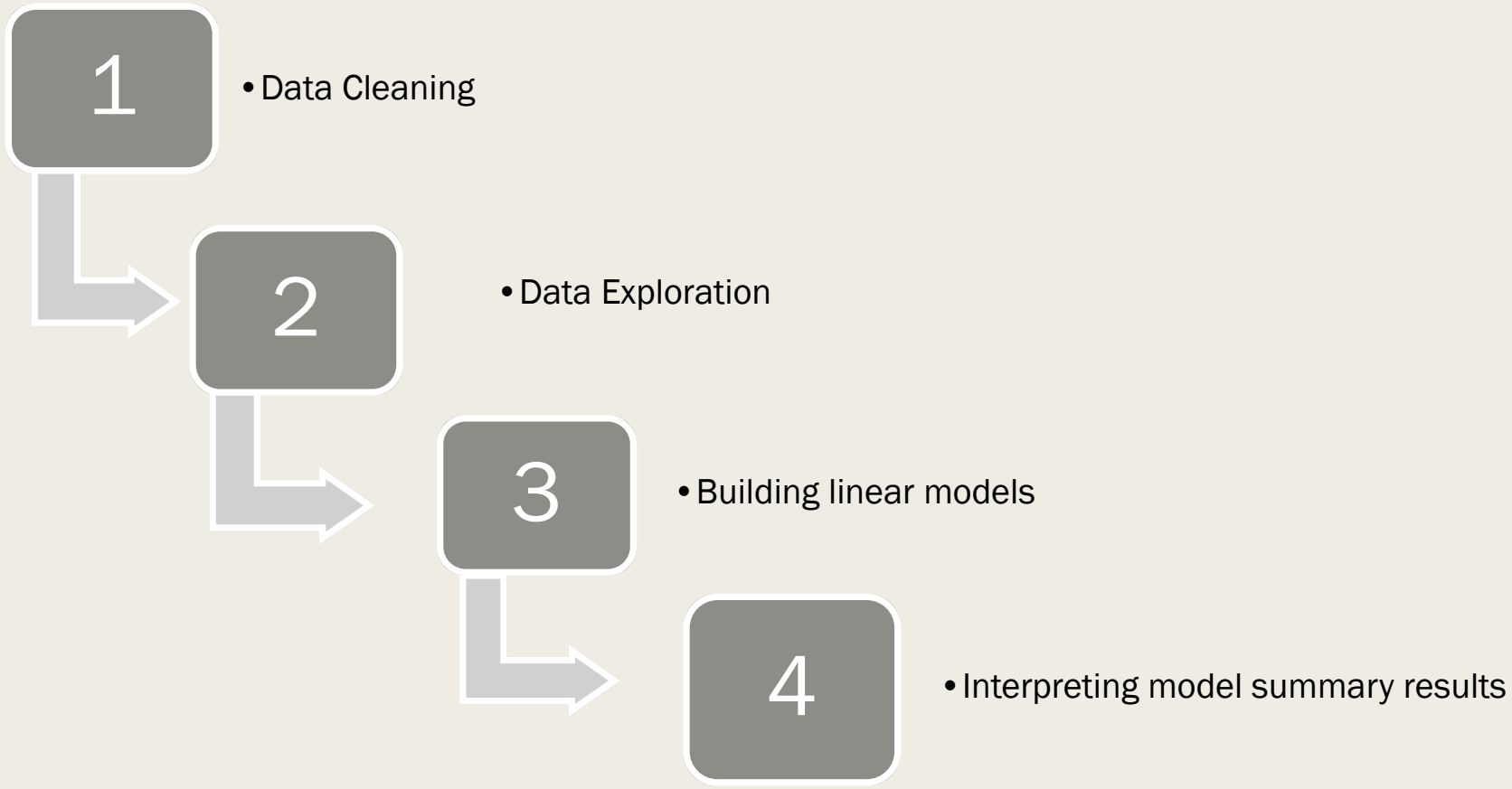
PREDICTING HOUSE PRICES IN KING COUNTY

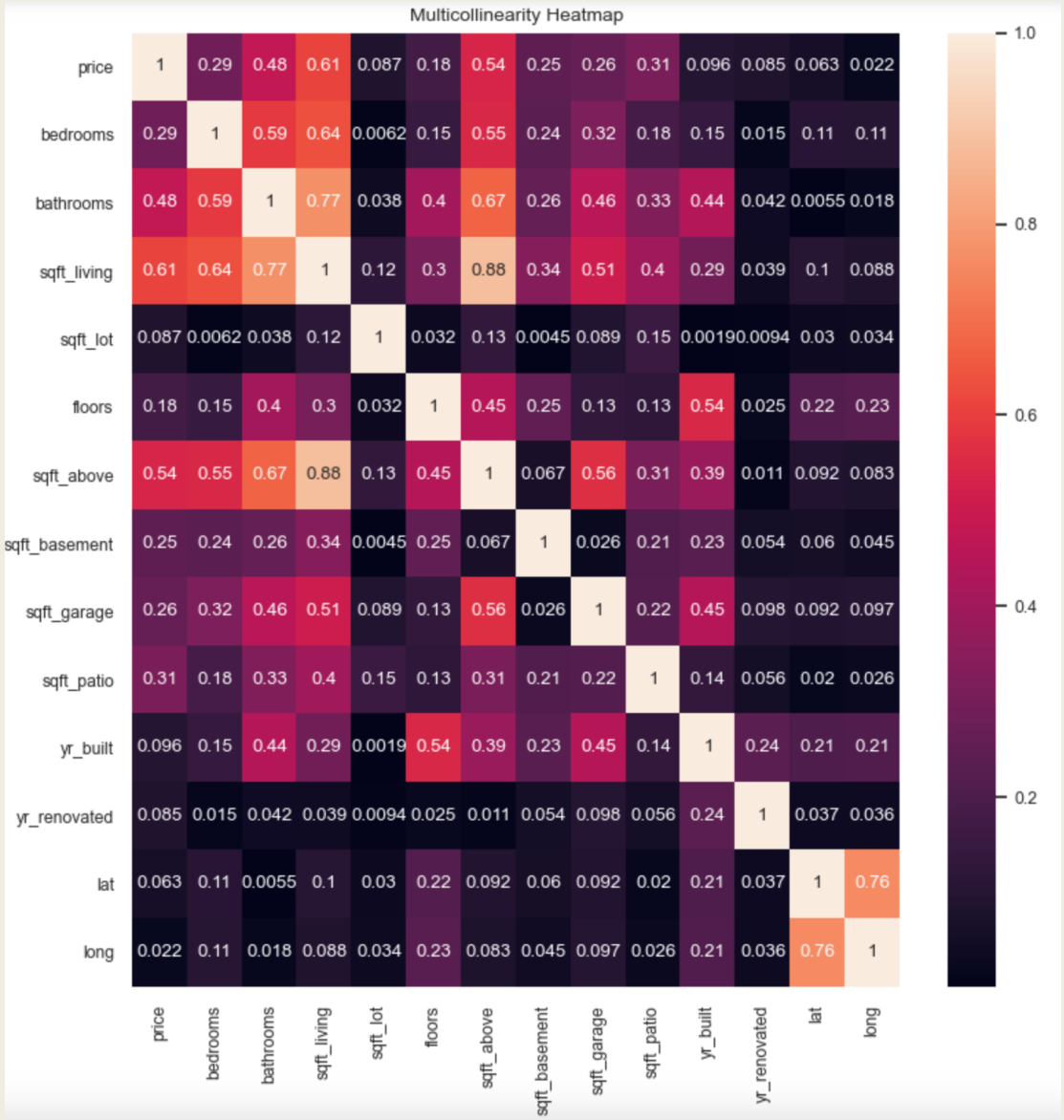
Deanna Gould

DATA

The data for this model utilizes the King County dataset, which includes information about the square footage, location, condition, and heating systems amongst other features. Before cleaning the data, there were 30,155 homes that were in the dataset.

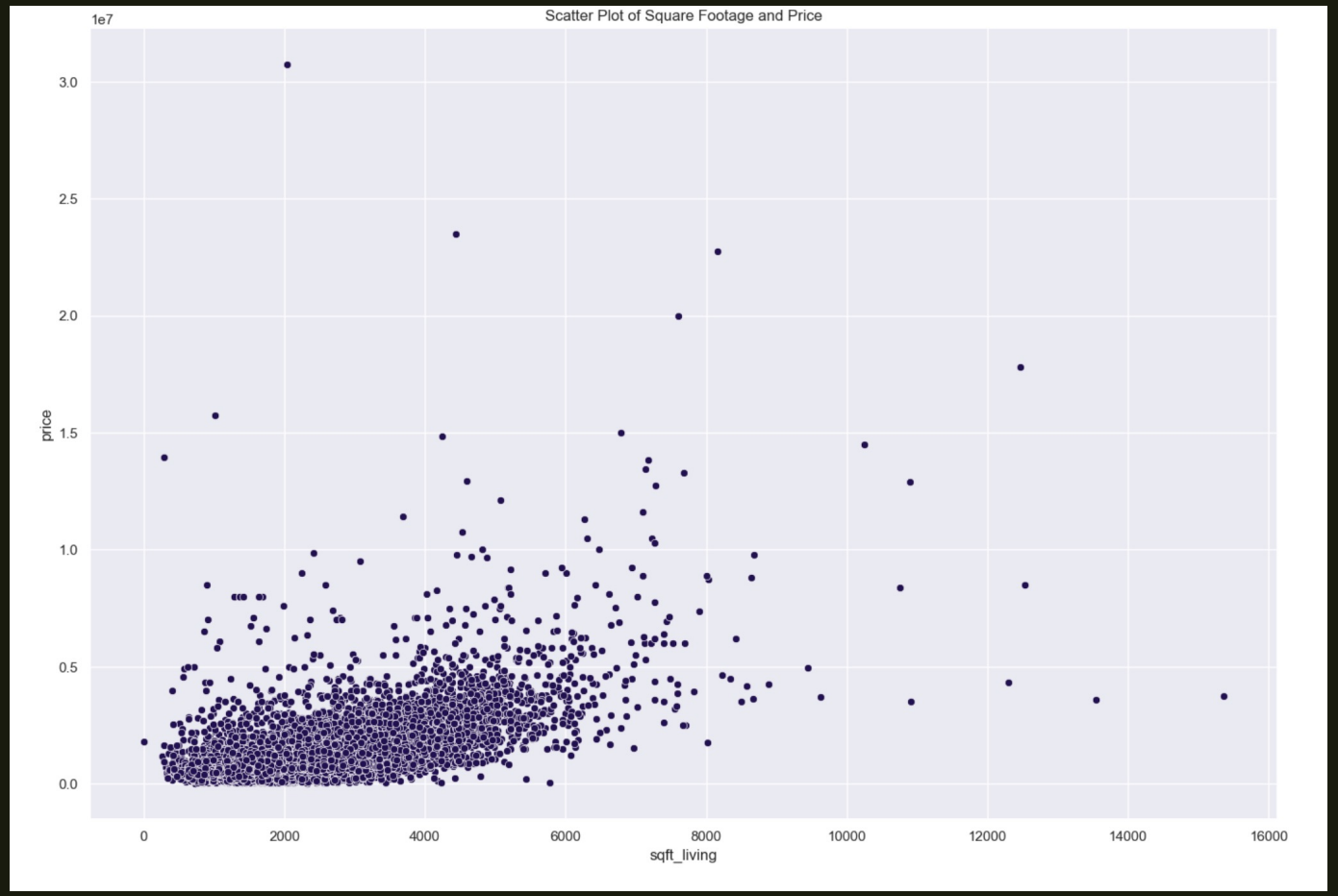
Modeling Process





Correlations

- Having features that are too highly correlated (multicollinearity) won't give accurate results. Features that are highly correlated to price but not *too* correlated to one another will be included in the model.
- Variables lighter than color in the leftmost column have higher correlations to price.



Is bigger always better?

- Investing in homes based on the target market
- Too big of homes can have a worse impact on price
- While outliers were removed from the model, square footage still has the highest correlation to price.

Final Model Results

- The final R-squared of the model was 0.657, which means the model can account for 65.7% of the variability in price.
- The Mean Absolute Error (MAE) of the model was $169,097.32 USD, which is the average error in predicting home prices.
- To determine a feature's eligibility in the model, the P-value needed to be below 0.05.

Recommendations

- Create more bedrooms if possible
- Expand the home
- Have a similar amount of bathrooms to bedrooms
- Maintain or renovate the home to be in good or very good condition.

Condition, square footage, and the number of bathrooms and bedrooms are the strongest predictors of house prices.

Other Possibilities

- This model can be recreated using the steps in the Jupyter notebook, but if I were to redo this project, I would potentially change some of the steps in my process.
- Something to keep in mind as a real estate investment firm – is the goal to improve homes in poor condition? Or is the goal to buy the best homes?
- How would the model change if outliers were not removed?

Thank You!