HIGH LEVEL DESIGN (HLD)

Booking Data Analysis

(AirBnB Booking Analysis)

Written By / Author	Gaurav Rajgor
Document Version	HLD-V1.0
Last Revised Date	12/06/2022

Document Version Control

Date Issued	Version	Description	Author
12/06/2022	HLD-V1.0	First Version of Complete LLD	Gaurav Rajgor

Contents

Abstract

Airbnb is a corporation based in the United States that operates an online marketplace for lodging, primarily homestays for holiday rentals, and tourism activities. It essentially connects travelers with local hosts who want to rent out their homes with others looking for lodging in that area. This platform, on the other hand, allows hosts to list their available space and earn extra income through rent, while also allowing travelers to book unique homestays from local hosts, saving them money and providing them the opportunity to interact with locals.

The travel industry is advancing with the role of Data Science and Analytics in a world of rising new technology and innovation. Data analysis can help them understand their business in a new light and improve the quality of service by identifying the company's weak points. This study demonstrates how various analyses can help businesses make better decisions and analyze customer trends and satisfaction, which can lead to new and improved products and services. Various analyses, such as exploratory data analysis and descriptive analysis, were performed on a variety of use cases to obtain key insights from this data, which will be used to make business decisions.

1 Introduction

This document will be used for documenting High-level designs of project.

1.1 Purpose of the Document

The purpose of this plan is to

- Describe different design approaches.
- Describe different analysis approaches based on variety of Use Cases.
- Describe third party components/tools required for the system.
- Present complete Process Flow followed for this project.

1.2 Objective of HLD

- To provide an overview of the entire system.
- To provide introduction of Problem Perspective & Statement, Data Requirements, Tools used and many more.
- To provide a module-wise breakup of the entire system.

1.3 Scope of HLD

The High-Level Design documentation covers the structure of the system as the application/database architecture, application flow and technology architecture.

2 General Description

2.1 Listings Product Perspective & Problem Statement

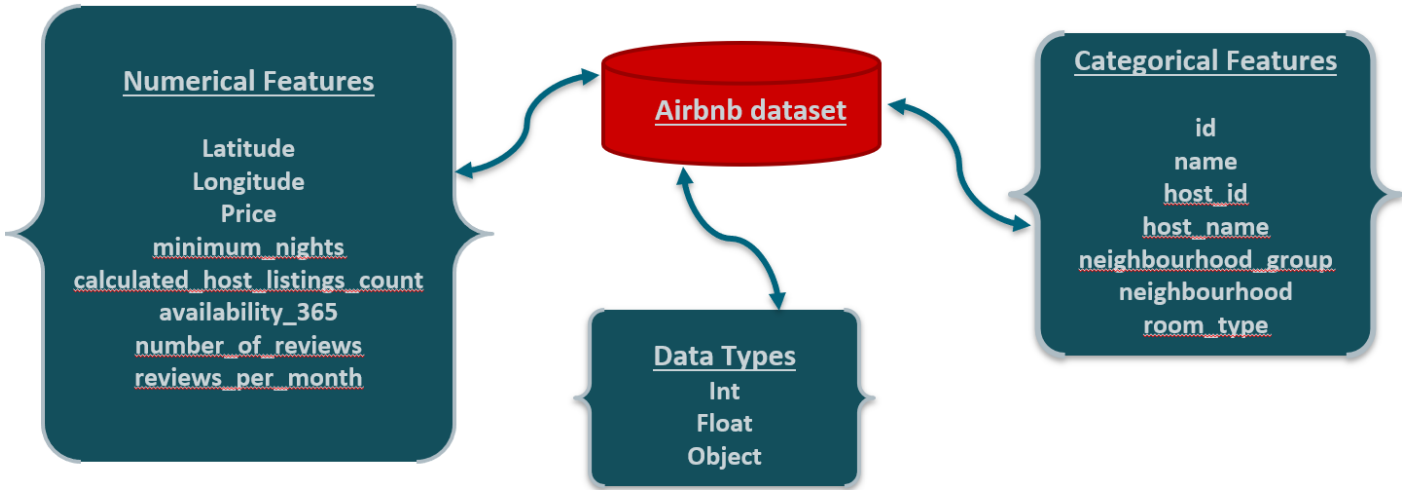
Since the last few decades, the travel industry has been an essential reflection of the economy, and Airbnb housing pricing ranges are of great importance to both Hosts and Travelers. We are examining many components with diverse use cases in this project, which includes many features of Airbnb listings. It not only aids in identifying the meaningful links between features, but it also allows us to do our own research and present our findings.

The project's goal is to do exploratory data analysis, data pre-processing, data cleaning and imputation, and finally, to apply various Data Visualization techniques to gain meaningful insight from the given data. This project aims apply some amazing Python Libraries such as Plotly and Seaborn which will give a boost to our visual understanding of the data.

2.2 Data Requirements

Data Requirement completely depend on our problem.

- In this project, to perform analysis, we are using datasets that are provided by ALMABETTER.
- We make a use of those different datasets as per the requirement and the problem statement.
- The features which are taken into consideration are:
- Some of the important features are:



Column	Description
id	listing ID
name	name of the listing
host_id	host ID
host_name	name of the host
neighbourhood_group	location
neighbourhood	area
latitude	latitude coordinates
longitude	longitude coordinates
room_type	listing space type
price	
minimum_nights	amount of nights minimum
number_of_reviews	number of reviews
last_review	latest review
reviews_per_month	number of reviews per month
calculated_host_listings_count	amount of listing per host
availability_365	number of days when listing is available for booking

2.3 Tools Used

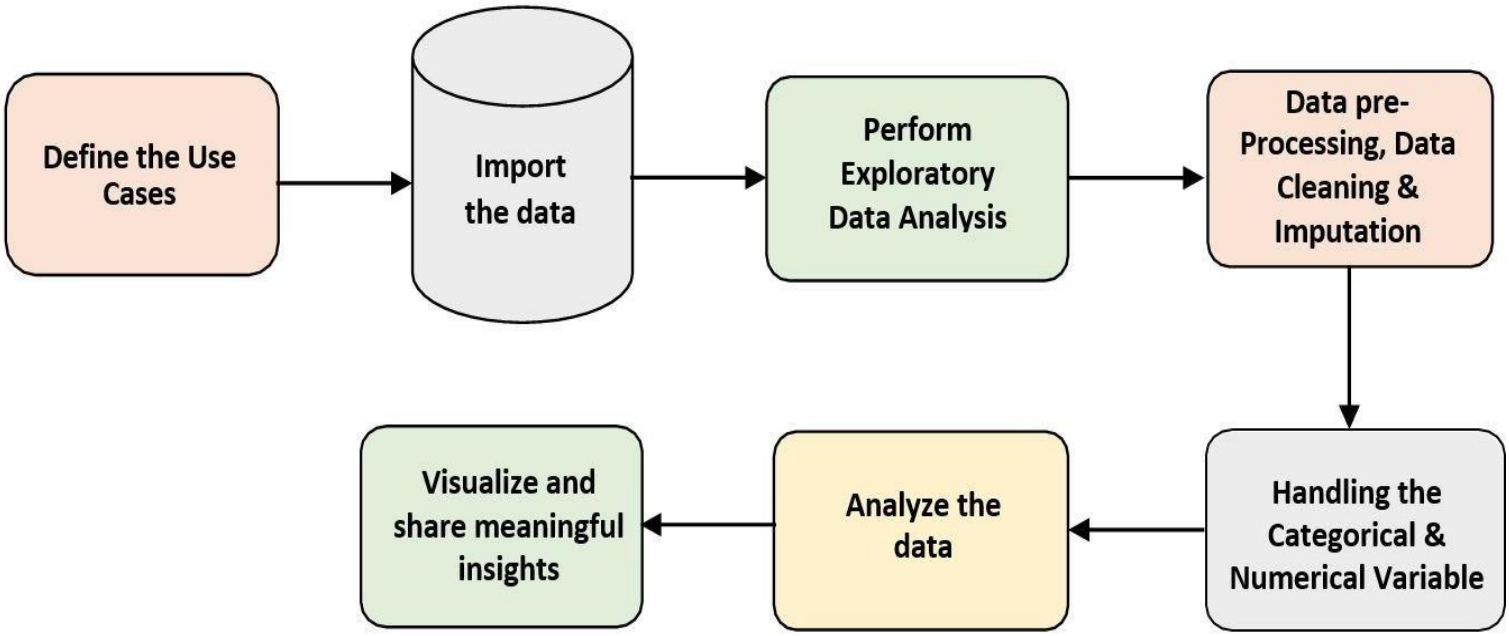
- Jupyter Notebook is used as IDE.
- Pandas and NumPy are used for Data Manipulation & Pre-processing and Mathematical functions respectively.
- Exploratory data analysis is automated by dataprep.
- For visualization of the plots, Matplotlib, Seaborn, Plotly are used.
- GitHub is used as version control system

2.4 Constraints

The analysis must be user friendly, the code must be nice and clear, and EDA should be automated as much as feasible to save time. Furthermore, users should not be required to have any of the coding knowledge as the insights they are looking for are mentioned in-detail with respective visuals.

3 Architecture

3.1 Process Flow



3.2 Error Handling / Exception Handling:

We designed this project in such a way that the entire script is tested and run multiple times to ensure that no errors occur during the process flow.
In addition, we have removed unnecessary warnings to avoid confusion by using the filter warnings class from the warning's module.

4 Conclusion

Approximately sixteen different use cases are analyzed for the given dataset in this analysis project to help make better business decisions and analyze customer trends and satisfaction, which can lead to new and better products and services. It was discovered that the majority of bookings were made for "Mission Bay," with around "22 percent," followed by "Pacific Beach," "La Jolla," with "15.5 percent," and "12 percent," respectively. We have also analyzed that, and we have Total "11091" which are both "Budget Hotel" and "Affordable."

Furthermore, we discover the Top Earners (Host), the relationship between Monthly Earning and Prices, Price Comparison in Terms of Property Type and Room Type, and Guests' Preference in Terms of Property Type and Room Type. Furthermore, we examined the Maximum Number of Bookings, Customer Reviews/Comments, Airbnb Amenities, and many other factors.