

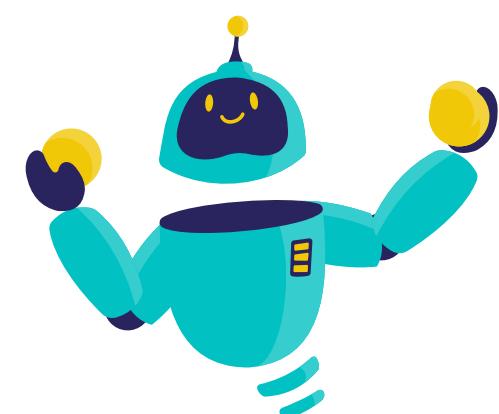
# Airbnb Price Prediction for Cape Town

A Data Science Approach to  
Dynamic Pricing Optimization

Oscar Msizi Mzobe  
21/07/2025

# TABLE OF CONTENTS

- 01 Project Overview**
- 02 Problem Statement**
- 03 Objectives**
- 04 Data Description**
- 05 Data Cleaning**
- 06 Exploratory Data Analysis**
- 07 Model Performances Comparison**
- 08 Recommendations**
- 09 Conclusion**



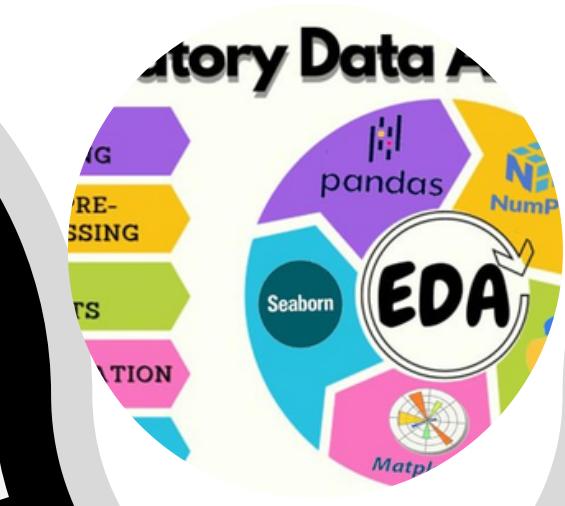
# The Journey

Understanding the Problem



Data Collection & Cleaning

Exploratory Data Analysis (EDA)



Feature Engineering & Modeling

Insights & Recommendations



# Project Overview



## Objective

- develop a machine learning model to predict Airbnb listing prices in Cape Town using property, location, and host data.



## Scope

- analyze Airbnb data from March 2025 to uncover key pricing drivers, clean and transform data, and build predictive models for accurate price forecasting.



## Impact

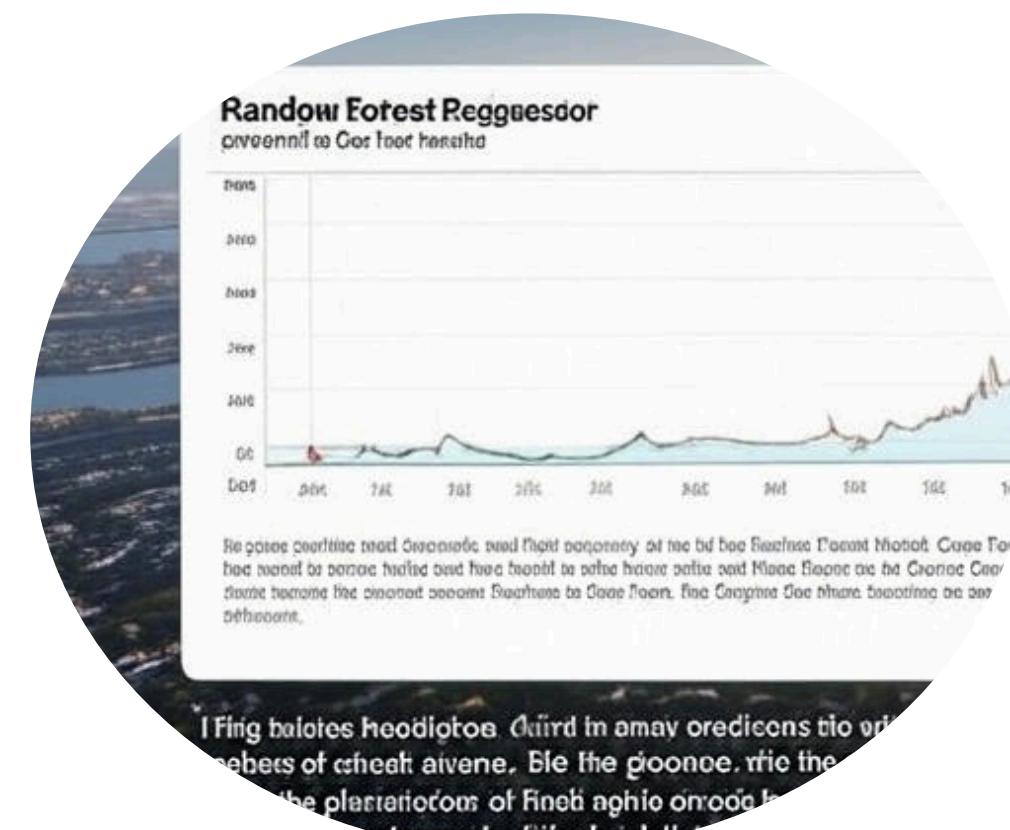
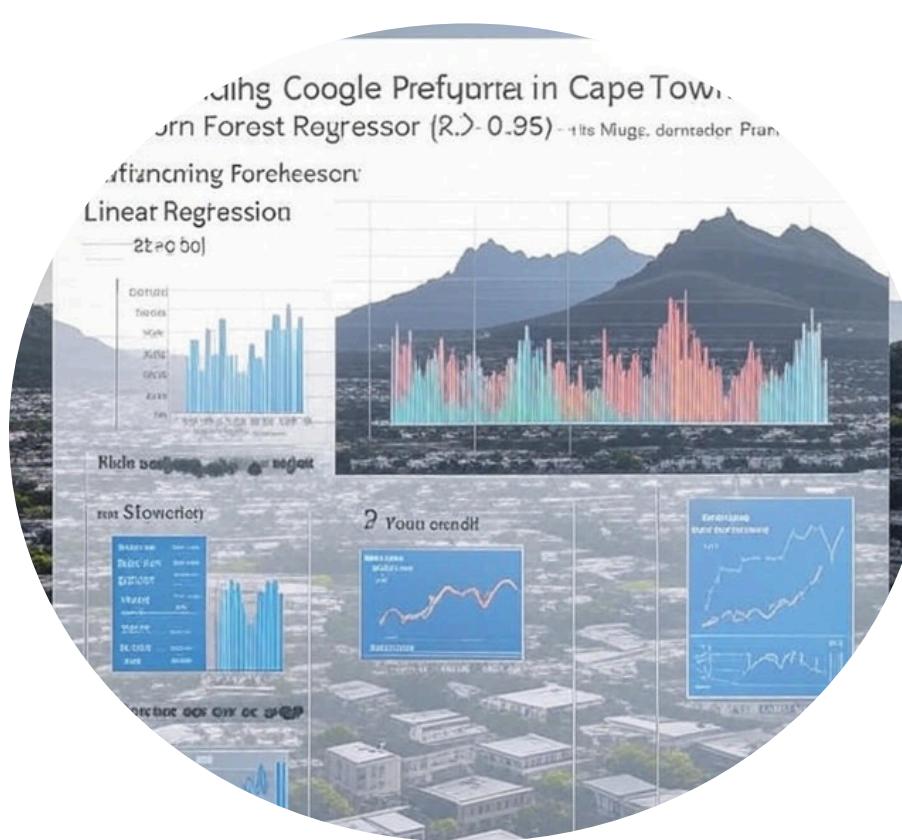
- empower hosts to optimize pricing strategies and help travelers make informed booking decisions, improving the efficiency of Cape Town's short-term rental market.

# Problem Statement

- Traditional pricing relies on intuition or comparisons.
- Lack of tools to factor in complex dynamics.
- Need for a machine learning model to assist in strategic pricing.



# Objectives



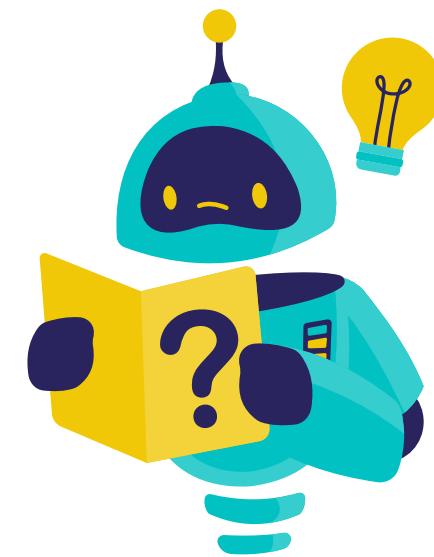
Analyze key factors influencing Airbnb pricing.

Develop a predictive model using machine learning.

Offer actionable insights for both hosts and travelers.

# Dataset Description

- Source: (<https://data.insideairbnb.com/south-africa/wc/cape-town/2025-03-19/visualisations/listings.csv>)
- Features:
  - Price (target)
  - Room type, location coordinates
  - Availability, review metrics
- Size:
  - Total Rows: 25,882
  - Total Columns: ~18

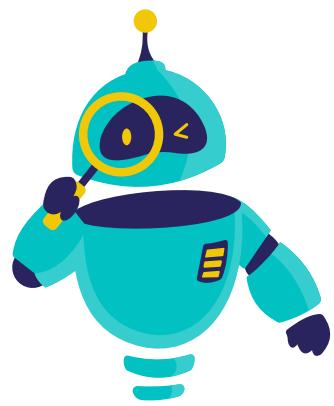


# Data Cleaning

- Removed nulls, duplicates.
- Cleaned currency symbols from price fields.
- Handled outliers and inconsistent values.



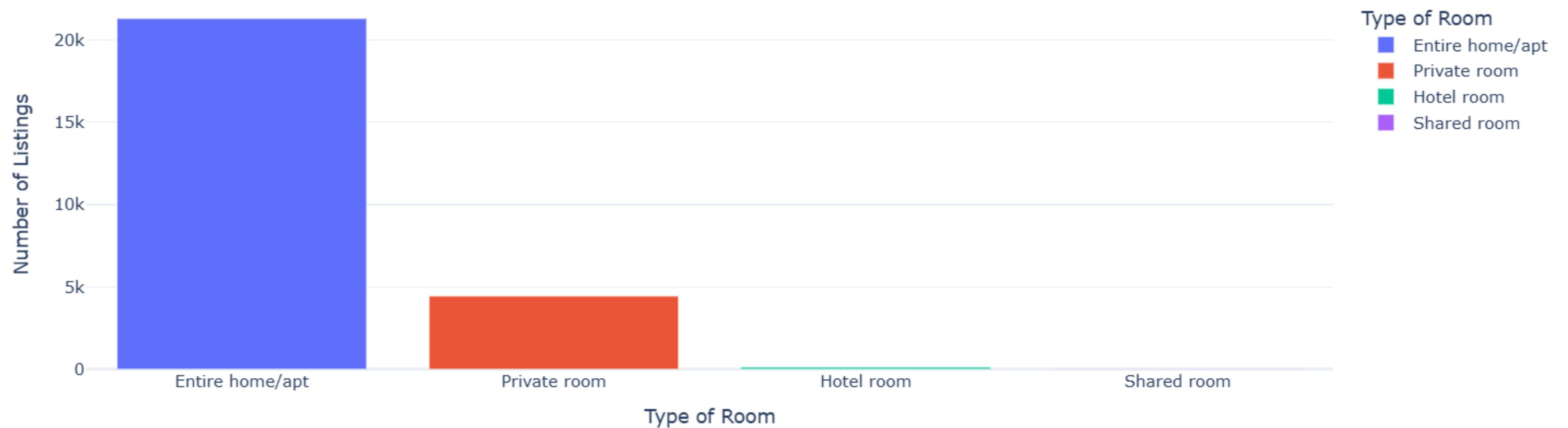
# **EXPLORATORY DATA ANALYSIS**



# Type of Room Distribution



Distribution of Room Types



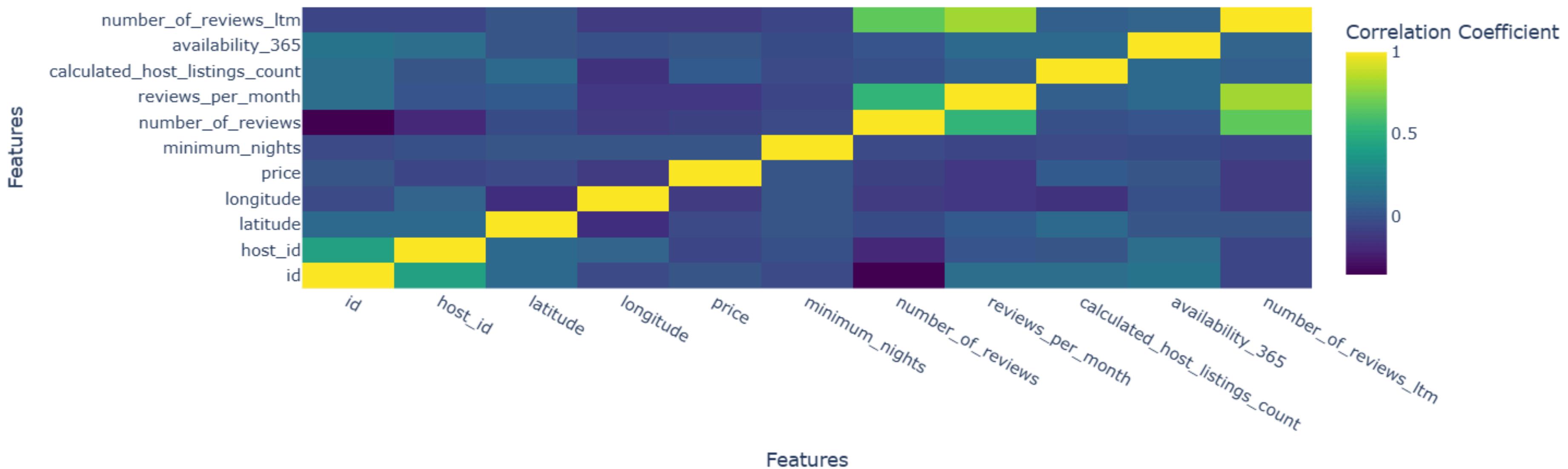
# The proportion of each room type



# Correlation Map



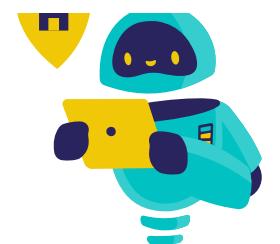
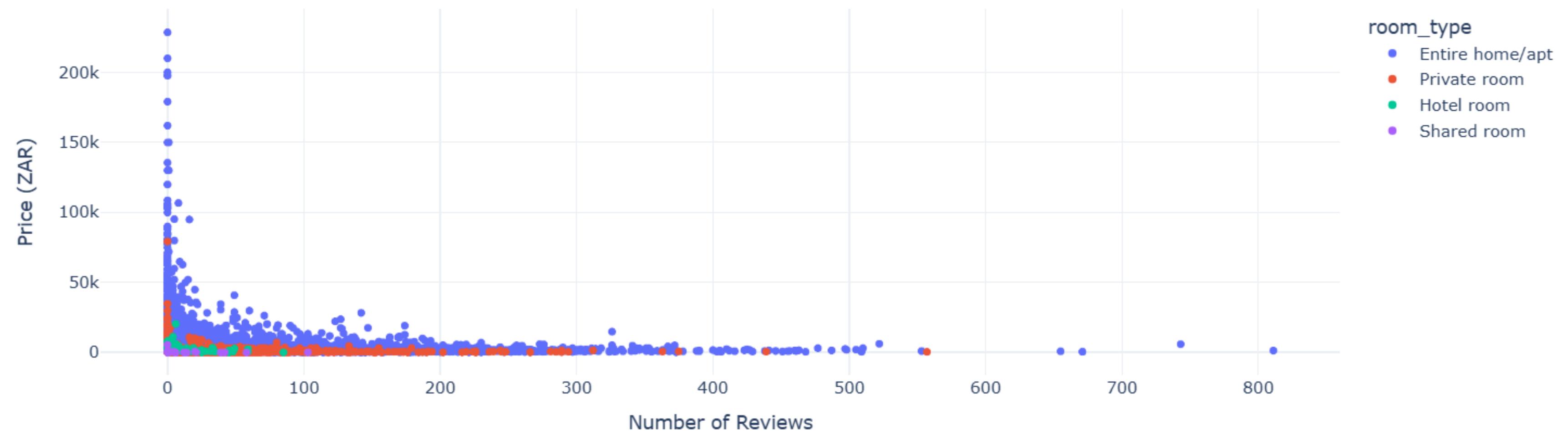
## Correlation Matrix of Numerical Features



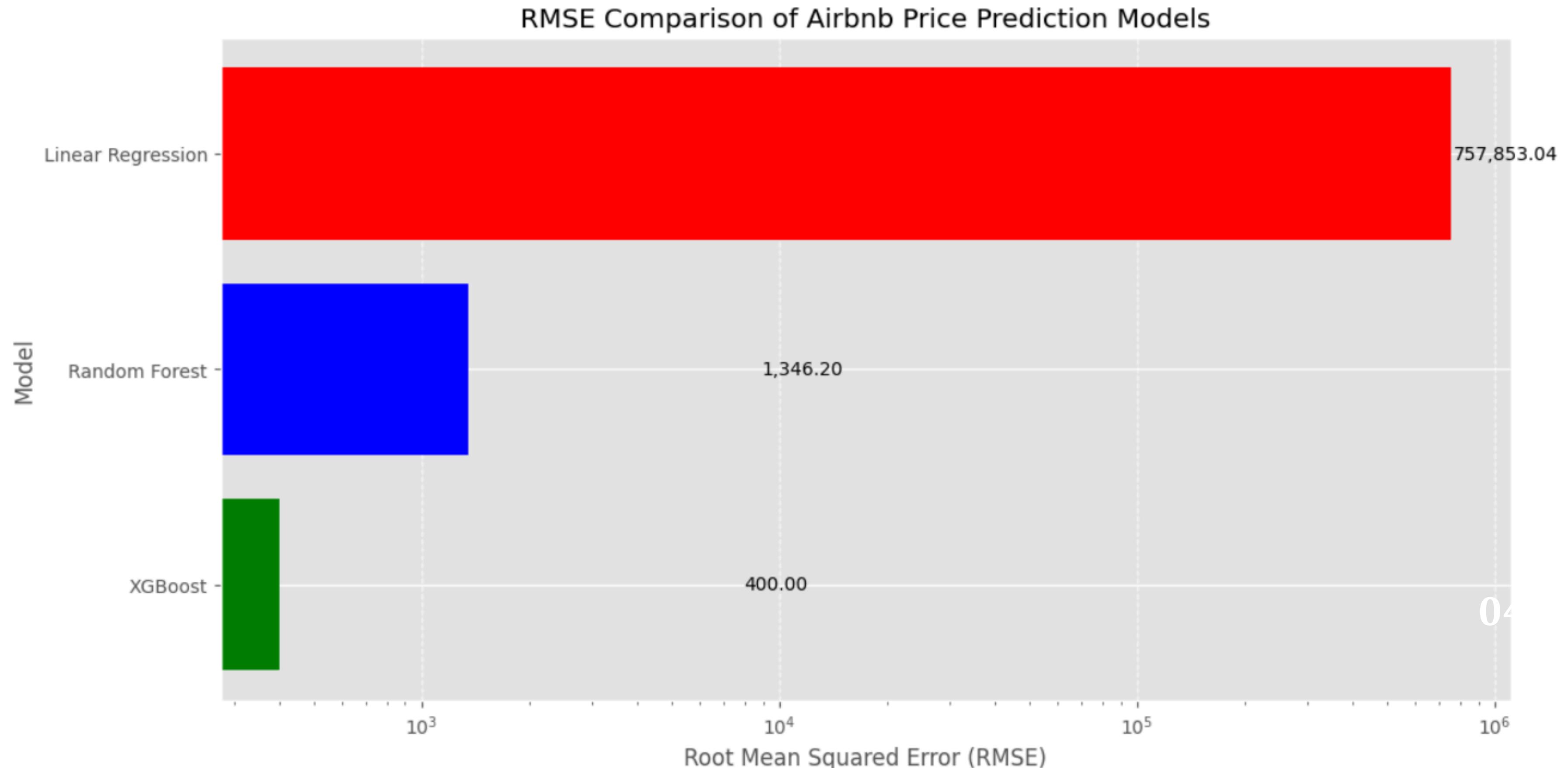
# Scatter Plot of Price vs Reviews



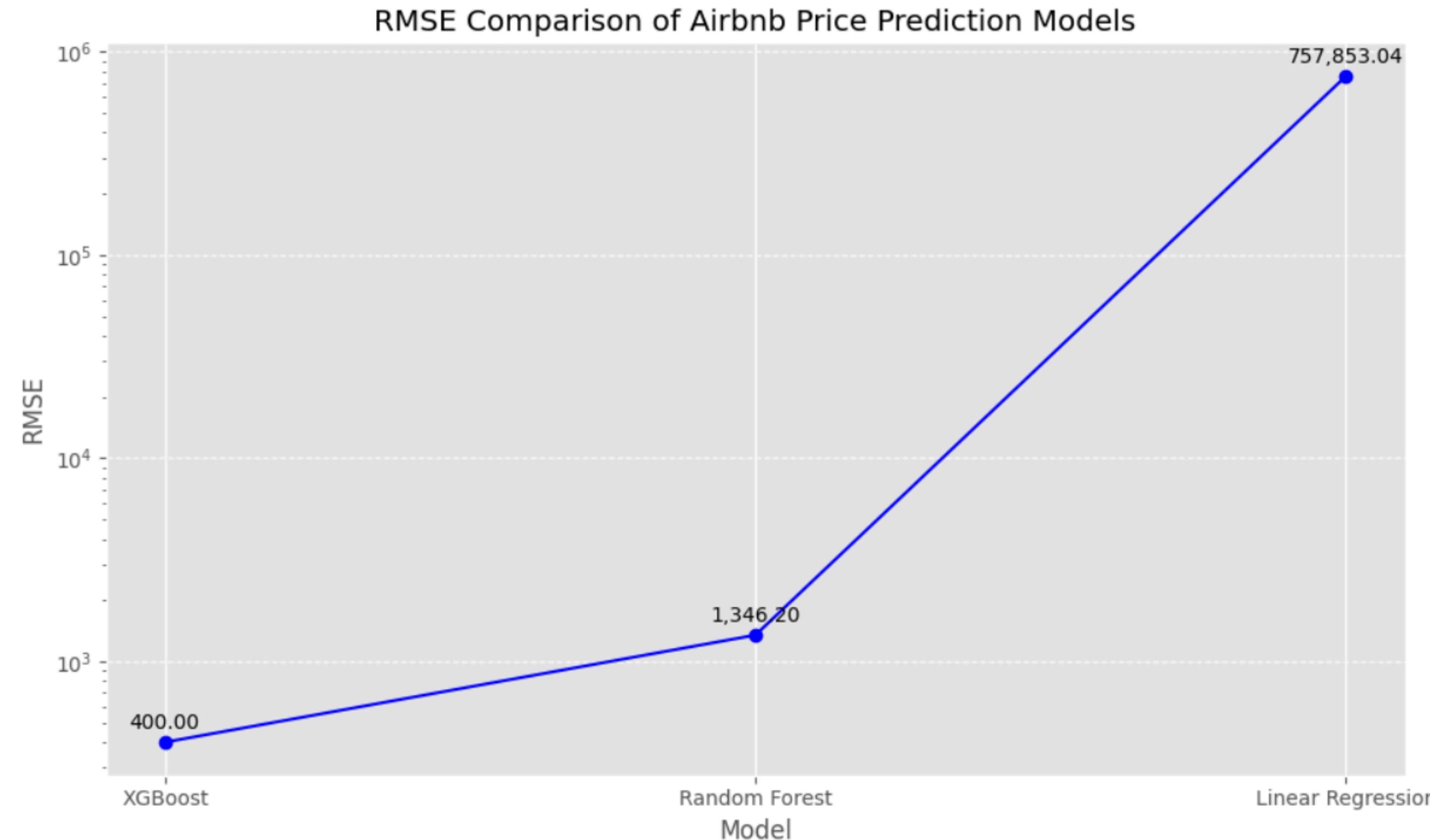
Price vs. Number of Reviews



# Model Performances (RMSE)

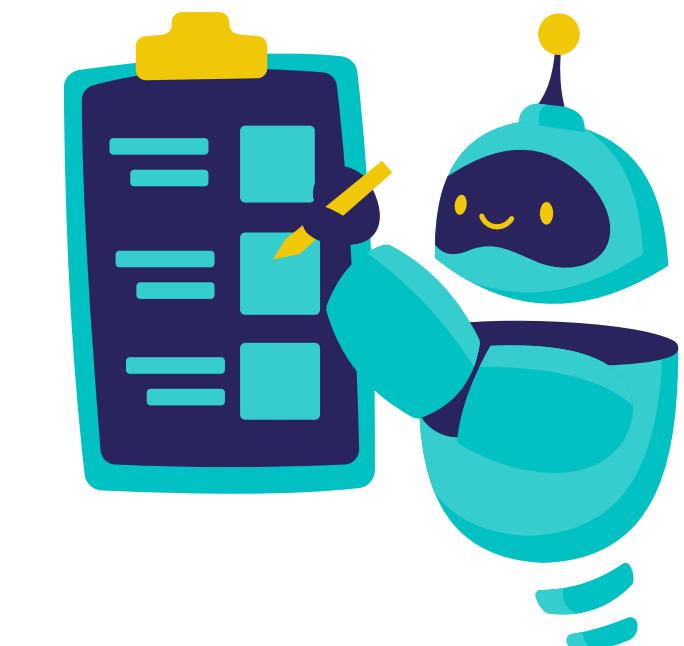


# Model Performances (Line Plot)



# Recommendations

- Hosts should consider:
  - Reducing minimum nights
  - Optimizing listing near city hotspots
  - Leveraging review engagement
- Use ML model to guide price adjustments seasonally.



# Conclusion

- Machine learning can enhance dynamic pricing.
- Insights help balance affordability and profitability.
- Future work: Add time-series features and real-time pricing APIs.

