Amazon Sales Data Analysis

LOW LEVEL DESIGN(LLD)

SIMRAN GODSEY

VERSION:1.0
DATE 31/01/2023

Document Version Control

Date Issued	Version	Description	Author
31-01-2023	1.0	First Version on complete LLD	Simran Godsey

Project Introduction

This is a project about Amazon Sales Data Analysis. The report critically evaluates how service-based organizations -Amazon use Management information systems to attain competitive advantage through efficient management and acquisition of information. The purpose of this project is to analyse Amazon Sales Data to obtain meaningful information. To do that, a Sales dataset is provided, which includes sales amount, list price, cost price, etc.

Table of Contents

Abstract

Amazon Sales data refers to sales, high performing sellers and several other data points. There are millions of Amazon sellers around the world. 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 Total Sales, Sales Quantity, Total Profit, Sales, Last Year Sales and other metrics.

By analysing these different metrics, we will be able to increase and improve our performance. It can also help us to better understand the market trends and customers' buying behaviours and help us to know what the customers really want.

In the world of rising new technology and innovation, E-commerce industry is advancing with the role of Data 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 analyse 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.

This dataset provides a huge amount of information about the Profit, Revenue, Cost, Unit Sold and other information Across Various Region and Country. Based on the Information the ultimate goal is to showcase the Sales trend month wise, year wise and Quarter wise and find important insights highlighting key indicators and metrics that influence customer choice

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.

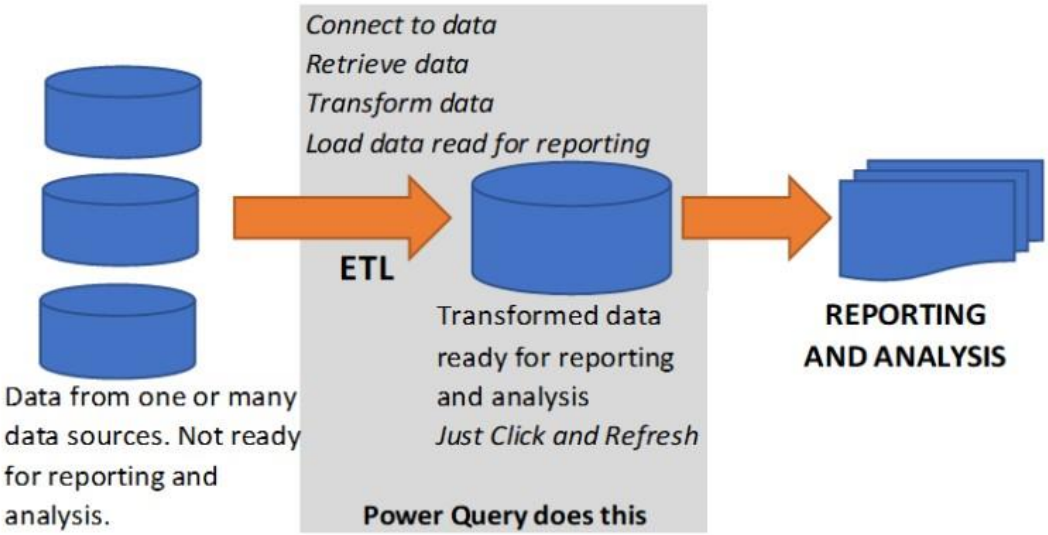
The main objective of the project is to analyse the various aspects with different use cases which covers many aspects of Amazon sales. It helps in not only understanding the meaningful relationships between attributes but also allows us to do our own research and come up with our findings.

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 how different analysis helps to make better business decisions and help analyse customer trends and satisfaction, which can lead to new and better products and services.

2.Architecture



ETL (extract, transform and load) in Power BI use the preparation of data sets for analysis by removing irregularities in the data. It also involves data visualization to draw meaningful patterns and insights.

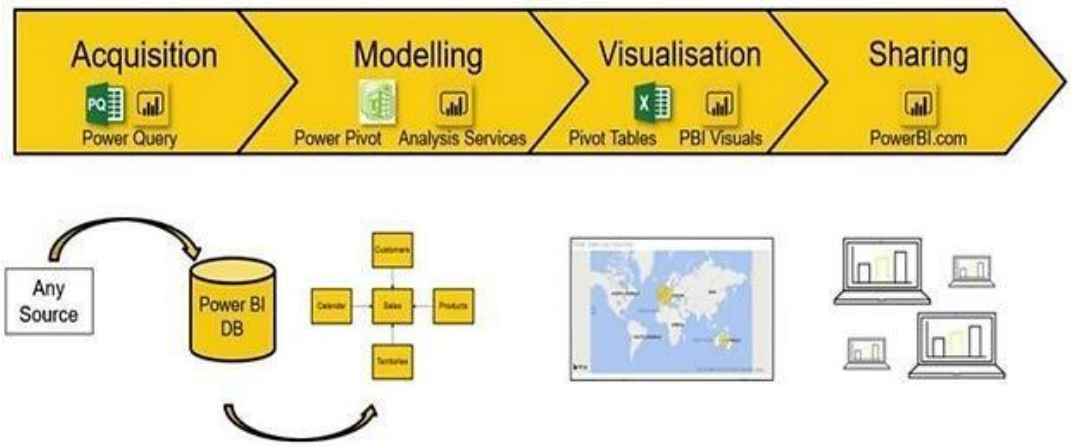
Based on the results of ETL, companies also make business decisions, which can have repercussions later.

If ETL is not done properly then it can damage the business a lot in many ways such as loss of clients which whom we are working, the decision making will go completely wrong, and many more issues.

If done well, it may improve the efficacy of everything we do next.

Below are the following steps to follow for ETL:

1. Data Sourcing
2. Data Cleaning
3. Data Modelling
4. Data Visualization



3. Architecture Description

3.1 Data Sourcing

The dataset is in CSV (Comma Separated Values) format. MS Excel is to load the data.

This dataset is publicly available for research purposes.

Title: Amazon Sales Data Analysis.csv

Source: Git Hub

3.2 Data Overview

- The dataset is of size 12.4 KB
- It includes a single file in ".csv" format.
- Number of rows/records: 100
- Number of attributes: 14

3.3 Data Description

The following attributes describes the dataset.

- Region
  - ► Description: Name of the Region
  - ► Datatype: string
- Country

- ► Description: Name of different Countries
- ► Datatype: string

- Item Type
  - ► Description: Different Product type sales in Amazon
  - ► Datatype: string
- Sales Channel
  - ► Description: Mode of shopping Online or Offline.
  - ► Datatype: string
- Order Priority
  - ► Description: Priority of Sales Range between low to high.
  - ► Datatype: string
- Order Date
  - ► Description: Date of the Order.
  - ► Datatype: Date
- Order ID
  - ► Description: Order Id of the Varieties of Product.
  - ► Datatype: string
- Ship Date
  - ► Description: Date when the product is dispatched.
  - ► Datatype: date
- Unit Solds
  - ► Description: Number of Unit sold per product.
  - ► Datatype: Int

- Unit Price
  - ► Description: Unit Price of the Product.
  - ► Datatype: Int

- Unit Cost
  - ► Description: Unit cost of the Product.
  - ► Datatype: string

- Total Revenue
  - ► Description: Amount Incurred after selling different Products.
  - ► Datatype: Decimal

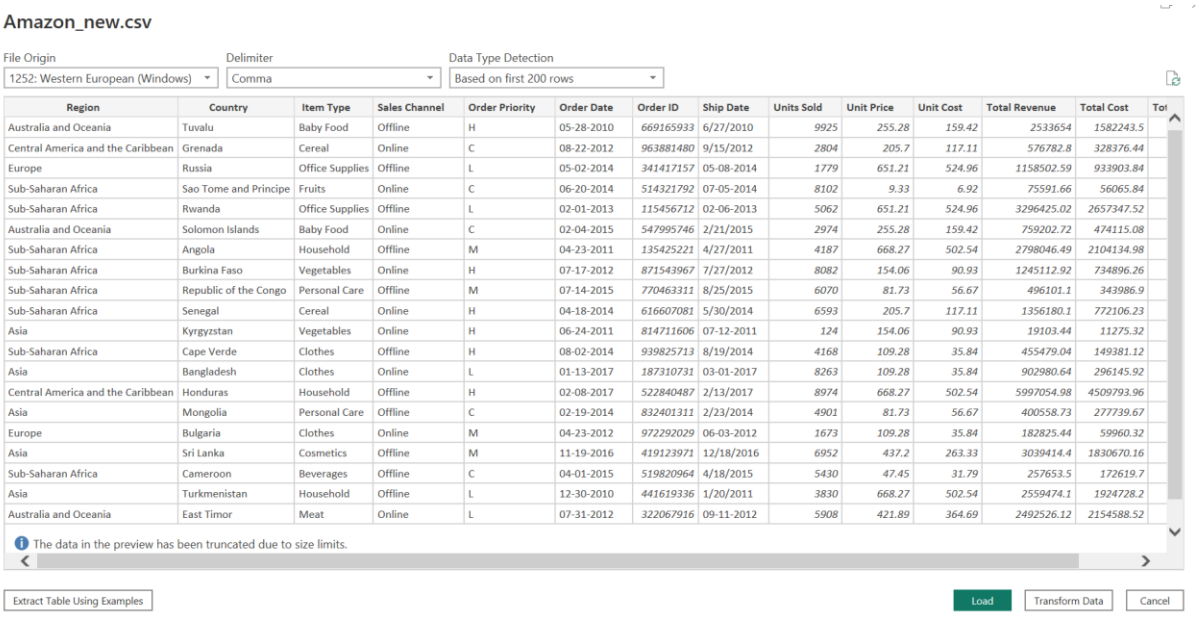
- Total Cost
  - ► Description: Total Cost Incurred by the company for making Products
  - ► Datatype: Decimal

- Total Profit
  - ► Description: Profit Earned by the Company after subtracting all the expenses from the revenue.
  - ► Datatype: Decimal

3.4 Data Loading in Power BI Query Editor

Power Query is the data connectivity and data preparation technology that enables end users to seamlessly import and reshape data from within a wide range of Microsoft products, including Excel, Power BI, Analysis Services, data verse, and more with the following characteristics.

- There can be multiple rows and columns in the data.
- Each row represents a sample of data,
- Each column contains a different variable that describes the samples (rows).
- The data in every column can be a different type of data like numbers, strings, dates, Boolean etc.



Amazon_new.csv

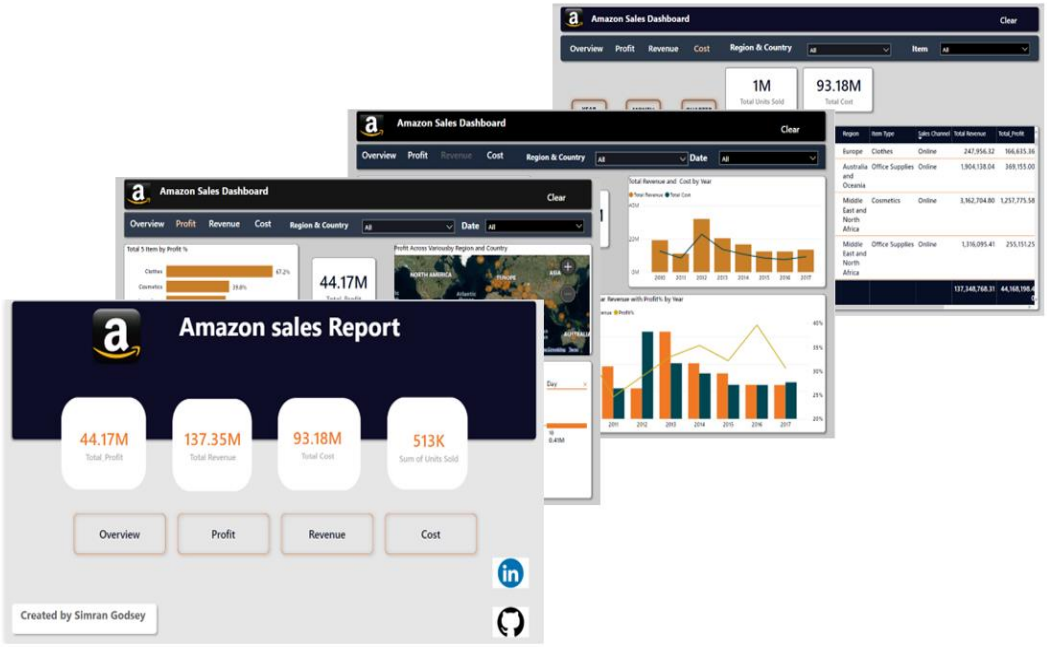
File Origin: 1252: Western European (Windows) | Delimiter: Comma | Data Type Detection: Based on first 200 rows

Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Tot
Australia and Oceania	Tuvalu	Baby Food	Offline	H	05-28-2010	669165933	6/27/2010	9925	255.28	159.42	2533654	1582243.5	
Central America and the Caribbean	Grenada	Cereal	Online	C	08-22-2012	963881480	9/15/2012	2804	205.7	117.11	576782.8	328376.44	
Europe	Russia	Office Supplies	Offline	L	05-02-2014	341417157	05-08-2014	1779	651.21	524.96	1158502.59	933903.84	
Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	06-20-2014	514321792	07-05-2014	8102	9.33	6.92	75591.66	56065.84	
Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	02-01-2013	115456712	02-06-2013	5062	651.21	524.96	3296425.02	2657347.52	
Australia and Oceania	Solomon Islands	Baby Food	Online	C	02-04-2015	547995746	2/21/2015	2974	255.28	159.42	759202.72	474115.08	
Sub-Saharan Africa	Angola	Household	Offline	M	04-23-2011	135425221	4/27/2011	4187	668.27	502.54	2798046.49	2104134.98	
Sub-Saharan Africa	Burkina Faso	Vegetables	Online	H	07-17-2012	871543967	7/27/2012	8082	154.06	90.93	1245112.92	734896.26	
Sub-Saharan Africa	Republic of the Congo	Personal Care	Offline	M	07-14-2015	770463311	8/25/2015	6070	81.73	56.67	496101.1	343986.9	
Sub-Saharan Africa	Senegal	Cereal	Online	H	04-18-2014	616607081	5/30/2014	6593	205.7	117.11	1356180.1	772106.23	
Asia	Kyrgyzstan	Vegetables	Online	H	06-24-2011	814711606	07-12-2011	124	154.06	90.93	19103.44	11275.32	
Sub-Saharan Africa	Cape Verde	Clothes	Offline	H	08-02-2014	939825713	8/19/2014	4168	109.28	35.84	455479.04	149381.12	
Asia	Bangladesh	Clothes	Online	L	01-13-2017	187310731	03-01-2017	8263	109.28	35.84	902980.64	296145.92	
Central America and the Caribbean	Honduras	Household	Offline	H	02-08-2017	522840487	2/13/2017	8974	668.27	502.54	5997054.98	4509793.96	
Asia	Mongolia	Personal Care	Offline	C	02-19-2014	832401311	2/23/2014	4901	81.73	56.67	400558.73	277739.67	
Europe	Bulgaria	Clothes	Online	M	04-23-2012	972292029	06-03-2012	1673	109.28	35.84	182825.44	59960.32	
Asia	Sri Lanka	Cosmetics	Offline	M	11-19-2016	419123971	12/18/2016	6952	437.2	263.33	3039414.4	1830670.16	
Sub-Saharan Africa	Cameroon	Beverages	Offline	C	04-01-2015	519820964	4/18/2015	5430	47.45	31.79	257653.5	172619.7	
Asia	Turkmenistan	Household	Offline	L	12-30-2010	441619336	1/20/2011	3830	668.27	502.54	2559474.1	1924728.2	
Australia and Oceania	East Timor	Meat	Online	L	07-31-2012	322067916	09-11-2012	5908	421.89	364.69	2492526.12	2154588.52	

The data in the preview has been truncated due to size limits.

Extract Table Using Examples | Load | Transform Data | Cancel

3.5 Data to Insights through Visualizations and Data Analysis



4. Deployment in Power BI Service