



KING COUNTY

# KING COUNTY HOUSING PROGRAM

BY GROUP 1

# Outline

---

- 01 Project Overview
- 02 Business Understanding
- 03 Objectives
- 04 Data Understanding
- 05 Data Visualization
- 06 Data Modeling
- 07 Conclusions
- 08 Recommendations



# Project Overview



We aim to develop a model to predict King County house prices using features like square footage and waterfront location.

Through data analysis and preprocessing, we'll refine our dataset and highlight key factors driving house prices.



# Business Understanding



Listing prices is vital for buyers and sellers to make informed decisions. Factors considered include; location, size and layout, condition just to mention a few

It's essential to identify clients' expectations, preferences, and needs with regards to property evaluation, to tailor the agency's services effectively.

# Objectives

---

1. To identify the key features that significantly influence house prices in King County
2. To develop a model that accurately estimates house prices based on the identified features
3. To evaluate the performance of the developed model in estimating house prices in King County



# Data Understanding

---



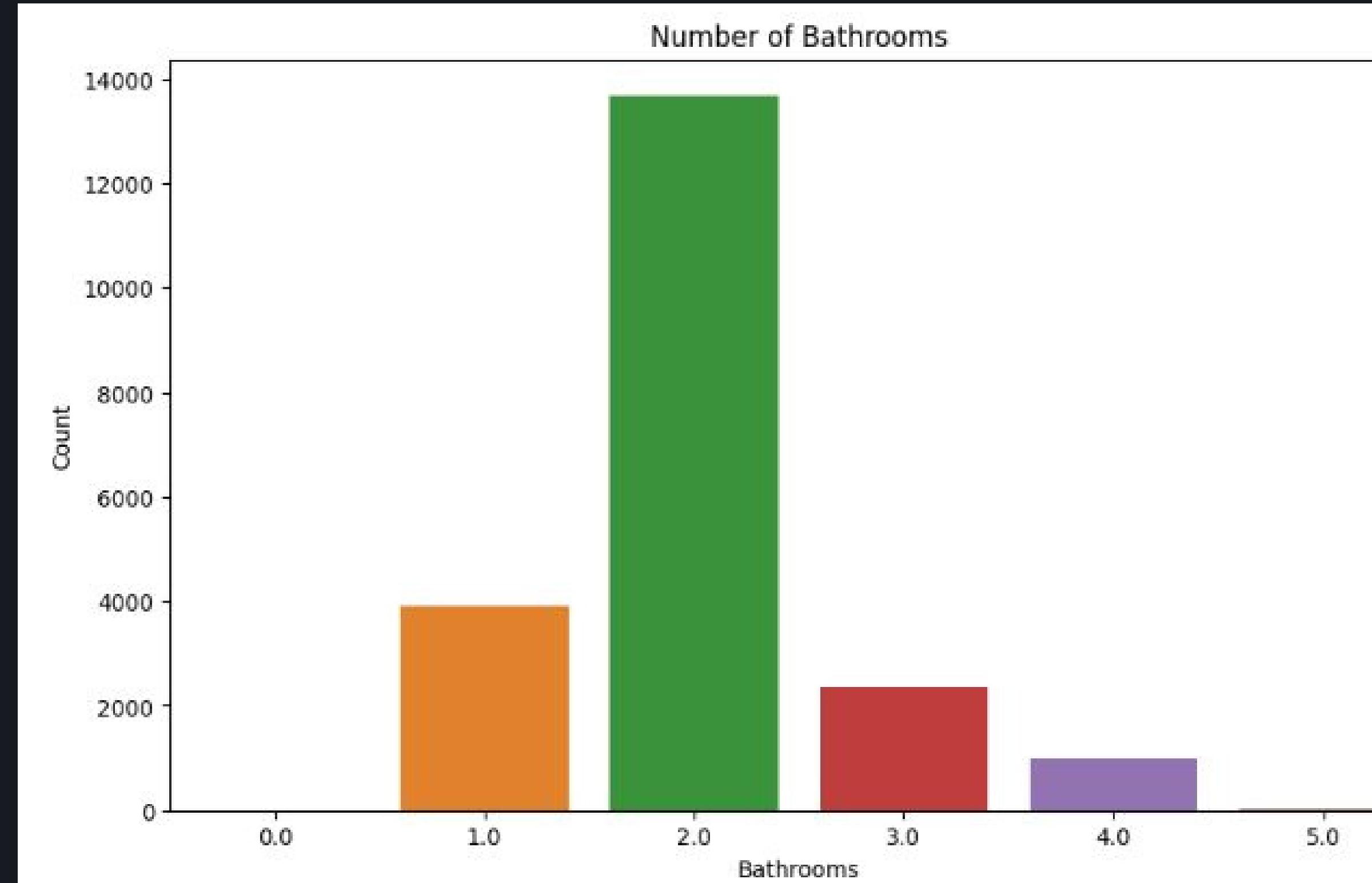
The dataset source was from a file named file named `kc_house_data.csv`. The datasets can be grouped into categorical and numerical data

Data set includes a record of :

- Squarefoot
- Price
- Grade
- Bedrooms
- Waterfront
- Condition

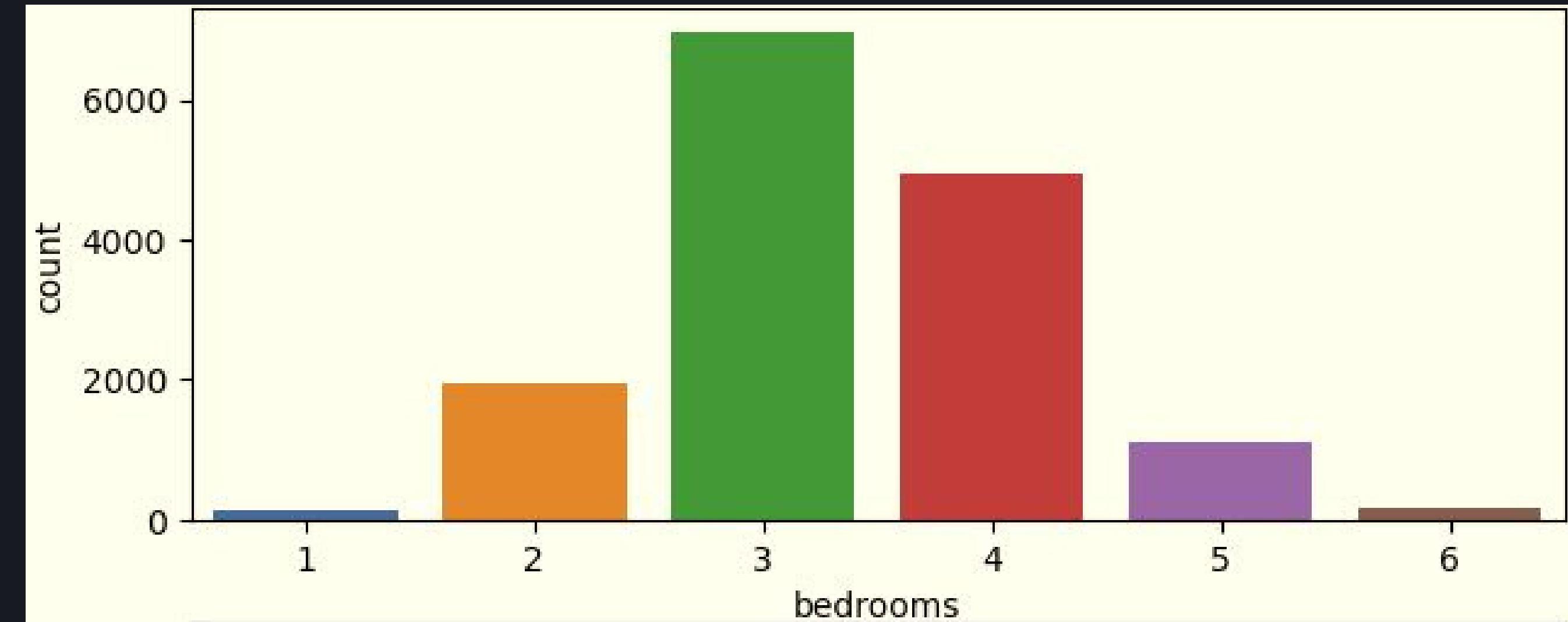
# Exploratory Data Analysis: Bathroom count

- Majority of the houses have 2 bathrooms.



# Exploratory Data Analysis: House Configuration

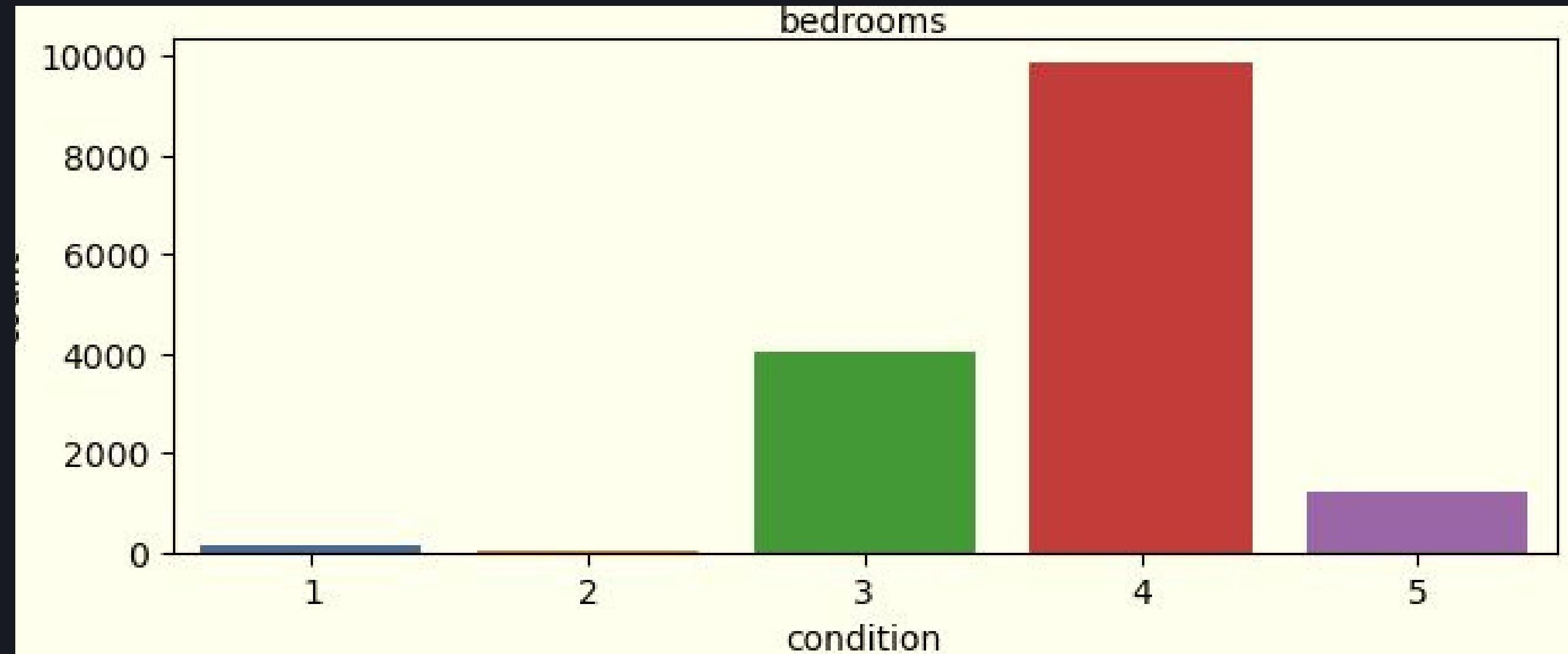
- The most common house configuration is a 3 bedroomed house.



• • •  
• • •  
• • •

# Exploratory Data Analysis: House Condition

- Houses with an average condition are the most prevalent.

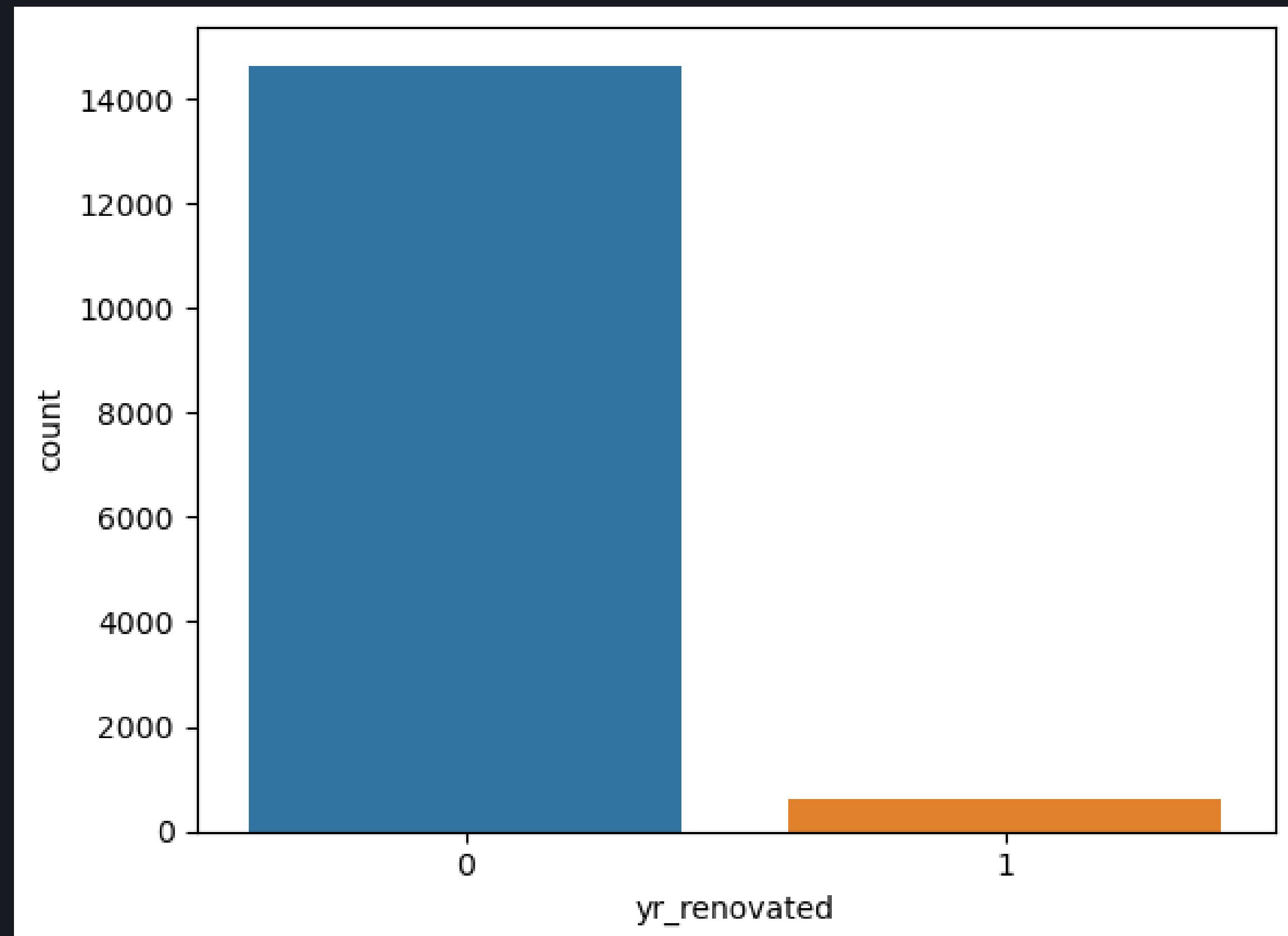


• • •  
• • •  
• • •

# Exploratory Data Analysis: House Renovation

- We find that most houses have not been renovated

• • •  
• • •  
• • •



# Data Modeling

## 1. Linear Regression

### Model results

R-squared: 0.438

P value : 0

### Model evaluation

- 44% of the variation in the price can be explained by the model relative to its mean.
- We reject the null hypothesis because the p- values round to 0

# Data Modeling

2. Multiple Linear Regression  
(features strongly correlating  
price)

## Model results

R-squared 0.529

## Model evaluation

- 53% of the variation in the price can be explained by the model relative to its mean.

# Data Modeling

## 3. Random Forest Regression Model

### Model results

Mean Squared Error Train: 0.01

Mean Squared Error Test: 0.08

R-Squared Train: 0.96

R-Squared Test: 0.71

### Model evaluation

- Training Set R-Squared (R<sup>2</sup>): 0.96, signifying a high ability to explain variability in the training data.
- Test Set R-Squared (R<sup>2</sup>): 0.71, reflecting good but slightly reduced explanatory power on unseen data, indicating some degree of overfitting

# Conclusion

---

- Both linear regression and random forest models were trained, with the latter performing better
- Most houses in King County had 2 bathrooms and 3 bedrooms. The houses with more bathrooms and bedrooms had the highest prices.
- Waterfront-view houses have a wider price distribution with higher prices.
- Some houses underwent renovations with dates recorded, but changes in price, grade, condition, and renovation costs aren't detailed.



# Recommendations

- Factors like bathrooms, bedrooms, having a waterfront and an excellent view have a positive effect on the house prices.
- Analyzing the influence of grade, condition, and renovation on house prices streamlines potential return assessments for property investors and renovators





KING COUNTY

# THANK YOU

