Low Level Design

Amazon Sales Analysis

Written By	Freeva Dsouza
Document Version	0.2
Last Revised Date	15th February 2022

DOCUMENT CONTROL

Change Record:

VERSION	DATE	AUTHOR	COMMENTS
0.1	19th January 2022	Freeva Dsouza	Introduction and Project
0.2	25th January 2022	Freeva Dsouza	Architecture & Architecture description appended and Updated.
0.3	15th February 2022	Freeva Dsouza	Architecture & Architecture description appended and Updated.

Contents

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 Sales Analysis. 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.

2. Project

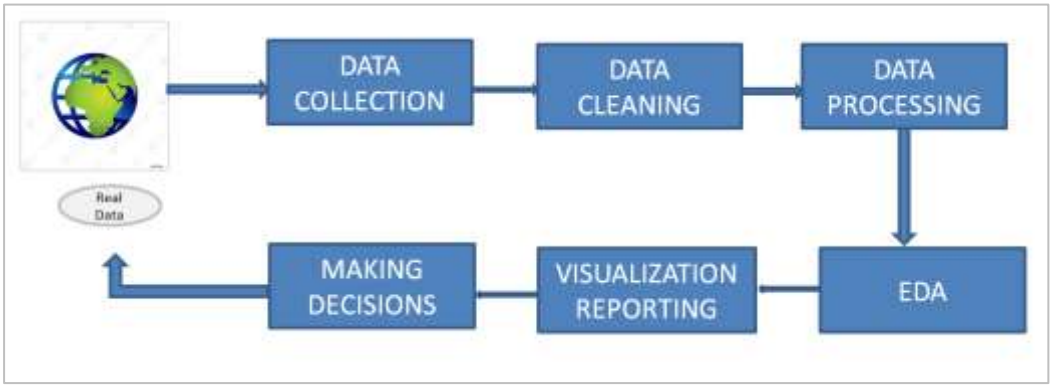
2.1 Objective of the Project

The objective of this project is to analyze the sales data and find key metrics and factors and show meaningful relationship between the attributes. The stakeholders want to know the overall sales trend, % of growth year-wise and month-wise , product analysis and Customer analysis.

2.2 Data Description

- CustKey: Unique ID of all the customers.
- Discount Amount: The amount deducted from the selling price of the items.
- Invoice Date: The date on which the goods have been billed and the transaction officially recorded.
- Invoice Number: A unique sequential code systematically assigned to the invoices.
- Item Class: Group of items that share similar properties.
- Item: Name of the product.
- Line Number: The number that uniquely identifies the item in the sales document.
- List Price: The suggested retail price of the product.
- Order Number: The number generated against each purchase.
- Promised Delivery date: The date you promised the customer you would deliver the products.
- Sales Amount: Total sales generated -> Sales Price*Sales quantity
- Sales cost Amount: Total cost price for the quantity of items sold.
- Sales Margin Amount: The amount a company makes from a sale of a service or product.
- Sales Price: Selling price of each product.
- Sales quantity: Number of each products sold
- Sales Rep: Unique ID generated for each Sales Representative.
- U/M: Units of measure of items sold.

3. Architecture



3.1 Architecture Description

1. **Raw Data Collection:** The dataset was provided by the company in the form of CSV file.

2. **Data cleaning:** Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. The data is cleaned using Pythons Pandas library.

   The process included:
   1. Removing unwanted observations
   2. Fixing structural errors
   3. Managing unwanted outliers
   4. Handling missing data

3. **Data Processing:** In this process the data is converted from a given form to a much more usable and desired form i.e. making it more meaningful and informative. It is emphasized upon removing redundancy in data and made it suitable for performing exploratory data analysis.

4. **EDA:** In this process various insights and observations from the data were found. Also, KPIs (Key performance metrics) i.e. the factors which are mainly responsible for facilitating efficient sales management were defined. A few outliers were detected within the distribution of various attributes and properly examined them at each step. Various charts and plots have been prepared based on meaningful data. Using Pythons Matplotlib, Seaborn, Plotly libraries several highly interactive bar charts, trends have been created.

5. **Visualization/Reporting:**
   Firstly, imported the datasets into Tableau public and organized the data. Prepared worksheets on various sections of data and established some meaningful mathematical relations between the numerical parameters available in the data. Used various plots to visualize data deeply and intensively and derived various observations based on visual analysis. Prepared a dashboard altogether to organize all the worksheets in a systematic and synchronized manner and then published it on Tableau Public.