Airbnb Dynamic Pricing Recommendation Engine

1. Introduction

In the competitive vacation rental industry, pricing plays a critical role in maximizing host revenue while maintaining competitive positioning. Airbnb hosts often struggle to determine optimal listing prices based on multiple factors such as location, seasonality, property type, guest capacity, and reviews. This project focuses on building a data-driven Airbnb Dynamic Pricing Recommendation Engine capable of predicting ideal listing prices using machine learning and visualizing key trends using Power BI.

2. Abstract

This project analyzes the Airbnb Price Prediction Dataset from Kaggle and leverages machine learning techniques to recommend optimal listing prices. After comprehensive exploratory data analysis (EDA) and feature engineering, a Linear Regression model was trained to predict listing prices. The model output was stored in an output file, and an interactive Power BI dashboard was developed to display insights, pricing trends, and high-performing locations. The solution demonstrates how data-driven pricing strategies can help hosts maximize earnings while maintaining competitiveness.

3. Tools Used

Tool	Purpose
Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-Learn)	Data cleaning, EDA, model training
Power BI	Dashboard creation and visualization
Excel / CSV	Data inspection and final result storage
Joblib	Saving trained model
Kaggle	Dataset source

4. Steps Involved in Building the Project

Step 1: Data Collection

Used the **Airbnb Price Prediction Dataset** sourced from **Kaggle**, containing listing details such as price, room type, property type, city, host ratings, number of reviews, and location coordinates.

Step 2: Data Cleaning

- Removed duplicate records.
- Cleaned and normalized the `price` column by removing currency symbols.
- Converted date fields such as `first_review`, `last_review`, and `host_since` into datetime format.
- Converted boolean fields such as `instant_bookable` and `host_identity_verified` into binary values (1/0).
- Filled missing numerical values such as `review_scores_rating` and `reviews_per_month` with appropriate imputation strategies.
- Handled outliers by removing listings with unrealistic price points.

Step 3: Exploratory Data Analysis (EDA)

- Analyzed price distributions and applied log transformation to handle skewness.
- Compared price variations by room type, property type, and city.
- Identified correlation between features such as reviews, ratings, number of beds, and price.
- Observed seasonal and location-based pricing trends.

Step 4: Feature Engineering

- Engineered features such as log-transformed price, month of last review, and availability indicators.
- Derived review-based insights and location-based trends for model training.

Step 5: Model Development

- Applied a **Linear Regression model** to predict the price of each listing.
- Trained the model on preprocessed features and evaluated performance using MAE and R^2 score.
- The trained model demonstrated strong capability in explaining price variations based on listing characteristics.
- Final predictions were generated and stored in an output CSV file.

Step 6: Power BI Dashboard Development

- Imported cleaned dataset and model predictions into Power BI.
- Designed visualizations for price trends by city, room type, and review ratings.
- Developed an interactive dashboard with slicers for filtering by city, room type, and capacity.

- Key insight: **Listings in New York had the highest predicted prices, followed by Los Angeles**.
- Average review ratings were consistently around **4million+**, indicating strong customer experiences.

5. Conclusion

The Airbnb Dynamic Pricing Recommendation Engine successfully predicts optimal listing prices based on listing characteristics using a Linear Regression model. The Power BI dashboard provides actionable visual insights, helping users compare pricing across cities, room types, and review levels. By accurately forecasting optimal pricing, this solution empowers property owners to make informed decisions, increase revenue, and stay competitive in the dynamic rental market.

Future enhancements may include integrating advanced machine learning models like XGBoost or Gradient Boosting, seasonal demand adjustments, competitor-based pricing strategies, and automated real-time pricing deployment via API.