

# King County Housing Market



Real Estate Trends and Insights

## Main Project Objective:

The project aims to provide a deeper understanding of the factors that influence housing prices in the area.

- Empower agents to make informed decisions
- Offer guidance on pricing strategies,
- highlight crucial features that drive property values
- optimize pricing strategies
- identify ideal locations and potential investment opportunities.



# Overview

The King County House Sales dataset is a comprehensive collection of real estate data from the King County area, encompassing information on residential properties sold over a specific time period. It consists of a range of variables, including price, location, size, condition, amenities, and etc. hence offering a detailed perspective on the housing market in King County

Our analysis utilizes multiple linear regression to identify the significant factors that contribute to house prices in the King County area.

# Business Problem Understanding

---

The project aims to provide a deeper understanding of the factors that influence housing prices in the area. By Determining the Factors affecting the House Market in King County, we will help real estate agents to achieve the following:



Pricing  
Strategy



Market  
Analysis



Investment  
Decisions



Risk  
Assessment



Marketing and  
Sales



Policy  
Development

# Data Understanding

---

## Data Sources

- We used a CSV data file called [kc\\_house\\_data.csv](#)
- We complemented our data source with a bit of online research

## Data preparation and cleaning

Our dataset has 21597 rows and 21 columns.

- There are 3 data types in the dataset: Floats, Integer, & strings
- In the dataset contains 6 columns in we categorical(string) data types while 15 variables were numerical.
- Since we needed most of the columns in numerical form for linear regression and feature engineering we ended up changing most of our data into numerical form.
- In our initial analysis of the data we realized that we had a bit of missing data so we dropped some of the rows to avoid losing data.
- We also realized that 86 of our columns were duplicates which we also dropped.

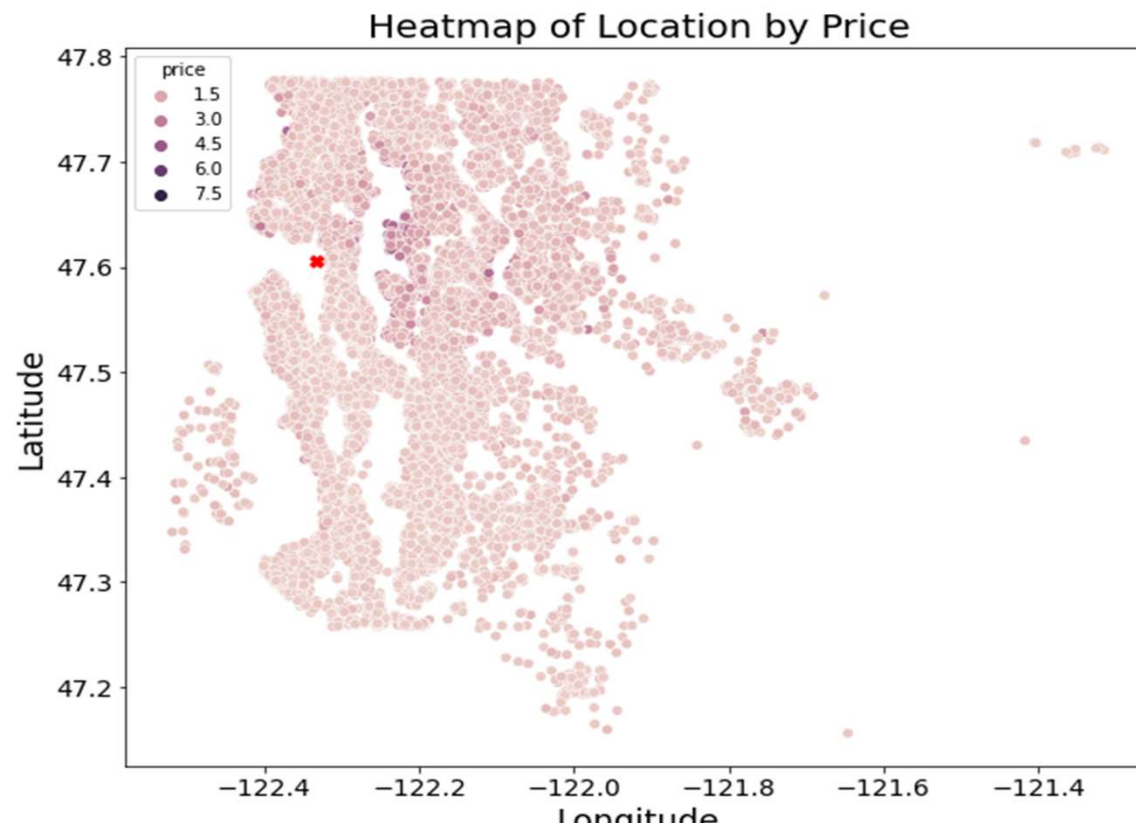
# Data Understanding

## Initial Insights

During our initial data analysis we realized that :

- ❖ That the Average price of a house in King county is \$540,296
  - The maximum price is \$7,700,000 and the minimum price is \$78,000
- ❖ The average number of bedrooms of a house in King County is 3 bedrooms
  - The maximum number of bedrooms is 33 and the minimum is 1
- ❖ The average number of bathrooms is 2
  - The maximum number of bathrooms is 8 and the minimum number of bathrooms is 0.5
- ❖ The average number of floors in house in King county is 1.49
  - The highest having 3.5 floors and the lowest having 1

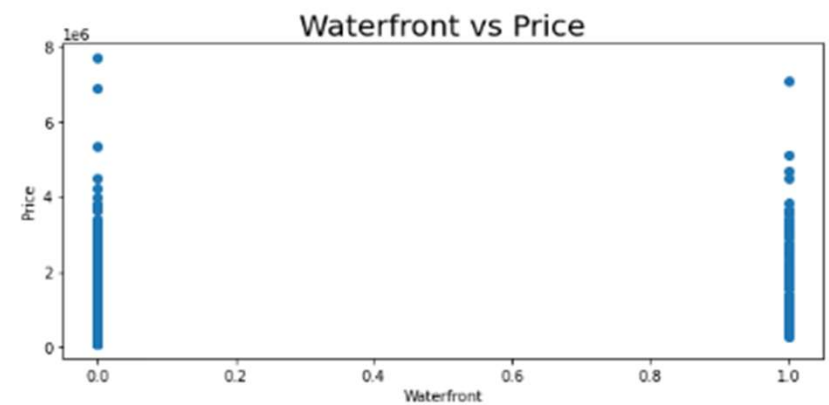
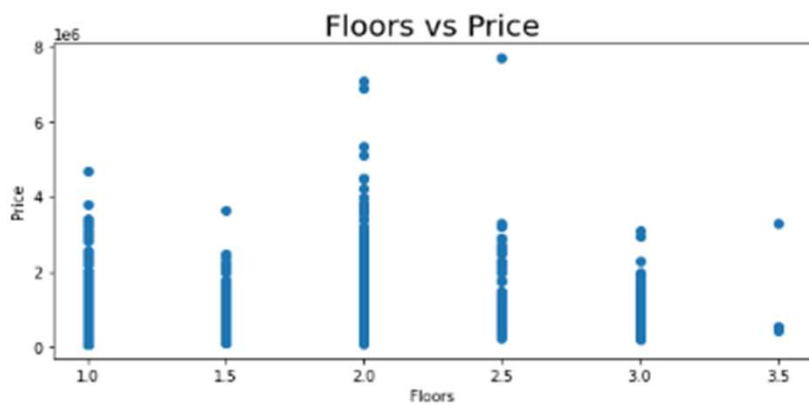
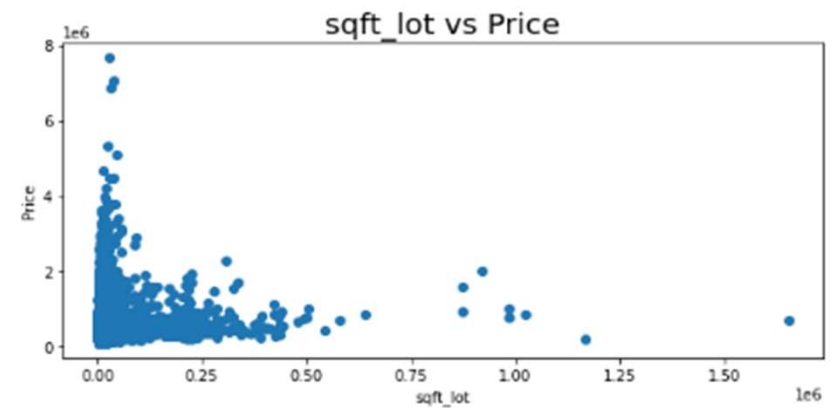
This is a heat map of price by location using longitude and latitude





# EDA

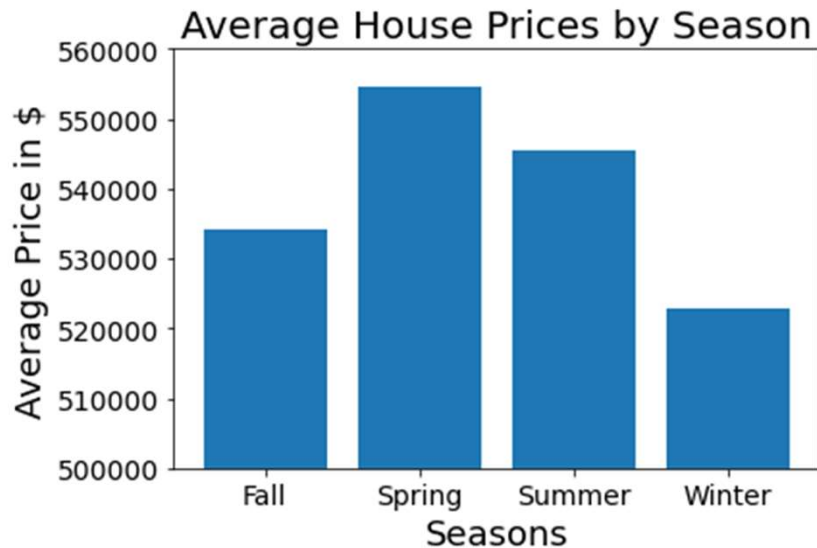
In this step we checked the correlation between the independent variable and the dependent variables.



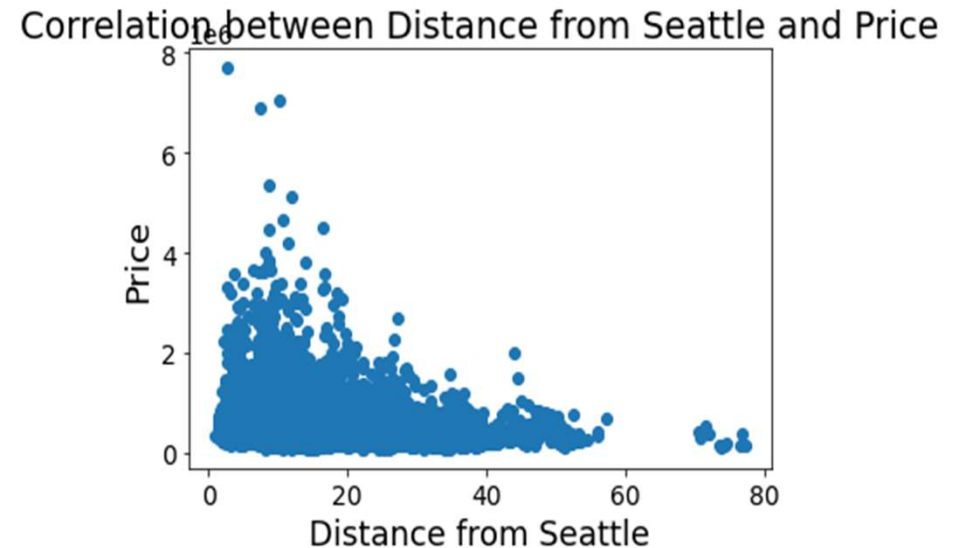


# Feature Engineering

We used the date column to feature engineer seasons and then calculated the average price per season.



We picked the coordinates of seattle and calculated the distance from the city centre using longitude and latitude



## Linear regression and Modelling

```

=====
                        OLS Regression Results
=====
Dep. Variable:          log(price)    R-squared:                0.679
Model:                  OLS          Adj. R-squared:            0.679
Method:                 Least Squares  F-statistic:              3601.
Date:                   Fri, 02 Jun 2023  Prob (F-statistic):      0.00
Time:                   19:52:15       Log-Likelihood:           -3157.7
No. Observations:       15301         AIC:                     6335.
Df Residuals:           15291         BIC:                     6412.
Df Model:               9
Covariance Type:        nonrobust
=====
                        coef    std err          t      P>|t|      [0.025    0.975]
-----
const                9.7495      0.140     69.692     0.000      9.475    10.024
bedrooms            -0.0689      0.004    -18.906     0.000     -0.076    -0.062
sqft_living          0.0002     9.16e-06    26.697     0.000      0.000      0.000
floors               0.0510      0.006      8.493     0.000      0.039      0.063
distance_from_seattle -0.0193      0.000    -75.776     0.000     -0.020    -0.019
Age_of_house         0.0023      0.000      8.181     0.000      0.002      0.003
waterfront_YES       0.5604      0.028     19.903     0.000      0.505      0.616
season_Spring         0.0455      0.005      8.684     0.000      0.035      0.056
log(sqft_living)      0.4569      0.021     21.781     0.000      0.416      0.498
log(Age_of_house)    -0.0761      0.013     -5.927     0.000     -0.101    -0.051
=====
Omnibus:              168.115    Durbin-Watson:           1.993
Prob(Omnibus):         0.000    Jarque-Bera (JB):        211.374
Skew:                  -0.178    Prob(JB):                1.26e-46
Kurtosis:              3.453    Cond. No.:               1.34e+05
=====

```

Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 1.34e+05. This might indicate that there are strong multicollinearity or other numerical problems.

# Linear Regression and Modelling

- 1.Distance from Seattle: Houses farther from Seattle are cheaper.
  - 2.Bedrooms: More bedrooms increase the house price.
  - 3.Space/Land: Larger houses or properties have higher prices.
  - 4.Seasonal Variation: Spring has the highest house prices. Sell in spring/summer, buy in fall/winter.
  - 5.Square Footage of Living Space: Larger living space leads to higher house prices.
  - 6.Age of the House: Older houses tend to be more expensive, possibly due to historical or architectural value.
- Note: Market conditions and specific property features can also impact prices

# RECOMMENDATIONS:

Below are the areas that we feel that the real estate agents will be able to assist their buyers and sellers with.

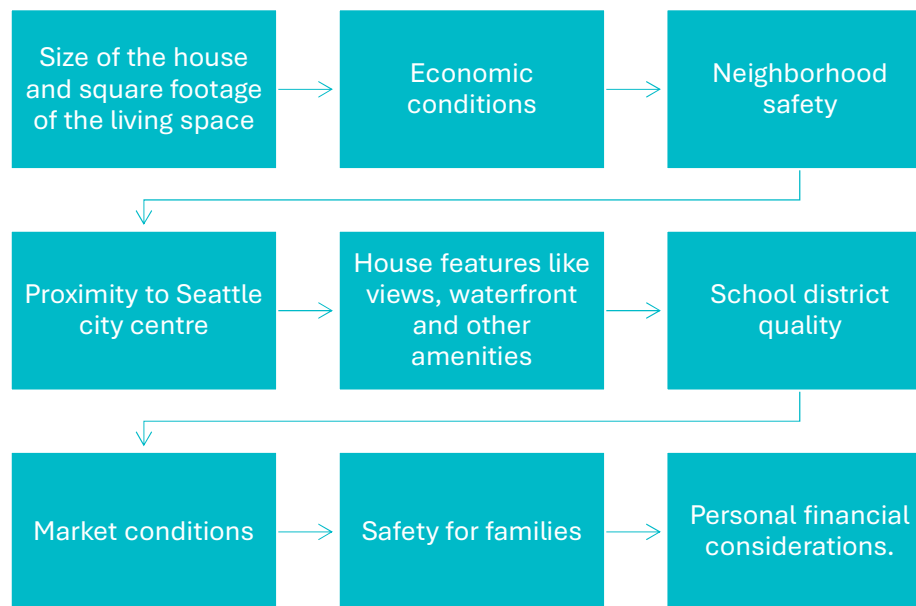
## Seller:

- We recommend renovating the houses before putting them up for sale
- We also recommend that the houses are put up for sale in the spring or summer months. This will help you get the best prices.

## Buyers:

- In this case we recommend buying in the winter and fall months in order to get the best prices(cheaper by nearly \$30,000).
- For buyers who are buying the houses for sale the following are the amenities to consider: The living space, the lot space, the school district,

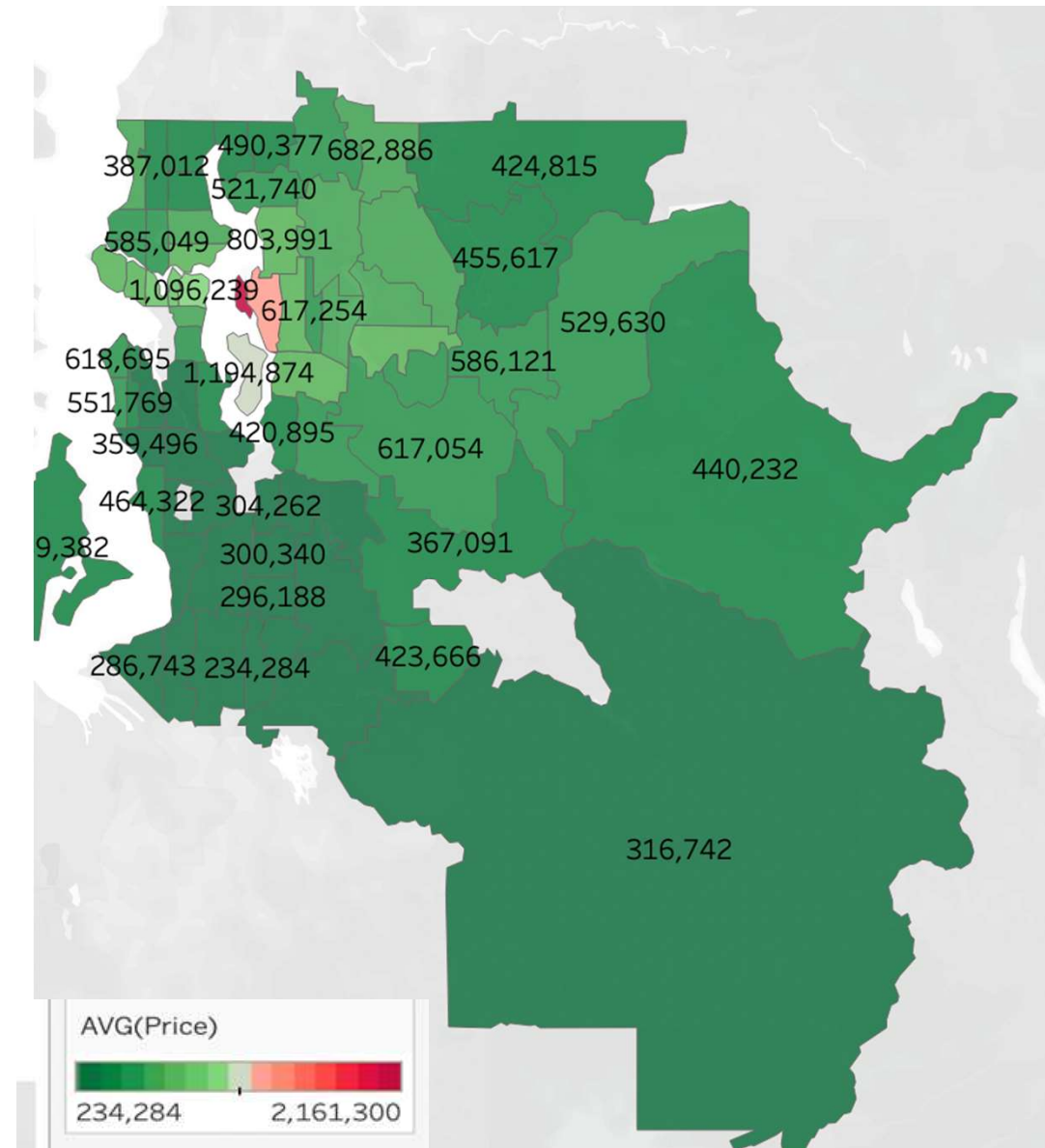
When examining the factors that influence house prices in King County, several key aspects come into play. The factors below are likely to greatly influence the Pricing, the buyers' Choices and the Timing of their house purchases in King County.



## What's Next

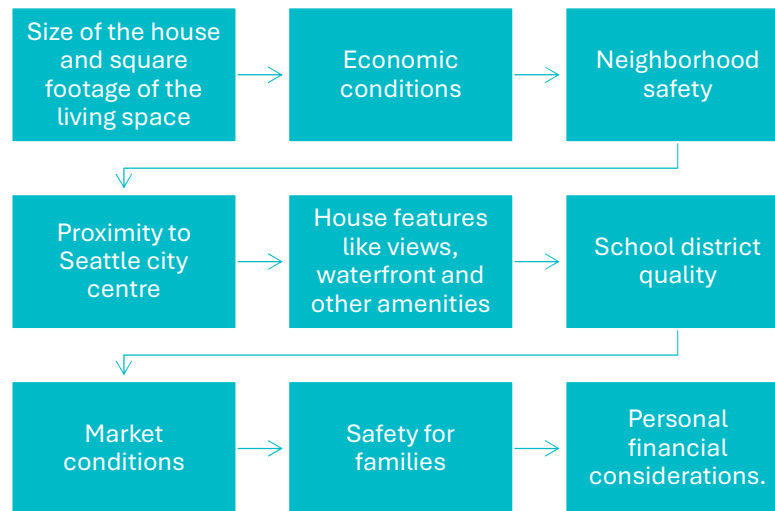
We carried out a simple tableau visualization, this allowed us to see:

- That the houses get gradually more expensive the closer they are to the city centre but that area also has some of the best schools so next is to research the correlation of the zip codes to the housing prices.



# What's Next

1. The above tableau analysis left us wanting to do further analyses of:



2. Explore other data sources.



# References

<https://www.niche.com/places-to-live/search/best-zip-codes-for-families/c/king-county-wa/>

<https://github.com/WatiriVivian/Phase2-DataScience-Project>