High-Level Design (HLD)

Amazon Sales Data Analysis

Revision Number -1

Last Date of Revision – 30/08/2024

Purba Nag

Document Version Control

Date issued	Version	Description	Author

30/08/2024	1.0	First Version of Complete HLD	Purba Nag

Contents

Abstract:

In the E-commerce industry, organizations strive to establish core competencies by developing and maintaining a distinctive process for gathering personal information about customers and their purchasing habits. This report provides a critical evaluation of how Amazon, a service-based organization, employs Management Information Systems (MIS) as a powerful tool for achieving a competitive advantage through efficient information management and acquisition. Effective sales management is crucial in today's market for predicting business performance and future prospects. Companies that implement proper sales management have demonstrated greater growth and are better positioned to focus on key products and customer segments, resulting in customer retention, attractive offers, and reduced losses. With the increasing competition in the market from new companies with better management systems, it is imperative to implement effective sales management to remain competitive.

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 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

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.

2 General Description

2.1 Product Prescriptive & Problem Statement

The project aim is to analyze Amazon sales Data from 2017 to 2019 to know more in-depth information about the items that are in high demand, items that are generating high profits which items should not be sold or has less demand not purchased by the customers, and how much stock we have to maintain for the further sales, etc.

2.2 Tools Used

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

3 Design Details

3.1 Functional Architecture

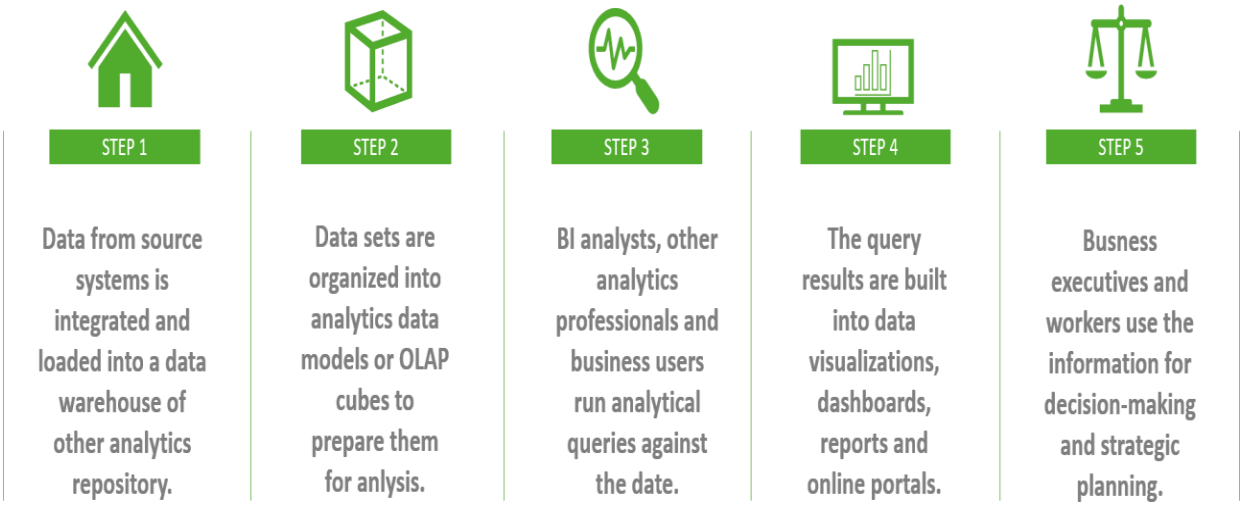
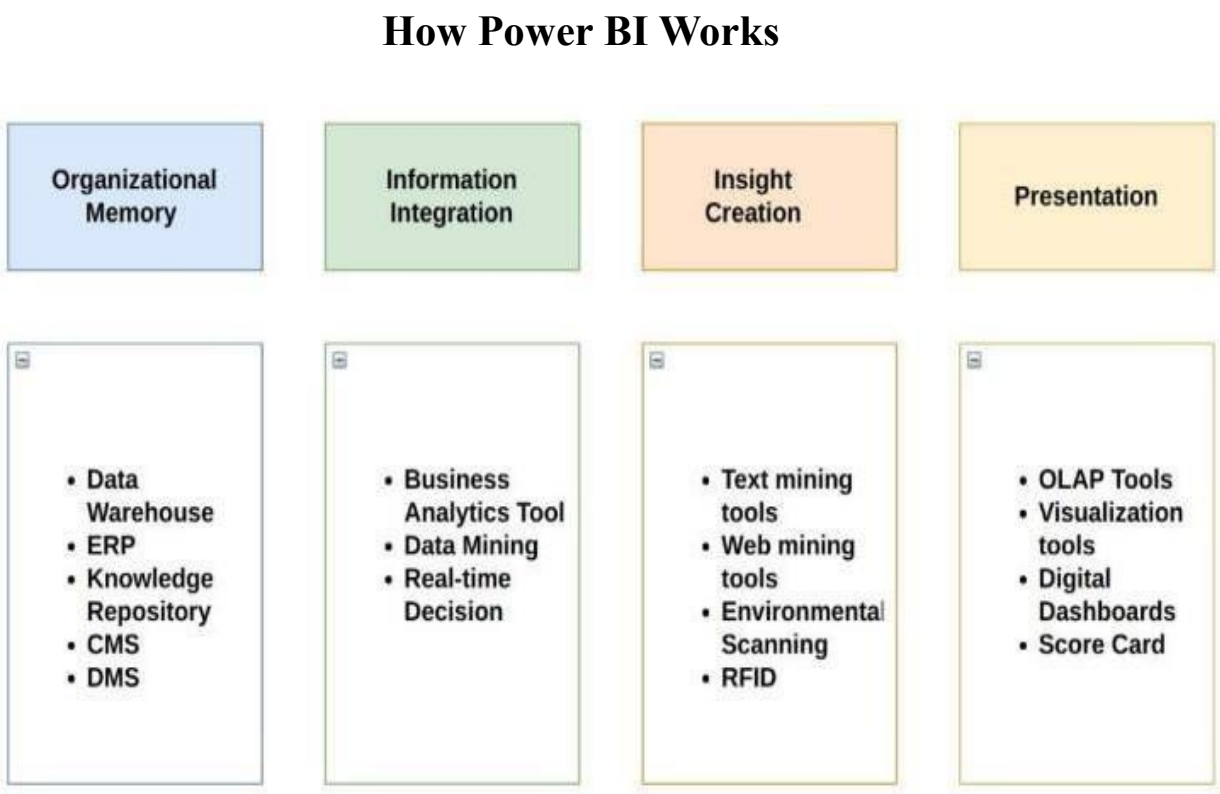


Figure 1: Functional Architecture of Business Intelligence



3.2 Optimization

1. 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.
2. 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.
3. 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).

4. 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. Computers can process integers and Booleans (t/f) much faster than strings. Boolean>Int>Float>Date>DateTime>String.

4. KPI

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

4.1 KPIs (Key Performance Indicators)

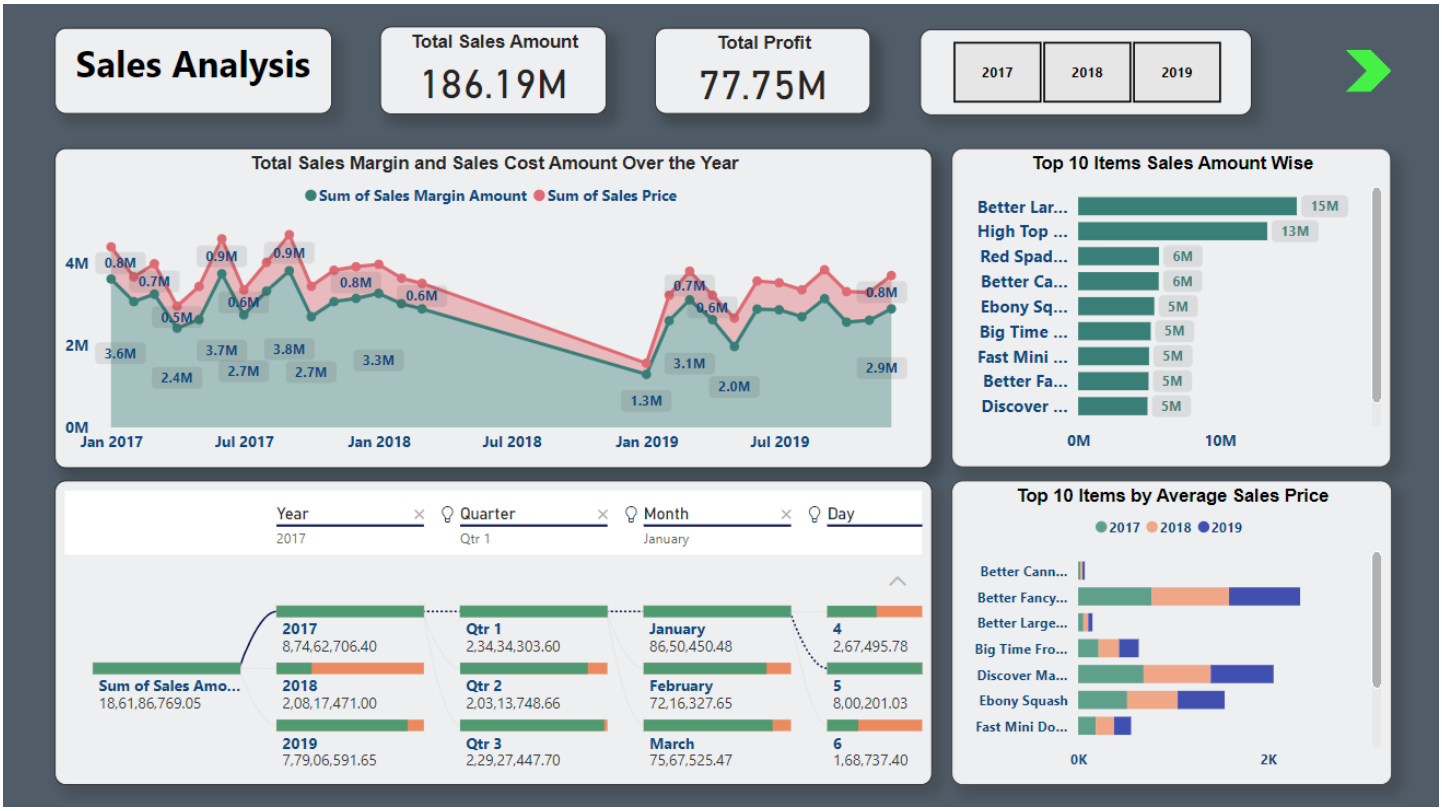
Key indicators displaying a summary of Sales Data and its relationships with different metrics.

1. Yearly, Quarterly, Monthly Ups and Downs in Sales & Profits.
2. Items That Generated Highest Sales, Profit etc.
3. Top 5 Items that generated highest Sales and Top 5 Items by Quantity.
4. Bottom 5 Items that generated Lowest Sales and Bottom 5 Items by Quantity.
5. Forecasting.

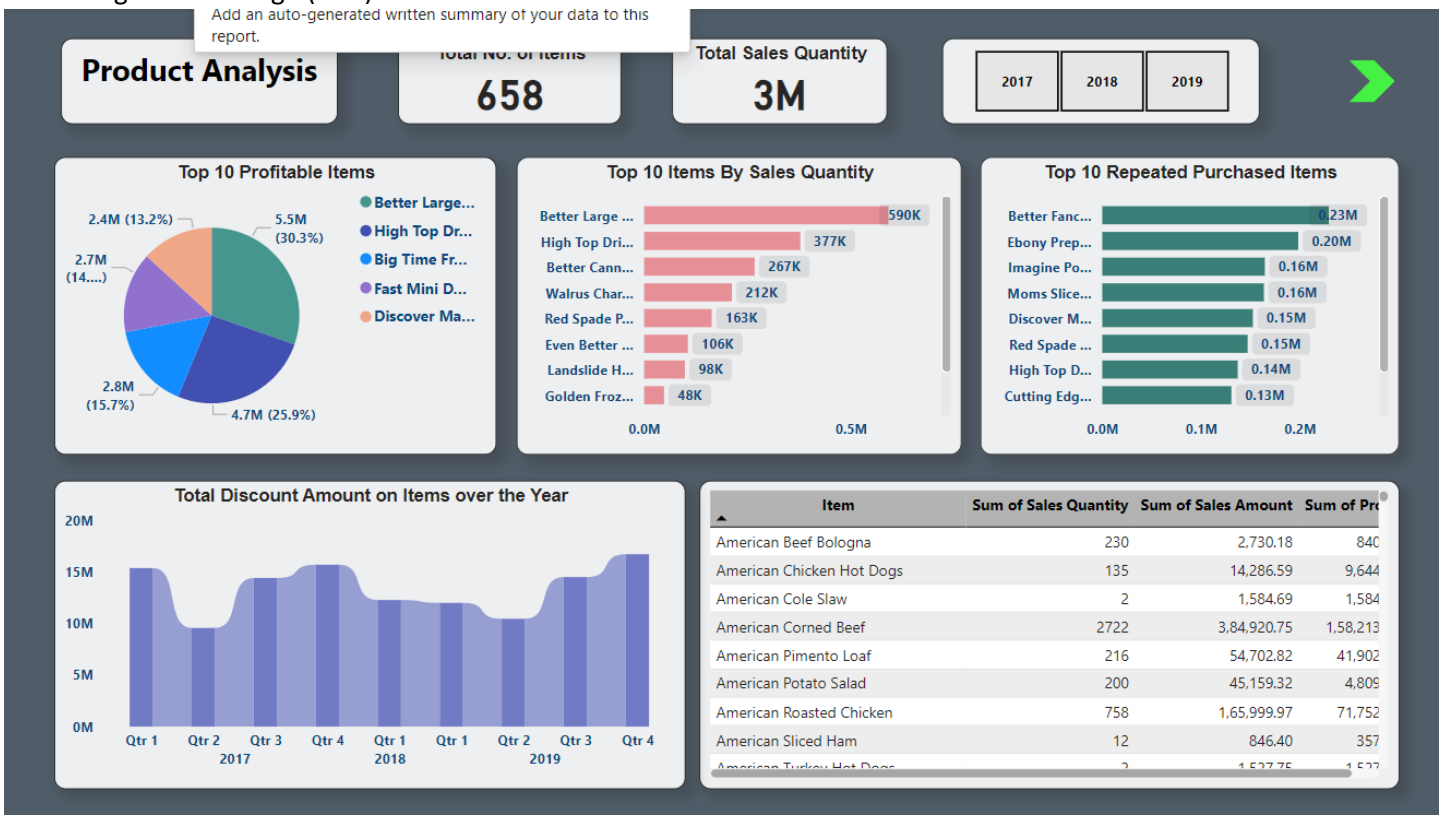
5 Deployment

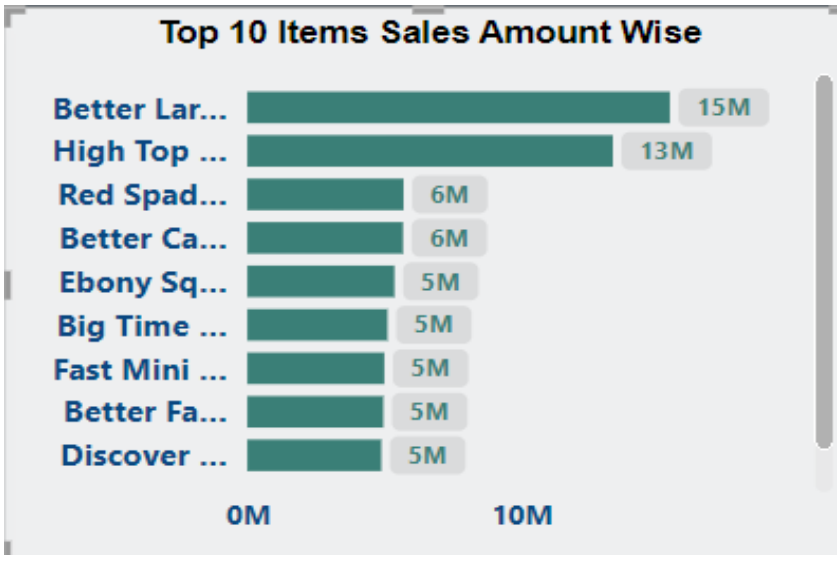
Prioritizing data and analytics couldn't come at a better time. Your company, no matter what size, is already collecting data and most likely Analyzing just a portion of it to solve business problems, gain competitive advantages, and drive enterprise transformation. With the explosive growth of enterprise 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 at scale, as well as organizing, orchestrating, and unifying disparate sources of data for business users and experts alike to author and consume content.

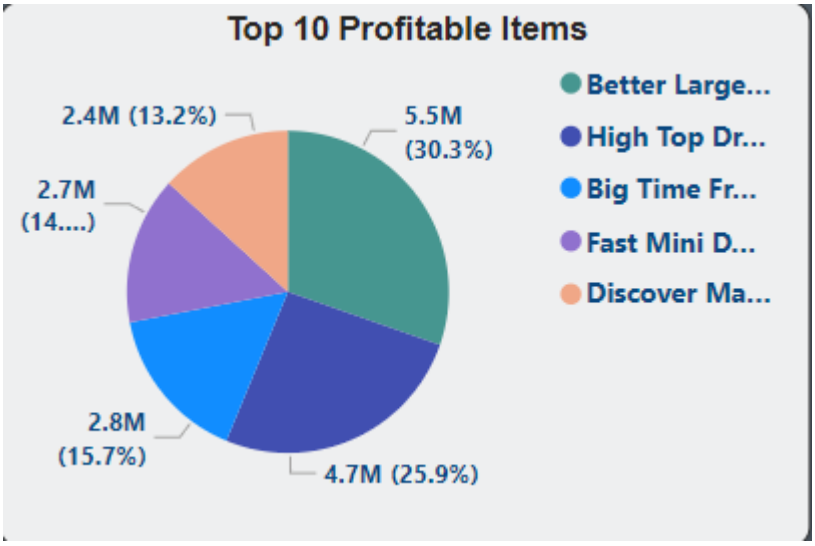
Power BI prioritizes choice in flexibility to fit, rather than dictate, your enterprise architecture. Power BI Desktop and Power BI Service leverage your existing technology investments and integrate them into your IT infrastructure to provide a self-service, modern analytics platform for your users.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

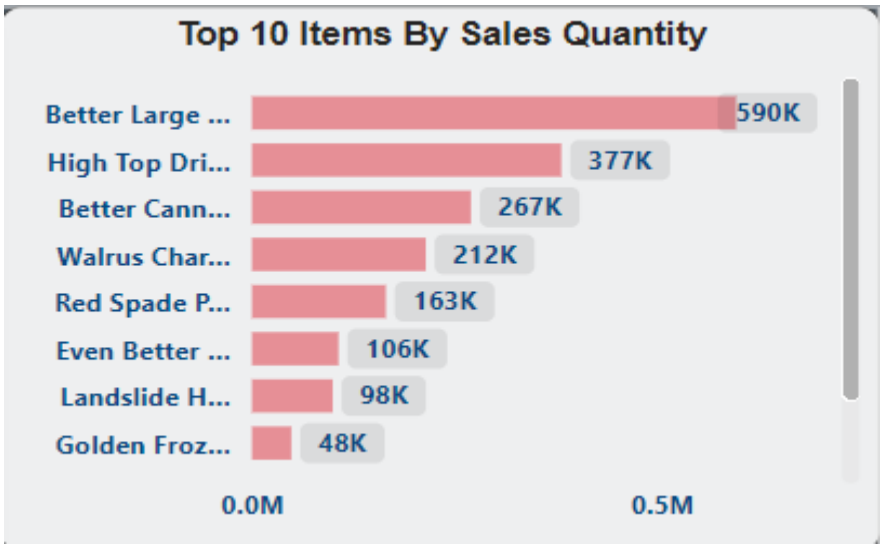

Sales Analysis
Total Sales Amount
186.19M
Total Profit
77.75M
2017
2018
2019
Total Sales Margin and Sales Cost Amount Over the Year
Sum of Sales Margin Amount
Sum of Sales Price
4M
2M
0M
Jan 2017
Jul 2017
Jan 2018
Jul 2018
Jan 2019
Jul 2019
Top 10 Items Sales Amount Wise
Better Lar... 15M
High Top ... 13M
Red Spad... 6M
Better Ca... 6M
Ebony Sq... 5M
Big Time ... 5M
Fast Mini ... 5M
Better Fa... 5M
Discover ... 5M
0M
10M
Year
2017
Quarter
Qtr 1
Month
January
Day
Sum of Sales Amo...
18,61,86,769.05
2017
8,74,62,706.40
2018
2,08,17,471.00
2019
7,79,06,591.65
Qtr 1
2,34,34,303.60
Qtr 2
2,03,13,748.66
Qtr 3
2,29,27,447.70
January
86,50,450.48
February
72,16,327.65
March
75,67,525.47
4
2,67,495.78
5
8,00,201.03
6
1,68,737.40
Top 10 Items by Average Sales Price
2017
2018
2019
Better Cann...
Better Fancy...
Better Large...
Big Time Fro...
Discover Ma...
Ebony Squash
Fast Mini Do...
0K
2K

12 High-Level Design (HLD)

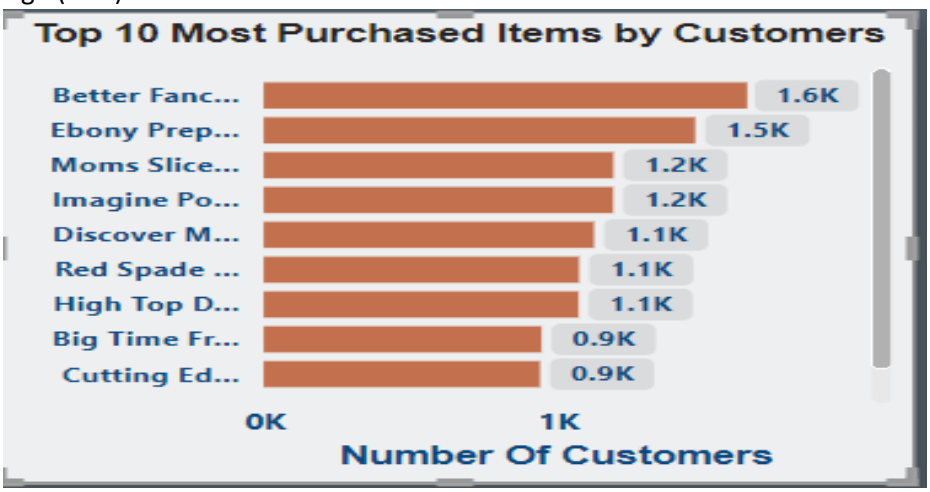


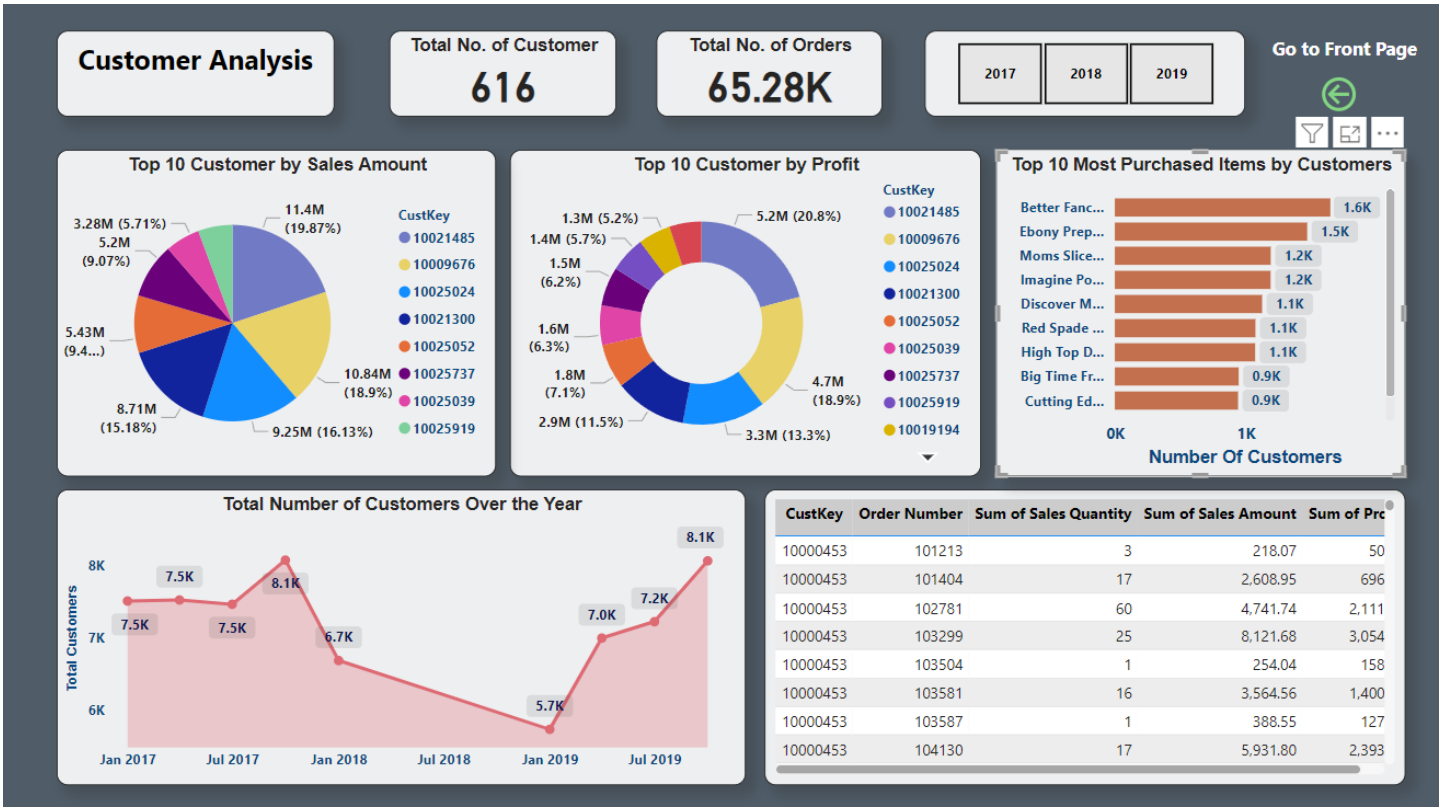



Top 10 Profitable Items
2.4M (13.2%)
5.5M
(30.3%)
2.7M
(14....)
2.8M
(15.7%)
4.7M (25.9%)
Better Large...
High Top Dr...
Big Time Fr...
Fast Mini D...
Discover Ma...

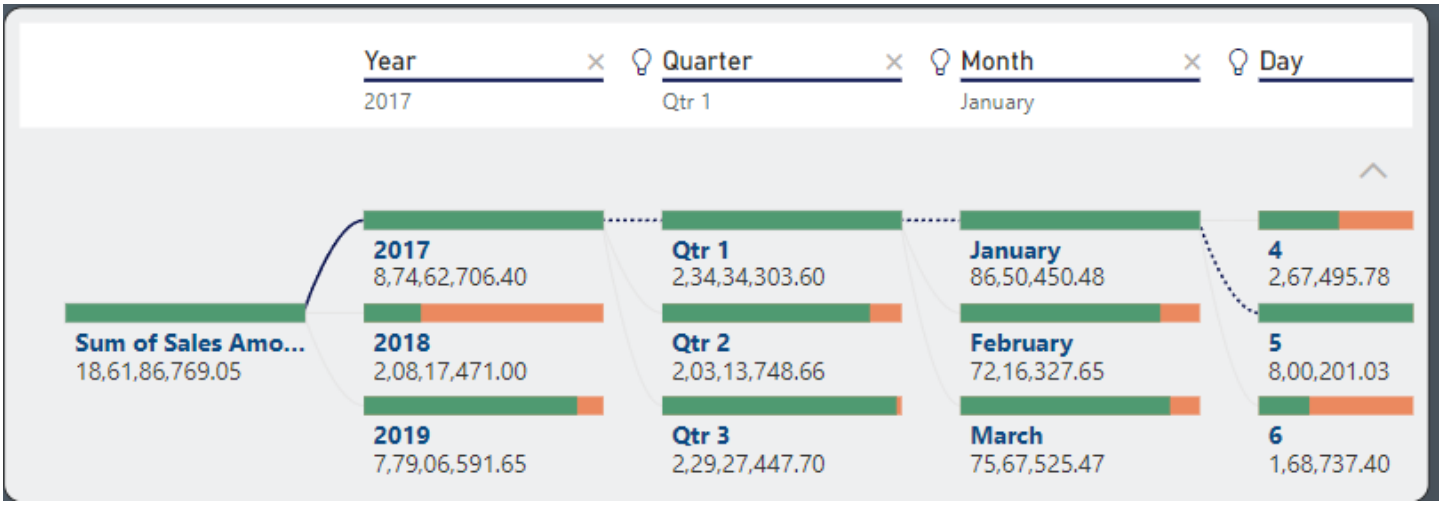

Top 10 Items By Sales Quantity
Better Large ...
590K
High Top Dri...
377K
Better Cann...
267K
Walrus Char...
212K
Red Spade P...
163K
Even Better ...
106K
Landslide H...
98K
Golden Froz...
48K
0.0M
0.5M

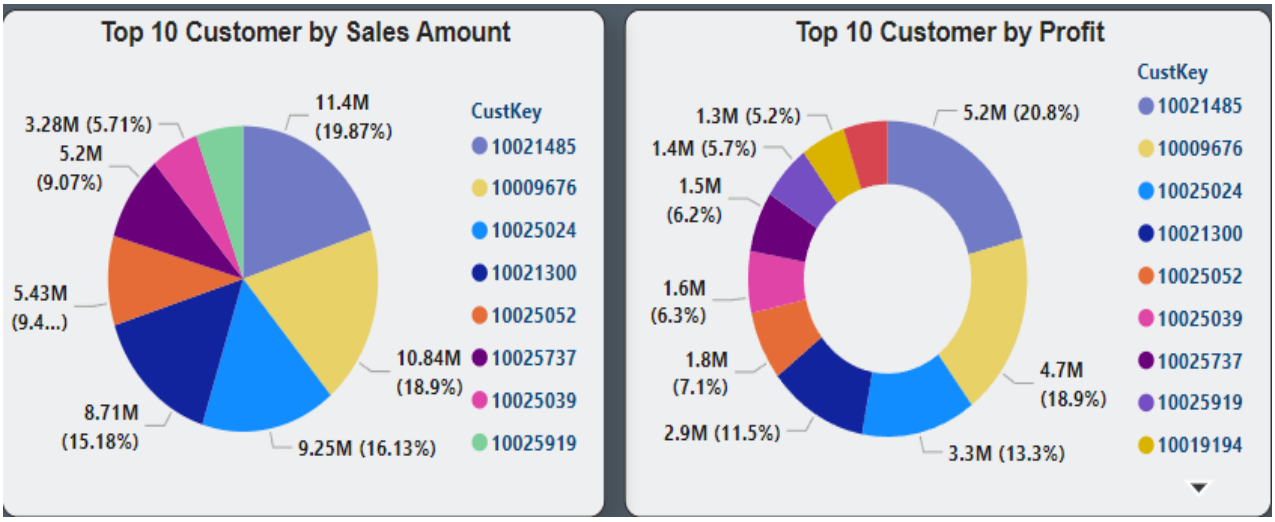
14 High-Level Design (HLD)





CustKey	Order Number	Sum of Sales Quantity	Sum of Sales Amount	Sum of Pro
10000453	101213	3	218.07	50
10000453	101404	17	2,608.95	696
10000453	102781	60	4,741.74	2,111
10000453	103299	25	8,121.68	3,054
10000453	103504	1	254.04	158
10000453	103581	16	3,564.56	1,400
10000453	103587	1	388.55	127
10000453	104130	17	5,931.80	2,393


Year
2017
Quarter
Qtr 1
Month
January
Day
Sum of Sales Amo...
18,61,86,769.05
2017
8,74,62,706.40
2018
2,08,17,471.00
2019
7,79,06,591.65
Qtr 1
2,34,34,303.60
Qtr 2
2,03,13,748.66
Qtr 3
2,29,27,447.70
January
86,50,450.48
February
72,16,327.65
March
75,67,525.47
4
2,67,495.78
5
8,00,201.03
6
1,68,737.40


Top 10 Customer by Sales Amount
11.4M (19.87%)
3.28M (5.71%)
5.2M (9.07%)
5.43M (9.4...)
10.84M (18.9%)
8.71M (15.18%)
9.25M (16.13%)
CustKey
10021485
10009676
10025024
10021300
10025052
10025737
10025039
10025919
Top 10 Customer by Profit
1.3M (5.2%)
5.2M (20.8%)
1.4M (5.7%)
1.5M (6.2%)
1.6M (6.3%)
1.8M (7.1%)
2.9M (11.5%)
4.7M (18.9%)
3.3M (13.3%)
CustKey
10021485
10009676
10025024
10021300
10025052
10025039
10025737
10025919
10019194