* Proposal on Housing Price Prediction

I assume that as the number of bedrooms and bathrooms of the houses increases, the price of the houses increases. The second assumption is that as the square feet of the houses is large, the price of the houses is high. The last assumption is that as the year when the houses are built is more recent, the price of the houses becomes more increased. In summary, the number of bedrooms and bathrooms and the area of the houses are proportional to the price of the houses. On the other hand, year that the houses were constructed is inversely proportional to the price of the houses. Therefore, I will investigate the correlation between the variables affecting prices of houses.

Background knowledge [1]

1. The number of bedrooms and bathrooms are the most important factors appraising the value of the houses. The second important factor affecting the prices of the houses is not the total area of the houses, but the area of livable space. The last important factor is the year when the houses were built. If the year of construction is not too old, the number of appliances needed for repair are not too much and buyer of the houses can save money.

My study on the residents living in king county Seattle Washington

Modifying the table

1. Original table about prices of houses was downloaded from <https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DA0101EN-SkillsNetwork/labs/FinalModule_Coursera/data/kc_house_data_NaN.csv>



1. At first, I dropped irrelevant columns not affecting the prices of houses, such as ‘id’, ‘date’, ‘view’, ‘condition’, ‘grade’, ‘zipcode’, ‘lat’, and ‘long’ as a data cleaning process.



A table with numbers and letters

Description automatically generated

1. Columns of price, bedrooms, bathrooms, sqft\_above, sqft\_basement, yr\_built, and yr\_renovated were selected from the original table. And the price column of the houses was moved to the right most column.



A table with numbers and letters

Description automatically generated

1. If some houses were renovated, built years were replaced by the renovated year, and ‘yr\_built’ and ‘yr\_renovated’ columns were deleted. The ‘year of construction’ of the houses is calculated by subtracting the year of construction from current year 2024.

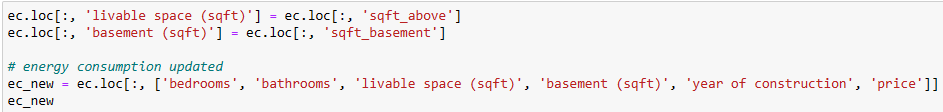
A white background with numbers and symbols

Description automatically generated

A screenshot of a computer

Description automatically generated

1. The livable space of the houses means the area apart from needless areas like basement, attics, and garages. This is the same as sqft\_above column in the table and the column is replaced with the new column named ‘livable space (sqft)’ in the table. A column named ‘sqft\_basement (sqft)’ was added to the table and the name of the column is changed to ‘basement (sqft)’. The objective of adding this column is to make sure if not livable area rarely affect the price of the houses.



A table with numbers and a number of rooms

Description automatically generated

1. The data were sorted in ascending order based on the prices of the houses.



A table with numbers and a number of rooms

Description automatically generated

1. The houses with age lower than or equal to 40 years were selected from the datasets. This is for the consideration of houses built recently.



A table with numbers and a number of rooms

Description automatically generated

* Analysis from the result tables

1. Data types of columns

A screenshot of a computer

Description automatically generated

A screenshot of a computer code

Description automatically generated

1. Draw a heatmap: show correlations of the variables affecting sales prices of the houses.

Before drawing the heatmap, correlation values can be calculated numerically.



A table with numbers and letters

Description automatically generated

Then, heatmap can be calculated in 2 dimensions.

A screenshot of a computer program

Description automatically generated

A chart of different shades of brown and green

Description automatically generated

Analysis from the heatmap

Although the area of basement was expected not to greatly affect the price of houses, the factor shows correlation higher than that of the number of bedrooms, which is presumably a dominant effect in this case. Differently from my assumption that recently constructed houses have high sales prices, this factor shows nearly zero effect on the prices of the houses. Other factors, such as the number of bedrooms and bathrooms and the area of livable and basement spaces show a tendency proportional to the prices of the houses.

Citation

1. The data was provided by Flatiron School for Data Science Immersive program phase final project.
2. Home Selling Tips, “8 critical factors that influence a home’s value” written by Gomez on June 4th 2022.