King County Housing Market Analysis

Group Members

- **Rose Kimondo**
- **Cynthia Atieno**
- **Richard Macharia**
- **Cynthia Njambi**

Group Members

- **Bestina Mutisya**

PHASE 2 GROUP 7

Overview

- In the King County (KC) housing market, multiple dynamics play a major role in shaping the prosperity of Real Estate companies.
- This project aims to help a Real Estate Development Company by providing a comprehensive analysis that focuses on the key factors that influence housing prices.
- By leveraging insights gained from the data from the King County House Sales dataset, we hope to advise the Real Estate Development Company on the most important characteristics/features to focus on when constructing homes, in order to generate more income and be as profitable as possible.

Business Problem

The Real Estate Development Company is venturing into the King County Housing market and would like to develop a sales strategy that makes it profitable by targeting houses that have the highest prices in the market.

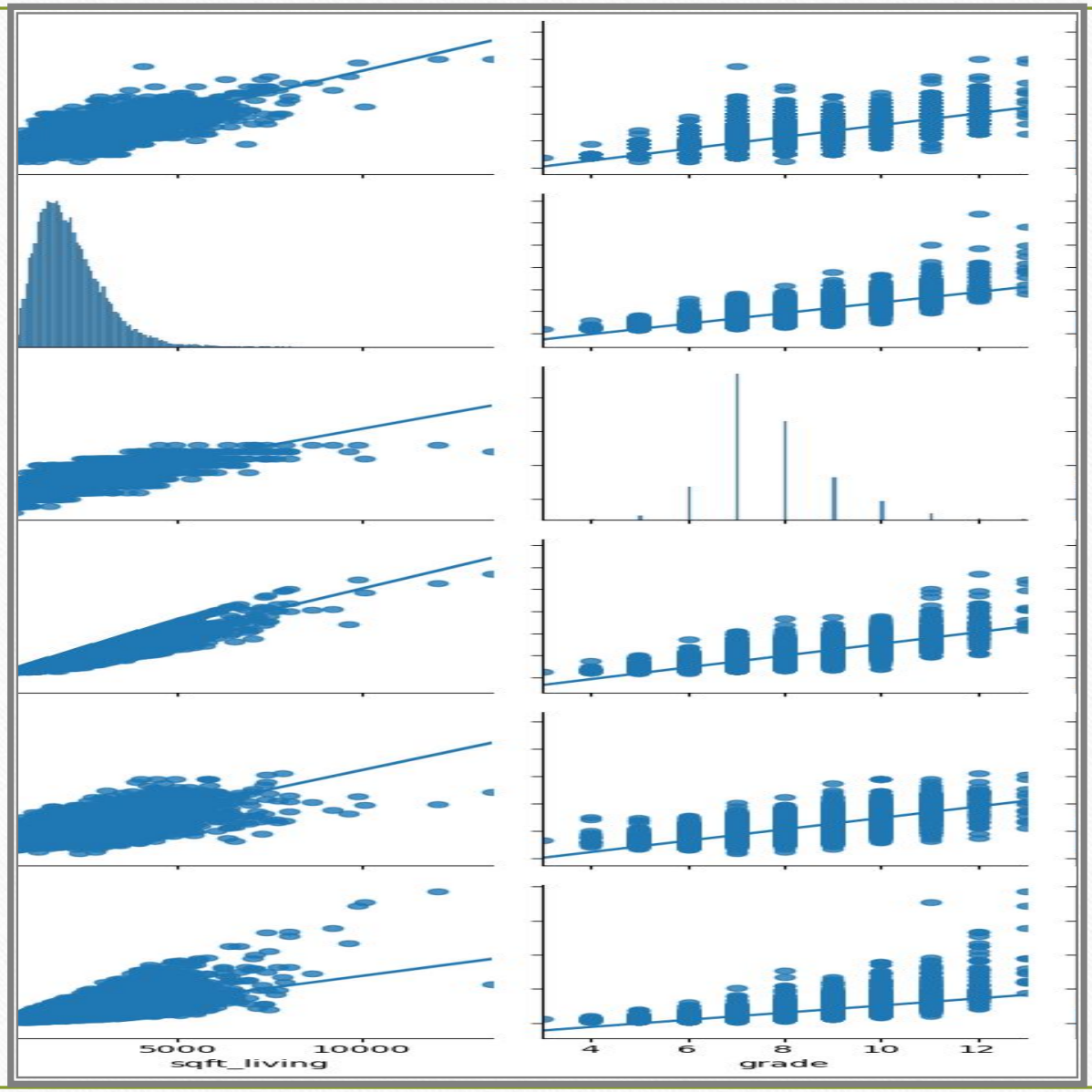
Data Understanding

- The King County House Sales dataset is central to this project, encompassing features pertinent to house sales in King County.
- The dataset is considered comprehensive for the project's aims, although additional features may be necessary to fully explain the relationships involved.
- However, the dataset has certain limitations that could impact the project's outcomes. Firstly, the data is not normally distributed, which might affect the model's reliability, especially at the extremes of the price distribution. To mitigate this issue, the data has been scaled. Additionally, some columns in the dataset contain missing values, which has been addressed to maintain data integrity and ensure accurate analysis.

Data Preparation

- Based on the correlation coefficients with price from the KC Housing Data dataset, the two features that have strong relationships with housing prices are:
  - Sqft_living – (Footage of the home) :- Sqft_living has a very strong positive correlation with price (0.7). This indicates that as the square footage of the home increases, its price increases.
  - Grade - Overall grade given to the housing unit, based on King County grading system :- Grade has a strong positive correlation with price (0.67). This indicates that as the grade of the home increases, its price increases.

Relationships between some selected features

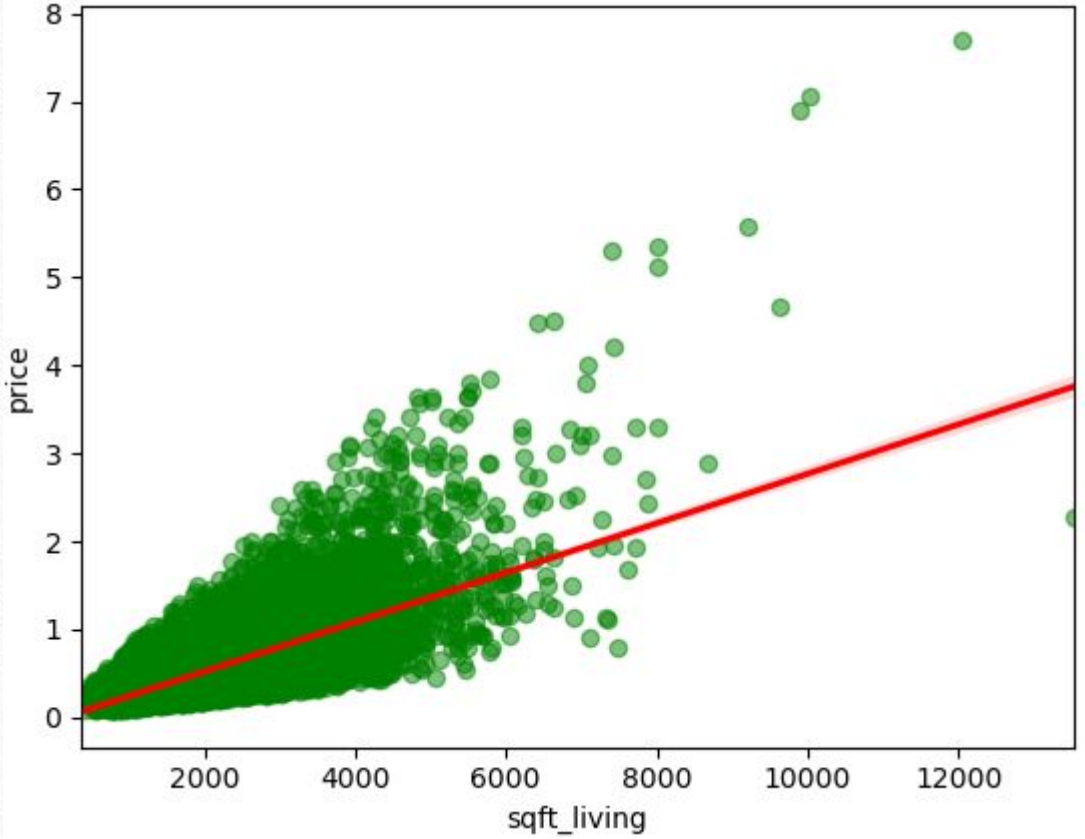


Data Analysis

Relationship between sqft_living and price

Trend Observation:
In the scatter plot with a regression line of square footage of the home (sqft_living) against price, you can observe a positive trend. As the average square footage of the home increases, the price of the house tends to increase as well.

Implication:
This suggests that houses with a higher square footage have higher prices in the King County housing market. They are priced higher.

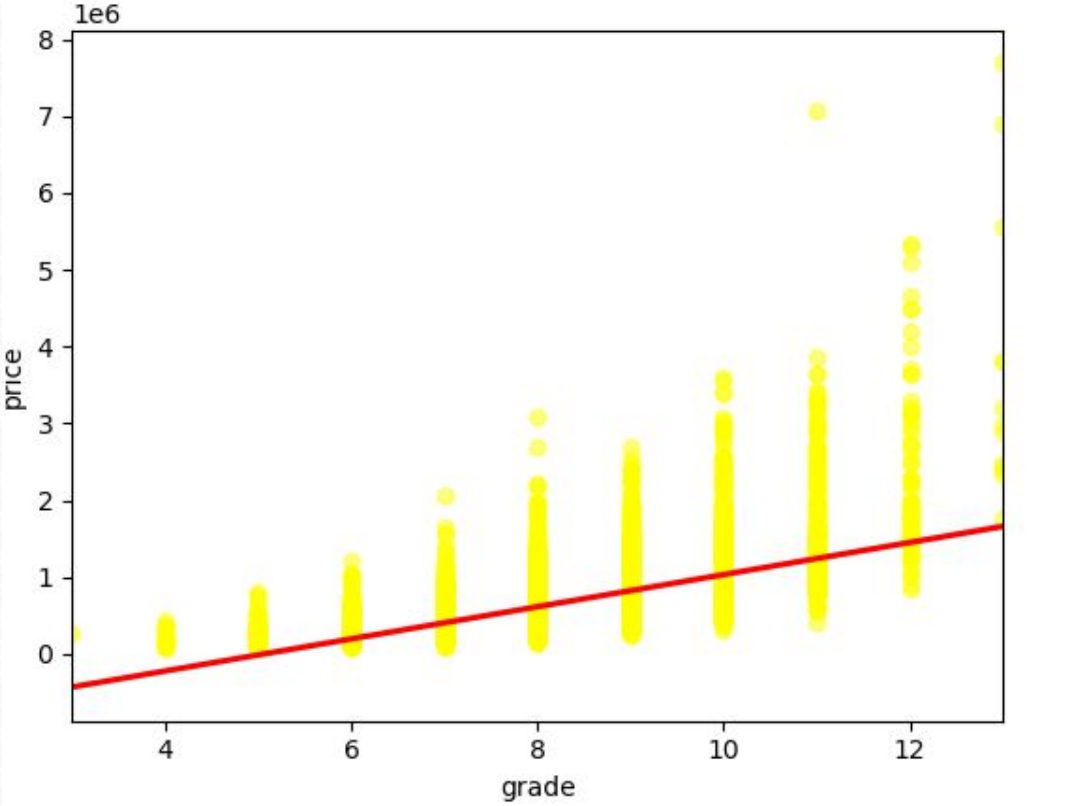


Data Analysis

Relationship between grade and price

Trend Observation:
In the scatter plot with a regression line of grade (overall grade given to the housing unit, based on King County grading system) against price, you can observe a positive trend. As the grade given to the home increases, the price of the house tends to increase as well.

Implication:
This suggests that houses with a higher grade have higher prices in the King County housing market. They are priced higher.



Model Summary

OLS Regression Results

Dep. Variable:	price	R-squared:	0.539
Model:	OLS	Adj. R-squared:	0.539
Method:	Least Squares	F-statistic:	5895.
Date:	Tue, 16 Jul 2024	Prob (F-statistic):	0.00
Time:	22:15:05	Log-Likelihood:	-2.0936e+05
No. Observations:	15117	AIC:	4.187e+05
Df Residuals:	15113	BIC:	4.188e+05
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-6.251e+05	1.6e+04	-39.181	0.000	-6.56e+05	-5.94e+05
bathrooms	-3.33e+04	4166.666	-7.991	0.000	-4.15e+04	-2.51e+04
sqft_living	197.0231	3.997	49.295	0.000	189.189	204.857
grade	1.079e+05	2746.353	39.286	0.000	1.03e+05	1.13e+05

Model Fit

```
mse 62009541655.03987
mae 161609.34538598044
R2 0.5323509874463058
Train Mean Squared Error: 62642802802.33004
Test Mean Squared Error: 62009541655.03987
Difference between Train and Test MSE is: 1.0 %)
```

The R squared value shows that our model explains 53.2% the relationship between our features and price based on the test data.

A mean absolute error of 161609 shows the predicted value is 161609 points away from the actual value, based on the test data.

However, the difference between the train and test mean squared error is now very little (1%). This indicates that we are not overfitting. This means that our model generalizes to new unseen data well.

Results

- Based on the regression results the features with strong relationships to sale prices are: - Sqft_living

  - Grade

- Their coefficients are:
  - Sqft_living: For each additional square foot of living space, the house prices increases by approximately 197,023
  - Grade: For each increase in grade, the house price increases by approximately 107, 900

Conclusion

- **Square Footage of Living Space (sqft_living)**: The coefficient suggests that for each additional square foot of living space, house prices are estimated to increase by $197.0231, holding other variables constant. This indicates that larger houses tend to command higher price
- **Grade of the House (grade)**: A higher grade is associated with higher house prices, with each unit increase in grade estimated to increase house prices by $107,900, holding other variables constant. Grades typically reflect the quality and features of the house, influencing buyer preferences and willingness to pay.

Conclusion

- The overall model has an R-squared value of 0.539, indicating that approximately 53.9% of the variance in house prices can be explained by the predictors (sqft_living, grade, bathrooms) included in the model. This suggests that while these predictors are significant, there are other factors not captured by the model that also influence house prices.
- The F-statistic is very high (5895) with a low p-value (0.00), indicating that the model is statistically significant. This implies that at least one of the predictors is significantly related to house prices, reinforcing the reliability of the model's predictions.
- Contrary to initial expectations, the coefficient suggests that each additional bathroom is associated with a decrease in house prices by $-33,300 , holding other variables constant. This unexpected finding may warrant further investigation into potential factors influencing this relationship.

Recommendations

- **Target Larger Houses:**
  - Since the model indicates that house price increases with the square footage of living space, the company should focus on acquiring or developing larger houses to maximize potential sales prices.
- **Focus on High-Grade Houses:**
  - Houses with higher grades are significantly associated with higher prices. The company should aim to develop or invest in properties with higher quality finishes, better construction, and more desirable features.
- **Use Predictive Analytics for Targeted Marketing:**
  - By using the model to predict high-priced houses, the company can create targeted marketing strategies to attract high-end buyers. This can involve highlighting the features that are most strongly associated with higher prices, such as larger living spaces and higher grades.