Low Level Design (LLD)

Travel Data Analysis

(AirBNB Data Analysis)

Written By / Author	Shivank Singh
Document Version	LLD-V1.0
Last Revised Date	07/10/2022

Document Version Control

Date Issued	Version	Description	Author
07/10/2022	LLD-V1.0	First Version of Complete LLD	Shivank Singh

Contents

Document Version Control 2
Abstract 4
1 **Introduction** 5
1.1 Why this Low-Level design document? 5
1.2 Scope 5
1.3 Constraints 5
2 **Technical Specifications** 6
2.1 Listings Dataset 6
2.1.1 Listings Dataset Overview 6
2.2 Review Dataset 6
2.1.1 Review Dataset Overview 7
3 **Architecture** 8
3.1 Architecture Description 8
3.1.1 Data Description 8
3.1.2 Define the Use Cases 8
3.1.3 Import the Dataset 9
3.1.4 Exploratory Data Analysis (EDA) 9
3.1.5 Data Pre-processing, Data Cleaning & Imputation (Handling the Categorical & Numerical Variables) 9
3.1.6 Analyse the Data 10
3.1.7 Visualize & Share Meaningful Insights 10
4 **Technology Stack** 11

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

1.1 Why this Low-Level design document?

The purpose of this LLD or a Low-Level Design (LLD) document is to give the internal logical design of the actual program code for Airbnb Data Analysis project. LLD describes the class diagrams with the methods and relations between classes and program specs. It describes the modules so that the programmer can directly code the program from the document. This document is intended for both the stakeholders and the developers of this project and will be proposed to the higher management for its approval.

The main objective of the project is to analyze the various aspects with different use cases which covers many aspects of AirBNB listings. It helps in not only understanding the meaningful relationships between attributes but it also allows us to do our own research and come-up with our findings.

1.2 Scope

Low-level design (LLD) is a component-level design process that follows a step-by step refinement process. This process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

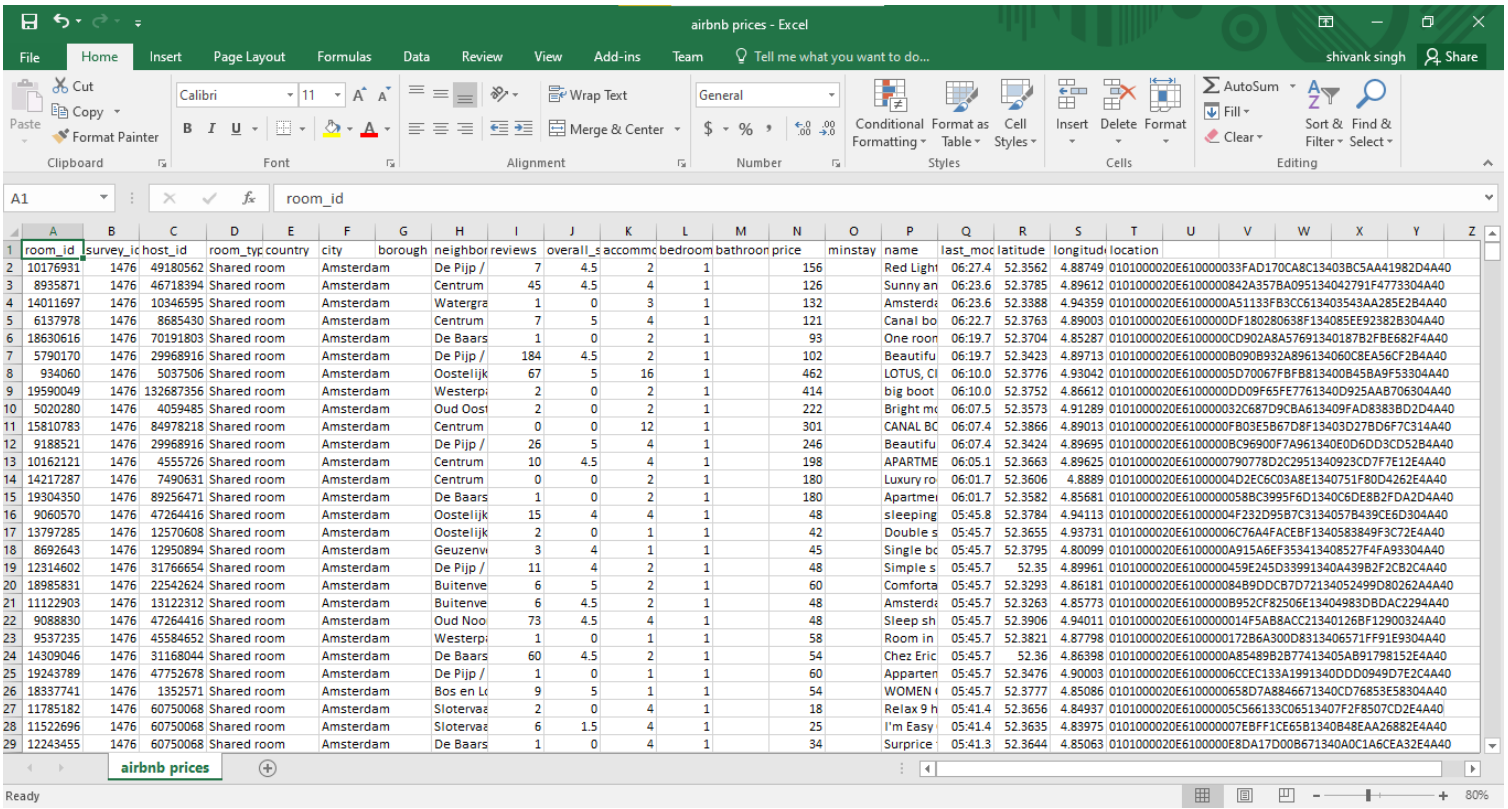
This study demonstrates the how different analysis help out to make better business decisions and help analysis customer trends and satisfaction, which can lead to new and better products and services.

1.3 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.

2 Technical Specifications

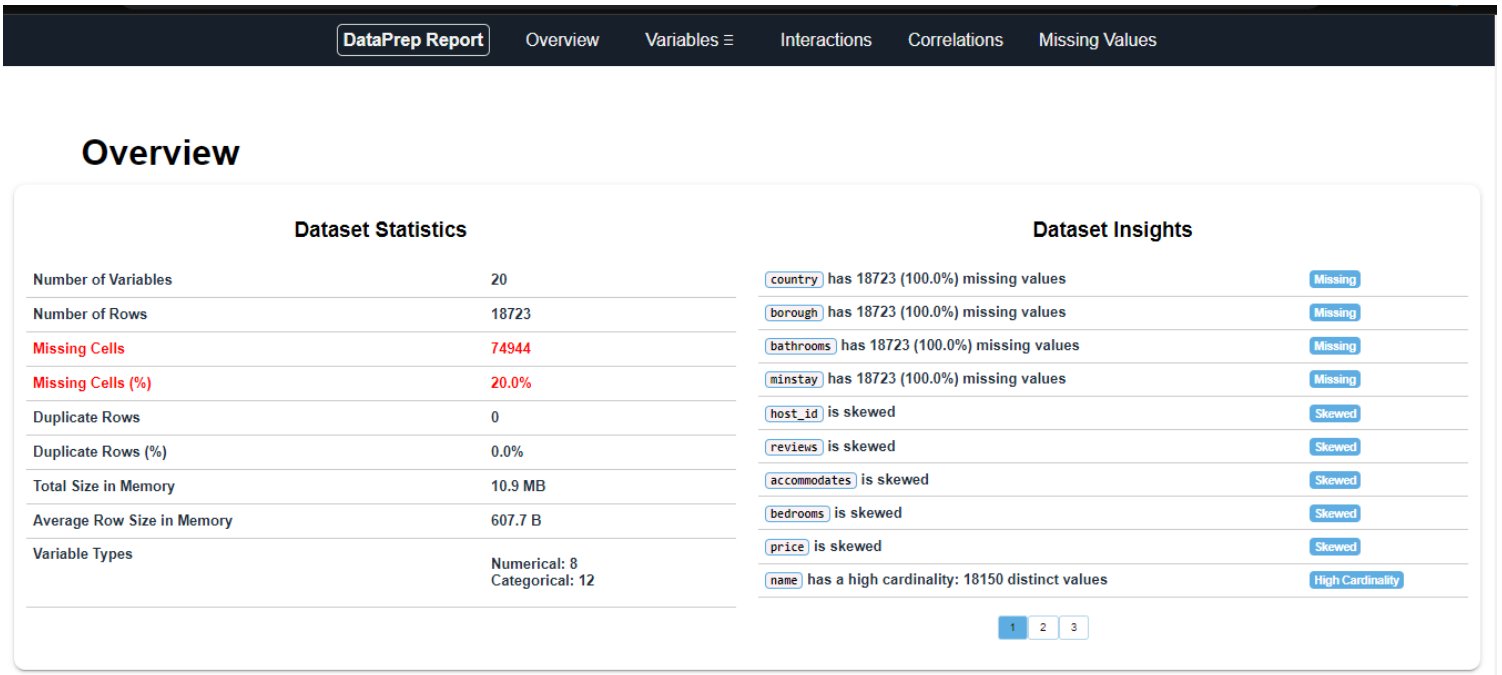
2.1 Listings Dataset



room_id	survey_id	host_id	room_typ	country	city	borough	neighbor	reviews	overall_s	accommo	bedroom	bathroon	price	minstay	name	last_moc	latitude	longitud	location
10176931	1476	49180562	Shared room		Amsterdam		De Pijp /	7	4.5	2	1		156		Red Light	06:27.4	52.3562	4.88749	0101000020E610000033FAD170CA8C13403BC5AA41982D4A40
8935871	1476	46718394	Shared room		Amsterdam		Centrum	45	4.5	4	1		126		Sunny an	06:23.6	52.3785	4.89612	0101000020E6100000842A357BA095134042791F4773304A40
14011697	1476	10346595	Shared room		Amsterdam		Watergra	1	0	3	1		132		Amsterda	06:23.6	52.3388	4.94359	0101000020E6100000A51133FB3CC613403543AA285E2B4A40
6137978	1476	8685430	Shared room		Amsterdam		Centrum	7	5	4	1		121		Canal bo	06:22.7	52.3763	4.89003	0101000020E6100000DF180280638F134085EE92382B304A40
18630616	1476	70191803	Shared room		Amsterdam		De Baars	1	0	2	1		93		One roon	06:19.7	52.3704	4.85287	0101000020E6100000CD902A8A5769134018 7B2FBE682F4A40
5790170	1476	29968916	Shared room		Amsterdam		De Pijp /	184	4.5	2	1		102		Beautifu	06:19.7	52.3423	4.89713	0101000020E6100000B090B932A896134060C8EA56CF2B4A40
934060	1476	5037506	Shared room		Amsterdam		Oostelijk	67	5	16	1		462		LOTUS, Cl	06:10.0	52.3776	4.93042	0101000020E61000005D70067FBFB813400B45BA9F53304A40
19590049	1476	132687356	Shared room		Amsterdam		Westerpa	2	0	2	1		414		big boot	06:10.0	52.3752	4.86612	0101000020E6100000DD09F65FE7761340D925AAB706304A40
5020280	1476	4059485	Shared room		Amsterdam		Oud Oost	2	0	2	1		222		Bright m	06:07.5	52.3573	4.91289	0101000020E61000032C687D9CBA613409FAD8383BD2D4A40
15810783	1476	84978218	Shared room		Amsterdam		Centrum	0	0	12	1		301		CANAL BC	06:07.4	52.3866	4.89013	0101000020E6100000FB03E5B67D8F13403D27BD6F7C314A40
9188521	1476	29968916	Shared room		Amsterdam		De Pijp /	26	5	4	1		246		Beautifu	06:07.4	52.3424	4.89695	0101000020E6100000BC96900F7A961340E0D6DD3CD52B4A40
10162121	1476	4555726	Shared room		Amsterdam		Centrum	10	4.5	4	1		198		APARTME	06:05.1	52.3663	4.89625	0101000020E6100000790778D2C2951340923CD7F7E12E4A40
14217287	1476	7490631	Shared room		Amsterdam		Centrum	0	0	2	1		180		Luxury ro	06:01.7	52.3606	4.8889	0101000020E61000004D2EC6C03A8E1340751F80D4262E4A40
19304350	1476	89256471	Shared room		Amsterdam		De Baars	1	0	2	1		180		Apartmei	06:01.7	52.3582	4.85681	0101000020E6100000058BC3995F6D1340C6DE8B2FDA2D4A40
9060570	1476	47264416	Shared room		Amsterdam		Oostelijk	15	4	4	1		48		sleeping	05:45.8	52.3784	4.94113	0101000020E61000004F232D95B7C3134057B439CE6D304A40
13797285	1476	12570608	Shared room		Amsterdam		Oostelijk	2	0	1	1		42		Double s	05:45.7	52.3655	4.93731	0101000020E61000006C76A4FACEBF1340583849F3C72E4A40
8692643	1476	12950894	Shared room		Amsterdam		Geuzenv	3	4	1	1		45		Single bc	05:45.7	52.3795	4.80099	0101000020E6100000A915A6EF353413408527F4FA93304A40
12314602	1476	31766654	Shared room		Amsterdam		De Pijp /	11	4	2	1		48		Simple s	05:45.7	52.35	4.89961	0101000020E6100000459E245D33991340A439B2F2CB2C4A40
18985831	1476	22542624	Shared room		Amsterdam		Buitenve	6	5	2	1		60		Comforta	05:45.7	52.3293	4.86181	0101000020E6100000084B9DDCB7D72134052499D80262A4A40
11122903	1476	13122312	Shared room		Amsterdam		Buitenve	6	4.5	2	1		48		Amsterda	05:45.7	52.3263	4.85773	0101000020E6100000B952CF82506E13404983DBDAC2294A40
9088830	1476	47264416	Shared room		Amsterdam		Oud Noo	73	4.5	4	1		48		Sleep sh	05:45.7	52.3906	4.94011	0101000020E6100000014F5AB8ACC21340126BF12900324A40
9537235	1476	45584652	Shared room		Amsterdam		Westerpa	1	0	1	1		58		Room in	05:45.7	52.3821	4.87798	0101000020E6100000172B6A300D8313406571FF91E9304A40
14309046	1476	31168044	Shared room		Amsterdam		De Baars	60	4.5	2	1		54		Chez Eric	05:45.7	52.36	4.86398	0101000020E6100000A85489B2B77413405AB91798152E4A40
19243789	1476	47752678	Shared room		Amsterdam		De Pijp /	1	0	1	1		60		Apparten	05:45.7	52.3476	4.90003	0101000020E6100000 6CCEC133A1991340DDD0949D7E2C4A40
18337741	1476	1352571	Shared room		Amsterdam		Bos en Lo	9	5	1	1		54		WOMEN (	05:45.7	52.3777	4.85086	0101000020E6100000658D7A8846671340CD76853E58304A40
11785182	1476	60750068	Shared room		Amsterdam		Slotervaa	2	0	4	1		18		Relax 9 h	05:41.4	52.3656	4.84937	0101000020E61000005C566133C06513407F2F8507CD2E4A40
11522696	1476	60750068	Shared room		Amsterdam		Slotervaa	6	1.5	4	1		25		I'm Easy	05:41.4	52.3635	4.83975	0101000020E6100000007EBFF1CE65B13408B48EAA26882E4A40
12243455	1476	60750068	Shared room		Amsterdam		De Baars	1	0	4	1		34		Surprice	05:41.3	52.3644	4.85063	0101000020E6100000E8DA17D00B671340A0C1A6CEA32E4A40

2.1.1 Listings Dataset Overview –

The Listings dataset consists of a table with 18722 records and 20 features. Features are distributed as 11 Continuous features and 9 Categorical features. There are a total 20% of records having Missing values.

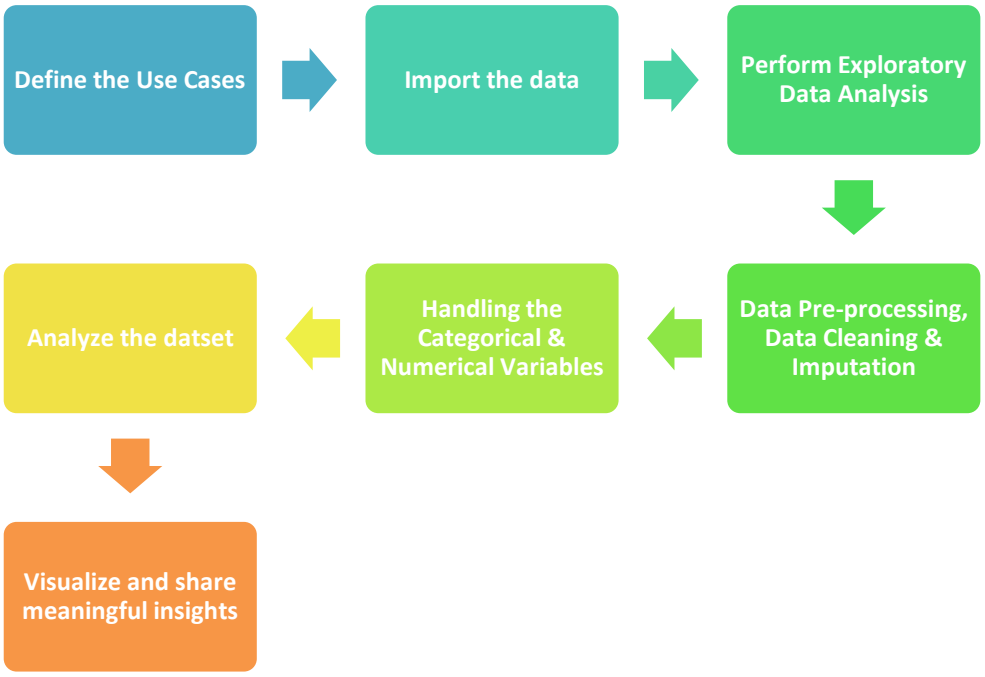


2.2 D-tale EDA Visual -

D-TALE Actions Visualize Highlight Settings

	room_id	survey_id	host_id	room_type	country	city	borough	neighborhood	reviews	overall_satisfaction	accommodates
0	10176931	1476	49180562	Shared room	Netherlands	Amsterdam	Centrum	De Pijp / Rivierenbuurt	7	4.50	2
1	8935871	1476	46718394	Shared room	Netherlands	Amsterdam	Centrum	Centrum West	45	4.50	4
2	14011697	1476	10346595	Shared room	Netherlands	Amsterdam	Centrum	Watergraafsmeer	1	0.00	3
3	6137978	1476	8685430	Shared room	Netherlands	Amsterdam	Centrum	Centrum West	7	5.00	4
4	18630616	1476	70191803	Shared room	Netherlands	Amsterdam	Centrum	De Baarsjes / Oud West	1	0.00	2
5	5790170	1476	29968916	Shared room	Netherlands	Amsterdam	Centrum	De Pijp / Rivierenbuurt	184	4.50	2
6	934060	1476	5037506	Shared room	Netherlands	Amsterdam	Centrum	Oostelijk Havengebied / Indische Buurt	67	5.00	16
7	19590049	1476	132687356	Shared room	Netherlands	Amsterdam	Centrum	Westerpark	2	0.00	2
8	5020280	1476	4059485	Shared room	Netherlands	Amsterdam	Centrum	Oud Oost	2	0.00	2
9	15810783	1476	84978218	Shared room	Netherlands	Amsterdam	Centrum	Centrum West	0	0.00	12
10	9188521	1476	29968916	Shared room	Netherlands	Amsterdam	Centrum	De Pijp / Rivierenbuurt	26	5.00	4
11	10162121	1476	4555726	Shared room	Netherlands	Amsterdam	Centrum	Centrum Oost	10	4.50	4
12	14217287	1476	7490631	Shared room	Netherlands	Amsterdam	Centrum	Centrum Oost	0	0.00	2
13	19304350	1476	89256471	Shared room	Netherlands	Amsterdam	Centrum	De Baarsjes / Oud West	1	0.00	2
14	9060570	1476	47264416	Shared room	Netherlands	Amsterdam	Centrum	Oostelijk Havengebied / Indische Buurt	15	4.00	4
15	13797285	1476	12570608	Shared room	Netherlands	Amsterdam	Centrum	Oostelijk Havengebied / Indische Buurt	2	0.00	1
16	8692643	1476	12950894	Shared room	Netherlands	Amsterdam	Centrum	Geuzenveld / Slotermeer	3	4.00	1
17	12314602	1476	31766654	Shared room	Netherlands	Amsterdam	Centrum	De Pijp / Rivierenbuurt	11	4.00	2
18	18985831	1476	22542624	Shared room	Netherlands	Amsterdam	Centrum	Buitenveldert / Zuidas	6	5.00	2
19	11122903	1476	13122312	Shared room	Netherlands	Amsterdam	Centrum	Buitenveldert / Zuidas	6	4.50	2
20	9088830	1476	47264416	Shared room	Netherlands	Amsterdam	Centrum	Oud Noord	73	4.50	4
21	9537235	1476	45584652	Shared room	Netherlands	Amsterdam	Centrum	Westerpark	1	0.00	1
22	14309046	1476	31168044	Shared room	Netherlands	Amsterdam	Centrum	De Baarsjes / Oud West	60	4.50	2
23	19243789	1476	47752678	Shared room	Netherlands	Amsterdam	Centrum	De Pijp / Rivierenbuurt	1	0.00	1

3 Architecture



3.1 Architecture Description –

3.1.1 Data Description –

As we have seen earlier, in our listings dataset, we have around 1.19 Lacs of records with 20 different features. Features are distributed as 10 Continuous features and 10 Categorical features and in our reviews dataset, we have around 3.44 Lacs of records with 6 different features among them there are 3 Continuous features and 3 Categorical features. These datasets are given in the form of Comma Separated Value (.csv) format.

3.1.2 Define the Use Cases –

At this stage, based on the given dataset and business problems we have defined the several Use Cases to perform the analysis on and this will definitely help out get the key insights from this data based on which business decisions will be taken. Furthermore, It helps in not only understanding the meaningful relationships between attributes but it also allows us to do our own research and come-up with our findings.

3.1.3 Import the Dataset –

As we have received the dataset in the form of Comma Separated Value (.csv) format, therefore we can import the same using Pandas read_csv( ) function.

READING CSV FILE

```
In [2]: df_airbnb = pd.read_csv('airbnb prices.csv')
```

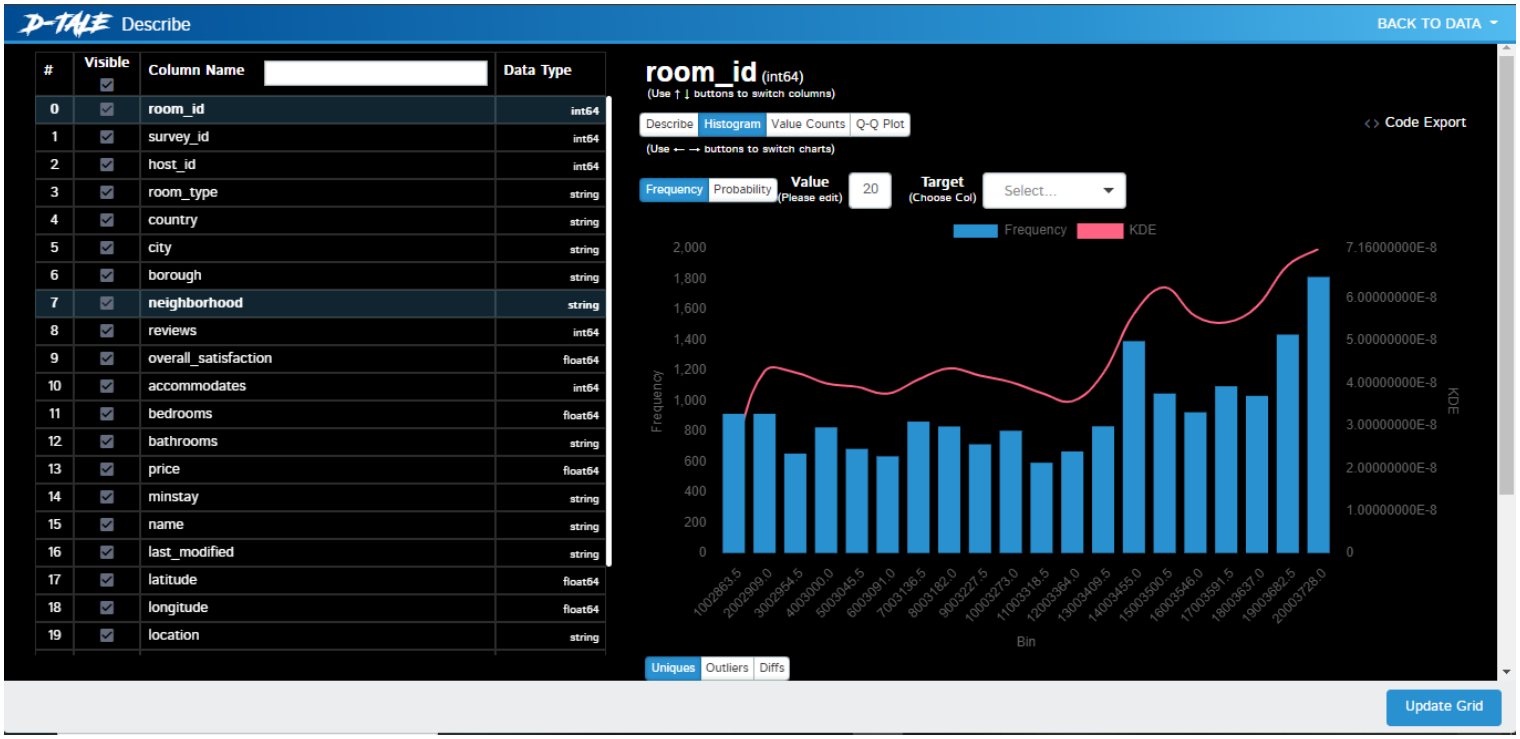
```
In [3]: df_airbnb.shape
Out[3]: (18723, 20)
```

```
In [4]: df_airbnb.head()
Out[4]:
```

	room_id	survey_id	host_id	room_type	country	city	borough	neighborhood	reviews	overall_satisfaction	accommodates	bedrooms	bathroor
0	10176931	1476	49180562	Shared room	NaN	Amsterdam	NaN	De Pijp / Rivierenbuurt	7	4.5	2	1.0	Na
1	8935871	1476	46718394	Shared room	NaN	Amsterdam	NaN	Centrum West	45	4.5	4	1.0	Na
2	14011697	1476	10346595	Shared room	NaN	Amsterdam	NaN	Watergraafsmeer	1	0.0	3	1.0	Na
3	6137978	1476	8685430	Shared room	NaN	Amsterdam	NaN	Centrum West	7	5.0	4	1.0	Na
4	18630616	1476	70191803	Shared room	NaN	Amsterdam	NaN	De Baarsjes / Oud West	1	0.0	2	1.0	Na

3.1.4 Exploratory Data Analysis (EDA) using D-Tale –

- "Exploratory Data Analysis" (EDA) is a "Data Exploration" step in the Data Analysis Process,where a number of techniques are used to better understand the dataset being used.
- Understanding the Dataset can refer to a number of things including but not limited to…
  - Extracting Important "Variables".
  - Identifying "Outliers", "Missing Values", or "Human Error".
  - Understanding the Relationships between variables.
  - Ultimately, maximizing our insights of a dataset and minimizing potential "Error" thatmay occur later in the process.
- In other words, it will gives you a better Understanding of the "Variables" andthe "Relationships" between them.
- Here, we make use of dataprep module to automate our EDA process.
- It provides the following information:
  - Overview: detect the types of columns in a DataFrame.
  - Variables: variable type, unique values, distinct count, missing values
  - Quartile statistics like minimum value, Q1, median, Q3, maximum, range, interquartilerange
  - Descriptive statistics like mean, mode, standard deviation, sum, median absolutedeviation, coefficient of variation, kurtosis, skewness.
  - Correlations: highlighting of highly correlated variables, Spearman, Pearson andKendall matrices
  - Missing Values: Bar Chart, Heatmap and spectrum of missing values.



3.1.5 Data Pre-processing,Data Cleaning & Imputation (Handling theCategorical & Numerical Variables) –

Data pre-processing is a process of preparing the raw data and making it suitable for our analysis purpose, where we have to do lot of Data Cleaning, handle the missing values by using appropriate imputation techniques and based on that variable nature i.e. either of Categorical & Numerical variable. Here, in this project, we have done the substitution/imputation of missing values using either mean, median or mode according to the nature of those variables. Moreover, we also removed the columns which are does not participate in our analysis.

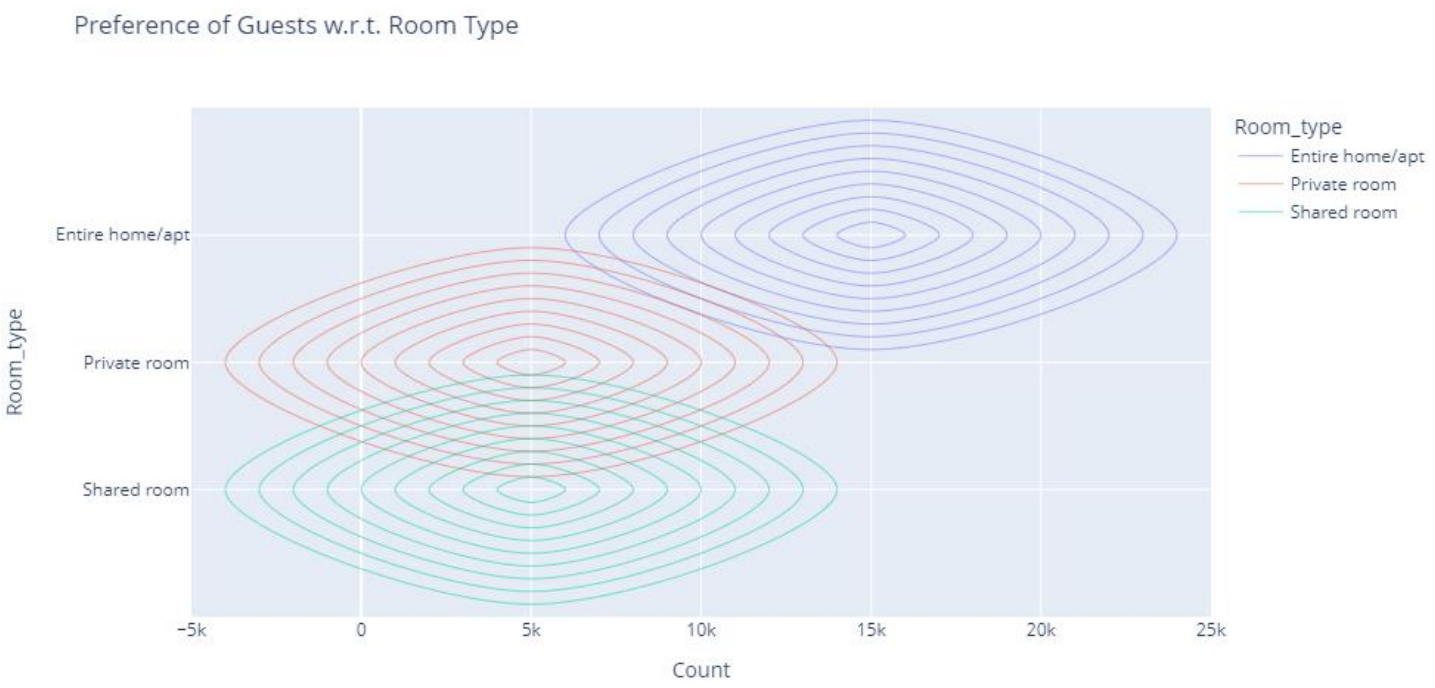
3.1.6 Analyse the Data –

Once the pre-processing is done, we are good to go with our actual analysis where we write lines of codes and logics to prepare our data as per the defined use cases.

3.1.7 Visualize & Share Meaningful Insights –

Finally, it is time to turn our data into some sort of visual representation. In short, Data visualization is the process of translating large data sets and metrics into charts, graphs and other visuals such as Bar Plot, Pie Chart, Heat map, Box Plot, Scatter Plot, and many more. The resulting visual representation of data makes it easier to identify and share insights about the information represented in the data.

Here is the beautiful glimpse of one of our visuals are –



All those different analysis help out to make better business decisions and help analyse customer trends and satisfaction, which can lead to new and better products and services.

4 Technology Stack

Data Manipulation & Mathematical Computation Library	Pandas, NumPy
Visualization Library	Matplotlib, Seaborn, Plotly, cufflinks,WordCloud, etc
EDA	Dataprep, D-tale
Dataset	.CSV Format
IDE	Jupyter Notebook