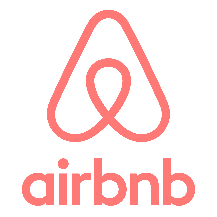
**High-Level Design (HLD)**

**Travel Data Analysis(Air BNB Case Studies)**



Last Revised Date: 25/08/2021

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Author | Comments |
| 1. | 29-Sep-21 | Manish Tiwari  Samruddhi Deshmukh  Subhabrata Nath | Adding Problem statement |
| 2. | 29-Sep-21 | Samruddhi Deshmukh | Adding Flowchart  Updating Index |
|  |  |  |  |

**Reviews:**

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Reviewer | Comments |
|  |  |  |  |

**Approval Status:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Version | Review Date | Reviewed By | Approved By | Comments |
|  |  |  |  |  |

**Index**

|  |  |
| --- | --- |
| **Content** | **Page No** |
| Abstract | 3 |
| 1 Introduction | 3 |
| 1.1 Scope of the HLD | 3 |
| 2 General Description | 3 |
| 2.1 Product Perspective & Problem Statement | 3 |
| 2.2 Tools used | 4 |
| 2.2.1 Tableau Desktop and Tableau Public | 4 |
| 3 Design Details | 5 |
| 3.1 Application Architecture | 5 |
| 3.2 Optimization | 6 |
| 3.2.1 Data strategy and Performance | 6 |
| 3.3 Architecture Description | 6 |
| 4 Dashboard | 7 |
| 4.1 KPIs (Key Performance Indicators) | 7 |
| 5 Deployment | 8 |
| 5.1 Steps of deployment | 8 |

**Abstract**

Observe the AirBNB travel data and find the relation between different attributes and mark the top host those are making more profit and find the particular places where most of the tourists like to visit. Airbnb is an online platform that allows people to rent short-term accommodation. Airbnb Travel Data is describe the listing activity and metrics in San Diego, California for 2019. Our main goal is to find Research Questions from the given data and make a dashboard that stands with our answers.

**1. Introduction**

**1.1 Scope of the HLD:**

The HLD system aims to present the functionality of the overall system. It aims to describe the process workflow along with a description of its various components. It describes the –

* Problem statement under consideration.
* Solution strategy towards the problem.
* The architecture of the process.
* Implementation of the solution.
* Tools used.
* Analysis of the KPI’s and coming out with suitable conclusions about the analysis.

The final analysis (dashboard) will be shared through PowerBI services.

**2. General Description**

**2.1 Product Perspective and Problem Statement**

The main problem statement is to explore the Travel dataset which includes multiple tables and find the relation between them and make an entity diagram that makes the relation between tables. Understand the problem statement and perform ETL to make measures and relationships between different attributes to find the Research Questions.

The main objective of the project includes :

* Analyzing raw dataset.
* Perform data cleaning and pre-processing operations on the raw data.
* Using Power-BI ETL tool to preprocess the raw data.
* Building visualizations on the cleaned features using a BI tool(Power-BI).
* Find different Research Questions and Find the answers from data.
* Deployment of the project on Power-BI Service.

**2.2 Tools Used**

* Data Pre-processing – Power BI ETL tools, Excel. Python library pandas.
* Data Visualization – Power BI.
* Dashboarding/ Deployment – Power BI Service.



Graphical user interface, icon

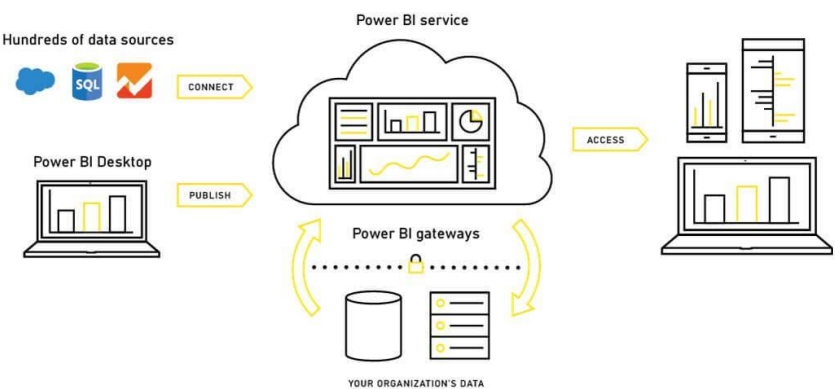
Description automatically generated



**2.2.1 Tableau Desktop and Tableau Public –**

The Power BI or Power BI Desktop is data visualization software that lets you see and understand data in minutes. The professional version of this can transform, process, and store huge volumes of data which is responsible for all the data-driven decision-making of an organization. **Power BI Desktop** can connect a variety of data sources whether it’s big data, a SQL database, a spreadsheet, or cloud apps like Google Analytics and Salesforce. **Power BI Desktop**has no limit to how many rows of data it can store, process or share. For **Power BI Desktop**, reports published can be stored on your **Local Drive** as well as on the **Power BI Service.**

The **Power BI Desktop** is essentially a free version of Power BI visualization software. It allows you to use most of the software functions. You can create visualizations and connect to CSV, Text, and Excel documents. Power BI Public is a free platform to publicly share and explore data visualizations online. Anyone can create visualizations using either Power BI Desktop Professional Edition or the free Public Edition.

**2.2.2 Power BI Architecture.**

**3. Design Details:**

**3.1 Application Architecture.**

* The following flow we will cover for the entire application.

A screen shot of a computer

Description automatically generated with low confidence

**3.2 Optimization**

**3.2.1 Data Strategy and Performance.**

* Removing duplicate records from the dataset.
* Handling null values.
* Performing feature encoding on the dataset.
* Building interactive filters in tableau to view data as required.
* Creating groups, hierarchy, and calculated fields for easy analysis.
* Crete Entity Diagram for creating connection different tables.
* Creating Custome Measure and field for making plots and find answers.
* Creating Dashboard for Supporting found answers.

**3.3 Architecture Description.**

* **Understanding Problem Statement:**

Find /Analyse relations between features and come out with the significant contributors to travel Data Analysis. The end goal includes building a Dashboard for user interaction and a Detailed project report for publishing our findings.

* **Collecting Data:**

Dataset is already available on the project dashboard. We will use the given dataset for building the project dashboard.

* **Exploring Data:**

Performing pre-processing with raw dataset and using the cleaned dataset for building graphs. We will use the plotting library and python data manipulation library like pandas for these tasks.

* **Data Optimization through Filtering:**

We will optimize the given dataset before starting the analysis. We have to drop duplicate columns and reduce the dataset dimensions as needed. We will also be encoding our data into numerical to categorical or vice versa as needed.

* **Table Calculation:**

We will calculate some new fields that will represent more granular results than raw data. We will find the relationship between attributes and perform the calculation as required.

* **Importing Data In BI Tool:**

We will use Power BI for creating the Visualization and Dashboard. Power BI is the most popular BI application for creating Dashboard from Microsoft and it is very much efficient in handling a large amount of data. We can plot and show different types of relationships between attributes. Also Power BI provides inbuild ETL tools to perform data preprocessing.

* **Filtering Data Attributes:**

We will use different types of filters on our data set for plotting purposes and getting the insight information. Using filters can speed up information finding and also it can produce more variety and granular reports.

* **Insight Creation Using BI Tools:**

This is the most crucial part of the complete process. Here we will find meaningful information by going deep into our dataset. We will plot different types of plots for every attribute on the Travel Data Analysis(AirBnb) dataset to show the relationship.

* **Creating Report for the Presentation:**

After we find all the possible information that can be found from the dataset we will create a detailed report for presenting and publishing our works. We will include all the visual plots and will write key findings from the plots.

* **Hosting the Dashboard on Cloud:**

After creating a Detailed Project report and creating the dashboard we will host our dashboard on the cloud platform for global use. For this, we will use the Power BI Service Platform. Here we can host our dashboard and it will create a sharable link for access globally.

**4. Dashboard**

For showing the report user can use our Dashboard. The dashboard will be implemented to display and indicate certain KPIs and relevant indicators.



The dashboard will show the relationship between different attributes and will tell the answers to research questions like who are the top earners among all hosts. From which place the most of travel booking are coming

**4.1 Key Performance Indicators (KPIs)**

KPIs will put a summary of the findings of the research questions like who are the top earners and from which place the most of booking is coming and e.tc.

**4.1.2** **Research Questions.**

* **Regarding the Host**
  + Who are the Top Earners?
  + Is there any relationship between monthly earning and prices?
* **Regarding the Neighbourhood**
  + Any particular location getting the maximum number of bookings.
  + Price relation with respect to location.
* **Regarding the Reviews**
  + Relationship between Quality and price.
* **Regarding Price**
  + Price Vs Amenities.
  + Price Vs Location.

**5. Deployment**

Power BI service will be used for deployment. It required an official email Id to create a free account. And publishing dashboard for making it publically available. There is some advantage of choosing this server it is required less maintenance, can be host directly form local machine, and the deployment process is also very easy. Deployment can be done from Power BI Workspace only.

**Benefits of Power BI:**

* Easy to use and Free.
* Fast and more reliable.
* No software installation is required for the deployment.
* Easy to share of the globally accessible link.
* The report can be saved locally.

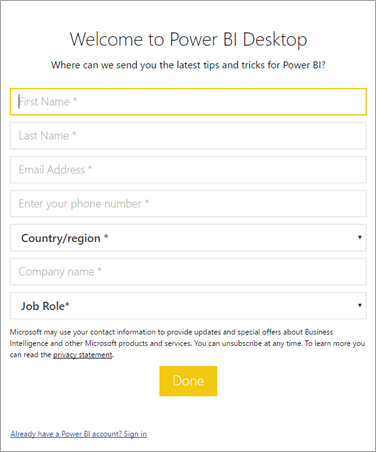
**Drawbacks of Power BI:**

* For Publishing Dashboard required official email id. Regular email id or Gmail Id will not work.
* Computational Expensive software. Required sufficient amount of RAM to run on Local machine.

**5.1 Steps of Deployment**

We can deploy our dashboard directly from our workspace, we must have registered on the Power BI service before that, It required an official email id.

**Steps 1 – Login**



The user has to enter his registered official email address and password for the login.

**Step 2 – Building the Dashboard**

Users need to Build the Dashboard before Deploying on Power BI Services. And have to save the dashboard on the Power BI workspace. We have to select and have to present the desire dashboard or worksheet that we want to deploy.

**Step 3 – Save the Dashboard**

**```**

**Put a dashboard picture here.**

**```**

**Step 4 – Publish to Power BI**

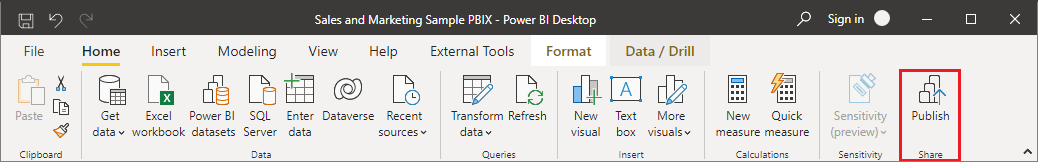
In Power BI Desktop, choose File > Publish > Publish to Power BI or select Publish on the ribbon.

Fig:xx

Here we have to select the destination. And after selecting the workspace have to click on Publish button. As shown in the above image(fig:xx)

**Steps 5 – Send to Cloud.**

After Selecting the workspace here workspace is our final dashboard. And click on Publish button we will get a confirmation of the successful deployment of our report on the Power BI cloud.

