Low Level Design (LLD)

(Analyzing Swiggy)

Written By / Author	Jaimin Koladiya
Document Version	LLD-V1.0
Last Revised Date	24/03/2023

Document Version Control

Date Issued	Version	Description	Author
24/03/2023	LLD-V1.0	First Version of Complete LLD	Jaimin Koladiya

Contents

Abstract

The online food ordering market includes foods prepared by restaurants, prepared by independent people, and groceries being ordered online and then picked up or delivered. The first online food ordering service, World Wide Waiter (now known as Waiter.com), was founded in 1995. Online food ordering is the process of ordering food from a website or other application. The product can be either ready-to-eat food or food that has not been specially prepared for direction consumption.

In the world of rising new technology and innovation, Food industry is advancing with the role of Data Science and Analytics. Data analysis can help them to understand their business in a quiet different manner and helps to improve the quality of the service by identifying the weak areas of the business. This study demonstrates the how different analysis help to make better business decisions and help analyze customer trends and satisfaction, which can lead to new and better products and services. Different analysis performed such as Extract, Transformed Load(ETL) Analysis and Descriptive Analysis on variety of use cases to get the key insights from this data based on which business decisions will be taken.

1 Introduction

1.1 Why this Low-Level design document?

The low level document (LLD) is essential for a project like "Analyzing Swiggy" because it provides a detailed description of the system's design and functionality. The LLD contains information on the system's architecture, data flow, and various components that work together to produce the desired results.

In this project, the LLD covers the Extract, Transform, and Load (ETL) process, which involves extracting data from the source file, transforming it into a suitable format, and loading it into a database. The LLD also includes the analysis of the Swiggy dataset, which provides insights into the delivery trends, popular cuisines, and top-rated restaurants.

The LLD acts as a guide for developers and other stakeholders involved in the project. It helps them understand how the system works, what components are involved, and how data flows through the system. The LLD also helps in identifying potential bottlenecks, performance issues, and areas for improvement.

Therefore, the low level document is essential for the "Analyzing Swiggy" project to ensure that the system is well-designed, efficient, and meets the project's requirements.

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 analyze 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, ETL 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 Bangalore Outlet Dataset

2.1.1 Swiggy Bangalore Outlet 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.

	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

2.1.2 Overview of Variables:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 118 entries, 0 to 117
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Shop_Name     118 non-null    object
 1   Cuisine       118 non-null    object
 2   Location      118 non-null    object
 3   Rating        118 non-null    object
 4   Cost_for_Two  118 non-null    object
dtypes: object(5)
memory usage: 4.7+ KB
```

3 Architecture

3.1 Architecture Description –

3.1.1 Data Description –

As we have seen earlier, in our Swiggy Bangalore Outlet 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.

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.

	Slide Type	Slide ⌄

Data Extraction

	Slide Type	Sub-Slide ⌄

```
data = pd.read_csv('Swiggy Bangalore Outlet Details.csv')
```

	Slide Type	Sub-Slide ⌄

```
data.head()
```

3.1.4 Extract, Transform, Load (ETL) –

- "Extract Transform Load" (ETL) 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 gives 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 Data Frame.
    - 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.

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.

	Slide Type	Slide

Data Cleaning

	Slide Type	Sub-Slide

```python
print(f"Checking Total null values in data :")
data.isna().sum().sum()
```

```
Checking Total null values in data :

0
```

	Slide Type	Sub-Slide

```python
print(f"Checking Total null Values in all columns of data :")
data.isna().sum()
```

```
Checking Total null Values in all columns of data :

Shop_Name       0
Cuisine         0
Location        0
Rating          0
Cost_for_Two    0
dtype: int64
```

|	Slide Type	Sub-Slide

```python
data.duplicated().sum()
```

```
0
```

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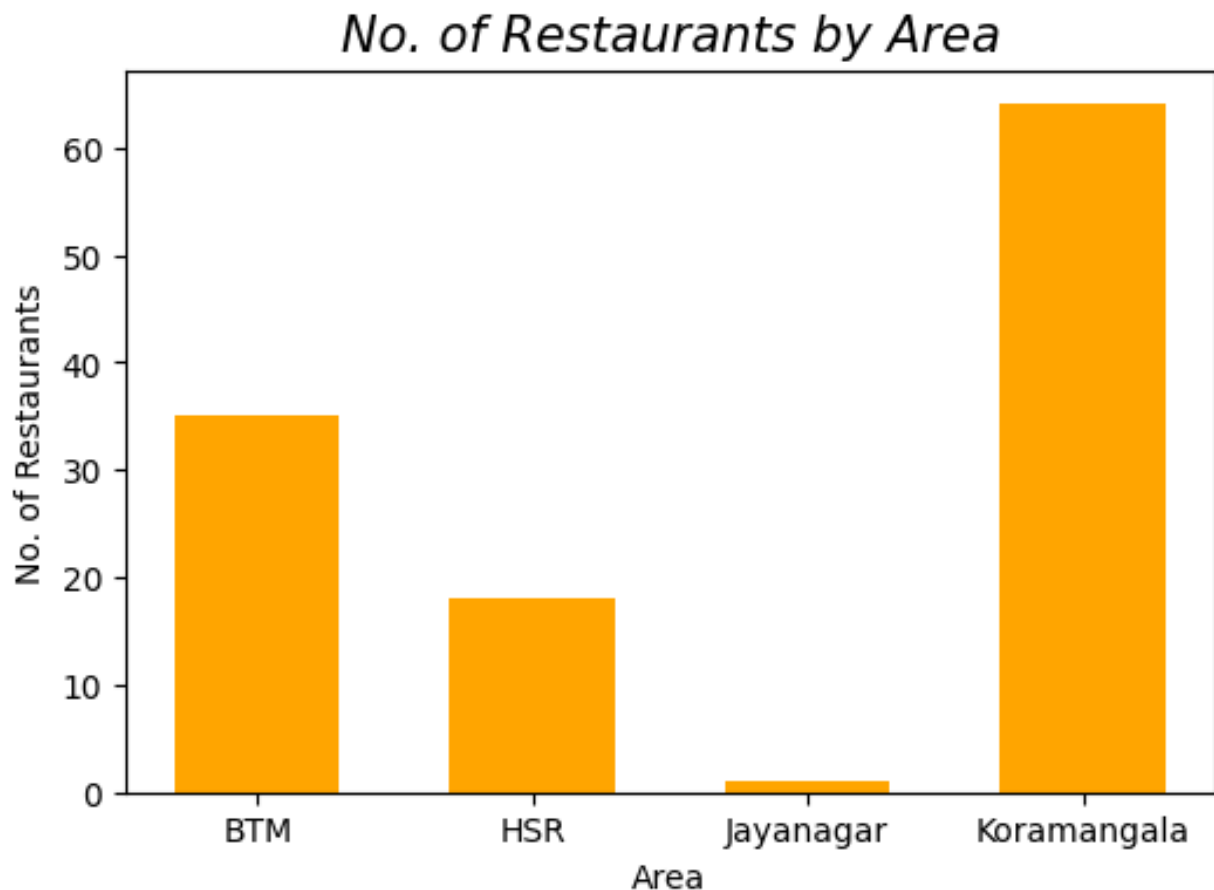
All those different analysis help out to make better business decisions and help analyse customer trends and satisfaction, which can lead to new and better products and services.

4 Technology Stack

Data Manipulation Library	Pandas
Visualization Library	Matplotlib, Seaborn
ETL	Pandas
Dataset	.CSV Format
IDE	Google Colab