High Level Design (HLD)

Churn Analytics

Revision Number: 1.2

Last date of revision: 10/07/2024

Ayush Pathania

**Document Version Control**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Issued** | **Version** | **Description** | **Author** |
| 07/07/2024 | 1.0 | Abstract, Introduction,  General Description | Ayush Pathania |
| 08/07/2024 | 1.1 | Design Detail, KPI, Deployment | Ayush Pathania |
| 10/07/2024 | 1.2 | Final Revision | Ayush Pathania |
|  |  |  |  |

**Contents**

Abstract

Customer churn is a critical concern for transportation companies as it directly impacts revenue and operational efficiency. Churn analytics involves the process of identifying customers who are likely to discontinue using a transportation service and understanding the underlying reasons for their departure. This study focuses on the application of churn analytics within the transportation sector, leveraging data from various sources to provide actionable insights.

In this research, we utilize historical data from a transportation company, which includes customer feedback in the form of commendations and complaints, service details, and usage patterns. The dataset comprises key variables such as agency, subject matter, issue detail, year, quarter, and branch/line/route, enabling a comprehensive analysis of customer experiences and their correlation with churn rates.

Our methodology involves the application of data preprocessing techniques to clean and structure the data, followed by exploratory data analysis (EDA) to identify trends and patterns. We employ advanced statistical methods and machine learning models to predict customer churn, considering factors like service quality, customer service interactions, and operational performance.

The results of this analysis are visualized through an interactive Power BI dashboard, which highlights crucial metrics including total customers lost, percentage of customers lost, value of recurring business loss, and percentage of recurring value loss. These visualizations provide transportation companies with a clear understanding of churn dynamics and enable them to implement targeted strategies for customer retention.

By identifying the key drivers of churn and their impact on the business, transportation companies can enhance service quality, improve customer satisfaction, and ultimately reduce churn rates. This study underscores the importance of churn analytics as a tool for sustaining customer loyalty and ensuring long-term profitability in the transportation industry.

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

Problem Statement: Churn Model helps identifying customers who are most likely to switch to different eCommerce website. Once identified the companies can take actions in order to keep its existing customers. Now the question is, how does Churn model identify these customers?

The model can be used to calculate the churn rate and depending on the nature of business, different metrics can be used. Few common metrics are –

Number of customers lost

Percent of customers lost

Value of recurring business lost

Percent of recurring value lost

Project perspective: The objective of this project is to analyze customer feedback data for various transit agencies to gain insights into customer satisfaction, identify frequent issues, and assess the financial impact of recurring business loss. The analysis will help in improving service quality and customer retention by addressing key pain points.

2.2 Tools Used

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







3 Design Details

3.1 Functional Architecture



Figure - 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.
* 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
* KPIs

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the Amazon sales, 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.

* Identify which regions or countries are generating the most profit. This information can be used to allocate more resources to those regions or countries.
* Identify which products are most profitable. This information can be used to increase marketing for those products or to develop new products that are similar.
* Identify which sales channels are generating the most revenue. This information can be used to invest in those sales channels or to develop new sales channels.
* Identify trends in the business, such as seasonal fluctuations in sales or changes in customer behavior. This information can be used to make predictions about future performance and to make adjustments to the business plan.
* Deployment

Prioritizing data and analytics are more important than ever. Every company, regardless of size, is already collecting data and is likely only analyzing a fraction 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 and flexibility to fit, rather than dictate, your enterprise architecture. Power BI Desktop, Power BI Premium, and Power BI Cloud Services leverage your existing technology investments and integrate into your IT infrastructure to provide a self-service, modern analytics platform for your users. With on-premises, cloud, and hybrid options, there is a version of Power BI to match your requirements. Below is a comparison of the three types:

* **Power BI Desktop** is a good option. If you need a more scalable and secure solution, Power BI Premium or Power BI Cloud Services may be a better choice.



* **Power BI Premium** is a paid version of Power BI that offers more features and capacity. It is a good option for larger businesses that need to scale their analytics workloads or who need to share reports with a large number of users.



* **Power BI Cloud Services** is a fully managed cloud-based version of Power BI. It is a good option for businesses that want to avoid the hassle of managing their own Power BI infrastructure.



Which version of Power BI is right for you depends on your specific needs and requirements. If you are looking for a self-service, modern analytics platform that is easy to deploy and use, Power BI Desktop is a good option. If you need a more scalable and secure solution, Power BI Premium or Power BI Cloud Services may be a better choice.