**Low-Level Design (LLD)**

**ANALYZING AMAZON SALES DATA**

**GAUTAM VORA**

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| **VERSION** | **DATE** | **AUTHOR** | **COMMENTS** |
| 0.1 | 23/05/2023 | Gautam Vora | Introduction and architecture defined |
| 0.2 | 25/05/2023 | Gautam Vora | Architecture & Architecture description appended and updated. |
|  |  |  |  |
|  |  |  |  |

**Contents**

**1. Introduction………………………………………………………………………………………………………… 03**

**1.1 What is Low-Level Design Document? …………………………………………………….. 03**

**1.2 Scope ……………………………………………………………………………………………………... 03**

**2. Architecture ……………………………………………………………………………………………………….. 04**

**3. Architecture Description …………………………………………………………………………………… 04**

**3.1 Data Description ……………………………………………………………………………………… 04**

**3.2 Web Scrapping ………………………………………………………………………………………… 07**

**4. Unit test cases ………………………………………………………………………………………………………07**

**1. Introduction**

**1.1 What is Low-Level design document?**

The goal of the LDD or Low-level design document (LLDD) is to give the internal logic design of the actual program code for the House Price Prediction dashboard. LDD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

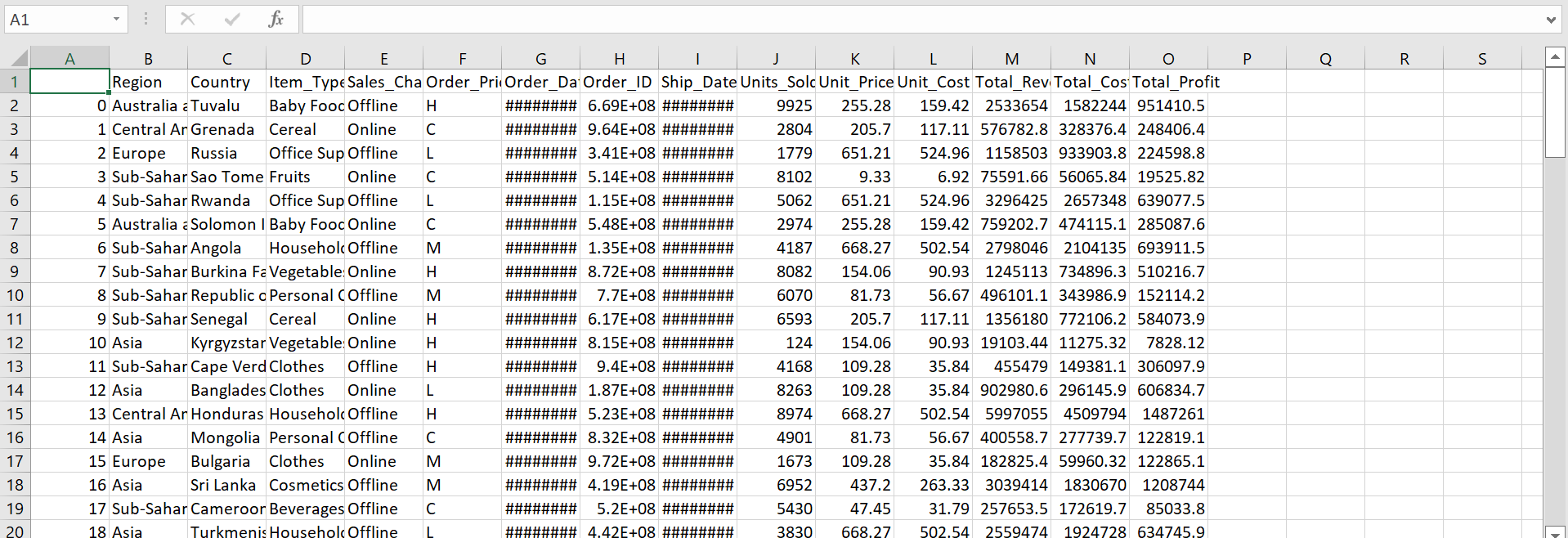
**1.2 Scope**

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The 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.

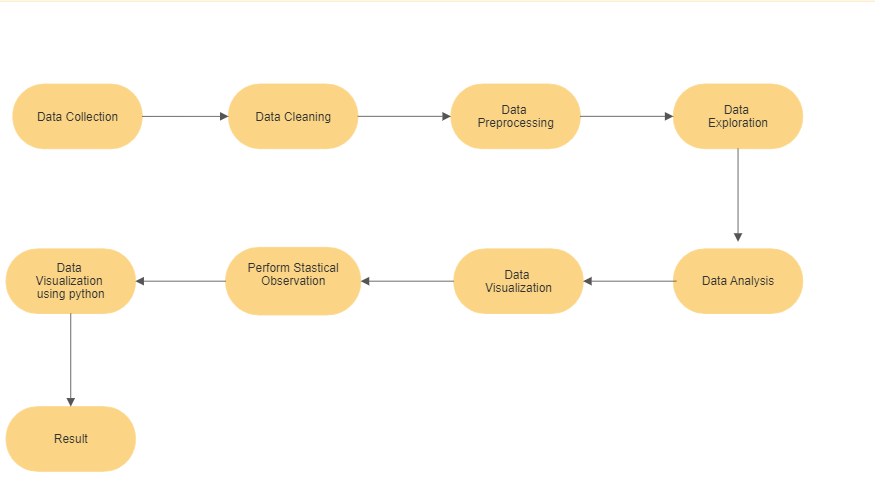
**1.Technical specifications Dataset**

The Dataset is taken from iNeuron’s provided dataset-

<https://drive.google.com/drive/folders/1FkmFVL8wlJmQWP1z52TD8PlhOJhitTyI?usp=sharing>



**2. Architecture**



**3. Architecture Description**

**3.1 Data Description**

**Data Collection -**

As we have seen earlier, in our Dataset preview, we have around hundreds of records with 14 different features. Features are distributed as Continuous features, Categorical features and Geographical feature in our Sales Dataset. The "100 Sales Records.csv" dataset was given in the form of Comma Separated Value (.csv) format and we load this dataset in Python.

**Data Cleaning -**

At this stage, based on the given knowledge unit and business problems we have formed the several Use examples to act the observations on and this will definitely help out get the key knowledge from this facts based on which business decisions will be taken. in addition, It helps in not only getting through knowledge the purposeful relationships between given properties, but it also lets us to do our own operation of making observations and come-up with our decisions in law. happily, on condition that knowledge unit has no missing values, for this reason we proceeded farther.

**Import the Dataset -**

In this process, we changed the data into highly understandable format by applying Python's Pandas and Numpy libraries.We changed categorical data into number-based dummy numbers that change/things that change in order to make it moregood for related to studying numbers/ number-based analysis which makes/gives meaning to the data up to higher extent. We drew attention to upon removing unnecessary thing in data and made it good for doing exploratory data analysis.

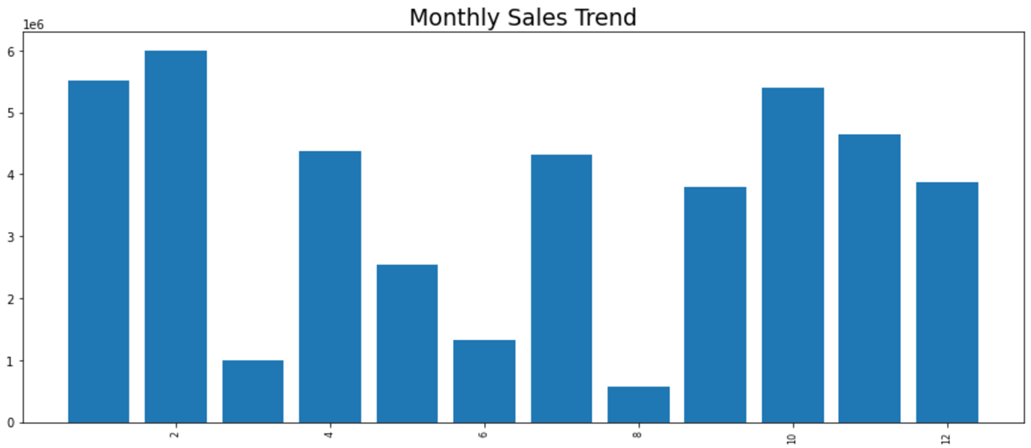
**Exploratory Data Analysis (EDA)-**

It is a "Data Exploration" step in the Data Analysis Process, where several 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 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.

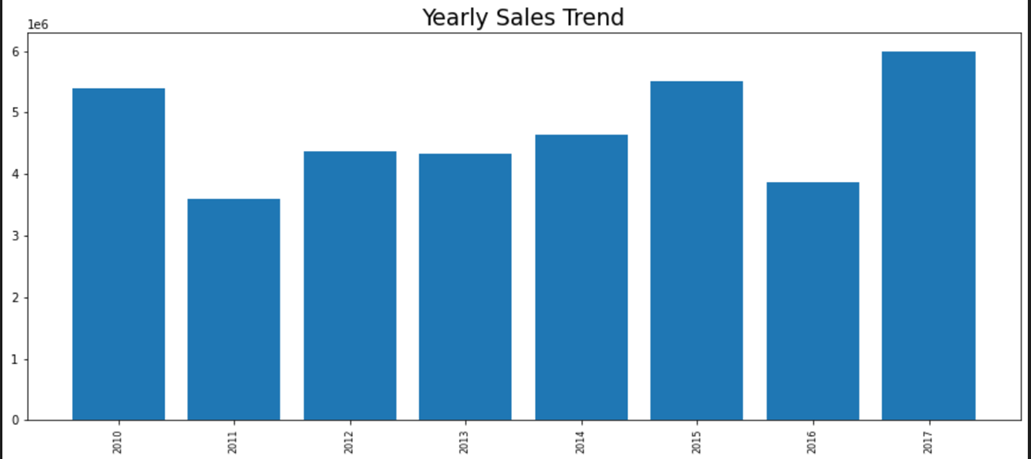
**Data Visualization Using Python-**

Initially, we extracted data from the datasets into the Python. We prepared various charts and plots based on meaningful data. We established several mathematical relations between the numerical attributes present in the data. We prepared various visuals for different sections of the data and filtered the charts according to various parameters to make it more user interactive and user friendly.. The resulting visual representation of data makes it easier to identify and share insights about the information represented in the data.

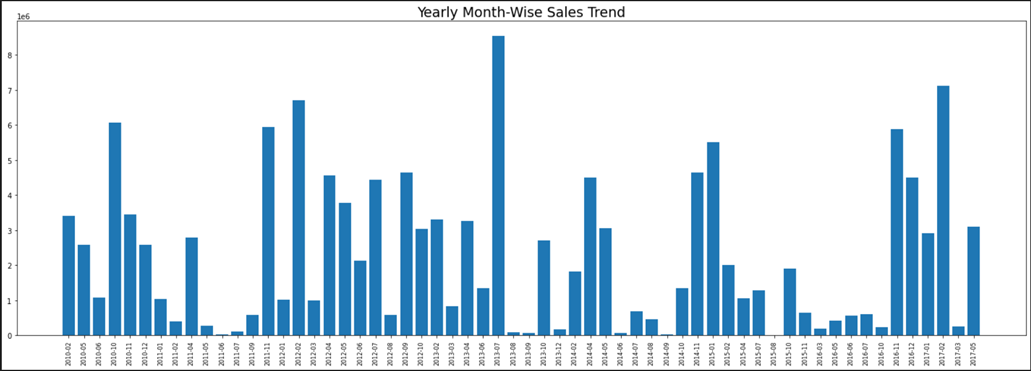
**1.Monthly Sales Trends :-**



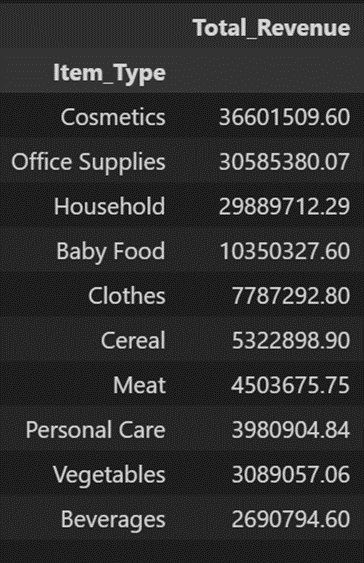
**2.Yearly Sales Trend**



**3.Yearly Month-Wise Sales Trend**



**4.Total Revenue**



**3.2. Web Scrapping**

Web scraping is a technique to automatically extract content and data from websites using bots. It is also known as web data extraction or web harvesting. Web scrapping is made simple now days, many tools are used for web scrapping. Some of python libraries used for web scrapping are Beautiful Soup, Scrapy, Selenium, etc.

|  |  |
| --- | --- |
| Data Manipulation and preprocessing | Pandas, Microsoft Excel |
| Visualization Tool | python |
| Dataset | .CSV format |
| IDE | Jupyter Notebook |

**4. Unit Test Cases**