HIGH LEVEL DESIGN DOCUMENT

Amazon Sales Data Analysis

BY
VIKASH KUMAR MAHAPATRA

Revision No & Date - 1.3 20/09/2022

Document Version Control

Date Issued	Version	Description	Author
15 SEPT 2022	1.0	Created Document And Add Document Version Control Created Content Page ,Abstract Page & Introduction Page	Vikash
16 SEPT 2022	1.1	Created General Description Page, Design Detail ,KPI Page, Add Page Number in Content Section	Vikash
20 SEPT 2022	1.2	Content added to KPI, Content added of Deployment, Add Page No	Vikash
20 SEPT 2022	1.3	Final Revision	Vikash

CONTENTS

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 have to 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 have to focus on etc. Sales management so helps in minimizing the losses. Also, the 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.

INTRODUCTION

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 before 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

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.

GENERAL DESCRIPTION

PRODUCT PERSPECTIVE & PROBLEM STATEMENT

As new e-commerce popping as well as their steady rise in competition from existing e-commerce players as well as new budding startup who have better management as well overall good systems analysis are giving them edge over the existing e-commerce players, it is now very important to have a proper sales management to run any business and to compete with these companies

The problem statement aims to analyze amazon sales from 2017-2019 to know more in depth information about the items that are in high demand, items that are generating high profit as well items that shouldn't be sold and how much stock have to maintained for further increase in the sales. To achieve the goal, we used a data set that is given and analyze most important parameter that are responsible for it. Also get some meaningful insights from the given information regarding Sales Data.

TOOLS

Business Intelligence tools and libraries works such as NumPy, Pandas, Seaborn, Matplotlib, MS-Excel, Power BI, Jupyter Notebook and Python Programming Language are used to build the whole framework.

DESIGN DETAILS

FUNCTIONAL ARCHITECTURE

HOW BI WORKS

ORGANIZATIONAL MEMORY	INFORMATION INTEGRATION	INSIGHT CREATION	PRESENTATION
Data Warehouse Enterprise resource planning(ERP) Knowledge Repository Content Management System(CMS)	Business Analytical Tools Data Mining Real Time Decision	Text Mining Tool Web Mining Tool Environmental Scanning RFID	Online Analytical Processing Tool Visualization Tool Digital Dashboard Score Card

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 & the use of accelerated views.

- Reduce the marks (data points) in your view
  - Practice guided analytics. There is no need to fit everything you plan to single view. Compile related views and connected them with action filtered to travel from overview to highly-granular views at the speed of thought.
  - Remove unneeded dimensions from the detail self.
  - Explore try displacing your data in different type 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 including 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 data 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.
    - 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.
  - Make groups with calculations. Like include filters, calculated groups load only named members of the domain, whereas Tableau's group function loads the entire domain.
  - Use Booleans or numeric calculations instead of string calculations. Computer can process integers and Booleans (t/f) much faster than strings (Boolean>Int>Float>Date-Time>String).

KPI

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators. 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 Indicator)

Key indicators displaying a summary of the Sales Data and its relationship with different metrics are as follows:-

- Yearly and Quarterly Ups in sales and profit.
- Items that generated highest sales.
- Yearly and Quarterly Down in sales and profit.
- Top 10 items sold.
- Least 10 items sold.

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 applied 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 helps track the business how influences 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 best visualization this role can be enhanced much better way and reaching the objectives of business problems.
- **Understanding current business insights and setting up goals:** Business can be understand from insight of KPI like finding some tangible goals, past performance, past strategy planning & many more factors.
- **Operational and Performance analysis:** Increase the productivity with the help of visualization techniques the clarity of KPIs depicting the trends of the productivity and guiding were to improve the productivity.

Power BI prioritizes choice in flexibility to fit, rather than dictate, your enterprise architecture. Power BI Desktop and Power BI Service leverage your existing technology investments and integrate them into your IT infrastructure to provide a self-service, modern analytics platform for your users. With on-premises, cloud, and hosted options, there is a version of Power BI to match your requirements.