



PHASE
GROUP 2

KING COUNTY HOUSING PROJECT

<https://github.com/bizzfuzz/dsc-phase-2>



Our team

This project was worked on by Group 2 DSFT-08 HYBRID members. The work was subdivided into the stages of a datascience life cycle i.e: Business & Data Understanding, Data Preparation, EDA, Regression & modelling and Recommendations & Conclusion.

Ian Musau

*Scram master
Regression & Modelling*

Jacinta Chepkemoi

Data Preparation

Tabitha Berum
*Business & Data
Understanding*

Christine Malinga

*Exploratory Data
Analysis*

Business Problem

- The client is King County Realty. A real estate company looking to understand the anatomy of a high priced home. Armed with a dataset of house sales in King County, the team of analysts look to define the most sought after characteristics in a home.

Objectives

- To build models that can easily interpret the analysis of the data.
- Identify the characteristics that make houses most attractive to customers?
- Identify features that would add the most value to a house?

With this information, the client will be able to pick the most promising properties to sell.

Data Understanding

- This project uses the King County House Sales dataset.
- It has 21597 rows and 21 columns.
- Key variables: price, square footage_living, bedroom, grade and construction.
- The dataset also includes additional information such as id, date, price, bathrooms, sqft_lot, floors, waterfront, view, condition, grade, sqft_above, sqft_basement, yr_built, yr_renovated, zipcode, lat, long, sqft_living15, sqft_lot15

Data Preparation

Objective: preparing housing dataset for analysis by handling missing values, duplicates and outliers

Data loading and inspection

Handling missing values

Handling outliers

Data cleaning

You paragraph text



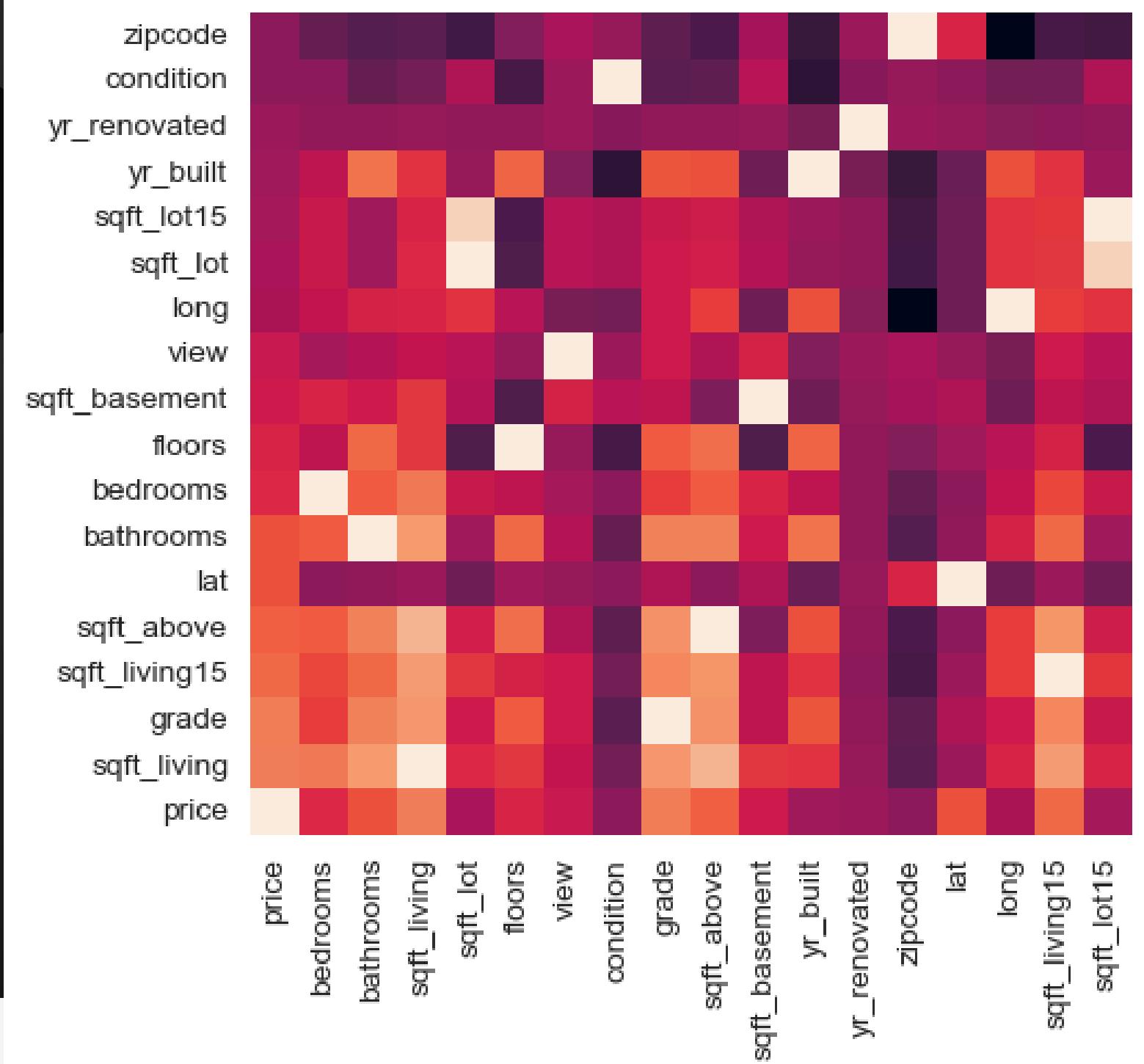
Exploratory Data Analysis

After examining and manipulating the data, these are the key data points:

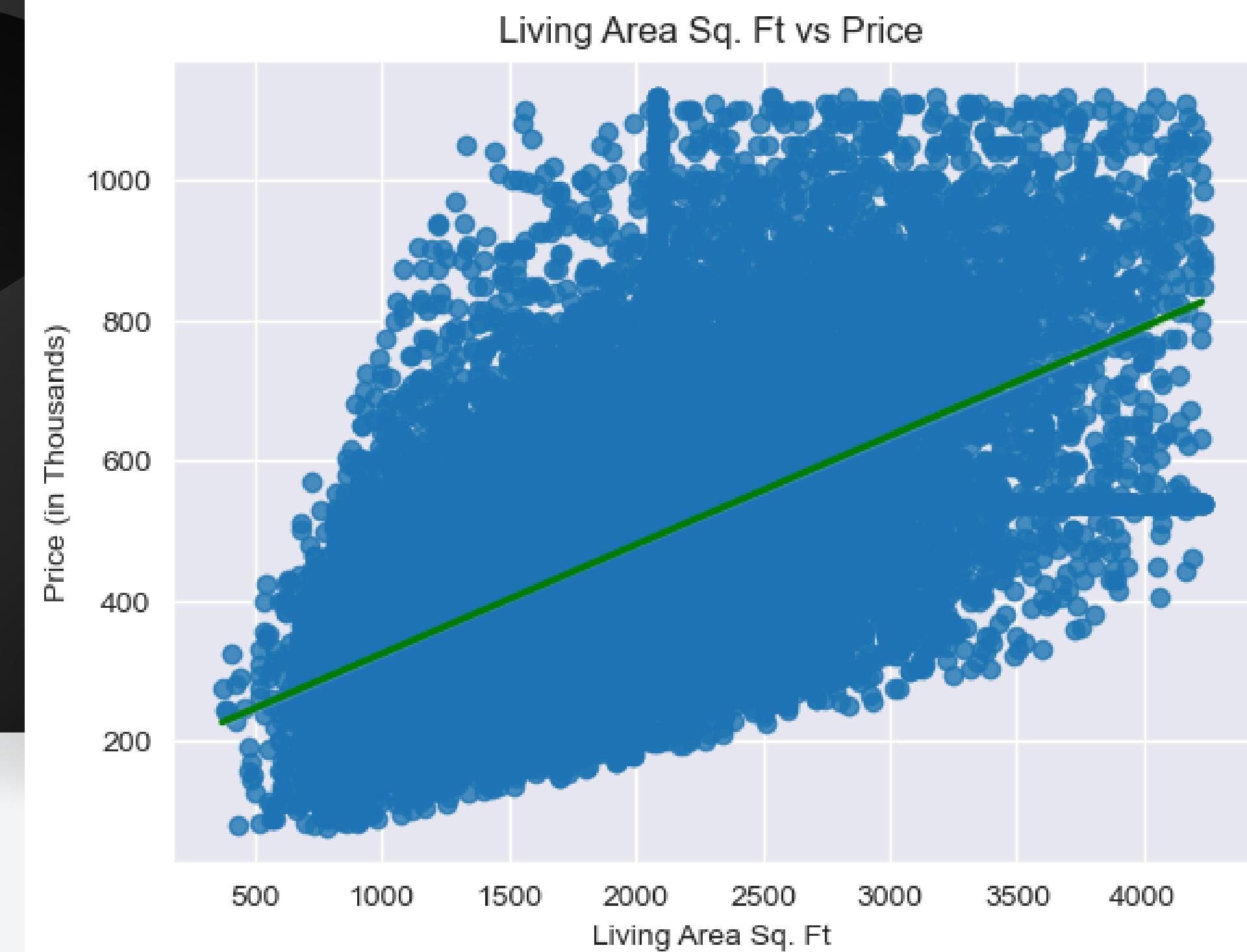
- The average house sold for \$480,000
- The price is most strongly linked to living area square footage



Correlation Heatmap



Price - Living Area Relationship



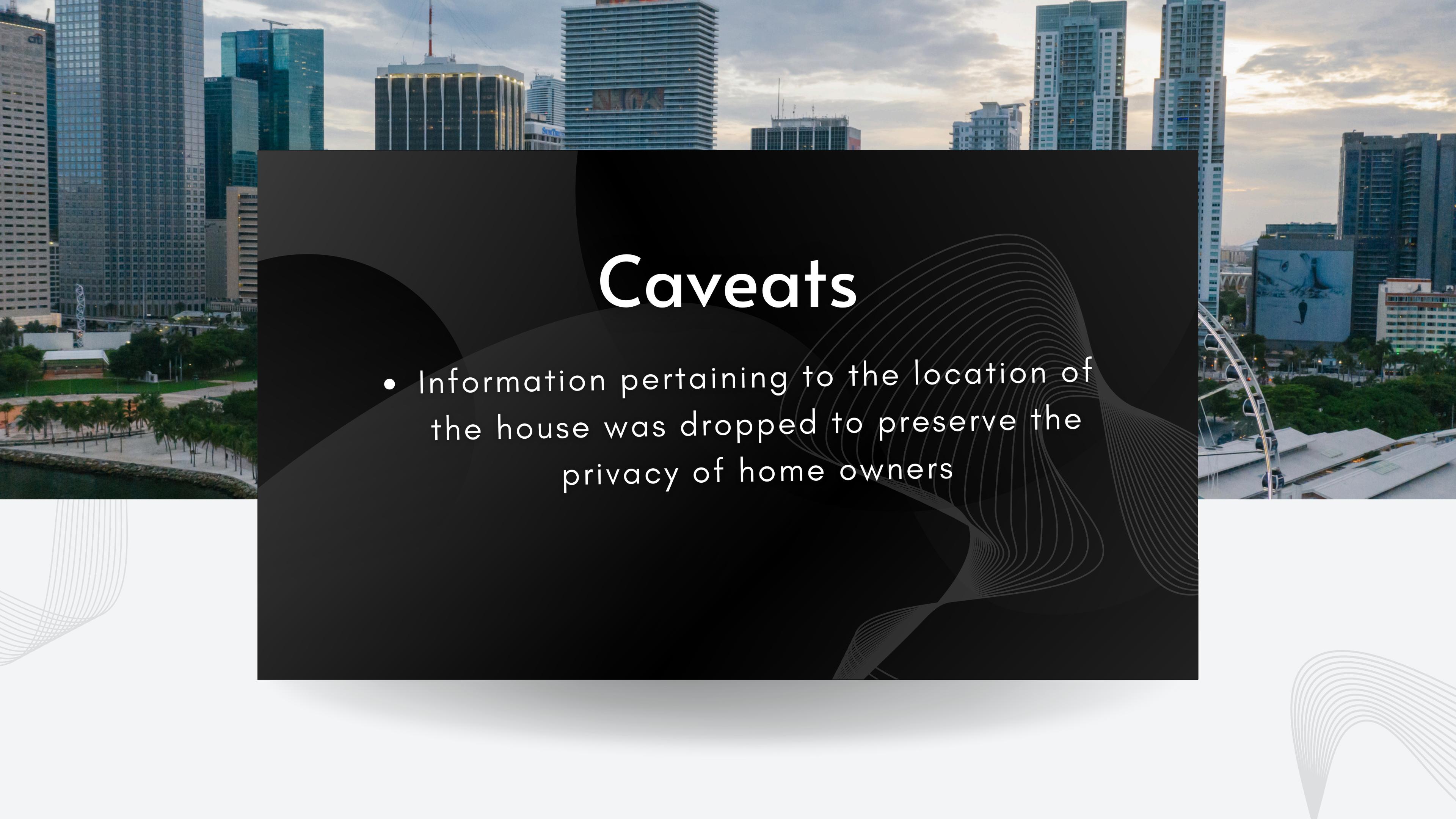
Data Modeling

After iterating the modeling process several times. We found the following patterns.

The background of the slide features a photograph of the Miami skyline at sunset. The sky is filled with warm, orange and yellow hues, transitioning into cooler blues and purples. Numerous skyscrapers of various heights are silhouetted against the bright sky. In the foreground, there's a dark, semi-transparent graphic element consisting of a large black rectangle with white wavy lines radiating from its center, resembling a stylized sun or a thought bubble.

Findings

- Each addition to the number of floors in a house increased the price by \$19,792
- Every bathroom accounted for an increase of \$16,470
- An increase in condition ranking raised the price by \$9,195

The background of the slide features a photograph of the Miami skyline at sunset. The sky is filled with warm, orange and yellow hues, transitioning into cooler blues and purples. Numerous skyscrapers of various heights are silhouetted against the bright sky. In the foreground, there's a dark, semi-transparent graphic element consisting of a large black rectangle with white wavy lines radiating from its center, resembling a stylized sun or a signal. The overall aesthetic is modern and professional.

Caveats

- Information pertaining to the location of the house was dropped to preserve the privacy of home owners

Conclusions

- Model Effectiveness: The S.L.R.M explains only about 33% of the variation in house prices while the MLRM explaining approximately 51% of the variance in house prices.
- Significant Variables: intercept and square footage living area coefficients in the SLRM are statistically significant. In the MLRM variables such as the number of floors, number of bathrooms, and condition of the house are statistically significant contributors to house prices.
- Price Increase Factors: Each floor adds approximately \$19,792.10, while each bathroom adds approximately \$16,470.20. Improving the condition of the house can also lead to higher prices, with each increase in ranking contributing approximately \$9,195.20.

Recommendations

- Improving the predictive power of the regression models by including variables that are statistically significant and dropping those that are not.
- Pay close attention to factors such as the number of floors, number of bathrooms, and condition of the house, as these variables have a significant impact on house prices.
- Consider collecting additional data or exploring alternative modeling techniques to further enhance the accuracy of house price predictions.



Group
Phase 2

Thank You

Gerente General