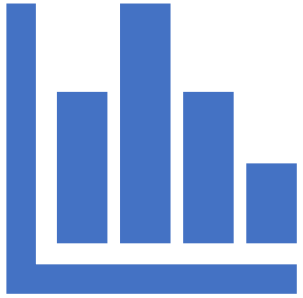
King County Housing

Affordable Housing Campaign

Predictive Modeling of House Prices

Overview



Multiple linear regression model to predict house prices in King County

Derive data driven recommendations on features that best predict house prices

Outline

Business Problem

Business Value

Data and Methods

Exploratory Data Analysis

Linear Regression Modeling

Results

Conclusion

Recommendations

Business Problem

- Scarcity of affordable housing in King County
- Misrepresentation of house grade by property developers
- Limited housing driving prices up
- Need for developers to focus on features that accurately predict housing prices and enhance affordability

Business Value

- Accurate prediction to guide developers on which house features to focus on
- Narrowing of features down to the most important ones using data driven insights
- Boosting development of affordable properties focused on resolving the housing scarcity affordably

Data and Methods

- King County Housing Data Set
- Contains data on 20 variables describing housing such as sale price, number of bedrooms, no. of bathrooms e.t.c
- Data cleaning
- Exploratory data analysis
- Linear regression modeling

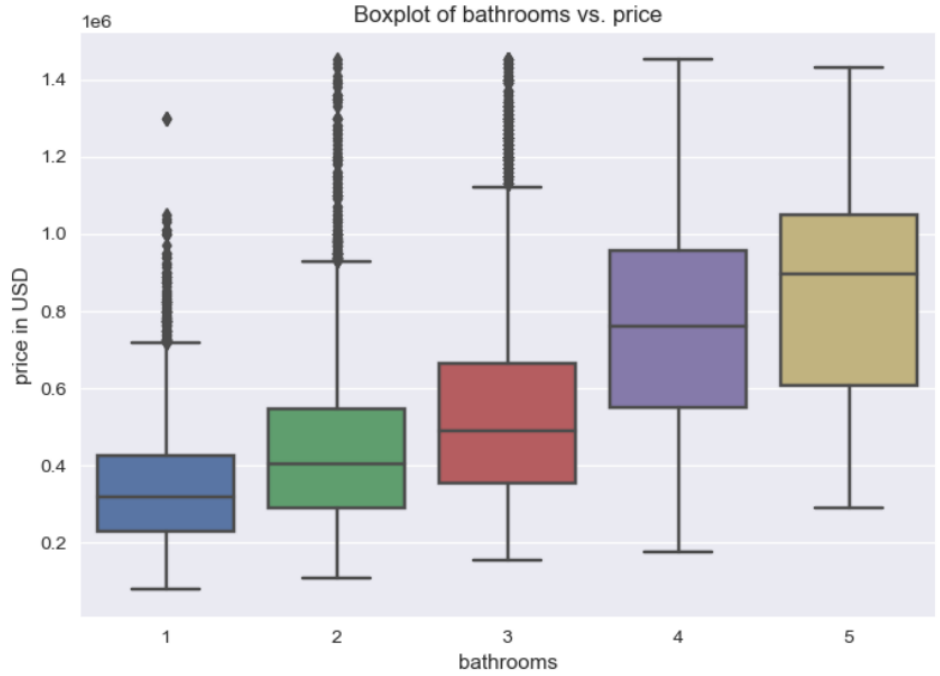
Exploratory Data Analysis

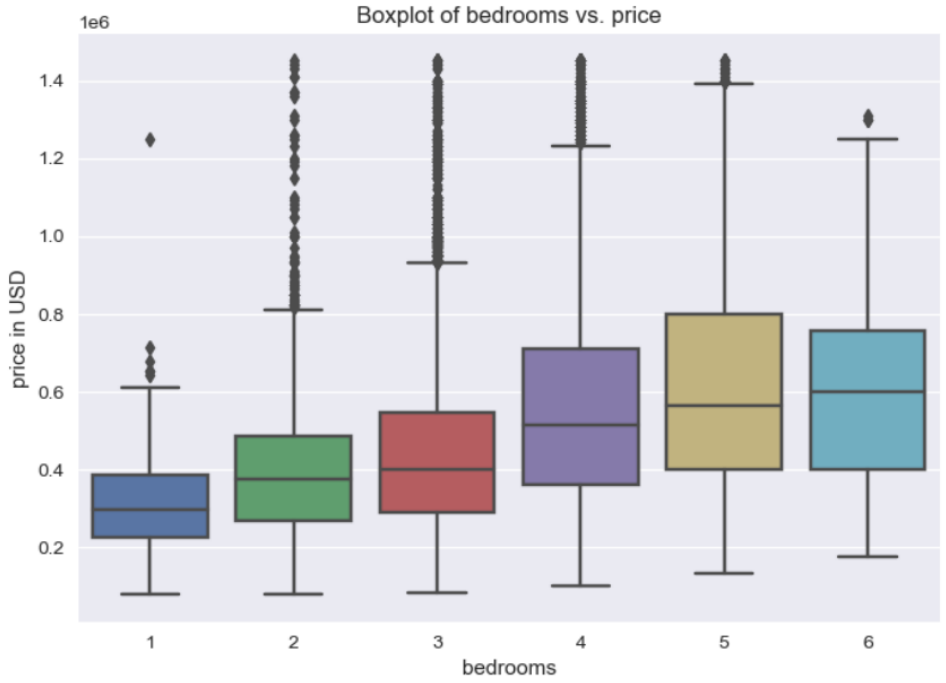
Waterfront

- Mean price of a house with a waterfront - $890,690
- Mean price of a house without a waterfront - $483,804
- Waterfront has a significant effect on the price with the mean price of houses with waterfront being almost double of those without.

Bedrooms and Bathrooms

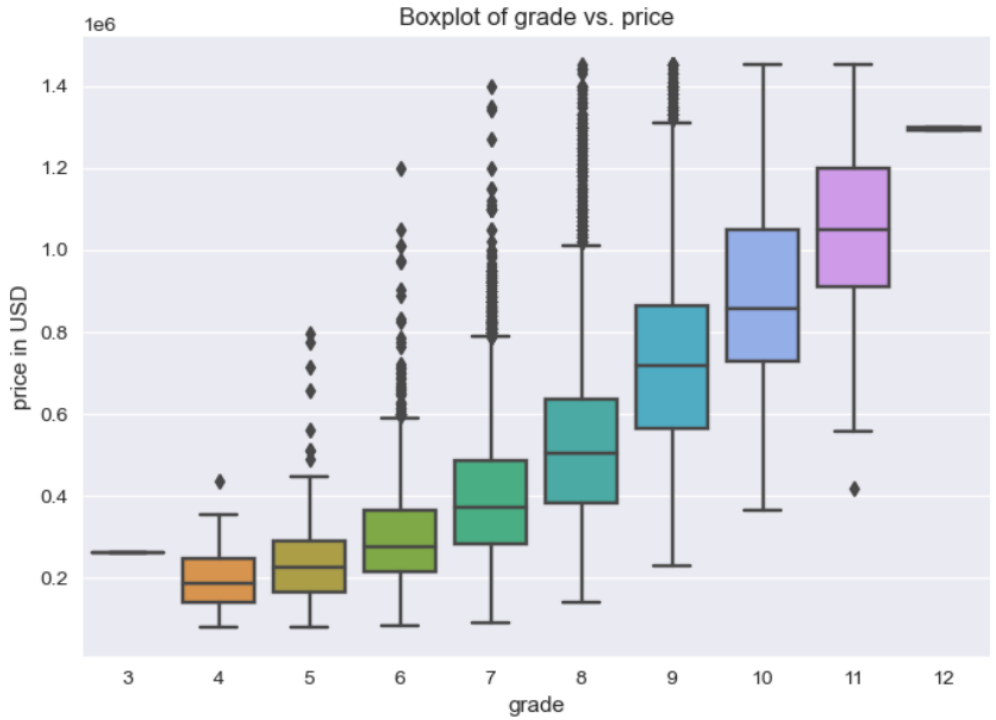
- As bedrooms increase so does the price.
- 5 bedrooms most preferred.
- As the bathrooms increase the price increases.

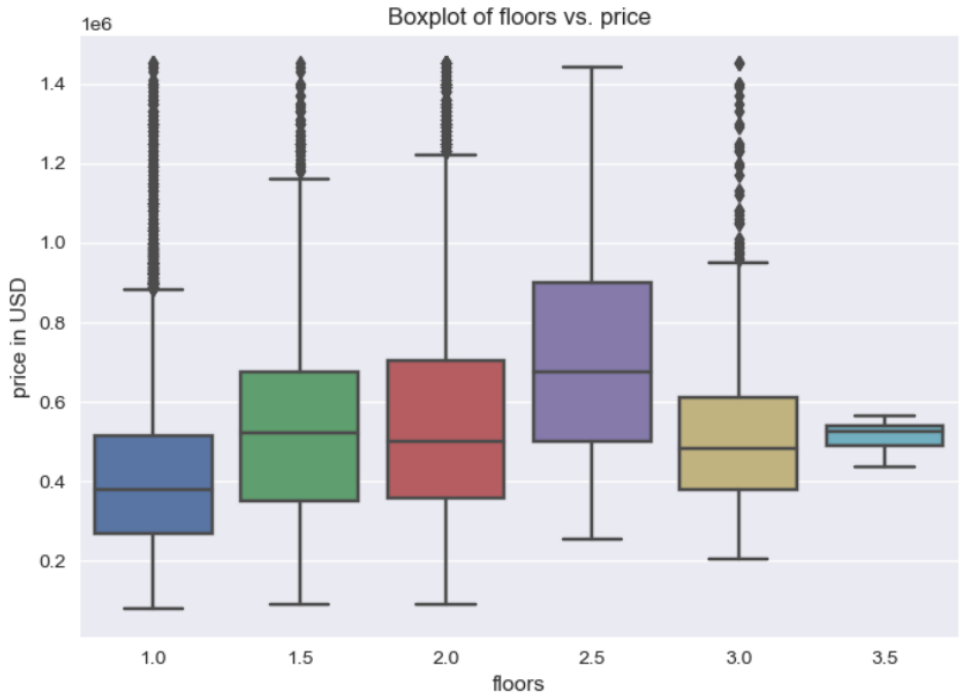




Floors and Grade

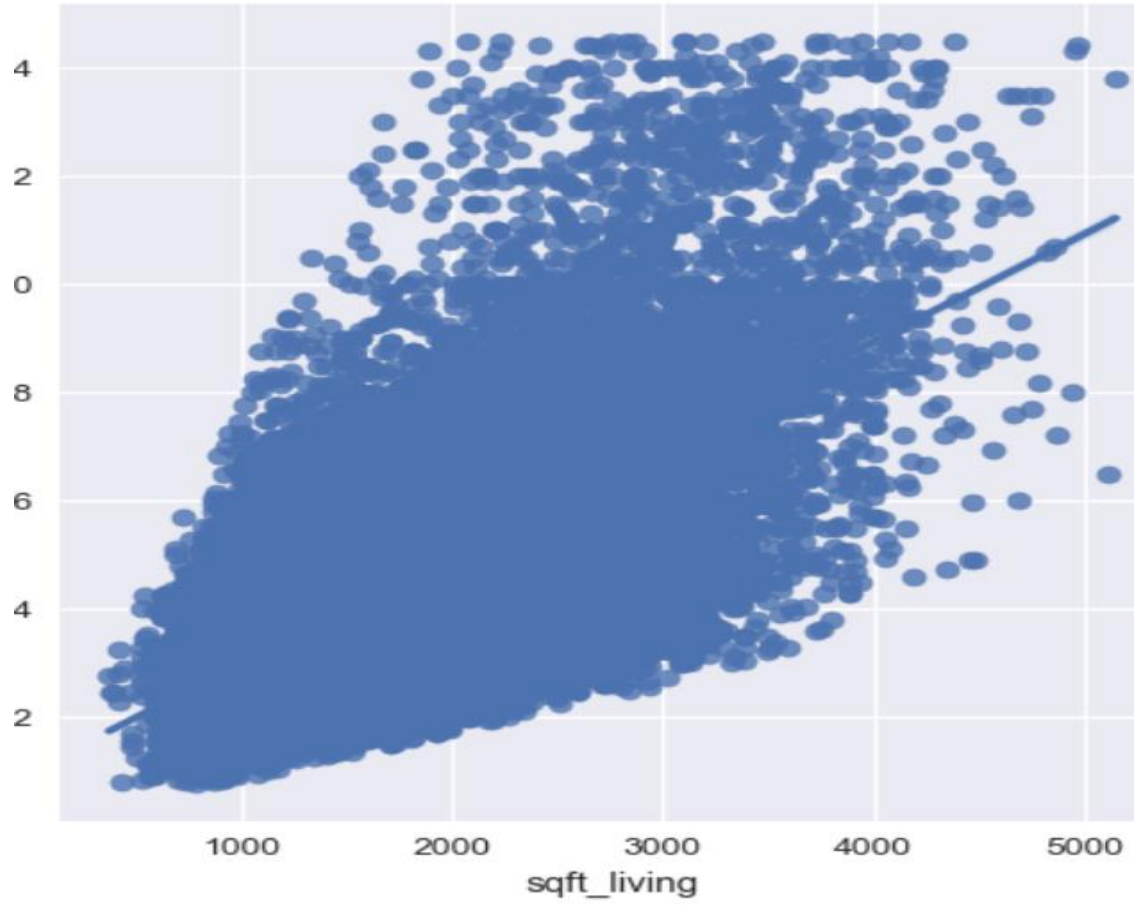
- Floors also seem to affect the price and 2.5 seems to be the most common.
- The grade is affecting the price increase.





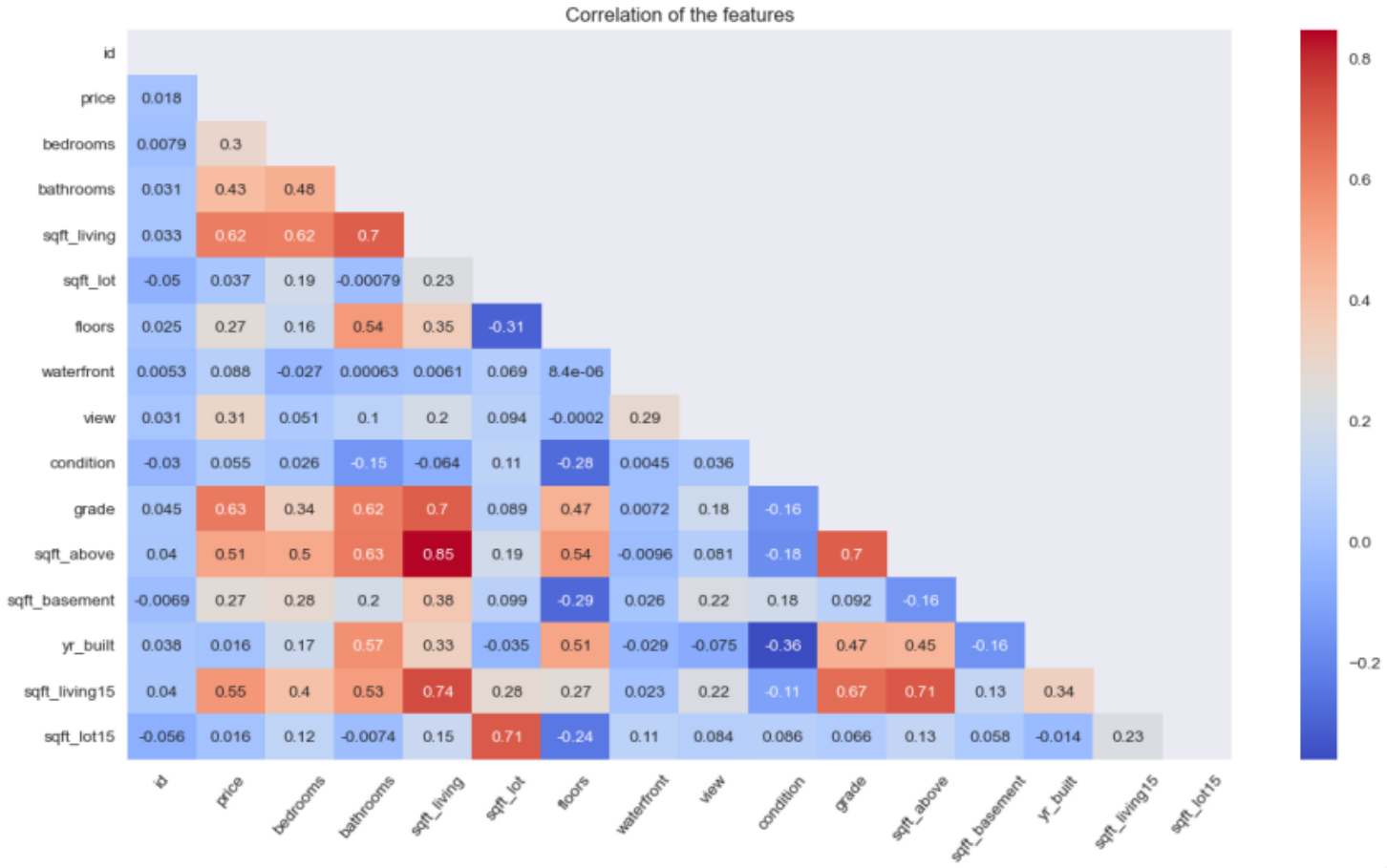
Square Foot Living

- There is a positive linear relationship between the square footage of living space and the housing price



Correlations and multicollinearity

- Warmer colors indicate high correlation and the cooler colors low correlation
- Multicollinearity: High correlation between two predictor variables



Linear Regression Modeling

- Selected features for modeling
  - Grade - overall grade given to the housing unit, based on King County grading system
  - Sqft_living – square footage of the home
  - Bathrooms – number of bathrooms
  - Bedrooms – number of the bedrooms in the home
  - Sqft_lot – square footage of the lot on which the home is built
  - Yr_built – Year the home was built

Multiple Linear Regression Model Results

Multiple Linear Regression Model Summary		
R-squared value: **0.582**	F-statistic: **4404**	P-value of F-statistic: **0.00**
Variable	**Coefficient value**	**P-value of Co-efficent**
Intercept	5.674e+06	0.00
Grade	-1.203e+05	0.00
Bathrooms	2.569e+04	0.00
Bedrooms	-2.228e+04	0.00
Sqft_living	127.79	0.00
Sqft_lot	-5.8597	0.00
Yr_built	-3188.98	0.00

Results

- A unit increase in square footage of living increases house price by $121.78 with all other variables held constant.
- A unit increase in square footage of the lot on which a house is built decreases house price by $ 5.85 with all other variables held constant
- Every additional bedroom decreases house price by $222,800 with all other variables held constant

Results

- Every additional bathroom increases house price by $25,690 with all other variables held constant
- A one-year increment in the year the house was built reduces house price by $3,188 holding all other variables constant.
- A unit increase in grade corresponds to a $120,300 increase in price, all other variables held constant.

Conclusion

- The number of bedrooms in a house and the housing grade have the largest influence on house sale price
- The number of bathrooms has a notable effect in the house sale price.
- Marginally, the year a house was built also has an influence on price
- The typical expensive house will have an average number of bedrooms, higher than average number of bathrooms, a high grading and a waterfront.

Recommendations

- For affordable housing in King County developers should:
  - Prioritize the construction of houses with an average grade rating to achieve a good balance between price and comfort.
  - Focus more on properties away from the waterfront where price tends to be very high.
  - Limit the number of bedrooms and bathrooms to the requirements of an average home buyer in King County

Further Analysis

- Investigate the impact of year renovated in relation to year built
- Investigate the cause of the price decrease with increase in the number of bedrooms
- Expand the dataset size to enhance model robustness

Thank You

Prepared by: Group 12