Amazon Sales Data Analysis

High Level Design (HLD)

Document Version Control

Date Issued	Version	Description	Author
04-05-2024	**1.0**	**First version of Complete HLD**	Jyoti Prava Samal

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 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 of 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
- ➢ Power Query – With Power Query, you can connect to many different data sources and transform the   data into the shape you want ➢ Power BI - An interactive data visualization software company focused on business intelligence.

2 General Description

2.1 Product Perspective & 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 Analyse Amazon Sales data to get a substantial data which will help in bringing changes in a business in the future. It will help to reveals flaws in the business model or in the way that one is going about conducting business. Sellers will be able to clearly see where they're losing money, what the problem is, and reduce their losses accordingly. It facilitates coming up with strategic solutions to problems. This project aims to provide visual understanding of the data using Microsoft Power Bi

2.2 Tools used

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.

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 Functional Architecture
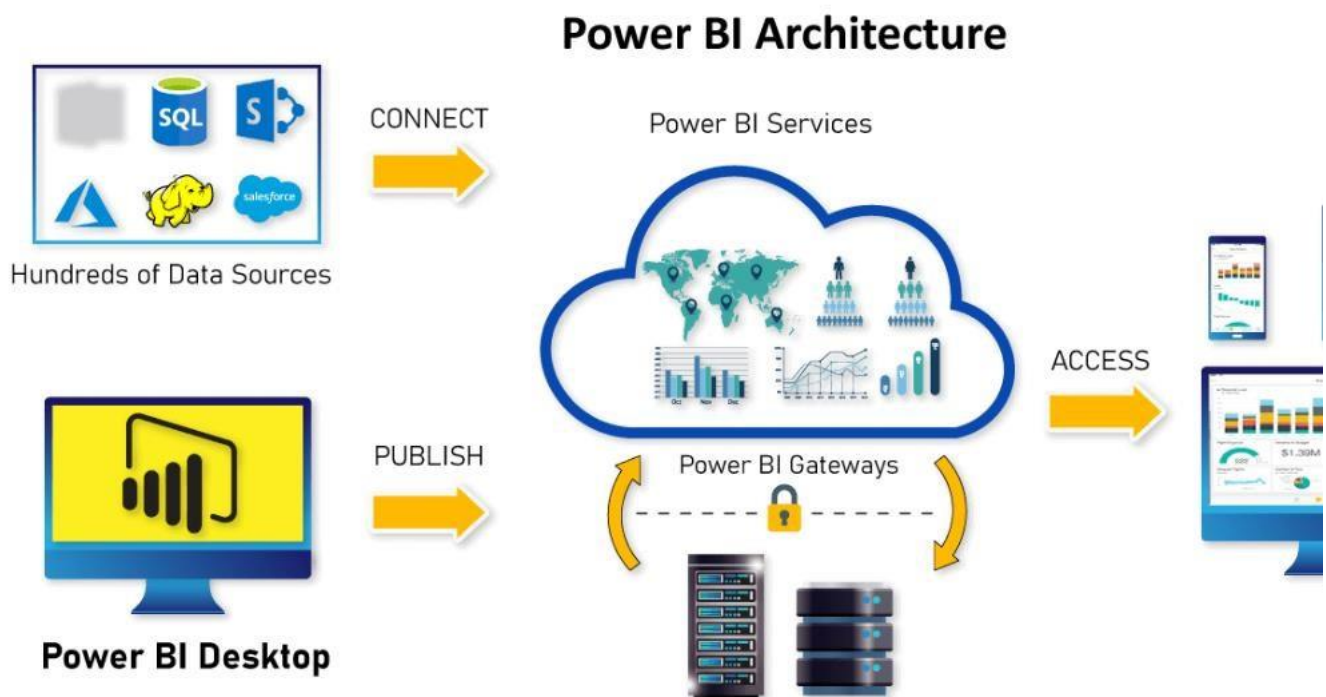
Power BI

Power BI is a business analytics solution that lets you visualize your data and share insights across your organization, or embed them in your app or website. Connect to hundreds of data sources and bring your data to life with live dashboards and reports. It provides interactive visualizations with self-service business intelligence capabilities, where end users can create reports and dashboards by themselves, without having to depend on any information technology staff or database administrator.

<u>Benefit of Business Intelligence</u>

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

3.2 Optimization

Your 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 andtype • 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 calculation.

•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.KPIs

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators about the sales of the products in various years.

Sales representative according to the number of the sales.

4.1 KPIs (Key Performance Indicator)

Key indicators displaying a summary of Sales Data and its relationships with different metrics.

1. Total Profit Distribution Yearly, Quarterly and Monthly.
2. Items That Generated Highest Revenue and Profit.
3. Top 5 Items that generated highest Profit in terms of percentage.
4. Total Profit Across Various Region and Country.
5. Total Revenue Contribution by Per Item.
6. Total Revenue and Total Cost by Item type.
7. Revenue generated Year over Year.
8. Last two years Revenue distribution with profit percentage.
9. Total Unit sold Across all the year.

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 analysing 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 Tableau at scale, as well as organizing, orchestrating, and unifying disparate sources of data for business users and experts alike to author and consume content.

5 Deployment