High Level Design (HLD)
STORE SALE PREDICTION

Revision Number – 1.0

Last Date of Revision – 28-08-2023

Amit Kumar

Document Version Control

Date	Version	Description	Author
26-04-2023	1.0	Abstract	Amit Kumar
27-04-2023	1.1	Design Flow	Amit Kumar
28-04-2023	1.2	Performance Evaluation	Amit Kumar

Contents	
Abstract	4
INTRODUCTION	5
Why this HLD documentation?	5
1 Description	5
1.1 **Problem Perspective**	5
1.2 Problem Statement	5
1.3 Proposed Solution	5
1.4 Solution Improvements	7
1.5 Technical Requirements	7
1.6 Data Requirements	7
1.7 Tools Used	7
1.8 **Constraints**	8
1.9 Assumptions	8
2 Design Flow	9
2.1 Modelling Process	9
2.2 Deployment Process	9
2.3 **Logging**	10
2.2 **Error Handling**	10
3 Performance Evaluation	11
3.1 **Reusability**	11
3.2 **Application Compatibility**	11
3.3 **Resource Utilization**	11
3.2 Deployment	11
Conclusion	11

Abstract

This High-Lavel Design Document encapsulates the development of a comprehensive system poised to revolutionize inventory management and customer demand forecasting within the retail landscape. In the contemporary retail ecosystem, shopping malls and Big Marts meticulously collect and archive individual item sales data, yielding invaluable insights into consumer behavior and product specifics. These data repositories, securely stored within a data warehouse, serve as reservoirs of opportunity. Our system embarks on a transformative journey through the realms of data, expertly harnessing its power. The journey unfolds through meticulously orchestrated steps: data ingestion from Kaggle datasets, data transformation for cleanliness and relevance, model building to extract meaningful patterns, and the establishment of an efficient batch prediction pipeline. We don't stop there; we extend this journey to the end-users with a well-crafted, user-friendly interface, bridging the gap between data and actionable insights. This document, a testament to our technical prowess, delves deep into the modular architecture, interfaces, algorithms, and visualizations that underpin this transformative solution, setting the stage for a future where anomalies and common patterns emerge as strategic assets in the world of retail..

1 Introduction

1.1 Why this High-Level Design Document?

The main purpose of this HLD documentation is to feature the required details of the project and supply the outline of the machine learning model and also the written code. This additionally provides the careful description on however the complete project has been designed end-to-end.

1.2 Description

Problem Perspective

Utilizing data warehousing techniques to analyze individual item sales data from shopping malls and Big Marts presents a promising prospect. This approach enables us to unearth valuable insights, including anomalies and recurrent patterns, within vast repositories of consumer information and product specifics. By harnessing advanced data mining and analytics, we can enhance demand forecasting accuracy and refine inventory management strategies, ultimately optimizing operations and customer satisfaction. This data-driven approach has the potential to revolutionize how retailers adapt to ever-evolving market dynamics and consumer preferences.

1.3 Problem Statement

Nowadays, shopping malls and Big Marts keep track of individual item sales data in order to forecast future client demand and adjust inventory management. In a data warehouse, these data stores hold a significant amount of consumer information and particular item details. By mining the data store from the data warehouse, more anomalies and common patterns can be discovered..

1.4. Project Solution

Developing a comprehensive solution for leveraging data warehousing in shopping malls and Big Marts is essential. First, we must establish robust data pipelines to ingest and store individual item sales data in the data warehouse.

This data repository should be designed for efficient retrieval and analysis, considering both consumer demographics and product attributes.

Next, employing advanced data mining and machine learning techniques, we can extract valuable insights from this data store. This includes identifying anomalies that might indicate theft or data entry errors and discovering common patterns that offer invaluable information for demand forecasting and inventory management.

1.5 Answer enhancements

Enhancing the solution for leveraging data warehousing in shopping malls and Big Marts involves ensuring data quality, scalability, predictive maintenance, personalization, real-time analytics, security, and compliance, fostering cross-functional collaboration, integrating external data sources, continuous monitoring, machine learning models, and customer feedback. By implementing these enhancements, the solution becomes more comprehensive, adaptable, and capable of addressing evolving challenges, enabling retailers to stay competitive and responsive to customer demands while optimizing operations and customer satisfaction through data-driven strategies.

1.6 Technical needs

There are not any hardware needs needed for victimization this application, the user should have an interactive device that has access to the web and should have the fundamental understanding of providing the input. And for the backend half the server should run all the package that's needed for the process and provided information to show the results.

1.7 Information needs

The info demand is totally supported the matter statement. and also, the information set is accessible on the Kaggle within the type of standout sheet(.xlsx), because the main theme of the project is to induce the expertise of real time issues, we have a tendency to once more mercantilism {the information into the prophetess data base and commerce it into csv format.

1.8 Tools Used

- Python 3.9 is employed because the programming language and frame works like numpy, pandas, sklearn,flask, streamlit and alternative modules for buildingthe model.
- Visual Studio Code is employed as IDE.

- Front end development is completed victimization HTML/CSS.

- Flask is employed for each information and backend readying.

- GitHub is employed for version management.

- Streamlit Cloud and Locallhost is used for Deployment

1.9 Constraints

Implementing a data warehousing solution for shopping malls and Big Marts faces constraints related to data privacy regulations, complex data integration from various sources, high infrastructure costs, cybersecurity risks, and the imperative need for data quality assurance. Adhering to privacy laws like GDPR and CCPA necessitates stringent compliance measures. Integrating data from diverse sources within stores requires careful planning. The substantial investment in hardware, software, and skilled personnel is a financial constraint. Ensuring data security amid growing cyber threats is vital, and maintaining data quality is paramount to avoid flawed insights and decisions.
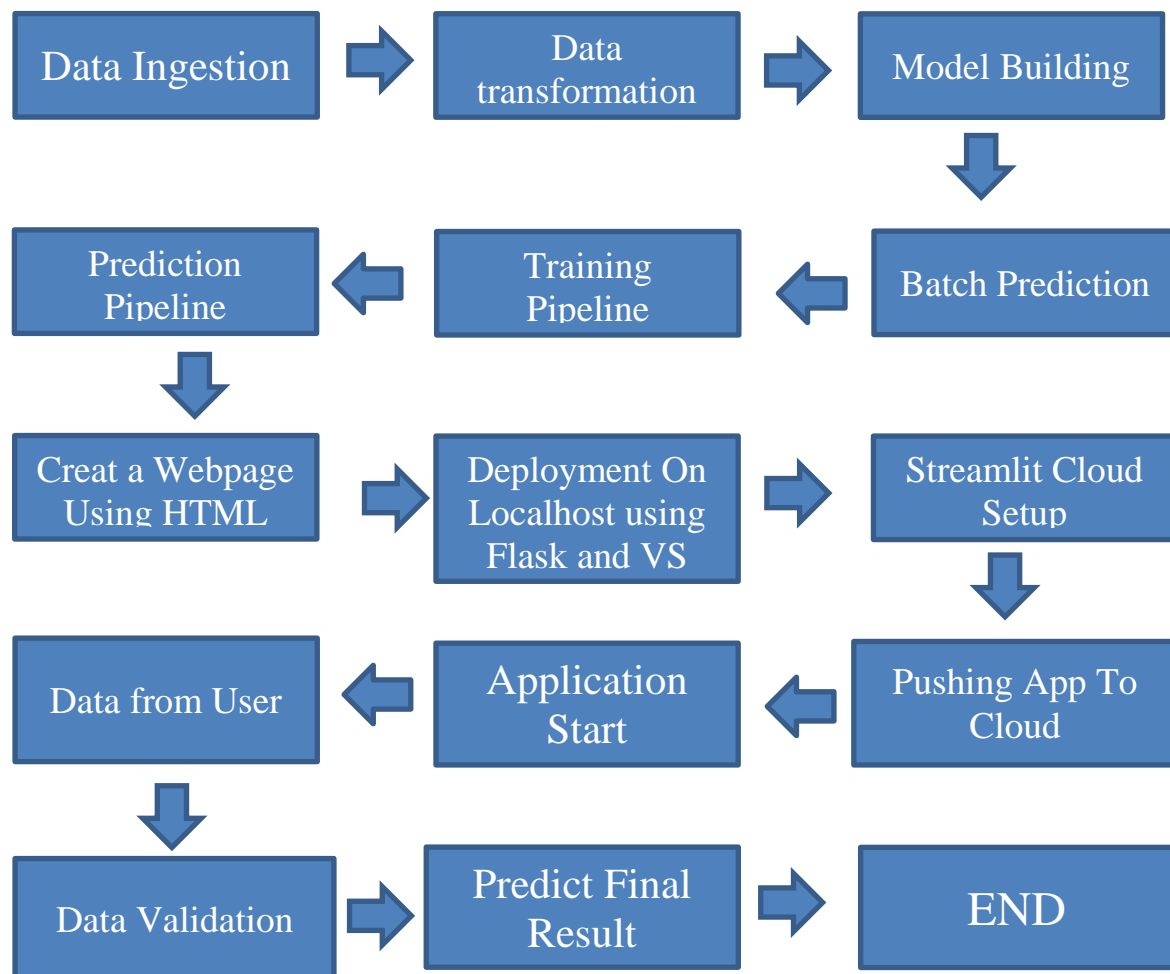
1.10 Assumptions

In developing a data warehousing solution for shopping malls and Big Marts, certain assumptions are made to guide the planning and implementation process. Assumptions may include the availability of data sources from various store operations, including point-of-sale systems and inventory management tools. It's assumed that data privacy regulations are adhered to and that customer data is handled in compliance with applicable laws. Additionally, there's an assumption that the necessary financial resources and skilled personnel are allocated for the construction and maintenance of the data warehousing infrastructure. Assumptions also encompass the expectation that data integration challenges can

be effectively addressed through careful planning and technical solutions, and

that ongoing investments in cybersecurity measures are made to protect sensitive customer information.

2.1 and 2.2 Design Flow and Deployment Process

Data Ingestion	→	Data transformation	→	Model Building

Prediction Pipeline	←	Training Pipeline	←	Batch Prediction

Creat a Webpage Using HTML	→	Deployment On Localhost using Flask and VS	→	Streamlit Cloud Setup

Data from User	←	Application Start	←	Pushing App To Cloud

Data Validation	→	Predict Final Result	→	END

2.3 Logging

Each step is being logged within the system that runs internally, that shows the date time and therefore the processed that has been performed, work is completed in several layers as information, DEBUG, ERROR, WARNINGS. this provides US the perceive of the logged info.

2.4 Error Handling

Once ever a slip is occurred, the reason are logged in its several log file, in order that the developer will rectify the error.

3 Performance analysis

3.1 Reusability

Elements of the code written is accustomed different applications and therefore the rest is changed and be reused.

3.2 Application Compatibility

The various parts for this project are exploitation python as associate interface between them. every element can have its own tasks to perform, and it's the work of the python to make sure correct transfer of data.

3.3 Resource Utilization

Once any task is performed, it'll doubtless; use all the process power offered till that performs is finished.

3.4 Deployment

The model is being deployed on local machine and Streamlit Cloud.

Conclusion

In conclusion, implementing a data warehousing solution in shopping malls and Big Marts holds immense potential for optimizing operations, enhancing customer experiences, and staying competitive in the dynamic retail landscape. By effectively harnessing data from various store operations and ensuring compliance with data privacy regulations, retailers can gain valuable insights into consumer behaviors and preferences. While constraints such as high infrastructure costs and cybersecurity risks are challenges to navigate, they can be mitigated with prudent planning and investment. With a commitment to data quality and ongoing adaptation to changing market dynamics, the future for these retail giants looks promising as they leverage data-driven strategies to meet customer demands and drive business success.