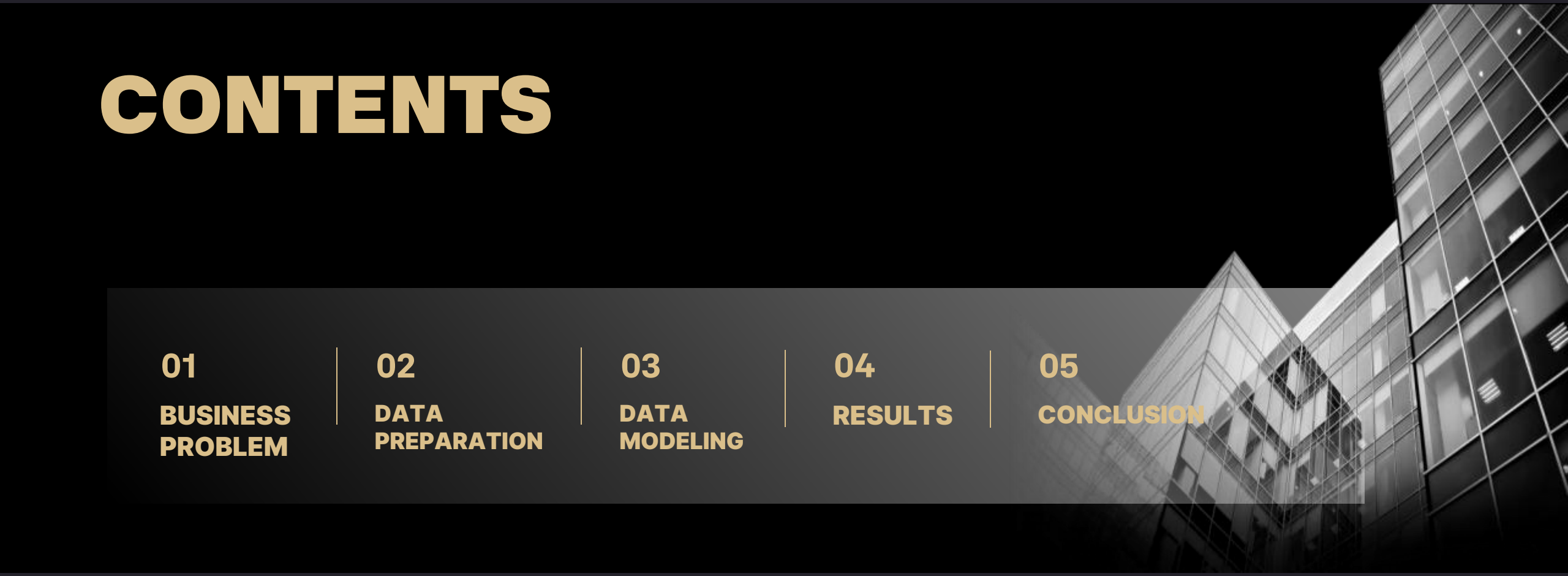
Simple fashion atmosphere

# KING COUNTY

## HOUSING MODEL

Simple fashion atmosphere

# CONTENTS




01
BUSINESS PROBLEM

# BUSINESS PROBLEM

- There is a rising market of home-buyers and home-sellers in King County, indicating a vibrant real estate landscape. However, navigating this market can be daunting for homeowners who are unsure about the key factors driving house prices.
- Our project equips homeowners with insights into King County's housing market, aiding in informed pricing strategies and property enhancements for maximizing sale potential.

# RESEARCH QUESTIONS

1. What are the current market trends and conditions in King County's real estate market?
2. What is the outlook for the King County real estate market in terms of future growth and stability?
3. What are the typical features and characteristics of homes that command higher prices in King County?
4. How do factors like property size, number of bedrooms and bathrooms, floor plan layout, and architectural style correlate with home prices?
5. Does the quality of view of houses have a significant effect on price?

# DATA UNDERSTANDING

- ✓ The dataset used is the King County House Sales dataset, containing property sale information in King County, Washington. It's crucial for analyzing factors affecting house prices in the area.

- ✓ Each row represents a single house sale, recording various property attributes. This data helps analyze market trends and provides insights to real estate professionals, homeowners, and buyers.

- ✓ The target variable is the "price" column. The main goal is to predict house prices based on features like square footage, bedrooms, bathrooms, location, and other relevant attributes.

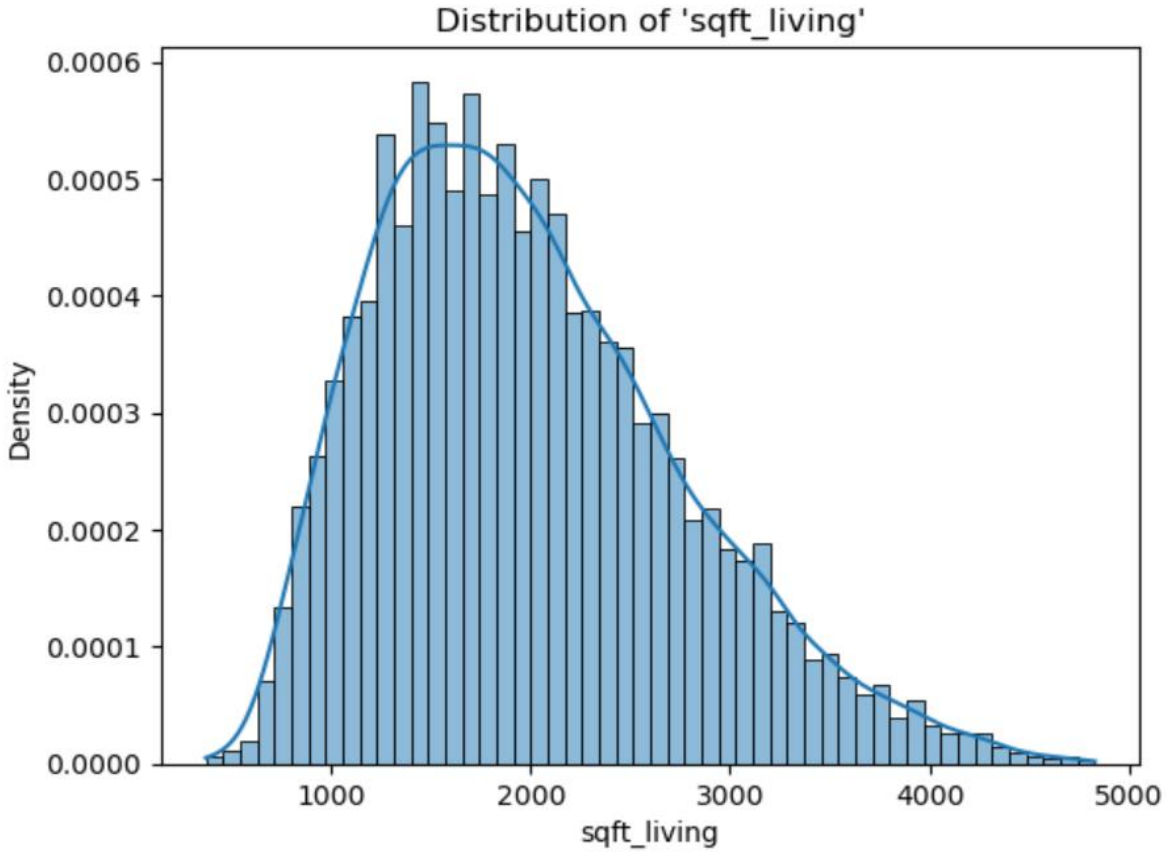

02
DATA PREPARATION

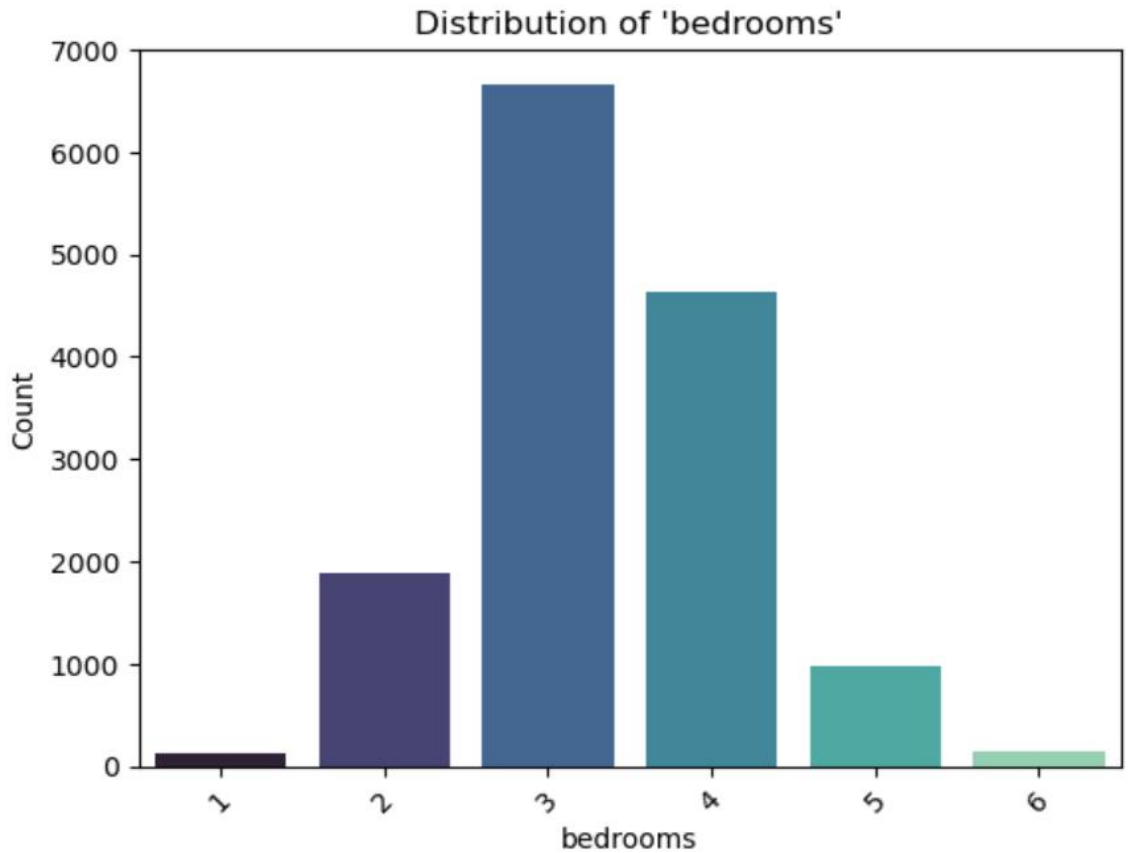
# DATA CLEANING

- ✓ Most of the features had outliers in them. Outliers can lead to incorrect inferences and conclusions about the underlying relationship between variables. So we removed them.
- ✓ Those features that contained missing values were also dropped so as not to compromise the integrity of the data.

# CHECKING FOR NORMAL DISTRIBUTION

- ✓ To affirm the integrity of our data, we visualized the density distributions of some of the features. Most of them had normal distributions, meaning our data was viable to be used for analysis.



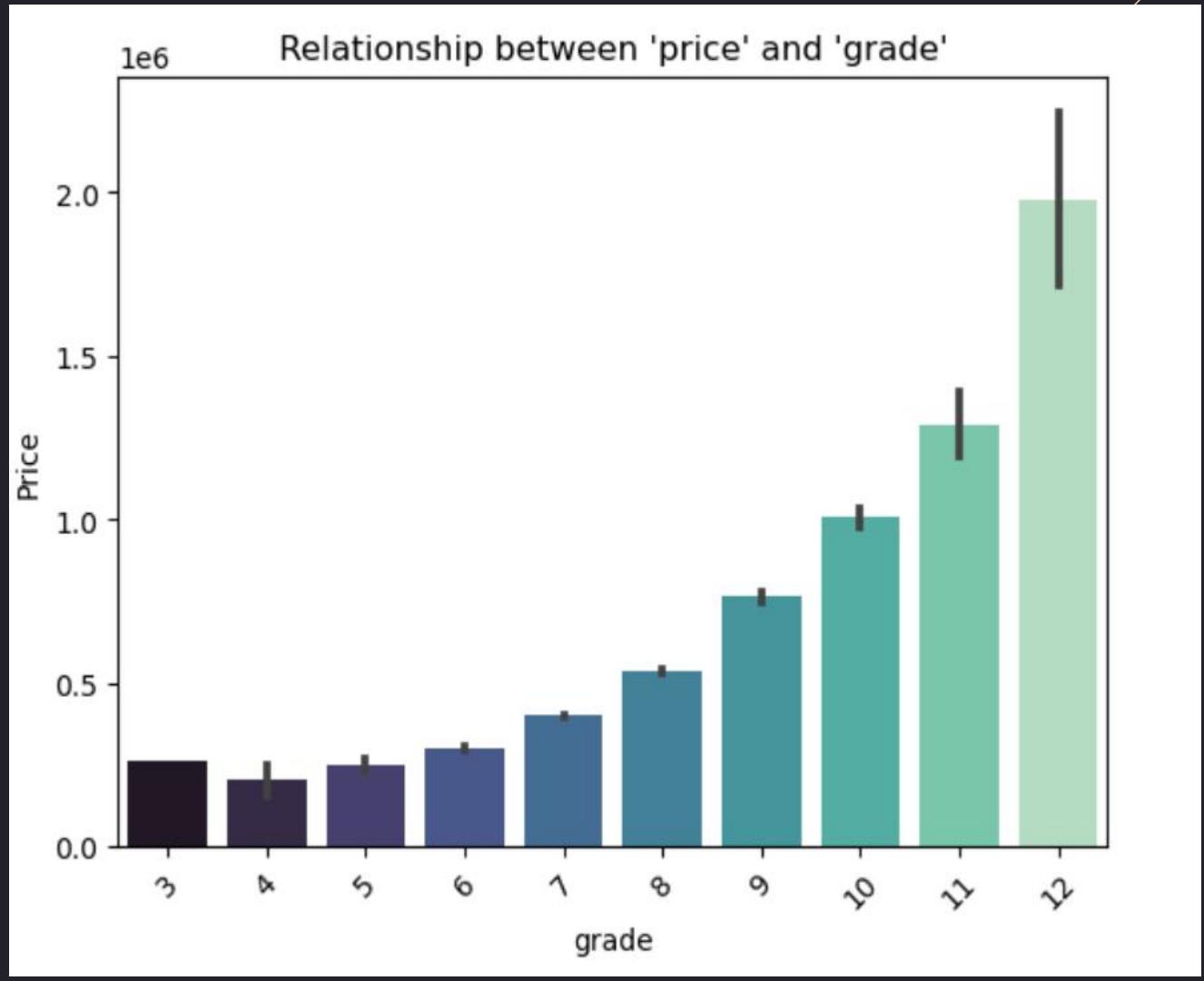



# DATA ANALYSIS

The following features were prominent from the exploratory data analysis conducted:

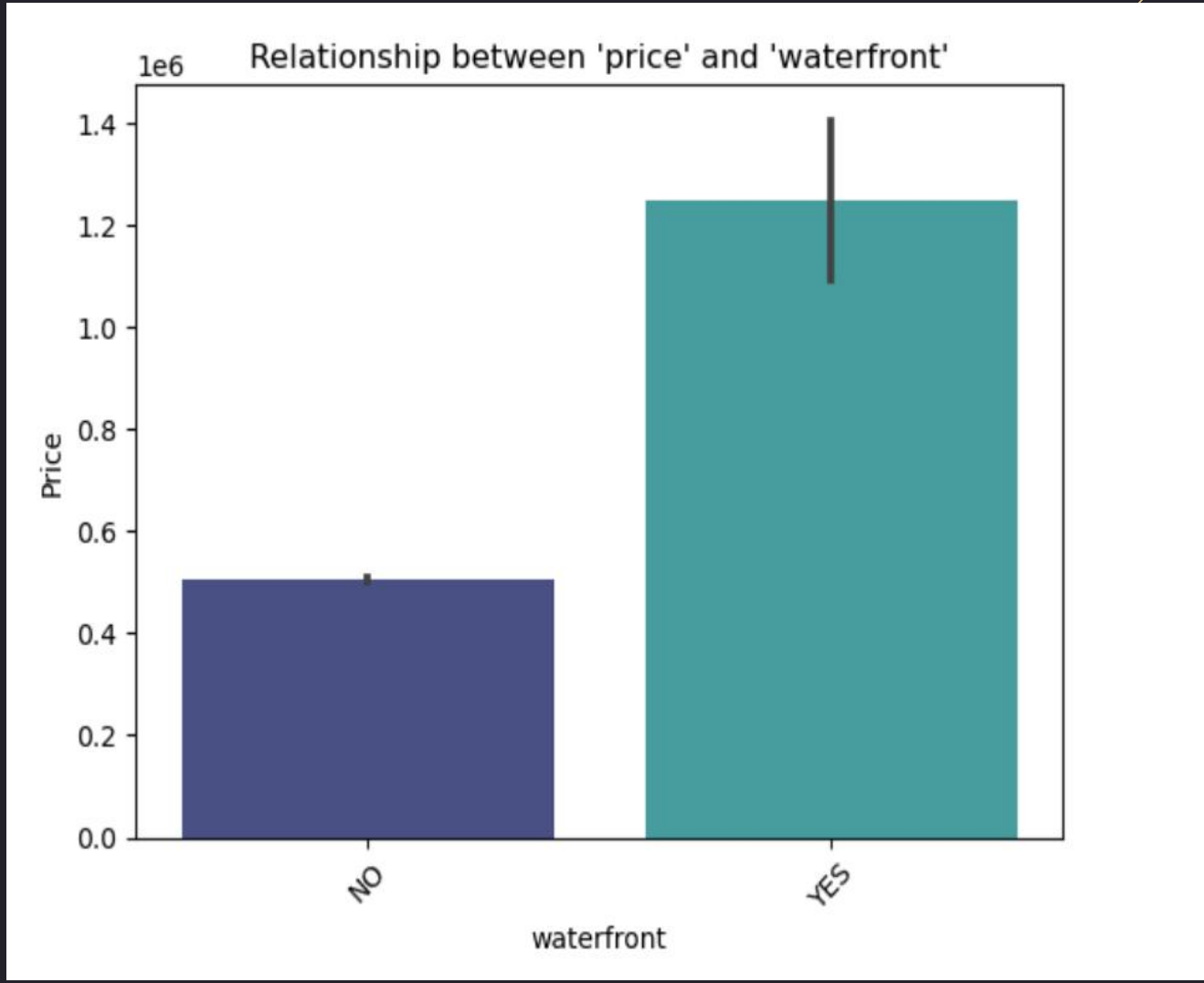
1. **GRADE**: As the grade of a house increased, so did its relative price. This indicates that the grade is a key aspect to take into consideration while setting the price of a house.

- Luxury and Excellent graded houses had the highest prices



# DATA ANALYSIS

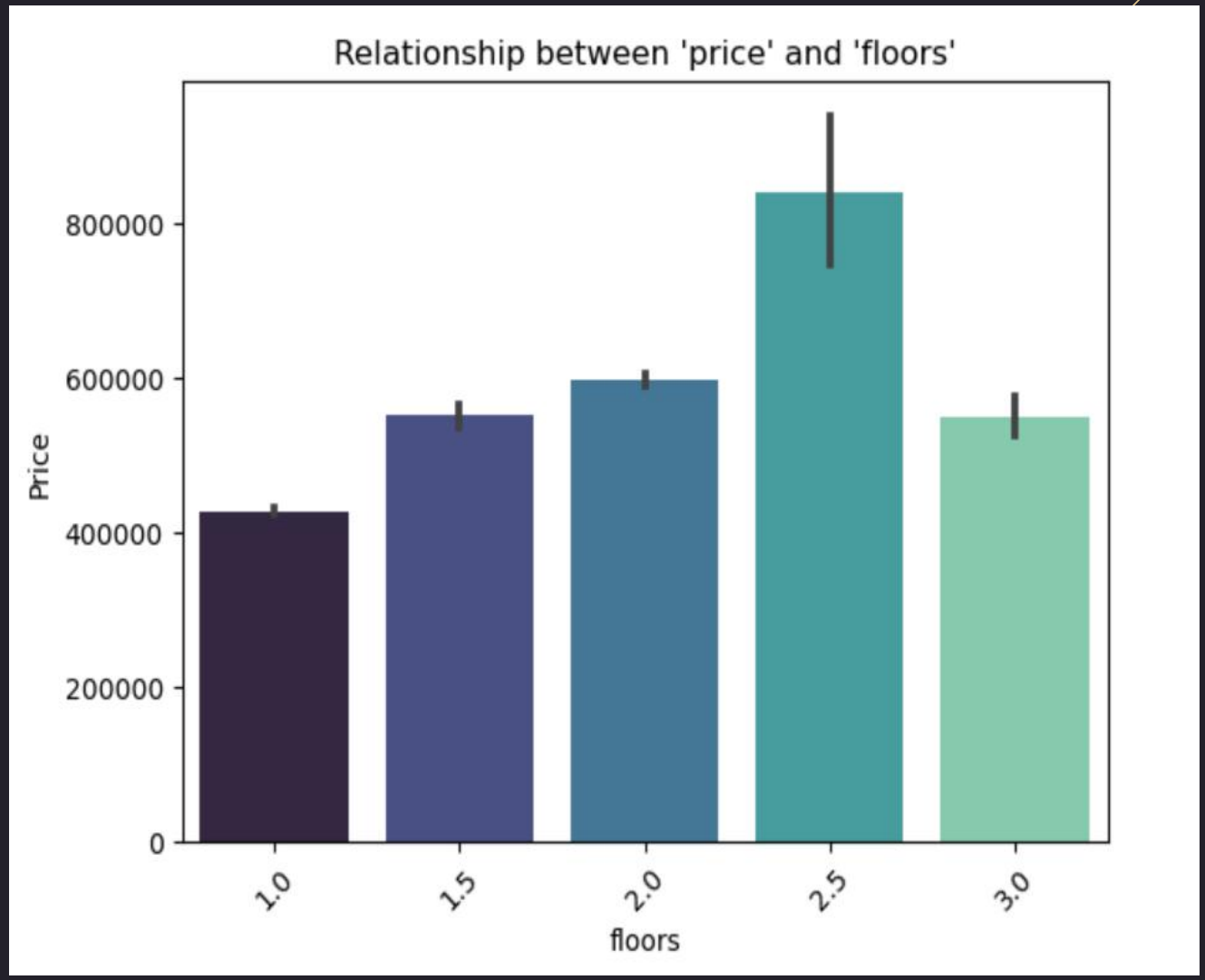
**2. WATERFRONT**: Houses with waterfronts recorded a higher price as compared to those without. It is almost guaranteed to expect a house by the waterfront to have a much higher price than an inland house.



# DATA ANALYSIS

**3. FLOORS**: Houses with a single floor have the lowest price while those with 2.5 floors are the most expensive. Those with 3 floors are likely to cost a bit less than houses with 2 floors.

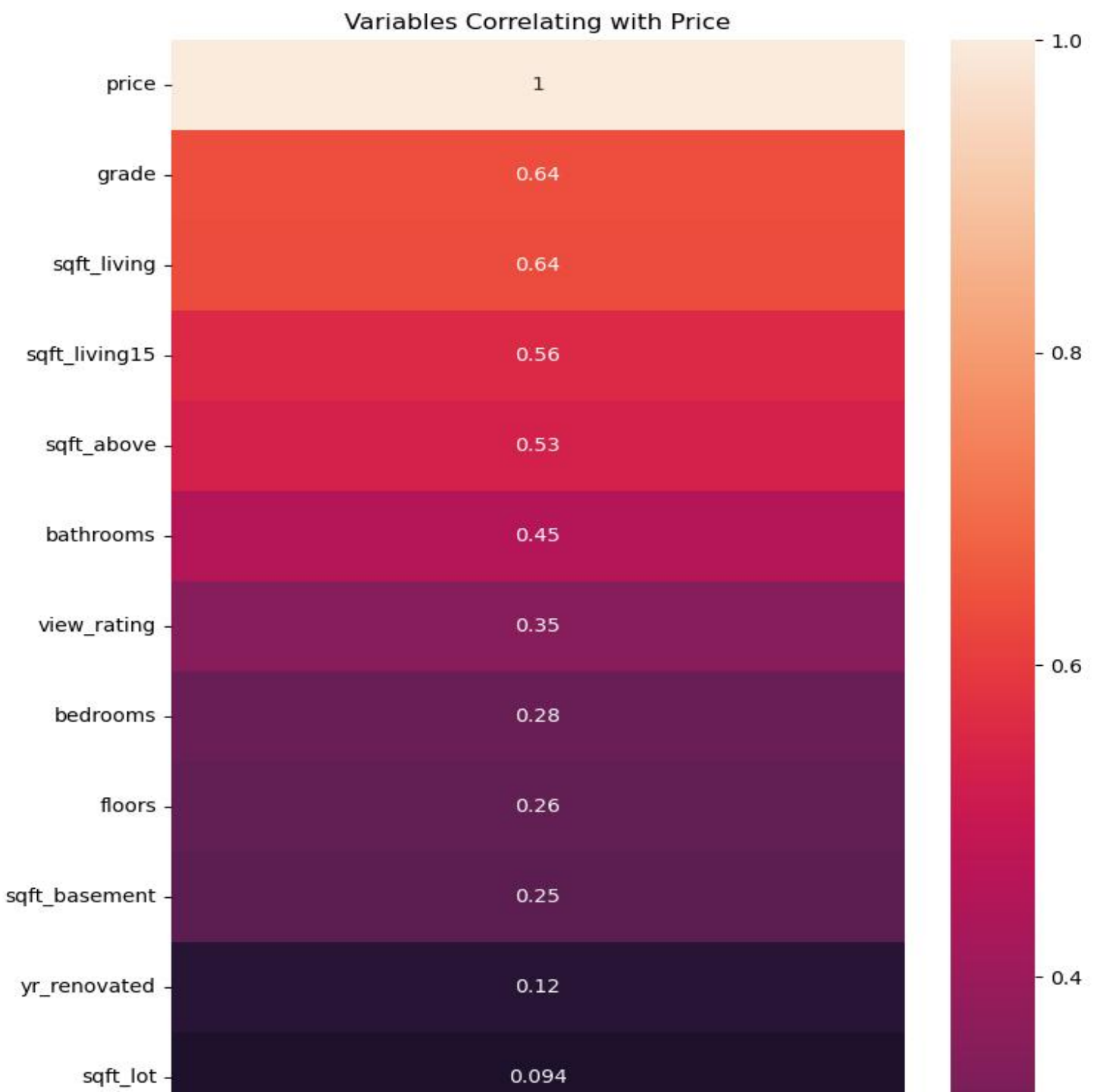
- 1 floor ~ $400K
- 3 floors ~ $500K
- 1.5 floors ~ $550K
- 2 floors ~ $600K
- 2.5 floors ~ $800K



# CORRELATION ANALYSIS

The heatmap on the left shows correlation of all features with price from highest to lowest.

- **Grade** and **square footage of living space** have the highest effect on price.
- The number of **bathrooms** as well is a notable feature that affects price.




03
DATA MODELING

# MODELING

We came up with 4 models to use in our project:

1. Simple linear regression model
2. Multiple linear regression model
3. One-Hot Encoded MLR
4. Log-Transformed MLR

# MODEL EVALUATION

One-Hot MLR

OLS Regression Results

| | | | |
|---|---|---|---|
| **Dep. Variable:** | price | **R-squared:** | 0.738 |
| **Model:** | OLS | **Adj. R-squared:** | 0.737 |

Log-MLR

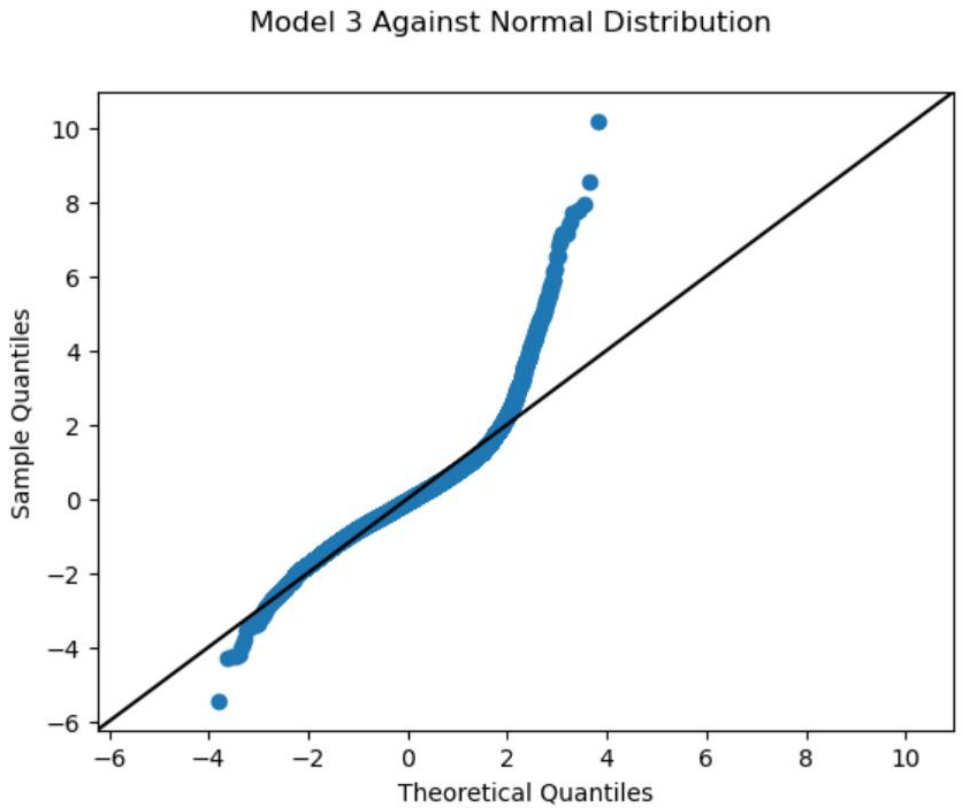
OLS Regression Results

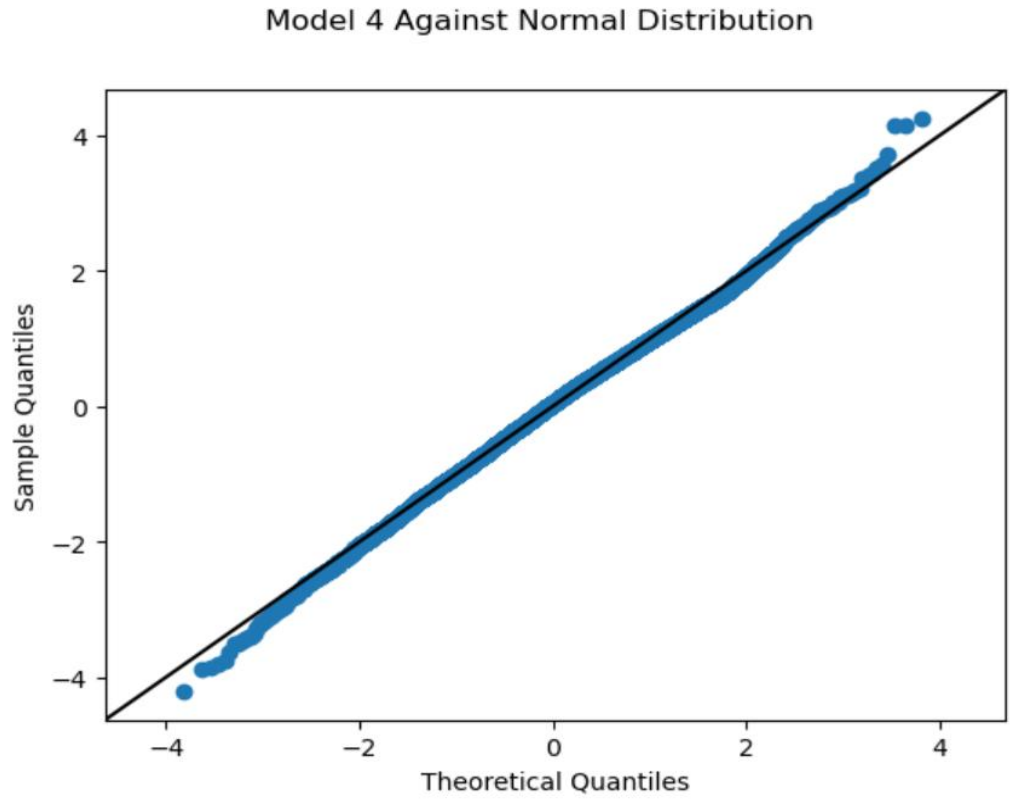
| | | | |
|---|---|---|---|
| **Dep. Variable:** | price | **R-squared:** | 0.784 |
| **Model:** | OLS | **Adj. R-squared:** | 0.783 |

- ✓ Model 4 (Log-MLR) outperformed the third model (One-Hot MLR) in terms of the R-squared value.
- ✓ In essence, the higher the R-squared value, the better the regression model fits the data, as it indicates how much of the variability in the dependent variable can be attributed to the independent variables in the model.
- ✓ Therefore our best pick was Model 4.

# MODEL EVALUATION

- In addition, Model 4 passed the normality test as shown by the Q-Q (quantile-quantile) plots below :




04
RESULTS

# MODEL RESULTS

1. Houses with waterfront have a 33% higher price than houses without.

2. Houses with excellent views have  a 14% higher price than houses with average views.

3. Houses with very good conditions have a 11% higher  price compared to houses with average conditions.

4. One additional bathroom will result in 7% increase in price.

# MODEL RESULTS

5. Houses with 2 bedrooms have a 4% increase in price than houses with one ,however, houses with more than 4 bedrooms have a decrease in price compared to houses with one.

6. The higher the grade the higher the price as houses with grade 12 have a 100% higher price than houses with grade 5.

7. Houses in Medina city have the highest price , 62% higher than houses in Seattle, Houses in  Mercer Island have a 25% higher price and houses in Bellevue have a 16% higher price.

05

# CONCLUSION

# INSIGHTS

Based on the analysis conducted, we have gathered the following insights:

1. Key Factors Affecting House Prices:
    - Square footage of living space appears to have the most significant impact on house prices, as evidenced by its high correlation with price.
    - Other features such as square footage of living space, grade, waterfront status, and year of renovation also show notable correlations with house prices, indicating their influence on property values.

# INSIGHTS

2. Significance of Property Characteristics:

- The number of bedrooms and bathrooms also plays a role in determining house prices, although their individual impacts may vary.
- Renovation status can affect house prices, with recently renovated properties likely commanding higher values.

# CONCLUSION

4. For every unit increase in the logarithmically transformed floor levels:
    - For houses with 1.5 floors, prices are expected to increase by about 6.56%.
    - For houses with 2 floors, prices are expected to increase by about 4.55%.
    - For houses with 2.5 floors, prices are expected to increase by about 7.94%.

# LIMITATIONS

- ✓ The dataset used for this analysis was limited to a specific geographic area and time period. It may not be representative of other locations or time periods, which could limit the generalizability of the results.

- ✓ The data in the dataset is from 2014 and 2015. Therefore, it may not be able to account for changes in the housing market since then. As a result, the model may not accurately predict the value of a house in 2024.

- ✓ While the model can identify relationships between variables, it cannot prove causality. Therefore, it's important to be cautious about making causal claims based solely on the results of this model.

- Overall, the analysis provides valuable insights for homeowners, real estate professionals, and potential buyers interested in the King County housing market. By understanding the key factors affecting house prices and acknowledging the limitations of the analysis, stakeholders can make informed decisions regarding pricing strategies, property enhancements, and investment opportunities. Further research and analysis may be warranted to explore additional factors or refine existing models for better predictive accuracy actionable insights.

"The world is changed by examples, not by opinions."

- Paulo Coelho

Simple fashion atmosphere

# THANK YOU

GROUP 3

- Claudia Sagini
- Philip Oluoch
- Ivyne Atieng
- Nashon Okumu
- Catherine Wangui

- Amani Wanene
- Simon Makumi
- Bradley Ouko
- Lynn Komen

Simple fashion atmosphere