HIGH LEVEL DESIGN

CUSTOMER SEGMENTATION

Revision Number: 2.0
Last date of revision: 28/09/2021

Atufa Shireen

Document Version Control

Date Issued	Version	Description	Author
28th Sep 2021	1.0	First Version of Complete HLD	Atufa Shireen
29th Sep 2021	1.1	Added Architectures	Atufa Shireen
5th Oct 2021	2.0	Second Version of Complete HLD	Atufa Shireen

Table of Contents

Abstract

As per pareto principle 80% of the revenue is derived from the 20% of the customers, and hence retaining the available customers is less expensive than onboarding new customers.
The customer analysis and segmentation can help in predicting the future transactions of the customers, their behaviour towards business products, relationship strength with the business, customer retention and churn rate, which can also help in targeted marketing.

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.

1.3 Definitions

TERM	DESCRIPTION
Customer Lifetime Value	The expected monetary value of a customer for a period of time.

RFM Model	Recency, frequency, monetary value (RFM) is a marketing analysis tool used to identify a firm's best clients based on the nature of their spending habits.
Recency	How recently a customer has made a purchase.
Frequency	How often a customer makes a purchase
Monetary Value	How much money a customer spends on purchases
Segmentation/Clustering	Dividing a company’s customers into groups that reflect similarity to decide how to relate customers in each segment in order to maximize the value of each customer to the business.
Database	Collection of all information monitored by this system
IDE	Integrated Development System
AWS	Amazon Web Services

2. General Description

2.1 Product Perspective

The project will allow business to understand its customer relationship strength, and create targeted marketing.

2.2 Problem Statement

To create a pipeline for customer segmentation.

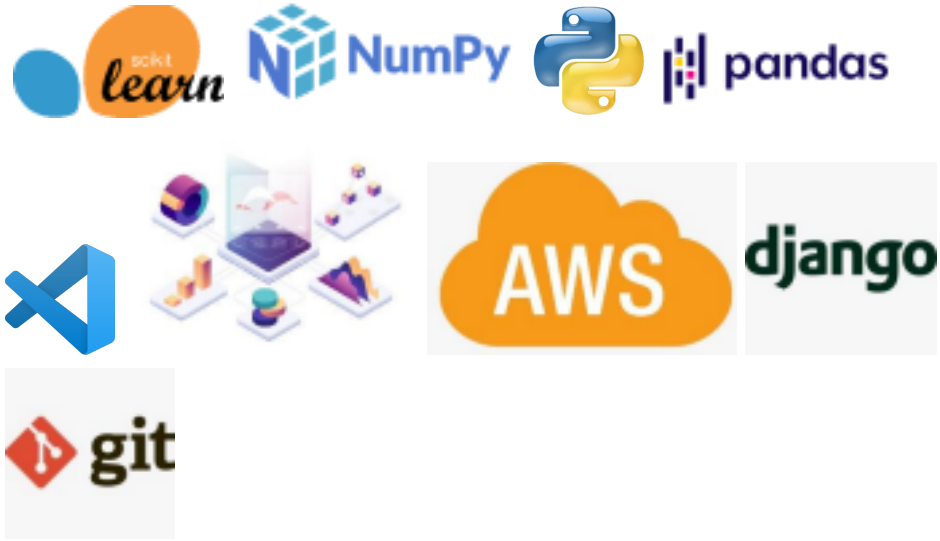
2.3 Proposed Solution

- Create an analysis report on relations between the features.
- Segment customers based on RFM model.
- Segment customers based on Cluster analysis.
- Calculate Lifetime Value of customers.
- Expected Number Of Transactions.
- Churn Rate.
-

3. Technical Requirements

3.1 Tools Used

Python programming language, libraries and frameworks such as Numpy, Pandas, Scikit-learn, lifetimes, matplotlib, seaborn, plotly, Django are used to build the whole model.



- PyCharm is used as an IDE.
- For visualization of the plots, Matplotlib, Seaborn and Plotly are used.
- AWS is used for deployment of the model.
- Tableau/Power BI is used for dashboard creation.
- SQL/MongoDB is used to retrieve, insert, delete, and update the database.
- Front end development is done using HTML/CSS/JS.
- Python Django is used for backend development.
- GitHub is used as a version control system.

3.2 Constraints

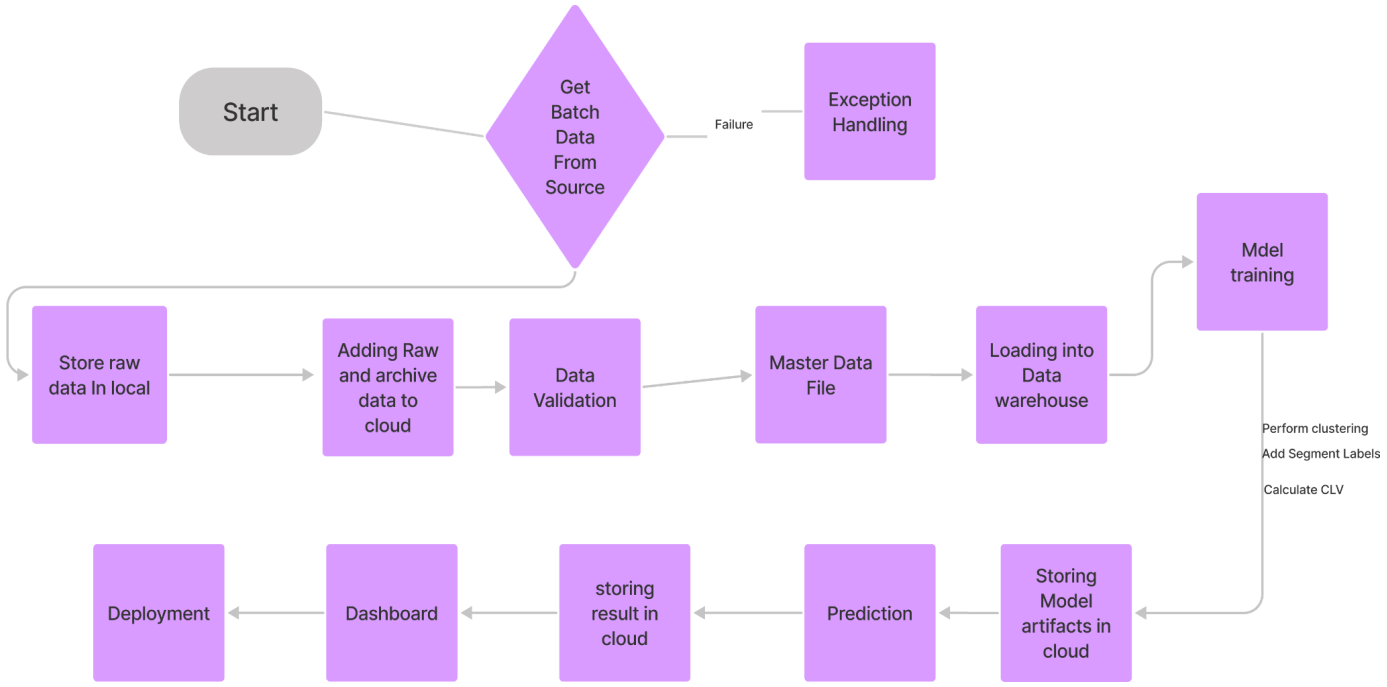
The Customer Segmentation system must be user friendly, as automated as possible and users should not be required to know any of the workings.

3.3 Assumptions

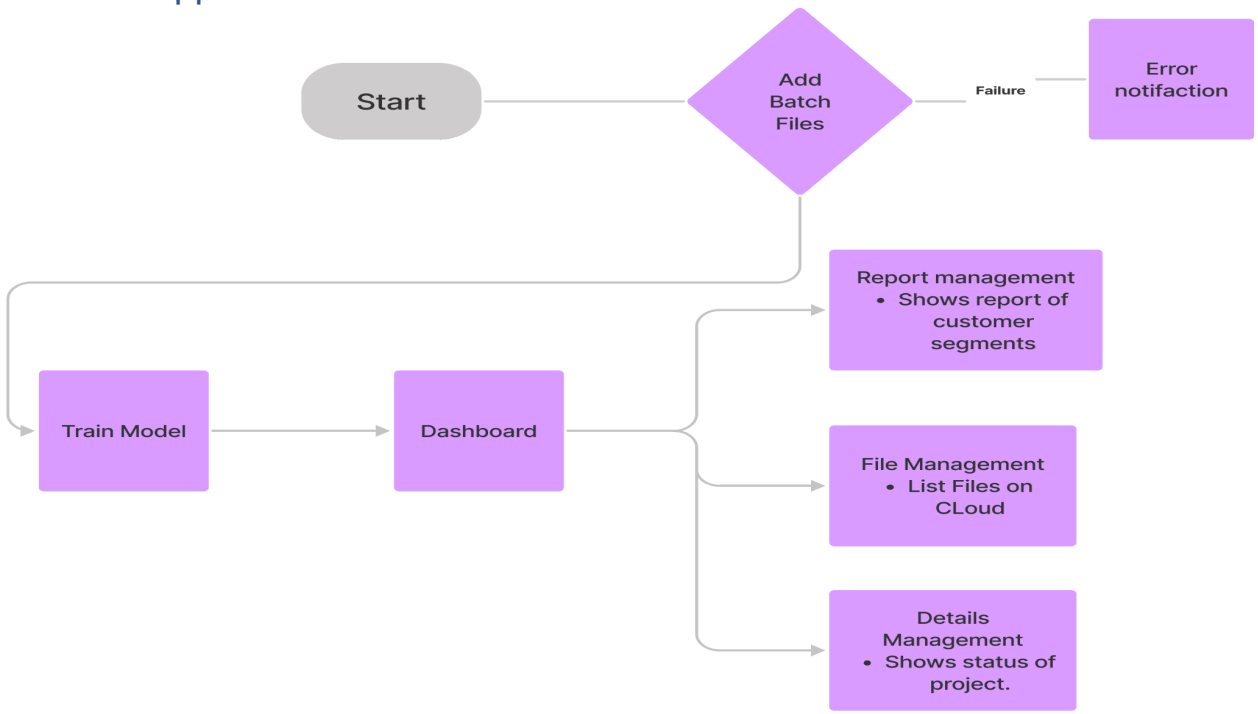
The main objective of the project is to classify the customers into segments based on the information provided by the user/business, by using Machine Learning techniques. It is also assumed that all aspects of this project have the ability to work together in the way the designer is expecting.

4. Design Details

4.1 Functional Architecture



4.2 Web Application Architecture



4.3 Database Design

System needs to store every request into the database and you need to store it in such a way that if you want to retrain a model it should be easy to retrain the model with new data as well.

Initial Step-By-Step Description:

1. The User adds the data of customers (to be segmented).
2. The User gives required information.
3. The system stores each and every data and output given by the user or received on request to the database.
4. Final data is loaded into warehouse as part of etl pipeline

4.4 Event log

The system should log every event so that the user will know what process is running internally.

Initial Step-By-Step Description:

1. The System identifies at what step logging is required.
2. The System should be able to log each and every system flow.
3. Developers can choose logging methods. You can choose database logging/ File logging as well.
4. System should not hang even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

4.5 Error Handling

Should errors be encountered, an explanation will be displayed as to what went wrong? An error will be defined as anything that falls outside the normal and intended usage.

4.6 Help

‘The ‘Help’ option is provided in the web application for guiding users regarding the maximum range of valid inputs required for segmenting.

4.7 Performance

Customer segmentation system is used for segmenting customers, it should be as accurate as possible. So that it will not mislead the user. Also, model retraining is very important to improve the performance.

4.8 Security

Since the Customer Segmentation system consists of businesses personal information, the information should be secured.

4.9 Reusability

The code written and the components used should have the ability to be reused with no problems.

4.10. Application compatibility

The different components for this project will be using Python as an interface between them. Each component will have its own task to perform, and it is the job of Python to ensure the proper transfer of information.

4.11. Resource utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

5. Deployment

6. Dashboards

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the segments.

7. KPIs (Key Performance Indicators)

Key indicators display a summary of the person's health and physique as compared to a normal individual with similar basic traits.

1. Number of times the customer has made purchases.
2. Number of times the customer has visited the business.
3. Date of joining and leaving of customers from the business.
4. Revenue or sale derived from customers.
5. Churn rate of customers.
6. Customer Lifetime Value.

8. References

1. RFM ANALYSIS : Clever Tap
2. CUSTOMER SEGMENTATION: Business Science
3. Automating Customer Segmentation: Habr
4. Case Study:How much a customer worth: Etsy
5. BGNBD Model: Papers
6. Lifetimes Package: Lifetimes
7. Calculating CLV: Stephan Oats

9. Conclusions

The segmentations group customers with similar RFM behaviour(i.e, A group can focus more on Monetary value or recency of customers) but are distinct from each other, which can help in understanding the customer's relation and calculate their Lifetime value and number of future transactions over a period of time within business.