

Phase 2 Project:  
Business Case Study using  
Data Science

Dennis Mwanzia  
Ian Macharia  
Samuel Igecha  
Pamela Otieno  
Pauline Ndung'u

# BACKGROUND

King County is a region in Washington state that includes Seattle and is known for its strong economy, diverse population, and natural beauty. Major employers include technology companies, healthcare providers, financial institutions, and government agencies. The real estate market in King County is highly competitive due to strong demand, limited supply, and high levels of income and wealth among buyers, with median home prices above the national average and increasing steadily in recent years.





THE OBJECTIVE OF THIS PROJECT IS TO  
APPLY OUR NEWLY ACQUIRED DATA  
SCIENCE SKILLS TO FORMULATE AND  
SOLVE A REAL BUSINESS PROBLEM.



WE HAVE BEEN PROVIDED WITH A RAW  
DATASET CONSISTING OF HOUSE SALES  
PRICES IN NORTHWESTERN COUNTY.



TASK IS TO FORMULATE A REAL BUSINESS  
CASE STUDY, ANALYZE THE DATA, AND  
PROVIDE VALUABLE INSIGHTS TO AID IN  
MAKING INFORMED BUSINESS DECISIONS.

## Business Understanding



The aim of this project is to provide useful business insights to a hypothetical client, Kings Wajuzi Developers who is interested in investing in the real estate market of Northwestern County.



Our goal is to analyze the provided dataset and formulate a business problem related to this domain, which can be solved using data science techniques.



By leveraging our expertise, we aim to provide valuable recommendations that can help our client in making informed decisions regarding their investment in the local real estate market.

# OBJECTIVES



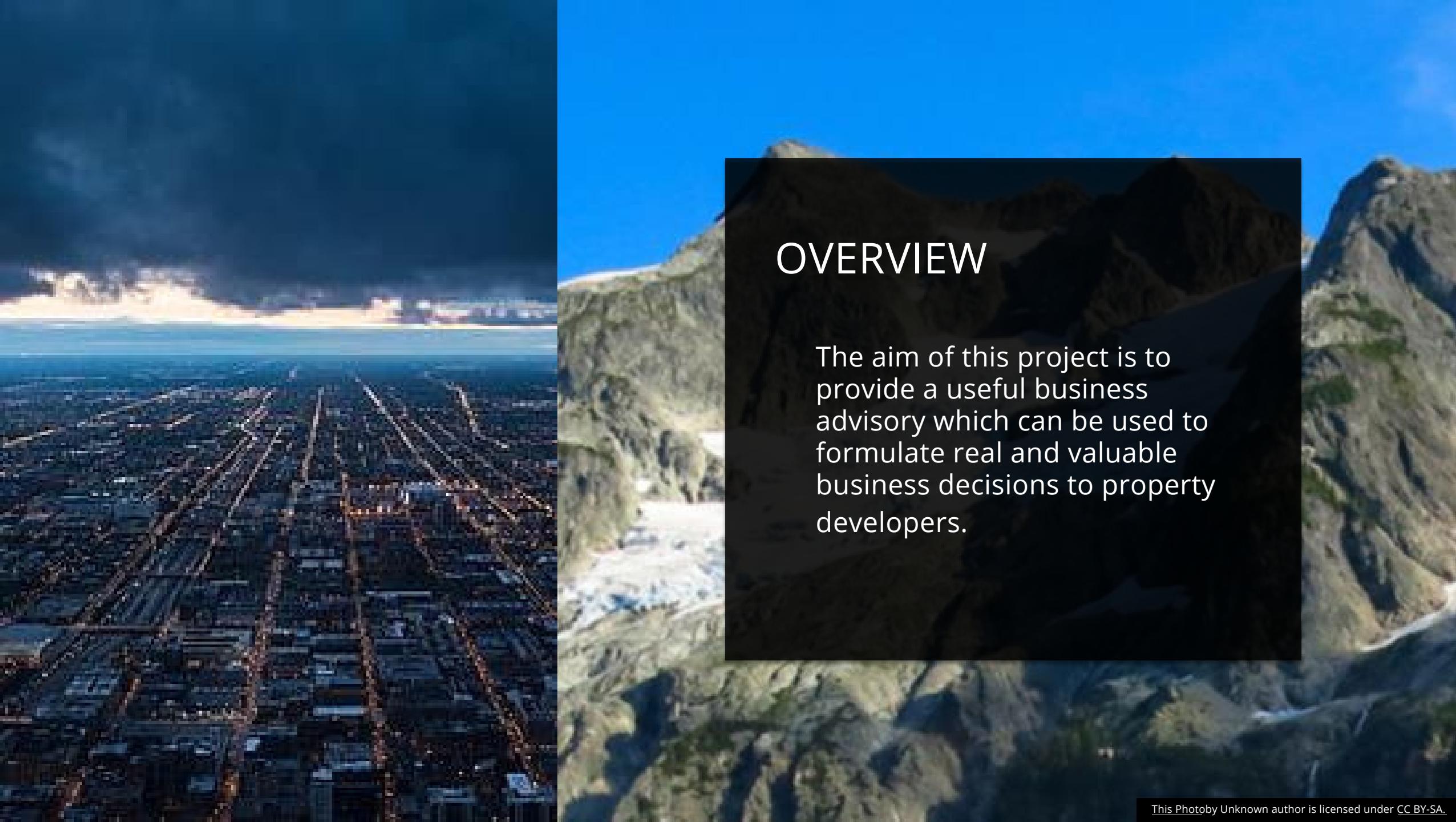
To determine the best season that favors sale of prices of homes.



To determine which locations are the most popular to home owners.



To determine which house features home owners look out for the most.



## OVERVIEW

The aim of this project is to provide a useful business advisory which can be used to formulate real and valuable business decisions to property developers.

The background of the slide is a nighttime photograph of a city skyline, likely San Francisco, featuring the Golden Gate Bridge and various lit-up buildings.

# TOOLS

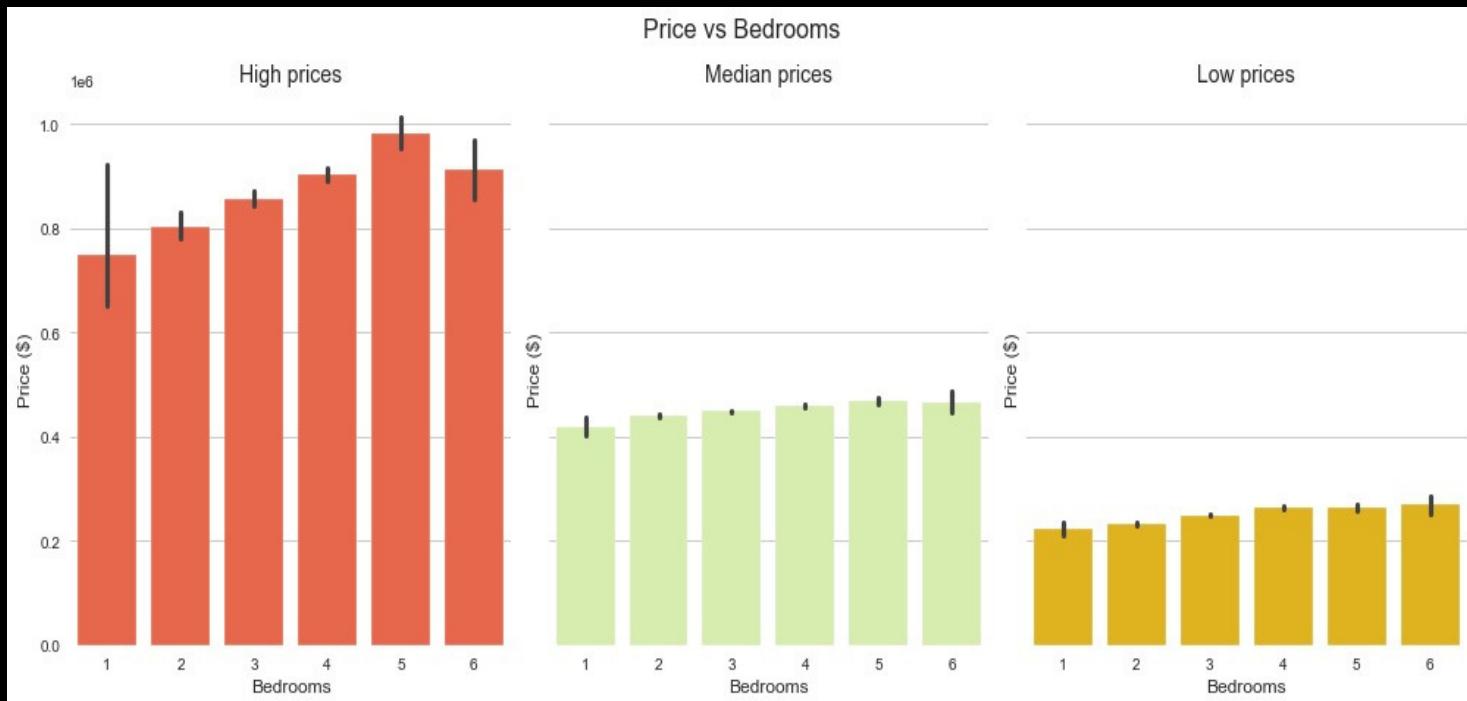
- Python
- Python Libraries

## DATA SOURCE:

- Kaggle.com

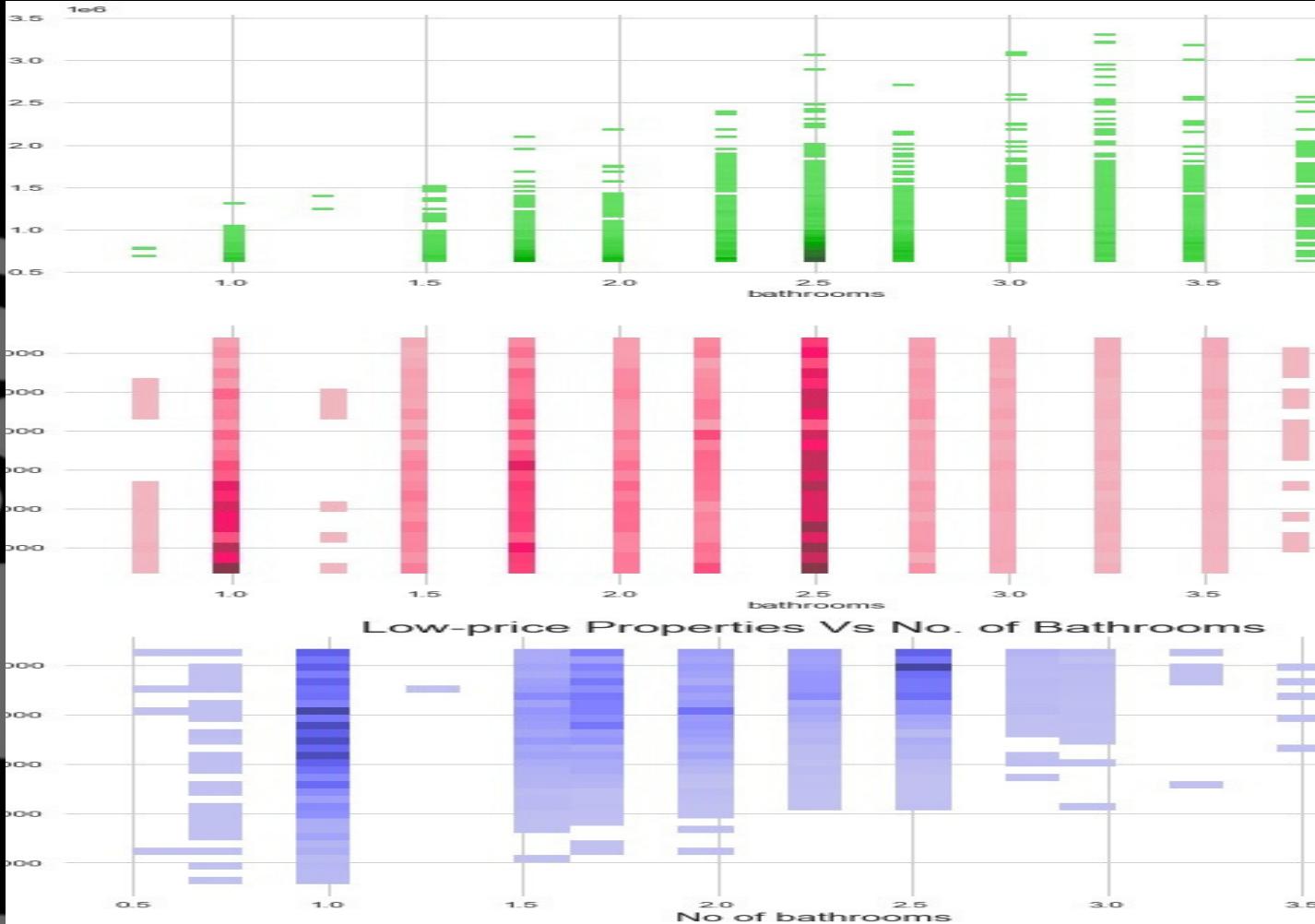
# DATA VISUALIZATION

# 1. How do different priced houses relate with the number of bedrooms?



Here, we can see that the larger the number of bedrooms, the higher the prices as well. For average and low-priced houses, there isn't much of an increase but more of constant with a slight increase in prices with the number of bedrooms.

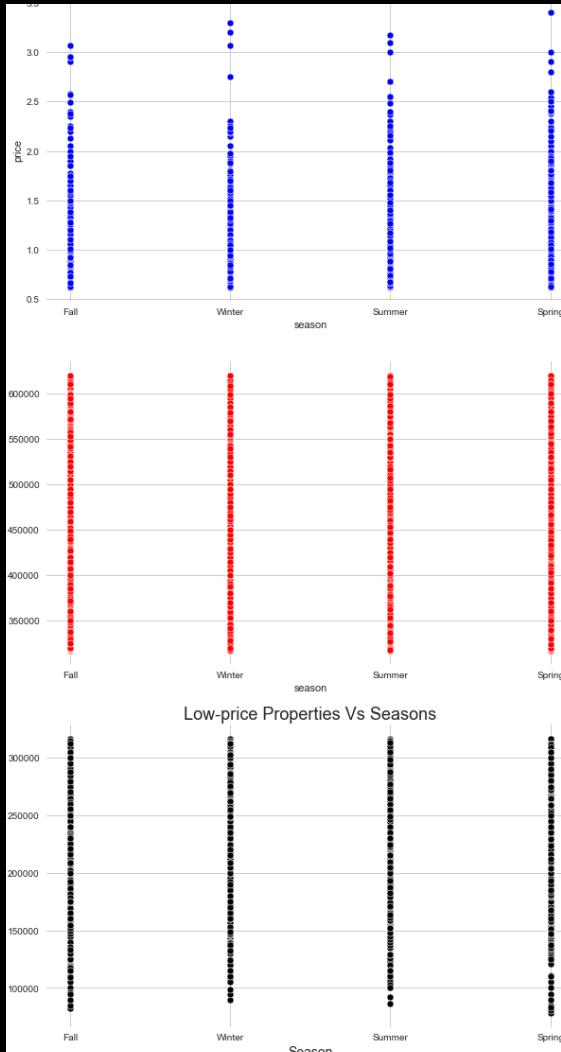
## 2. How do different priced houses relate with the number of bathrooms?



we observed that the majority of median and high-end properties had a higher density at two bathrooms with an additional bathtub, toilet or sink. From the high end we also noticed that the density was more towards the median showing that there is little difference that distinguished them from the median. For the low price range the density was more on one bedoomed houses indicating a prevalence of studio apartments.

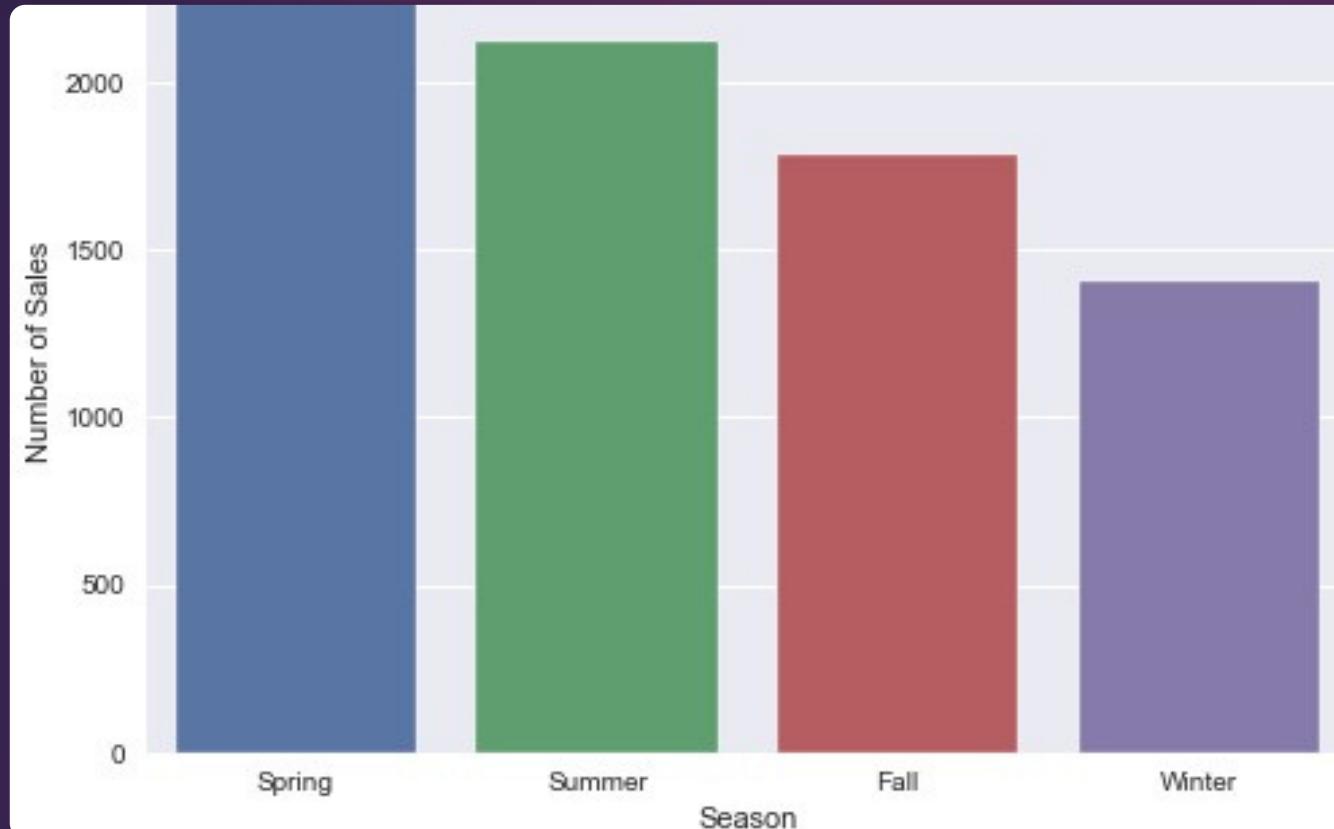
The histograms suggest that the number of bathrooms is positively correlated with the price of properties. However, we need to perform statistical analysis to confirm this relationship and identify the strength of the relationship.

### 3. How do different-priced houses relate with the seasons



- Based on this visualization, the scatter plots show the relationship between the seasons and prices of properties . From the count of properties sold in each season, it appears that spring & summer are the most popular seasons for homebuying while fall & winter have comparatively fewer sales.
- The scatter plots of high end , medium and low price properties against the season show that prices of properties are generally consistent across all seasons. However there's a slight increase in the number of high end properties sold in the spring and a decrease in the number of low-priced properties sold in the winter
- This pattern may be due to various factors, including the perceived desirability of certain seasons of buying and selling homes, the availability of listings during different seasons , or even the preferences of buyers and sellers e.g. the spring season may be associated with better weather conditions & more opportunities for homeviewing which may increase demand for high end properties. Conversely the winter season might be associated with lower sales of low-priced properties due to holiday expenses & colder weather.

## How do Seasons affect sales of Properties?



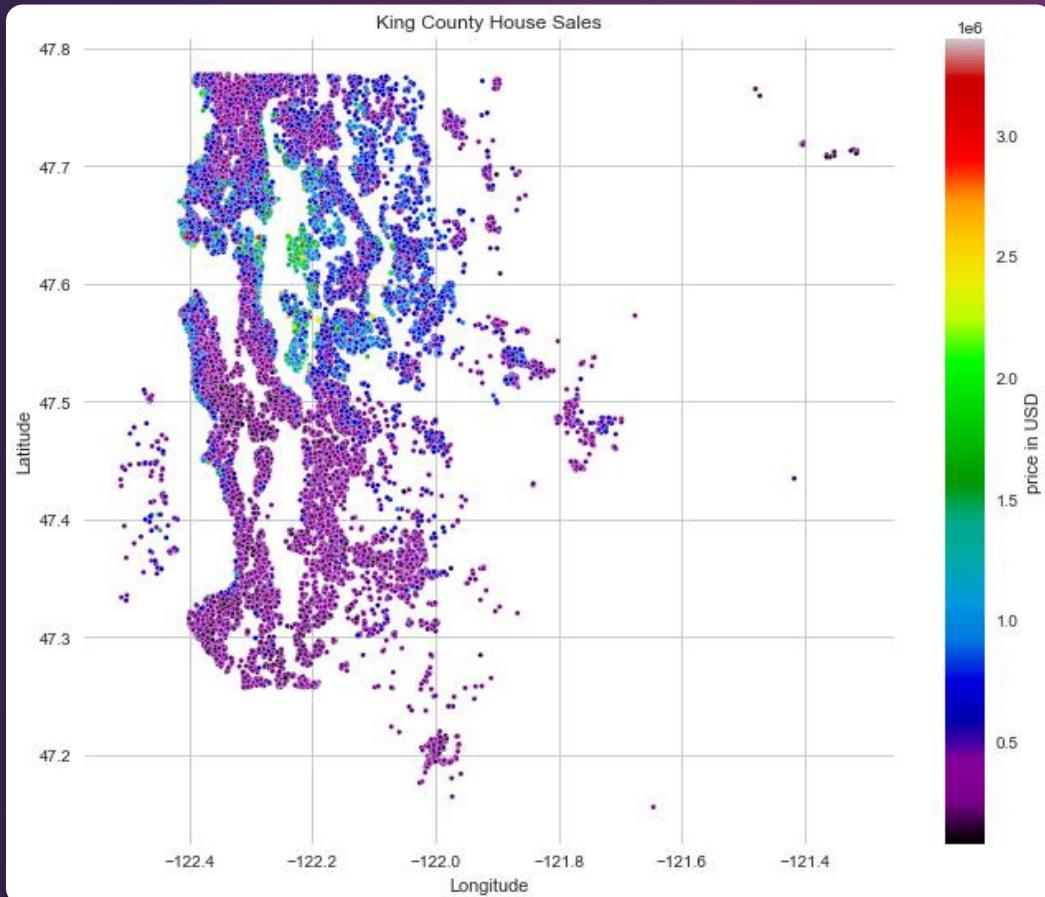
Looking at the sales data, we can see that the spring months of March, April, and May are the best for selling houses with a total of 6518 sales. This is likely due to the fact that Spring is a time of renewal and growth, and people may be more willing to make big changes such as buying a new home during this time. May is the most popular month for selling homes within Spring, with 2414 sales.

Summer, which comprises the months of June, July, and August, had a slightly lower total sales of 6328 compared to spring. July had the highest sales with 2211 followed by June with 2178 and August with 1939.

The fall season of September, October, and November had a total of 5056 sales, with October having the highest sales of 1876 followed by September with 1771 and November with 1409.

Lastly, the winter season of December, January, and February had the lowest total sales of 3695. December had the highest sales with 1470 followed by February with 1247 and January with 978. This may be due to the fact that Winter is a time when people tend to stay indoors and may be less inclined to go through the hassle of buying or selling a home.

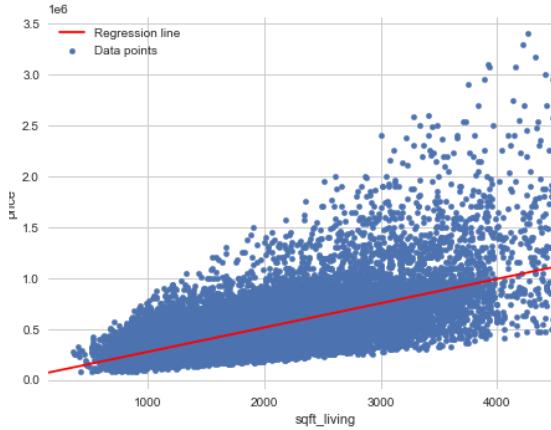
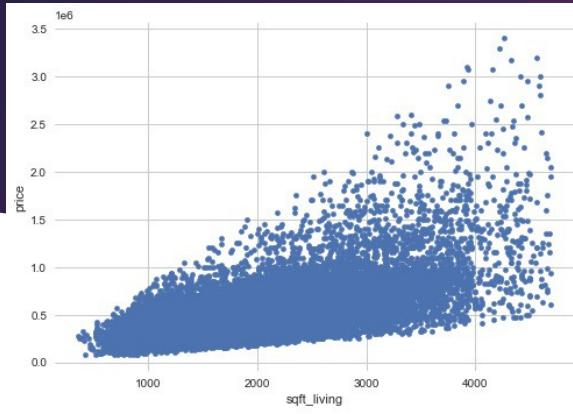
# How does Location affect the Sale Price of a house?



From this visualisation we can already draw some insights on houses' geographical locations. The highest house prices are concentrated in the areas with a latitude of 47.6 & longitude of around -122.25. There is a disparity with southern locations achieving lower house prices.



From this visualisation we can already draw some insights on how houses in different zipccodes affect prices of housing. Some zipcodes have highly priced houses and vice versa



There was strong linear relationship between the square footage of the living space of homes & price. With this in mind we built a baseline model and with our findings, we concluded that with the increase in square footage of living space, so does the increase in prices of homes. Take for example a house that has a living space of 1000sqft, we would expect the price to be about \$273,826

## Baseline Model Generation

# RECOMMENDATIONS

Based on all the data analysis and modelling we did we came up with the following recommendations;

## CRecommendations and conclusion

### Objective 1: Season of the year.

Sales of properties is highest in spring and lowest in winter. Property developers such as our client Kings Wajenzi developers should target to develop properties that will be market in the season of spring and Summer.

### Objective 2: Location

Location within King County is also a crucial factor, with significant price disparities among zip codes. The location with zip code 98112 has the most priced properties with 98001 having the least priced properties. We suggest that developers targeting high end clients should develop homes in zip codes 98112 , 98109 and 98105. Those targeting low income earners should target to develop properties in zip codes 98001, 98106 and 98108. Those targeting middle income earners should target other zip codes.

### Objective 3: House features

The most important predictor of house price is view. We observed that house prices with waterfront view are the most desirable. Houses that have a waterfront view have median price that is almost twice as those that dont have waterfront view. We suggest that developers target to develop homes in front of water bodies such as lakes and rivers as these will fetch highest prices.

### Bedrooms

Number of bedrooms is the second most important predictor of house prices. Properties with higher number of bedrooms fetch high prices and hence developers should target high number of bedrooms where they wish to fetch high price for properties. However, incremeament beyond 6 bedrooms does not guarantee highest prices. The optimal number of bedrooms is 5 bedroom houses which fetch highest prices.

### Bathrooms

Number of bathrooms is not an important predictor of houses. Both high and medium priced properties have an optimal number of bathrooms which is 2.5. Low priced properties have 1 bathroom as the optimal. Therefore, developers should ensure high and medium priced properties have an optimum of 2.5 bathrooms.

### Grade

The median house price increases with grade, suggesting a positive correlation. We believe that grade is a reliable indicator of price, so we recommend focusing on homes with a grade of 10 or higher.

### Square Footage and Condition

Square footage and the condition of the property are also significant factors that affect the price of a house and should be considered by developers when developing their properties.

### Further Analysis

Our model had shortcomings initially as it only explained less than 50% but after improvements, it was able to predict almost 100% of variations in the response variable. This is because of proper data understanding, cleanup and fine tuning it. However, to gain new insights, we can consider additional factors such proximity to the capital city of Washington State, the income per zip code, presence of schools, hotels, commuter trains, and entertainment facilities. Longer time span should also be considered as the data used here was for 1 year only.

# THANK YOU