2023

LOW LEVEL DESIGN DOCUMENT

AMAZON SALES DATA ANALYSIS

ISHAN CHAURASIA

INEURON

Document Version Control:

Data Issued	Version	Description	Author
27-07-2023	1.0	Final LLD	Ishan Chaurasia

Contents:

1 Introduction:

1.1 What is Low-Level Design Document?

The purpose of Low-Level Design Document is to provide the detailed logic design of the program code for the Sales Analysis Dashboard. LLDD shoes the class diagrams with the methods and relationships among classes and program specifications. It explains the modules so that the programmer can easily code the program from the document.

1.2 Scope:

The Low-Level Design is a process that breaks down the components of a system into smaller and more detailed parts. It can help to design the data structures, software architecture, source code, performance algorithm that are needed for the system. The data organization can be initially defined during the requirement analysis and then improved during the data design work.

1.3 Project Introduction:

This project aims to understand the sales management of e-commerce businesses. It focuses on how Amazon and other service-oriented companies use information systems to collect and analyze customer data and preferences. The project also examines how these companies gain a competitive edge by managing and acquiring information effectively. The project uses a sales dataset that contains various attributed such as sales amount, list price, cost price, etc., to perform data analysis and derive insights.

2 Problem Statement:

To cope with the growing competition and the demand for better distribution methods that can lower costs and boost profits, sales management has become a vital function in any commercial and business enterprise. I can help you with some data analysis on Amazon Sales Data Analysis using ETL. I can extract some Amazon datasets and transform them to find the sales trends by months, years and yearly month, I can also load the results and show you the key metrics and factors that influence the sales and how they are related to each other.

3 Dataset Information:

Invoice Date: The day on which the Invoice was generated.

Discount Amount: Total discount provided on any item.

Sales Amount: Total Sales Price of an Item.

Sales Margin Amount (Profits): Sales Margin Amount is the difference between Sales Cost

Amount & Sales Amount.

Sales Cost Amount: Total Cost Price of any Item.

Sales Price: Sales Price of any particular Item.

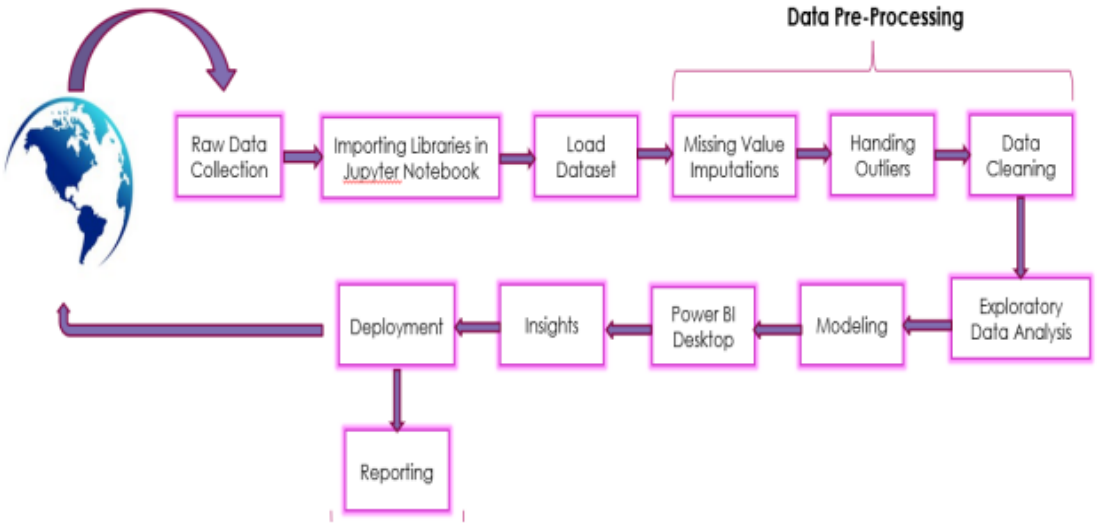
List Price: Basic Price of an Item as published on the price list.

Sales Rep: A person whose job is to sell products or services for a company.

U/M: Unit of Measure

CustKey: It is a Unique Number on the Invoice that is used to reference customers' accounts.

4 Architecture:



4.1 Architecture Design

1. Raw Data Collection

The dataset was taken from iNeuron's internship portal - project description document.

https://docs.google.com/spreadsheets/d/1h3EsPf-fTLrzpP7sGeyuRnGBXrdJRcXY/edit?usp=sharing&ouid=105519103382792804653&rtpof=true&sd=true

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:

a) Handling Null/Missing Values

b) Handling Skewed Data

c) 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.

a) Remove duplicate and irrelevant observations

b) Filter unwanted outliers

c) Renaming required attributes

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 in data analytics field. Because being a data analyst one should be well versed with ease and self-explanatory reports because your model will be used by several stakeholders who are not from a technical background.

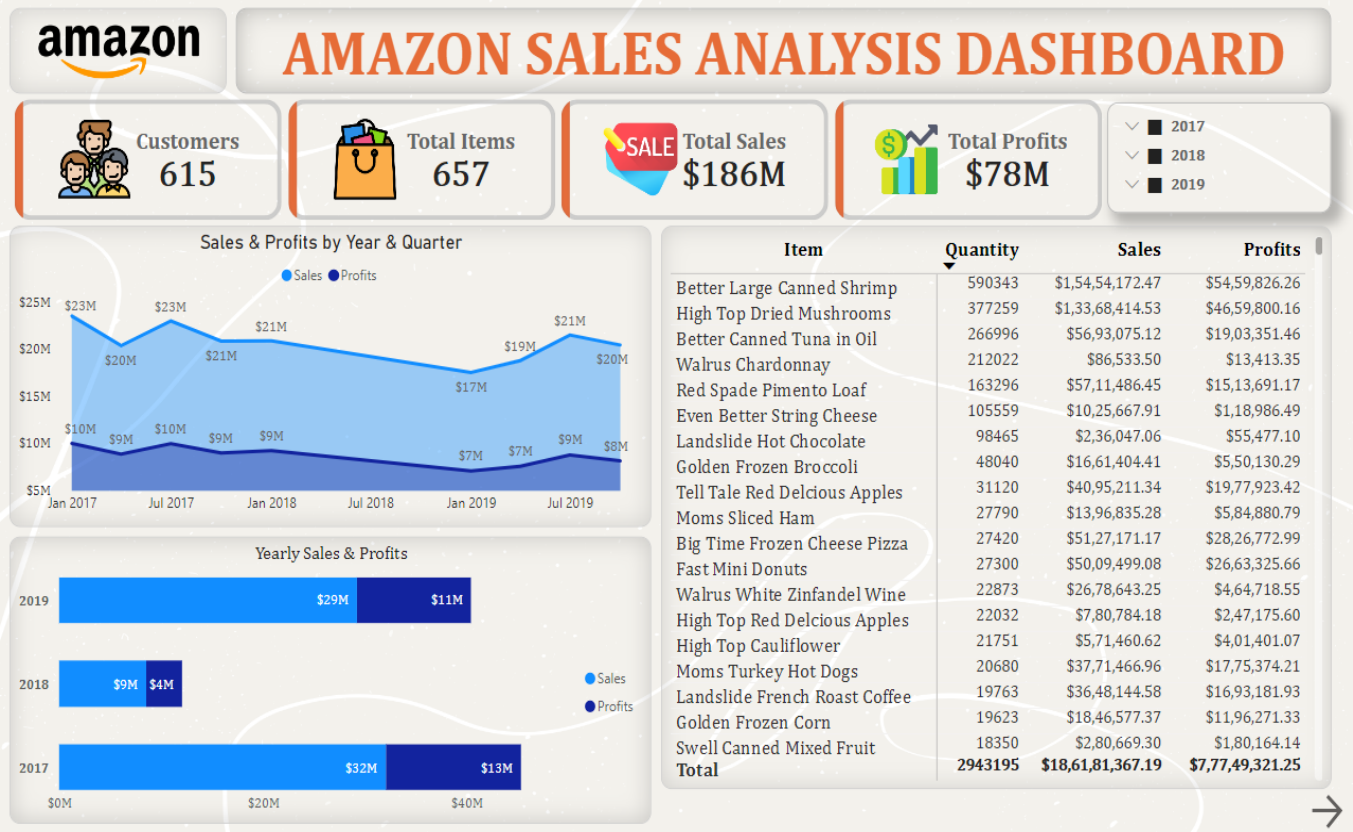
a) High Level Design (HLD) Document
b) Low Level Design (LLD) Document
c) Architecture
d) Wireframe
e) Detailed Project Report
f) PowerPoint Presentation

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 business processes. The data models are created to store the data 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.

5. Deployment



Item	Quantity	Sales	Profits
Better Large Canned Shrimp	590343	$1,54,54,172.47	$54,59,826.26
High Top Dried Mushrooms	377259	$1,33,68,414.53	$46,59,800.16
Better Canned Tuna in Oil	266996	$56,93,075.12	$19,03,351.46
Walrus Chardonnay	212022	$86,533.50	$13,413.35
Red Spade Pimento Loaf	163296	$57,11,486.45	$15,13,691.17
Even Better String Cheese	105559	$10,25,667.91	$1,18,986.49
Landslide Hot Chocolate	98465	$2,36,047.06	$55,477.10
Golden Frozen Broccoli	48040	$16,61,404.41	$5,50,130.29
Tell Tale Red Delcious Apples	31120	$40,95,211.34	$19,77,923.42
Moms Sliced Ham	27790	$13,96,835.28	$5,84,880.79
Big Time Frozen Cheese Pizza	27420	$51,27,171.17	$28,26,772.99
Fast Mini Donuts	27300	$50,09,499.08	$26,63,325.66
Walrus White Zinfandel Wine	22873	$26,78,643.25	$4,64,718.55
High Top Red Delcious Apples	22032	$7,80,784.18	$2,47,175.60
High Top Cauliflower	21751	$5,71,460.62	$4,01,401.07
Moms Turkey Hot Dogs	20680	$37,71,466.96	$17,75,374.21
Landslide French Roast Coffee	19763	$36,48,144.58	$16,93,181.93
Golden Frozen Corn	19623	$18,46,577.37	$11,96,271.33
Swell Canned Mixed Fruit	18350	$2,80,669.30	$1,80,164.14
Total	**2943195**	**$18,61,81,367.19**	**$7,77,49,321.25**