



PREDICT THE PRICE OF THE HOUSE

TALAL ALARABI

Intordication



-Problem statement:

Due to the information gap about house prices and the difference in house prices from one neighborhood to another, the individual may face a problem estimating house prices and may be subject to deception by real estate owners. Therefore, we will build a model that predicts house prices.

-Project goal:

The main goal of this project is to study the Estates market and to train a model to predict the prices of the estates of and testing it using linear regression models .

EDA



SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority

Data Cleaning :

1-delet null values

2-Delete duplicated

EDA

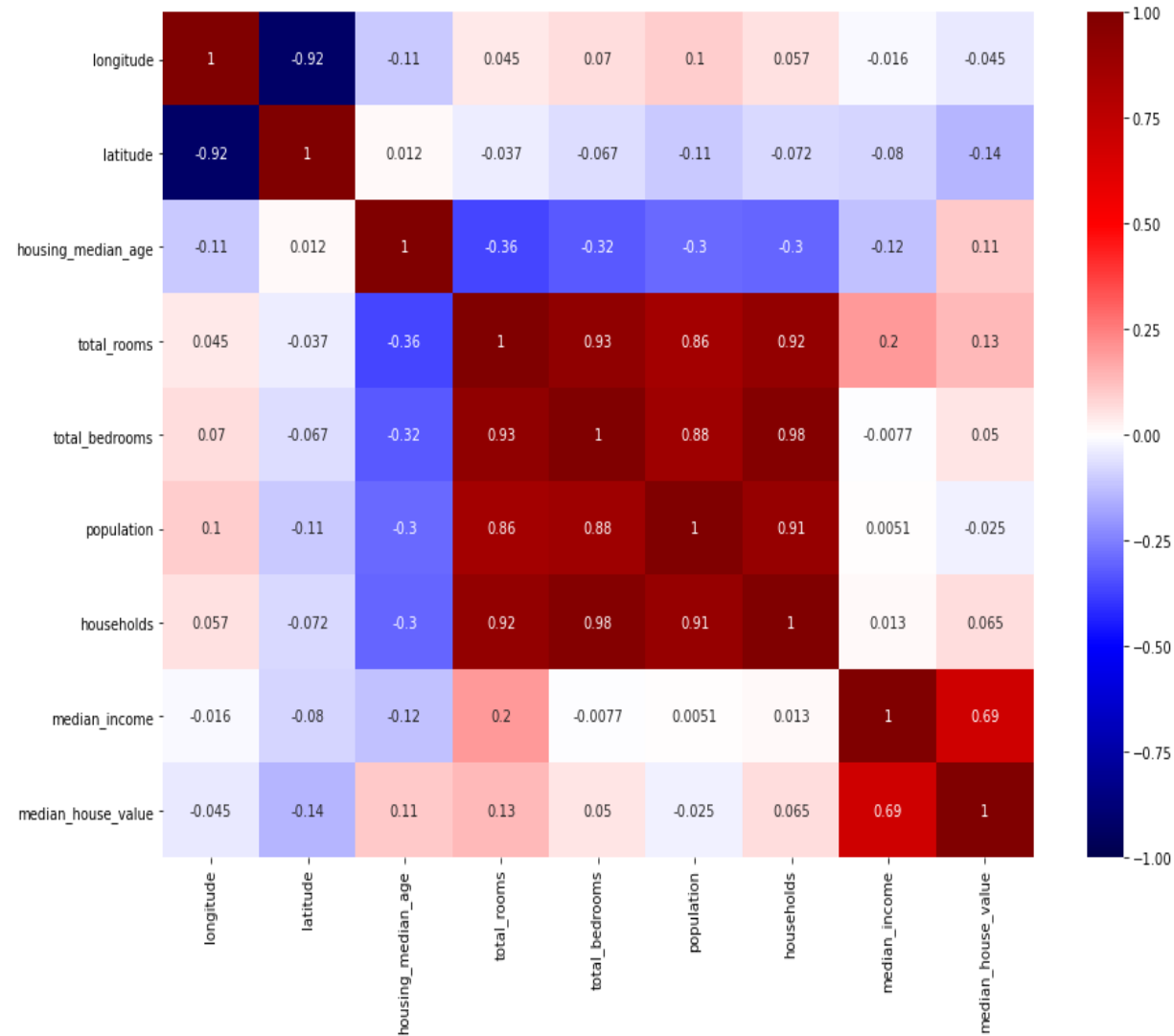
```
In [14]: df = pd.read_csv('housing.csv')  
df
```

Out[14]:

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value	ocean_proximity
0	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252	452600.0	NEAR BAY
1	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	358500.0	NEAR BAY
2	-122.24	37.85	52.0	1487.0	190.0	496.0	177.0	7.2574	352100.0	NEAR BAY
3	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431	341300.0	NEAR BAY
4	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462	342200.0	NEAR BAY
...
20635	-121.09	39.48	25.0	1665.0	374.0	845.0	330.0	1.5603	78100.0	INLAND
20636	-121.21	39.49	18.0	697.0	150.0	356.0	114.0	2.5568	77100.0	INLAND
20637	-121.22	39.43	17.0	2254.0	485.0	1007.0	433.0	1.7000	92300.0	INLAND
20638	-121.32	39.43	18.0	1860.0	409.0	741.0	349.0	1.8672	84700.0	INLAND
20639	-121.24	39.37	16.0	2785.0	616.0	1387.0	530.0	2.3886	89400.0	INLAND

20640 rows × 10 columns

EDA



-Heat Map

Regression Models



SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority

R-squared (Cross-Validation)

Linear Regression

0.65

Polynomial

0.43

Ridge

0.65

Regression Models

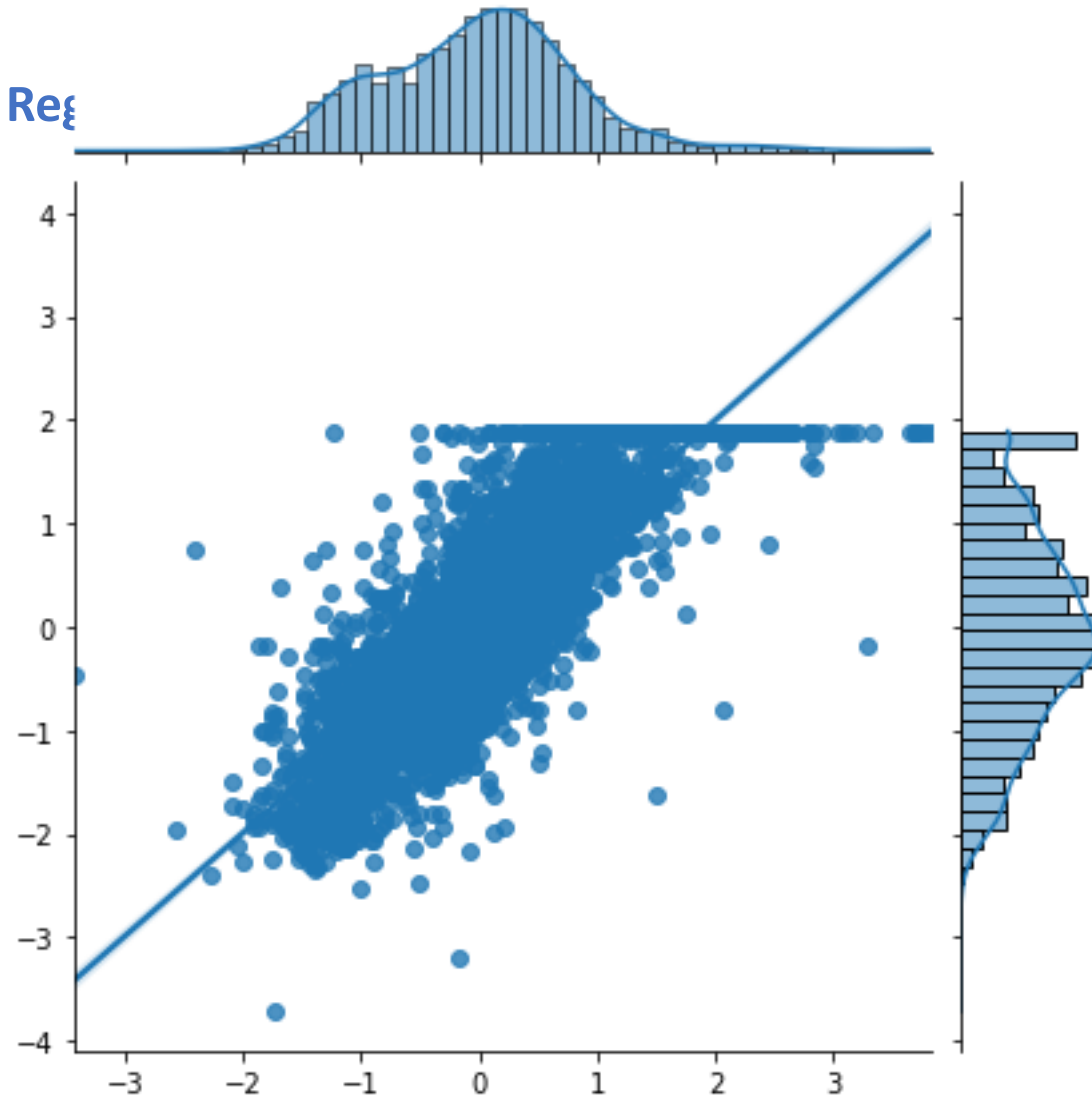


SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority

Prediction after the experiment(Use **Linear Reg**

Training: 65%

Testing: 64%



Conclusion

We have made a model capable of predicting house prices at a high accuracy, We will develop the model with more features by increasing the number of rows to be an excellent model

