TASK 02 – AI APPLICATIONS IN BUSINESS PROBLEM SOLVING

APPLICATION OF AI TECHNIQUES AND TOOLS IN CUSTOMER SEGMENTATION FOR A RETAILER

INTRODUCTION

The retail industry nowadays is driven by the constantly changing customer behaviour, growing trend for customization and competition (Desai, 2025). The landscape is dominated by the virtual shopping experience and omnichannel retail strategies which require the collection of large volumes of data across multiple channels (Desai, 2025). Customer data is collected from e-commerce platforms, POS systems, Loyalty Programmes and digital transactions/interactions.

However, comprehending your customer in its core is a struggle for most retailers despite the volume of data that is available. Generally, most retailers follow a typical marketing strategy designed to treat all the customers as one. This path leads to many missed opportunities to address brand loyalty as well us low engagement rates and unnecessary marketing expenditure.

Customer Segmentation provides the solution for this concern. This marketing strategy allows retailers to segment their customer base into groups based on common characteristics such as demographics, purchasing behaviour and engagement patterns, ensuring the ability to tailor the marketing campaigns, optimize the product offerings and deliver a personalized customer experience (Desai, 2025).

In order to reveal hidden patterns in customer data and automate segmentation, AI systems, specifically unsupervised machine learning can be utilized. Customers can be dynamically grouped now using algorithms such as K-Means clustering which associates similar patterns in behaviour and creates detailed customer profiles with insights that drive targeted business decisions, improves marketing and sales strategies and allows retailers to provide customized experiences (Desai, 2025).

DESIGN OF THE AI SOLUTION

CHOICE OF ALGORITHM

The desired outcome of customer segmentation is to recognize natural clusters without any pre-labelled categories in customer data. There are a number of approaches that can be used such as Dimensionality reduction + clustering, model-based segmentation, RFM + clustering but unsupervised learning is the best method.

Careful scrutinization of DBSCAN, Heirachial Clustering and K-Means clustering algorithms helped determine that latter would be ideal (Tabianan, Velu & Ravi, 2022).

This method allows retailers

- to compute very high volumes of customer data efficiently,
- define easy to interpret cluster centroids
- and makes it adaptable to a variety of business needs by adjusting the number of clusters.

This algorithm provides the ideal ration between practicality, scalability and interpretability for retailers (Tabianan, Velu & Ravi, 2022).

DATA REQUIREMENTS

Multi-dimensional data will be required by the retailer for effective segmentation. Transaction, engagement and demographic data to be considered (Tabianan, Velu & Ravi, 2022).

TRANSACTION DATA

a. Frequency of transactions over a period of time (No. of Purchases)

b. When was the last transaction (Recency)

c. Total amount spent by the customer (Monetary value)

d. Favoured products and the categories

DEMOGRAPHIC DATA

a. Geographic region (urban, suburban, rural)

b. Age, gender, income level and job category

ENGAGEMENT DATA

    a. Online data and history (Behaviour such as browsing history, website clicks and abandoned carts)

The data sources will be as follows:

1. From internal systems: loyalty points, POS systems and CRM databases
2. From digital platforms: Mobile Applications and e-commerce websites
3. From external sources: census information and 03rd party demographic providers

DATA PREPROCESSING

This process refines the data making it suitable for analysis, ML models and other applications (Bala & Behal, 2024).

DATA CLEANING:

- Imputing Values or remove incomplete records by sorting out missing demographic data.
- Clearing out duplicate or incorrect transactions

FEATURE ENGINEERING

- The three RFM Metrics:
  - Recency: When the last purchase was made?
  - Frequency: Frequency of transactions over a period of time (No. of Purchases)
  - Monetary: Total amount spent by the customer
- Added-value features:
  - Number of purchases made during promotions – to identify customers sensitive to discounts
  - Average basket size (Optional)
  - Preferred channel (Optional)

STANDARDIZATION/NORMALIZATION

- This is to ensure all features contribute equally to cluster formation as K-Means uses Euclidean distance that may result in biased results when it comes to assigning points in clusters where variables are on different scales.

DIMENSIONALITY REDUCTION (OPTIONAL)

- Principal Component Analysis (PCA) can be used to improve cluster separation, highlight variable features in clusters and reduce noise

TOOLS AND PLATFORMS

Implementation of the AI Solution will involve:

- Using Python as the main programming language

LIBRARIES

- *pandas* to handle missing values, delete duplicates, manging datasets (transaction & demographic), create new features (such as discount sensitivity), calculate RFM Metrics
- During preprocessing, *numpy* can be used for efficient mathematical operations.
- *Scikit-learn* – Preparing features on comparable scales, implementation of the alogorithm for clustering and evaluating results.
- *Seaborn, matplotlib* – profiling segments, PCA-reduced cluster visualizations, plotting distributions.

CLOUD AI SERVICES

Google Vertex AI, Azure ML or AWS Sage Maker.

Models have to be trained at scale for large data volumes, for real-time segmentation: automate the deployment pipelines, and as new data is made available clusters have to be updated dynamically and model performance have to be monitored.

DATABASES

For data storage, fast retrieval for preprocess and model input and supporting integration between AI solution and the retailer system – use SQL databases or a cloud data warehouses such as BigQuery.

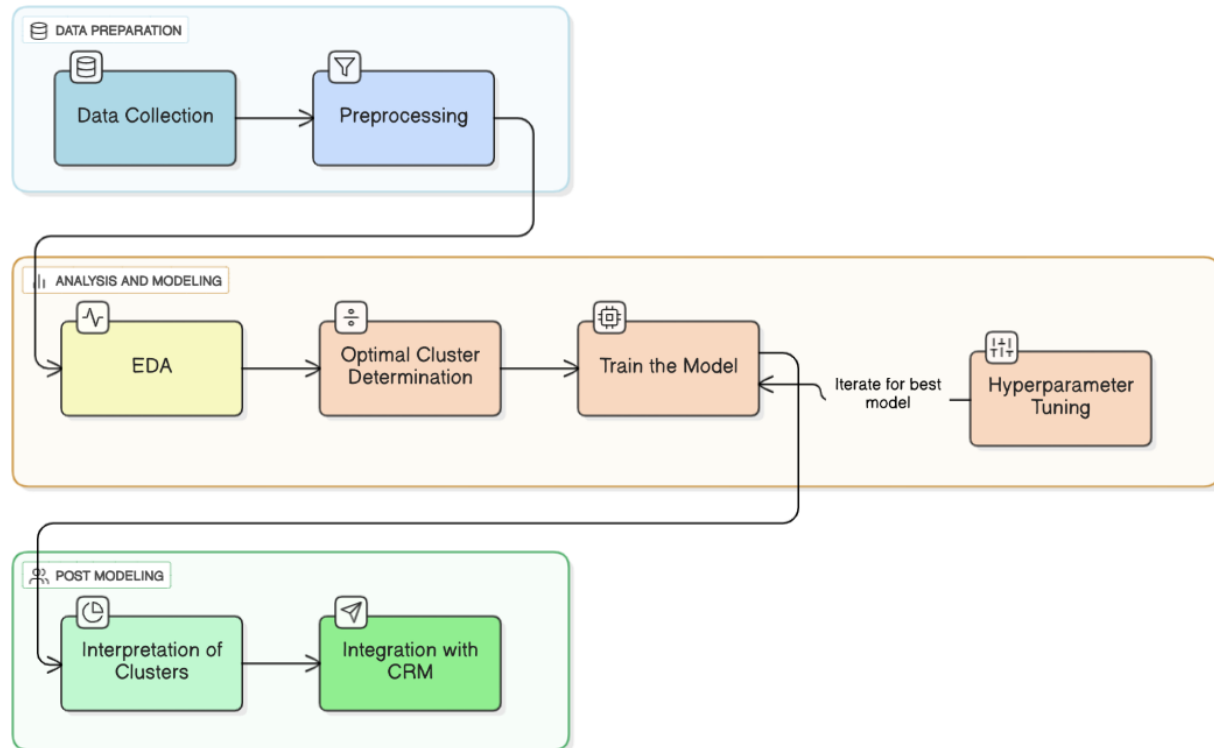## THE IMPLEMENTATION (CONCEPTUAL)

Figure 1: Flowchart of customer onboarding process.
Created by author using ERASER.io, 2025.

DATA COLLECTION

From POS, e-commerce platforms, loyalty programs, CRM

PREPROCESSING

Cleaning and normalizing data. For e,g. using RFM maetrics

EXPLORATORY DATA ANALYSIS (EDA)

Using Histograms (visualize income distribution), Heatmaps (analyse relationships between demographic and behavioural features), Scatterplots (identify natural separations in data)

OPTIMAL CLUSTER (K) DETERMINATION

Elbow Method – Graphical representation within-cluster sum of squares (WCSS) vs. number of clusters

Silhouette Score – Measurement of how similar a data point is to its own cluster Vs how different it to that of other clusters. Example: This analysis may suggest 10 optimal clusters

TRAIN THE MODEL

Use K=10 and Apply K-Means clustering and to ensure stable cluster assignments; run multiple initializations.

INTERPRETATION OF CLUSTERS INTO MEANINGFUL SEGMENTS:

Characteristics of the Clusters	Label
First time / New Customers: Highly Recent	New Customer
Purchases driven by discounts; Low frequency	Bargain Hunters
Attracted by promotions; Moderate expenditure	Promotion-sensitive Customers
Mid-level income; Average Frequency	General Customers
Premium preferences, High Expenditure; High Frequency	High-Value Loyal Customers

HYPERPARAMETER TUNING

Change the number of clusters, reassign the weightage of features based on retailer's requirements

INTEGRATION

Export and use in a CRM for loyalty rewards, personalized marketing campaigns and product recommendations.

IMPACT ON BUSINESS & EVALUATION

METRICS FOR EVALUATION

The evaluation relies on internal and external validation as there are no labelled outcomes in unsupervised learning:

INTERNAL METRICS

a. Davies-Bouldin Index (Higher Values – Mediocre Clustering)

    b. Silhouette Score (Better Separation = Closer to 1)

    c. How well the yielded clusters can be useful in the real world – Business Interpretability

    d. Are they practically useful? A/B testing to compare between a segmented and un-segmented group.

EXPECTED IMPACT ON BUSINESS

1. Better Purchase Recommendations – cross selling based on preferences of a cluster
2. Personalized Marketing Campaigns:
    i. Eg: Exclusive rewards, Early access to new launches – High Value, Loyal Customers
    ii. Discounted Campaigns – Bargain hunters
    iii. Welcome Discounts and offers – New Shoppers
3. Retaining Customers – Trigger campaigns to reactivate customers with lower frequency and higher recency.
4. Allocate marketing budget on personalized campaigns for the segments instead of blanket promotions.
5. Higher Profitability and Long-term loyalty sue to increased customer satisfaction, repetitive purchases and value.

Figure 2: Expected Impact on Business
Created by author, 2025.

CONCLUSION

The modern retail industry has transformed into a highly competitive environment, and retailers can get a competitive edge over their peers by using AI for customer segmentation. AI technologies namely unsupervised machine learning such as K-Means clustering, which allows for moving beyond traditional demographic segmentation and reveal nuanced insights of a customer's buying behaviour.

Using AI solutions ensures retailers are able to analyse large data sets rapidly to ensure once implemented, it is able to benefit the retailers. This solution needs quality data that has to be carefully pre-processed and undergo robust evaluation and it is able to yield stronger customer relationships through improved retention, optimized spending and personalized marketing campaigns.

Integrating AI powered customer segmentation into day-to-day activities of the retailer will enhance operational efficiency and more importantly foster loyal long-term customers and profitability. This competitive advantage will set retailers higher up among their peers in the modern data-driven and customer focused retail environment.