KC Houses Business Insights

Author: Eliel dos Santos Paes
Jundiaí 2019

Table of Contents

Abstract

The aim of the project is to define the best transactions opportunities within the portfolio. To find what houses should be bought, two approach were taken. First, a general comparison between houses with same condition and same region was done. With this approach the estimated profit was on average from 15% up to 65%. The second approach takes the effects of the houses features to refine the analysis trying to find high profitable opportunities. Based in the houses features, 10 hypothesis were checked, and for the validated hypothesis the estimated profit on average was 25%. Was found the 91% of the houses are in condition 3 and 4. That shows the people preference, and trading with those houses will be easier. Houses built after 2010 are all in condition 3, but the price can be higher than houses built before, even in better condition, what can generate good trade opportunities.

1. Introduction

From the historical data of houses sold between may 2014 and may 2015 in King County Washington State, USA, a data set was built and provide the information for this work. By help decision of houses buying process can improve the profit. So, this project intent to help in that decision.

2. Problem statement

The aim of this project is to generate insights to help in decision of house acquisition and or trade based on statistic variation of prices and houses features.

3. Methodology

Was adopted a criterion to defines a trade opportunities as following: the trade opportunities are all the houses that have the price lower than the median price, attending the premise ahead:

Premises and assumptions:

1 - The price comparison will be made only for houses with same condition and in the same region (zip code).

2 - The outliers will be removed from the data frame considering the price column, and will not be analyzed.

3 – The profit was estimated considering the houses should be sold by the median price.

The trade opportunities were selected using two different approach. The first one, was to get the total data frame grouped the data by condition and region and evaluate the median price for each case. Then, the price of each house was compared with the median price, for a given condition and region , and selecting the trade opportunities. In the

second approach a set of hypothesis, based in the houses features, was checked trying to find specifics and high profitable opportunities. The set of hypothesis are in the table bellow.

Table 1: Set of hypothesis.

Number	Hypothesis description
H1	Houses with waterfront are on average 20% more expensive.
H2	Houses with year built lower than 1950 are on average 20% cheaper
H3	Houses with basement are on average 40% more expensive.
H4	The YoY price, for year built, increase on average 10%.
H5	House with more than one bathroom are, on average, 15% more expensive
H6	Houses near to water, but without waterfront, are on average 20% cheaper than houses with waterfront.
H7	Houses with only one floor are on average 20% more expensive. Due to people likes houses without ladders
H8	Houses price increase with the increase of the living room area.
H9	Houses with year built higher than 2010 are on average 30% more expensive.
H10	Houses price increase with the increase of the lot area.

For the hypothesis that weren't refuted the profit was estimated with the criterion presented above.

4. Results

Median price per condition

As there are five different house conditions, it is important to know how the prices behaves depends the houses conditions and the number of houses per condition. Lets visualize that.

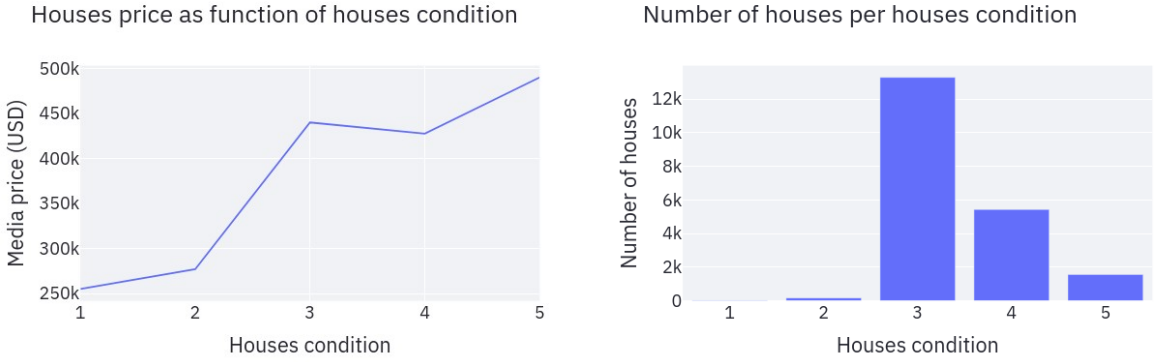


Figure 1: Median prices and number of houses per houses condition.

As wee can see, the house's price increases with the increase of the number that represents the house condition, indicating that the highest number the better house condition.

The most part of the houses are in the condition 3 and 4, being 65% of houses in condition 3 and 26% in condition 4.

4.1 House to be bought in the first approach.

Using the grouped data by condition and regions, lets define two new features for the houses:

Status: that defines if a house should be bought or not.

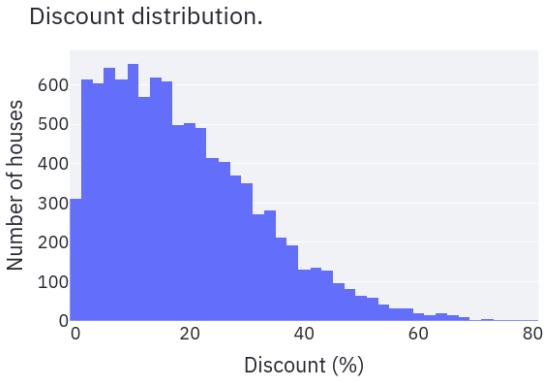
x% lower: that shows the discount, in other words, how much the house's price is lower the than median price for a given condition and region.

See below the data over view with the these two new features.

	price	condition	median_price	status	x% lower
0	221900	3	267000	Buy	16.8914
1	538000	3	416000	Not Buy	0
2	180000	3	449950	Buy	59.9956
3	604000	5	560000	Not Buy	0
4	510000	3	645000	Buy	20.9302
5	257500	3	275900	Buy	6.6691
6	291850	3	261000	Not Buy	0
7	229500	3	271875	Buy	15.5862
8	323000	3	350000	Buy	7.7143
9	662500	3	552000	Not Buy	0

4.1.1 Mapping the best opportunities.

Applying a filter in the data, taken only houses with status equal but, let's visualize the distribution of opportunities.



	x% lower
count	10113
mean	18.9126
std	13.5571
min	0.0001
25%	8.1046
50%	16.4286
75%	27.1845
max	79.5455

Figure 2: Discount distribution of houses that should be bought.

By the chart and the table we can conclude:

- There are 10113 houses with price lower than the median prices.
- For these 10113 houses on average the prices are about 19% lower than the median prices, for a given region and condition.
- 25% of the houses are with prices equal or lower than 27% of the median prices.

Based in the above conclusions, the data status will gain a new definition as follow:

- If the discount is 27% or higher, the status will be changed to Buy_SRP, that means Strongly Recommended Purchase.
- The set of houses out of the set of best opportunities, was called regular opportunities.

See below the data over view of this new definition

	price	condition	median_price	status	x% lower
0	221900	3	267000	Buy	16.8914
1	538000	3	416000	Not Buy	0
2	180000	3	449950	Buy_SRP	59.9956
3	604000	5	560000	Not Buy	0
4	510000	3	645000	Buy	20.9302
5	257500	3	275900	Buy	6.6691
6	291850	3	261000	Not Buy	0
7	229500	3	271875	Buy	15.5862
8	323000	3	350000	Buy	7.7143
9	662500	3	552000	Not Buy	0

Profit estimation

Estimating the profit, considering the houses should be sold by the median price.

Regular opportunities:

Minimal profit = 0%
Max profit = 37%
Averaged profit = 15%

Best opportunities:

Minimal profit = 37%
Max profit = 389%
Averaged profit = 65%

4.2 House to be bought in the second approach.

Now, let’s check the set of hypothesis. The aim of these hypothesis is to refine the analysis and try to find specific high profitable opportunities.

H1 - Houses with waterfront are on average 20% more expensive.

Value found = 97%

H1 confirmed

Actually, the prices of houses with waterfront are, in the average, about 97% higher than prices of houses without waterfront. Let's check price distribution for houses with waterfront.

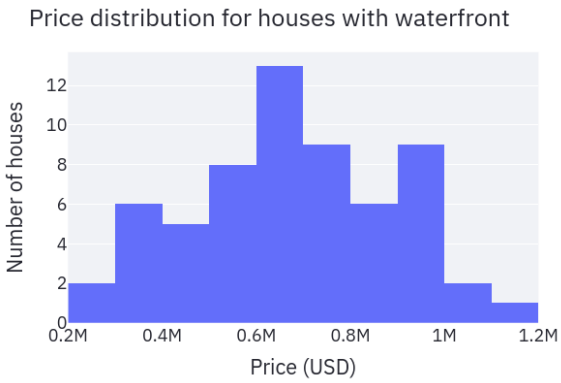


Figure 3: Distribution of houses with water front.

Houses with waterfront presents a quite homogeneous distribution with the averaged price of 655000. Considering only houses with waterfront and follows the premises defined in the methodology, let’s estimate the profit.

Profit estimation

Minimal profit = 2%
Max profit = 112%
Averaged profit = 25%

H2 - Houses with year built lower than 1950 are on average 20% cheaper.

Value found = 10%

H2 refuted

H3 - Houses with basement are on average 40% more expensive.

Value found = 17%

H3 refuted

H4 - The YoY price, for year built, increaseon average 10%.

Value found = 5%

H4 refuted

H5 - House that have more than one bathroom are on average 15% more expensive.

Value found = 44%

H5 Confirmed

Lets see how many houses there are in the portfolio for each amount of bathrooms.

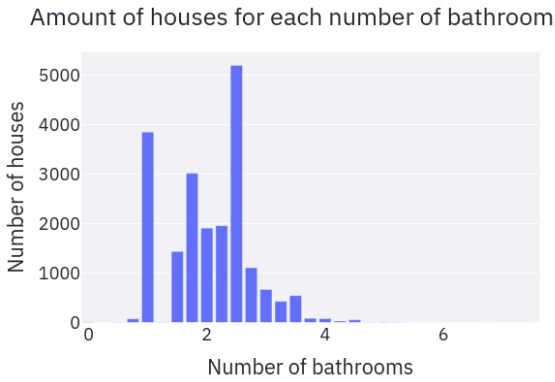


Figure 4: Amount of houses for each amount of bathrooms.

By the chart, we can see that the amount of houses where the number of bathrooms is higher or equal 1 and less or equal than 2.25 represents the major part of the houses, indicating people prefer these kind of houses. The recommendation is to buy this kind of houses, due to will be easier to trade.

Let’s estimate the profit comparing houses in the same region, with same condition and with same number of bathrooms.

Profit estimation

Minimal profit = 0%
Max profit = 300%
Averaged profit = 21%

H6 - Houses near to water, but without waterfront, are on average 20% cheaper than houses with waterfront.

There aren't houses that attends the hypothesis six.

H6 refuted

The idea of this hypothesis was to check if exists houses near to water, but without waterfront. That could be a good trade opportunity, once house with water front are on average 97% more expensive. So, would be interesting to buy a house in that condition, applies a renovation that convert the houses to have waterfront and sold after renovation.

To check the hypothesis six, o view in the map was done using green markers for houses with water front and blue markers for houses that attend the hypothesis six. As wee can see in the figure ahead, there aren't houses that attend the hypothesis six.

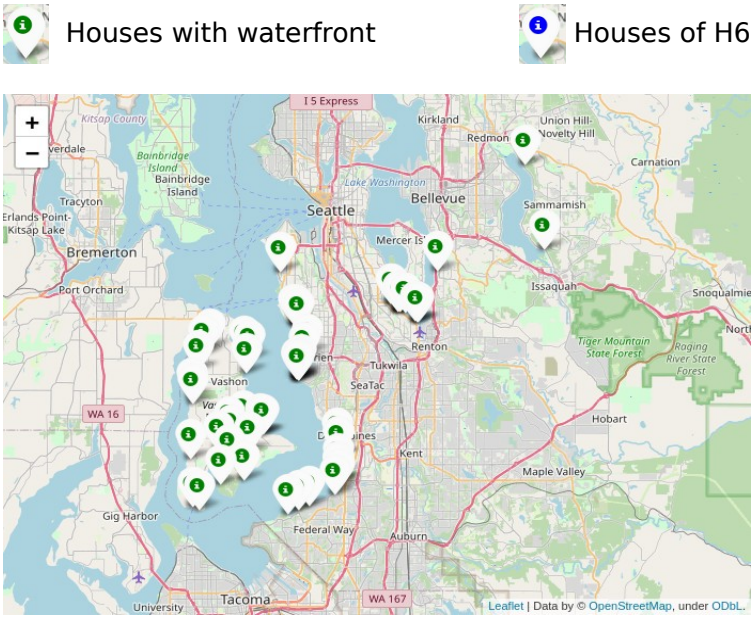


Figure 5: Map showing the location of houses with waterfront.

H7 - Houses with only one floor are on average 20% more expensive. Due to people likes houses without ladders.

Value found: -15 %

H7 refuted

H8 - Houses price increase with the increase of the living room area.

To check that hypothesis let's visualize the trend of the price as function of the living room area.

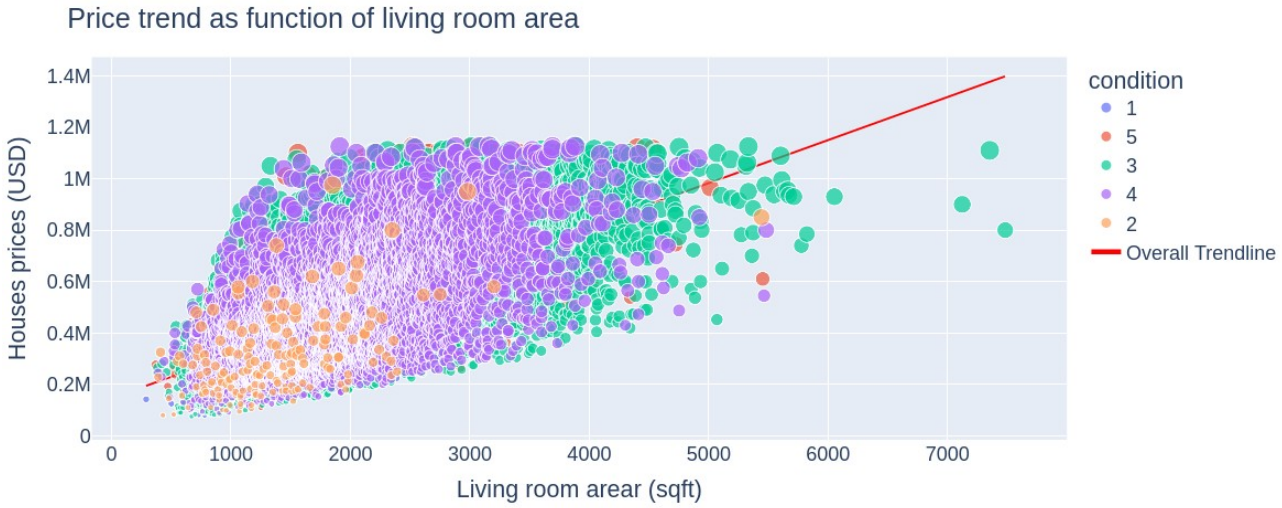


Figure 6: Price relationship with the living room area.

H8 confirmed.

In this case the profit was estimated with a different approach. It was considered that the house were bought when the prices were lower than the value of the trend line, considering same condition and region, and then sold by the price of the trend line.

Profit estimation

Minimal profit = 0%
Max profit = 258%
Averaged profit = 43%

H9 - Houses with year built higher than 2010 are on average 30% more expensive.

Value found: 28 %

H9 refuted

Although the hypothesis was refuted, there are opportunities to be considered, once the averaged prices, for houses built after 2010, are about 28% higher than averaged price for houses built before 2010. Lets visualize it.

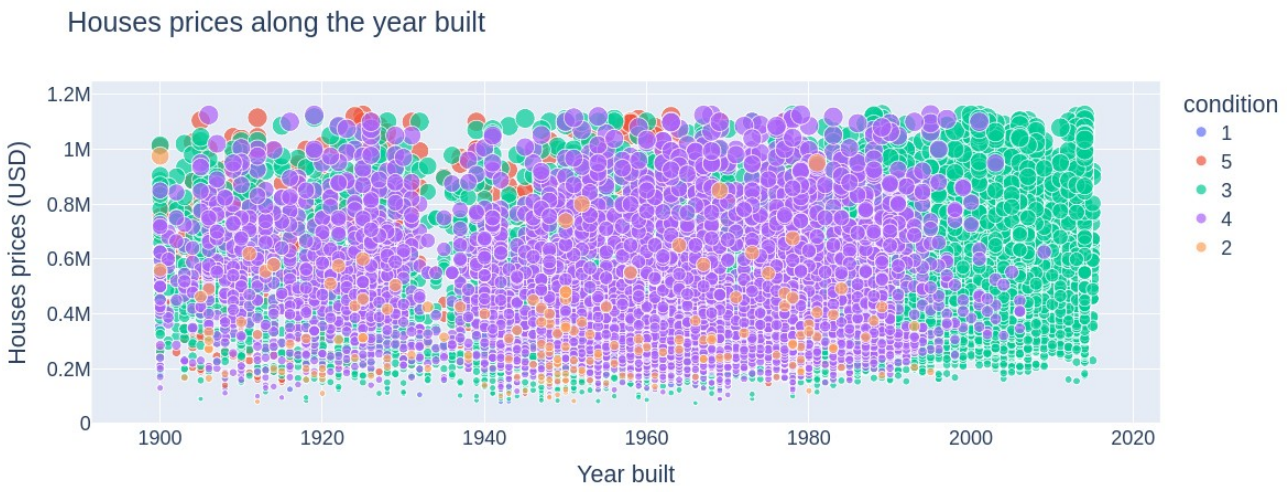


Figure 7: Houses price along the year built.

From the chart is possible see that the condition of house built after 2010 are in major 3, but the prices of these houses can be higher than the price of houses built before 2010, even with better condition. Let's estimate the profit buying only houses with year built 2010 and higher, making the general estimation.

Profit estimation

Minimal profit = 0%
Max profit = 170%
Averaged profit = 21%

H10 - Houses price increase with the increase of the lot area.

H10 refuted

See the chart below

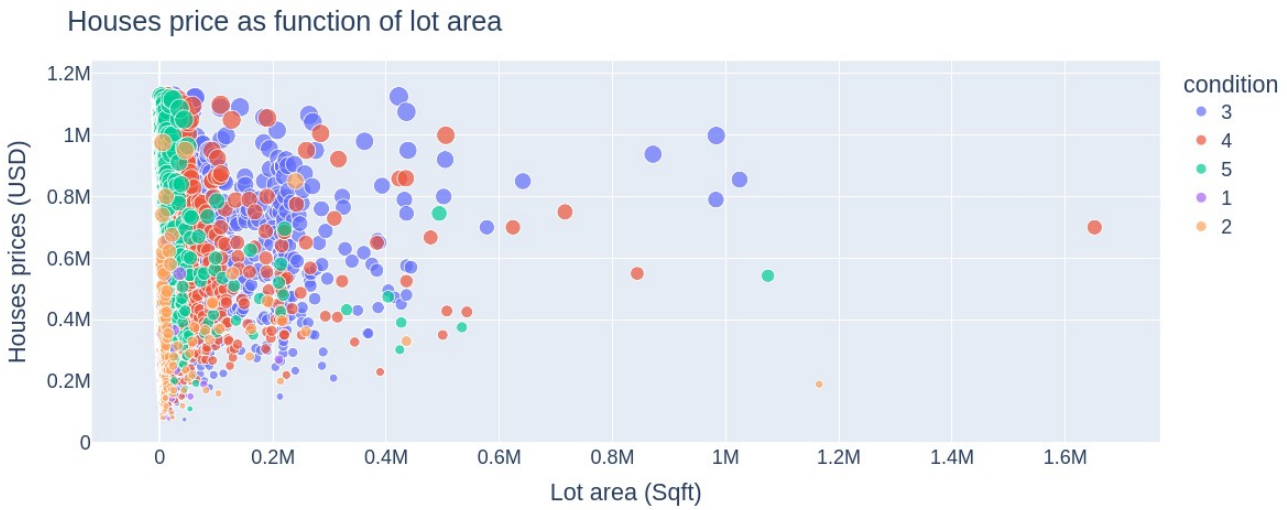


Figure 8: Relationship between houses prices and lot area.

4.2.1 Building data sets with houses that attends the confirmed hypothesis.

The table ahead shows the hypothesis that were confirmed, the number of houses that attend each hypothesis and the averaged estimated profit.

Table 2: Results of the validated hypothesis

Hypothesis	Number of houses	AVG estimated Profit
H1	24	25%
H5	8314	21%
H8	10927	42%
H9*	284	21%

* The H9 hypothesis was refuted but the data were analyzed due to potential profitable characteristics found.

For each case, and included the general approach shown in the beginning of the report, a data set was generated. The files were named as sample_”name”.csv. For instance, for houses that attend the hypothesis H1 the file was named as “sample_h1.csv”, and so on. The files are in the dataset folder, and was used in the dashboard built in the streamlit library.

5. Discussion

The most part of the houses are in the condition 3 and 4, being 65% of houses in condition 3 and 26% in condition 4. So, there are more trade opportunities with houses in conditions 3 and 4. The business should focus in that kind of houses, due to the highest probability of trade. The conditions 1 and 2 should be discarded, or very well analyzed in each case before buy a house. Considering condition 5, let’s say those houses are high level houses and there is a specific group of clients for them, and the House Rocket should focus a small part of the houses in those houses condition.

In the general approach were found 10113 trade opportunities. So, the number of houses found in the hypothesis, or probably in total, inside the set of houses taken for the first approach. So, looking for data from the hypothesis, we are looking for specific opportunities.

The hypothesis 1 (H1) presents the highest profitable approach. Although, the estimated profit found in the hypothesis 8 is approximately the average of the estimated profit of the general approach and probably don't represent the effect of the living room area in the price.

In terms of number of bathrooms, the must part of the houses has 1 or more and 2.5 or less bathrooms, indication the preference of people for that feature.

All data from the first and second approach are in the files samples in the dataset folder as csv files, and dashboard of the house can be visualized in the map using the streamlit application in the web site: https://hr-analytics-dashboard.herokuapp.com/

6. Conclusion

The companies should focus the business in houses with condition 3 and 4, that represents 91% of total portfolio.

Houses with condition 1 and 2 should be discarded, or very well analyzed before be bought.

The major part of the houses has at lest one bathroom. The companies should focus the business in houses the have at least one bathroom to increase the probability of trade.

Considering the definition of profit estimation, houses with water front presents the highest estimated profit, being on average 25%.

Houses built after 2010, are in condition 3, but the price can be higher than houses built before, even in better condition. Buying houses built after 2010, can create good and profitable opportunities.

7. Next Steps

To compare the approach of median price with the averaged price for the profit estimation.

Create prediction models for help in the decision of house to be bought.

Create a prediction models for help to find the best moment to buy the houses.