Link for the streamlit app: https://house-rocket-insight-project.herokuapp.com/

****

**1. Business Understanding/Problem.**

House Rocket is a company which main porpoise is obtains profit buying and selling houses localized at King County, state of Seattle, USA.

The object of this project is providing answer of the main questions made by the company’s CEO, which are:

* Which houses the company should buy, and for which price?
* Once the houses have been bought, what is the best moment to sell them, and for how much?

The answer for those questions is presented in two different methods:

* Reports recommending which houses to buy, for how much and when to sell them.
* A website created by the data analyst showing a table of the most interested houses to buy and other table with information about house prices to sell and when to them. Also a collection of hypotheses and a dashboard is presented in order to auxiliary the CEO decisions.

The tools used for this project are: Python 3.8, Pycharm, Jupyter Notebook, Streamlit and Heroku.

**2. Data Understanding.**

The data collected to put this project into operation is found at website: <https://www.kaggle.com/harlfoxem/housesalesprediction>.

Dataset list:

|  |  |
| --- | --- |
| **Attributes** | **Meaning** |
| id | Unique ID for each house sold |
| date | Date of the home sale |
| price | Price of each home sold |
| bedrooms | Number of bedrooms |
| bathrooms | Number of bathrooms,  where:  0.25 bathroom is a hallway shower or a freshen-up room with a single sink and mirror;  .5 accounts for a bathroom with a toilet but no shower;  .75 bath is a bathroom that contains one sink, one toilet and a shower or a bath;  and 1 is a full bath contains at least one sink, one toilet, a shower and a bath |
| sqft\_living | Square footage of the apartments interior living space |
| sqft\_lot | Square footage of the land space |
| floors | Number of floors. The .5 floor is a partial floor added to allow for more space. |
| waterfront | A dummy variable for whether the apartment was overlooking the waterfront or not, where 1 = the house has water view; 0 = the house does not have water view. |
| view | An index from 0 to 4 of how good the view of the property was |
| condition | An index from 1 to 5 on the condition of the apartment |
| grade | An index from 1 to 13, where 1-3 falls short of building construction and design, 7 has an average level of construction and design, and 11-13 have a high quality level of construction and design |
| sqft\_above | The square footage of the interior housing space that is above ground level |
| sqft\_basement | The square footage of the interior housing space that is below ground level |
| yr\_built | The year the house was initially built |
| yr\_renovated | The year of the house’s last renovation |
| zipcode | What zipcode area the house is in |
| lat | Latitude |
| long | Longitude |
| sqft\_livining15 | The square footage of interior housing living space for the nearest 15 neighbors |
| sqft\_lot15 | The square footage of the land lots of the nearest 15 neighbors |

**3. Business Assumptions.**

After carefully research, some assumptions are taken based on several information obtained at website <https://www.kaggle.com/harlfoxem/housesalesprediction/discussion>.

Those assumptions lead to identify possible outliers’ existent on dataset. Such as:

* Any house which contains no bathrooms or bedrooms is considered outlier, therefore it is excluded.
* Any house which the number of bedrooms is higher than 11 is considered outlier, therefore it is excluded.

Furthermore, some assumptions were made to identify the profit range by selling houses.

* According the website: <https://www.prnewswire.com/news-releases/average-us-home-seller-profits-hit-65-500-in-2019--another-new-high-300991828.html>, the minimal profit made by selling houses In US is 10%, the maximum is 45%. This profit range flouts due the houses characteristics, such as: location, size, number of bedrooms and others.

**4. Solution Strategy.**

As the company’s CEO wants answer for two different questions, the solution strategy is divided into 2 parts:

First part: Which houses to buy?

1º - Collect data from kaggle’s website.

2º - Group them by region (zipcode), due this attribute is extremely influential on the house’s price.

3º - Find the house’s price median by region.

4º - Recommend that the houses with prices inferior to the median value should be bought, and the condition is minimal 3.

5º - Filter those houses, that should be bought, by size, number of floors, number of bedrooms, and number of bathrooms, in order to identify the level of recommendation of each house.

Second part: When to sell the houses and for how much?

1º - After the company buys the houses, the data is grouped by region and seasons.

2º - Inside each region and seasons, it is calculated the median price.

3º - If the buy price is higher than median price plus season and recommendation to buy is regular, than the sell price will be equal the buy price plus 10 %.

If the buy price is higher than median price plus season and recommendation to buy is high, than the sell price will be equal the buy price plus 12.5 %.

If the buy price is higher than median price plus season and recommendation to buy is very high, than the sell price will be equal the buy price plus 15 %.

If the buy price is lower than median price plus season and recommendation to buy is regular, than the sell price will be equal the buy price plus 30 %.

If the buy price is lower than median price plus season and recommendation to buy is high, than the sell price will be equal the buy price plus 37.5 %.

If the buy price is lower than median price plus season and recommendation to buy is very high, than the sell price will be equal the buy price plus 45 %.

4º - It is specified the best moment to sell based on the profit by season.

5º - It is specified the best moment to sell based on the profit by season and recommendation to buy in general and individual houses.

6º - It is specified the total profit by buying and selling houses.

**5. Top 10 Data Insights.**

**Hypothesis 01:** Houses which has water view are 20% more expensive, in general.

**False:** Houses with water view are 212.57668803323867 percent more expensive.

**Hypothesis 02:** Houses that was built before 1955 are 50% cheaper, in general.

**False:** Houses that was built before 1955 are -0.7757205525248732 percent cheaper.

**Hypothesis 03:** Houses without basement are 40% bigger them house with basement, related to total area (sqft\_lot).

**False:** Houses without basement are 22.483151526642544 percent bigger them houses with basement.

## Hypothesis 04: The growth of house prices YoY (Year over Year) (May 2014 compared to May 2015) is 10%, in general.

**False:** The total houses price YoY (Year over Year) suffered a decrease of -62.79177358882806 percent.

**Hypothesis 05:** Houses with 3 bathrooms have a growth MoM (month over Month) of 15%.

**False:** The total houses price MoM (month over Month) suffered a decrease of -9.953899240174858 percent.

**Hypothesis 06:** Houses with number of bedrooms above 8 have a number of bathrooms 40% higher than houses with number of bedrooms between 5 and 8, and 94% higher than houses with number of bedrooms between 1 and, 4 on average.

**True:** Houses with number of bedrooms above 8 have a number of bathrooms 39.9514563106796 percent higher than houses with number of bedrooms between 5 and 8, and 94.48676155875182 higher than houses with number of bedrooms between 1 and 4.

**Hypothesis 07:** Houses with 7 bedrooms has the total area (sqft\_lot) bigger between 132 to 320 percent than houses with 8 to 11 bedrooms, on average.

**True:** Houses with 7 bedrooms has the total area (sqft\_lot) bigger between 132.29431644290653 and 320.17243208828523 percent than houses with 8 to 11 bedrooms.

**Hypothesis 08:** Renovated Houses have living rooms 12% bigger than houses not renovated, on average.

**True:** Renovated Houses have living rooms 12.132344286788795 percent bigger than houses not renovated, on average.

**6. Financial Results.**

 House Rocket Company would have a profit of almost 19 percent, which are more than $771 million, if applies this data analytics method.

**7. Conclusion.**

In conclusion, it is possible to identify that the application of data analytics project at dataset from House Rocket Company was very successful, providing a huge profit opportunity based on which houses to buy and when to sell.

**8. Next Steps.**

Other project that can be made with this dataset is the exploration data analyses, which identify the best’s attributes in order to apply machine learning algorithms, with the objective to predict the price of futures houses to buy.