

KING COUNTY REAL ESTATE ANALYSIS

GROUP 3





Our Team

CRYSTAL WANJIRU

SCRUM

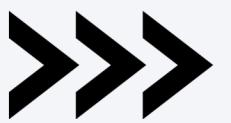
RISANUEL OBONDO

TIMOTHY LENKU

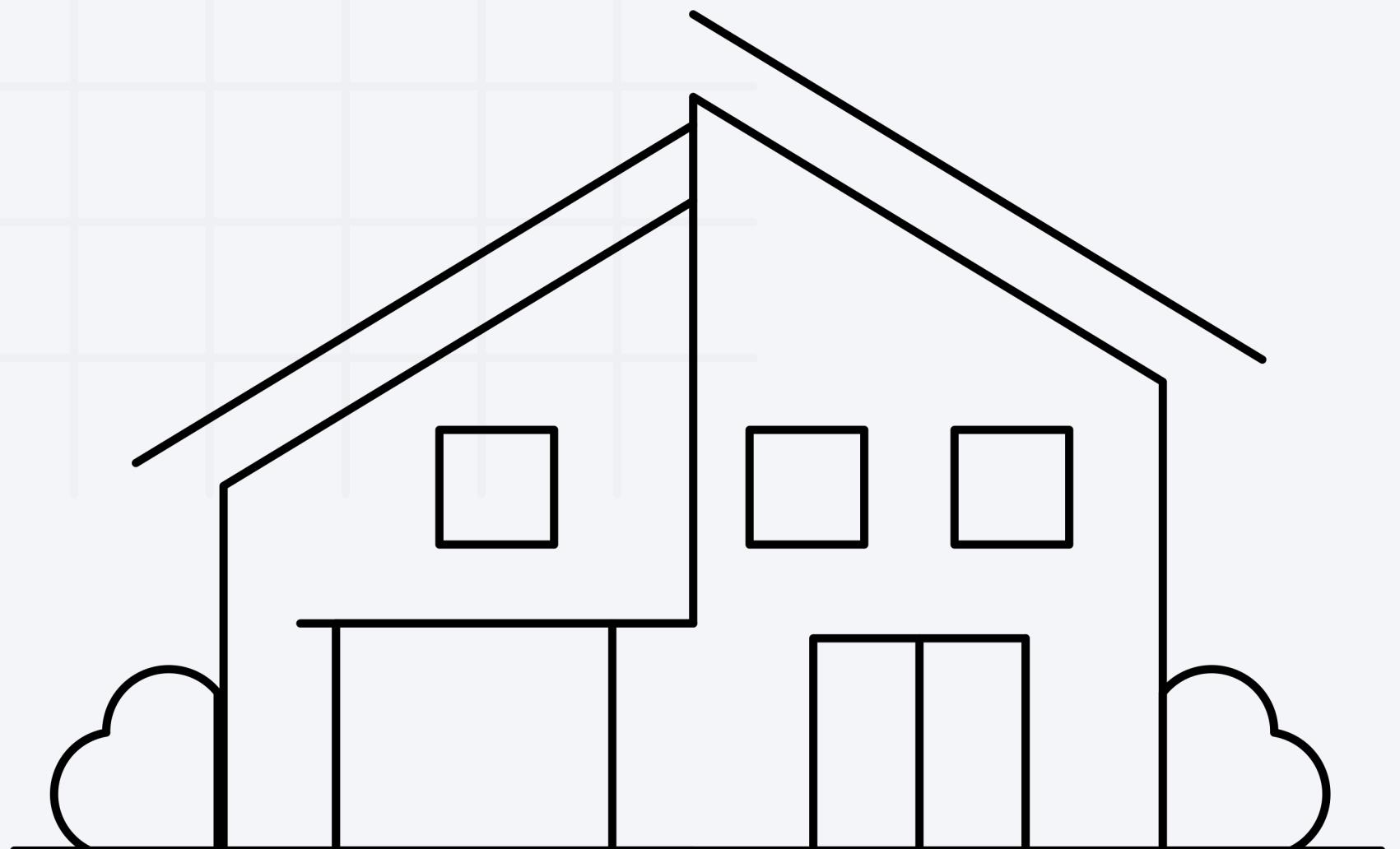
WILSON MUTUNGU

LAUREEN CHEPKOECH

VICTORINE IMBUHILA



INTRODUCTION



King County, Washington, is a real estate powerhouse, pivotal to the region's economic growth. Influenced by location appeal, property features, and market trends, the diverse housing market in King County thrives due to a robust economy, tech industry presence, and its scenic beauty, offering a range of housing options, attracting both homebuyers and investors.



PROBLEM STATEMENT

The booming King County real estate market lacks clear, actionable insights into the specific factors influencing property prices, hindering the ability of investors to make informed decisions.



Our aim is to offer or provide homeowners and real estate professionals with dependable insights for navigating the ever-changing King County real estate market with confidence.



OBJECTIVES

01

To determine the main features that affect the value of a home. See how specific features relate to the housing price

03

To quantify the effect of the above features on the value of a home.

02

To check if there is a difference in the price of the renovated houses and those that were not renovated

04

To Develop a model to predict the home value.



Data understanding

This project uses King County
house sales dataset



Data Composition

This data contains information about the houses in the northwestern county. This is found in kc_house_data.csv.

Data Analysis

We use Pandas and NumPy for data cleaning and analysis, Tableau, seaborn and Matplotlib for data visualization

Modelling

We use Scikit Learn together with StatsModels for creating the machine learning model



DATA CLEANING

01

02

03

DUPLICATES

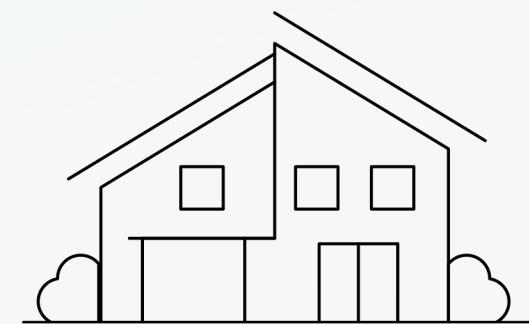
We drop the entries with duplicated IDs. This is to avoid redundancy that might skew results which can lead to inaccurate analysis.

MISSING RENOVATION VALUES

The missing values indicate that particular houses have never been renovated. Therefore, we fill the null values with zeros to accurately reflect the absence of any renovation for those properties.

MISSING WATERFRONT VALUES

We fill the missing value according to the zip code. We fill the missing values with the mode after grouping by zip code and the data remains representative of the general trend.





Data Analysis

Key Factors Affecting House Prices

1. Positive Correlations

- Bathrooms
- Living Spaces
- Grades
- Ground Space

2. Multicollinearity Alert

There is multicollinearity between features

-e.g. High Correlation Between sqft Living and sqft Above

	Correlation Matrix														
id	1	-0.019	0.0017	0.0041	-0.014	-0.13	0.018	-0.024	0.0066	-0.012	0.0049	0.021	-0.011	-0.011	0.0029
price	-0.019	1	0.32	0.53	0.7	0.09	0.26	0.034	0.67	0.61	0.32	0.053	0.12	0.12	0.18
bedrooms	0.0017	0.32	1	0.53	0.59	0.034	0.18	0.023	0.37	0.49	0.3	0.16	0.018	0.018	0.16
bathrooms	0.0041	0.53	0.53	1	0.76	0.088	0.5	-0.13	0.67	0.69	0.28	0.51	0.047	0.047	0.16
sqft_living	-0.014	0.7	0.59	0.76	1	0.17	0.35	-0.062	0.76	0.88	0.43	0.32	0.051	0.05	0.2
sqft_lot	-0.13	0.09	0.034	0.088	0.17	1	-0.00580	0.0087	0.11	0.18	0.015	0.052	0.00510	0.0052	-0.035
floors	0.018	0.26	0.18	0.5	0.35	-0.0058	1	-0.27	0.46	0.52	-0.24	0.49	0.00270	0.0026	-0.25
condition	-0.024	0.034	0.023	-0.13	-0.062	-0.0087	-0.27	1	-0.15	-0.16	0.17	-0.36	-0.056	-0.056	0.13
grade	0.0066	0.67	0.37	0.67	0.76	0.11	0.46	-0.15	1	0.76	0.16	0.45	0.015	0.015	0.049
sqft_above	-0.012	0.61	0.49	0.69	0.88	0.18	0.52	-0.16	0.76	1	-0.053	0.42	0.02	0.02	-0.21
sqft_basement	-0.0049	0.32	0.3	0.28	0.43	0.015	-0.24	0.17	0.16	-0.053	1	-0.13	0.066	0.066	0.82
yr_builtin	0.021	0.053	0.16	0.51	0.32	0.052	0.49	-0.36	0.45	0.42	-0.13	1	-0.2	-0.2	-0.16
yr_renovated	-0.011	0.12	0.018	0.047	0.051	0.00510	0.0027	-0.056	0.015	0.02	0.066	-0.2	1	1	0.045
renovated	-0.011	0.12	0.018	0.047	0.05	0.00520	0.0026	-0.056	0.015	0.02	0.066	-0.2	1	1	0.045
basement	0.0029	0.18	0.16	0.16	0.2	-0.035	-0.25	0.13	0.049	-0.21	0.82	-0.16	0.045	0.045	1
	id	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	condition	grade	sqft_above	sqft_basement	yr_builtin	yr_renovated	renovated	basement

Data Analysis

Square Feet Living Area vs. House Price

- Positive correlation observed.
- Larger living area means a higher house price.
- Highlights the relationship between living space and price.





Modelling



Baseline model

Statistically significant model.

Adjusted R squared - 51%

RMSE = \$264,562

Multiple Regression Model

Statistically significant model

Adjusted R squared = 60%

RMSE = \$239,783.

Polynomial Regression

Statistically significant model

Adjusted R squared = 64%

RMSE = \$225,788

Recursive Feature Elimination

Statistically significant model

Adjusted R squared = 61%

RMSE = \$237,989



Recommendations

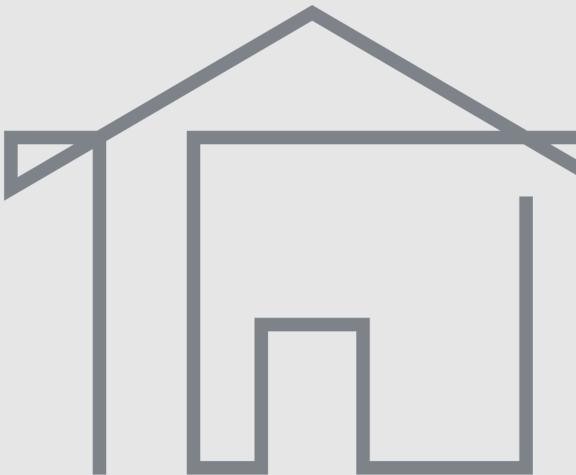


01

1. Renovations Command High Prices- Renovated houses see a predicted price increase of \$129,800 compared to non-renovated ones.

02

Start with Fair/Good Condition Properties -Focus on properties in fair or good condition for maximum improvement potential.



03

Feature Enhancement: Invest in improving key features of the properties like adding a basement or floor.

04

Reconsider the pricing strategy for houses with a higher number of bedrooms. Additional bedrooms lead to lower prices



Next steps



01 Employ more advanced techniques

To investigate variation of house price, it requires a more advanced modeling techniques such as regularization



02 Model Deployment

For practical use, whether for predictions, decision support, or other applications. Ensure that the model is integrated into your workflow or system as needed.



Thank You



QUESTIONS

>>>