KING COUNTY HOUSES SALES DATA

Scope:

King County

1. Overview
2. Data Understanding
3. Modelling
4. Regression Results
5. Conclusions and Recommendations
6. Thank you

Overview

King County

King county is in Washington state a population of 2,269,675 according to the 2020 census, and it is projected to grow to 2,487,000 by 2025.

The project

The aim of this project as a data scientist is to report on the King County Houses Sales Data and create models that will be essential in predicting the price of houses.

Data Understanding

65%

of the data collected was used.

The data columns used were:

- Price
- Sqft_lot
- Sqft_living
- Sqft_above
- Bedrooms
- Bathrooms
- Condition
- Grade

Data Understanding

Price
This is the selling price of a house.

Sqft_lot
This is the square footage of the lot where a house is built on.

Sqft_living
Square footage of living space in the home

Sqft_above
Square footage of house apart from basement

Bedrooms
Number of bedrooms

Bathrooms
Number of bathrooms

Condition
How good the overall condition of the house is.

Grade
Overall grade of the house. Related to the construction and design of the house.

Exploratory Data Analysis (EDA)

Price

From the King County Houses Data:

$ 2,000,000

Maximum Price

From our analysis this was the maximum price of the houses in King County.

$ 82,000

Minimum Price

From our analysis this was the minimum price of the houses in King County.

$ 477,707

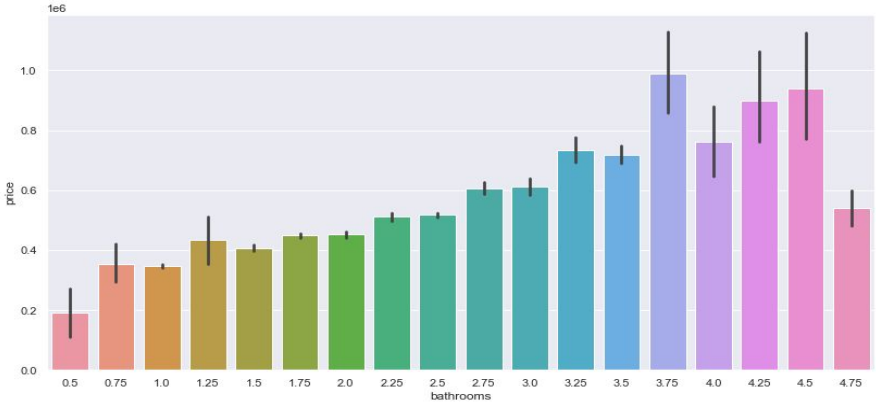
Average Price

From our analysis this was the average price of the houses in King County.

Bathrooms and Bedrooms(EDA)

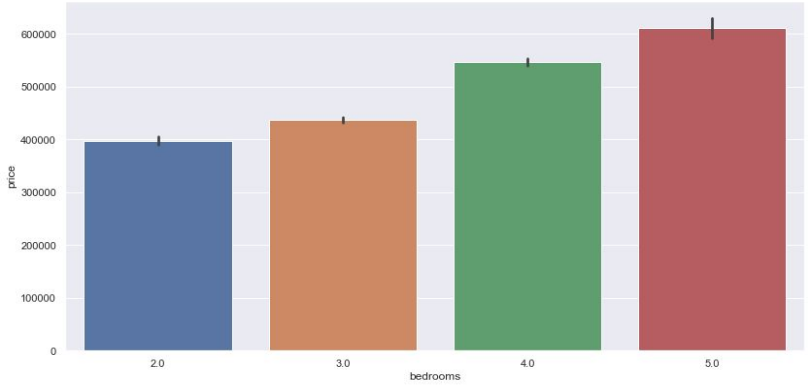
King County

Bathrooms



Most houses have 2.5 bathrooms.
From this analysis we can see that the houses with 3.75 bathrooms have the highest price.

Bedrooms

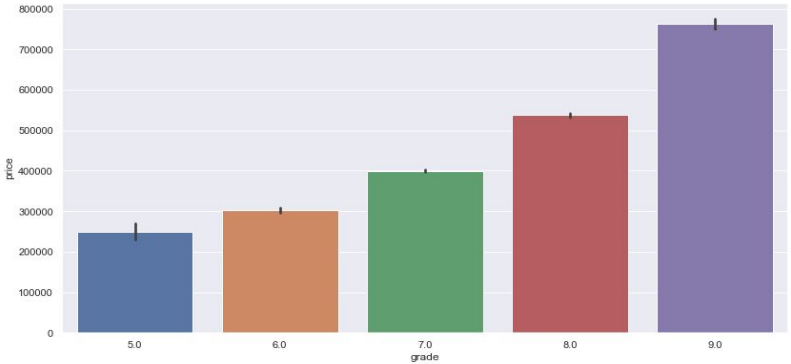


Most houses have 3 bedrooms.
From this analysis we can see that the houses with 5 bedrooms have the highest price.

Grade and Condition(EDA)

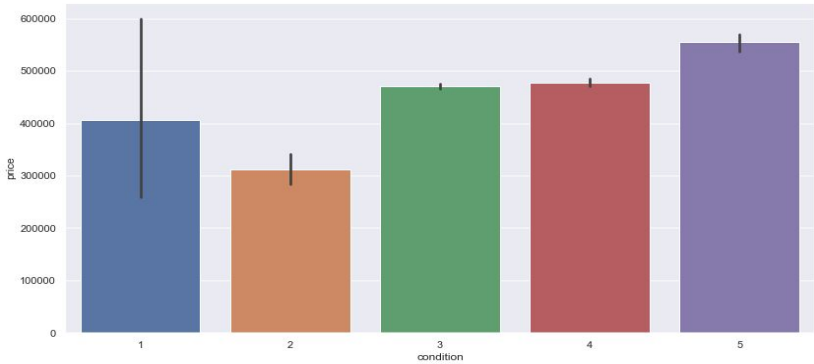
King County

Grade



Most houses were designed at grade 7.
The above analysis concludes that the higher a house grade the higher the price.

Condition



Most houses were of average condition.
Houses in a very good condition had the highest price according the above analysis.

Square Footage(EDA)

King County

10,236 Sq.ft	1,922 Sq.ft	13,987 Sq.ft
Lot area	**Living space area**	**Above ground area**
Average square footage of the lot.	Average square footage of living space in the home.	Average square footage of house apart from basement.

For each sqft_lot the price is $ 32.
For each sqft_living the price is $ 208.
For each sqft_above the price is $ 251.
Average price for each square footage $ 170.

Modelling

King County

Simple Linear Regression

We created a model to predict price using Sqft_above.

Simple Linear Regression

We created another model to predict price using Sqft_living.

Multiple Linear Regression

We created a model to predict price using sqft_lot + sqft_living + sqft_above + bathrooms + bedrooms + grade

Regression Results

Price Vs Sqft_above.

R-squared = 20%
Mean Absolute Error = $154,581

The model explains about 20% of the variability of price around its mean.

Since we are predicting house value based on prices , this model is significantly making relatively large errors.

Price Vs Sqft_living.

R-Squared = 30%
Mean Absolute Error = $142,318

The model explains about 30% of the variability of price around its mean.

Since we are predicting house value based on prices , this model is also significantly making relatively large errors.

Prive vs Sqft_lot + Sqft_living + Sqft_above + Bathrooms + Bedrooms + Grade.

R-Squared = 40%

Mean absolute error = $129,493

This is better as compared to the previous models. Our model is slowly getting better at predictions.

Conclusion & Recommendations.

Conclusion

After analyzing King County data, our final model would suggest the main factors in increasing property value are sqft foot living of the property as well as its grade. Linear regression was probably not the best tool to use to get the most out of this data set so in the future we would like to use different more powerful machine learning tools in order to make this a more accurate predictive model.

Recommendations

- The real estate agency should explore properties that occupy a large square foot of the land since most people would rather source a relatively large proportion of land
- As per our Exploratory Analysis we see that houses with good construction and design are likley to be more in demand which in turn generates more revenue from purchasing and renting, therefore the agency should check on property managers who are investing on the property architectural design
- The real estate agency should look into properties with 3 or more bedrooms since more people are likely to occupy them

Thanks!

Contact us:

DSC-FT4-1.2
P.O Box 28860 - 00100,
Nairobi.

contact@moringaschool.com