

# EDA on Airbnb Listings

Presented by: Nihar Suman

Guided by: Naina Devi

Techno India University | B.Tech CSE-AI (3rd Year)



# Agenda

“The main agenda of my project is to analyze Airbnb listings and predict the price of a property based on various factors like neighbourhood, room type, and reviews. The project also visualizes trends and patterns to help understand pricing behavior.”

# Introduction to the Project

## The Rise of Airbnb and Predictive Pricing

Airbnb is a leading platform connecting travelers with unique accommodations worldwide. For both hosts and guests, accurate pricing is **paramount**.

- This project analyzes extensive Airbnb data, using machine learning (**Scikit-learn**) to model price determination and **Streamlit** to deliver the results via an interactive dashboard.
- Identify market drivers and pricing behavior.
  - Establish a fair and competitive pricing model.



# Technical Ecosystem: Tools & Technologies

We employed a robust set of tools for data handling, model training, and deployment, ensuring a scalable and efficient solution.



## Core Libraries

**Pandas**, **NumPy** for data manipulation; **Matplotlib**, **Seaborn** for visualization; **Scikit-learn** for predictive modeling; **Joblib** for model serialization.



## Deployment & Interface

**Streamlit** for rapid development of the interactive web application; **Render** for cloud deployment and accessibility.



## Data & Environment

Dataset: **New York Airbnb Listings 2024** provides rich features for analysis. Development in **Jupyter Notebook** and **VS Code**, version control with **GitHub**.

# Project Working

 Airbnb Project

Select a section

- Home / Dataset Overview
- Outliers in Price
- Price Distribution
- Minimum Nights Distribution
- Number of Reviews Distribution
- Availability 365 Distribution
- Correlation Heatmap
- Number of Reviews vs Price
- Room Type vs Average Price
- Price Dependency on Neighbourhood
- Geographical Distribution of Listings
- Pairplot Overview
- Predict Price

## Airbnb Dataset Overview

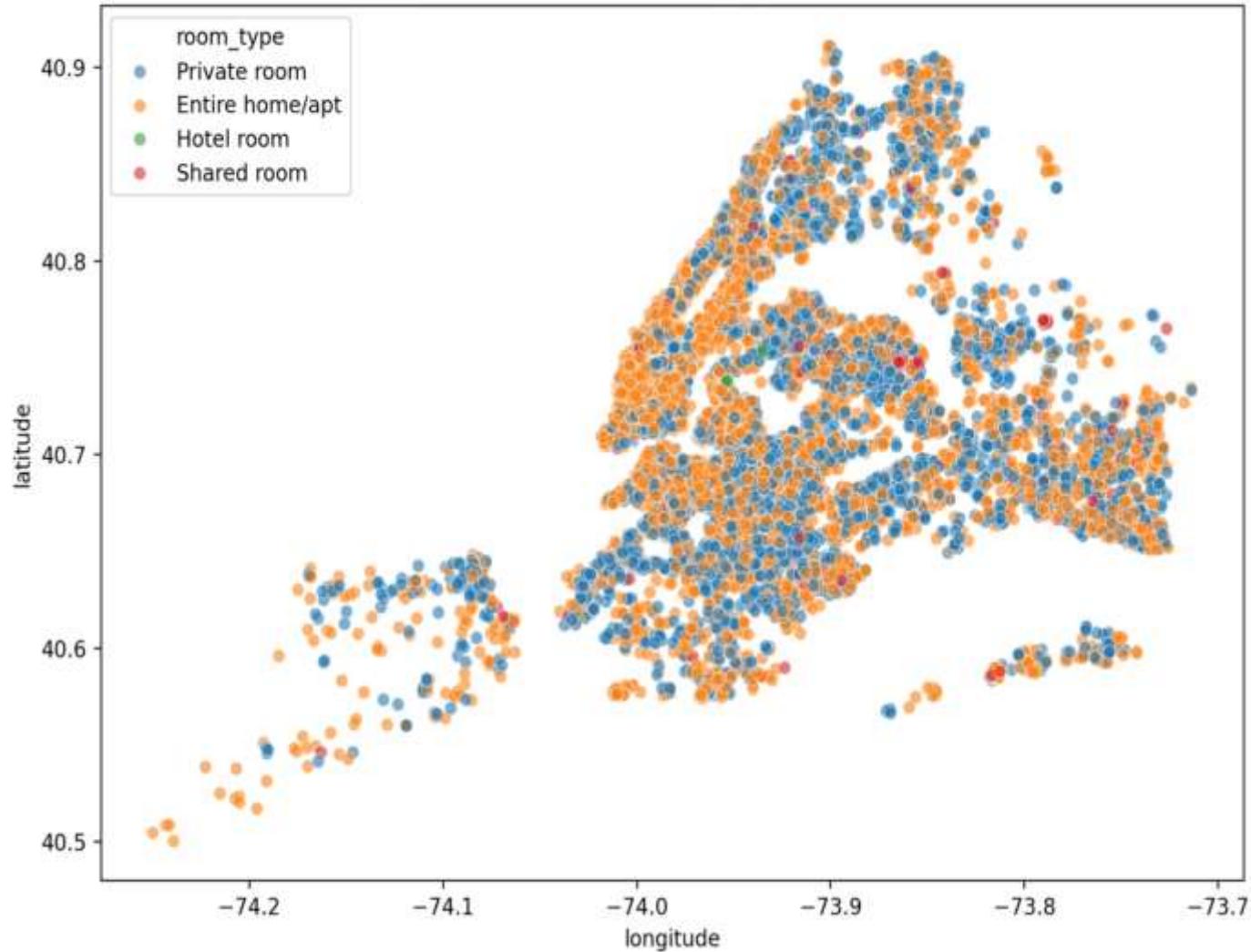


This section gives an overview of the dataset used for Airbnb price prediction.

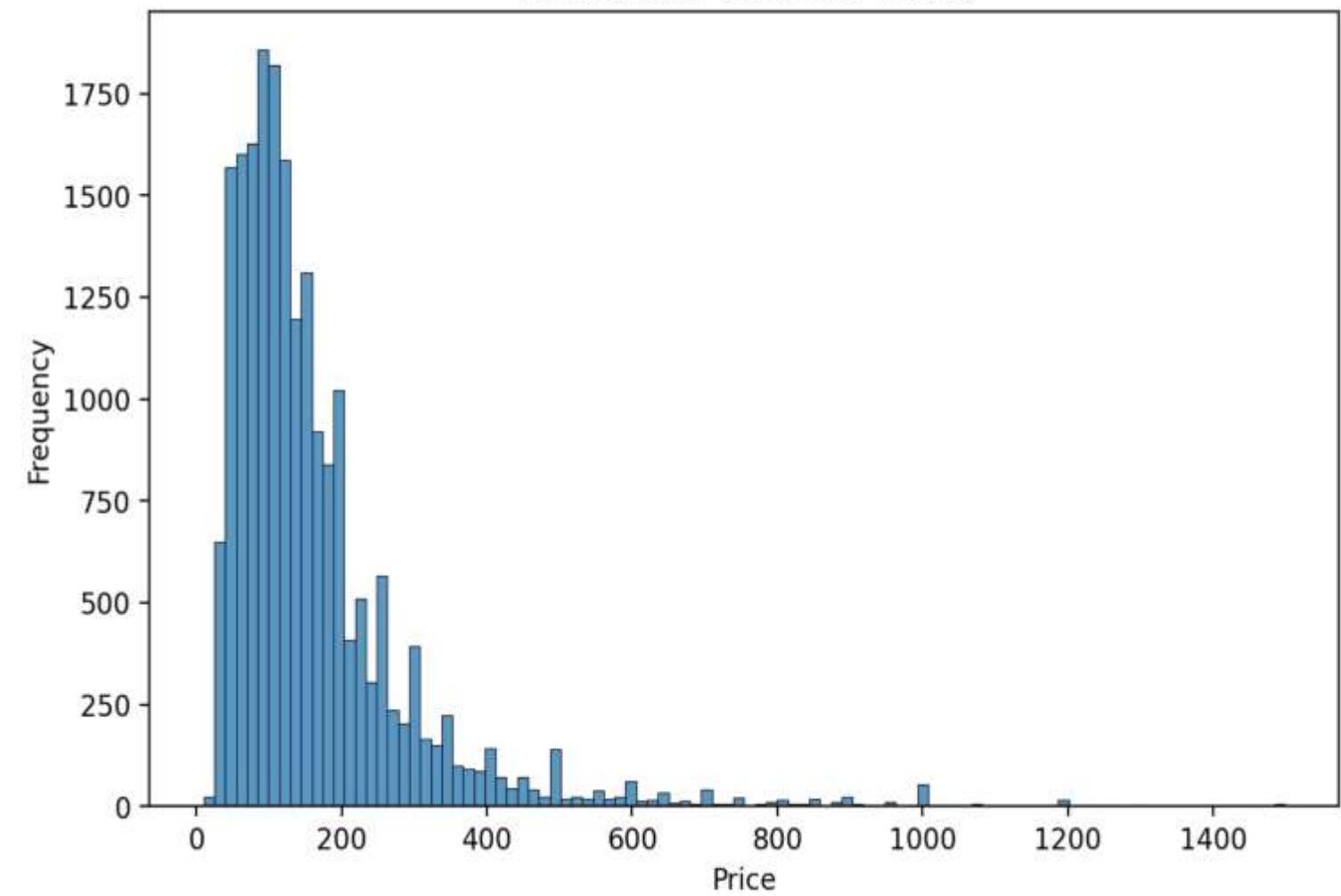
Show first 10 rows of dataset

		name	host_id	host_name	neighbourhood_group	neighbour
0	1,312,228	Rental unit in Brooklyn · ★5.0 · 1 bedroom	7,130,382	Walter	Brooklyn	Clinton Hill
1	45,277,537	Rental unit in New York · ★4.67 · 2 bedrooms · 1 bed · 1 bath	51,501,835	Jeniffer	Manhattan	Hell's Kitchen
2	9710000000000000000	Rental unit in New York · ★4.17 · 1 bedroom · 2 beds · 1 bath	528,871,354	Joshua	Manhattan	Chelsea
3	3,857,863	Rental unit in New York · ★4.64 · 1 bedroom · 1 private bath	19,902,271	John And Catherine	Manhattan	Washington Heights
4	40,896,611	Condo in New York · ★4.91 · Studio · 1 bed · 1 bath	61,391,963	Stay With Vibe	Manhattan	Murray Hill
5	49,584,983	Rental unit in New York · ★5.0 · 1 bedroom · 1 bath	51,501,835	Jeniffer	Manhattan	Hell's Kitchen
6	45,277,537	Rental unit in New York · ★4.67 · 2 bedrooms · 1 bed · 1 bath	51,501,835	Jeniffer	Manhattan	Hell's Kitchen
7	9710000000000000000	Rental unit in New York · ★4.17 · 1 bedroom · 2 beds · 1 bath	528,871,354	Joshua	Manhattan	Chelsea
8	3,857,863	Rental unit in New York · ★4.64 · 1 bedroom · 1 private bath	19,902,271	John And Catherine	Manhattan	Washington Heights
9	40,896,611	Condo in New York · ★4.91 · Studio · 1 bed · 1 bath	61,391,963	Stay With Vibe	Manhattan	Murray Hill

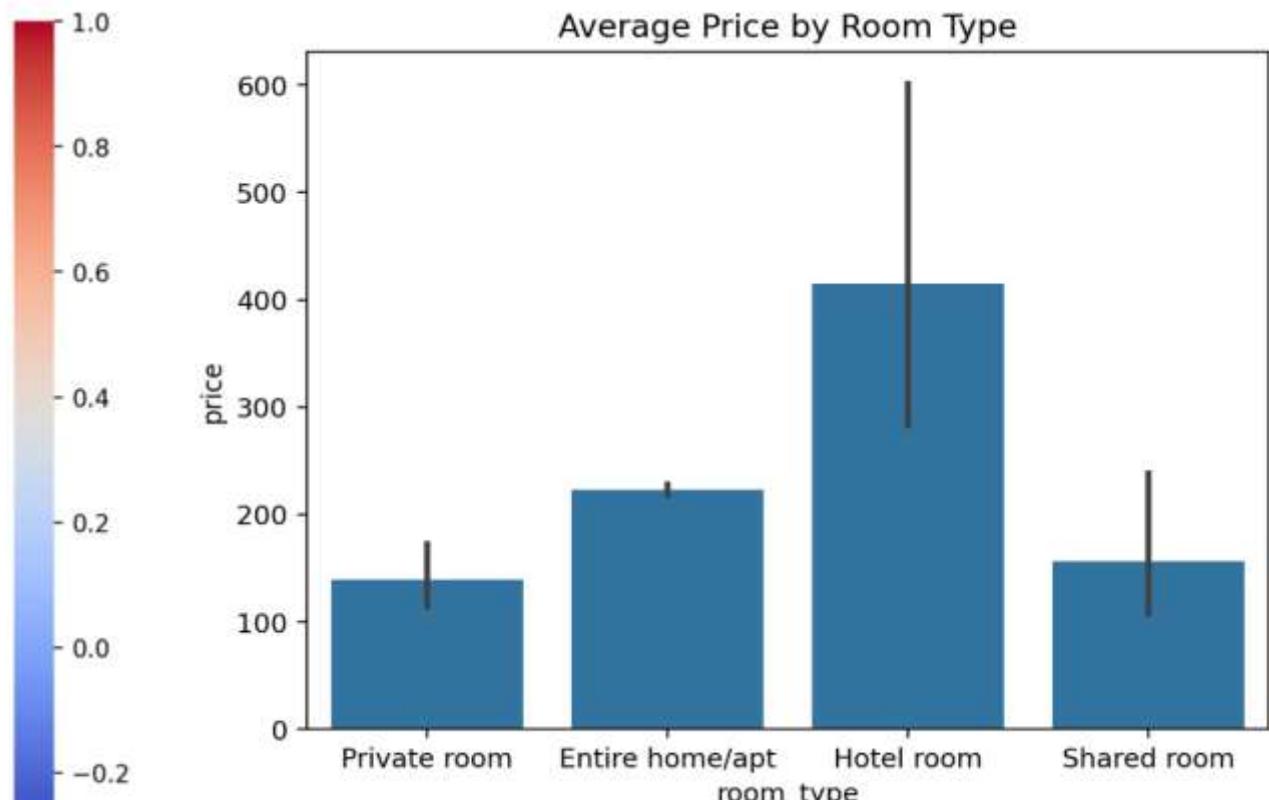
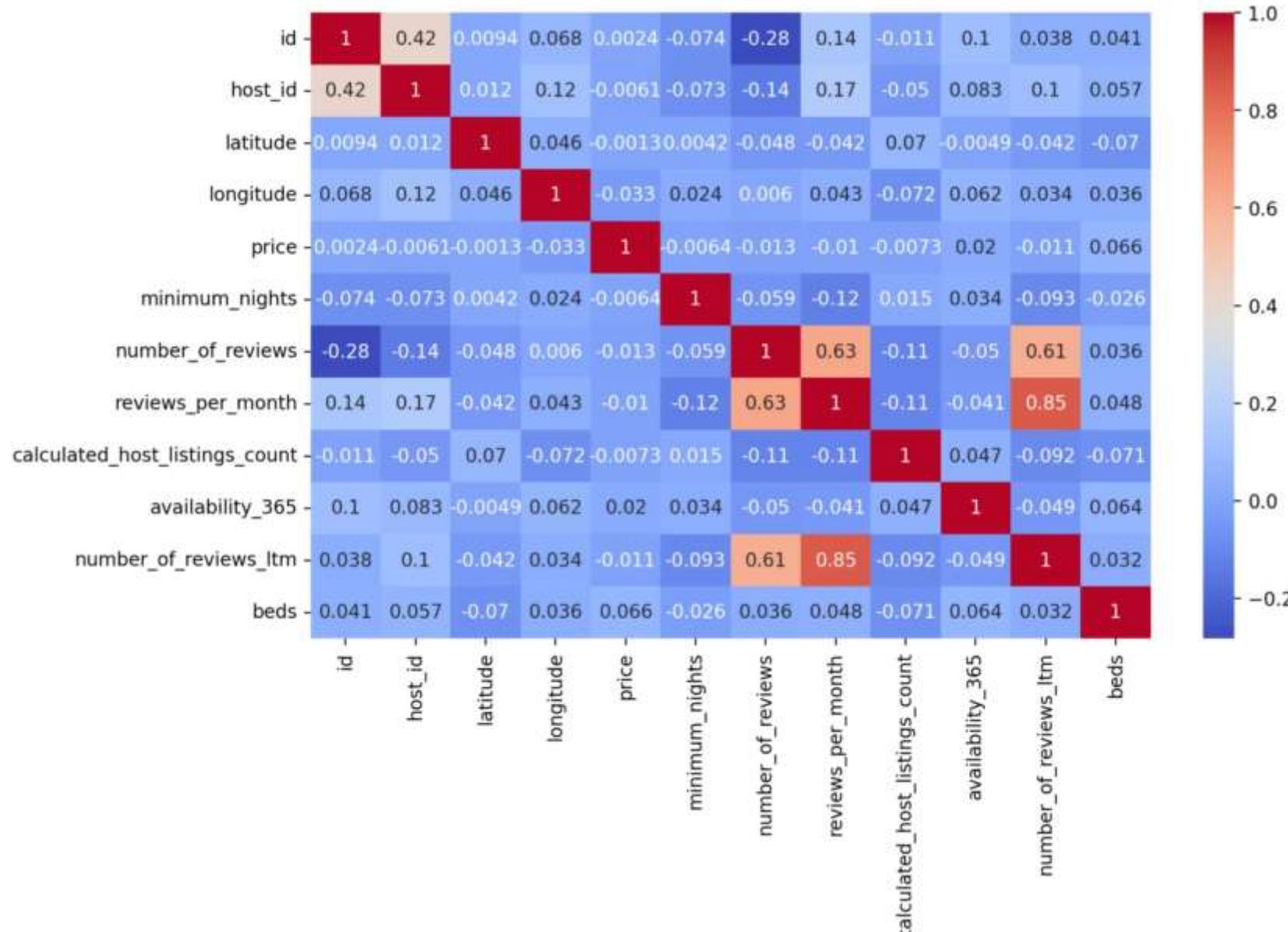
Geographical Distribution by Neighbourhood Group



Distribution of Airbnb Prices



Displays correlation between numerical columns such as price, reviews, and availability.



# Real-World Applications and Value

This project offers tangible benefits across multiple sectors.

## Traveler Benefits

Assists travelers in easily identifying affordable stays and optimizing their budget, ensuring value for money.

## Host Pricing Strategy

Empowers hosts to set **competitive and dynamic pricing** based on data-driven market analysis and local conditions.

## Tourism & Real Estate

Provides valuable data for urban planners, real estate analysts, and tourism agencies to understand market trends.

# Challenges Encountered and Solutions

Developing a robust data product required overcoming several technical and data quality obstacles.

## Data Quality & Consistency

**Challenge:** Missing values (NaN) and inconsistent feature formats.

**Solution:** Rigorous data preprocessing, imputation, and feature engineering in Pandas.

## Model Accuracy

**Challenge:** Initial model exhibited high error rates.

**Solution:** Improved through feature selection, hyperparameter tuning, and multiple validation iterations using Scikit-learn.

## Price Outliers

**Challenge:** Presence of extreme price values affecting visualizations and prediction accuracy.

**Solution:** Applied logical filtering (e.g., excluding listings above a certain price threshold) to stabilize trends and ensure realistic results.

## Deployment Complexity

**Challenge:** Configuration and dependency issues during Render deployment.

**Solution:** Created a proper requirements.txt file, adjusted file paths, and configured environment variables for stable deployment.

# Future Scope

To elevate the project into a comprehensive, production-ready solution, we plan to implement several advanced features.



## Advanced ML Models

Integrate and benchmark sophisticated algorithms like **XGBoost** and **Gradient Boosting** for improved prediction accuracy.



## Geographic Expansion

Scale the project to include listings from other major cities or countries, creating a global prediction service.



## Real-Time Data Feeds

Integrate with actual Airbnb APIs to fetch and analyze real-time data, enhancing market relevance.



## UI/UX Enhancement

Enhance the Streamlit dashboard with interactive visualizations, intuitive layouts, and responsive design for improved user engagement.

## Access the Live Deployment

The complete interactive dashboard is deployed and accessible via Render:

[Project Dashboard Link](#)