**High-Level Design Document for Airbnb**

**Document Version Control**

| **Version** | **Date** | **Description** | **Author** |
| --- | --- | --- | --- |
| 1.0 | 2024-09-23 | Initial Draft | [Mahesh Shetty] |
| 2.0 | 2024-10-10 | Updates on Scope & KPIs | [Hrishikesh Panchmukh] |

**Abstract**

This project analyzes Airbnb listing activity and metrics in San Diego for 2019. It aims to uncover insights about host earnings, geographical trends, pricing strategies, and review impacts to better understand what drives success in the short-term rental market. The study will answer key questions around top-earning hosts, price-location dynamics, booking frequency by neighborhood, and relationships between price and amenities. Findings will be used to identify patterns, KPIs, and meaningful relationships between attributes.

**Introduction**

Since its launch in 2008, Airbnb has expanded traveling opportunities by offering more personalized and unique accommodation options. This dataset provides critical information to explore host activities, geographical availability, guest reviews, and pricing strategies within San Diego for 2019.

The goal of this project is to analyze these attributes to find correlations, predict trends, and support decision-making for future hosts and guests. Key questions include identifying top earners, examining the impact of location and amenities on price, and understanding the role of reviews in influencing booking success.

**Why This High-Level Design Document?**

This document provides a structured overview of the project to ensure alignment across stakeholders. It defines the scope, tools, objectives, and methodologies, helping ensure a smooth and consistent flow from data exploration to insights and deployment.

**Scope**

This project will focus on:

1. Identifying key patterns in Airbnb listings in San Diego.
2. Drawing insights on the relationship between price, reviews, and amenities.
3. Examining geographical patterns to understand which neighborhoods perform best.
4. Establishing KPIs for host success and market trends.
5. Preparing visualizations for better interpretation and communication.

**General Description**

**Product Perspective & Problem Statement**

Understanding Airbnb listing data is vital for hosts seeking to optimize earnings, for guests making better booking decisions, and for analysts tracking the health of the rental market. Housing prices are a major economic indicator, and understanding the pricing dynamics within short-term rental markets is critical for hosts, guests, and policymakers.

The goal is to find meaningful relationships between attributes (e.g., location, price, and reviews) and present them through KPIs and visual dashboards.

**Tools Used**

1. **Python**: Data cleaning, transformation, and analysis (Pandas, NumPy, Matplotlib, Seaborn).
2. **Power BI / Tableau**: Creating visual dashboards to present insights and KPIs.
3. **Jupyter Notebooks**: Interactive data exploration.
4. **GitHub**: Code version control and documentation.

**Design Details**

**Functional Architecture**

1. **Data Collection & Cleaning**
   * Handle missing values (e.g., reviews, prices, availability).
   * Remove outliers for better statistical modeling.
2. **Exploratory Data Analysis (EDA)**
   * Descriptive statistics and visualization (distribution of prices, reviews, and availability).
3. **Research Questions & Hypothesis Testing**
   * Identify top-earning hosts.
   * Assess correlation between location and price.
   * Examine relationships between reviews, quality, and pricing.
4. **Feature Engineering**
   * Create new features like price per guest or booking frequency by month.
5. **Visualization & Reporting**
   * Create dashboards with key metrics: Average price per neighborhood, Top hosts, etc.
6. **Optimization & Recommendations**
   * Suggest strategies for improving host earnings based on findings.

**Optimization**

* **Outlier Detection & Removal**: Filter out extreme values in price and reviews for better model performance.
* **Feature Selection**: Focus on attributes with high predictive power (e.g., location, amenities).
* **Data Scaling & Transformation**: Normalize variables where appropriate for consistent analysis.

**KPIs (Key Performance Indicators)**

1. **Host Performance**
   * **Top Earning Hosts**: Total earnings by host.
   * **Average Earnings per Month**: Relationship between monthly earnings and pricing strategy.
2. **Neighborhood Insights**
   * **Neighborhood with Most Bookings**: Frequency of listings booked in each area.
   * **Average Price per Neighborhood**: Identify high-demand vs. low-demand areas.
3. **Pricing Dynamics**
   * **Price vs. Amenities**: Impact of amenities on listing price.
   * **Price vs. Location**: How proximity to attractions influences prices.
4. **Review Metrics**
   * **Average Review Score per Listing**: Relationship between reviews and bookings.
   * **Price vs. Quality of Reviews**: Do higher-rated properties charge more?
5. **Overall Market Trends**
   * **Occupancy Rate**: Percentage of available nights booked per listing.
   * **Booking Frequency**: Monthly trends in bookings and cancellations.

**Deployment**

1. **Reporting & Dashboards**
   * Develop interactive dashboards in Power BI/Tableau for host performance and market insights.
2. **Documentation**
   * Maintain project documentation and notebooks on GitHub.
3. **Presentation & Recommendations**
   * Prepare a summary report and present findings to stakeholders with actionable recommendations.

**Conclusion**

This high-level design document lays the foundation for analyzing the 2019 Airbnb data in San Diego. It outlines the project’s objectives, methodology, KPIs, and tools to ensure a comprehensive analysis of the dataset. Through visualizations and insights, the project aims to provide actionable recommendations for improving host strategies and guest experiences.