# Architecture

## (Swiggy Data Analysis)

| | |
|---|---|
| Written By / Author | Hazel A |
| Document Version | V1.0 |
| Last Revised Date | 25/07/2022 |

## Document Version Control:

| Date | Version | Author | Comments |
|---|---|---|---|
| **25/07/2022** | V1.0 | Hazel A | First Draft |
| | | | |

## Approval Status:

| Version | Review Data | Reviewed By | Approved By | Comments |
|---|---|---|---|---|
| V1.0 | | | | |

# Contents

# 1 Introduction

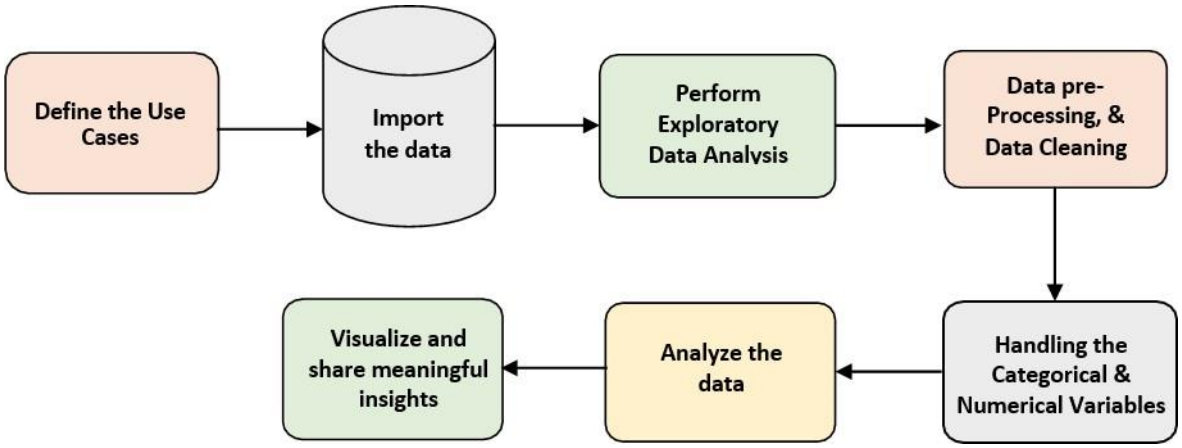
## 1.1 Why this Architecture design document?

The goal of this document is to provide a detailed architecture design of the Airbnb Data Analysis Project, focusing on each of our architecture's attributes.
This document will discuss the project's history as well as the architecturally significant function requirements. This document's goal is to assist the development team in determining how the system will be structured at the highest level.

## 1.2 Scope

Architecture Design Document (ADD) is a step-by-step refinement process in architecture design. The method can be used to create data structures, software architecture, source code, and, ultimately, performance algorithms. In general, design principles can be defined during requirement analysis and refined during architectural design work.

# 2 Architecture




## 2.1 Architecture Description –

### 2.1.1 Data Description –

Swiggy dataset, we have around 118 records with 5 different features. Features are distributed as 2 Continuous features and 3 Categorical features. These datasets are given in the form of Comma Separated Value (.csv) format.

### 2.1.2 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 help 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.

### 2.1.3 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.

```
In [2]:
1 df_swiggy= pd.read_csv("E:\PGP GL\Ineuron/Swiggy Bangalore Outlet Details.csv")
2 df_swiggy
3
```

Out[2]:

| | Shop_Name | Cuisine | Location | Rating | Cost_for_Two |
|---|---|---|---|---|---|
| 0 | Kanti Sweets | Sweets | Koramangala, Koramangala | 4.3 | ₹ 150 |
| 1 | Mumbai Tiffin | North Indian, Home Food, Thalis, Combo | Sector 5, HSR | 4.4 | ₹ 400 |
| 2 | Sri Krishna sagar | South Indian, North Indian, Fast Food, Beverag... | 6th Block, Koramangala | 4.1 | ₹ 126 |
| 3 | Al Daaz | American, Arabian, Chinese, Desserts, Fast Foo... | HSR, HSR | 4.4 | ₹ 400 |
| 4 | Beijing Bites | Chinese, Thai | 5th Block, Koramangala | 4.1 | ₹ 450 |
| ... | ... | ... | ... | ... | ... |
| 113 | Wok Paper Scissors | Pan-Asian, Chinese, Asian | JNC Road, Koramangala | 3.9 | ₹ 219 |
| 114 | Savoury Restaurant | Arabian, Middle Eastern, North Indian, Grill, ... | Madiwala, BTM | 4.1 | ₹ 600 |
| 115 | Royal Treat | North Indian, Chinese, Seafood, Biryani | 5th block Koramangala, Koramangala | 4.2 | ₹ 193 |
| 116 | Thali 99 | North Indian | Koramangala, Koramangala | 4.3 | ₹ 200 |
| 117 | Mani's Dum Biryani | Andhra, Biryani | 1st Block, Koramangala | 4.2 | ₹ 400 |

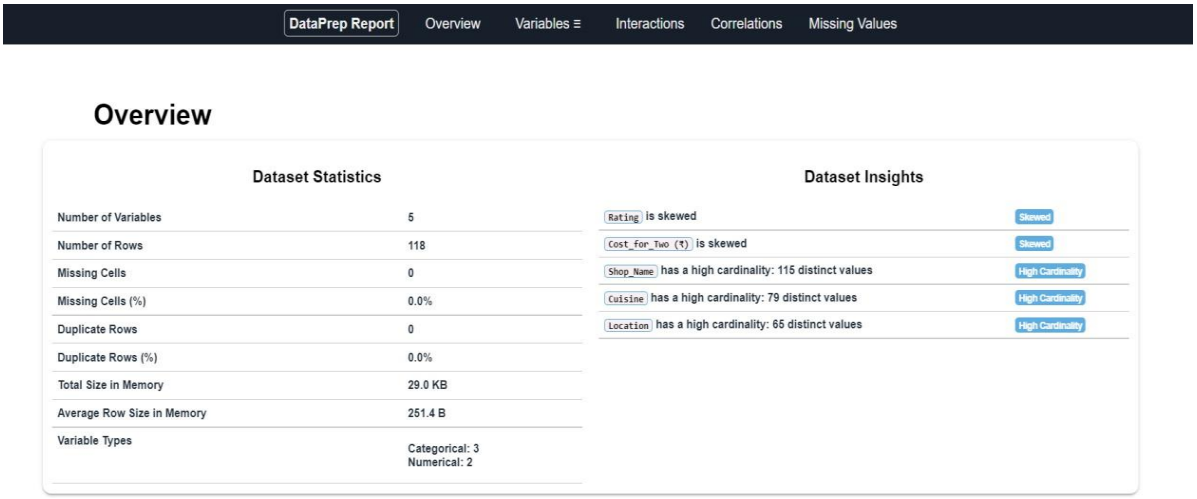
118 rows × 5 columns

### 2.1.4 Exploratory Data Analysis (EDA) –

- Data Analysis Process that employs a variety of techniques to better understand the dataset under consideration.
- • Understanding the Dataset can mean a variety of things, including but not limited to...
- Obtaining Important "Variables"
- Detecting "Outliers," "Missing Values," or "Human Error."
- Understanding the Interrelationships of Variables
- Finally, we want to maximise our insights from a dataset while minimising potential "Error" that may occur later in the process.
- • In other words, it will help you understand the "Variables" and the "Relationships" between them.
  • In this case, we use the dataprep module to automate our EDA process.

• It includes the following details:

1. Overview: determine the column types in a DataFrame.
2. Variables include variable type, distinct values, distinct count, and missing values.
3. Minimum value, Q1, median, Q3, maximum, range, and interquartile range are examples of quartile statistics.
4. Mean, mode, standard deviation, sum, median absolute deviation, coefficient of variation, kurtosis, and skewness are examples of descriptive statistics.
5. Correlations: highlighting highly correlated variables using the Spearman, Pearson, and Kendall matrices




### 2.1.5 Data Pre-processing, Data Cleaning & Imputation (Handling the Categorical & Numerical Variables) –

Data pre-processing is the process of preparing raw data for analysis, where we have to do a lot of data cleaning, handle missing values by using appropriate imputation techniques, and based on the variable nature, which can be categorical or numerical. In this project, we substituted/imputation of missing values using either mean, median, or mode depending on the nature of the variables.

### 2.1.6 Analyse the Data –

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.

### 2.1.7 Visualize & Share Meaningful Insights –

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 help analyse customer trends and satisfaction, which can lead to new and better products and services.