High Level Design (HLD)

Airbnb Data Analysis

Revision Number: 1.0

Rushikesh Husan Ramteke

Document Version Control

Date Issued	Version	Description	Author
12th October 2022	1.0	First Version of Complete HLD	Rushikesh Ramteke

Contents

Abstract

Since 2008, guests and hosts have used **Airbnb** to expand on travelling possibilities and present more unique, personalized way of experiencing the world. Today, Airbnb became one of a kind service that is used and recognized by the whole world. Data analysis on millions of listings provided through Airbnb is a crucial factor for the company. These millions of listings generate a lot of data - data that can be analysed and used for security, business decisions, understanding of customers' and providers' (hosts) behaviour and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

This dataset has around 18723 observations in it with 20 columns and it is a mix between categorical and numeric values.

This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions.

1 Introduction

1.1 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:
  - o security
  - o reliability
  - o maintainability
  - o portability
  - o reusability
  - o application
  - o compatibility
  - o resource
  - o utilization
  - o serviceability

1.2 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.

2 General Description

2.1 Product Perspective & Problem Statement

Since 2008, guests and hosts have used Airbnb to expand on travelling possibilities and present more unique, personalized way of experiencing the world. This dataset describes the listing activity and metrics in Amsterdam, Netherland for 2019. Content This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions.

The objective of the project is to perform data visualization techniques to understand the insight of the data. This project aims to apply **Exploratory Data Analysis (EDA**) and **Business Intelligence tools** such as **Power BI** to get a visual understanding of the data.

Objectives: Research Questions

Regarding the Host

- Who are top earners
- Is there any relationship between monthly earning and prices

Regarding the Neighbourhood

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

Regarding the reviews

- Relationship between Quality and Price

Regarding Price

- Price vs amenitites
- Price vs location

Find key metrics and factors and show the meaningful relationships between attributes.

2.2 Tools used

Business Intelligence tools and libraries works such as Python-Numpy, Pandas, Seaborn, Matplotlib, Excel, Power BI are used to build the whole framework.

3 Design Details

3.1 Functional Architecture

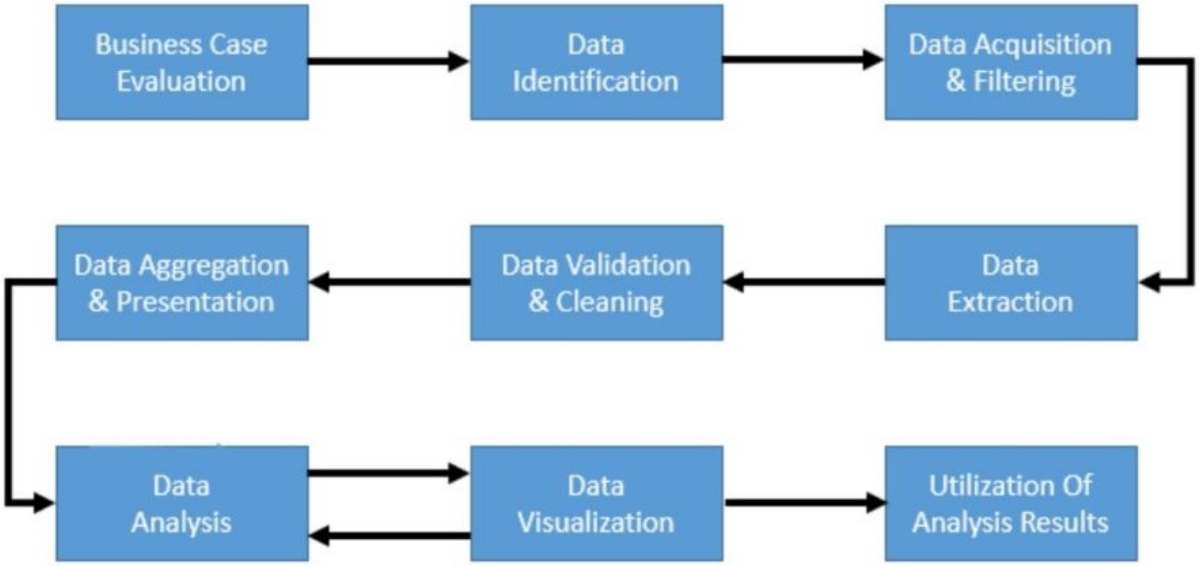


Figure 1: Functional Architecture of Data Analysis

3.1.1 Dataset/ Data Acquisition

The dataset had information regarding the reviews with respect to listing id. This data had all the information regarding the listings. It had Host name, location, neighbourhood, price, review score and number of review, latitude, longitude ,room type.

You can find the dataset on the given link: https://drive.google.com/drive/folders/1ANkgtAT0Pdp2r86IxFKv9vKYmnsYjJDO?usp=sharing

3.1.2 Exploratory Data Analysis (EDA)

- **EDA Using Python**

Exploratory Data Analysis is an approach to analyse the datasets to summarize their main characteristics in form of visual methods.
EDA is nothing but a data exploration technique to understand various aspects of the data. The main aim of EDA is to obtain confidence in a data to an extent where we are ready to engage a machine learning model.
EDA is important to analyse the data it's a **first steps in data analysis process**.
Exploratory data analysis help us to finding the errors, discovering data, mapping out data structure, finding out anomalies. Exploratory data analysis is important for business process because we are preparing dataset for deep through analysis that will detect you business problem.

Steps Involved in EDA :-

- Data Sourcing
- Data Cleaning
- Univariate analysis with visualisation
- Bivariate analysis with visualisation
- Derived metrics

Data Sourcing: Data Sourcing is the process of gathering data from multiple sources as external or internal data collection.
There are two major kind of data which can be classified according to the source: 1. Public data 2. Private data

Data Cleaning : After collecting the data , the next step is data cleaning. Data cleaning means that you get rid of any information that doesn't need to be there and clean up by mistake.
Data Cleaning is the process of clean the data to improve the quality of the data for further data analysis and building a machine learning model. The benefit of data cleaning is that all the incorrect and irrelevant data is gone and we get the good quality of data which will help in improving the accuracy.

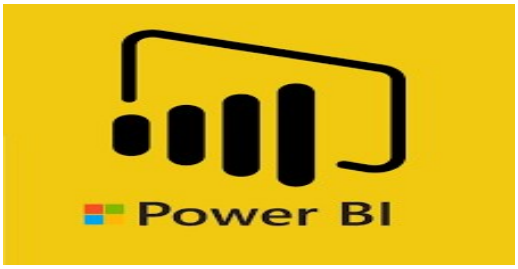
analysis with visualisation: Visualisation is the presentation of the data in the graphical or visual form to understand the data more clearly. Visualisation is easy to understand the data. Easily analyse the data and summarize it. Easily understand the features of the data. Help to find the trend or pattern of the data. Help to get meaningful insights from the data.

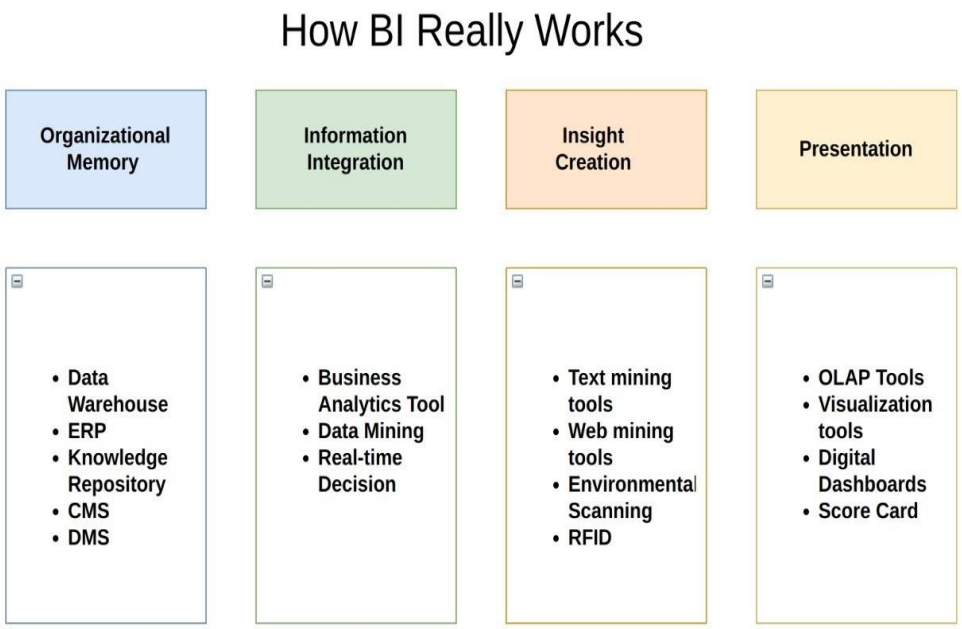
- Important Charts for Visualisation:
  1.Histogram
  2.Bar Chart
  3.Box plot
  4.pie chart
  5.Heatmap
  6.Scatter plot
  7.Line chart etc…

3.1.3 Business Intelligence Tool

- **Power BI :**

Power BI is an interactive data visualization software product developed by Microsoft with a primary focus on business intelligence.
Microsoft Power BI is used to find insights within an organization's data. Power BI can help connect disparate data sets, transform and clean the data into a data model and create charts or graphs to provide visuals of the data. All of this can be shared with other Power BI users within the organization.
The **Power BI dashboard** is a feature only available in **Power BI Service**. Since a Power BI dashboard is limited to one page, it only contains the highlights of a story. You cannot create a dashboard on Power BI Desktop. In a dashboard, visualizations are generated from reports, and each report is based on one dataset.





The overall process of Power BI is ETL which stands for extract, transform and load. The data is first extracted from the data sources in Power BI. The data is then transformed into the required format. Once you are done with shaping the data, you can load the data in the Data Catalog.

step-by-step process of Power BI:

Data Connection: It is easy to import data into Power BI or even upload a custom file. We have imported an excel file having a Airbnb booking data.

Data Transformation: After loading the data, it needs to undergo pre processing based on the requirements. Data cleaning, dealing with invalid values, null values, missing values. etc…

Data Modeling: It enhances the data in order to achieve accurate insights and analytics.

Data Visualization: In this step, the visualization type is chosen as per the data

Data Reports: You can publish and export the reports securely.