

(Houses Price Prediction Using Linear Regression)

By :

Shatha Hameed Alharbi

For :

The Online Data Science Bootcamps for SDAIA Academy

Dr._oumaima oueslati

October , 2021

Motivation :

most people know the benefit of owning a house, because buying a house is considered the most utilized and profitable investment.

an accurate prediction on the house price is important to prospective homeowners, developers, investors, appraisers and other real estate market participants, such as, mortgage lenders and insurers.

Price prediction using Linear Regression is the title of my project in this online data science Bootcamp for Sadaia Academy .

Design:

- The work concentrate on the commercial field.
- A huge amount of data is collected in the companies and commercial establishments , these data may contain full information about the price and customers.
- Discovering the relationship between this information it help the workers in this to understand the needing of the customers and allow them to predict the price.

In this project , I solved the question :

- How companies can predict the prices depending on the data which they have collected?

Data :

- For this purpose, I work using data about the price of houses.
- This file is collected from [www.Kaggle.com](https://www.kaggle.com).
- The data contains 21613 record.
- The data contains 21 columns, which are:
(id,date,price,bedrooms,bathrooms,sqft_living,sqft_lot,floors,waterfront,view,condition,grade,sqft_above,sqft_basement,yr_built,yr_renovated,zipcode,lat,long,sqft_living15,sqft_lot15).
- Eight of the columns will represent the features, which are:

(id,date,bedrooms,bathrooms,sqft_living,sqft_lot,floors,waterfront,view,condition,grade,sqft_above,sqft_basement,yr_built,yr_renovated,zipcode,lat,long,sqft_living15,sqft_lot15).

-The “price” column will represent the target.

Model Selection :

-I approached this problem with supervised machine learning (Linear Regression),
We know it's a regression task because we are being asked to predict a numerical outcome (Price).

- Score metric will be used to evaluate the model.

Tools :

-Jupyter and excel will be used in the work.

-Sklearn , panda , numpy , matplotlib and seaborn libraries will be used in the work.

Result :

The result for Train and Test data after applying the model :

```
from sklearn.linear_model import LinearRegression
LinearRegressionModel = LinearRegression(fit_intercept=True, normalize=False, copy_X=True, n_jobs=1)
LinearRegressionModel.fit(X_train, y_train)
print('Linear Regression Train Score is : ', LinearRegressionModel.score(X_train, y_train)*100,"%")
print('Linear Regression Test Score is : ', LinearRegressionModel.score(X_test, y_test)*100,"%")
y_pred = LinearRegressionModel.predict(X_test)
```

Linear Regression Train Score is : 70.26008959815655 %

Linear Regression Test Score is : 69.3340039939012 %