CUSTOMER SEGMENTATION USING ML

PROBLEM DEFINITION

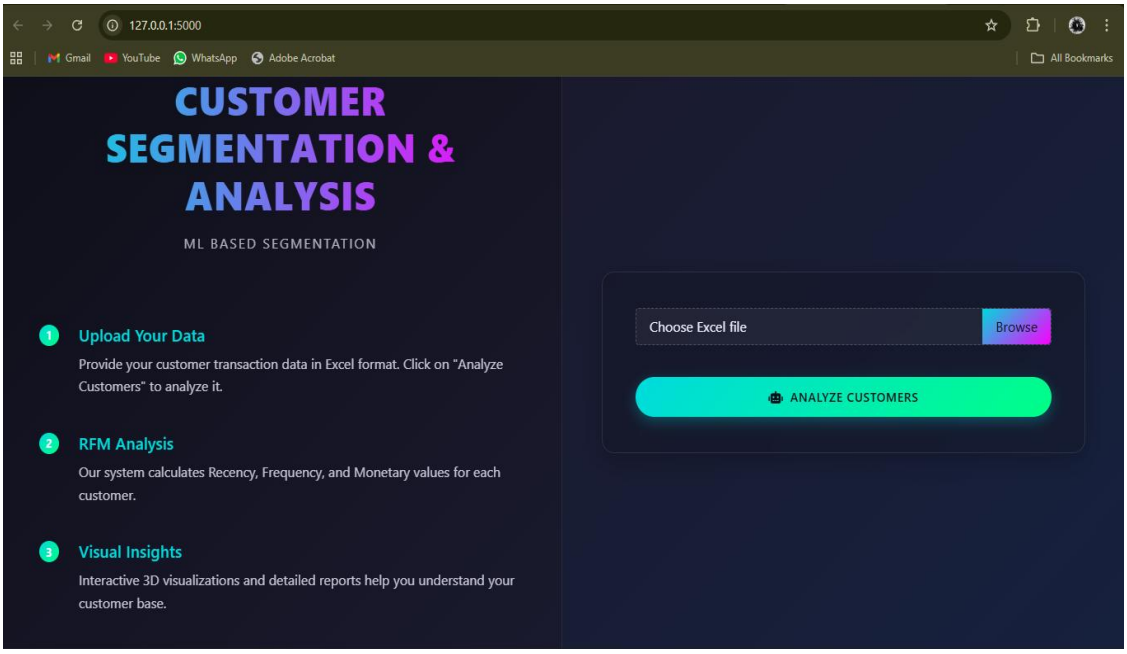
Businesses face challenges in effectively understanding and targeting their diverse customer base due to the vast amount of unstructured customer data they collect. Without a systematic approach to segment customers based on key characteristics, behaviors and marketing leads to reduced customer satisfaction and engagement. Traditional segmentation methods are time-consuming, manual, and lack scalability, making them insufficient for modern business needs. This project addresses the need for an automated, data-driven segmentation solution using Machine Learning algorithms like k-means clustering, allowing businesses to analyze large customer datasets and generate well-defined customer segments. By grouping customers with similar attributes, this approach enables personalized marketing, improves customer retention, and supports data-driven decision-making.

PROPOSED SYSTEM

The Proposed system for the “Customer Segmentation Using ML” uses RFM (Recency, Frequency, Monetary) analysis to measure customer behavior in quantifying three important metrics: how recent the last purchase of customers was (Recency), how frequently they buy (Frequency), and how much they buy (Monetary). This effective marketing tool segments customers according to their buying behavior, allowing companies to label their customers for tailored engagement initiatives. For segmentation, the system uses K-means clustering with the number of clusters optimized using the elbow method. The elbow method examines the within-cluster sum of squares (WCSS) for varying numbers of clusters and chooses the point at which further clusters yield decreasing returns. K-means then clusters customers with similar RFM attributes into separate segments that expose natural patterns in buying behavior. The system is built with Python's data science stack (Pandas for data processing, Scikit-learn for machine learning, and Matplotlib/Plotly for visualization), the system converts raw transaction data into actionable insights via an easy-to-use interface. This pipeline provides automated segmentation with interactive visualizations, enabling businesses to take data-driven marketing decisions and optimize customer relationship management strategies. The integration of RFM analysis and solid clustering techniques offers a scalable solution for analyzing customer value over various datasets.

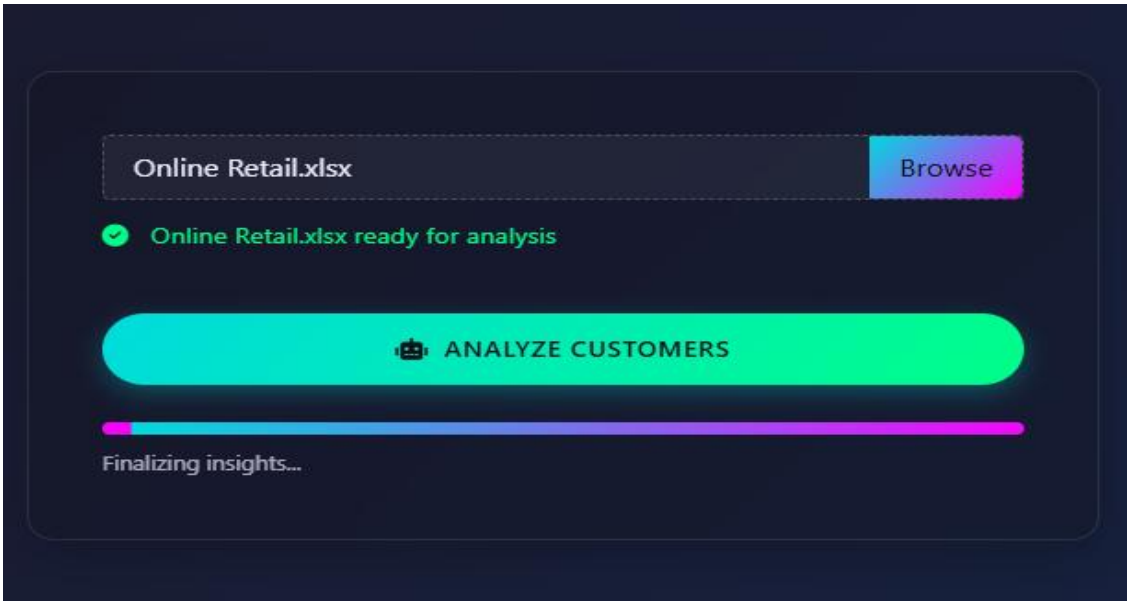
OUTPUT OF THE PROJECT

1. Homepage of Customer Segmentation system



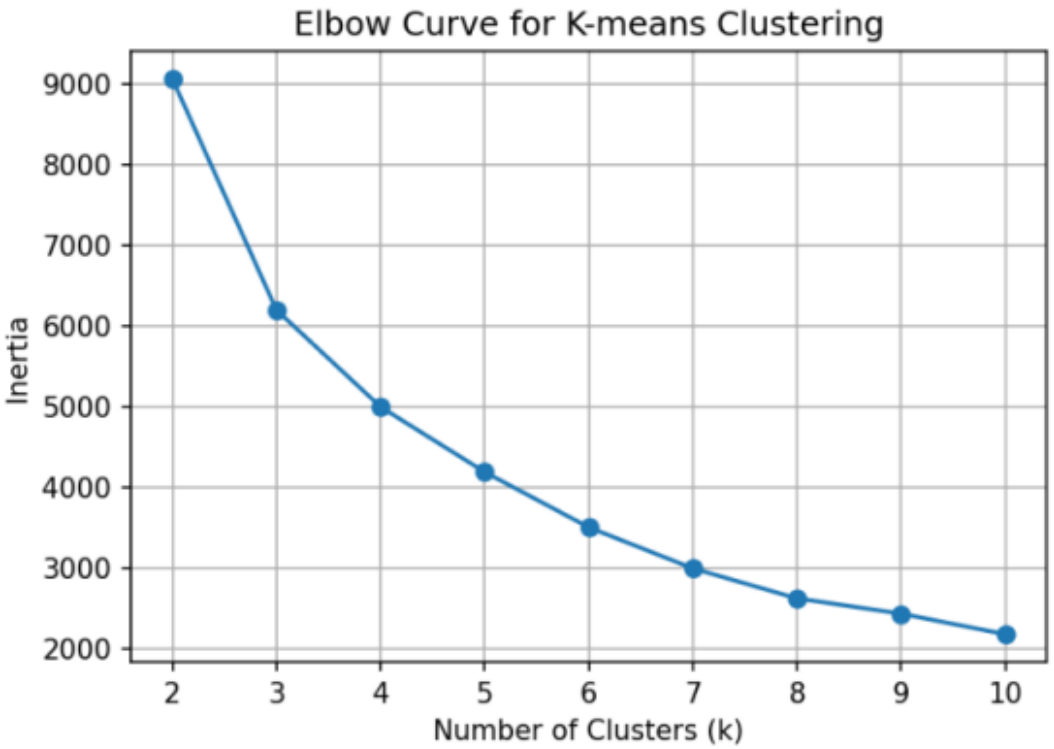
The User Interface or Homepage of Customer Segmentation system consists of a workflow of what you can do on the system and a simple section to upload files for segmentation. The workflow contains three simple steps of uploading the data, RFM analysis and visualizing the results.

2. Uploading the files for Segmentation



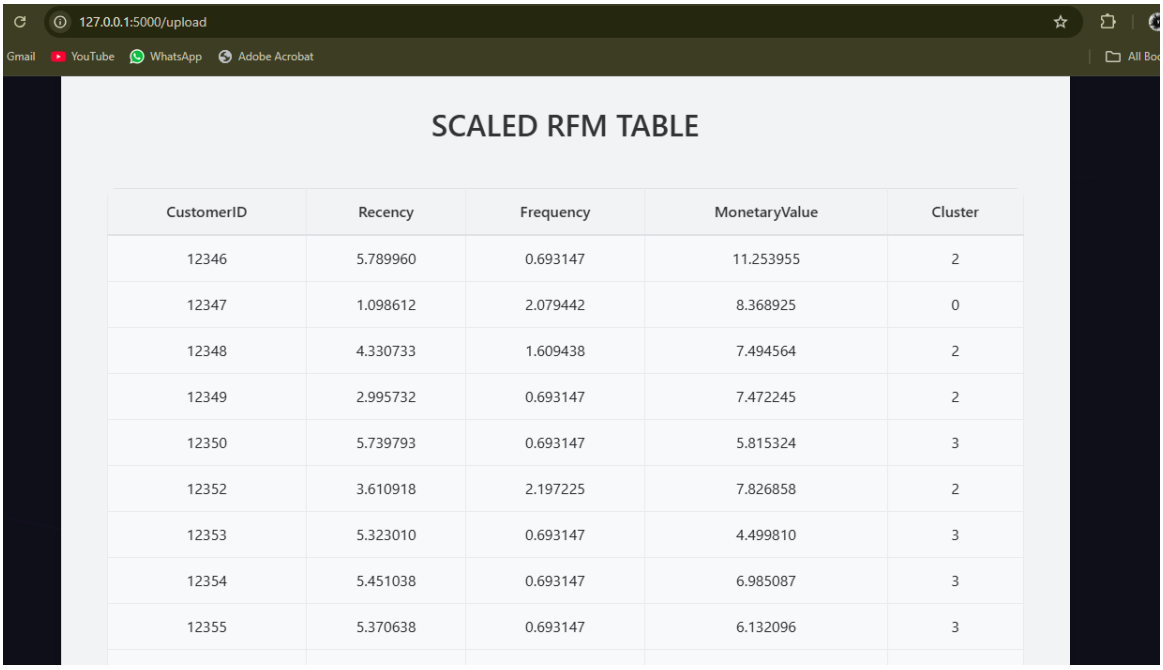
The upload section enables the users to upload the files for the further step of segmentation. Here, the upload section ONLY takes the files in XLSX format. It doesn't allow to input any other format of files. Once the file is ready we can click on ANALYZE CUSTOMERS button to start the segmentation process.

3. Checking the Optimum number of Clusters



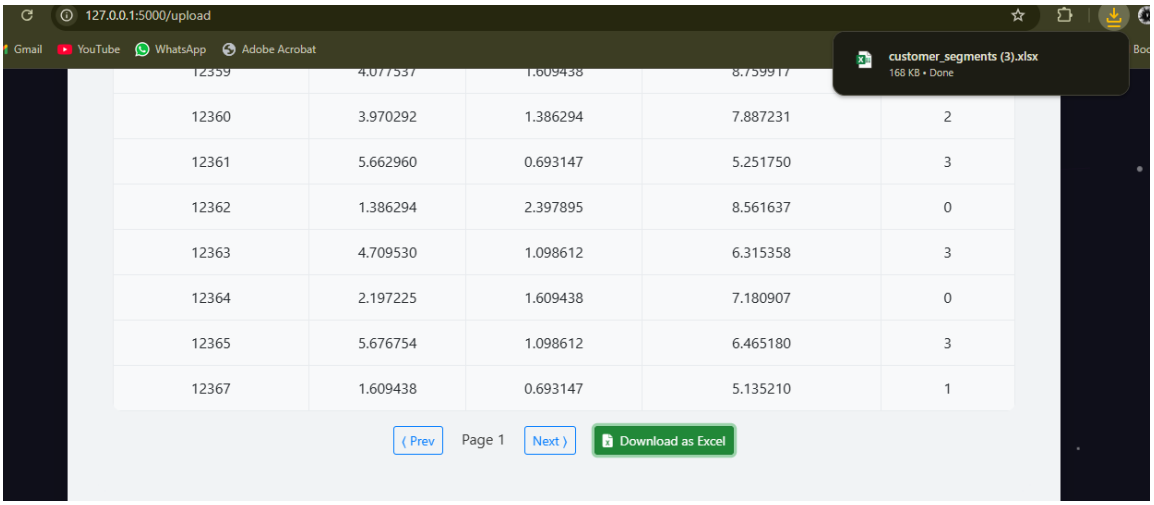
This image shows how to decide how many number of clusters to be formed which is based on the dataset. In this case, we use a special function called Elbow Method which in terms uses WCSS(within cluster sum of squares) to find out the optimum number of clusters value. In this case, it has shown that clusters will be best in form when number of clusters is 4.

4. Scaled RFM table

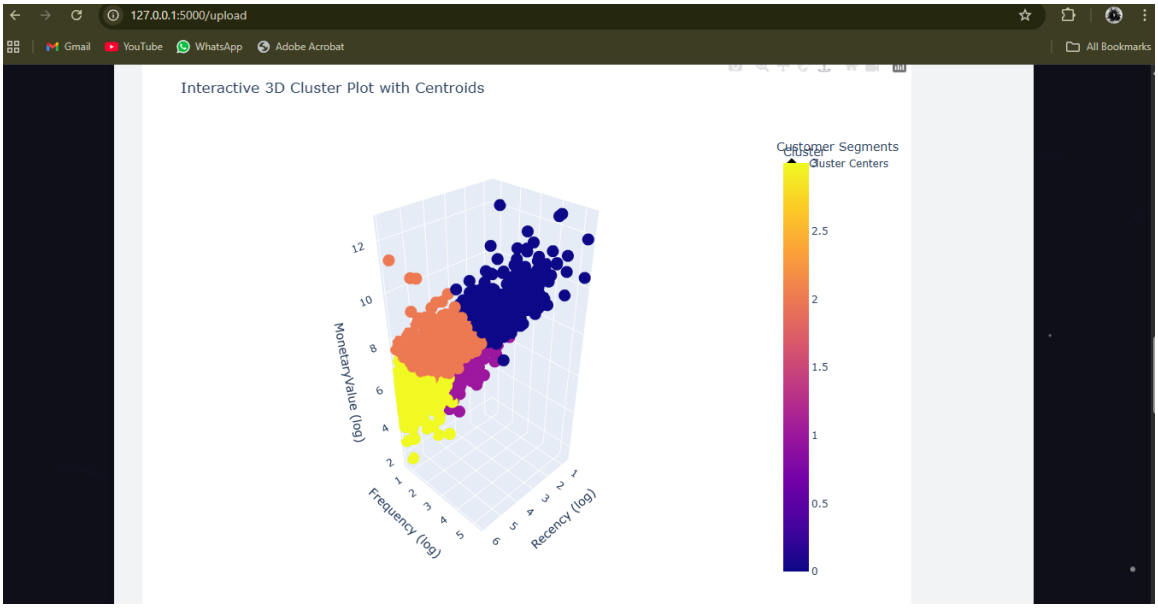


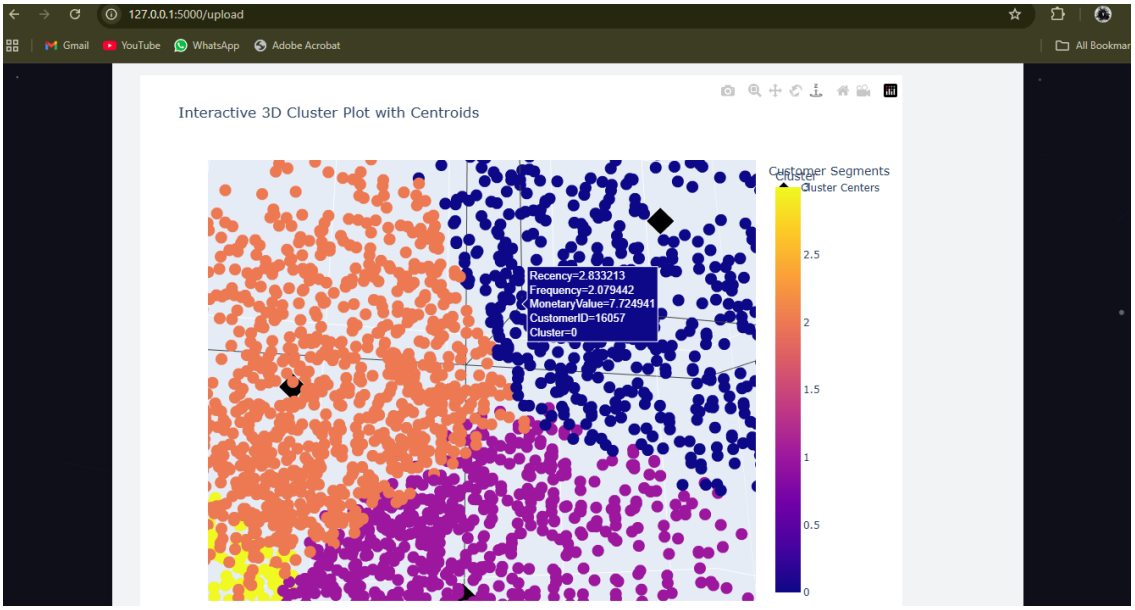
CustomerID	Recency	Frequency	MonetaryValue	Cluster
12346	5.789960	0.693147	11.253955	2
12347	1.098612	2.079442	8.368925	0
12348	4.330733	1.609438	7.494564	2
12349	2.995732	0.693147	7.472245	2
12350	5.739793	0.693147	5.815324	3
12352	3.610918	2.197225	7.826858	2
12353	5.323010	0.693147	4.499810	3
12354	5.451038	0.693147	6.985087	3
12355	5.370638	0.693147	6.132096	3

This image shows the Scaled RFM values table where the data of the customer transactions is converted into RFM values (Recency, Frequency, Monetary). Each of the RFM value is again scaled into a short-range using Log Transformation. The table shows the RFM value of each CustomerID and even it tells which cluster group it belongs to. We can also download the table in Excel format which is shown below.



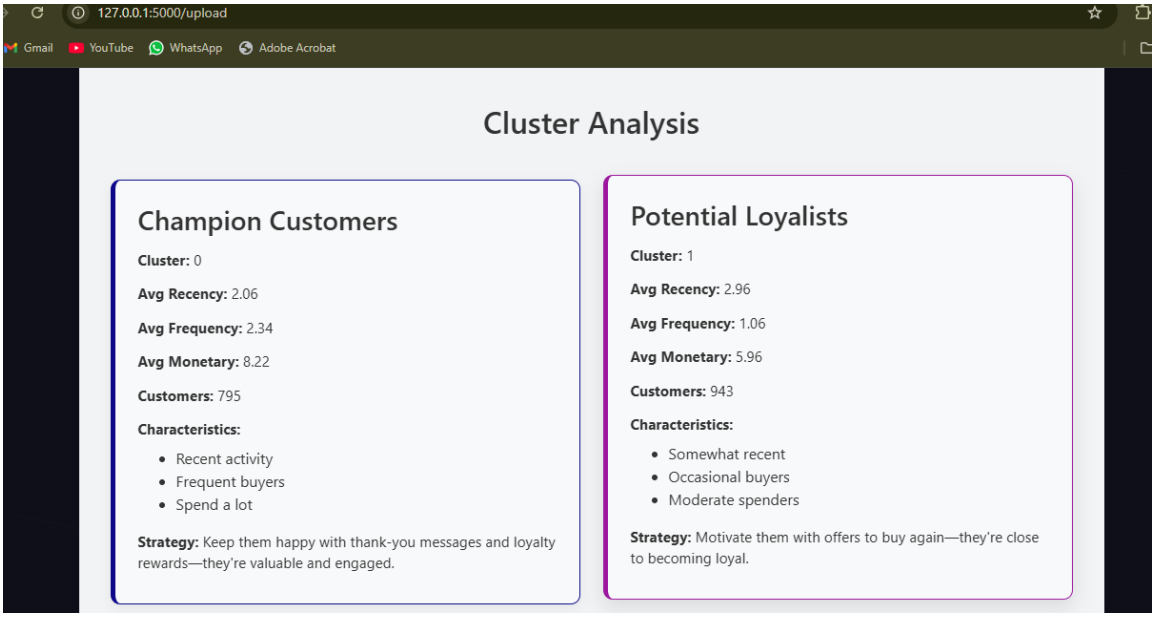
5. Visualizing the clusters

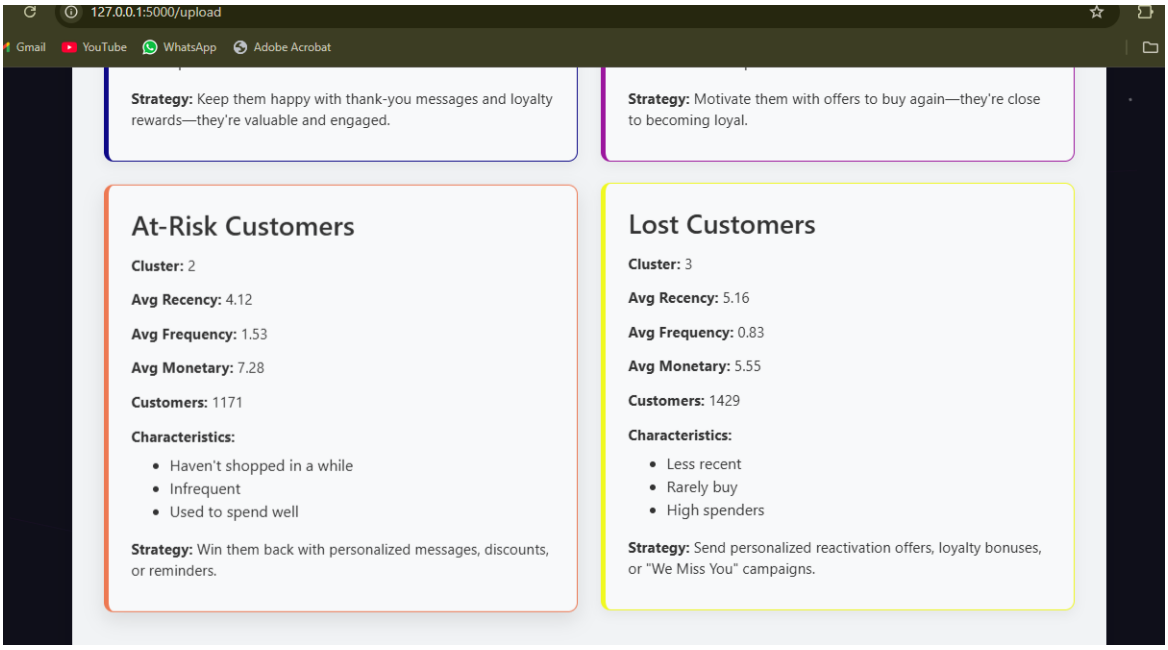




The above pictures show 4 different clusters of the customers data where you can see the Recency, Frequency, Monetary (RFM) values of each data point individually with the CustomerID. The black coloured data points in between the clusters are centroids.

6. Cluster Analysis





The final part of visualization is analysis of customer groups. The clusters which are formed before are analysed in this cluster analysis section automatically. It provides the information about the clusters like average RFM values, how many customers are there in particular cluster, it’s characteristics like High spenders, Frequent visitors etc. and also provide strategies to improve the business.