EDA on Airbnb Listings

Presented by: **Nihar Suman**

Guided by: Miss 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

This project focuses on analyzing and predicting Airbnb listing prices using Python.
The main goal is to understand how various factors such as **location, room type, number of reviews, and availability** influence the overall price of a listing.
Through **data cleaning, visualization, and pattern identification**, the dataset was explored to uncover useful insights and trends.
After analysis, a **machine learning model** was developed to predict Airbnb prices accurately based on key input features.
Finally, an **interactive Streamlit web application** was created, allowing users to explore the data visually and generate real-time price predictions in an easy-to-use interface.

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.

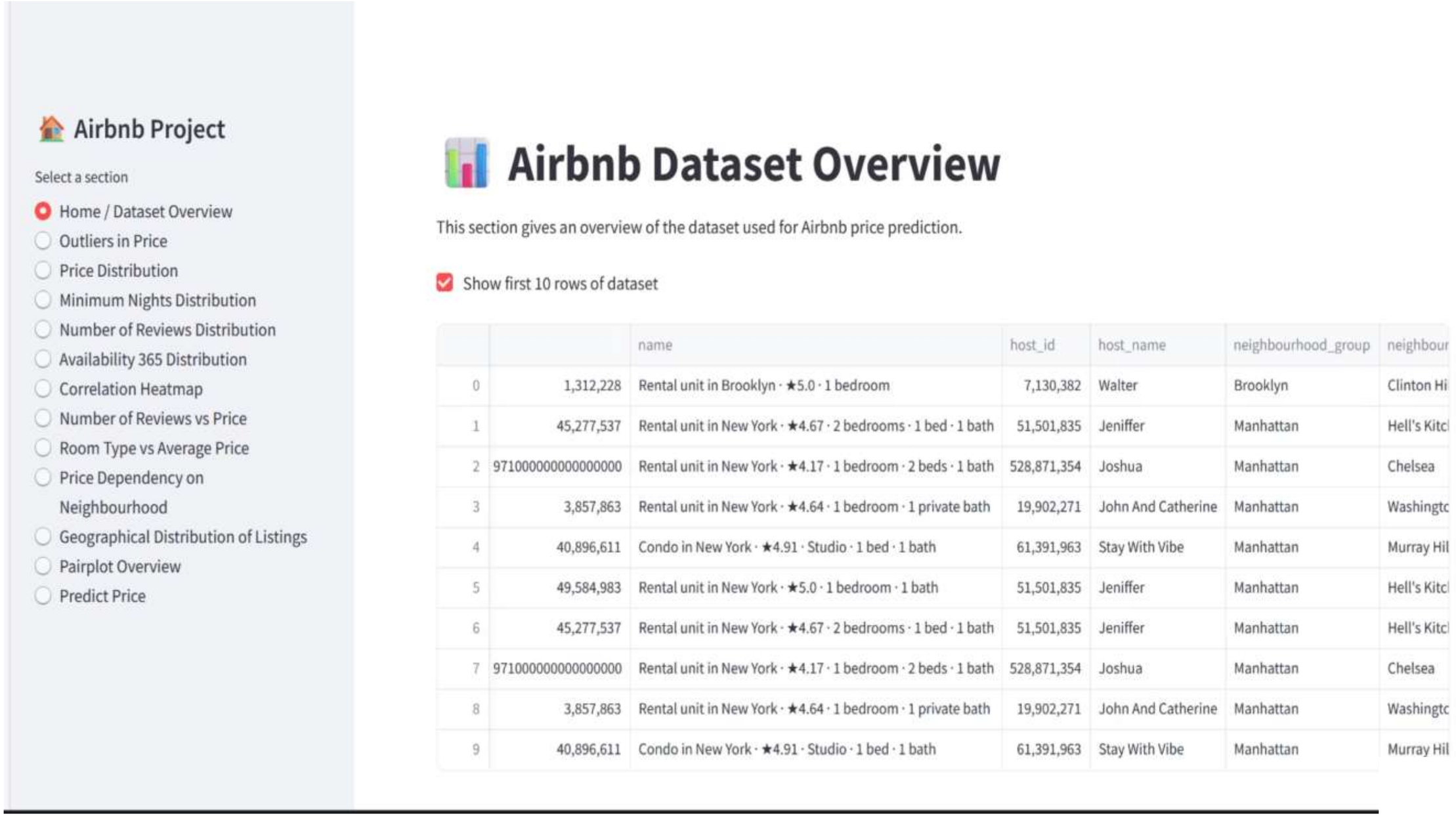
Deployment & Interface

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

Data & Environment

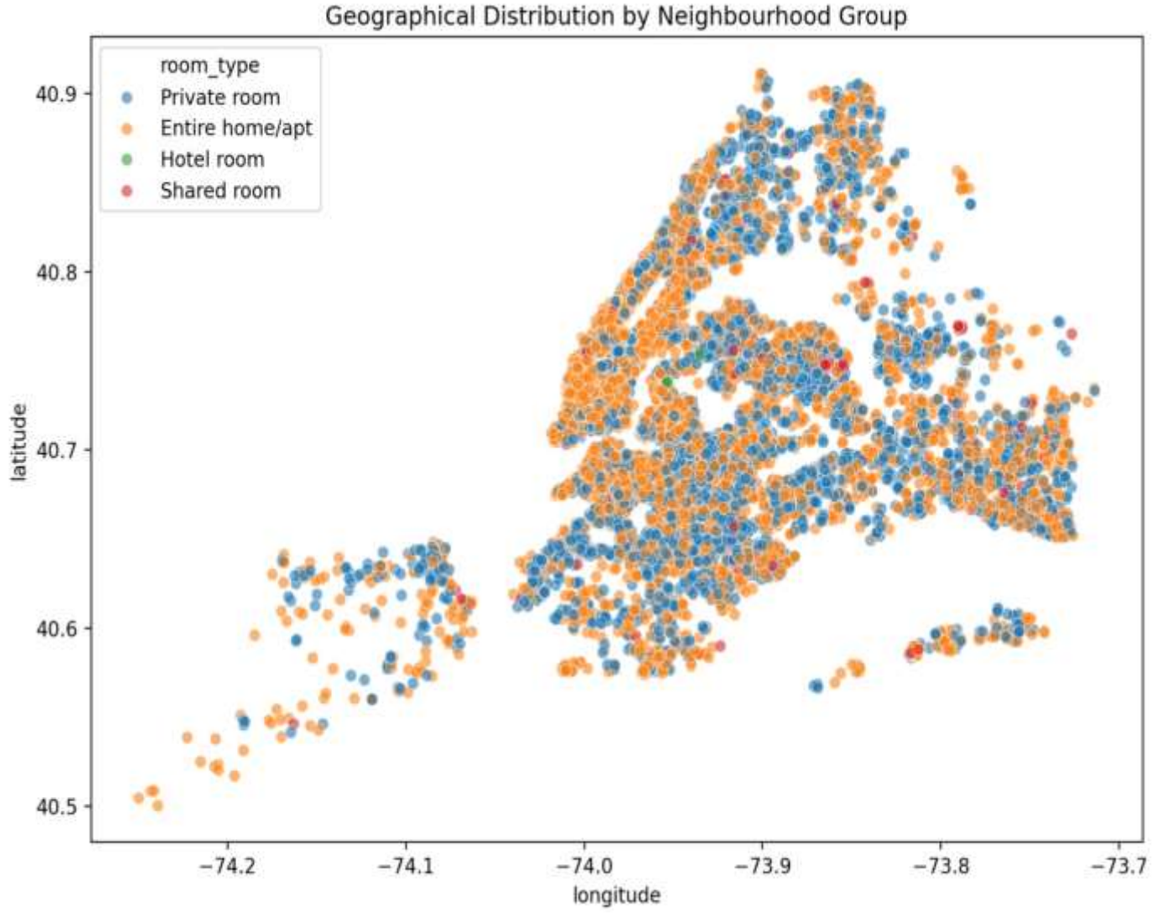
Dataset: **New York Airbnb Listings 2024** provides rich features for analysis. Development in **Jupyter Notebook** and **VS Code**, version control with **GitHub**.

Project Working

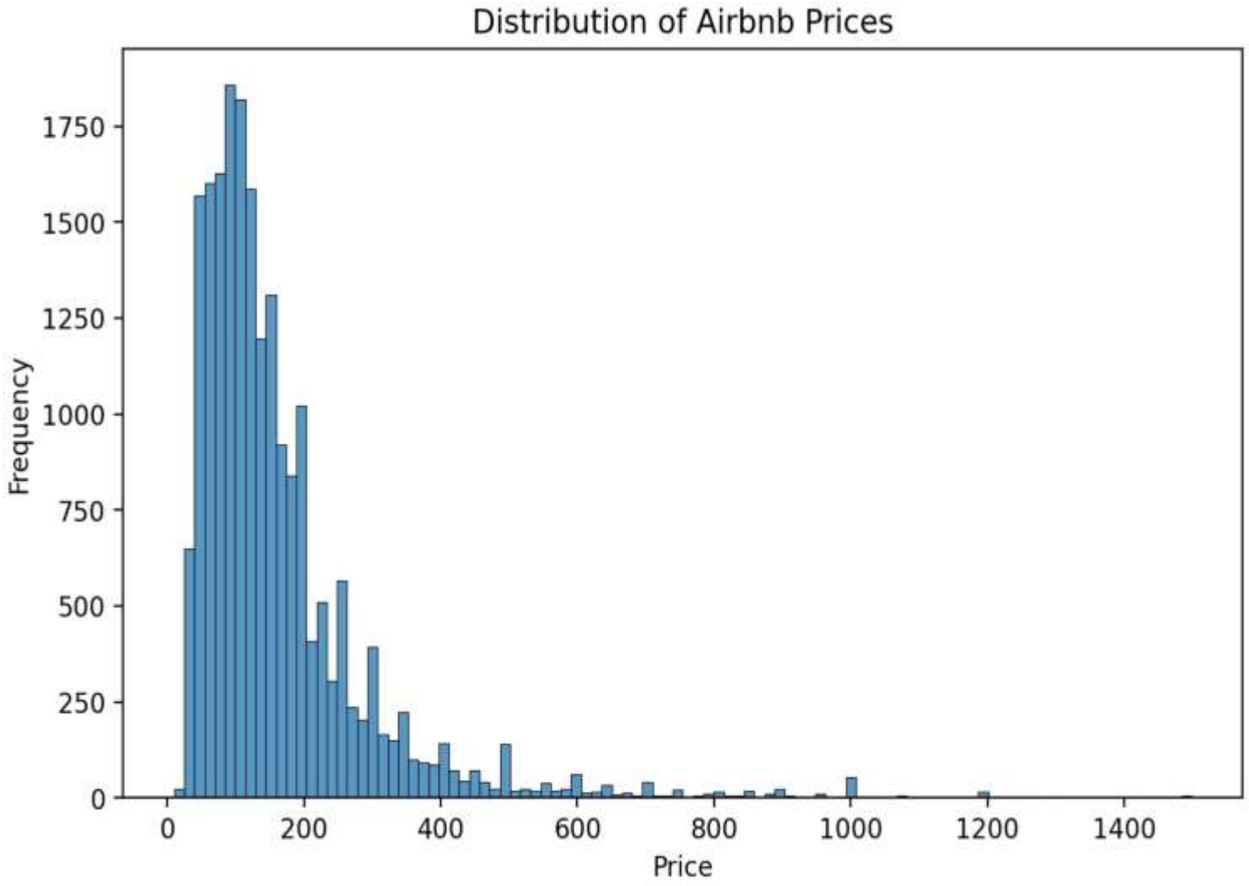


		name	host_id	host_name	neighbourhood_group	neighbour
0	1,312,228	Rental unit in Brooklyn · ★5.0 · 1 bedroom	7,130,382	Walter	Brooklyn	Clinton Hi
1	45,277,537	Rental unit in New York · ★4.67 · 2 bedrooms · 1 bed · 1 bath	51,501,835	Jeniffer	Manhattan	Hell's Kitc
2	971000000000000000	Rental unit in New 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	Washingtc
4	40,896,611	Condo in New York · ★4.91 · Studio · 1 bed · 1 bath	61,391,963	Stay With Vibe	Manhattan	Murray Hil
5	49,584,983	Rental unit in New York · ★5.0 · 1 bedroom · 1 bath	51,501,835	Jeniffer	Manhattan	Hell's Kitc
6	45,277,537	Rental unit in New York · ★4.67 · 2 bedrooms · 1 bed · 1 bath	51,501,835	Jeniffer	Manhattan	Hell's Kitc
7	971000000000000000	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	Washingtc
9	40,896,611	Condo in New York · ★4.91 · Studio · 1 bed · 1 bath	61,391,963	Stay With Vibe	Manhattan	Murray Hil

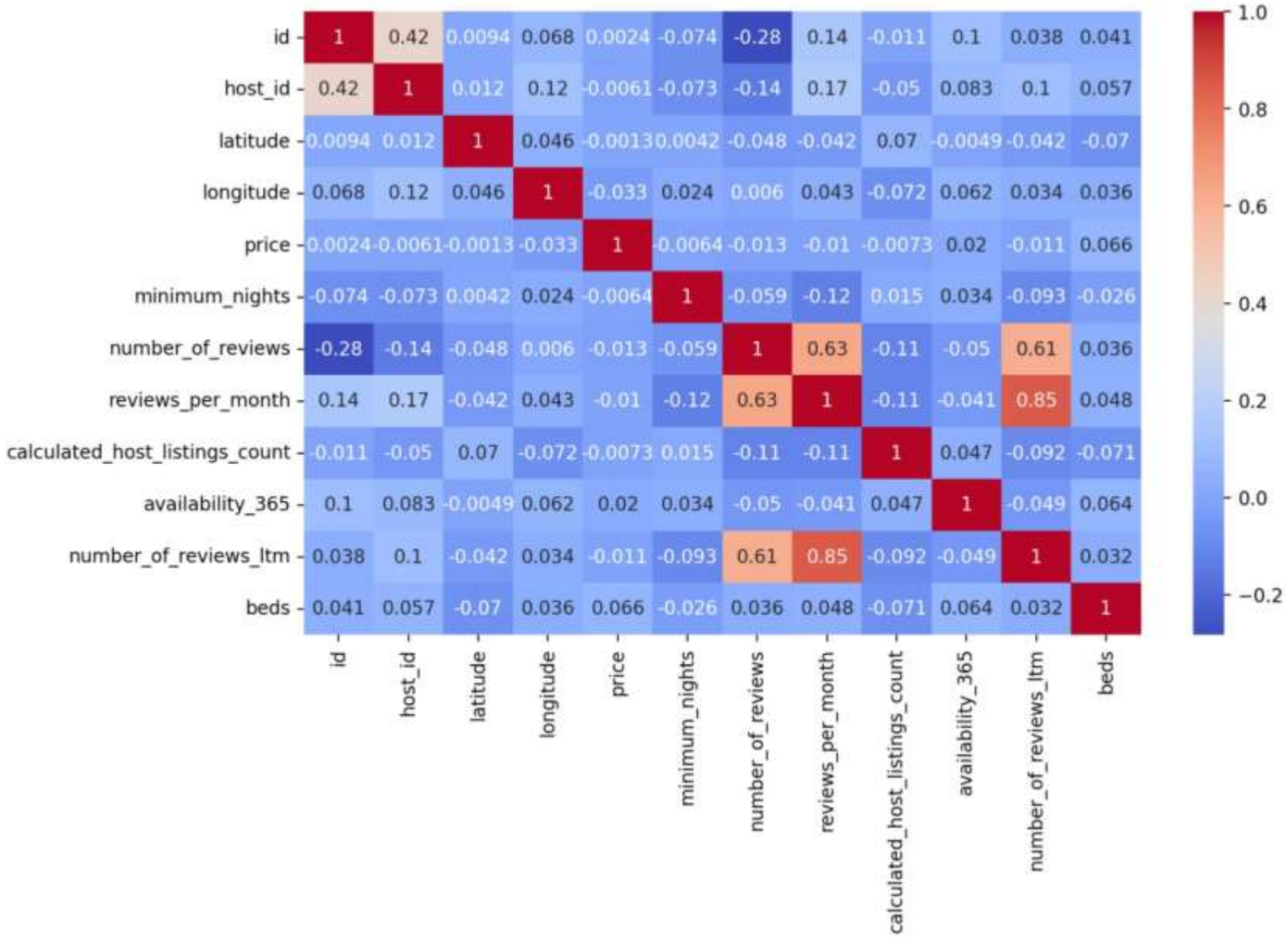
Distribution of Airbnb listings across city areas by room type.



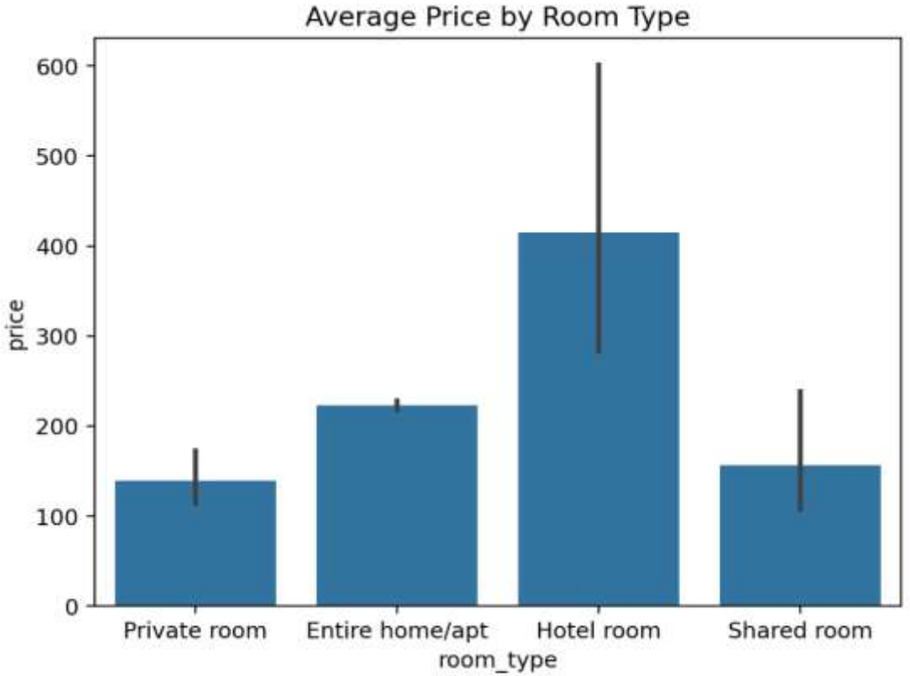
Price distribution of Airbnb listings — most are in the lower price range



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



Average Airbnb price by room type — hotel rooms cost the most, shared rooms the least.



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.

Project Link

Access the Live Deployment:

The complete interactive dashboard is deployed and accessible via **Streamlit Cloud**:

Working Link:- https://airbnb-listings-nihar18.streamlit.app/

THANK YOU