



# KC REAL ESTATE AGENCY

DATA DRIVEN REAL ESTATE



# INTRODUCTION

Real estate simply put, refers to the land, plus any permanent man-made additions, such as houses and other buildings. Any additions or changes to the land that affects the property's value are called an improvement. Once land is improved, the total capital and labor used to build the improvement represent a sizable fixed investment.

Real estate market is a trillion dollar industry globally that is valued at USD 3.69 trillion as of 2021 and is expected to have an annual growth rate of 5.2% till 2030. This is a very healthy growth rate that ensures profitable returns to potential investors in the field.

America has a population of approximately 340 million people. All these people require shelter as it is a basic necessity, which means these are all home-owners or potential home owners.

# BUSINESS PROBLEM

As we can see the industry is very lucrative but also has the ability to cause great losses when house pricing gets out of control. This means that potential investors have to make smart and informed decisions so as to safeguard their money, they want to get maximum value, no matter how much is spent. After all who likes to be known to make poor investment decisions and what better way to ensure this than with data-driven decision making.

With this in mind we want to help our stakeholders KC Real Estate Agency to make the right decisions with their money as they venture to bring more, better and nicer homes to the 340 million Americans.

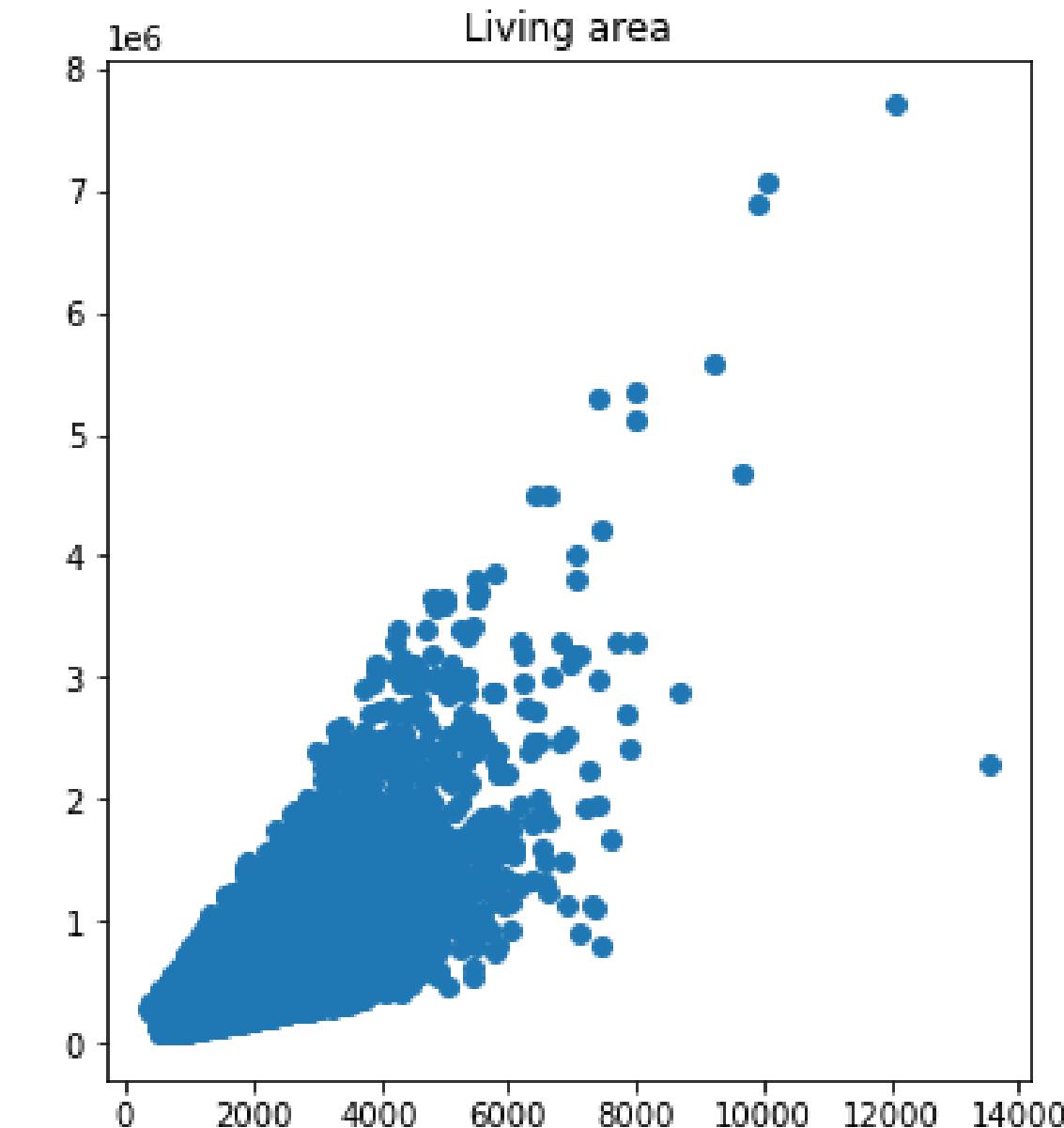
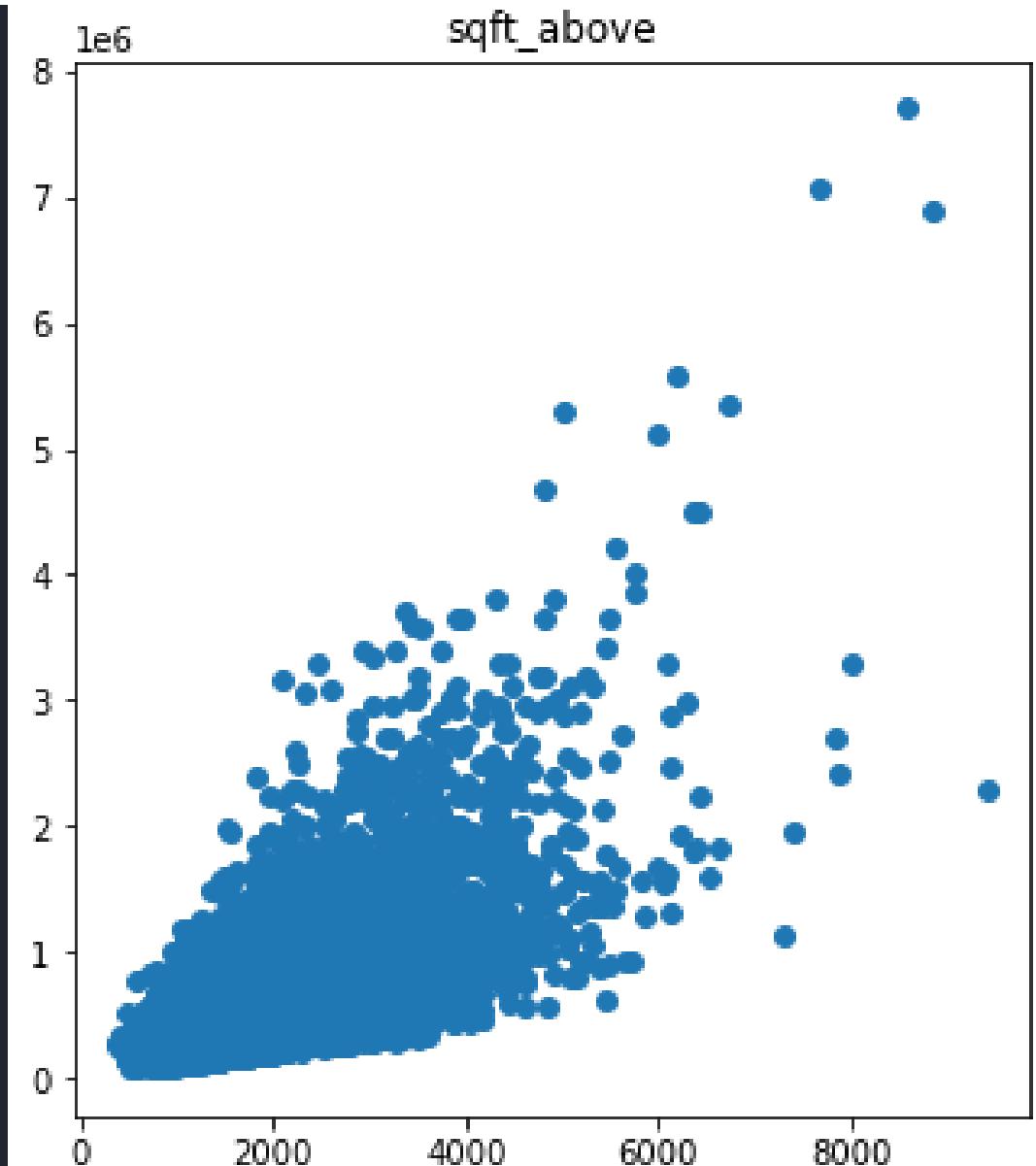
# OBJECTIVES

- 1. To determine which variables affect house pricing, more specifically those that increase the value of a house.**
- 2. Factors that affect grade of a house. We want to invest in only high grade houses despite the price.**
- 3. To identify similar areas of spatial clustering based on the price.**

# OBJECTIVE 1

We see that sqft\_above and sqft\_living have a relatively strong positive correlation. Thus we plot them individually and check their linearity in relation to house prices.

The 2 seem very similar with `sqft\_living` with just a bit less variance of the two



# MODEL INTERPRATATION

- Our model seems to be statistically significant overall due to the fact that our F-statistic probability is less than our alpha value(0.05).
- The model explains 49.7% of the variance in price. This shows our model isn't that much of an improvement in terms of variance of the data when compared to our earlier model with just sqft\_living.
- Our y-intercept and all our slopes(coefficients) are statistically significant(the p-values are less than alpha).
- Our intercept is -48,590 meaning that a home with 0 square feet for living and above ground it would cost -USD48,590
- Our coefficient for sqft\_above is -18.31. This means that for an additional square foot, we expect the price to go up by -18.31
- Our coefficient for sqft\_living is 299. This means that for an additional square foot, we expect the price to go up by USD299

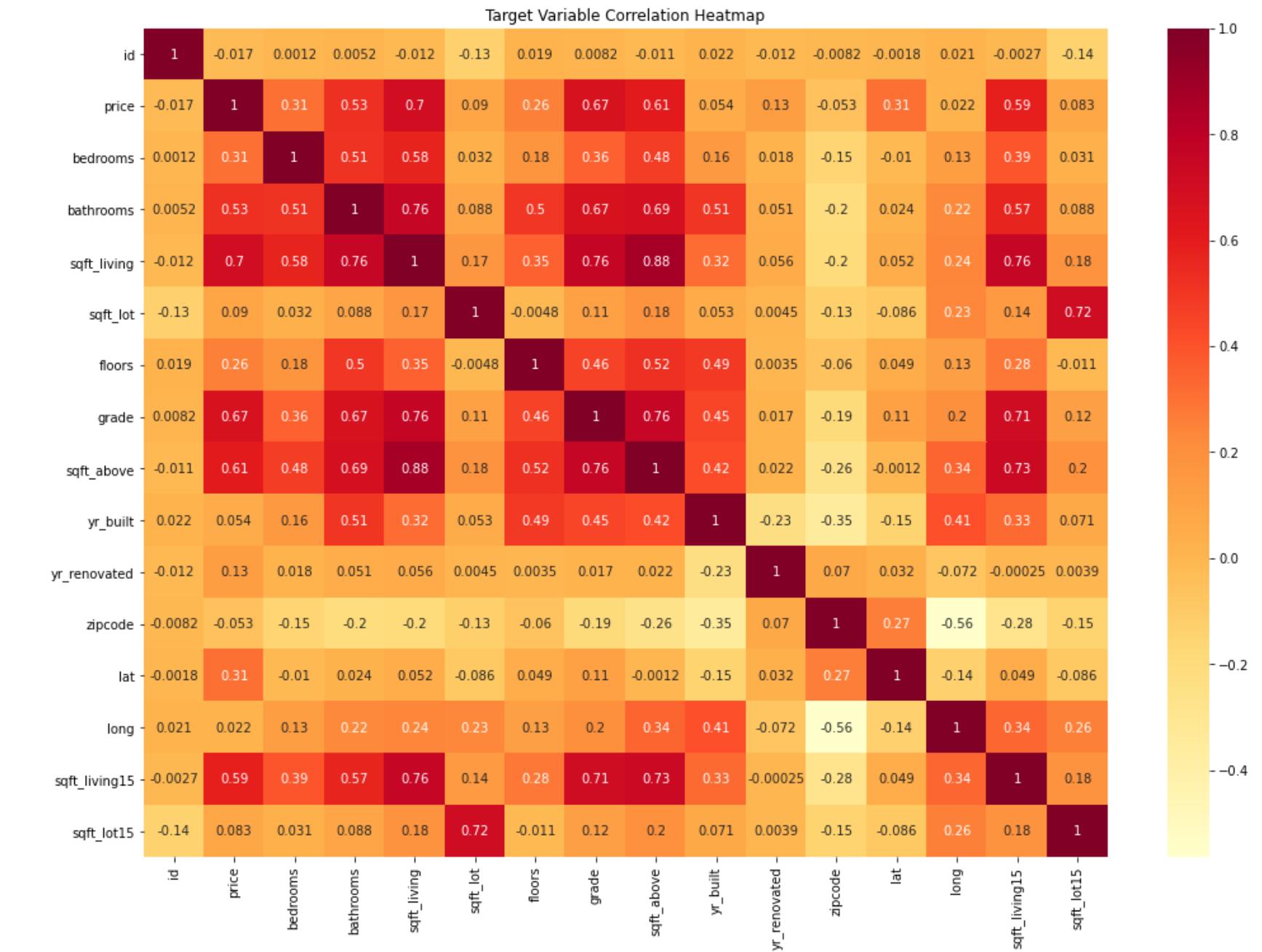
# OBJECTIVE 1 CONCLUSION

Square foot living and square foot that is above the ground floor(additional floors) independent variables are statistically significant in determining the dependent variable price. sqft\_living explains 49.6% of the variance in price and sqft\_above explains 37.1% of the variance in price. Hence the realtor might consider taking the sqft\_living and sqft\_above as determinants of fixing the value of the house in regard to price.

# OBJECTIVE 2

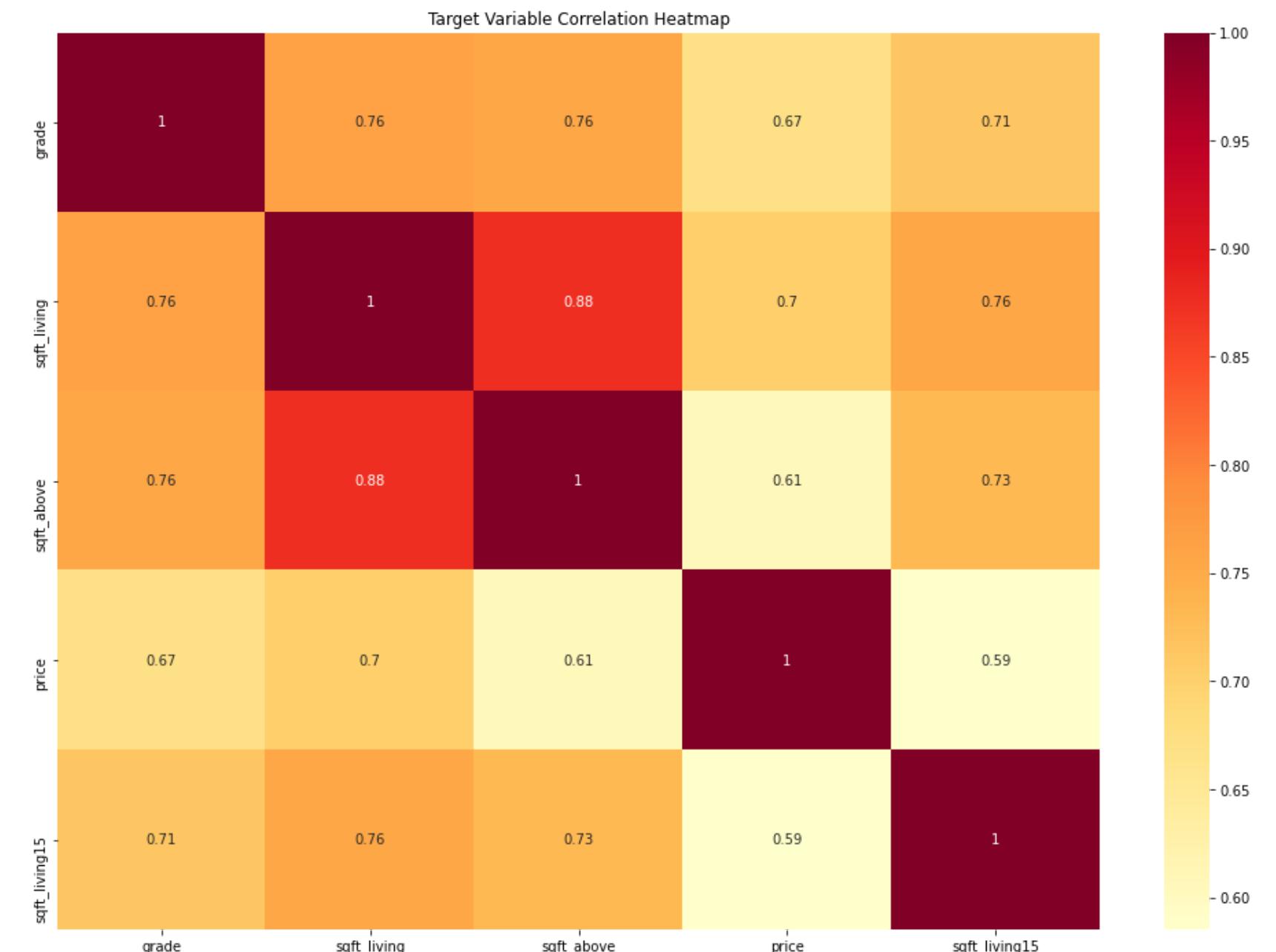
The second objective was to check factors of housing and their relationship to grading so as to see which ones will increase the grade of a house. This can be shown by a positive correlation.

Here we can see a correlation heat map with all the parameters



# OBJECTIVE 2

To determine the link between the dependent variable, grade, and the independent variables, sqft\_living, sqft\_above, price, and sqft\_living15, we used a multiple linear regression analysis.



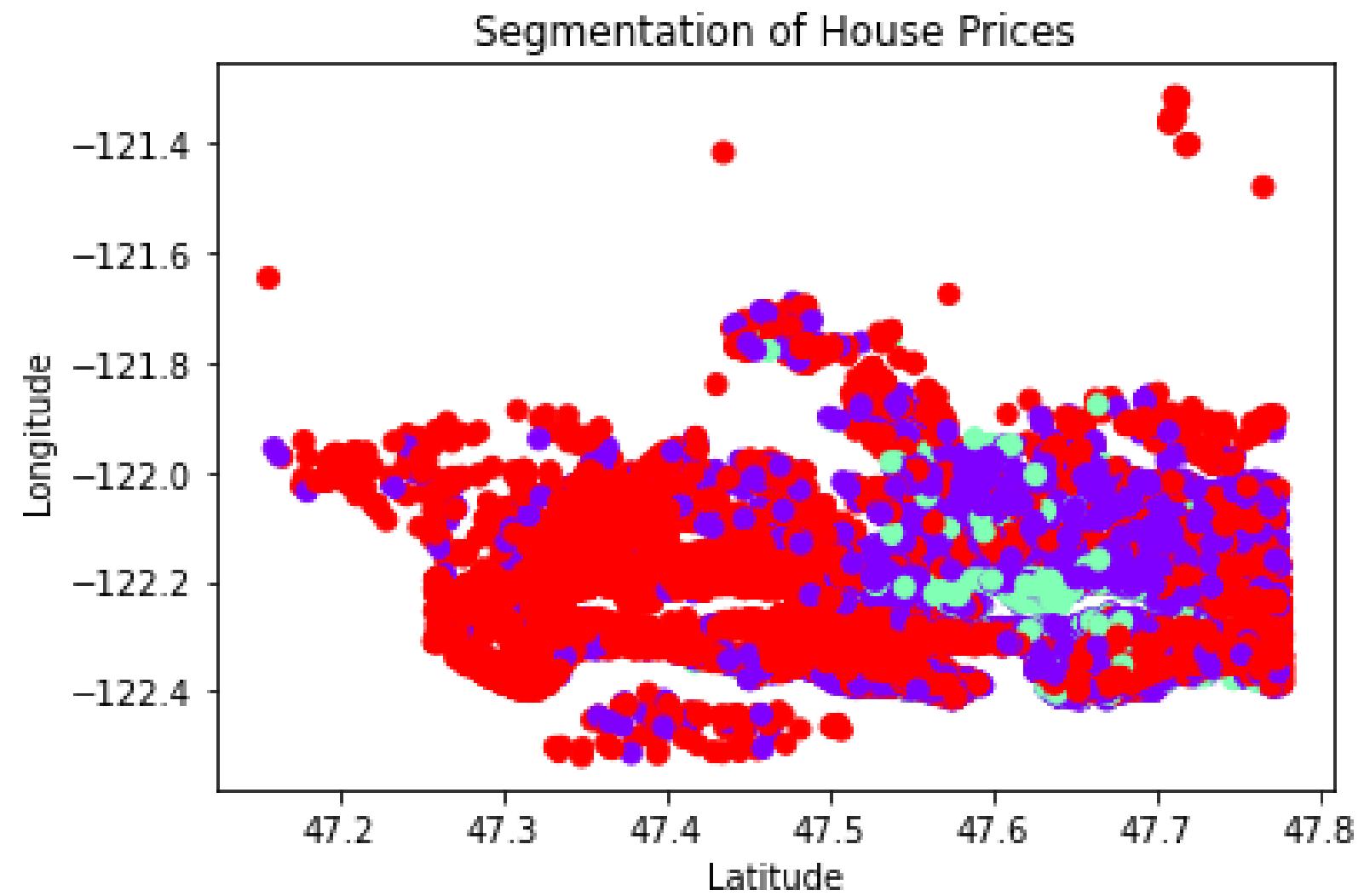
# OBJECTIVE 2 CONCLUSION

The grade of a house of a house has a high correlation with the size of the living room, the square footage of the house apart from the basement and price. This should be put under consideration if the intention is only on high-grade houses for profit maximization.

# OBJECTIVE 3

To identify areas of similar spatial clustering, we are going to use Segmentation analysis. Segmentation analysis is used to divide a dataset into distinct groups using an identified criteria. In this case, we are going to perform a segmentation analysis on house prices based on the latitude and longitude (The location of the property).

This is particularly helpful in creating 'neighborhoods' based on house prices. The real estate company can therefore know the house prices in the various geographic regions and where the real value for money would be.



# OBJECTIVE 3 CONCLUSION

**Given that Cluster 1 has the highest average home price, it is likely to have expensive or luxurious residences. The lowest average house price in Cluster 0 indicates that there are more reasonably priced or less expensive residences there. Cluster 2 is situated in the middle. Cluster 0 can symbolize homes in less affluent locations, whereas Cluster 1 might stand for homes in luxury neighbourhoods**

**These clusters may be helpful to investors in locating possible investment possibilities. Investment opportunities in inexpensive housing may exist in Cluster 0, whilst luxury real estate may be sought after in Cluster 1 meaning considerations should be on Cluster 1**

# NEXT STEPS

1.

Now that we know the factors that affect the price and grade of a house, the follow-up step is to find which areas have external developments(malls, developed roads) that would make people want to buy homes there.

2.

Observe areas where a lot of people are moving and find ways in which the real estate agency can invest there.

3.

To collect further parameters that may affect the price and grade of a house in order to make adjustments on the R-squared.

# OUR TEAM

**Dominic Mugo**

Scrum Master

[dominic.njue@student.  
moringaschool.com](mailto:dominic.njue@student.moringaschool.com)



**Jacquiline Tulinye**

Collaborator

[jacquiline.tulinye@student.  
moringaschool.co  
m](mailto:jacquiline.tulinye@student.moringaschool.com)



**Dennis Mutuku**

Collaborator

[dennis.mutuku@student.  
moringaschool.co  
m](mailto:dennis.mutuku@student.moringaschool.com)



**Onesphoro Kibunja**

Collaborator

[onesphoro.kibunja@sti  
dent.moringaschool.co  
m](mailto:onesphoro.kibunja@student.moringaschool.com)

