Low Level Design (LLD)

Amazon Sales Data

Revision Number - 1.2

Last Date of Revision–04/11/2022

Vijit Kumar

Document Control

Date	Version	Description	Author
02/11/2022	1.0	Introduction, Problem Statement	Vijit Kumar
02/11/2022	1.1	Dataset Information, Architecture Description	Vijit Kumar
03/11/2022	1.2	Final Revision	Vinay Pandhe

Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Amazon sales data trend. It will explain the necessary steps which have to be followed before any analysis can begin. LLD 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. This document is intended for both the stakeholders and the developers of the system and will be proposed to the higher management for its approval.

The LLD will be focusing on the below objectives:

- Problem Understanding.
- Data Acquisition.
- Data Pre-Processing and Exploratory Analysis

- Development of models
- Auditing accuracy and retrain if require
- Finalizing the model
- Dashboard report for important activities

Scope

The LLD documentation presents the detailed structure of the Investment analytics for each of its individual components. The goal of LLD is to give the internal logical design of the actual program code. Low-level design is created based on the high-level design. The LLD documentation contains the complete description of the model used along with the comparisons of the proposed model/library compared with a baseline(existing) model against a set of metrics.

Project Introduction

Sales management has gained importance to meet increasing competition and the need for improved methods of distribution to reduce cost and to increase profits. Sales management today is the most important function in a commercial and business enterprise.

The given dataset contains sales data.

Constraints

Our analysis is done based on a limited dataset provided.The analysis is done on year vs sales trends, month vs sales trends and month year vs sales trends.

Risks

Document specific risks that have been identified or that should be considered.

Out of Scope

Delineate specific activities, capabilities, and items that are out of scope for the project.

- **Technical specifications**

Dataset

The Dataset is taken from iNeuron’s provided dataset-

In [3]:
```
pd.read_csv(r'C:\Users\hp\Desktop\AmazonSalesDataAnalysis\AmazonSalesData.csv')
```

```
C:\Users\hp\AppData\Local\Temp\ipykernel_13796\3643685868.py:1: DtypeWarning: Columns (22) have mixed types. Specify dtype opti
on on import or set low_memory=False.
  pd.read_csv(r'C:\Users\hp\Desktop\AmazonSalesDataAnalysis\AmazonSalesData.csv')
```

Out[3]:

	Unnamed: 0	CustKey	DateKey	Discount Amount	Invoice Date	Invoice Number	Item Class	Item Number	Item	Line Number	...	Sales Amount	Sales Amount Based on List Price	Sales Cost Amount	Sales Margin Amount	Sales Price
0	0	10000481	2017-04-30	-237.91	2017-04-30	100012	NaN	NaN	Urban Large Eggs	2000	...	237.91	0.00	0.00	237.91	237.910000
1	1	10002220	2017-07-14	368.79	2017-07-14	100233	P01	20910	Moms Sliced Turkey	1000	...	456.17	824.96	0.00	456.17	456.170000
2	2	10002220	2017-10-17	109.73	2017-10-17	116165	P01	38076	Cutting Edge Foot-Long Hot Dogs	1000	...	438.93	548.66	0.00	438.93	438.930000
3	3	10002489	2017-06-03	-211.75	2017-06-03	100096	NaN	NaN	Kiwi Lox	1000	...	211.75	0.00	0.00	211.75	211.750000
4	4	10004516	2017-05-27	96627.94	2017-05-27	103341	P01	60776	High Top Sweet Onion	1000	...	89248.66	185876.60	0.00	89248.66	196.150901
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
65277	65277	10017638	2018-03-21	505.78	2018-03-21	226497	P01	13447	High Top Oranges	8000	...	569.90	1075.68	239.95	329.95	63.322222
65278	65278	10017638	2018-03-21	410.75	2018-03-21	226497	P01	25906	Landslide White Sugar	38000	...	462.81	873.56	423.55	39.26	231.405000
65279	65279	10017638	2018-03-21	876.16	2018-03-21	226497	P01	61856	Moms Potato Salad	227001	...	987.20	1863.36	574.00	413.20	123.400000
65280	65280	10017638	2018-03-21	24226.77	2018-03-21	226498	P01	17801	Better Fancy Canned Sardines	1000	...	27297.51	51524.28	16188.90	11108.61	758.264167
65281	65281	10017638	2018-03-21	24479.26	2018-03-21	226498	P01	27550	Imagine Popsicles	4000	...	27582.02	52061.28	14234.22	13347.80	574.625417

65282 rows × 23 columns

The dataset consists of 65282 individual data. There are 23 columns in the dataset which are described below.

•Problem Statement

Sales management has gained importance to meet increasing competition and the need
for improved methods of distribution to reduce cost and to increase profits. Sales management today is the most important function in a commercial and business enterprise.

To understand the sales trends –
This dataset contains sales of the product and we have to find -

- Month vs sales trends
- Year vs sales trends

Month year vs sales trends
Find key metrics and factors and show the meaningful relationships between attributes.
Do your own research and come up with your findings

- **Exploratory Data Analysis (EDA)**
- **Modelling**
- **Deployment**
- **Data Cleaning**
  - **Data Pre- Processing**
  - **Raw Data Collection**
  - **Architecture**

Reporting

- **Architecture Description**
  - **Raw Data Collection-** The Dataset was taken from iNeuron provided Project Description Document.
  - https://drive.google.com/drive/folders/1FkmFVL8wlJmQWP1z52TD8PlhOJhitTyI?usp=sharing
  - **Data Pre-Processing**

Before building any model, it is crucial to perform data pre-processing to feed the correct data to the model to learn and predict. Model performance depends on the quality of data to the model to train.

This Process includes-

- Handling Null/Missing Values

- **Data Cleaning**

  Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

  - Remove duplicate or irrelevant to bservations
  - Filter unwanted outliers
  - Renaming required attributes

- **Exploratory Data Analysis (EDA)**

  Exploratory Data Analysis refers to the critical process of performing initial investigations on data to discover patterns, spot anomalies, test hypothesis

  and to check assumptions with the help of summary statistics and graphical representations.

- **Reporting**

  Reporting is a most important and underrated skill of a data analytics field. Because being a Data Analyst you should be good in easy in report because your model will be used by many stakeholders who are not from technical background.

  - High Level Design Document(HLD)
  - Low Level Design Document(LLD)
  - Architecture
  - Wireframe
  - Detailed Project Report
  - Power Point Presentation

- **Modeling**

  Data Modeling is the process of analyzing the data objects and their relationship to the other objects. It is used to analyze the data requirements that are required for the business processes. The data models are created for the data to be stored in a database. The Data Model's main focus is on what data is needed and how we have to organize data rather than what operations we have to perform.

- **Deployment**

I have created three dashboard which are month vs sales trend, year vs sales trends and month years vs sales trends respectively