

# Low Level Design (LLD)

## (Swiggy Data Analysis)

## Document Version Control

Date Issued	Version	Description	Author
25/07/2022	LLD-V1.0	First Version of Complete LLD	Hazel Abraham

## Contents

Document Version Control	2
<b>Abstract</b>	4
<b>1 Introduction</b>	5
1.1 Why this Low-Level design document?	5
1.2 Scope	5
1.3 Constraints	5
<b>2 Technical Specifications</b>	6
2.1 Swiggy Dataset	6
2.1.1 Swiggy Dataset Overview	6
2.1.2 Overview of Variables	7
<b>3 Architecture</b>	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
<b>4 Technology Stack</b>	11

## Abstract

The market for ordering food online comprises dishes made by independent chefs, restaurants, and consumers who order goods online for pick-up or delivery. The act of ordering meals from a website or other application is known as online food ordering. Food that hasn't been specially prepared for directed eating or food that is ready to consume can both be considered products.

The food sector is advancing with the help of data science and analytics in the realm of emerging new technology and innovation. By highlighting the weak points of the company, data analysis can help them view their business from a very different perspective and enhance the level of service they provide. This study highlights how various analyses can be used to improve corporate decisions and assess consumer trends and satisfaction, which can result in the development of fresh, improved goods and services. In order to extract the important insights from this data on the basis on which business decisions would be made, several analyses, such as exploratory data analysis and descriptive analysis, were performed on a variety of use cases.

# 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 Swiggy 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 analyse the various aspects with different use cases which covers many aspects of Swiggy Food Delivery Service.

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

## 1.3 Constraints

The analysis must be user readable 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 Swiggy Dataset –

	A	B	C	D	E
1	Shop_Name	Cuisine	Location	Rating	Cost_for_Two
2	Kanti Sweets	Sweets	Koramangala, Koramangala	4.3	₹ 150
3	Mumbai Tiffin	North Indian, Home Food, Thalys, Combo	Sector 5, HSR	4.4	₹ 400
4	Sri Krishna sagar	South Indian, North Indian, Fast Food, Beverages, Jain	6th Block, Koramangala	4.1	₹ 126
5	Al Daaz	American, Arabian, Chinese, Desserts, Fast Food, Mughlai, North Indian	HSR, HSR	4.4	₹ 400
6	Beijing Bites	Chinese, Thai	5th Block, Koramangala	4.1	₹ 450
7	Kitchens of Punjab	North Indian	Koramangala 4th Block, Koramangala	4.2	₹ 350
8	99 VARIETY DOSA AND PAV BHAIJI	Fast Food, North Indian, Chinese	BTM 2nd Stage, BTM	4.1	₹ 200
9	La Pino'z Pizza	Italian	BTM, BTM	3.9	₹ 500
10	Hotel Manu	South Indian, Kerala, Chinese, North Indian	HSR, HSR	4.1	₹ 350
11	Yumlane Pizza	Pizzas, Italian, Mexican	9th Main road, Koramangala	3.8	₹ 150
12	Ambur Star Briyani	Chinese, South Indian, North Indian, Desserts, Fast Food, Kerala, Andhra Pradesh	outer ring road, BTM	4.1	₹ 500
13	Cake Box	Desserts	Koramangala, Koramangala	4	₹ 247
14	Meghana Foods	Chinese, Andhra, Biryani, Seafood	5th Block, Koramangala	4.3	₹ 550
15	Momoz	Chinese	5th Block, Koramangala	4.3	₹ 450
16	A2B - Adyar Ananda Bhavan	South Indian, Chinese, Desserts, North Indian	7th Block, Koramangala	4.2	₹ 450
17	Shawarma Inc	Arabian, Fast Food	1st Main, Koramangala	4.1	₹ 150
18	WarmOven Cake & Desserts	Desserts, Beverages	Koramangala, Koramangala	4.1	₹ 200
19	Sri Lakshmi Dhaba	North Indian	Bommanahalli, BTM	3.7	₹ 200
20	Falahaar & Kota Kachori	North Indian	6th block, Koramangala	4.2	₹ 300
21	Shree Khana Khazana	Indian, Rajasthani	Sector 4, HSR	4.1	₹ 350
22	Just Bake - Cakes & confectioners	Desserts, Bakery	BTM 1st stage, BTM	4.3	₹ 300
23	Maa Di Hatti	Chinese, Healthy Food, North Indian	Jakkasandra Extn, Koramangala	4	₹ 129
24	Hotel Godavari	North Indian, Chinese, Hyderabad	Marutnagar Main Road, BTM	4	₹ 400
25	Rolls On Wheels	Fast Food	1st Block, Koramangala	4.5	₹ 300
26	New Udipi Grand	Chinese, Jain, North Indian, South Indian	HSR, HSR	4.3	₹ 150
27	Swad Punjab da	Indian	BTM, BTM	4.1	₹ 250

#### 2.1.1 Swiggy Dataset Overview –

The Listings dataset consists of a table with 118 records and 5 features. Features are distributed as 2 Continuous features and 3 Categorical features. There are a total 0% of records having Missing Values. In short, there are no Missing Values present in the dataset.

### Overview

Dataset Statistics		Dataset Insights	
Number of Variables	5	<u>Shop_Na_</u> has a high cardinality: 115 distinct values	High Cardinality
Number of Rows	118	<u>Cuisine</u> has a high cardinality: 79 distinct values	High Cardinality
Missing Cells	0	<u>Location</u> has a high cardinality: 65 distinct values	High Cardinality
Missing Cells (%)	0.0%	<u>Cost_fo_</u> has constant length 5	Constant Length
Duplicate Rows	0		
Duplicate Rows (%)	0.0%		
Total Size in Memory	45.5 KB		
Average Row Size in Memory	395.1 B		
Variable Types	Categorical: 5		

## 2.1.2 Overview of Variables:

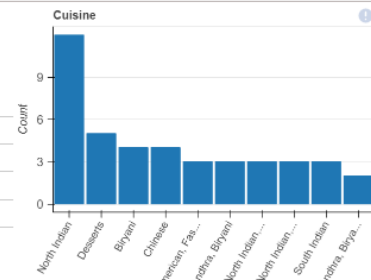
## Variables

Sort by Feature order ☐ Reverse orderShop\_Name  
categorical[Show Details](#)

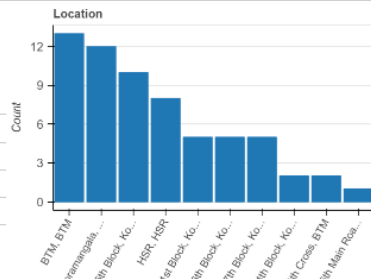
Approximate Distinct Count	115
Approximate Unique (%)	97.5%
Missing	0
Missing (%)	0.0%
Memory Size	9.3 KB

Cuisine  
categorical[Show Details](#)

Approximate Distinct Count	79
Approximate Unique (%)	67.0%
Missing	0
Missing (%)	0.0%
Memory Size	10.6 KB

Location  
categorical[Show Details](#)

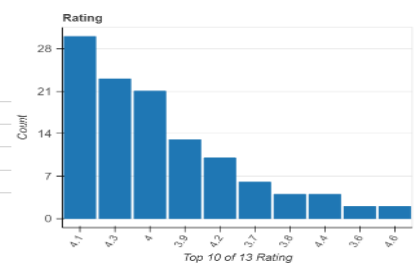
Approximate Distinct Count	65
Approximate Unique (%)	55.1%
Missing	0
Missing (%)	0.0%
Memory Size	9.8 KB



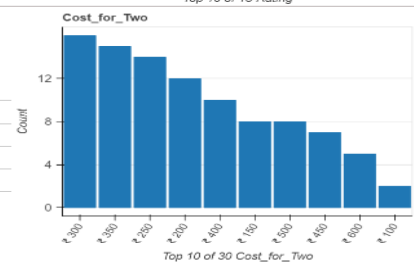
## Rating

Rating  
categorical[Show Details](#)

Approximate Distinct Count	13
Approximate Unique (%)	11.0%
Missing	0
Missing (%)	0.0%
Memory Size	7.8 KB

Cost\_for\_Two  
categorical[Show Details](#)

Approximate Distinct Count	30
Approximate Unique (%)	25.4%
Missing	0
Missing (%)	0.0%
Memory Size	12.4 KB



### 3 Architecture



#### 3.1 Architecture Description –

##### 3.1.1 Data Description –

We have 118 entries with a total of five different attributes in our Swiggy dataset, as we saw earlier. Two continuous features and three categorical features make up the distribution of features. The format of these datasets is Comma Separated Value (.csv).

##### 3.1.2 Define the Use Cases –

At this point, depending on the provided dataset and business problems, we have created a number of Use Cases to conduct the analysis on, and this will undoubtedly assist in getting the important insights from this data on the basis of which business decisions will be made. Furthermore, it enables us to do independent study and present our findings as well as identifying the significant links between features.

##### 3.1.3 Import the Dataset –

3.1.4 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, Thalís, Combo	Sector 5, HSR	4.4	₹ 400
2	Sri Krishna sagar	South Indian, North Indian, Fast Food, Beverag...	8th 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



### 3.1.5 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

DataPrep Report

Overview

Variables ≡

Interactions

Correlations

Missing Values

Overview

Dataset Statistics

Number of Variables	5
Number of Rows	118
Missing Cells	0
Missing Cells (%)	0.0%
Duplicate Rows	0
Duplicate Rows (%)	0.0%
Total Size in Memory	29.0 KB
Average Row Size in Memory	251.4 B
Variable Types	Categorical: 3 Numerical: 2

Dataset Insights

Rating is skewed	Skewed
Cost_for_Two (*) is skewed	Skewed
Shop_Name has a high cardinality: 115 distinct values	High Cardinality
Cuisine has a high cardinality: 79 distinct values	High Cardinality
Location has a high cardinality: 65 distinct values	High Cardinality

### 3.1.6 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. Furthermore, we removed the columns that did not contribute to our analysis.

### 3.1.7 Analyse the Data –

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

### 3.1.8 Visualize & Share Meaningful Insights –



In a nutshell, data visualisation 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.:

## 4 Technology Stack

<b>Data Manipulation Library</b>	Pandas
<b>Visualization Library</b>	Matplotlib, Seaborn, Plotly, etc
<b>EDA</b>	dataprep
<b>Dataset</b>	.CSV Format
<b>IDE</b>	Jupyter Notebook