High Level Design (HLD) Airbnb Data Analysis

Revision Number: 1.0

Last date of revision: 19/06/2023

Anugrah Mohan

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| **19th June 2023** | 1.0 | First Version of Complete HLD | Anugrah Mohan |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Contents**

[Document Version Control 2](#_bookmark0)

[Abstract 3](#_bookmark1)

1. [Introduction 5](#_bookmark2)
   1. [Why this High-Level Design Document? 5](#_bookmark3)
   2. [Scope 5](#_bookmark4)
2. [General Description 6](#_bookmark5)
   1. [Product Perspective & Problem Statement 6](#_bookmark6)
   2. [Tools used](#_bookmark8) 7
3. [Design Details 7](#_bookmark9)
   1. [Functional Architecture 7](#_bookmark10)
   2. [Optimization 8](#_bookmark11)
4. [KPIs 9](#_bookmark12)
   1. [KPIs (Key Performance Indicators) 9](#_bookmark13)
5. [Deployment 9](#_bookmark14)

# Abstract

The Airbnb Data Analysis Project aims to explore and analyze a dataset from Airbnb, a popular online marketplace for short-term rentals. The project focuses on gaining insights and understanding various aspects of the Airbnb ecosystem, such as pricing trends, property types, geographic patterns, and other property related characteristics.

The analysis begins with data preprocessing, including data cleaning, handling missing values, and transforming variables for further analysis. Exploratory data analysis techniques are then employed to identify patterns, trends, and relationships within the dataset. Visualizations, statistical summaries, and data aggregations are utilized to provide a comprehensive understanding of the dataset.

# Introduction

## Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

* + - Present all of the design aspects and define them in detail
    - Describe the user interface being implemented
    - Describe the hardware and software interfaces
    - Describe the performance requirements
    - Include design features and the architecture of the project
    - List and describe the non-functional attributes like:
      * Security
      * Reliability
      * Maintainability
      * Portability
      * Reusability
      * Application compatibility
      * Resource utilization
      * Serviceability

## Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

# General Description

## Product Perspective & Problem Statement

## Product Perspective:

The Airbnb Data Analysis Project is positioned as a data-driven analytical solution that aims to provide valuable insights and actionable recommendations for both Airbnb hosts and potential guests. It leverages the vast amount of data available on the Airbnb platform to uncover trends, patterns, and relationships that can inform decision-making and enhance the user experience.

By analyzing the data, the project aims to offer a comprehensive understanding of the Airbnb ecosystem, including factors influencing pricing, popular property types and neighborhoods and the impact of location on booking demand. This perspective empowers hosts to optimize their listings, pricing, and customer experience, while also helping potential guests make informed decisions when searching for accommodations.

## Problem Statement:

The Airbnb Data Analysis Project seeks to address several key problems and challenges faced by hosts and potential guests in the short-term rental market. These problems include:

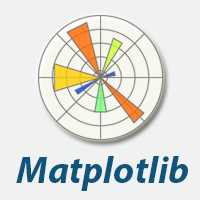
* Pricing Optimization: Hosts often struggle to determine the optimal pricing for their Airbnb listings. They face challenges in understanding the factors that influence pricing, such as property location, amenities, and demand. This project aims to analyze the data to uncover insights into pricing trends and factors affecting pricing decisions.
* Property and Neighborhood Selection: Potential guests face difficulties in selecting the most suitable property and neighborhood for their stay. They need assistance in identifying popular neighborhoods, preferred property types, and features that align with their preferences and requirements. The project aims to provide data-driven insights to facilitate informed decision-making during the property selection process.
* Host Performance and Customer Satisfaction: Hosts strive to enhance their performance and provide exceptional experiences for their guests. They need insights into the relationship between property reviews, and customer satisfaction. The project aims to analyze these factors to help hosts improve their offerings and guest satisfaction.
* Understanding Booking Demand: Hosts require a deeper understanding of the factors influencing booking demand, including location-based preferences, seasonality, and customer reviews. Analyzing the data can help identify patterns and trends in booking demand, enabling hosts to optimize availability, pricing, and promotional strategies.

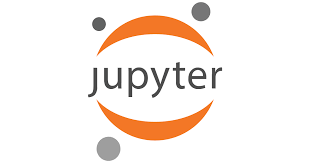
By addressing these problems, the Airbnb Data Analysis Project seeks to provide practical solutions and recommendations that enable hosts to optimize their listings, improve customer satisfaction, and ultimately maximize their revenue. It also empowers potential guests with valuable insights for making informed decisions and enhancing their overall Airbnb experience.

## Tools used

Business Intelligence tools and libraries such as Numpy, Pandas, Matplotlib, Seaborn, Plotly, Folium, Pandas profiling, D – Tale, jupyter notebook are used to build the whole framework.





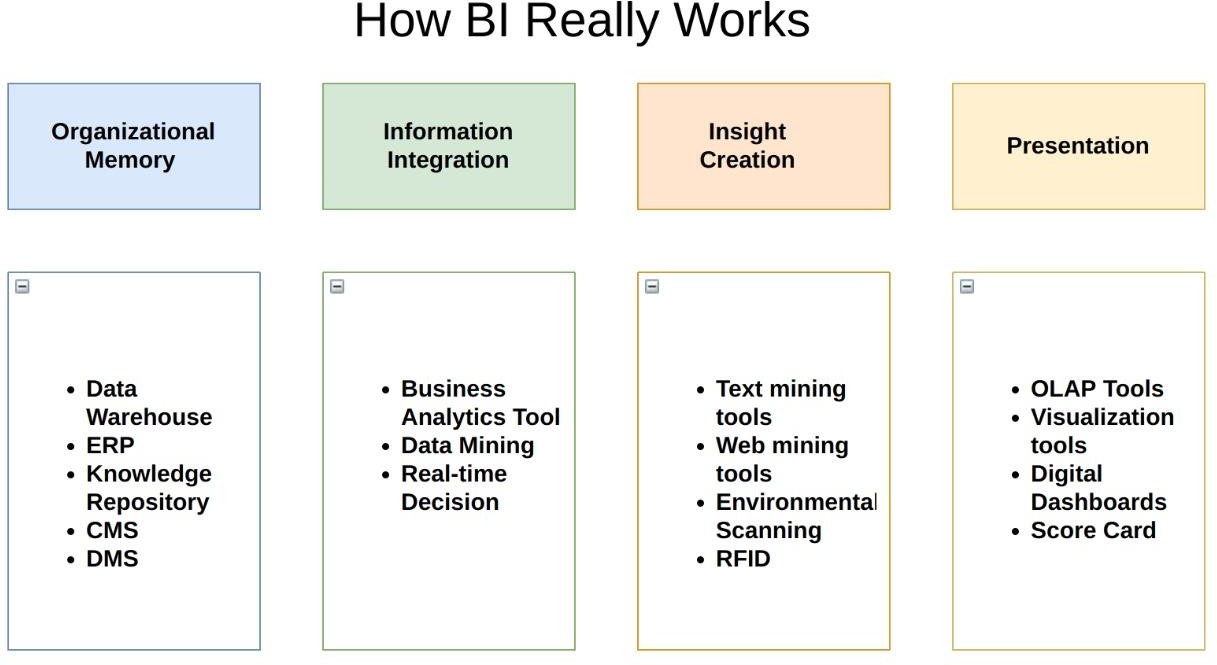


# Design Details

## Functional Architecture



Figure 1: Functional Architecture of Business Intelligence



## Optimization

### Your data strategy drives performance

* + - Minimize the number of fields
    - Minimize the number of records
    - Optimize extracts to speed up future queries by materializing calculations, removing columns and the use of accelerated views

### Reduce the marks (data points) in your view

* + - Practice guided analytics. There’s no need to fit everything you plan to show in a single view. Compile related views and connect them with action filters to travel from overview to highly-granular views at the speed of thought.
    - Remove unneeded dimensions from the detail shelf.
    - Explore. Try displaying your data in different types of views.

### Limit your filters by number and type

* + - Reduce the number of filters in use. Excessive filters on a view will create a more complex query, which takes longer to return results. Double-check your filters and remove any that aren’t necessary.
    - Use an include filter. Exclude filters load the entire domain of a dimension, while include filters do not. An include filter runs much faster than an exclude filter, especially for dimensions with many members.

### Optimize and materialize your calculations

* + - Perform calculations in the database
    - Reduce the number of nested calculations.
    - [Where possible, use MIN or MAX instead of AVG](http://onlinehelp.tableau.com/current/pro/online/windows/en-us/help.htm#calculations_aggregation.html). AVG requires more processing than MIN or MAX. Often rows will be duplicated and display the same result with MIN, MAX, or AVG.

# KPIs

Visualizations will be implemented to display and indicate certain KPIs and relevant indicators for the Airbnb properties.

## KPIs (Key Performance Indicators)

Key indicators displaying a summary of the Airbnb dataset and its relationship with different metrics

1. Regarding the Host

* Who are top earners
* Who earns the least

2. Regarding the Neighborhood

* Any particular location getting maximum number of bookings
* Price relation with respect to location

3. Regarding the reviews

* Relationship between Quality and Price

4. Regarding Price

* Price vs amenities
* Price vs location

# Deployment

The deployment section of the high-level design document for the Airbnb Data Analysis project outlines the approach for deploying the project's components and making the analysis accessible to stakeholders.

1. Deployment Architecture:

The Airbnb Data Analysis project will be deployed as a Jupyter Notebook-based solution. The Jupyter Notebook environment allows for interactive data analysis, code execution, and visualizations.

The project can be deployed on a local machine or a cloud-based platform that supports Jupyter Notebooks, such as Google Colab, Microsoft Azure Notebooks, or AWS SageMaker.

2. Data Loading and Preparation:

The project requires the Airbnb dataset to be loaded and preprocessed before analysis. The Jupyter Notebook should include the necessary code to load the dataset and perform data cleaning steps, such as handling missing values, data type conversions, and feature engineering if required.

Detailed documentation should be provided within the Notebook to guide users on how to load their own dataset or update the existing dataset path.

3. Data Analysis and Visualization:

The Jupyter Notebook should include sections that perform exploratory data analysis, statistical analysis, and data visualizations using libraries like Pandas, NumPy, Matplotlib, Seaborn, Plotly or folium.

Each analysis step should be well-documented and include clear explanations of the insights gained from the analysis.

4. Results and Reporting:

The project should provide a clear presentation of the analysis results and insights derived from the data.

The Jupyter Notebook should include visualizations, summary statistics, and any relevant findings that answer the research questions or address the problem statement.

The Notebook should also include markdown cells or narrative explanations to provide context and interpretation for the results.

5. Sharing and Collaboration:

To facilitate collaboration and knowledge sharing, the Jupyter Notebook can be shared with stakeholders and team members via online platforms like GitHub.