Airbnb Data Analysis

High Level Design Document

Divyansh Verma
Data Analyst Intern
iNeuron AI

Table of Contents

Document Version Control

Version	Date	Author	Description
1.0	2024-07-15	Divyansh Verma	Initial Draft

Abstract

This document provides a high-level design for the AirBnB data analysis project. The project involves analyzing the listing activity and metrics for AirBnB in Amsterdam, Netherlands, for the year 2019. The analysis aims to derive insights regarding hosts, neighborhoods, reviews, and prices using various data visualization and statistical techniques.

1. Introduction

The primary objective of this project is to perform a comprehensive data analysis of AirBnB listings in Amsterdam for 2019. The analysis focuses on several research questions to understand the dynamics of the AirBnB market in this region, including top earners, neighborhood popularity, and the relationship between prices, amenities, and customer satisfaction.

1.1. Why this High-Level Design Document?

This document aims to provide a clear and structured overview of the project's design. It outlines the scope, tools, design details, and the functional architecture needed to achieve the project objectives. The document serves as a guide for stakeholders to understand the project framework and the methodology used for data analysis and visualization.

1.2. Scope

The scope of this project includes:

- Data collection and cleaning.
- Exploratory Data Analysis (EDA) to understand the dataset.
- Analyzing host earnings, neighborhood popularity, review quality, and price dynamics.
- Visualizing the data using Python and PowerBI.
- Providing insights and recommendations based on the analysis.

2. General Description

The project involves analyzing a dataset of AirBnB listings in Amsterdam, Netherlands, for the year 2019. The dataset includes various attributes such as host information, neighborhood details, review counts, price, and amenities. The analysis aims to answer specific research questions related to hosts, neighborhoods, reviews, and prices.

2.1. Product Perspective & Problem Statement

Since 2008, AirBnB has provided a unique way of experiencing travel. This project aims to analyze the listing activity and metrics for AirBnB in Amsterdam to derive insights into host earnings, neighborhood popularity, and the relationship between prices, amenities, and customer satisfaction.

2.2. Tools Used

- **Programming Languages:** Python
- **Libraries:** Pandas, NumPy, Matplotlib, Seaborn, Statsmodels
- **Visualization Tools:** PowerBI
- **Data Source:** CSV file hosted on GitHub

3. Design Details

3.1. Data Collection and Cleaning

- Importing the dataset from GitHub.
- Checking for null values and handling them appropriately.
- Converting data types if necessary.

3.2. Exploratory Data Analysis (EDA)

- Analyzing the distribution of each column.
- Calculating summary statistics.
- Identifying unique values and data inconsistencies.

3.3. Analysis by Research Questions

1. Regarding the Host

- Calculate total earnings per host.
- Identify top 10 earners.
- Analyze the correlation between average price per listing and total earnings.

2. Regarding the Neighborhood

- Count the number of bookings per neighborhood.
- Identify top neighborhoods by bookings.
- Calculate the average price per neighborhood.

3. Regarding the Reviews

- Analyze the correlation between overall satisfaction and price.

4. Regarding Price

- Analyze the relationship between price and the number of bedrooms.
- Create heatmaps to visualize price and bookings by location.

3.4. Data Visualization

- Create various plots using Matplotlib and Seaborn to visualize the data.
- Use PowerBI to create interactive maps and dashboards.

3.5. Functional Architecture

The functional architecture consists of the following components:

- **Data Ingestion:** Loading the dataset from GitHub.
- **Data Cleaning:** Handling missing values and data type conversions.
- **Data Analysis:** Performing calculations and statistical analysis.
- **Data Visualization:** Creating plots and dashboards for visual representation of insights.

3.6. Optimization

- Using efficient Pandas operations to handle large datasets.
- Leveraging vectorized operations for faster computations.
- Utilizing PowerBI for interactive and optimized data visualizations.

4. KPIs

- Number of top-earning hosts identified.
- Correlation between price and earnings.
- Number of bookings per neighborhood.
- Correlation between review quality and price.
- Insights derived from heatmaps and visualizations.

5. Deployment

- The analysis code will be executed in a Jupyter Notebook environment.
- Visualizations will be shared through PowerBI dashboards.
- The final report will be documented and shared with stakeholders.

By following this high-level design document, the project aims to deliver a comprehensive analysis of the AirBnB listings in Amsterdam, providing valuable insights into the market dynamics and helping stakeholders make informed decisions.