

High Level Design (HLD)

Sales Data Analysis

Revision Number: 1.0

Last date of revision: 10/02/2024

Bodapatla Vinay Kumar Reddy

Document Version Control

[illegible]

Contents

Document Version Control.....	2
Abstract.....	3
1 Introduction.....	4
1.1 Why this High-Level Design Document?.....	4
1.2 Scope.....	5
2 General Description.....	6
2.1 Product Perspective & Problem Statement.....	6
2.2 Tools used.....	6
3 Design Details.....	7
3.1 Functional Architecture.....	7
3.2 Optimization.....	8
4 KPIs.....	9
4.1 KPIs (Key Performance Indicators).....	9
5 Deployment.....	9

Abstract

Diwali sales data holds a profound significance as it mirrors economic trends and captures the attention of both buyers and sellers in the festive season. Unlike a typical home buyer's dream description, where details like basement ceiling height may not be prioritized, this dataset from a Diwali sales competition underscores that various factors beyond product features significantly influence purchase decisions.

Accurately predicting Diwali sales becomes a complex task, as buyers are not solely focused on product specifications or conventional attributes. The dataset suggests that numerous factors play a pivotal role in determining sales outcomes during the festive season. Beyond product features, considerations such as marketing strategies, pricing discounts, and the overall festive experience contribute to the dynamics of Diwali sales. Successfully navigating these factors requires a nuanced understanding of the diverse elements that shape consumer behavior and impact sales during the Diwali period, making the analysis and prediction of Diwali sales an intricate and essential aspect of retail strategy.

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.

2 General Description

2.1 Product Perspective & Problem Statement

Diwali sales data holds paramount significance, reflecting economic trends during the festive season and capturing the attention of both buyers and sellers. In this project, the objective is to predict sales outcomes leveraging explanatory variables that encompass diverse aspects of festive retail. The primary focus is on utilizing advanced data visualization techniques, employing Python libraries such as NumPy, Matplotlib, and Seaborn, to extract valuable insights from the dataset.

The project aims to harness the capabilities of NumPy for efficient numerical operations, Matplotlib for creating insightful visualizations, and Seaborn for enhancing the overall aesthetic and interpretability of the graphics. Through the combined power of these libraries, the project seeks to uncover trends, patterns, and correlations within the Diwali sales data. Visual narratives generated with Matplotlib and Seaborn will provide stakeholders with a comprehensive understanding of the factors influencing sales during the festive period. This project not only navigates the intricacies of Diwali sales but also demonstrates the prowess of Python's data analysis ecosystem in extracting actionable insights from complex datasets.

2.2 Tools used

Business Intelligence tools and libraries such as Numpy, Pandas are used to build the whole framework.



3 Design Details

3.1 Functional Architecture

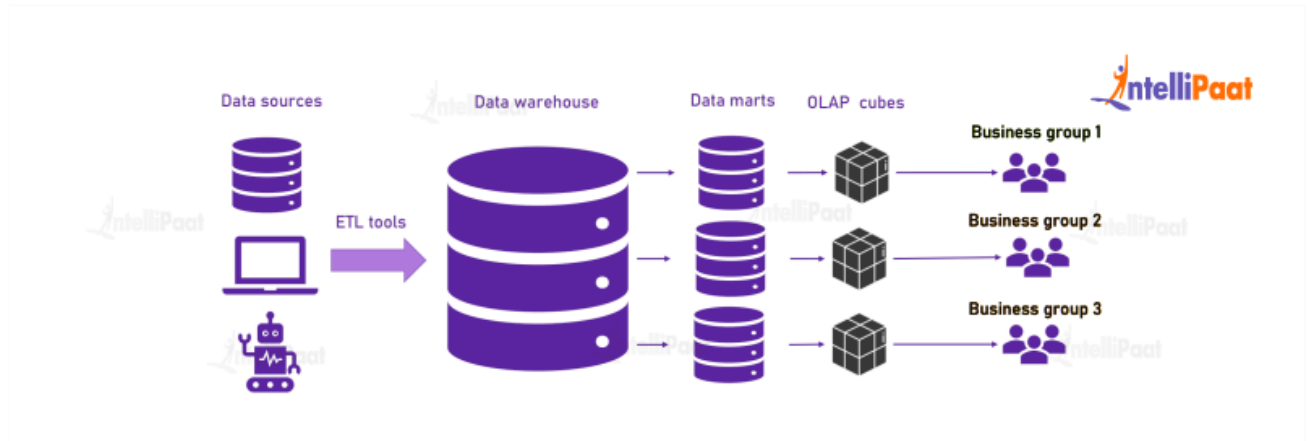
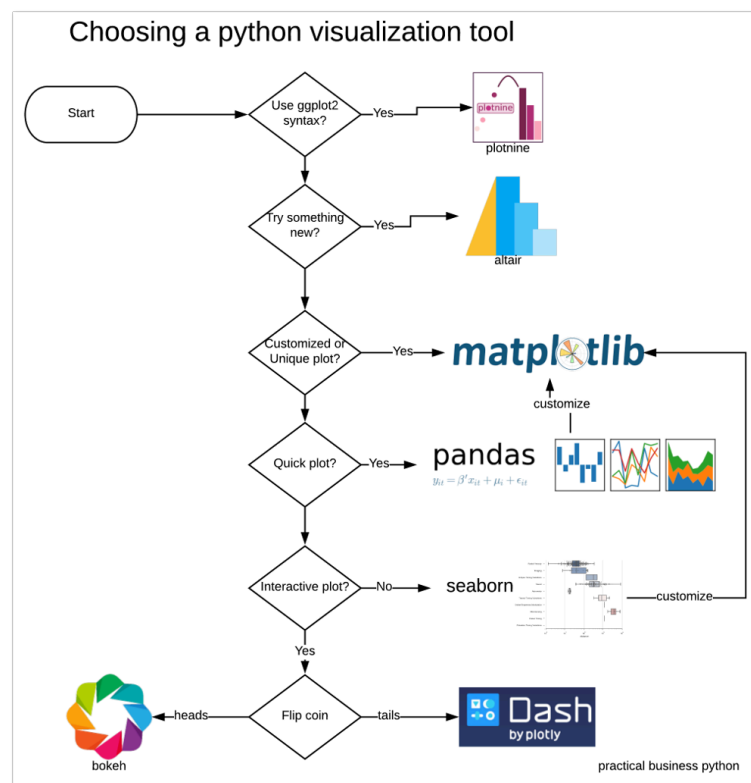


Figure 1: Functional Architecture of Business Intelligence



3.2 Optimization

In the context of Python data analytics, employing NumPy, Matplotlib, and Seaborn presents a robust framework for optimizing performance. These libraries are instrumental in implementing various strategies that enhance efficiency and streamline data visualization.

NumPy Optimization:

- **Minimizing Data Complexity:** NumPy excels in handling numerical operations efficiently, minimizing the complexity of data structures. Its array operations enable streamlined computations, contributing to enhanced performance.

Matplotlib and Seaborn Visualizations:

- **Reducing Data Points:** Matplotlib and Seaborn provide powerful visualization tools. Streamlining data points in visualizations helps improve rendering speed and overall visualization efficiency.

Guided Analytics Principles:

- **Interconnected Views:** Following guided analytics principles, create interconnected views using Matplotlib and Seaborn. This involves reducing the number of fields and records, and employing action filters to facilitate seamless navigation between related views.

Efficient Filter Handling with NumPy:

- **Strategic Filter Usage:** When dealing with filters, leverage NumPy strategically. Utilize continuous date filters or numeric/Boolean filters to optimize processing speed, taking advantage of NumPy's efficient numerical capabilities.

NumPy in Calculations:

- **Performing Calculations:** NumPy's efficient array operations make it ideal for performing calculations directly in arrays. This minimizes nested operations and reduces granularity, contributing to faster computation.

Numerical Function Optimization with NumPy:

- **Prioritizing MIN or MAX over AVG:** In numerical computations, prioritize NumPy functions like MIN or MAX over AVG where applicable. These functions typically require less processing, contributing to improved performance in calculations.

By meticulously incorporating these strategies with NumPy, Matplotlib, and Seaborn, your Python data analytics workflow becomes not only efficient but also optimized for delivering high-performance insights. The synergy of these libraries empowers data professionals to extract meaningful information swiftly and effectively.

4 KPIs

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 the Diwali sales and its relationship with different metrics

1. We can see that most of the buyers are females and even the purchasing power of females are greater than men
2. We can see that most of the buyers are of the age group between 26-35 yrs female.
3. We can see that most of the orders & total sales/amount are from Uttar Pradesh, Maharashtra and Karnataka respectively
4. We can see that most of the buyers are married (women) and they have high purchasing power
5. We can see that most of the sold products are from Food, Clothing and Electronics category

5 Deployment

Prioritizing data analytics is imperative in the contemporary business landscape, transcending company size. Python, with its robust libraries—Pandas, Numpy, Matplotlib, and Seaborn—becomes an instrumental ally in the analysis of Amazon Diwali Sales data. The data undergoes an Extract, Transform, Load (ETL) process using Microsoft Excel, followed by Python's prowess in data cleaning and visualization. This streamlined approach ensures meaningful insights, aligning seamlessly with the prevailing trend of self-service analytics. The visualizations generated offer an interactive exploration of sales trends, empowering stakeholders with a deeper comprehension of the Amazon Diwali Sales landscape. Python's adaptability allows for a dynamic analysis of intricate datasets, fostering a culture of data-driven decision-making. This concise and efficient workflow encapsulates the overarching strategy of harnessing analytics tools for actionable business intelligence, contributing to a more informed and responsive organizational environment.