House Price Prediction in King County

Multiple Linear Regression Model

Janaki Purushothaman

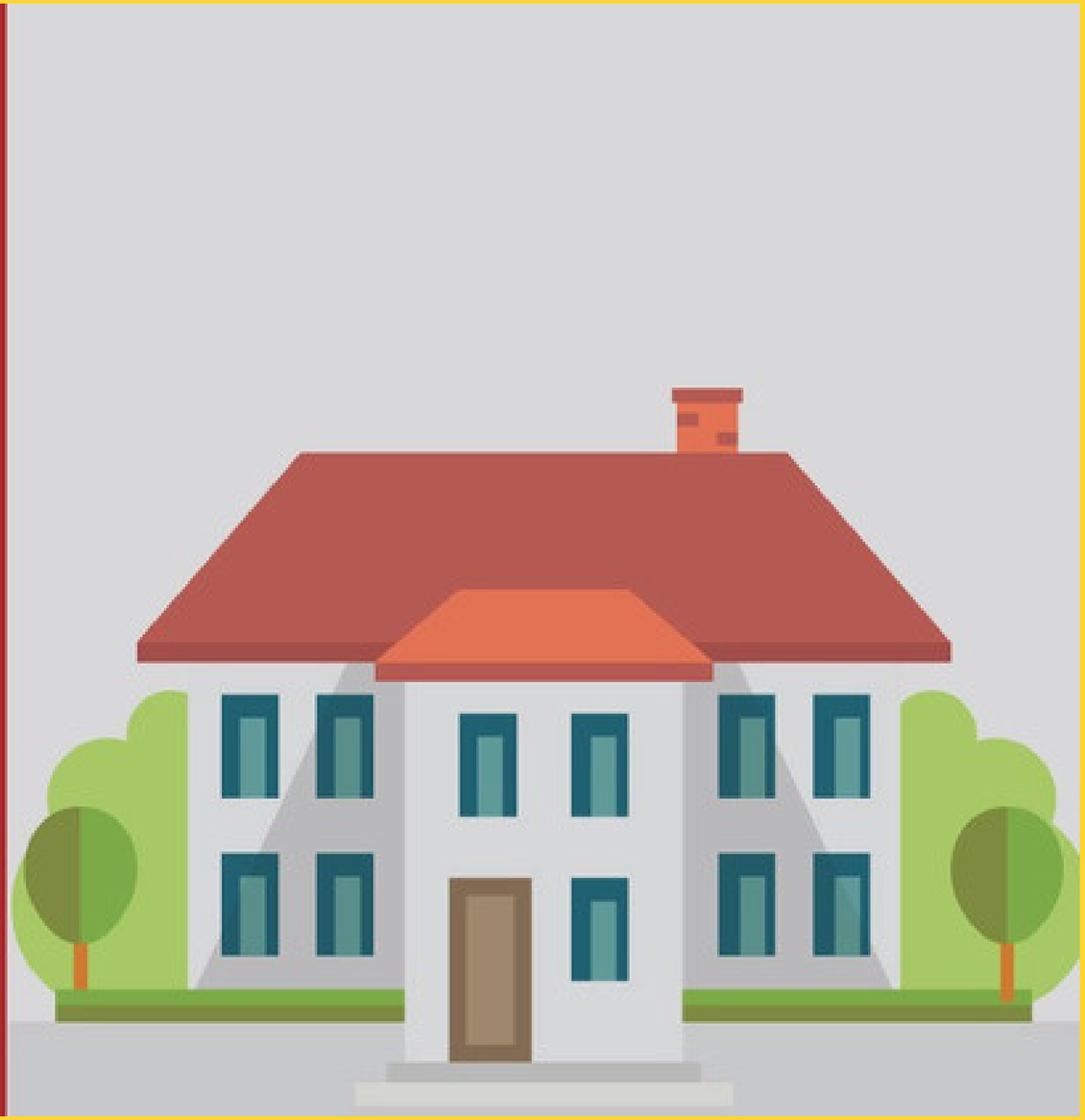
01/31/2021

What We'll Discuss

TOPIC OUTLINE

- **Introduction**
- **Business Problems**
- **Data**
- **Approach/Methodologyy**
- **Analysis**
- **Results**
- **Conclusion**
- **Recommendations**
- **Next Step**

HOUSES FOR SALE



Introduction

This project focuses on predicting the house and investigating house sales in the King County area.

Business Problems

- To build a multiple linear regression model with the given dataset to predict the sale price of the house as accurately as possible
- Analysis on providing prediction data to the real estate agents for homeowners who want to buy homes

Data

This project uses the 2014-2015 King County House Sales dataset, which contains around 21k house sale data.

APPROACH/METHODOLOGY

1. **OBTAIN**
2. **SCRUB**
3. **EXPLORE**
4. **MODEL**
5. **INTERPRET**

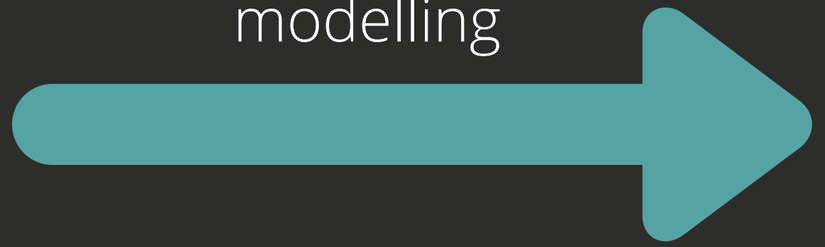
Analysis

House Features

Baseline Model

Model 1

Initial data exploration
modelling



**58%
Accuracy**

Feature Engineering and Testing

Model 2

Linear with Price

57% Accuracy

Applying Log Transformation

Model 3

Normalization

58% Accuracy

Multicollinearity Checks

Model 4

Multicollinearity

55% Accuracy

Applying Categorical Terms

Model 5

Categorical

56% Accuracy

Applying Interaction Terms

Model 6

Interactions

64% Accuracy

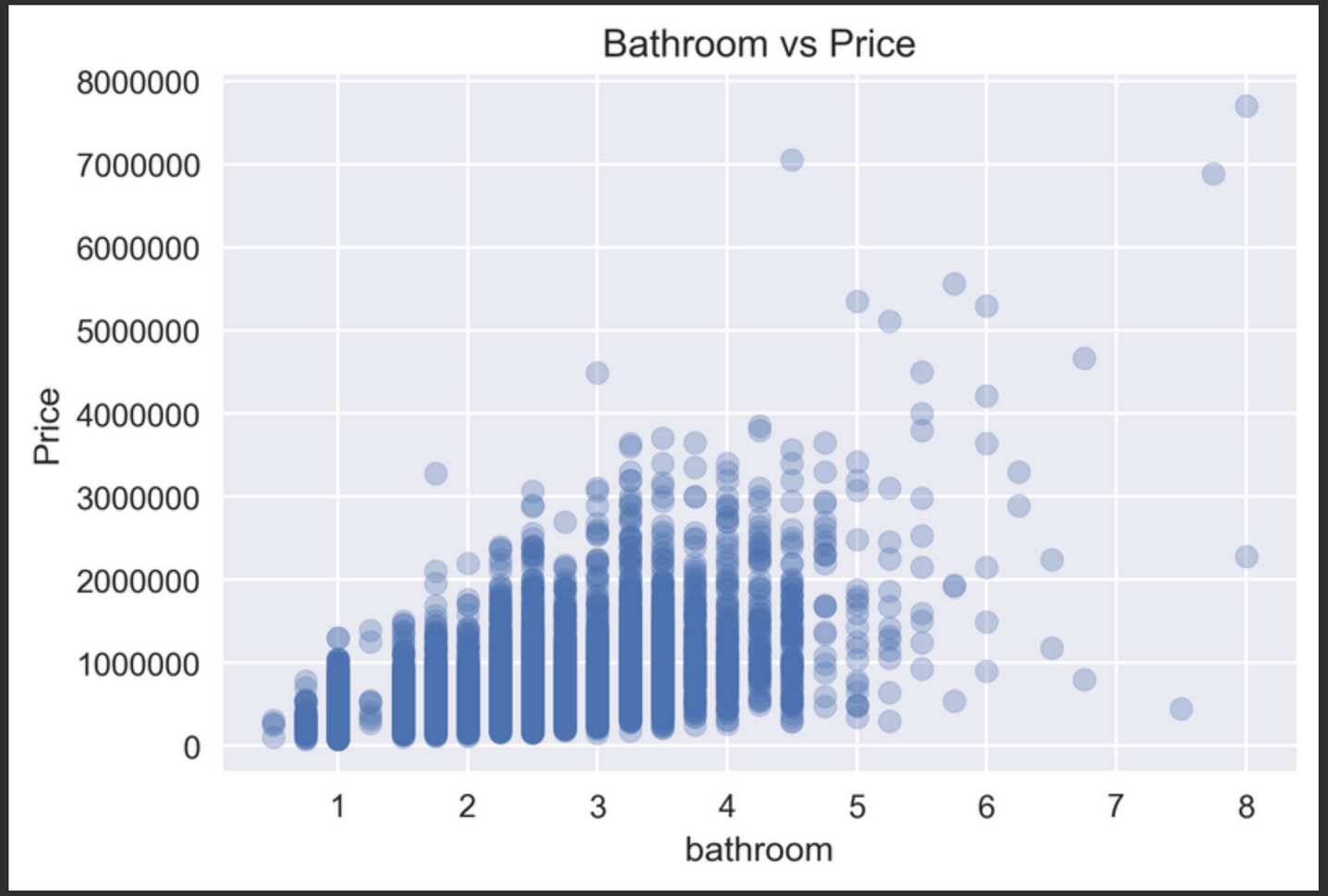
Results:

Model	Model 6
Features	9
R2 Score	0.64
RMSE Value	0.31
Square Footage Living	0.108
Bathroom	0.082

Square Footage Vs Price

The most distinctive feature is the square footage, where the larger house will be more expensive. It is noticeable that the graph is getting broader as the square foot increases. So for every 1% increase in square footage, price increases by 0.11%

Bathroom Vs Price



From the graph, it is clear that as the number of bathrooms increases the house price increases. So for every 1% increase in the bathroom, price increase by 0.08%

Conclusion:

- Our model RMSE for the train and test set are similar, which indicates that our model will perform well on different data
- Our model can be used as a predictor for the Real Estate Agents to help homeowners to buy a house or to analyze the various criteria that stimulate the housing price

Recommendations:

- If you are looking for housing that won't make your bank account fragile, then go for the housing with a minimal bathroom so that you could share them
- It's advisable to shrink on square footage that would make our house purchase a very cost-effective one.

NEXT STEP

Linear Assumptions	Multicollinearity
Distance from economic hubs	Zip code data
Outliers	Handling outliers

THANK YOU!

EMAIL:
janakipurushothamman@gmail.com

GITHUB:
@JanakiGanesh