# HIGH LEVEL DESIGN DOCUMENT FOR AIRBNB DATA ANALYSIS

# DOCUMENT VERSION CONTROL

| Date Issue | Version | Description | Author |
|---|---|---|---|
| 09/03/2023 | 1 | Initial HLD – V 1.0 | Abdul Jaweed |
| 25/06/2023 | 2 | Final HLD -  V 2.0 | Abdul Jaweed Aalmeen Khan Safa Ashraf |

# CONTENT

- DOCUMENT VERSION

- ABSTRACT

- INTRODUCTION

- WHY THIS HIGH LEVEL DESIGN DOCUMENT

- SCOPE

- GENERAL DESCRIPTION

- PROBLEM STATEMENT

- PRODUCT PERSPECTIVE

- PROPOSED SOLUTION

- TECHNICAL REQUIREMENTS

- TOOLS USED

- TABLEAU INTRODUCTION

- DESIGN DETAILS

- OPTIMIZATION

- ARCHITECTURE DESCRIPTION

- TABLEAU DASHBOARD INTRODUCTION

- AIRBNB DATA ANALYSIS VISUALIZATION DASHBOARD

# ABSTRACT

Airbnb is an American company that facilitates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. It basically connects travelers with local hosts who want to rent out their homes with people who are looking for accommodations in that locality. On the other hand, this platform enables hosts to list their available space and earn extra income in the form of rent and it also enables travelers to book unique homestays from local hosts, saving them money and giving them a chance to interact with locals. In the world of rising new technology and innovation, Travel industry is advancing with the role of Data Science and Analytics. Data analysis can help them to understand their business in a quite different manner and helps to improve the quality of the service by identifying the weak areas of the business. This study demonstrates how different analyses help out to make better business decisions and help analyze customer trends and satisfaction, which can lead to new and better products and services. Different analyses performed such as Exploratory Data Analysis and Descriptive Analysis on a variety of use cases to get the key insights from this data based on which business decisions will be taken.

# 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 THE DESIGN ASPECTS AND DEFINE THEM IN DETAIL
- DESCRIBE THE USER INTERFACE BEING IMPLEMENTED
- DESCRIBE THE HARDWARE AND SOFTWARE INTERFACE
- DESCRIBE THE PERFORMANCE REQUIREMENT
- INCLUDE DEFINE FEATURE AND ARCHITECTURE OF THE PROJECT
- LIST AND DESCRIBE THE NON FUNCTIONAL ATTRIBUTES

  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

## PROBLEM STATEMENT

The objective of the project is to perform an exploratory data analysis, data pre-processing, data cleaning & imputation and at the end, apply different Data Visualization techniques to get the meaningful insight from the given data. This project aims to apply some amazing Python Libraries such as Plotly and WordCloud which will give a boost to our visual understanding of the data.

### The main objective of the project includes :
- Analyzing raw dataset.
- Perform data cleaning and preprocessing operations on the raw data.
- Building visualizations on the cleaned features using a BI tool.
- Building Dashboard using Tableau

# TECHNICAL REQUIREMENTS

The solution can be a cloud-based or application hosted on an internal server or even be hosted on a local machine. For accessing this application below are the minimum requirements:

- Good internet connection.
- Web Browser.

For training model, the system requirements are as follows:

- +4 GB RAM preferred
- Operation System: windows, Linux, Mac
- Tableau / Jupyter Notebook / Google Colab
- S3 buckets for cloud storage for the collected data.

## TOOLS USED

Python Libraries such as NumPy, Pandas, Matplotlib, Seaborn, and MS-Excel and Business Intelligence tools such as Tableau, Colab / Jupyter Notebook and Python Programming Language are used to build the whole framework.

## TABLEAU

Tableau is a leading data visualization and business intelligence tool developed by Tableau Software, now a part of Salesforce. It helps users analyze and visualize data to gain valuable insights and make data-driven decisions. Here's an overview of Tableau:

**Data Connection:** Tableau allows users to connect to various data sources, including spreadsheets, databases, cloud-based platforms, and big data sources. It supports both structured and unstructured data.

**Data Visualization:** Tableau offers a wide range of interactive and visually appealing charts, graphs, maps, and dashboards to present data in a meaningful way. Users can drag and drop data fields to create visualizations and customize them with formatting options.

**Dashboard Creation:** With Tableau, users can create interactive dashboards by combining multiple visualizations onto a single canvas. Dashboards provide a holistic view of data and allow for quick insights and exploration.

**Analytics and Calculation:** Tableau provides advanced analytics capabilities, including statistical functions, forecasting, clustering, and trend analysis. Users can also create calculated fields and perform ad-hoc calculations using formulas and expressions.

**Data Blending and Joining:** Tableau enables users to blend and join data from multiple sources, allowing for comprehensive analysis across different datasets. It provides intuitive data blending features to merge and consolidate data for analysis.

**Collaboration and Sharing:** Tableau allows users to share their visualizations and dashboards with others in the organization through Tableau Server or Tableau Online. This enables collaboration and facilitates data-driven decision-making across teams.

**Mobile Support:** Tableau offers mobile applications for iOS and Android devices, allowing users to access and interact with their visualizations on the go. The mobile app provides a responsive and optimized experience for mobile devices.

**Integration and Extensibility:** Tableau integrates with various data sources, including Microsoft Excel, SQL databases, cloud platforms like Amazon Web Services (AWS) and Google Cloud, and more. It also supports custom extensions and APIs for further customization and integration with other tools.

Tableau provides a user-friendly interface that caters to both technical and non-technical users, making it popular in organizations of all sizes and industries. Its powerful data visualization capabilities and intuitive workflows make it a valuable tool for data analysis and reporting.

## DESIGN DETAILS

**Layout:** When designing a dashboard in Tableau, it's important to create a well-organized layout. Use a grid-based approach to align and position your visualizations consistently. Group related visualizations together using containers or layout containers. Use text boxes and titles to provide context and headings to separate different sections of your dashboard.

**Color:** Tableau offers a variety of color palettes to choose from, or you can create your own custom color scheme. Use color strategically to highlight important data points or to create visual emphasis. Ensure that the colors you choose are easily distinguishable and accessible for all users.

**Typography**: Choose a clear and readable font for your text elements in Tableau. Consistency is key, so use the same font for titles, labels, and descriptions throughout your dashboard. Adjust the font size and weight to create a visual hierarchy, with headings and titles standing out more prominently.

**Visualization:** Tableau provides a wide range of visualization options, including bar charts, line charts, scatter plots, maps, and more. Select the appropriate visualization type that effectively represents your data and insights. Avoid overcrowding your dashboard with too many visualizations. Instead, focus on the key messages and use visualizations that provide the most meaningful representation of your data.

**Interactivity:** Tableau offers various interactive features that can enhance the user experience. Utilize filters, parameters, and actions to allow users to interact with the data and explore different perspectives. Implement tooltips and drill-down capabilities to provide additional details and context on specific data points.

**Performance:** To optimize the performance of your Tableau dashboard, consider the following strategies:

**Data Source Optimization:** Optimize your data source by removing unnecessary columns, aggregating data where possible, and filtering out irrelevant data.

**Extracts**: Use Tableau data extracts (.tde or .hyper) for faster performance, especially when dealing with large datasets. Extracts can be pre-aggregated and optimized for specific use cases.

**Calculations:** Minimize the use of complex calculations and heavy computations in your dashboard. Use calculated fields efficiently and consider precomputing calculations where appropriate.

**Filters and Context**: Apply filters strategically to limit the amount of data being processed and displayed in real-time. Context filters can help improve performance by reducing the dataset before other calculations are performed.

**Data Blending:** If you are blending data from multiple sources, try to minimize the data blending operations or consider pre-joining the data outside of Tableau for better performance.

**Accessibility:** Ensure that your Tableau dashboard is accessible to all users by following these guidelines:

**Color Contrast:** Use colors with sufficient contrast for text and visual elements to ensure readability.

**Alternative Text:** Provide alternative text descriptions for images and visualizations for users who rely on screen readers.

**Keyboard Accessibility:** Ensure that all interactive elements, such as filters and buttons, can be accessed and operated using a keyboard alone.

**Data Tables:** Include data tables or provide an option to export data in a tabular format for users who prefer consuming data in a structured way.

By incorporating these design details and optimization techniques, you can create an effective and visually appealing Tableau dashboard that delivers insights and enhances data exploration.

## ARCHITECTURE DESCRIPTION.

The architecture for an Airbnb data analysis project typically involves several components and steps to process and analyze the data. Here's a high-level description of the architecture for an Airbnb data analysis project:

**Data Collection:** The first step is to gather the necessary data from Airbnb. This can include data on listings, reviews, bookings, host information, and more. The data can be obtained through Airbnb's public APIs, data scraping, or by downloading publicly available datasets.

**Data Storage:** Once the data is collected, it needs to be stored in a suitable data storage system. Common options include relational databases (such as MySQL, PostgreSQL, or SQL Server) or NoSQL databases (such as MongoDB or Elasticsearch). The choice of data storage depends on factors such as the volume of data, the complexity of relationships, and the requirements of the analysis.

**Data Cleaning and Preprocessing:** Before performing analysis, the collected data needs to be cleaned and preprocessed. This involves tasks such as handling missing values, removing duplicates, standardizing data formats, and transforming the data into a suitable structure for analysis.

**Data Transformation and Integration:** In some cases, additional data sources might be required to enrich the analysis. This could include external datasets such as geographical data, demographic data, or weather data. Data transformation and integration techniques, such as data joins and merges, are used to combine different datasets and create a unified view for analysis.

**Data Analysis and Visualization:** Once the data is prepared, various data analysis techniques can be applied to gain insights. This can involve statistical analysis, exploratory data analysis, data mining, and machine learning algorithms. Visualization tools, such as Tableau, Power BI, or Python libraries like Matplotlib or Seaborn, are used to create visual representations of the data and communicate findings effectively.

**Deployment and Reporting:** The analyzed results and visualizations can be deployed in different ways to share insights with stakeholders. This can include creating interactive dashboards, generating reports, or building web-based applications. The choice of deployment method depends on the intended audience and their requirements for accessing and interacting with the analysis results.

**Ongoing Monitoring and Iteration:** Data analysis projects are often iterative processes, where insights and findings lead to further questions and refinements. It's important to monitor the performance of the analysis over time, update the data as new information becomes available, and iterate on the analysis to gain deeper insights or address emerging questions.

Overall, the architecture of an Airbnb data analysis project involves data collection, storage, cleaning, preprocessing, analysis, visualization, and deployment stages. The specific tools and technologies used can vary based on the project requirements and the expertise of the data analysis team.

## TABLEAU DASHBOARD

Tableau is a leading data visualization and business intelligence tool that empowers users to create interactive and visually compelling dashboards and reports. Here's a Tableau version of the description:

Tableau is a robust data analysis and visualization platform that enables users to explore and present data in a user-friendly and intuitive manner. It offers a wide range of features and capabilities that make it a popular choice for businesses and analysts.

**Data Connectivity:** Tableau supports connectivity to various data sources, including spreadsheets, databases, cloud-based platforms, and web services. It provides native connectors to popular data sources such as Excel, SQL databases, Salesforce, and Google Analytics, allowing users to seamlessly connect and integrate their data.

**Visual Exploration:** With Tableau, users can easily drag and drop data fields onto the canvas to create dynamic visualizations. The platform offers a vast library of charts, graphs, maps, and other visual elements that can be customized to convey insights effectively. Users can interact with the visualizations, apply filters, and drill down into the data to explore different perspectives and uncover hidden patterns.

**Dashboard Creation:** Tableau allows users to combine multiple visualizations into interactive dashboards. Users can arrange and organize their visualizations in a layout that suits their needs, enabling them to create comprehensive and visually appealing dashboards. Dashboard elements can be linked to create interactivity, allowing users to explore the data and make discoveries.

**Advanced Analytics:** Tableau provides advanced analytics capabilities that empower users to perform calculations, apply statistical functions, and perform complex data modeling. Users can create calculated fields, use parameters for dynamic analysis, and leverage built-in statistical functions for

deeper insights. Tableau also supports integration with popular programming languages such as R and Python for advanced analytics and machine learning.

**Sharing and Collaboration:** Tableau offers various options for sharing and collaborating on dashboards and reports. Users can publish their visualizations to Tableau Server or Tableau Online, enabling secure access for authorized users within their organization. Dashboards can also be shared via Tableau Public or embedded into websites and applications, allowing for broader dissemination of insights.

**Performance Optimization:** Tableau provides features to optimize performance when working with large datasets or complex visualizations. Techniques such as data extracts, data blending, and data source filters can be employed to enhance performance. Tableau also offers options for caching, data aggregation, and incremental data updates to improve responsiveness and reduce query times.

Tableau offers a user-friendly interface, powerful data visualization capabilities, and a robust set of features for data exploration and analysis. Its versatility and flexibility make it a valuable tool for individuals and organizations across various industries to derive insights from their data and make data-driven decisions.

# AIRBNB DATA ANALYSIS VISUALIZATION DASHBOARD