# High Level Design

## Travel Data Analysis

## AirBNB Dataset

Revision number: 1.0

Last date of revision: 15/10/2021

Shivank Singh

# Document Version Control

| Date Issued | Version | Description | Author |
|---|---|---|---|
| 10/02/2022 | HLD-V1.0 | First Version of Complete HLD | Shivank Singh |
| | | | |

# Contents

# Abstract

AirBNB began in 2008 when two designers who had space to share hosted three travelers looking for a place to stay. Now, millions of hosts and travelers choose to create a free AirBNB account so they can list their space and book unique accommodation anywhere in the world. In addition, AirBNB experience hosts share their passions and interests with both travelers and locals.

AirBNB helps make sharing easy, enjoyable, and safe. We verify personal profiles and listings, maintain a smart messaging system so hosts and guests can communicate with certainty, and manage a trusted platform to collect and transfer payments.

AirBNB has provided many travelers a great, easy and convenient place to stay during their travels. Similarly, it has also given an opportunity for many to earn extra revenue by listing their properties for residents to stay. However, with so many listings available with varying prices, how can an aspiring host know what type of property to invest in if his main aim is to list it in AirBNB and earn rental revenue? Additionally, if a traveler wants to find the cheapest listing available but with certain features, he prefers like 'free parking' etc., how does he know what aspects to look into to find a suitable listing? There are many factors, which influence the price of a listing. Which is why we aim to find the most important factors that affect the price and more importantly the features that is common among the most expensive listings. This will allow an aspiring AirBNB host to ensure that his listing is equipped with those important features such that he will be able to charge a higher price without losing customers. Moreover, a traveler will also know the factors to look into to get the lowest price possible while having certain features he prefers.

In the arena of rising new generation and innovation, Travel enterprise is advancing with the function of Data Science and Analytics. Data analysis can assist them to understand their business in a quiet distinct way and helps to improve the exceptional of the provider by using identifying the vulnerable areas of the business. This examine demonstrates the how distinct analysis assist out to make higher business choices and help examine customer tendencies and pride, that may lead to new and higher products and services. Different evaluation accomplished along with Exploratory Data Analysis and Descriptive Analysis on type of use instances to get the important thing insights from these records primarily based on which enterprise decisions might be taken.

# 1. Introduction

In this age of online business, we have found that there are plenty of new companies coming and creating competition in different fields. Here we will talk about the hotel and stay chain business and particularly we will going to have a discussion on “AirBNB”. We have received a problem statement for analyzing the data of AirBNB stating, since 2008, guests and hosts have used AirBNB to expand on travelling possibilities and present unique, personalized way of experiencing the world. This dataset describes the listing activity and metrics in Amsterdam, Netherlands for 2017. Keeping the above statement in mind, we have done a Data Analysis and Visualization of Data using Python (Jupyter Notebook).

## 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

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

## 1.3 Scope of HLD

This HLD covers all areas of system.

# 2. General Description

## 2.1 Product Perspective & Problem Statement

Travel industries are having important mirrored image of the economy from beyond few a long time, and AirBNB housing fee levels are of terrific interest for both Hosts and Traveler. In this task, we are analyzing the diverse aspects with extraordinary use instances, which covers many components of AirBNB listings. It facilitates in no longer simplest understanding the meaningful relationships between attributes however, it also allows us to do our personal studies and come-up with our findings.

The goal of the mission is to perform an exploratory data analysis, information pre-processing, statistics cleansing & imputation and on the quit, practice exclusive Data Visualization techniques to get the meaningful insight from the given records. This assignment targets practice some high-quality Python Libraries including Plotly and Seaborn a good way to deliver a boost to our visual information of the information.

## 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 iNeuron Company.
- 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:

| Name | Description |
|---|---|
| **host_id** | Unique ID of the AirBNB |
| **name** | Name of the AirBNB |
| **neighborhood** | Name of the Neighborhood |
| **reviews** | Reviews count |
| **overall_satisfaction** | AirBNB Rating out of 5 Star |
| **room_type** | AirBNB Room Type |
| **last_modified** | Date and Time Modification of listing |
| **price** | AirBNB Price per day |
| **latitude** | Latitude of AirBNB |
| **Longitude** | Longitude of AirBNB |

## 2.3 Tools Used

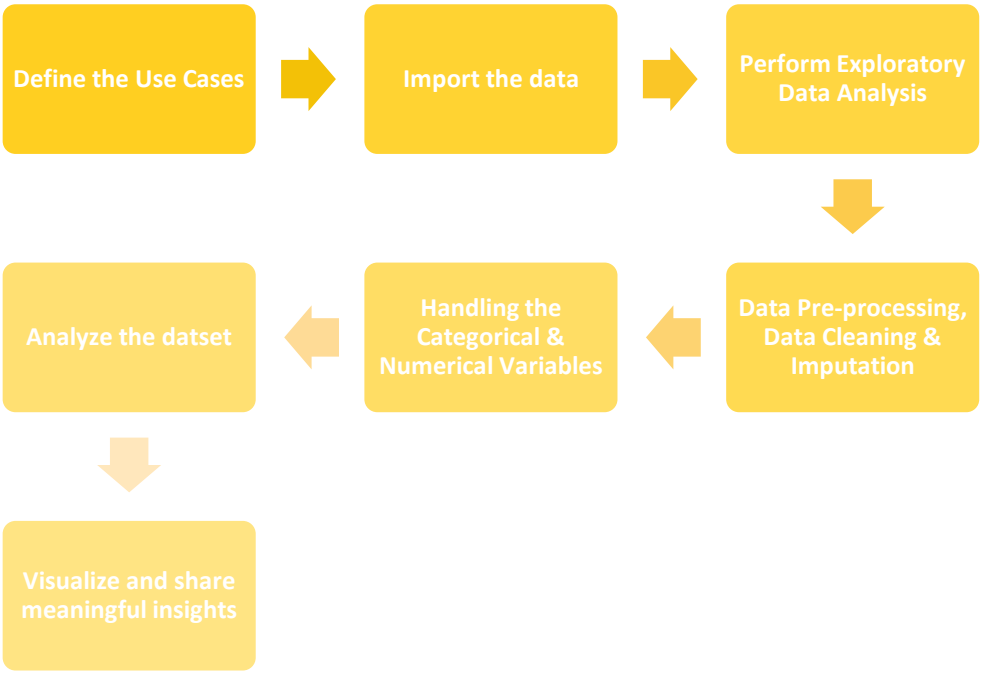
- Jupyter Notebook is used as IDE.
- Pandas and NumPy are used for Data Manipulation & Pre-processing and Mathematical functions respectively.
- Exploratory data analysis is automated by dataprep and D-tale.
- For visualization of the plots, Matplotlib, Seaborn, Plotly and Cufflinks are used.
- GitHub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


## 2.4 Constraints

The analysis must be user friendly, code must be neat & clean, EDA must be automated as much as possible because it will save huge amount of time. Moreover, 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. Design Details

## 3.1 Process Flow




## 3.2 Error Handling / Exception Handling:

We have designed this challenge in this sort of way that, entire script is tested and runs a couple of times to ensure that there is no blunders happened at some stage in manner drift.

Additionally, we have got also dismiss the un-important warnings to avoid confusion by means of the use of filter out warnings elegance from warnings module.

# 4. Conclusion

1. We have encounter the missing values have a count of 20% of the overall % of the dataset. The missing values was found in "Country", "Borough", "Bathroom", "Name" and "Minstay". We have handled the missing value by adding the cells with the relevant information according to the name of the feature.
2. We have also found that who are the top earners in Amsterdam also done the visualization for the same.
3. We tried to find out that is there any relationship between monthly earning and prices or not.
4. We created the logic to get information of any particular location getting maximum number of bookings.
5. We have found the price relation with respect to location.
6. We have visualize the relationship between Quality and Price.
7. We have plotted the graph for Price vs Amenities.
8. Also tried to visualize Price vs location.

**Other Findings:**

9. We have find out the relationship between Room_type and neighborhood.
10. We found that the preference of guest in comparison with Room Type is majorly Entire Home/Apt.
11. We encountered that which is the Cheapest AirBNB and what is its cost.
12. We also encountered the Most Expensive AirBNB.
13. We found that which are the top five Location having most booking on AirBNB.
14. We have tried to visualize Price relationship with respect to Neighborhood.
15. We have created a Map using the Latitude and Longitude Features of the data set and successfully plotted the accurate locations of all the AirBNB Available in Amsterdam.
16. We have visualize the maximum number of modification with respect to location that was encountered in different neighborhoods across Amsterdam, Netherlands.
17. We have found the modification with respect to room type and we are able to see that entire home/apt has maximum number of modifications.
18. We can clearly see the names of the entire neighborhood having AirBNB in Amsterdam by using Word Cloud plot, which looks beautiful.
19. Finally with the help of D-tale EDA library we have crosschecked the report weather the findings were accurate or not. In addition, we can clearly see that the findings that we have made was accurate and satisfactory.

# 5. References


1. How AirBNB works, insights into Business & Revenue Model
2. AirBNB - Wikipedia
3. Ineuron.ai
4. Youtube.com
5. Python Documentations available on web.