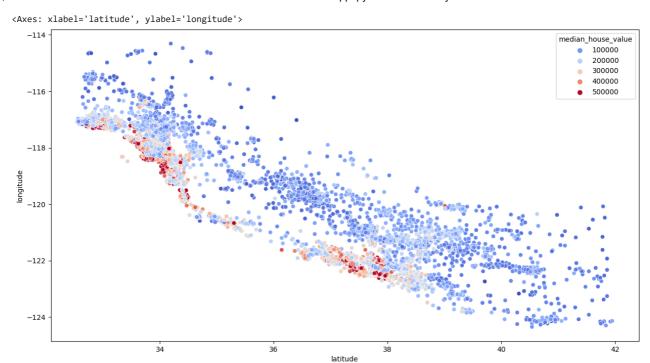
<ipython-input-16-2fd49b6ee71a>:2: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a f sns.heatmap(train_data.corr(), annot=True, cmap= "YlGnBu")



```
# Import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
train_data['total_rooms'] = np.log(train_data['total_rooms'] + 1)
train_data['total_bedrooms'] = np.log(train_data['total_bedrooms'] + 1)
train_data['population'] = np.log(train_data['population'] + 1)
train_data['households'] = np.log(train_data['households'] + 1)
train_data.hist(figsize=(15,8))
```

```
array([[<Axes: title={'center': 'longitude'}>,
                  <Axes: title={'center': 'latitude'}>,
                <Axes: title={'center': 'housing_median_age'}>],
[<Axes: title={'center': 'total_rooms'}>,
                  <Axes: title={'center': 'total_bedrooms'}>,
                 <Axes: title={'center': 'population'}>],
                [<Axes: title={'center': 'households'}>,
                  <Axes: title={'center': 'median_income'}>
                  <Axes: title={'center': 'median_house_value'}>]], dtype=object)
                              longitude
                                                                                       latitude
                                                                                                                                        housing_median_age
        5000
                                                                                                                        2500
                                                                6000
        4000
                                                                                                                        2000
        3000
                                                                4000
                                                                                                                        1500
        2000
                                                                                                                        1000
                                                                2000
        1000
                                                                                                                         500
               -124
                      -122
                              -120
                                      -118
                                              -116
                                                      -114
                                                                             34
                                                                                             38
                                                                                                      40
                                                                                                                                     10
                                                                                                                                             20
                                                                                                                                                     30
                                                                                                                                                                    50
                             total_rooms
                                                                                   total_bedrooms
                                                                                                                                              population
                                                                8000
                                                                                                                        8000
        6000
                                                                6000
                                                                                                                        6000
        4000
                                                                4000
                                                                                                                        4000
        2000
                                                                2000
                                                                                                                        2000
                                                                   0
                                                                                                                           0
                                                    10
                                                                                                                                                                    10
                             households
                                                                                   median income
                                                                                                                                        median house value
                                                                6000
        8000
                                                                                                                        3000
        6000
                                                                4000
                                                                                                                        2000
train_data = train_data.join(pd.get_dummies(train_data.ocean_proximity)).drop(['ocean_proximity'], axis=1)
                                                                     - 1
                                                                                                              \perp
plt.figure(figsize= (15,8) )
sns.heatmap(train_data.corr(), annot=True, cmap= "YlGnBu" )
       <Axes: >
                                                                                                                                                                     1.00
                   longitude
                                                 -0.11
                                                                                            -0.019
                                                                                                     -0.048
                                                                                                                      -0.048
                                                                                                                                        -0.48
                    latitude
                               -0.92
                                                         -0.036
                                                                  -0.071
                                                                           -0.14
                                                                                   -0.091
                                                                                            -0.077
                                                                                                      -0.14
                                                                                                              -0.45
                                                                                                                       0.34
                                                                                                                               -0.017
                                                                                                                                                 -0.16
                                                                                                                                                                     0.75
                                                                                    -0.24
        housing_median_age
                               -0.11
                                                         -0.31
                                                                  -0.27
                                                                           -0.24
                                                                                            -0.11
                                                                                                                       -0.24
                                       -0.036
                                                 -0.31
                total_rooms -
                                                                                    0.93
                                                                                                                       -0.016
                                                                                                                               -0.0084
                                                                                                                                        -0.018
                                                                                                                                                                     0.50
             total_bedrooms -
                                       -0.071
                                                 -0.27
                                                                            0.9
                                                                                            -0.026
                                                                                                                       -0.046
                                                                                                                                        -0.018
                  population
                                        -0.14
                                                 -0.24
                                                                                            -0.008
                                                                                                     -0.021
                                                                                                                      -0.073
                                                                                                                               -0.014
                                                                                                                                        -0.061
                                                                                                                                                 -0.016
                                                                                                                                                                     0.25
                                       -0.091
                                                                  0.98
                                                                                                                      -0.088
                                                                                                                               -0.011
                 households -
                                                 -0.24
                                                                                                                                                                    0.00
                              -0.019
                                       -0.077
                                                 -0.11
                                                                  -0.026
                                                                          -0.008
                                                                                                                       -0.24
                                                                                                                              -0.0081
             median_income -
        median_house_value -
                              -0.048
                                       -0.14
                                                                          -0.021
                                                                                                                       -0.49
                                                                                                                                                                     -0.25
                <1H OCEAN -
                                       -0.45
                                                                                                                       -0.61
                                                                                                                               -0.014
                                                                                                                                        -0.32
                                                                                                                                                 -0.34
                    INLAND -
                              -0.048
                                                 -0.24
                                                         -0.016
                                                                  -0.046
                                                                          -0.073
                                                                                   -0.088
                                                                                            -0.24
                                                                                                     -0.49
                                                                                                              -0.61
                                                                                                                               -0.011
                                                                                                                                         -0.24
                                                                                                                                                 -0.26
                                                                                                                                                                     -0.50
                    ISLAND -
                                       -0.017
                                                        -0.0084
                                                                          -0.014
                                                                                   -0.011
                                                                                           -0.0081
                                                                                                              -0.014
                                                                                                                      -0.011
                  NEAR BAY -
                              -0.48
                                                         -0.018
                                                                  -0.018
                                                                          -0.061
                                                                                                              -0.32
                                                                                                                       -0.24
                                                                                                                                                 -0.14
                                                                                                                                                                     -0.75
               NEAR OCEAN -
                                        -0.16
                                                                          -0.016
                                                                                                              -0.34
                                                                                                                       -0.26
                                                                                                                                        -0.14
                                                           rooms
                                                                                                               OCEAN
                                                                                                                                 ISLAND
                                                                                                                                         NEAR BAY
                                                                                                                                                  OCEAN
                                                  housing_median_age
                                                                   bedrooms
                                                                                                      median_house_value
                                                          total
                                                                                                               <1H
                                                                                                                                                  NEAR
```

```
plt.figure(figsize=(15,8))
sns.scatterplot(x="latitude", y="longitude", data=train_data, hue="median_house_value", palette="coolwarm")
```



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")

<axes:></axes:>																
longitude -	1	-0.92	-0.11	0.033	0.064	0.11	0.059	-0.019	-0.048	0.32	-0.048	0.0094	-0.48	0.045	0.1	-0.071
latitude -	-0.92	1	0.015	-0.036	-0.071	-0.14	-0.091	-0.077	-0.14	-0.45	0.34	-0.017	0.36	-0.16	-0.11	0.14
housing_median_age -	-0.11	0.015	1	-0.31	-0.27	-0.24	-0.24	-0.11	0.11		-0.24		0.26		-0.038	-0.037
total_rooms -		-0.036	-0.31	1	0.95	0.87	0.93	0.21	0.16		-0.016	-0.0084	-0.018		0.41	-0.38
total_bedrooms -		-0.071	-0.27	0.95	1	0.9	0.98	-0.026			-0.046		-0.018		0.68	-0.56
population -	0.11	-0.14	-0.24	0.87	0.9	1	0.93	-0.008	-0.021	0.12	-0.073	-0.014	-0.061	-0.016	0.58	-0.63
households -		-0.091	-0.24	0.93	0.98	0.93	1	0.0034	0.075		-0.088	-0.011			0.64	-0.66
median_income -	-0.019	-0.077	-0.11	0.21	-0.026	-0.008	0.0034	1	0.69	0.17	-0.24	-0.0081			-0.51	0.35
median_house_value -	-0.048	-0.14	0.11	0.16		-0.021		0.69	1	0.26	-0.49	0.021	0.16	0.14	-0.2	0.12
<1H OCEAN -	0.32	-0.45				0.12		0.17	0.26	1	-0.61	-0.014	-0.32	-0.34		-0.13
INLAND -	-0.048	0.34	-0.24	-0.016	-0.046	-0.073	-0.088	-0.24	-0.49	-0.61	1	-0.011	-0.24	-0.26	-0.097	0.18
ISLAND -	0.0094	-0.017		-0.0084		-0.014	-0.011	-0.0081		-0.014	-0.011	1	-0.0056			0.0071
NEAR BAY -	-0.48	0.36	0.26	-0.018	-0.018	-0.061			0.16	-0.32	-0.24	-0.0056	1	-0.14	-0.012	-0.015
NEAR OCEAN -		-0.16		0.0025	0.015	-0.016	0.018		0.14	-0.34	-0.26		-0.14	1	0.034	-0.038
bedroom_ratio -	0.1	-0.11	-0.038	0.41	0.68	0.58	0.64	-0.51	-0.2		-0.097		-0.012		1	-0.76
household_rooms -	-0.071	0.14	-0.037	-0.38	-0.56	-0.63	-0.66	0.35	0.12	-0.13	0.18	0.0071	-0.015	-0.038	-0.76	1
	longitude -	latitude -	using_median_age -	total_rooms -	total_bedrooms -	population -	households -	median_income -	edian_house_value -	<1H OCEAN -	INLAND -	ISLAND -	NEAR BAY -	NEAR OCEAN -	bedroom_ratio -	household_rooms -

from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

1.00

0.75

0.50

- 0.25

0.00

-0.25

-0.50

- -0.75

```
X_train, y_train = train_data.drop(['median_house_value'], axis=1), train_data['median_house_value']
X_train_s = scaler.fit_transform(X_train)
reg = LinearRegression()
reg.fit(X_train_s, y_train)
     ▼ LinearRegression
     LinearRegression()
test_data = X_test.join(y_test)
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 = test_data.join(pd.get_dummies(test_data.ocean_proximity)).drop(['ocean_proximity'], axis=1)
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']
X_test_s = scaler.transform(X_test)
reg.score(X_test_s, y_test)
     0.6662930852567805
from sklearn.ensemble import RandomForestRegressor
forest = RandomForestRegressor()
forest.fit(X_train_s, y_train)
     ▼ RandomForestRegressor
     RandomForestRegressor()
forest.score(X_test_s, y_test)
     0.8055368164709462
from sklearn.model_selection import GridSearchCV
forest = RandomForestRegressor()
param_grid = {
    "n_estimators": [100, 200, 300],
    "min_samples_split": [2, 4],
    "max_depth": [None, 4, 8]
grid_search = GridSearchCV(forest, param_grid, cv=5,
                          scoring="neg_mean_squared_error",
                          return_train_score=True)
{\tt grid\_search.fit(X\_train\_s,\ y\_train)}
                  GridSearchCV
      ▶ estimator: RandomForestRegressor
            ▶ RandomForestRegressor
grid search.best estimator
                          RandomForestRegressor
     RandomForestRegressor(min_samples_split=4, n_estimators=300)
grid_search.best_estimator_.score(X_test_s, y_test)
```

0.8074470465364803

• ×