LOW LEVEL DESIGN (LLD)

Booking Data Analysis

(AirBnB Booking Analysis)

Written By / Author	Gaurav Rajgor
Document Version	LLD-V1.0
Last Revised Date	10/06/2022

Document Version Control

Date Issued	Version	Description	Author
10/06/2022	LLD-V1.0	First Version of Complete LLD	Gaurav Rajgor

Contents

Abstract

Airbnb is a corporation based in the United States that operates an online marketplace for lodging, primarily homestays for holiday rentals, and tourism activities. It essentially connects travelers with local hosts who want to rent out their homes with others looking for lodging in that area. This platform, on the other hand, allows hosts to list their available space and earn extra income through rent, while also allowing travelers to book unique homestays from local hosts, saving them money and providing them the opportunity to interact with locals.

The travel industry is advancing with the role of Data Science and Analytics in a world of rising new technology and innovation. Data analysis can help them understand their business in a new light and improve the quality of service by identifying the company's weak points. This study demonstrates how various analyses can help businesses make better decisions and analyze customer trends and satisfaction, which can lead to new and improved products and services. Various analyses, such as exploratory data analysis and descriptive analysis, were performed on a variety of use cases to obtain key insights from this data, which will be used to make business decisions.

1 Introduction

1.1 Why this Low-Level design document?

This LLD or Low-Level Design (LLD) document's purpose is to provide the internal logical design of the actual program code for the Airbnb Data Analysis project. LLD describes class diagrams with methods and relationships between classes and program specifications. It describes the modules in detail so that the programmer can code the program directly from the document. This document is intended for both stakeholders and project developers, and it will be submitted to upper management for approval.

The project's main goal is to analyze various aspects with different use cases that cover many aspects of Airbnb listings. It not only aids in identifying the meaningful links between features, but it also allows us to do our own research and present our findings.

1.2 Scope

Low-level design (LLD) is a component-level design process that involves iterative refinement. This method can be used to create data structures, software architecture, source code, and, ultimately, performance algorithms. Overall, the data organization can be defined during the requirement analysis phase and then refined during the data design phase.

This study demonstrates how various analyses can help businesses make better decisions and analyze customer trends and satisfaction, which can lead to new and improved products and services.

1.3 Constraints

The analysis must be user friendly, the code must be neat and clean, and EDA should be automated as much as possible to save time. Furthermore, users should not be required to have any coding knowledge because the insights they seek are detailed with accompanying visuals.

2 Technical Specifications

2.1 Listings Dataset –

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	reviews_per_month	calculated_host_listings_count	availability_365
2	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.6475	-73.97237	Private room	149	1	9	19-10-2018	0.21	6	365
3	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.7536	-73.98377	Entire home/apt	225	1	45	21-05-2019	0.38	2	355
4	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.809	-73.9419	Private room	150	3	0			1	365
5	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.6851	-73.95976	Entire home/apt	89	1	270	05-07-2019	4.64	1	194
6	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.7985	-73.94399	Entire home/apt	80	10	9	19-11-2018	0.1	1	0
7	5099	Large Cozy 1 BR Apartment In Midtown East	7322	Chris	Manhattan	Murray Hill	40.7477	-73.975	Entire home/apt	200	3	74	22-06-2019	0.59	1	129
8	5121	BlissArtsSpace!	7356	Garon	Brooklyn	Bedford-Stuyvesant	40.6869	-73.95596	Private room	60	45	49	05-10-2017	0.4	1	0
9	5178	Large Furnished Room Near B'way	8967	Shunichi	Manhattan	Hell's Kitchen	40.7649	-73.98493	Private room	79	2	430	24-06-2019	3.47	1	220
10	5203	Cozy Clean Guest Room - Family Apt	7490	MaryEllen	Manhattan	Upper West Side	40.8018	-73.96723	Private room	79	2	118	21-07-2017	0.99	1	0
11	5238	Cute & Cozy Lower East Side 1 bdrm	7549	Ben	Manhattan	Chinatown	40.7134	-73.99037	Entire home/apt	150	1	160	09-06-2019	1.33	4	188
12	5295	Beautiful 1br on Upper West Side	7702	Lena	Manhattan	Upper West Side	40.8032	-73.96545	Entire home/apt	135	5	53	22-06-2019	0.43	1	6
13	5441	Central Manhattan/near Broadway	7989	Kate	Manhattan	Hell's Kitchen	40.7608	-73.98867	Private room	85	2	188	23-06-2019	1.5	1	39
14	5803	Lovely Room 1, Garden, Best Area, Legal rental	9744	Laurie	Brooklyn	South Slope	40.6683	-73.98779	Private room	89	4	167	24-06-2019	1.34	3	314
15	6021	Wonderful Guest Bedroom in Manhattan for SINGLES	11528	Claudio	Manhattan	Upper West Side	40.7983	-73.96113	Private room	85	2	113	05-07-2019	0.91	1	333
16	6090	West Village Nest - Superhost	11975	Alina	Manhattan	West Village	40.7353	-74.00525	Entire home/apt	120	90	27	31-10-2018	0.22	1	0
17	6848	Only 2 stops to Manhattan studio	15991	Allen & Irina	Brooklyn	Williamsburg	40.7084	-73.95352	Entire home/apt	140	2	148	29-06-2019	1.2	1	46
18	7097	Perfect for Your Parents + Garden	17571	Jane	Brooklyn	Fort Greene	40.6917	-73.97185	Entire home/apt	215	2	198	28-06-2019	1.72	1	321
19	7322	Chelsea Perfect	18946	Doti	Manhattan	Chelsea	40.7419	-73.99501	Private room	140	1	260	01-07-2019	2.12	1	12
20	7726	Hip Historic Brownstone Apartment with Backyard	20950	Adam And Chari	Brooklyn	Crown Heights	40.6759	-73.94694	Entire home/apt	99	3	53	22-06-2019	4.44	1	21
21	7750	Huge 2 BR Upper East Cental Park	17985	Sing	Manhattan	East Harlem	40.7969	-73.94872	Entire home/apt	190	7	0			2	249
22	7801	Sweet and Spacious Brooklyn Loft	21207	Chaya	Brooklyn	Williamsburg	40.7184	-73.95718	Entire home/apt	299	3	9	28-12-2011	0.07	1	0
23	8024	CBG CtyBGd HelpsHaiti rm#1:1-4	22486	Lisel	Brooklyn	Park Slope	40.6807	-73.97706	Private room	130	2	130	01-07-2019	1.09	6	347
24	8025	CBG Helps Haiti Room#2.5	22486	Lisel	Brooklyn	Park Slope	40.6799	-73.97798	Private room	80	1	39	01-01-2019	0.37	6	364
25	8110	CBG Helps Haiti Rm #2	22486	Lisel	Brooklyn	Park Slope	40.68	-73.97865	Private room	110	2	71	02-07-2019	0.61	6	304
26	8490	MAISON DES SIRENES1,bohemian apartment	25183	Nathalie	Brooklyn	Bedford-Stuyvesant	40.6837	-73.94028	Entire home/apt	120	2	88	19-06-2019	0.73	2	233
27	8505	Sunny Bedroom Across Prospect Park	25326	Gregory	Brooklyn	Windsor Terrace	40.656	-73.97519	Private room	60	1	19	23-06-2019	1.37	2	85
28	8700	Magnifique Suite au N de Manhattan - vue Cloitres	26394	Claude & Sophie	Manhattan	Inwood	40.8675	-73.92639	Private room	80	4	0			1	0
29	9357	Midtown Pied-a-terre	30193	Tommi	Manhattan	Hell's Kitchen	40.7672	-73.98533	Entire home/apt	150	10	58	13-08-2017	0.49	1	75
30	9518	SPACIOUS, LOVELY FURNISHED MANHATTAN BEDROON	31374	Shon	Manhattan	Inwood	40.8648	-73.92106	Private room	44	3	108	15-06-2019	1.11	3	311
31	9657	Modern 1 BR / NYC / EAST VILLAGE	21904	Dana	Manhattan	East Village	40.7292	-73.98542	Entire home/apt	180	14	29	19-04-2019	0.24	1	67
32	9668	front room/double bed	32294	Ssameer Or Trip	Manhattan	Harlem	40.8225	-73.95104	Private room	50	3	242	01-06-2019	2.04	3	355
33	9704	Spacious 1 bedroom in luxe building	32045	Teri	Manhattan	Harlem	40.8131	-73.95466	Private room	52	2	88	14-06-2019	1.42	1	255

2.1.1 Listings Dataset Overview –

The Listings dataset consists of a table with 48895 records and 16 features. Features are distributed as 10 Continuous features and 6 Categorical features. There are a total 2.6% of records having Missing values.

My Report | Overview | Variables ☰ | Interactions | Correlations | Missing Values

Overview

Dataset Statistics

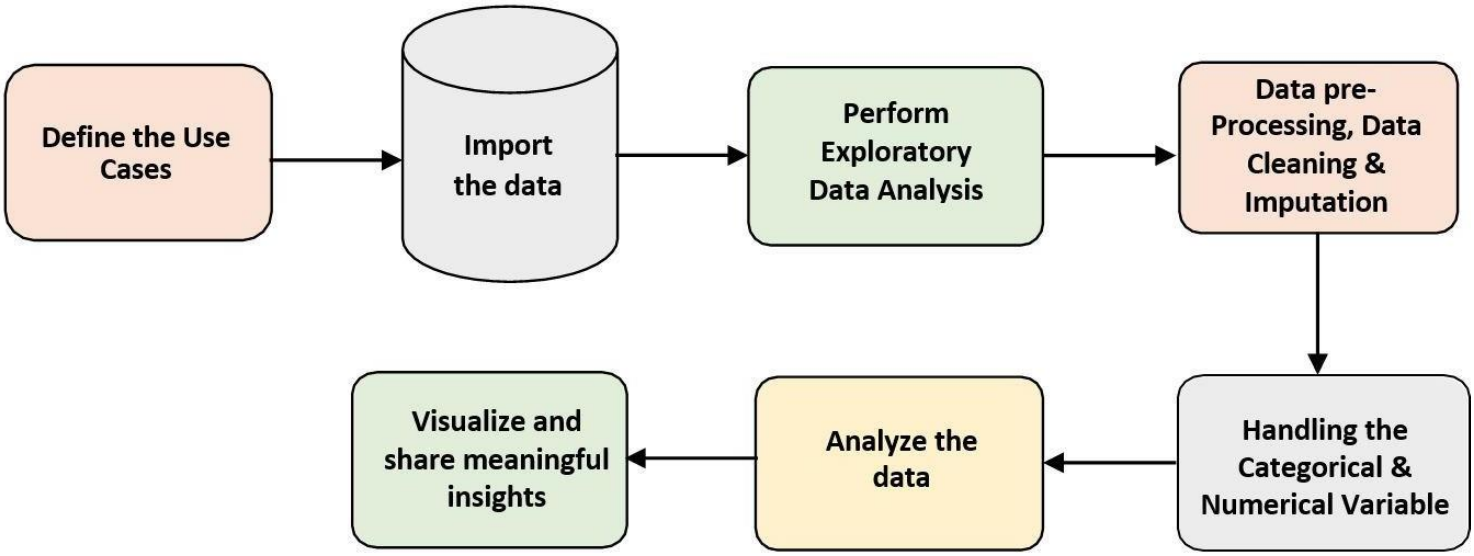
Number of Variables	16
Number of Rows	48895
Missing Cells	20141
Missing Cells (%)	2.6%
Duplicate Rows	0
Duplicate Rows (%)	0.0%
Total Size in Memory	23.5 MB
Average Row Size in Memory	504.1 B
Variable Types	Numerical: 10 Categorical: 6

Dataset Insights

`last_review` has 10052 (20.56%) missing values	Missing
`reviews_per_month` has 10052 (20.56%) missing values	Missing
`host_id` is skewed	Skewed
`longitude` is skewed	Skewed
`price` is skewed	Skewed
`minimum_nights` is skewed	Skewed
`number_of_reviews` is skewed	Skewed
`reviews_per_month` is skewed	Skewed
`calculated_host_listings_count` is skewed	Skewed
`availability_365` is skewed	Skewed

1 2

3 Architecture



3.1 Architecture Description –

3.1.1 Data Description –

As we have seen earlier, in our listing's dataset, we have around 48 thousand of records with 16 different features. Features are distributed as 10 Continuous features and 6 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, we have defined several Use Cases to perform the analysis on based on the given dataset and business problems, and this will undoubtedly help us get the key insights from this data on which business decisions will be made. Furthermore, it not only aids in understanding the meaningful relationships between attributes, but it also allows us to conduct our own research and come to our own conclusions.

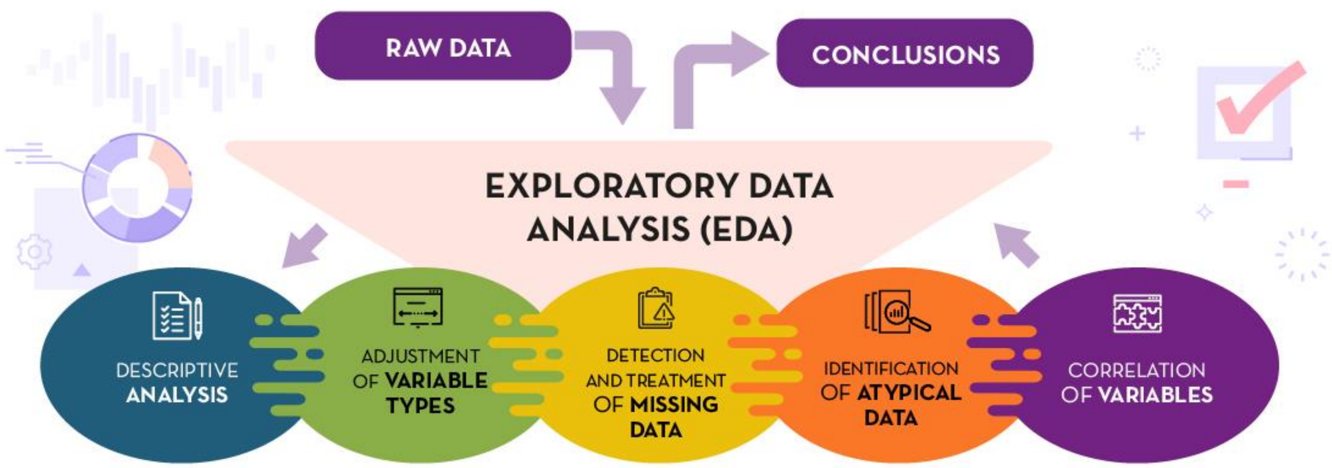
3.1.3 Import the Dataset –

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.

```
#Printing the dataset
airbnb_df=pd.read_csv(file_path)
pd.DataFrame(airbnb_df)
```

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	reviews_per_month	calculated_host_
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	9	2018-10-19	0.21	
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	2019-05-21	0.38	
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	0	NaN	NaN	
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	270	2019-07-05	4.64	
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	9	2018-11-19	0.10	

3.1.4 Exploratory Data Analysis (EDA) –



- "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" that may occur later in the process.
- In other words, it will give you a better Understanding of the "Variables" and the "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, interquartile range
    - Descriptive statistics like mean, mode, standard deviation, sum, median absolute deviation, coefficient of variation, kurtosis, skewness.
    - Correlations: highlighting of highly correlated variables, Spearman, Pearson and Kendall matrices
    - Missing Values: Bar Chart, Heatmap and spectrum of missing values.

My Report | Overview | Variables ☰ | Interactions | Correlations | Missing Values

Overview

Dataset Statistics

Number of Variables	16
Number of Rows	48895
Missing Cells	20141
Missing Cells (%)	2.6%
Duplicate Rows	0
Duplicate Rows (%)	0.0%
Total Size in Memory	23.5 MB
Average Row Size in Memory	504.1 B
Variable Types	Numerical: 10 Categorical: 6

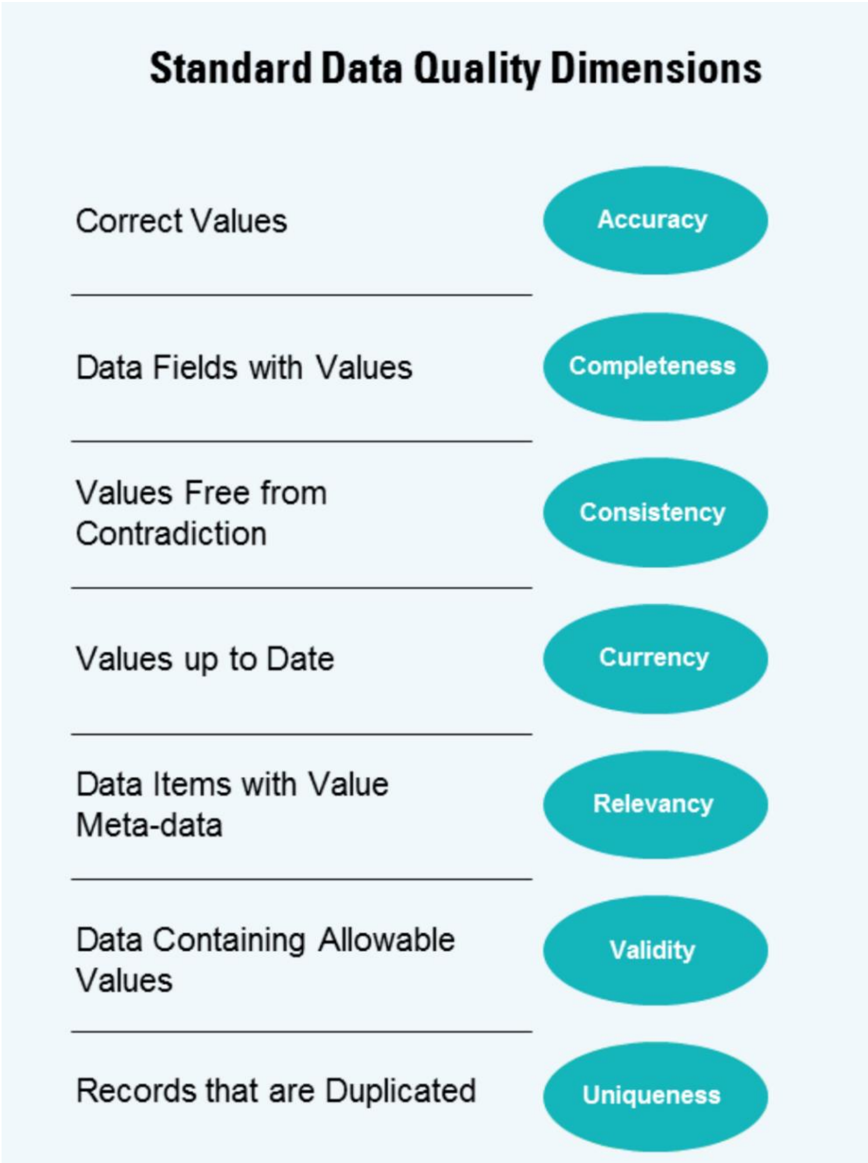
Dataset Insights

last_review has 10052 (20.56%) missing values	Missing
reviews_per_month has 10052 (20.56%) missing values	Missing
host_id is skewed	Skewed
longitude is skewed	Skewed
price is skewed	Skewed
minimum_nights is skewed	Skewed
number_of_reviews is skewed	Skewed
reviews_per_month is skewed	Skewed
calculated_host_listings_count is skewed	Skewed
availability_365 is skewed	Skewed

1 2

3.1.5 Data Pre-processing, Data Cleaning & Imputation (Handling the Categorical & 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.



Ref. KPMG Virtual Internship

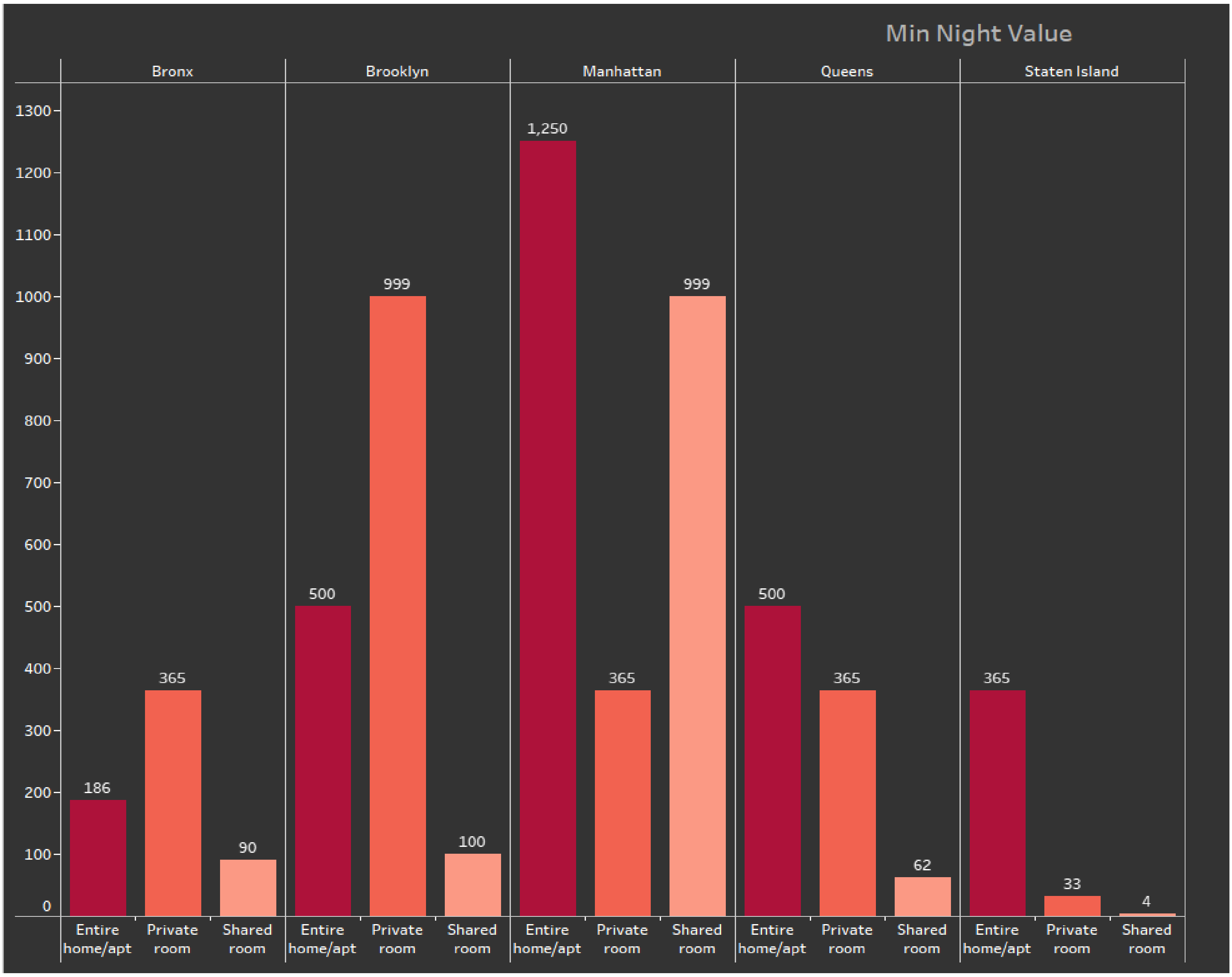
3.1.6 Analyze the Data –

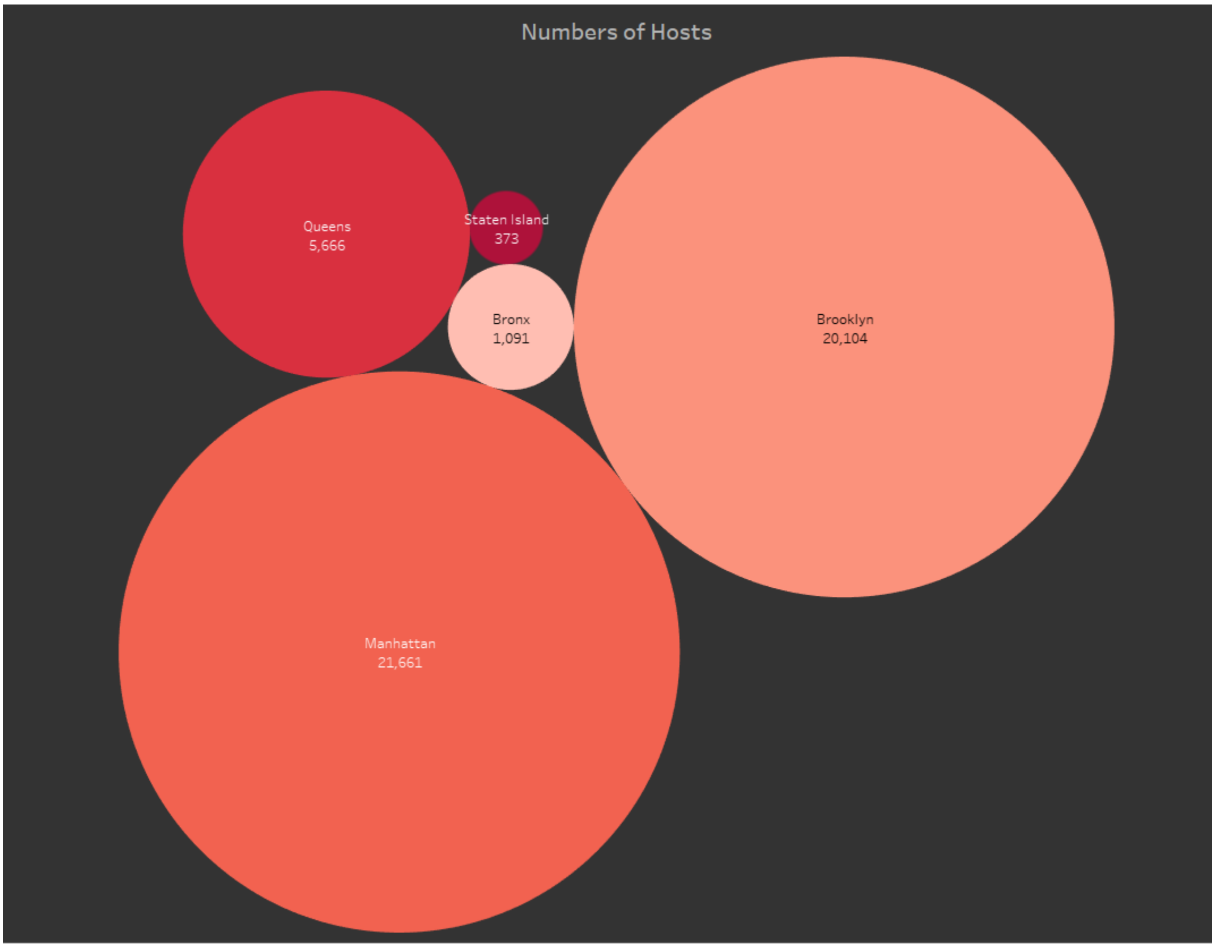
Once the pre-processing is complete, we can begin our actual analysis, in which we write lines of code and logic to prepare our data in accordance with the defined use cases.

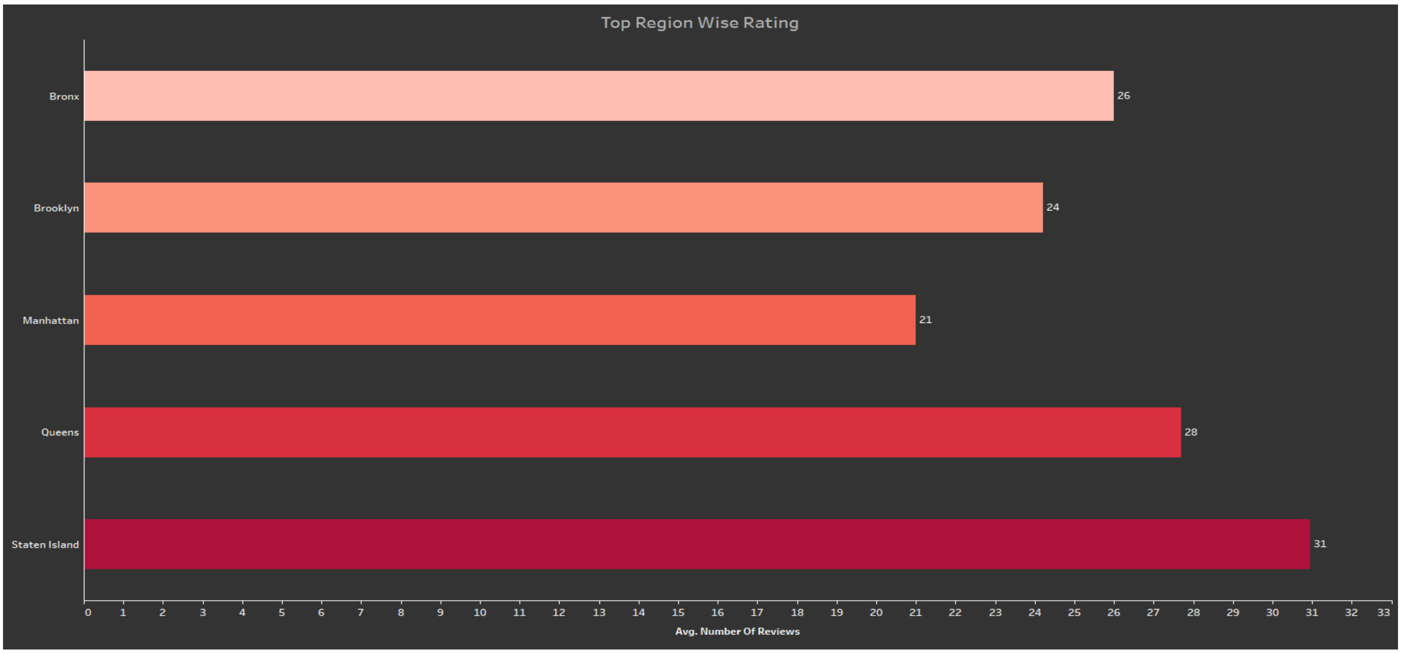
3.1.7 Visualize & Share Meaningful Insights –

Finally, it's time to transform our data into a visual representation. In a nutshell, data visualization is the process of converting large data sets and metrics into charts, graphs, and other visuals such as the Bar Plot, Pie Chart, Heat Map, Box Plot, Scatter Plot, and others. The resulting visual representation of data makes it easier to identify and share insights about the data's information.

Here is a beautiful preview of one of our visuals –




Numbers of Hosts
Queens
5,666
Staten Island
373
Bronx
1,091
Brooklyn
20,104
Manhattan
21,661


Top Region Wise Rating
Bronx
26
Brooklyn
24
Manhattan
21
Queens
28
Staten Island
31
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33
Avg. Number Of Reviews

4 Technology Stack

Data Manipulation & Mathematical ComputationLibrary	Pandas, NumPy
Visualization Library	Matplotlib, Seaborn, Plotly, etc
EDA	dataprep
Dataset	.CSV Format
IDE	Jupyter Notebook, Google Collab