

Low Level Design (LLD)

Analyzing Amazon Sales Data

Written By / Author	Lokesh Attarde
Document Version	LLD-V1.0
Last Revised Date	26/08/2022

Document Version Control:

Date Issued	Version	Description	Author
26/08/2022	LLD-V1.0	First Version of Complete LLD	Lokesh Attarde

Reviews:

VERSION	DATE	REVIEWER	COMMENTS
V1.0	26/08/2022	Lokesh Attarde	Document, Version Control and Architectural description added.

Approval Status:

VERSION	REVIEW DATE	REVIEWED BY	APPROVED BY	COMMENTS
V1.0	26/08/2022	Lokesh Attarde	Lokesh Attarde	

Contents

Document Version Control	2
Abstract	4
1 Introduction	5
1.1 Why this Low-Level design document?	5
1.2 Scope	5
2 Technical Specifications	6
2.1 Amazon Sales Dataset	6
2.2 Amazon Sales Dataset Overview	6
3 Architecture	7
3.1 Architecture Description	7
3.1.1 Data Collection	7
3.1.2 Data Cleaning	7
3.1.3 Import the Dataset	7
3.1.4 Exploratory Data Analysis (EDA)	8
3.1.5 Data Visualization using Power BI	9
4 Technology Stack	10

Abstract

Amazon Sales data refers to sales, high performing sellers and several other data points. There are millions of Amazon sellers around the world. Nearly half of them are self-employed and live off their ecommerce/retail businesses (47%), and 22% earn income from their Amazon businesses alone. Amazon sales data Analysis focuses on the process of analysing consumer behaviour, sales, and several other attributes in order to make improved, data-driven decisions. It is key to successfully sustaining their businesses and earning profits and for this purpose, they analyse different metrics like sales, Sales Quantity, Discount rate, Sales over years etc. By analysing different metrics, you will be able to increase and improve your performance in terms of sales, Items to be sold and discount rates etc. Analysis of the sales data the main factor that contributes to sellers improving their business and increasing their revenue. They can better understand the market trends and customers' buying behaviours and help them cater to what the customers really want.

In the world of rising new technology and innovation, E-commerce industry is advancing with the role of Data Science and Analytics. Data analysis can help them to understand their business in a quiet different manner and helps to improve the quality of the service by identifying the weak areas of the business. This study demonstrates the how different analysis help to make better business decisions and help analyze customer trends and satisfaction, which can lead to new and better products and services. Different analysis performed to get the key insights from this data based on which business decisions will be taken.

1 Introduction

1.1 Why this Low-Level design document?

The purpose of this LLD or a Low-Level Design (LLD) document is to give the internal logical design of the actual program code for Amazon Sales Data Analysis project. LLD describes the class diagrams with the methods and relations between classes and program 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 this project and will be proposed to the higher management for its approval.

The main objective of the project is to analyse the various aspects with different use cases which covers many aspects of amazon sales. It helps in not only understanding the meaningful relationships between attributes but it also allows us to do our own research and come-up with our findings.

1.2 Scope

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

This study demonstrates the how different analysis help out to make better business decisions and help analyse customer trends and satisfaction, which can lead to new and better products and services.

2 Technical Specifications

2.1 Amazon Sales Dataset –

Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit	Day Name	Month	Year	Quarter
Australia and Oceania	Tuvalu	Baby Food	Offline	H	28-05-2010	6.7E+08	27-06-2010	9925	255.28	159.42	2533654	1582243.5	951410.5	Friday	5	2010	2
Central America and the Caribbean	Grenada	Cereal	Online	C	22-08-2012	9.6E+08	15-09-2012	2804	205.7	117.11	576782.8	328376.44	248406.36	Wednesday	8	2012	3
Europe	Russia	Office Supplies	Offline	L	02-05-2014	3.4E+08	08-05-2014	1779	651.21	524.96	1158502.59	933903.84	224598.75	Friday	5	2014	2
Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	20-06-2014	5.1E+08	05-07-2014	8102	9.33	6.92	75591.66	56065.84	19525.82	Friday	6	2014	2
Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	01-02-2013	1.2E+08	06-02-2013	5062	651.21	524.96	3296425.02	2657347.52	639077.5	Friday	2	2013	1
Australia and Oceania	Solomon Islands	Baby Food	Online	C	04-02-2015	5.5E+08	21-02-2015	2974	255.28	159.42	759202.72	474115.08	285087.64	Wednesday	2	2015	1
Sub-Saharan Africa	Angola	Household	Offline	M	23-04-2011	1.4E+08	27-04-2011	4187	668.27	502.54	2798046.49	2104134.98	693911.51	Saturday	4	2011	2
Sub-Saharan Africa	Burkina Faso	Vegetables	Online	H	17-07-2012	8.7E+08	27-07-2012	8082	154.06	90.93	1245112.92	734896.26	510216.66	Tuesday	7	2012	3
Sub-Saharan Africa	Republic of the Congo	Personal Care	Offline	M	14-07-2015	7.7E+08	25-08-2015	6070	81.73	56.67	496101.1	343986.9	152114.2	Tuesday	7	2015	3
Sub-Saharan Africa	Senegal	Cereal	Online	H	18-04-2014	6.2E+08	30-05-2014	6593	205.7	117.11	1356180.1	772106.23	584073.87	Friday	4	2014	2
Asia	Kyrgyzstan	Vegetables	Online	H	24-06-2011	8.1E+08	10-07-2011	124	154.06	90.93	19103.44	11275.32	7828.12	Friday	6	2011	2
Sub-Saharan Africa	Cape Verde	Clothes	Offline	H	02-08-2014	9.4E+08	19-08-2014	4168	109.28	35.84	455479.04	149381.12	306097.92	Saturday	8	2014	3
Asia	Bangladesh	Clothes	Online	L	13-01-2017	1.9E+08	01-03-2017	8263	109.28	35.84	902980.64	296145.92	606834.72	Friday	1	2017	1
Central America and the Caribbean	Honduras	Household	Offline	H	08-02-2017	5.2E+08	13-02-2017	8974	668.27	502.54	5997054.98	4509793.96	1487261.02	Wednesday	2	2017	1
Asia	Mongolia	Personal Care	Offline	C	19-02-2014	8.3E+08	23-02-2014	4901	81.73	56.67	400558.73	277739.67	122819.06	Wednesday	2	2014	1
Europe	Bulgaria	Clothes	Online	M	23-04-2012	9.7E+08	03-06-2012	1673	109.28	35.84	182825.44	59960.32	122865.12	Monday	4	2012	2
Asia	Sri Lanka	Cosmetics	Offline	M	19-11-2016	4.2E+08	18-12-2016	6952	437.2	263.33	3039414.4	1830670.16	1208744.24	Saturday	11	2016	4
Sub-Saharan Africa	Cameroon	Beverages	Offline	C	01-04-2015	5.2E+08	18-04-2015	5430	47.45	31.79	257653.5	172619.7	85033.8	Wednesday	4	2015	2
Asia	Turkmenistan	Household	Offline	L	30-12-2010	4.4E+08	20-01-2011	3830	668.27	502.54	2559474.1	1924728.2	634745.9	Thursday	12	2010	4
Australia and Oceania	East Timor	Meat	Online	L	31-07-2012	3.2E+08	11-09-2012	5908	421.89	364.69	2492526.12	2154588.52	337937.6	Tuesday	7	2012	3
Europe	Norway	Baby Food	Online	L	14-05-2014	8.2E+08	28-06-2014	7450	255.28	159.42	1901836	1187679	714157	Wednesday	5	2014	2
Europe	Portugal	Baby Food	Online	H	31-07-2015	8.6E+08	03-09-2015	1273	255.28	159.42	324971.44	202941.66	122029.78	Friday	7	2015	3
Central America and the Caribbean	Honduras	Snacks	Online	L	30-06-2016	8E+08	26-07-2016	2225	152.58	97.44	339490.5	216804	122686.5	Thursday	6	2016	2
Australia and Oceania	New Zealand	Fruits	Online	H	08-09-2014	1.4E+08	04-10-2014	2187	9.33	6.92	20404.71	15134.04	5270.67	Monday	9	2014	3
Europe	Moldova	Personal Care	Online	L	07-05-2016	7.4E+08	10-05-2016	5070	81.73	56.67	414371.1	287316.9	127054.2	Saturday	5	2016	2
Europe	France	Cosmetics	Online	H	22-05-2017	9E+08	05-06-2017	1815	437.2	263.33	793518	477943.95	315574.05	Monday	5	2017	2
Australia and Oceania	Kiribati	Fruits	Online	M	13-10-2014	3.5E+08	10-11-2014	5398	9.33	6.92	50363.34	37354.16	13009.18	Monday	10	2014	4
Sub-Saharan Africa	Mali	Fruits	Online	L	07-05-2010	6.9E+08	10-05-2010	5822	9.33	6.92	54319.26	40288.24	14031.02	Friday	5	2010	2

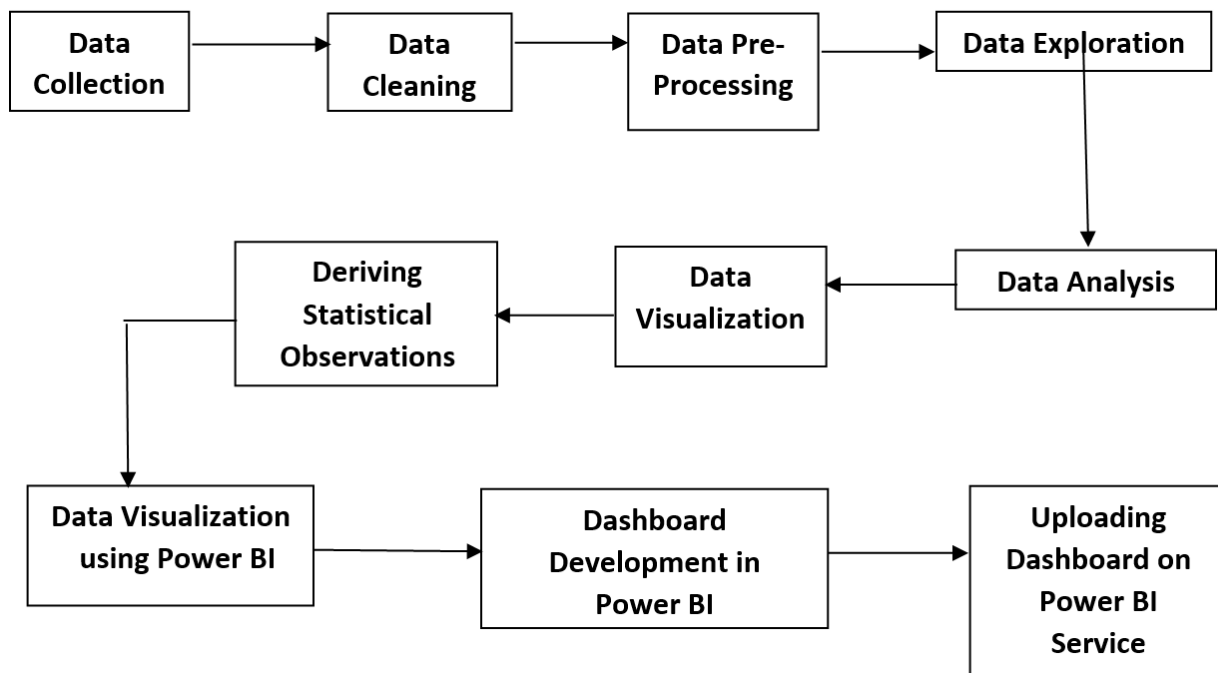
2.2 Amazon Sales Dataset Overview –

The Listings dataset consists of a table with 100 records and 18 features. Features are distributed as 8 Continuous features, 9 Categorical features and 1 Geographical feature. There are no Missing values.

Dataset Statistics		Dataset Insights	
Number of Variables	18	Unit Cost is skewed	Skewed
Number of Rows	100	Total Revenue is skewed	Skewed
Missing Cells	0	Total Cost is skewed	Skewed
Missing Cells (%)	0.0%	Total Profit is skewed	Skewed
Duplicate Rows	0	Country has a high cardinality: 76 distinct values	High Cardinality
Duplicate Rows (%)	0.0%	Order Date has a high cardinality: 100 distinct values	High Cardinality
Total Size in Memory	59.2 KB	Ship Date has a high cardinality: 99 distinct values	High Cardinality
Average Row Size in Memory	606.1 B	Order Priority has constant length 1	Constant Length
Variable Types	Categorical: 9 GeoGraphy: 1 Numerical: 8	Order Date has constant length 10	Constant Length
		Ship Date has constant length 10	Constant Length

Variables	
Sort by	Feature order
	<input type="checkbox"/> Reverse order

3 Architecture



3.1 Architecture Description –

3.1.1 Data Collection –

As we have seen earlier, in our Dataset preview, we have around hundreds of records with 18 different features. Features are distributed as 8 Continuous features, 9 Categorical features and 1 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 Power BI.

3.1.2 Data Cleaning –

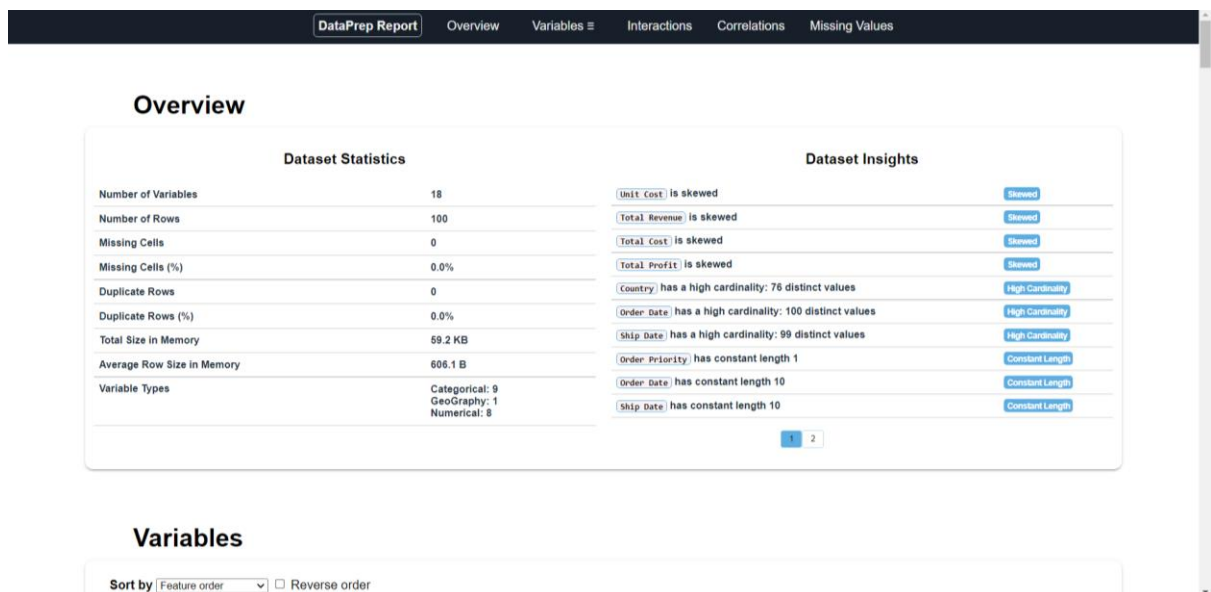
At this stage, based on the given dataset and business problems we have defined the several Use Cases to perform the analysis on and this will definitely help out get the key insights from this data based on which business decisions will be taken. Furthermore, It helps in not only understanding the meaningful relationships between attributes, but it also allows us to do our own research and come-up with our findings. Fortunately, provided dataset has no missing values, hence we proceeded further.

3.1.3 Import the Dataset –

In this process, we transformed the data into highly understandable format by applying Python’s Pandas and Numpy libraries. We eliminated the unnecessary rows and columns. We transformed categorical data into numerical dummy variables in order to make it more suitable for statistical numerical analysis which renders meaning to the data up to higher extent. We emphasized upon removing redundancy in data and made it suitable for performing exploratory data analysis.

3.1.4 Exploratory Data Analysis (EDA) –

- "Exploratory Data Analysis" (EDA) 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 give 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.



3.1.5 Data Visualization using Power BI -

Initially, we extracted data from the datasets into the Power BI Desktop. 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. We then prepared dashboards by comprising all the prepared visuals altogether in order to make the visualizations more systematic and synchronized manner. Eventually, published it on Web using Power BI Service and generated a shareable link for our dashboard.

The resulting visual representation of data makes it easier to identify and share insights about the information represented in the data.

Here is the beautiful glimpse of our Dashboard –



All those different analyses help out to make better business decisions and help analyse customer trends and satisfaction, which can lead to new and better products and services.

4 Technology Stack:

Data Manipulation & Pre-processing	Pandas, Microsoft Excel and Power Query
Visualization Tool	Microsoft Power BI Desktop
EDA	dataprep
Dataset	.CSV Format
IDE	Jupyter Notebook