

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front parallelogram is blue and the back one is a light green. Both are oriented diagonally, with their top-left corners pointing towards the top-left of the slide.

# **King County Houses Multiple Regression Analysis**



# Business Overview

- The housing demand in Washington State is high, with Seattle being the leading city in terms of population.
- As such, Tella Real Estate Agency, based in King County, has undertaken a research to find out the best performing metrics when it comes to house sale prices determination.
- By identifying the key factors that affect the sale prices of the houses, the agency can develop more effective marketing strategies, help sellers get the right buyers and make better investment decisions.



# Business Problem

- The agency has a dataset of home features collected from the various house sales in King County.
- The agency wants to understand the relationship between these features and the sale house prices of houses in King County, and particularly, which features most affect the prices.



# Objectives

- To understand the relationship between the various features and house sale prices.
- To determine the peak of house sale season.
- To build a multiple linear regression model that identifies the most important features that influence the sale price of a house.
- To use the model to gain insights into the home sales market and improve decision-making processes.

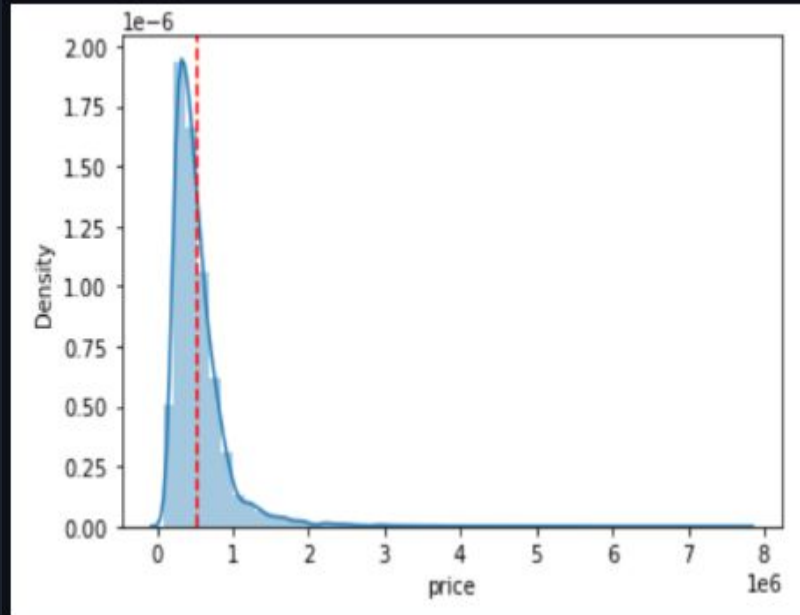


# Data Understanding

- I used King County Houses data for analysis.
- The dataset has 21 columns translating to 21 observed features and 21,597 rows representing the number of house sales between May 2014 and May 2015.

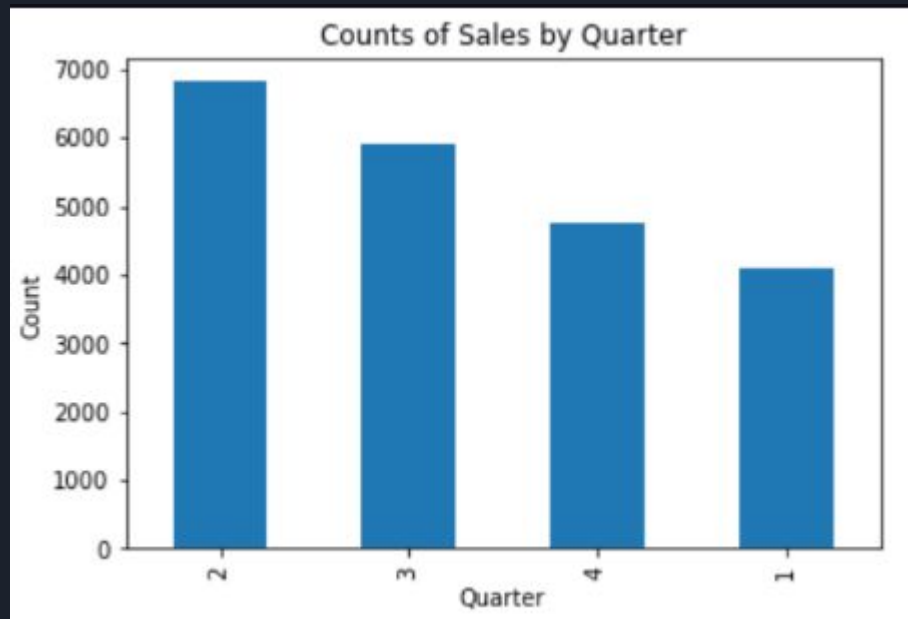
# Price Distribution

- This indicates that most houses are low-priced.



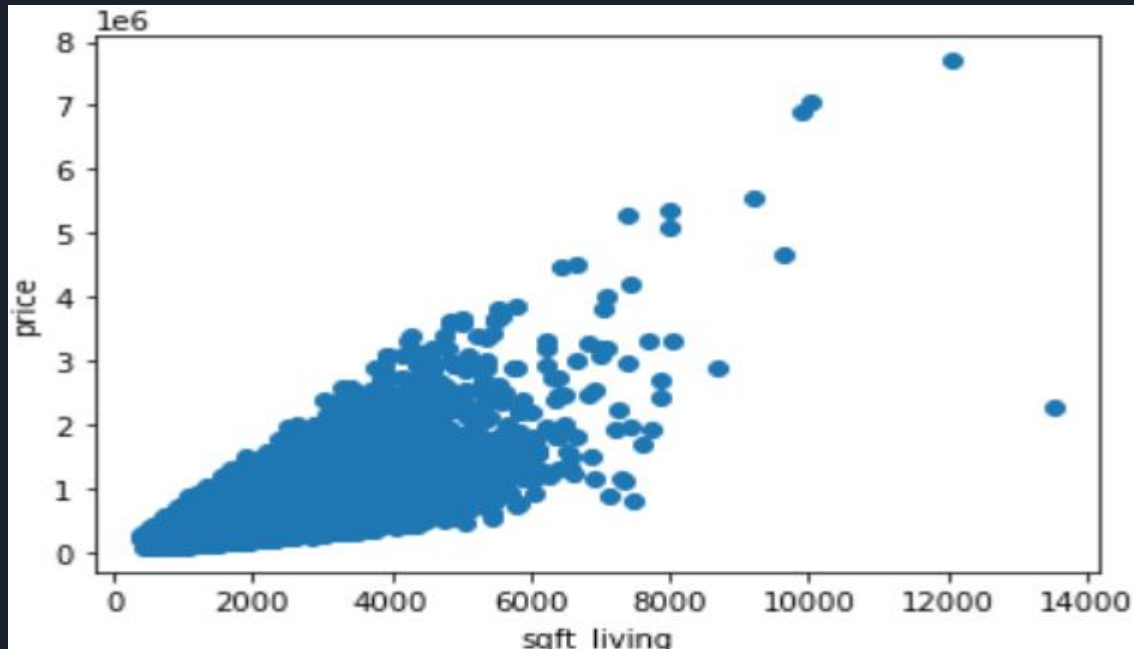
# Peak House Sale Season

- Most house sales are made in the Spring Season.



# Baseline Model

- The baseline model is a linear regression of price against square foot living.
- The model could predict 49.3% of the prices, so I went ahead to examine the other models that would predict better price.







# Best performing model

- The best model explained 61.7% of house prices.
- The predictor variables in the model were:

Bedrooms, sqft\_living, grade, sqft\_basement,  
yr\_renovated, age, waterfront, view, condition,  
sqft\_lot, sqft\_lot15



# Conclusion

- The variables that have a major influence on the price of the house are; square foot living, age of the house, condition of the house, if the house is on a waterfront and has an excellent view.
- For those looking for economical housing options, it might be wise to consider sacrificing spacious living quarters or arsenic waterfront view.
- The variables that have the least influence on the price of the house are; grade, number of bedrooms, sqft lot, sqft basement and sqft lot15.



# Recommendation

- Renovate their houses since this increases the value of the house.
- Ensure that the houses are in good condition before putting it into the market for sale.
- Increase square footage of living space.
- Put up their houses for sale in peak season which is the Spring Season.



# Thank You

Betshua Kerubo

Github: <https://github.com/BetshuaK/Phase-2>