Low Level Design

Analyzing Amazon Sales Data

Revision Number: 1.2

Last date of revision: 10/08/2023

Harshit Kumar Pathak

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| 08/08/2023 | 1.0 | Introduction, Problem Statement, Dataset Information | Harshit Kumar Pathak |
| 09/08/2023 | 1.1 | Architecture | Harshit Kumar Pathak |
| 10/08/2023 | 1.2 | Final Revision | Harshit Kumar Pathak |
|  |  |  |  |

**Contents**

[Document Version Control 2](#_bookmark0)

1 [Introduction](#_bookmark2) 4

1.1 [What is Low-Level Design Document?](#_bookmark3) 4

* 1. [Scope](#_bookmark4) 4

1. Problem Statement ………………………………………………………… 4
2. Dataset Information………………………………………………………… 5
3. Architecture …………………………………………………………………. 6

4.1 Architecture Description ……………………………………………… 7

1. **Introduction**
   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. **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. 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.

Do ETL: Extract-Transform-Load some Amazon dataset and find for me Sales-trend -> month-wise, year-wise, yearly-month wise. Find key metrics and factors and show the meaningful relationships between attributes.

**3. Dataset Information:**

The dataset contains Sales data of Amazon for the year 2010 to 2017

***Amazon Sales Data Contains*:**

1. **Region** - Continent or Region of the Sales Country

2. **Country** - Name of the Country of Sales happened

3. **Item Type** - It is The Category of the product

4. **Sales Channel** - Mode of the Sales (Online or Offline)

5. **Order Priority** - It is Priority of the product *(H > C > L > M)*

6. **Order Date** - Date of order registered

7. **Order** - Unique id number of order to track the order

8. **Ship Date** - Date of the ordered product got shipped

9. **Unit Sold** - Amount of quantity sold for that order

10. **Unit Price** - Amount sold to costumer per Unit

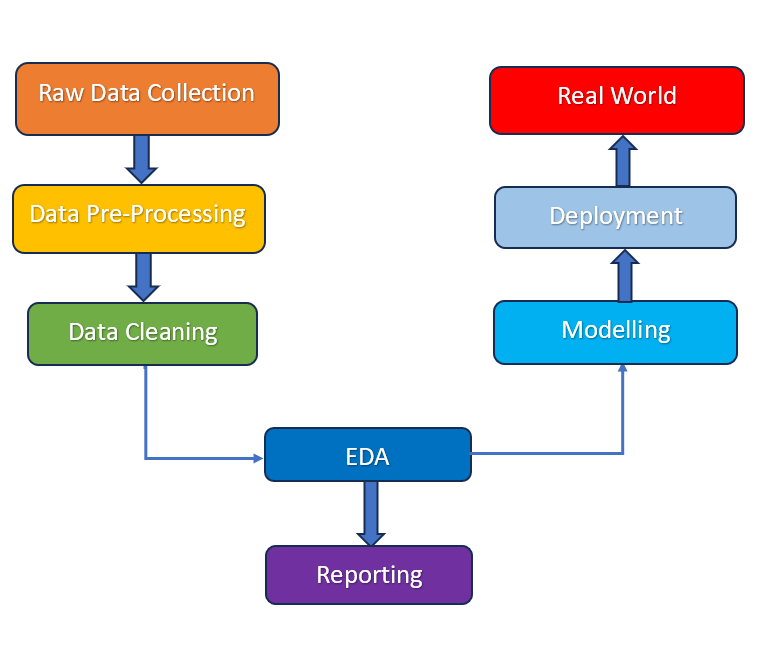
11. **Unit Cost** - Original price of the Unit

12. **Total Revenue** - Total revenue of the order

13. **Total Cost** - Original total cost of sold units

14. **Total Profit** - Profit gained in that order (Total revenue – Total Cost)

**4. Architecture:**



4.1 Architecture Description:

1. **Raw Data Collection –**

The dataset was taken from Project Description Document.

Data Set Link - [Amazon Sales Records (2).csv - Google Drive](https://drive.google.com/file/d/1tvNcSh1Ayfkv7NIE2oKqIMOLlfedNJvM/view)

**2. 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 fed to the model to train.

This Process includes-

1. Handling Null/Missing Values
2. Handling Skewed Data
3. Outliers Detection and Removal

**3. Data Cleaning –**

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

1. Remove duplicate or irrelevant observations
2. Filter unwanted outliers
3. Renaming required attributes
4. Creating Month, Year features from Order Date feature

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

**5. Reporting –**

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

1. High Level Design Document (HLD)
2. Low Level Design Document (LLD)
3. Architecture
4. Wireframe
5. Detailed Project Report

**6. Modelling –**

Data Modelling 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 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.

**7. Deployment–**

At the end, I created a Dashboard in PowerBI

