Amazon Sales Data Analysis

High-Level Design (HLD)

Soumyadip Tikader

Abdul Atif

Version: 1.0

Date: 28-01-2023

Project Title	Analyzing Amazon Sales Data
Technologies	Business Intelligence
Domain	E-commerce
Project Difficulties Level	Advanced

Document Version Control

Date Issued	Version	Description	Author
26-01-2023	1.0	First Version on complete HLD	Soumyadip Tikader
28-01-2023	1.1	Revised Version	Abdul Atif

Abstract

Organizations under the E-commerce industry seek to attain core competence by creating and sustaining a unique process to collect personal information about customers and their purchasing trends. The report critically evaluates how a service-based organization - Amazon uses Management information systems as a vibrant tool in attaining competitive advantage through efficient management and acquisition of information. As in today's market without proper sales management, it's very hard to predict how the business is running and how it will be in future. Many companies with proper sales management have shown better growth as they already know which item they must focus on, which product needs some improvement etc. Sales Management helps in maintaining its customer base for a longer time by providing them attractive offers, as they already have the information's like who are their top customers, whom they must focus on etc. Sales Management also helps in minimizing the losses. Also, Competition is increasing day by day as many new companies are coming with better management systems and giving tough competition due to that it is now very important to have a proper Sales Management to run any business and to compete with these companies.

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.

World is moving towards digitalization with great pace. E-commerce websites, online trading is expanding gradually. In this project we are analysing data for Amazon e-commerce website, it's time to time sales, sales trends, and other factors such as customers connected with their website. Our main aim is so analysing data, to find useful insights which will be used to take further business decisions.

Table of Contents

1. Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding.

This document is also intended to help detect contradictions prior to coding and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all the design aspects and define them in detail.
- Describe the user interface being implemented.
- Describe the hardware and software interfaces.
- Describe the performance requirements.
- Include design features and the architecture of the project.
- List and describe the non-functional attributes like
    - Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - Resource utilization
    - Serviceability

1.2 Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly technical terms which should be understandable to the administrators of the system.

1.3 Definitions

- ETL – Extract, Transform, Load
- EDA – Exploratory Data Analysis

- CSV file – Comma-separated values file, opened in MS Excel
- Fair Share – An administratively set data rate per time frame that is considered fair.
- Postgres SQL Server – A database management system.
- Power Query – With Power Query, you can connect to many different data sources and transform the data into the shape you want.
- ER – Entity Relation Diagram
- Power BI - An interactive data visualization software company focused on business intelligence.

2. General Descriptions

2.1 Product Perspective and 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.

The objective of the project is to perform data visualization techniques to understand the insight of the data. This project aims apply various Business Intelligence tools such as Tableau or Power BI to get a visual understanding of the data.

The goal of this project is to analyse sales of Amazon for the growth of sales and company, based on profit at particular cities and sales per month and per year that describes the trend of sales. To achieve the goal, we used a data set that is formed by taking into consideration some of the information of 100 record of sales. The problem is based on the given information about sales trend of Amazon.

2.2 Tools Used

a. Microsoft Excel

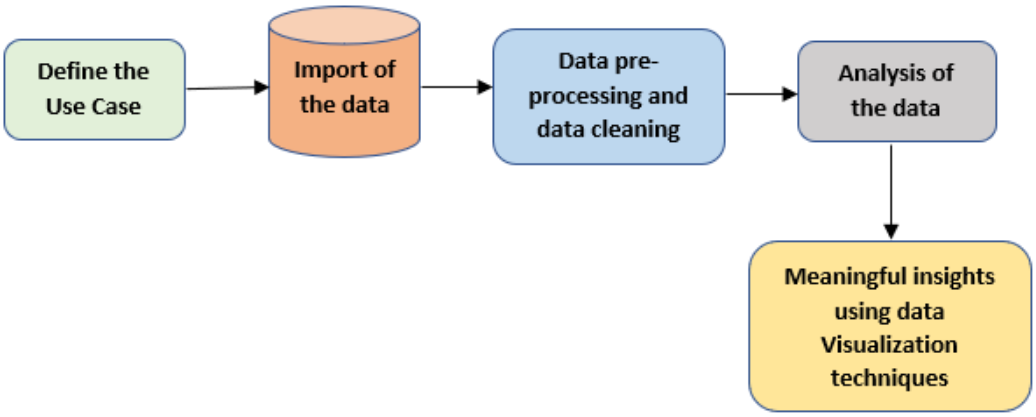
Microsoft Excel is a spreadsheet developed by Microsoft for Windows, macOS, Android and iOS. It features calculation or computation capabilities, graphing tools, pivot tables, and a macro programming language called Visual Basic for Applications. Excel forms part of the Microsoft Office suite of software.
Microsoft Excel is used for loading the data in CSV format, basic data cleaning and filter operations to execute the program.
MS Excel file was loaded into Power BI software.

b. Microsoft Power BI

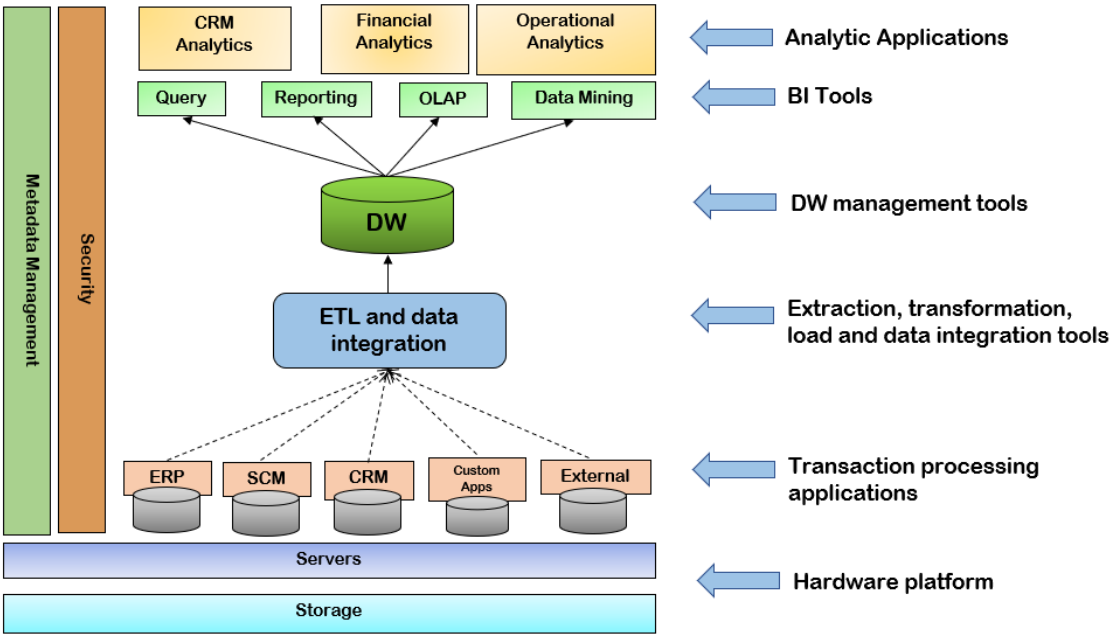
Power BI is an interactive data visualization software product developed by Microsoft with a primary focus on business intelligence. It is part of the Microsoft Power Platform. Power BI is a collection of software services, apps, and connectors that work together to turn unrelated sources of data into coherent, visually immersive, and interactive insights. Data may be input by reading directly from a database, webpage, or structured files such as spreadsheets, CSV, XML, and JSON.
With the help of Power BI, we have done the complete analysis and visualization.

3. Design Details

3.1 Process Flow



3.2 Functional Architecture



The current business environment is constantly evolving. The global economic scenario is providing opportunities as well as challenges. The factors affecting the business environment are consumer needs, globalization, and government policies, etc.

In such a business environment, an organization basically has four action steps. The organization can be reactive, anticipative, adaptive, or/and proactive. For this, the organization can develop a new strategy, get into partnerships, etc.

Today most the businesses are having computerized business support. This support is in form of a decision support system, business analysis, etc.
The main objective of business intelligence is to bridge the gap between an organization's current status and its desired position. Business intelligence helps the organization achieve commercial success along with sound financial management.

Business intelligence is a framework designed to support the decision-making process. This framework combines architecture, database, analytical tools and applications.

Business analytics forms an integral part of business intelligence.
Framework of Business Intelligence.
More and more businesses are moving towards business intelligence. The reason for this movement is the business environment. Organizations are forced to capture, store and interpret data. This data is at the core of business success.
Organizations require correct information for any decision-making process.

Business intelligence combines data warehousing, business analytics, performance, strategy and user interface. Business receives data from various sources. This data is capture in the data warehouse where it is stored, organized and summarized as per further utilization. Authorized users can access this data and work on it to get desired results. This result than are shared to executives for decision-making process. These data results can be published through dashboards or share points.

Business Intelligence Architecture and Components
The main components of business intelligence are data warehouse, business analytics and business performance management and user interface.
Data warehouse holds data obtained from internal sources as well as external sources. The internal sources include various operational systems.

Business analytics creates a report as and when required through queries and rules. Data mining is also another important aspect of business analytics.

Business performance management is a linkage of data with business objectives for efficient tracking. This business performance is then broadcasted to an executive decision- making body through dashboards and share-point.

Benefit of Business Intelligence

The benefits of Business intelligence are as follows:

- Business intelligence is faster more accurate process of reporting critical information.
- Business intelligence facilitates better and efficient decision-making process.
- Business intelligence provides timely information for better customer relationship management.
- Business intelligence improves profitability of the company.
- Business intelligence provides a facility of assessing organization's readiness in meeting new business challenges.
- Business intelligence supports usage of best practices and identifies every hidden cost.

Business intelligence usage can be optimized by identifying key projects on which company would like to focus. This process of highlighting key projects is called business intelligence governance.

3.3 Optimization

Our data strategy drives performance

- Minimize the number of fields
- Minimize the number of records
- Optimize extracts to speed up future queries by materializing calculations, removing columns and the use of accelerated views

Reduce the marks (data points) in our view

- Practice guided analytics. There's no need to fit everything you plan to show in a single view. Compile related views and connect them with action filters to travel from overview to highly granular views at the speed of thought.
- Remove unneeded dimensions from the detail shelf.
- Explore. Try displaying your data in different types of views.

Limit your filters by number and type

- Reduce the number of filters in use. Excessive filters on a view will create a more complex query, which takes longer to return results. Double-check your filters and remove any that aren't necessary.
- Use an include filter. Exclude filters load the entire domain of a dimension, while include filters do not. An include filter runs much faster than an exclude filter, especially for dimensions with many members.
- Use a continuous date filter. Continuous date filters (relative and range-of-date filters) can take advantage of the indexing properties in your database and are faster than discrete date filters.
- Use Boolean or numeric filters. Computers process integers and Booleans (t/f)

much faster than strings.

- Use parameters and action filters. These reduce the query load (and work across data sources).

Optimize and materialize your calculations

- Perform calculations in the database
- Reduce the number of nested calculations.
- Reduce the granularity of LOD or table calculations in the view. The more granular the calculation, the longer it takes. o LODs - Look at the number of unique dimension members in the

calculations

- Table Calculations - the more marks in the view, the longer it will take to calculate. Where possible, use MIN or MAX instead of AVG. AVG requires more processing than MIN or MAX. Often rows will be duplicated and display the same result with MIN, MAX, or AVG

4. KPI's

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the disease. As and when the system starts to capture the historical/periodic data for a user, the dashboards will be included to display charts over time with progress on various indicators or factors.

KPIs (Key Performance Indicators)

Key indicators displaying a summary of the Crop Production area and production information based on various parameters –

- Region of respective order.
- Country of respective order.
- Type of item of the respective order.
- Sales Channel i.e., Offline or Online.
- Order Date
- Order ID
- Ship Date
- Units Sold
- Unit Price
- Unit Cost
- Total Revenue
- Total Cost
- Total Profit

5. Deployment

Prioritizing data and analytics couldn't come at a better time. Your company, no matter what size, is already collecting data and most likely analyzing just a portion of it to solve business problems, gain competitive advantages, and drive enterprise transformation. With the explosive growth of enterprise data, database technologies, and the high demand for analytical skills, today's most effective IT organizations have shifted their focus to enabling self-service by deploying and operating Power BI Visualization at scale, as well as organizing, orchestrating, and unifying disparate sources of data for business users and experts alike to author and consume content.

- **Patterns in business operations:** Data visualization techniques help us to determine the patterns of business operations. By understanding the problem statement and identifying the solutions in terms of pattering and applying them to eliminate one or more of the inherent problems.
- **Identify business trends and relate to data:** These techniques help us identify market trends by collecting the data on Day-To-Day business activities and preparing trend reports, which help track the business how influences on the market. So that we could understand the competitors and customers. Certainly, this helps to long-term perspective.
- **Storytelling and Decision making:** Knowledge of storytelling from available data is one of the niche skills for business communication, specifically for the Data Science domain which is playing a vital role. Using the best visualization this role can be enhanced much better way and reach the objectives of business problems.
- **Understand the current business insights and set the goals:** Businesses can understand the insight of the business KPIs, finding tangible goals and business strategy planning, therefore they could optimize the data for business strategy plans for ongoing activities.
- **Operational and Performance analysis:** Increase productivity with the help of visualization techniques the clarity of KPIs depicting the trends of the productivity of the manufacturing unit, and guiding were to improve the productivity of the plant.

6. References

- https://en.wikipedia.org/wiki/Amazon_(company)