HIGH LEVEL DESIGN (H.L.D)

Customer Segmentation and Clustering

iNeuron

Internship Project

# Document Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Date Issue | Version | Description | Author |
| 8-10-2024 | 1 | Initial HLD – V1.0 | Gaurav Ghosh |

Contents

[Document Version Control 1](#_Toc188530634)

[ABSTRACT 3](#_Toc188530635)

[1 Introduction 4](#_Toc188530636)

[1.1 Why this High-Level-Design Document? 4](#_Toc188530637)

[1.2 Scope 4](#_Toc188530638)

[1.3 Definitions 4](#_Toc188530639)

[2 General Description 5](#_Toc188530640)

[2.1 Product Perspective 5](#_Toc188530641)

[2.2 Problem statement 5](#_Toc188530642)

[2.3 Proposed Solution 5](#_Toc188530643)

[2.4 Further Improvement 5](#_Toc188530644)

[2.5 Data Requirement 5](#_Toc188530645)

[2.6 Tools Used 6](#_Toc188530646)

[2.7 Constraints 6](#_Toc188530647)

[2.8 Assumptions 6](#_Toc188530648)

[3 Design Details 7](#_Toc188530649)

[3.1 Process Flow 7](#_Toc188530650)

[3.1.1 Model Training and Evaluation 7](#_Toc188530651)

[3.1.1 Deployment Process 7](#_Toc188530652)

[3.2 Event Log 8](#_Toc188530653)

[3.3 Error Handling 8](#_Toc188530654)

[4 Performance 9](#_Toc188530655)

[4.1 Reusability 9](#_Toc188530656)

[4.2 Application of Compatibility 9](#_Toc188530657)

[4.3 Resource Utilization 9](#_Toc188530658)

[4.4 Deployment 9](#_Toc188530659)

[5 Conclusion 10](#_Toc188530660)

# ABSTRACT

In today’s fast-paced and competitive market, understanding customer behaviour is crucial for businesses to stay ahead. Consumer preferences vary significantly across different age groups, income levels, and spending habits, creating diverse market segments. To address these variations effectively, businesses require a structured approach to analyse and categorize their customers.

This project leverages **unsupervised learning techniques** to cluster customers based on the key attributes: **MonetaryValue, Frequency** and **Recency.** By grouping customers into meaningful clusters, the project aims to provide actionable insights that can help businesses tailor their marketing strategies, product offerings, and services to better meet the unique needs of each segment.

This data-driven approach not only enhances customer understanding but also aids in resource optimization and improving overall customer satisfaction in a dynamic marketplace.

# 1 Introduction

## 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.

This HLD will:

* Describe the hardware and software interfaces.
* Describe the performance requirement.
* 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** | **Discription** |
| CSV | Comma\_Seperated Values |
| Database | Collection of data in a spreadsheet |
| IDE | Integrated Development Environment |
|  |  |

# General Description

## 2.1 Product Perspective

The Customer Segmentation and Clustering solution is a machine learning project designed to analyse market trends and customer behaviour using available datasets, enabling businesses to derive optimal strategies and solutions.

## 2.2 Problem statement

To create a ML solution for understanding market using available dataset and implement in following ways:

* Identify distinct customer segments.
* Tailor marketing strategies
* Optimize resource allocation

## 2.3 Proposed Solution

The proposed solution involves **Customer Segmentation and Clustering** to analyse the market and address the outlined problem. The model will begin by training on the provided CSV dataset, examining the various columns to identify patterns and relationships.

It will analyse attributes to uncover meaningful insights. By leveraging these patterns, the model will determine the optimal segmentation, grouping customers into clusters based on their age, income, and spending behaviour. This clustering approach ensures a tailored understanding of customer segments, enabling businesses to strategize effectively.

## 2.4 Further Improvement

The model can be further enhanced to analyse various types of market datasets, enabling it to derive optimal solutions and address a wide range of market dynamics effectively.

## 2.5 Data Requirement

Data Requirement completely depends on our problem statement.

* We need dataset that has the following.
  + StockCode
  + InvoiceData
  + Price
  + Quantity
* The dataset contains a minimum of 1,000 rows to ensure a comprehensive and accurate analysis by the model.
* Different types of graphical representation is used:
  + Elbow Graph
  + Silhouette Score
  + Scatter plot matrix (2D & 3D)

## 

## 2.6 Tools Used

* Python Programming language
* Python libraries:
  + Numpy
  + Pandas
  + Sklearn
    - K-means
    - silhouette\_score
  + Matplotlib
  + os
* PyCharm is used as IDE.
* For visualisation of the plot matplotlib.pyplot and mpl\_toolkits.mplot3d (Axes3D).
* Heroku is used for deployment.
* CSV is used as dataset.
* GitHub is used as version control system.

## 2.7 Constraints

The Customer Segmentation and Clustering model produces outputs that are both user-friendly and accurate, making them easy to interpret and apply effectively.

## 2.8 Assumptions

* Clustering data requires a large dataset.
* Customer segmentation is only used for marketing.
* Segmentation results are static.

# Design Details

## 3.1 Process Flow

### 3.1.1 Model Training and Evaluation

### 3.1.1 Deployment Process

## 3.2 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 required

2. The System should be able to log each and every system flow.

3. Developer can choose logging method. 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.

## 3.3 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.

# Performance

The Customer Segmentation and Clustering model is designed to analyse customers based on their StockCode, Quantity, InvoiceDate and Price. This enables businesses to identify optimal solutions that accurately address customer needs and preferences based on these attributes.

## 4.1 Reusability

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

## 4.2 Application of 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 the Python to ensure proper transfer of information.

## 4.3 Resource Utilization

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

## 4.4 Deployment

* Github

# Conclusion

In conclusion, the Customer Segmentation and Clustering model provides a powerful tool for understanding and categorizing customers based on key attributes such as gender, age, income, and spending score. By leveraging machine learning techniques, this approach delivers actionable insights that help businesses tailor their strategies to meet the diverse needs of their customer base. With accurate and user-friendly outputs, the model ensures effective decision-making, enabling companies to optimize their offerings and maintain a competitive edge in a dynamic market.