# High Level Design

## Analyzing Amazon Sales Data

Revision number: 1.0

Last date of revision: 26/08/2022

# Document Version Control

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

# Contents

Document Version Control ..... 2

Abstract ..... 4

1. Introduction ..... 5

    1.1. Purpose of the Document ..... 5
    1.2. Objective of HLD ..... 5
    1.3. Scope of HLD ..... 5

2. General Description ..... 6

    2.1 Product Perspective & Problem Statement ..... 6
    2.2 Data Requirements ..... 6
    2.3 Tools Used ..... 7

3. Design Details ..... 8

    3.1 Functional Architecture ..... 8
    3.2 Error Handling ..... 8
    3.3 Optimization ..... 8

4. KPI's ..... 9

    4.1. KPI's (Key Performance Indicators) ..... 10

5. Deployment ..... 11

# 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 analyzing consumer behavior, 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 analyze different metrics like sales, Sales Quantity, Discount rate, Sales over years etc. By analyzing 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 behaviors 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

This document will be used for documenting High-level designs of project.

## 1.1 Purpose of the Document

The purpose of this plan is to

- Describe different design approaches.
- Describe different analysis approaches based on variety of Use Cases.
- Describe third party components/tools required for the system.
- Present complete Process Flow followed for this project.

## 1.2 Objective of HLD

1. To provide an overview of the entire system.
2. To provide introduction of Problem Perspective & Statement, Data Requirements, Tools used and many more.
3. To provide a module-wise breakup of the entire system.

## 1.3 Scope of HLD

This HLD covers all areas of system.

# 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 Analyze 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 Data Requirements

Data Requirement completely depend on our problem.

- In this project, to perform analysis, we are using datasets that is provided by iNeuron Intelligence Pvt. Ltd.
- We make a use of those datasets as per the requirement and the problem statement.
- The features which are taken into consideration are:
- Some of the important features are:

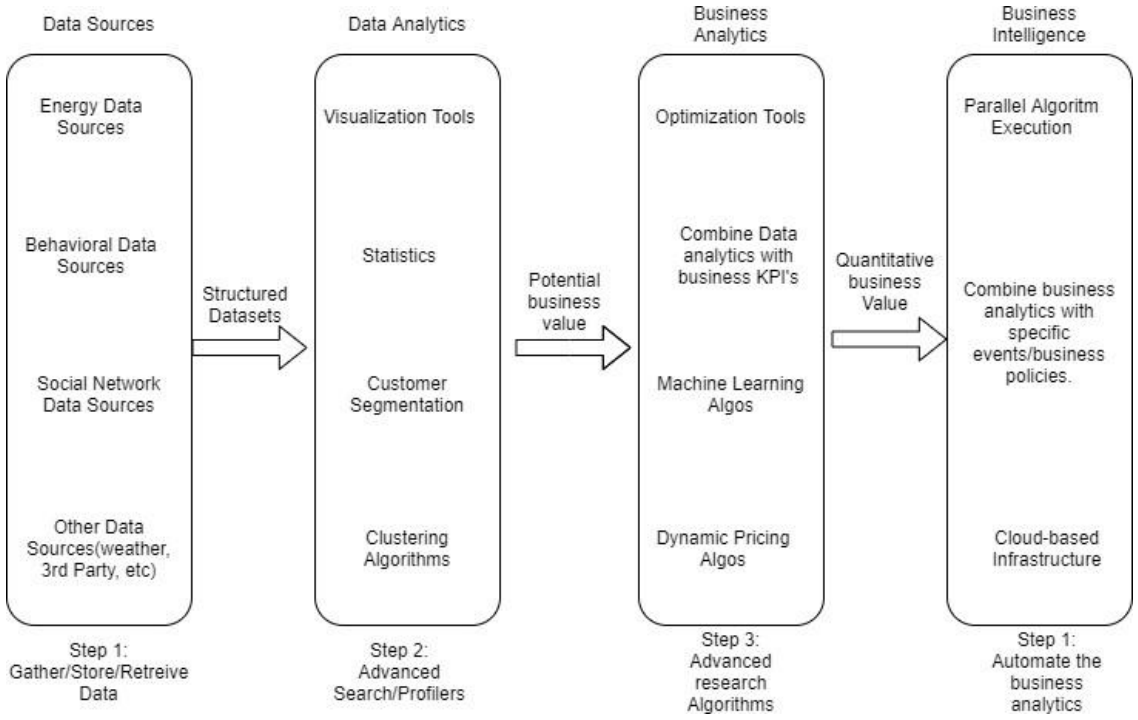
| Name | Description |
|---|---|
| **Region** | Region Name from which order was placed |
| **Country** | Country Name from which order was placed |
| **Item Type** | Product Category from which Items were ordered |
| **Sale Channel** | Mode of a sale |
| **Order Priority** | Delivery mode opted by the customer |
| **Order Date** | Date on which order was placed |
| **Order ID** | A unique order ID pertaining to each order placed |
| **Ship Date** | Date on which order expected to be delivered |
| **Units Sold** | Quantity of units sold |
| **Unit Price** | Selling Price per unit of a product |
| **Unit Cost** | Purchased cost per unit of a product |
| **Total Revenue** | Total Revenue/Sales made |
| **Total Cost** | Total cost which has invested |
| **Total Profit** | Total Profit which is made after deducting the Total Cost from Total Revenue |

## 2.3 Tools Used

- Pandas, Microsoft Excel and Power Query were used for Data Manipulation & Pre-processing.
- Microsoft Power BI is used as a Business Intelligence tool.
- GitHub is used as version control system.
- These are all the tools and libraries were used to build the whole framework.

# 3. Design Details

## 3.1 Functional Architecture




## 3.2 Error Handling:

We have designed this project in such a way that, at any step if any error occurs, the Business Intelligence interface will not terminate, instead it should catch that error and display it with proper explanation.

## 3.3 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 your 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.
    - 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 Power BI's group function loads the entire domain.
- Use Booleans or numeric calculations instead of string calculations. Computers can process integers and Booleans (t/f) much faster than strings.

  Boolean>Int>Float>Date>DateTime>String

# 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 Indicators)

Key indicators displaying a summary of the Housing Price and its relationship with different metrics

1. Revenue vs Budget
2. Revenue & Budget by Year and Month
3. Revenue & Profit by Weekday
4. Top 10 Items by Revenue, Profit & Orders
5. Total Profit by Sales Channel
6. Total Profit by Order Priority

# 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. Analyzing Data in Microsoft Power BI empowers you to understand your data through natural language queries that allow you to ask questions about your data without having to write complicated formulas. It will provide answers with stunning visuals such as Charts along with Tooltips, Tables with Conditional Formatting, Maps that can then be placed into the report.