KING COUNTY HOUSING MARKET

OPENING A REAL ESTATE AGENCY

An analysis of the market and recommendations for an agency firm entering the Seattle/King County Area.

Business Case

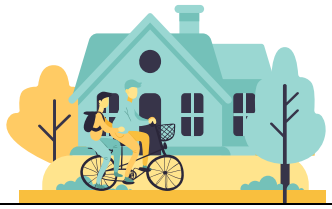
I was hired by **My Home Real Estate Agency firm** that considers entering the King County/Downtown Seattle market.

I'm working for a Real Estate Agency firm that considers entering the King County/Downtown Seattle market. However, since they don't know the area very well, the Agency hired me to study the market.

THE BUSINESS MODEL

- Agency is formed by **Real Estate Agents**
- They earn a **Commission (when sold Above Minimum Price)**
- **Revenue comes from Commission**

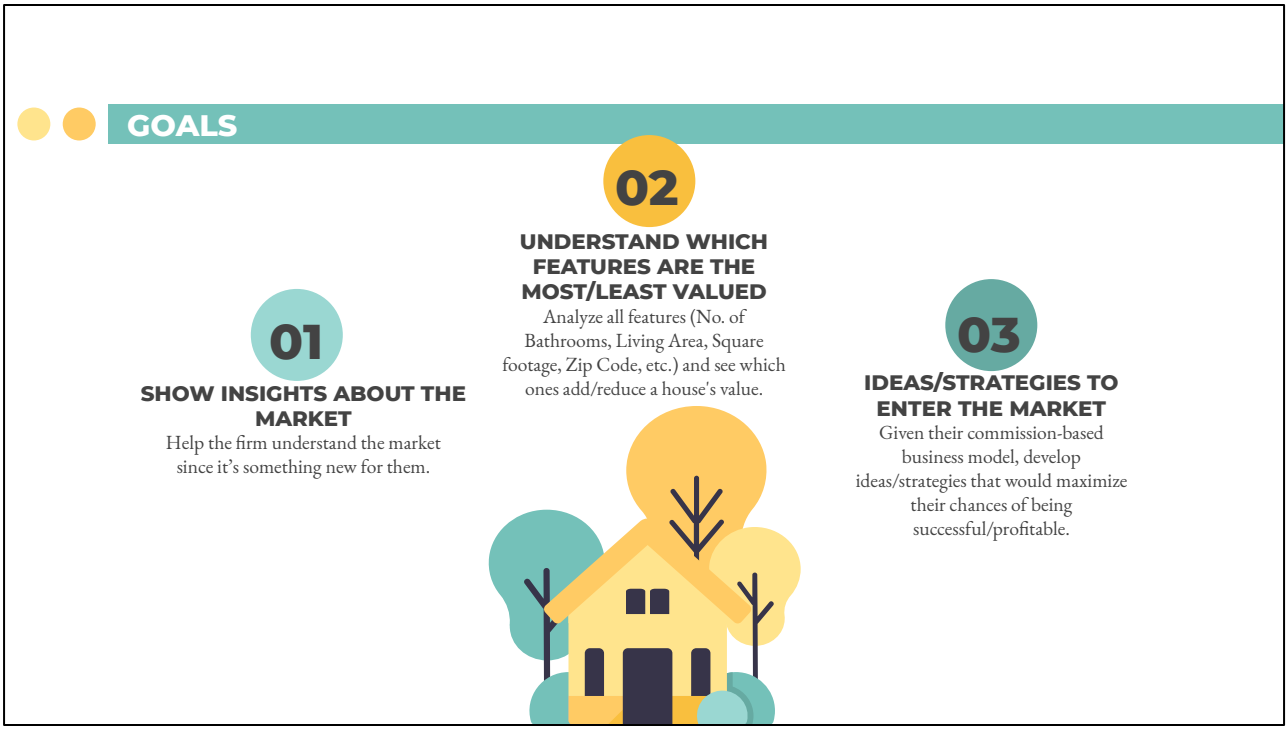
HOUSE SOLD AT	COMMISSION	AGENT KEEPS
At/Below Minimum Price	5% of Sale Price	Zero
Above Minimum	7% of Sale Price	40% of Commission




The Business Model is the following:

- The Agency hires Real Estate Agents to sell houses and pay them a Commission (see table).
- All revenue comes from these commissions generated by Agents. They don't know the No. of Agents they should hire, nor the Houses they should focus on selling.

No other information was given, and I made some assumptions during the analysis. You can find all the Assumptions in the Appendix of this presentation.



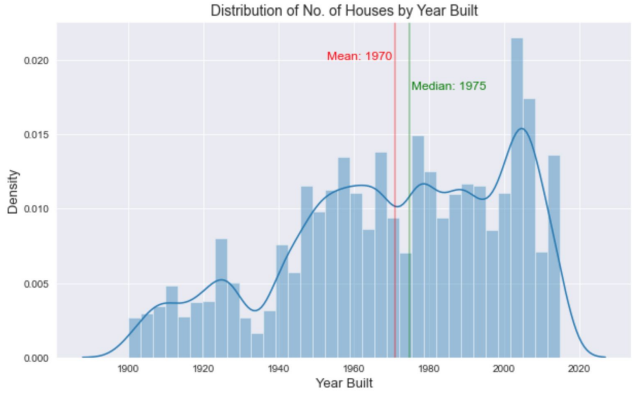
My goals are the following:

- Show some insights about the market:
  - Help the Agency understand the market.
  - Check whether it's a Healthy Market.
  - Identify areas with the most valuable houses.
- Analyze which features are most/least valued:
  - On a more granular level, help the Agency understand which House Features add/reduce value.
  - With that information, the Agency can focus on certain houses and help agents negotiate with clients since they know what most clients care about.
- Ideas/Strategies to enter the market:
  - Given their business model and all assumptions, come up with ideas and strategies. For example, identify the most profitable Zip Codes given their costs or the No. of Agents they should hire to increase Return on Investment

Understanding the Market

UNDERSTANDING THE MARKET

- **Growing Number of new houses**, with most of them built in 1970-1975.
- **The most expensive Zip Codes** are near Downtown Seattle
  - However, there's some variability within Zip Codes.

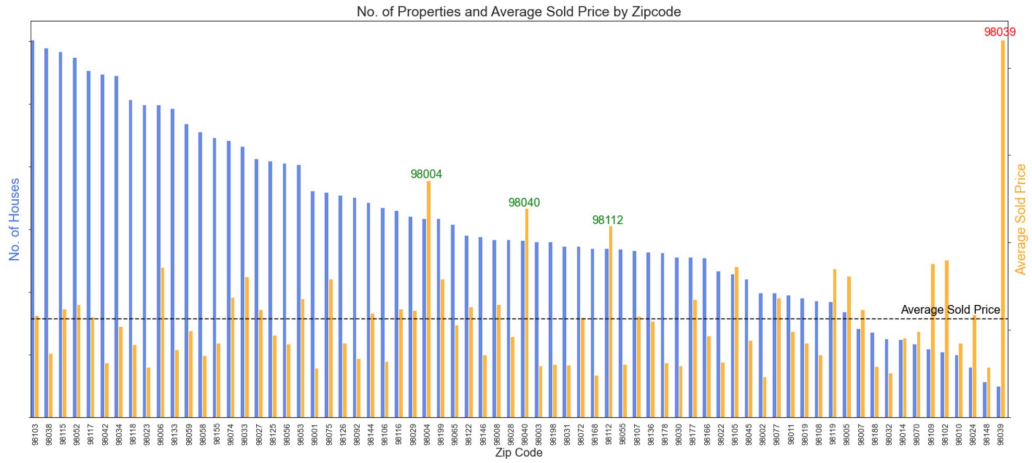


(Chart on the left) Growing No. of New Houses: The chart shows the Density Distribution of houses by Year Built. It's good to see an upward trend on the Year Built of Sold Houses. This information implies a demand for new houses because nobody would build homes in an area with no buyers.

(Chart on the right) Houses near Downtown Seattle are more expensive: The chart is a scatter plot of Sold Price by Distance from Downtown Seattle (in km). It's expected that places near economic centers are more expensive, and it's no different for Seattle. However, that doesn't mean that Houses in Zip Codes near Downtown are all more expensive. There's variability within Zip Codes.

UNDERSTANDING THE MARKET

- Some **Zip Codes have more sold houses than others.**
- The **Average Sold Price is higher in certain Zip Codes.**



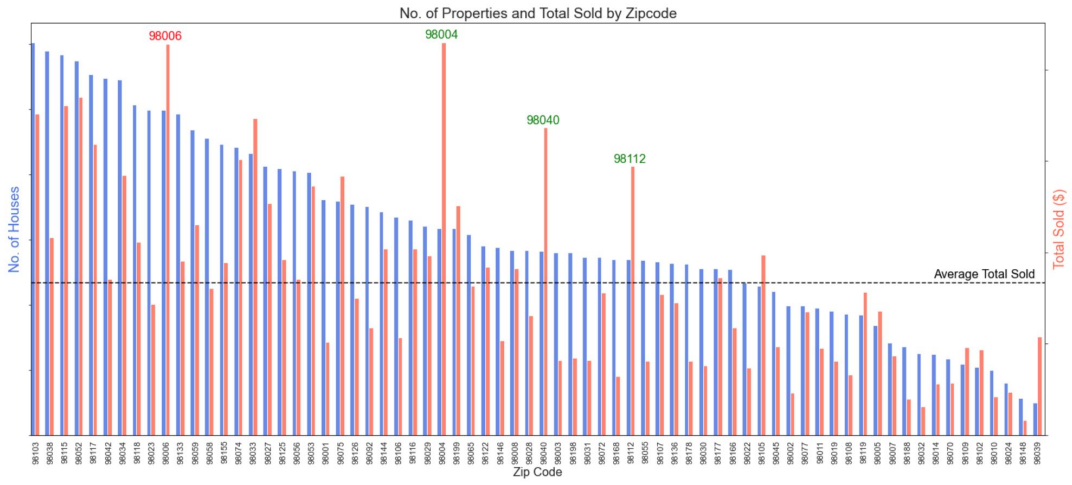
Blue Bars: Indicate the No. of Houses on the dataset. Since the dataset includes only Sold Houses, it shows the Zip Codes with the most No. of Sold Houses.
Orange Bars: Indicate the Average Sold Price by Zip Code. It's just the Sum of Sold Price divided by the No. of Houses in a Zip Code. It's a good way to see which Zip Codes sell for more.
Dashed-Line: Indicates the Average Sold Price including all Zip Codes. Sum of Average Sold Price by Zip Code divided by the No. of Zip Codes.

In the chart, four Zip Codes stand out: 98004, 98040, 98112 and 98039.

- 98039 is an outlier, the No. of Houses is very low, so the average might not be very representative, more data is needed to confirm the high value.
- However, the other three Zip Codes show very good numbers. These might be good opportunities for the Agency.

UNDERSTANDING THE MARKET

- It's possible to spot a few of them with the **highest sum of Sold Price.**



Blue Bars: Indicate the No. of Houses on the dataset. Since the dataset includes only Sold Houses, it shows the Zip Codes with the most No. of Sold Houses.
Red Bars: Indicate the Total Sold on the Zip Code. It's the Sum of Sold Price by Zip Code.
Dashed-Line: Indicates the Average Total Sold by Zip Code. It's the Sum of Sold Price by Zip Code divided by the Total No. of Zip Codes.

In the chart, four Zip Codes stand out: 98006, 98004, 98040, 98112:

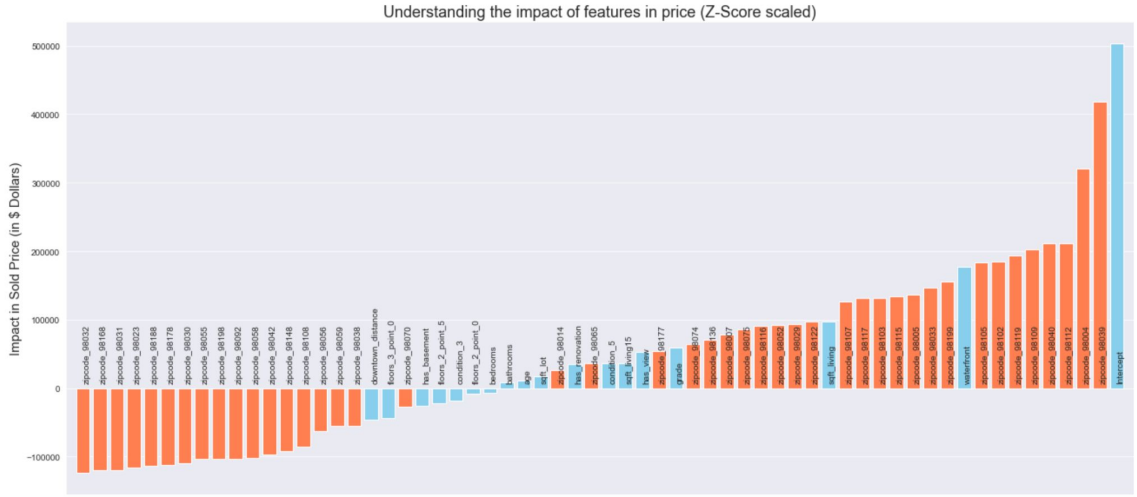
- It's interesting to see that the outlier from the previous chart is gone here, which supports my idea that the previous number was not representative.
- It's great to see the same three Zip Codes from the previous chart here, another good indication.
- There's a new Zip Code here, 98006, but it's interesting to see that it almost has the same sum as 98004, but with more houses. From a business standpoint, that's not good since it means more work for the same amount of money. Most likely, the Agency would need to hire more Agents to manage 98006 than 98004.

Most/Least Valuable House Features

In this section, after creating a Linear Regression Model fitted to predict the Sold Price given House Features, I'm analyzing the coefficients/weights of each feature. With that information, I can compare the importance of each.

MOST/LEAST VALUABLE FEATURES - ALL FEATURES

Zip Code is the most relevant feature (**Zip Code Bars are Red**).

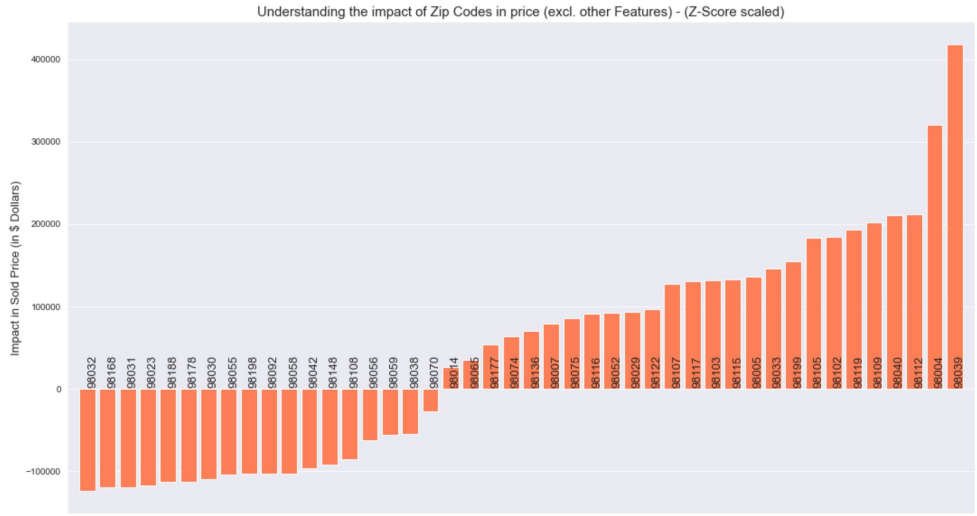


Red Bars: Indicate Zip Codes. Some have a positive impact in price, while others have a negative impact.
Blue Bars: Indicate House Features and the Intercept.

In this chart, it's easy to see that Zip Code (red bars) is the most important feature. In a few scenarios other features will have more impact.

MOST/LEAST VALUABLE FEATURES - ZIP CODES

Most Valuable: **98039 and 98004** Least Valuable: **98032 and 98168**

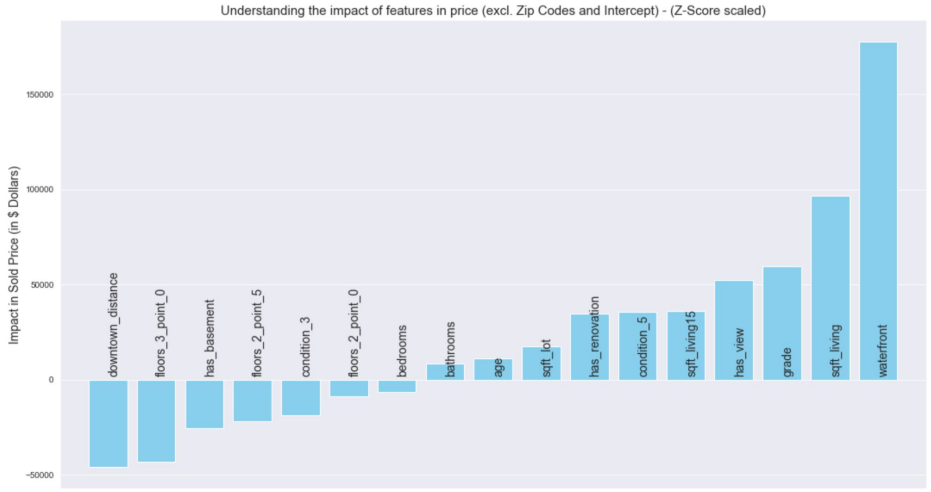


After excluding other features and keeping only Zip Codes, it's possible to compare them.
As stated on the slide, 98039 and 98004 are the most valuable Zip Codes, while 98032 and 98168 are on the other end of the spectrum.

One important caveat is that Zip Codes have their particularities. For example, some of them might have higher Square Footage than others, and to compensate for that, the model chooses to reduce the price for that specific Zip Code. That would bring some misunderstanding to the analysis. The correct analysis would be to look at all variables to have a good sense of the features. Nonetheless, looking at the coefficients is insightful, but it's not the ultimate way to analyze them, it's a simplification.

MOST/LEAST VALUABLE FEATURES - EXCL. ZIP CODES

Most Valuable: **Waterfront and Square Footage** Least Valuable: **High Downtown Distance and 3 Floors**



In this chart, after excluding Zip Codes and the Intercept, it's possible to see the most/least important features.

Most Valuable:

- Houses with a Waterfront are more valuable. The presence of a waterfront adds $177,850 to the sold price.
- Square Footage is also important: Bigger houses are more expensive, which is expected. But, more specifically, the price goes up $32 by Square Footage added.

Least Valuable:

- High distance from Downtown Seattle, the price goes down by $1,520 for every km.
- Houses with Three Floors have their price reduced by $43,279.

MOST/LEAST VALUABLE FEATURES - MARGINAL PRICE CHANGE

Continuous Features

feature_(1)	marginal_price_change_regular_scale_(2)
age	154.0
bathrooms	2945.0
bedrooms	-1456.0
downtown_distance	-1520.0
grade	6706.0
sqft_living	32.0
sqft_living15	13.0

*"Adding one unit of **(1)** changes price by **(2)** dollars"*

Binary Features

feature_(1)	marginal_price_change_(2)
condition_3	-18671.0
condition_5	35816.0
floors_2_point_0	-8505.0
floors_2_point_5	-21654.0
floors_3_point_0	-43279.0
has_basement	-25671.0
has_renovation	34971.0
has_view	52219.0
waterfront	177850.0

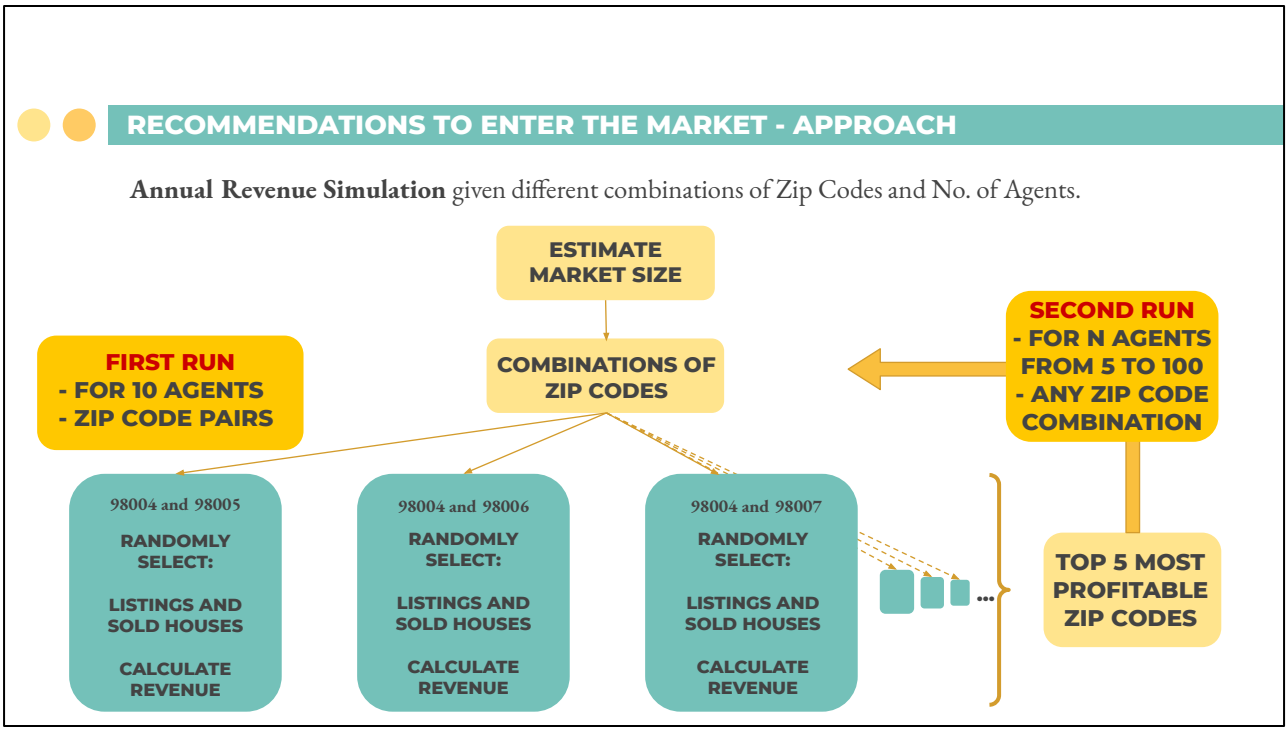
*"If house has **(1)** , price changes by **(2)** dollars"*

In this slide, feature coefficients are shown in their own scale for better comprehension. Nonetheless, it's better to look at Z-Scaled numbers since they're more representative of the dataset.

For example, it's cool to see that one Kilometer added to the distance from downtown Seattle reduces the price by 1520. However, we don't know the distribution of that distance. Are most houses located near Downtown or not? However, stakeholders like to look at these numbers since they are easier to understand than Z-Scores.

Recommendations to Enter the Market

In this part of the presentation, I'm going through my recommendations/ideas after running simulations given the business-case.



The approach is straightforward. I'm modeling the business by taking all Revenue/Cost Inputs and coming up with Assumptions.

I run simulations on various combinations of No. of Agents and Zip Codes where the Agency will be present. Based on the Average Success Rate, and by randomly picking houses and summing that 10% of them will be sold above the Minimum Price, I can estimate Revenue. Moreover, since I have the No. of Agents, their salaries, and commissions, I can also estimate Cost, which gets me to Net Income.

Here are the steps:

- To keep simulations a bit more realistic, I'm taking the Average Success Rate and assuming some deviation will smooth out as the No. of Agents go up (I assume there's a normal distribution with a mean of 30% and a standard deviation of 30%) and removing outliers from the dataset.
- Since the dataset includes only Sold Houses, and I'm trying to estimate Revenue, it's crucial to consider that Agents will not sell some houses. To achieve that, I'm estimating the Housing Market Size by extrapolating the No. of Sold Houses and the Success Rate from Agents. For example, if an Agent has a Success Rate of 50%, and he/she sells 10 houses, then there were 20 listings managed by the Agent during the year.
- With all of that in hand, I start by running simulations using all possible pairs of Zip Codes for 10 agents only. Since I know the No. of Houses each Agent will sell and the success rate, I can tell the No. of Listings by Agent. Therefore, I randomly pick which houses Agents will manage and sell. Finally, with that

- information, I can estimate the Net Income.
- Using the results, I identify the Top 5 Zip Codes by looking at the Estimated Net Income for each pair and summing them up by Zip Code.
    - For example, if A + B generated 10 dollars, A + C generated 15 dollars, and B + C generated 5 dollars.
    - My conclusion is that A is the best because if I isolate each Zip Code and sum revenue, I get the following results:
        - **A = 10 + 15 = 25**
        - B = 10 + 5 = 15
        - C = 15 + 5 = 20

  It's like playing basketball with different combinations of players and identifying that Player A is present every time the team scores more points.
- With the Top 5 Zip Codes in hand, I run the same simulation using any possible combinations of these 5 Zip Codes (including operating in all of them) + different No. of Agents from 5 to 100.

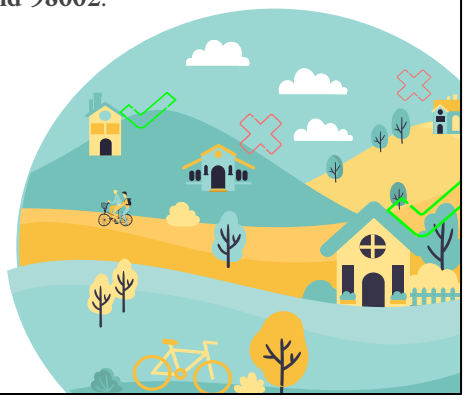
RECOMMENDATIONS TO ENTER THE MARKET - RESULTS

FIRST RUN (10 Agents and Zip Code Pairs):

- Most Profitable Combination: **98004 and 98039**.
- Most Profitable Zip Codes: **98004, 98040, 98112, 98039 and 98006**.
- Least Profitable Zip Codes: **98023, 98168, 98042, 98001 and 98002**.

SECOND RUN (Different No. of Agents on Top 5 Zip Codes):

- **Best Combination: 25 Agents Selling Houses only in 98004** will bring the highest Return on Investment (50%). It requires ~$4MM in investment, though.
- Low chance of good Return on Investment = **Rethink Business Model.**



The slide shows the results of the simulations.

One of the conclusions was that maybe the Business Model is not good. However, that could be a consequence of all of my assumptions. In an ideal world, I'd build the model with stakeholders who better understand the business, with all its inputs and constraints.

Conclusions

01 HEALTHY MARKET AND INSIGHTS

. Growing No. of New Houses.
. Houses near Downtown Seattle and Washington Lake have higher prices.

02 BEST/WORST HOUSE FEATURES

. Zip Code is the most impactful feature.
. Whether house has a Waterfront and Square Footage add more value.
. High distance from Downtown Seattle and Three Floors reduce value.

03 RECOMMENDATIONS

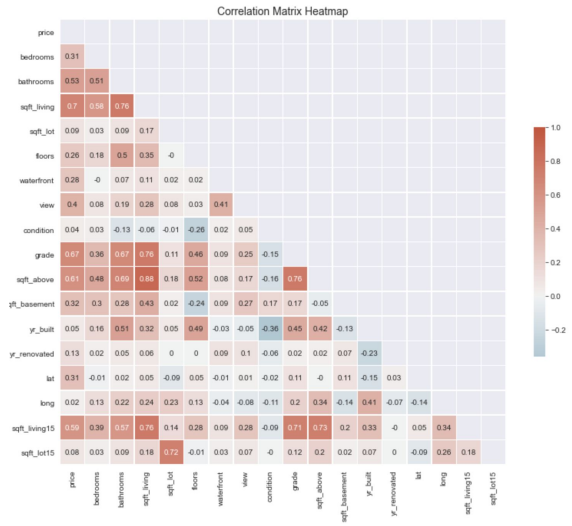
. 25 Agents selling 98004 Houses is the most profitable option.
. Most Profitable: 98004, 98040, 98112, 98039 and 98006.
. Focus on Most Valuable Features.
. Rethink Business Model.

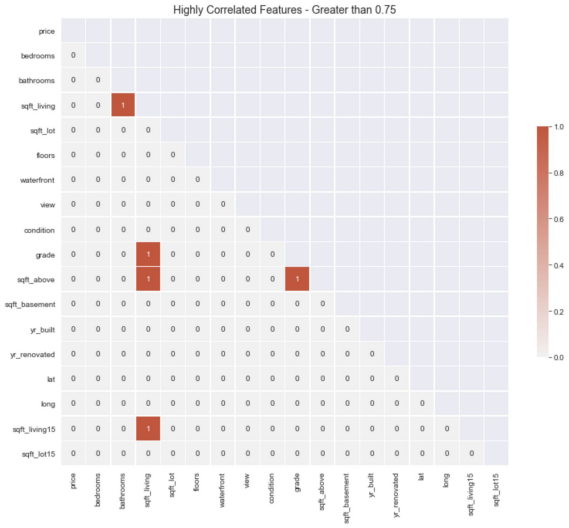
THANKS

Reno Neto - renoneto@gmail.com
GitHub:
https://github.com/renoneto

Appendix

APPENDIX: CORRELATION MATRIX - SOME HIGHLY CORRELATED FEATURES

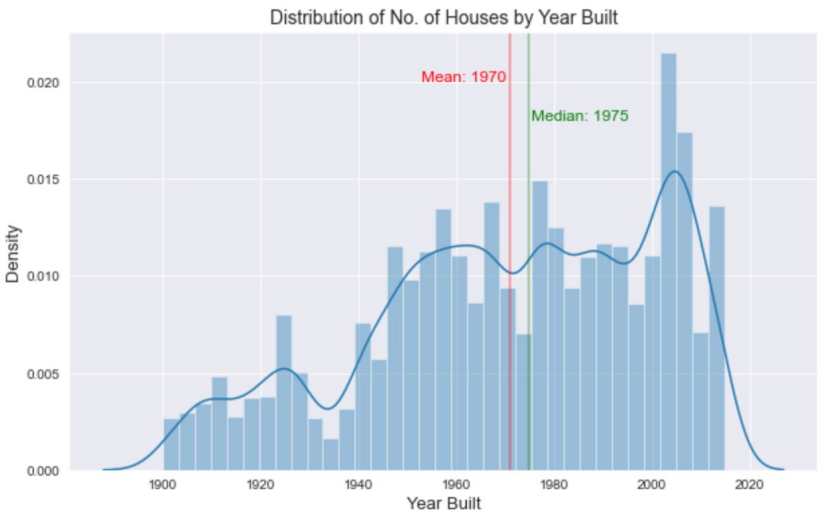




APPENDIX: DISTRIBUTION OF HOUSES BY YEAR BUILT

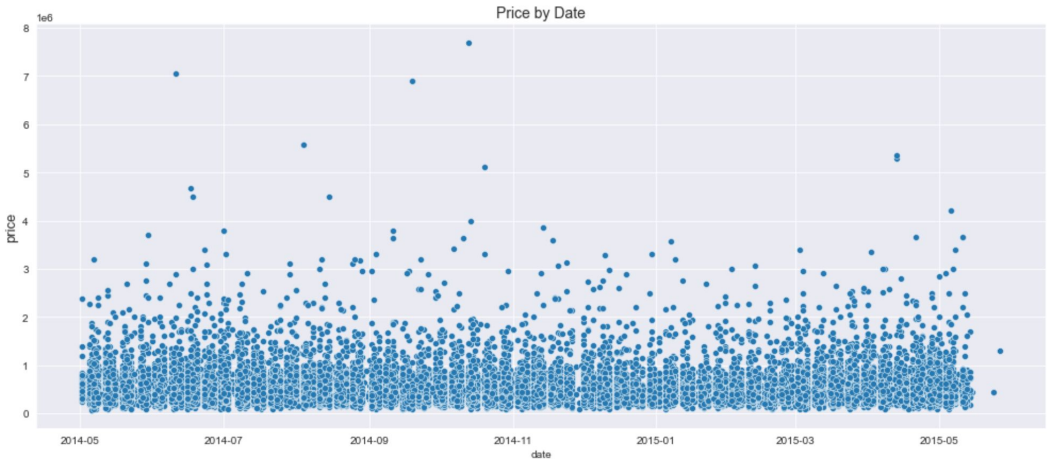
Even though the dataset is not showing all houses in the market. Based on the Year Built of sold houses, it's possible to see the *No. of Houses increasing as the Year Built goes up*. Possibly indicating th it's a **Healthy Market**, with new houses built over the years and buyers buying them, indicating that **there's a demand for new houses**.

Nonetheless, it's interesting to see that the Average is 1970, which is somewhat related to companies that moved to Seattle in the 70s.



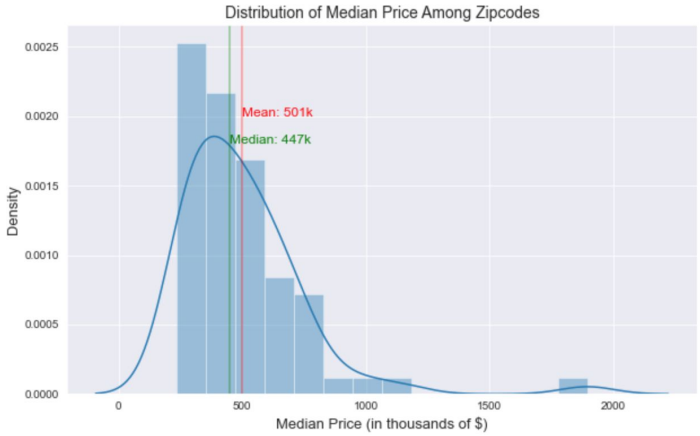
APPENDIX: SALE PRICE OVER TIME

It's **hard to see an increase in Sale Price over time**, which can be a *concerning point*. However, it would be better to have more data. One year, might not be enough time to show changes in price.



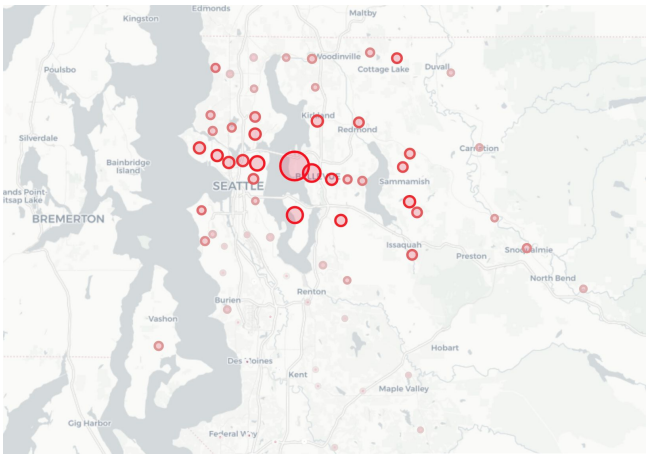
APPENDIX: MEDIAN PRICE ACROSS ZIP CODES

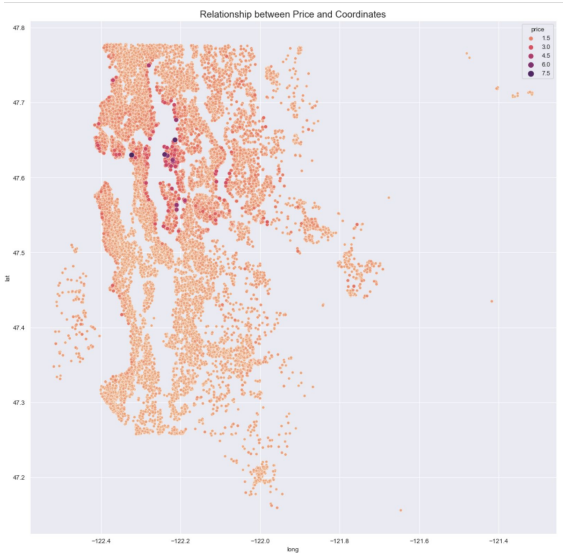
After calculating the Median Price across Zip Codes. I created the following distribution chart showing that the Median of Median Prices is $447k.



APPENDIX: AVERAGE PRICE BY ZIP CODE IN MAP

We can see that regions near Downtown Seattle and Lake Washington's shoreline are more expensive on both maps.





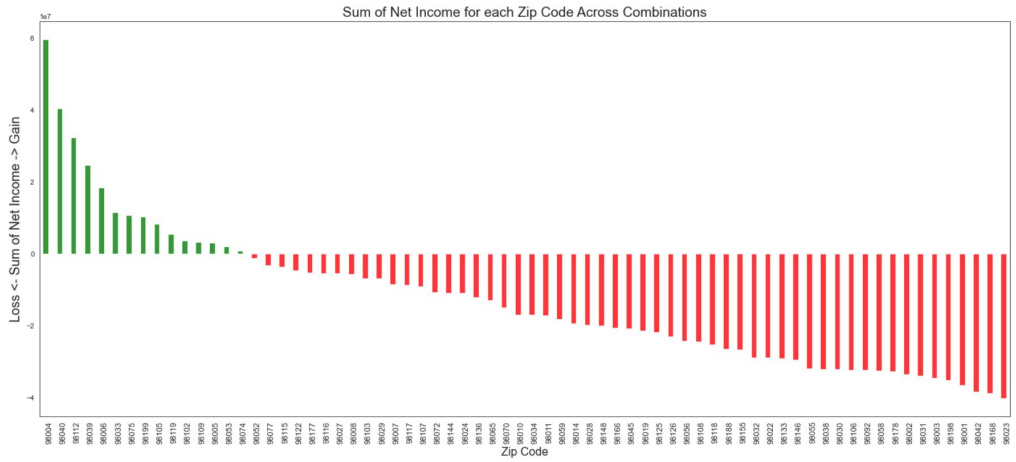
APPENDIX: RECOMMENDATIONS TO ENTER THE MARKET

Assumptions:

- **Estimated Market Size** based on Sold Houses and Avg. Success Rate by Agent.
  - No. of Houses in Dataset / Success Rate = **71,990 Houses in Market**
  - Sold Houses are representative of market.
  - All houses were sold by Agents.
- **Agent Information** in Seattle found online.
  - Annual Cost: $152,565 Salary for Intermediate Agents in Seattle.
  - Sell of Average 10 houses per year.
- Agency **cannot be present** in all Zip Codes.
  - Otherwise, would have to test all possible combinations (with 1, 2, 3, n Zip Codes).
- **Success Rate on Selling Above Minimum Price** is 10%.
- **Average Success Rate** is 30% - Std Dev of 30% as well.

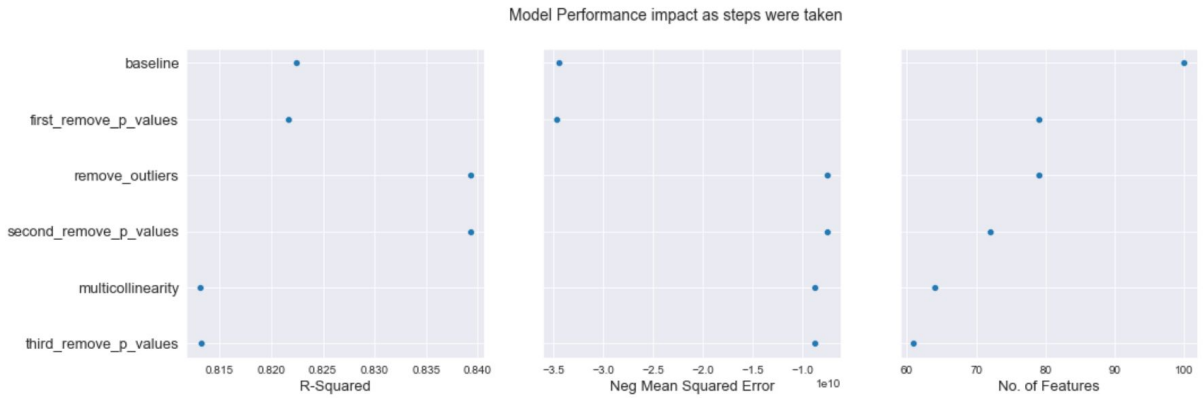
APPENDIX: NET INCOME BY ZIP CODE

After calculating the Net Income from different Zip Code pairs sold by only 10 Agents, I decided to sum all Net Incomes for each Zip Code. The idea was to find patterns. For example, when Zip Code A is present, Net Income tend to be Higher/Lower.



APPENDIX: PROGRESSION OF LINEAR REGRESSION MODEL METRICS

Below, you can see the progress of the Linear Regression Model. Starting with Baseline, where a model was fitted to the data. Then, steps were taken to improve the model (Removing High P-Value features, Outliers and deal with Multicollinearity).



APPENDIX: FINAL MODEL Q-Q PLOT AND SCATTER PLOT OF RESIDUALS

I could have taken more steps to increase the accuracy of the model. However, it would involve losing some of the interpretability of the model. I even tried doing log of price, which increased performance, but then it became hard to interpret coefficients. More could have been done with outliers, but it would involve removing a lot of data.