How to Maximize the Selling Price of your King County Home

Problem Statement

- High price range in King County
- Wide range of property types
- Finding an advantage

Methodology

Data

Model Selection - Multi-linear Regression

Iterations to fix overfitting

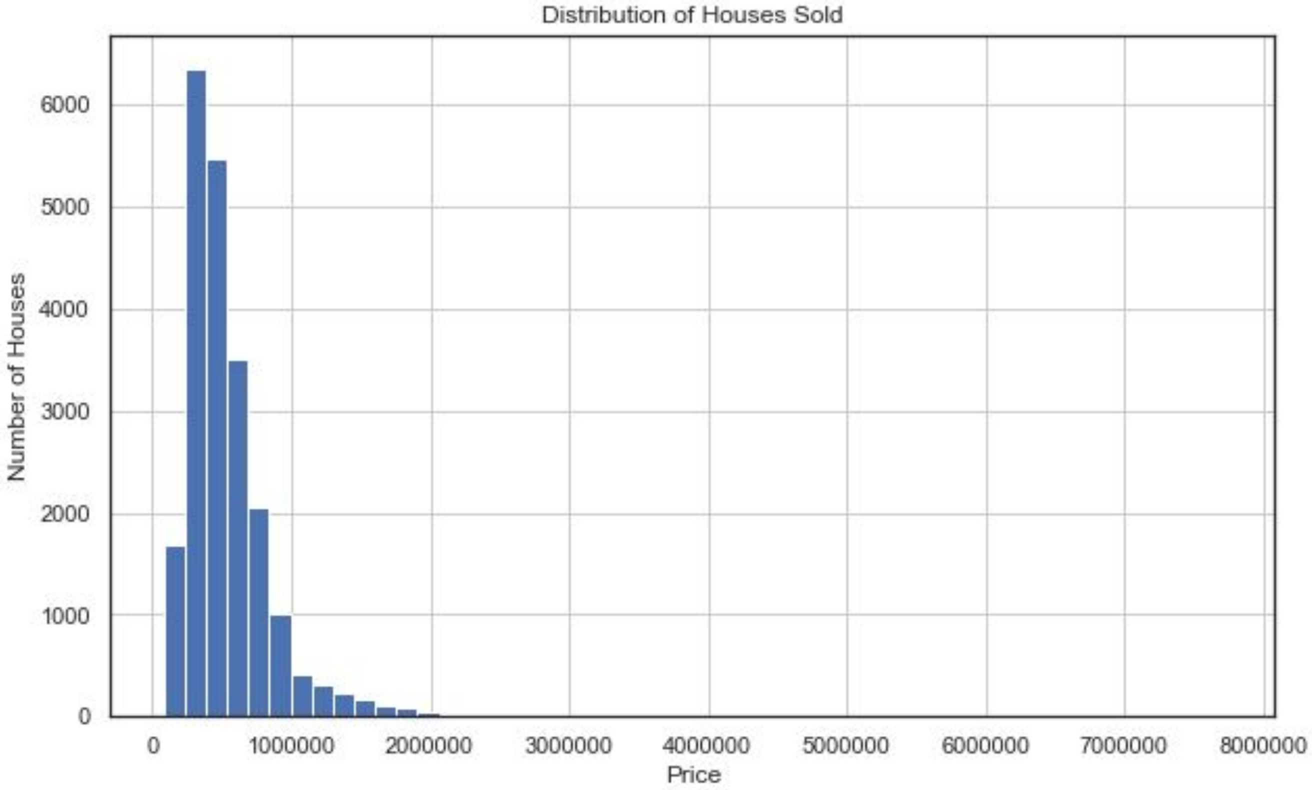
Recursive Feature Elimination

Predictive accuracy of 83%

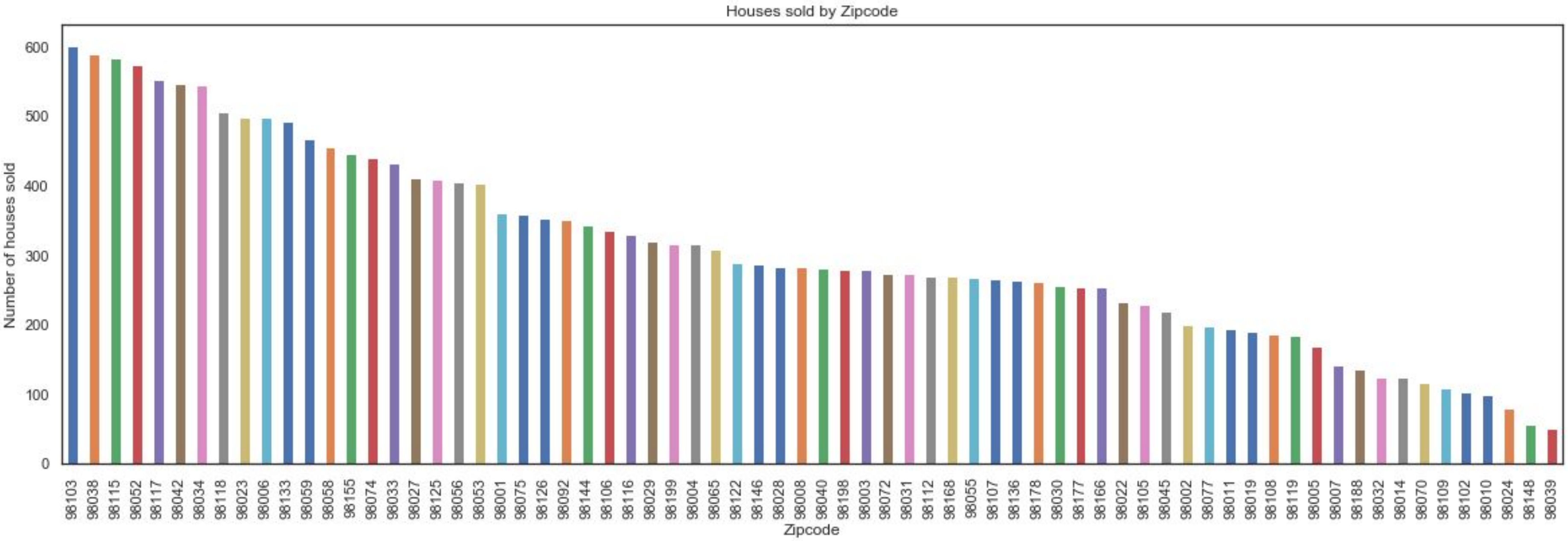
Business Value

1. What does the data show as trends in sales price?
2. What areas, by zip code, are selling the most? What areas are selling the least?
3. Should you buy on the waterfront?
4. Should you renovate?
5. How can maintenance and upkeep affect your final selling price?

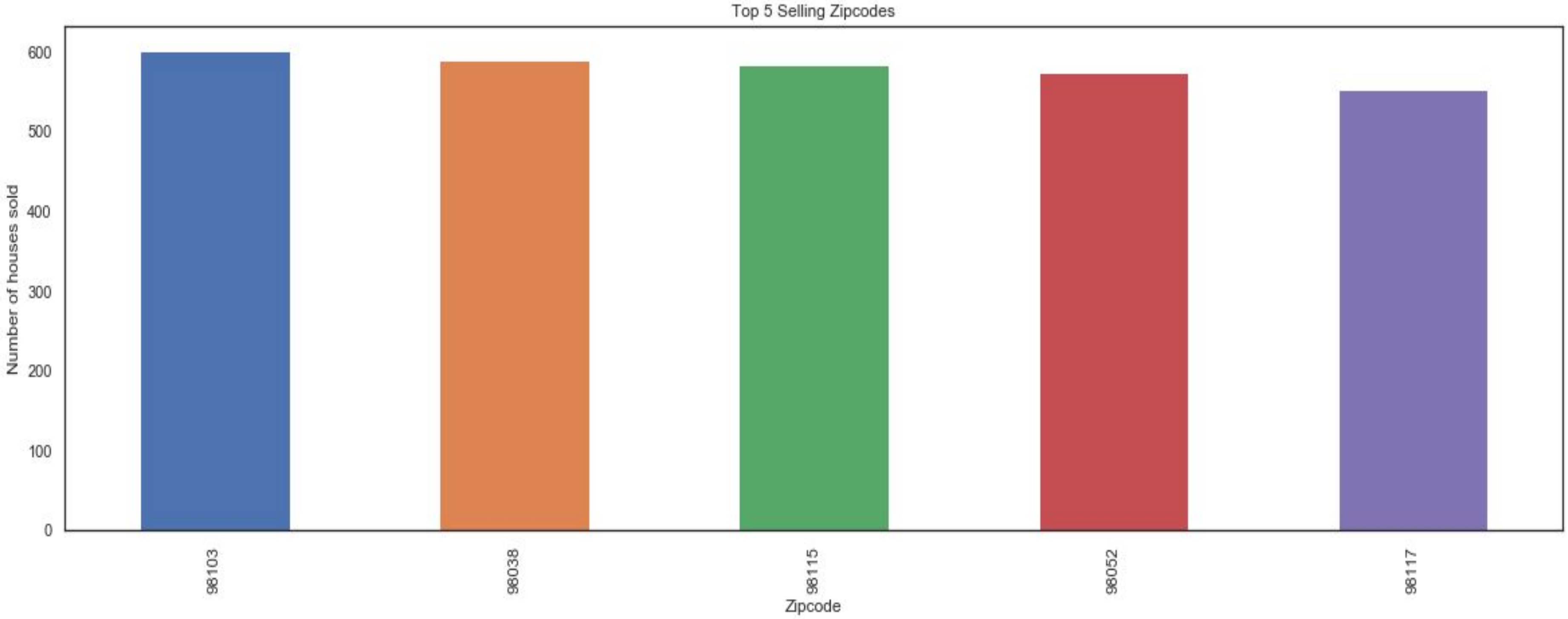
1. What to expect from the sales data 2014/2015



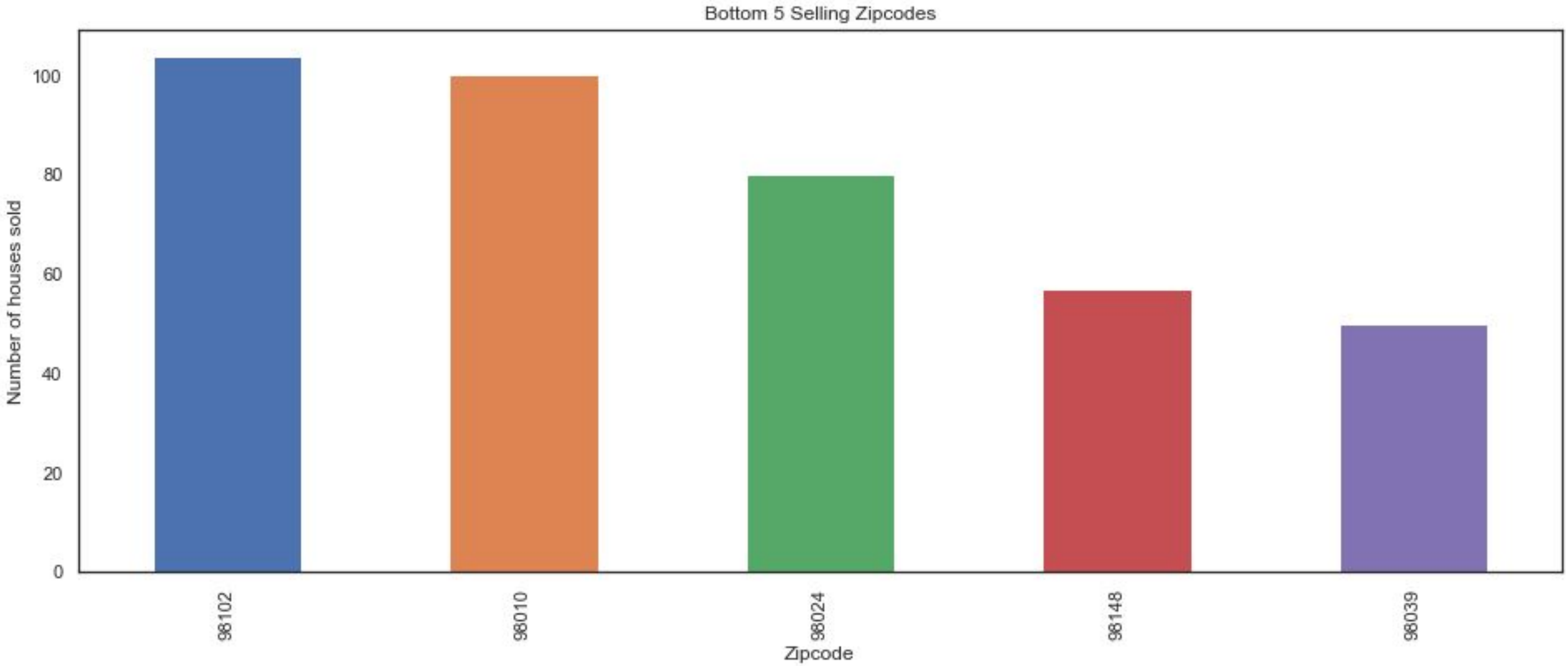
2. Houses Sold by Zip Code



Zip Codes with Most Sales



Zip Codes with Least Sales



3. Should You Buy on the Waterfront?

Average price of houses sold on the waterfront = $ 1.7 million

Average price of houses sold off the waterfront = $530,000

Difference of roughly $535,000

Average cost of flood insurance = $870/year

https://www.betterflood.com/average-cost-of-flood-insurance-in-washington-state/#Kingcounty

4. Should You Renovate?

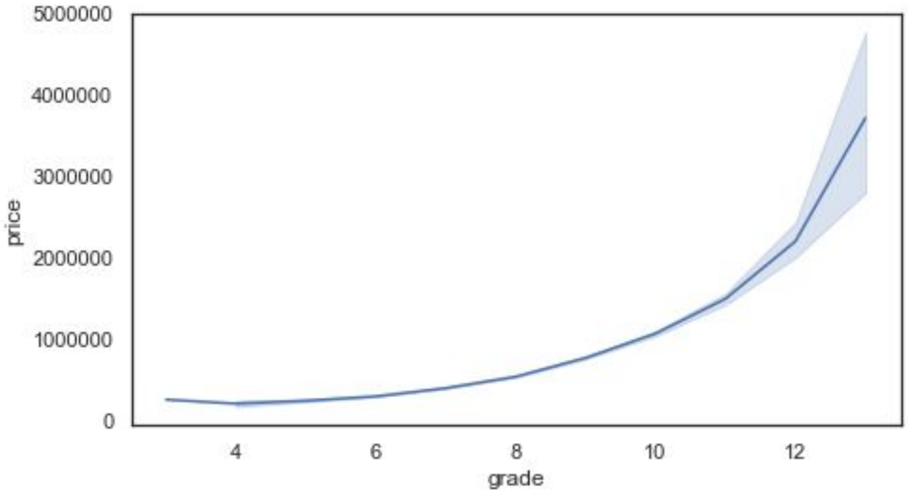
Average price of renovated houses sold = $ 780,000

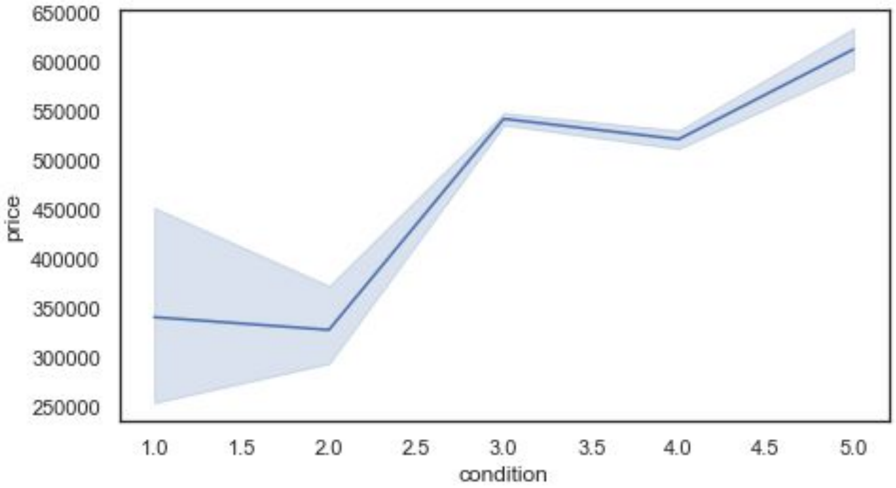
Average price of non-renovated houses sold = $533,000

Difference of roughly $250,000

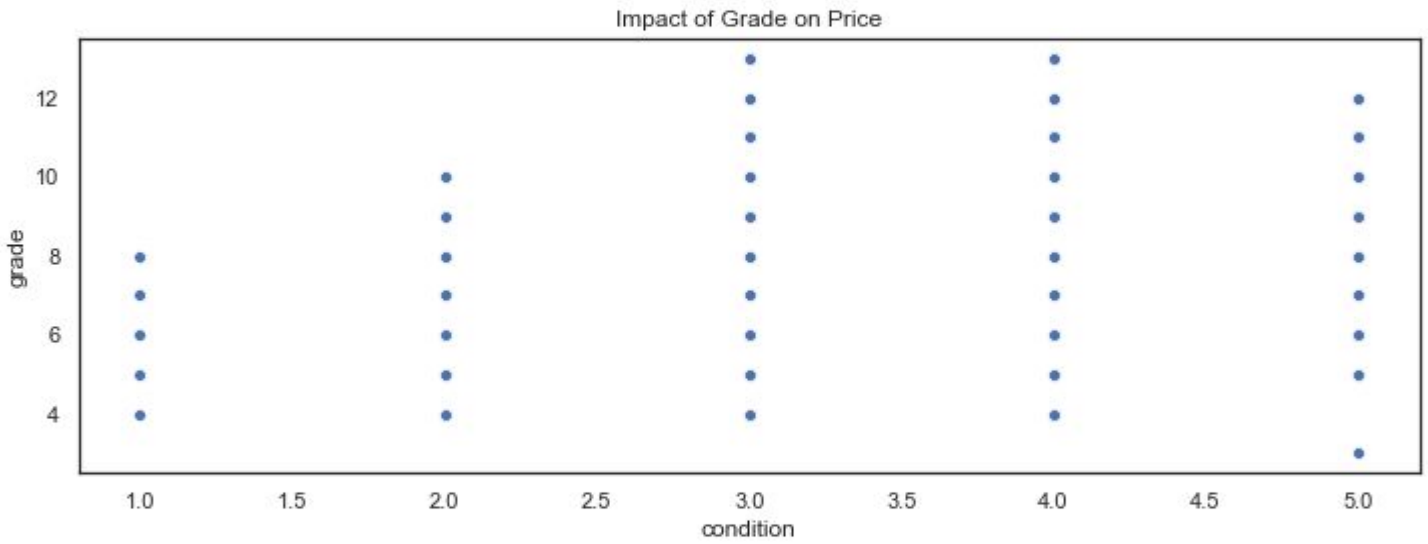
Average age of homes = 43 years old

5. How Much Effort Should Go into Maintenance?

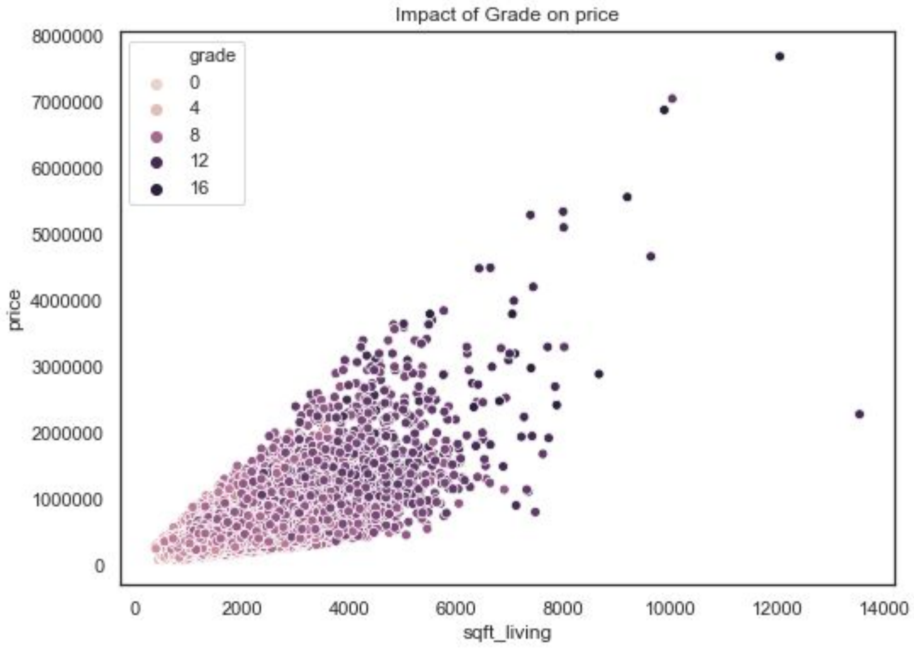


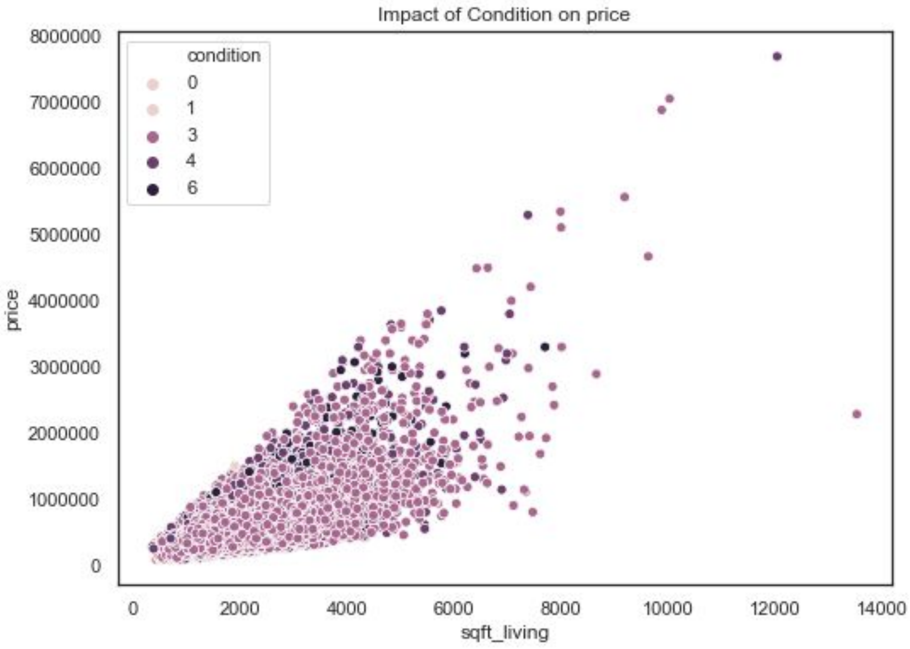


5. How Much Effort Should Go into Maintenance?



5. How Much Effort Should Go into Maintenance?





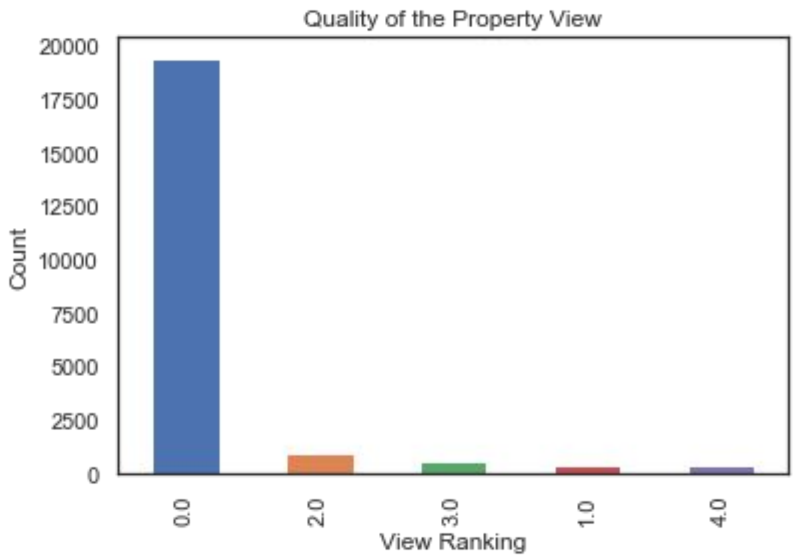
Key Takeaways

1. Depending on your homes features, expect to sell below $500,000.
2. Look for zip codes that are selling well, find out why certain zip codes are not selling well.
3. Waterfront sales are significantly higher
4. Renovation can increase sales price.
5. Understand the grade and condition ratings and how you can improve your ratings.

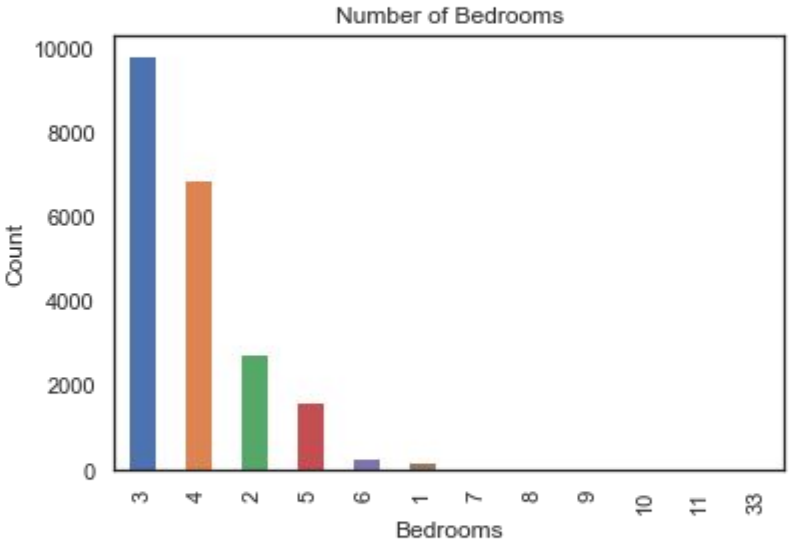
Future Work

- Use coefficients and predictive modeling to create a user friendly price predictor.
- Apply same model to updated and larger data sets.
- Gradient boosting.

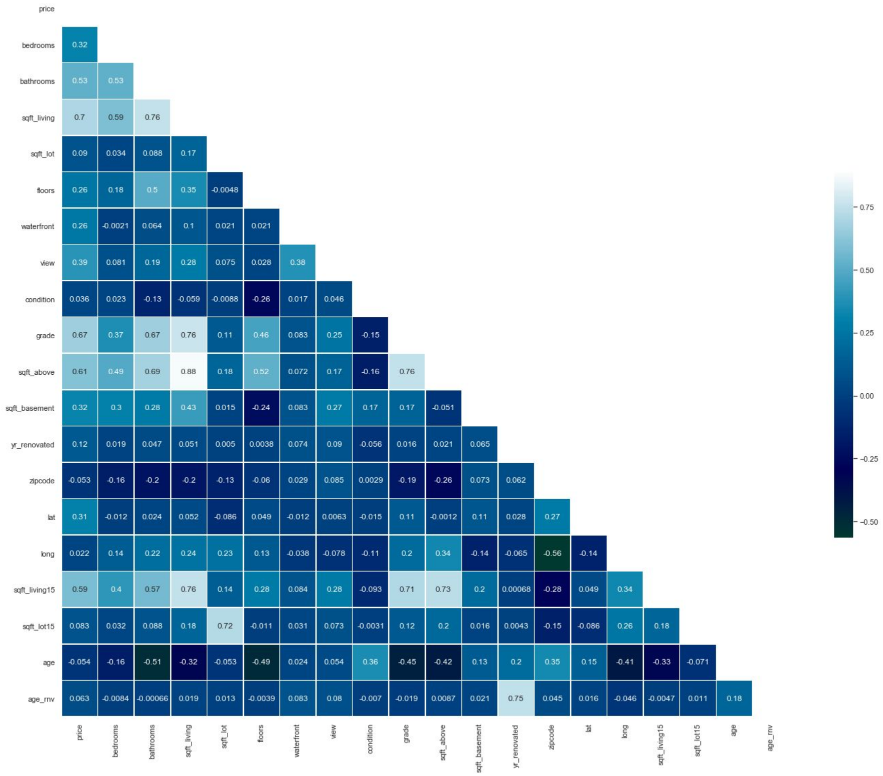
Index 1. - View Rankings



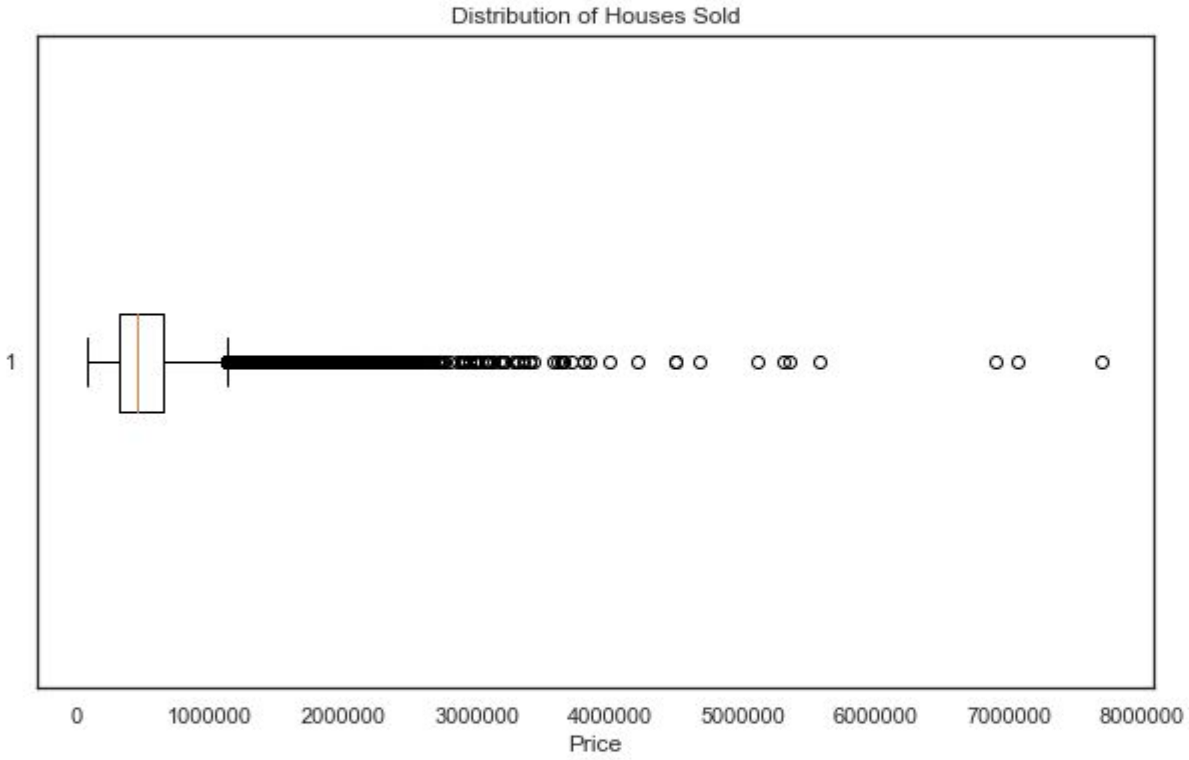
Index 2. - Number of Bedrooms in Homes



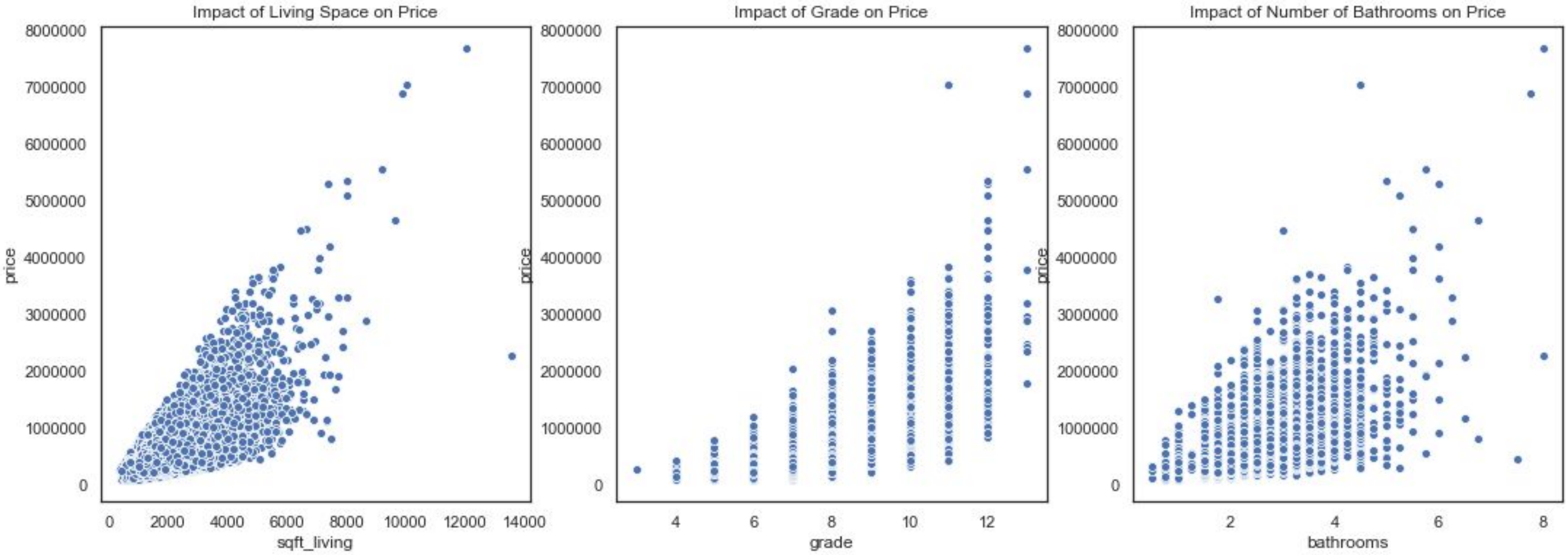
Index 3. - Multicollinearity Heat Map



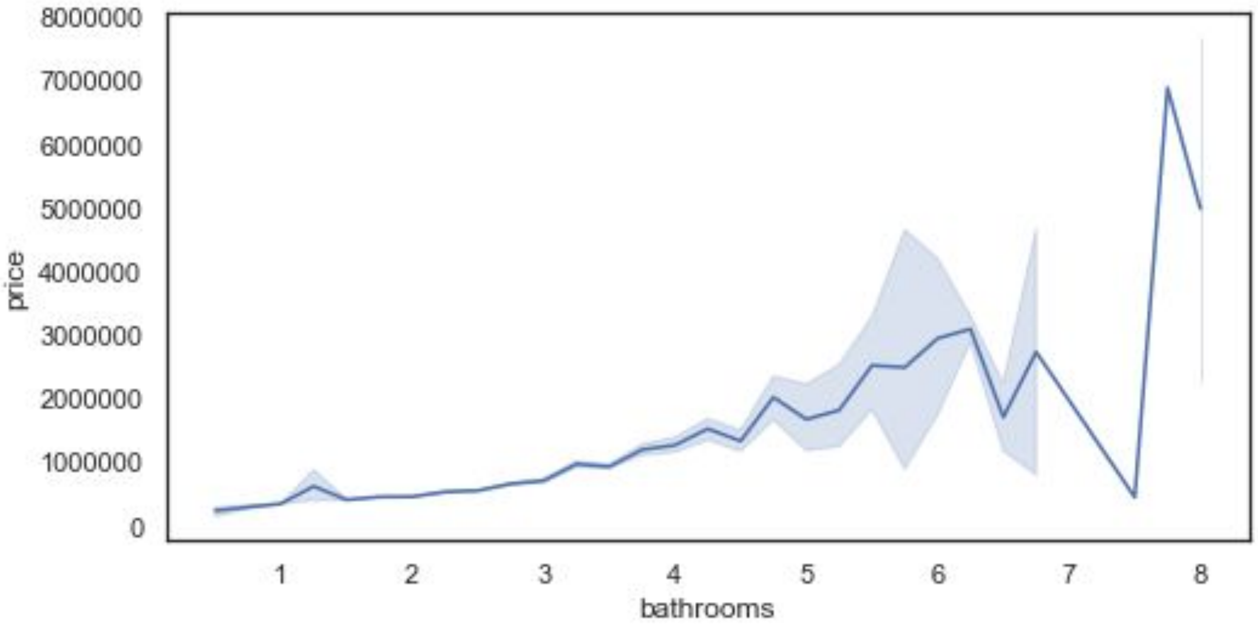
Index 4. - Distribution



Index 5. - Linear Relationships



Index 6. - Bathrooms



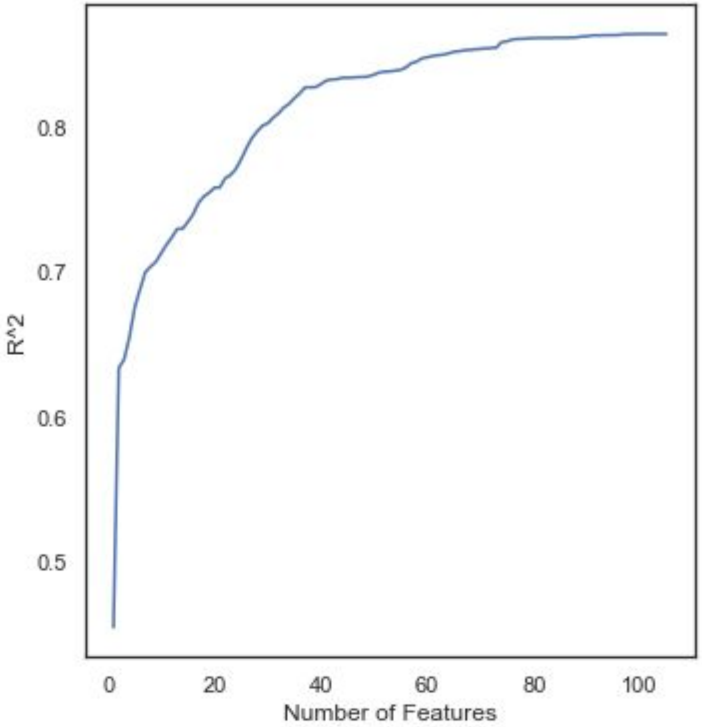
Index 7. - RFE Top 5

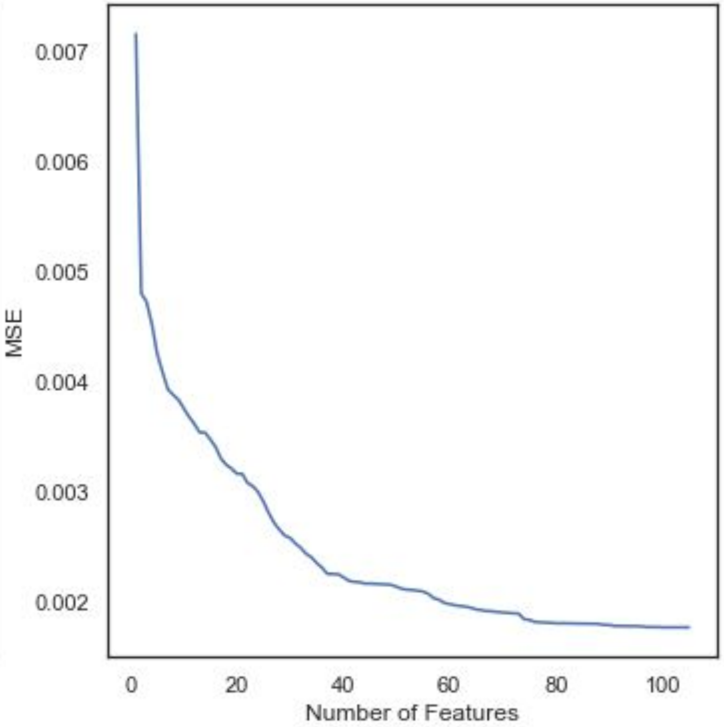
In [1236]:
```
# The top 5 features

selected_features.sort_values(by=['Scaled']).head()
```

Out[1236]:

	Column	Scaled
0	sqft_living	1
2	lat	2
60	zipcode_98039	3
28	waterfront_1	4
39	zipcode_98004	5

Index 8. - Feature Selection





Index 9. - OLS

In [1242]:
```
# test the accuracy of the model with the top 35 features using an OLS technique

predictors = sm.add_constant(X[selected_columns])
model = sm.OLS(Y, predictors).fit()
model.summary()
```

Out[1242]:

Dep. Variable:	price	**R-squared:**	0.834
Model:	OLS	**Adj. R-squared:**	0.834
Method:	Least Squares	**F-statistic:**	2412.
Date:	Sat, 06 Apr 2019	**Prob (F-statistic):**	0.00
Time:	12:21:18	**Log-Likelihood:**	35544.
No. Observations:	21597	**AIC:**	-7.100e+04
Df Residuals:	21551	**BIC:**	-7.063e+04
Df Model:	45		
Covariance Type:	nonrobust		