

Project Title	Shopper Spectrum: Customer Segmentation and Product Recommendations in E-Commerce
Skills take away From This Project	 Public Dataset Exploration and Preprocessing Data Cleaning and Feature Engineering Exploratory Data Analysis(EDA) Clustering Techniques Collaborative Filtering-based Product Recommendation Model Evaluation and Customer Segmentation Interpretation Streamlit
Domain	E-Commerce and Retail Analytics

Problem Statement

The global e-commerce industry generates vast amounts of transaction data daily, offering valuable insights into customer purchasing behaviors. Analyzing this data is essential for identifying meaningful customer segments and recommending relevant products to enhance customer experience and drive business growth. This project aims to examine transaction data from an online retail business to uncover patterns in customer purchase behavior, segment customers based on Recency, Frequency, and Monetary (RFM) analysis, and



develop a product recommendation system using collaborative filtering techniques.

★ Real-time Business Use Cases:

- Customer Segmentation for Targeted Marketing Campaigns
- Personalized Product Recommendations on E-Commerce Platforms
- Identifying At-Risk Customers for Retention Programs
- Dynamic Pricing Strategies Based on Purchase Behavior
- Inventory Management and Stock Optimization Based on Customer Demand Patterns

Problem Type:

- Unsupervised Machine Learning Clustering
- Collaborative Filtering Recommendation System

Project Tasks

Step 1:Dataset Collection and understanding

- Dataset Link
- Explore the dataset to understand the structure and data types.
- Identify missing values, duplicates, and unusual records.



★Dataset Description

Column	Description
InvoiceNo	Transaction number
StockCode	Unique product/item code
Description	Name of the product
Quantity	Number of products purchased
InvoiceDate	Date and time of transaction (2022–2023)
UnitPrice	Price per product
CustomerID	Unique identifier for each customer
Country	Country where the customer is based



Step 2: 📌 Data Preprocessing:

- Remove rows with missing CustomerID
- Exclude cancelled invoices (InvoiceNo starting with 'C')
- Remove negative or zero quantities and prices

Step 3 : ★ Exploratory Data Analysis (EDA):

- Analyze transaction volume by country
- Identify top-selling products
- Visualize purchase trends over time
- Inspect monetary distribution per transaction and customer
- RFM distributions
- Elbow curve for cluster selection
- Customer cluster profiles
- Product recommendation heatmap / similarity matrix

Step 4 : * Clustering Methodology:

1 Feature Engineering:



- Calculate **Recency** = Latest purchase date in dataset Customer's last purchase date
- Calculate **Frequency** = Number of transactions per customer
- Calculate **Monetary** = Total amount spent by customer
- 2 Standardize/Normalize the RFM values
- 3 Choose Clustering Algorithm (KMeans, DBScan, Hierarchial etc)
- 4 Use Elbow Method, Silhouette Score to decide the number of clusters
- 5 Run Clustering

Label the clusters by interpreting their RFM averages:

Cluster	Characteristics	Segment Label
High R, High F, High M	Regular, frequent, recent, and big spenders	High-Value
Medium F, Medium M	Steady purchasers but not premium	Regular
Low F, Low M, older R	Rare, occasional purchases	Occasional



High R, Low F, Low M	Haven't purchased in a long time	At-Risk

6 Visualize the clusters using a scatter plot or 3D plot of RFM scores.

Save the best performing model for streamlit usage

Recommendation System Approach:

- Use Item-based Collaborative Filtering
- Compute cosine similarity (or another similarity metric) between products based on purchase history (CustomerID–StockCode matrix)
- Return top 5 similar products to the entered product name

Streamlit App Features

@1 Product Recommendation Module

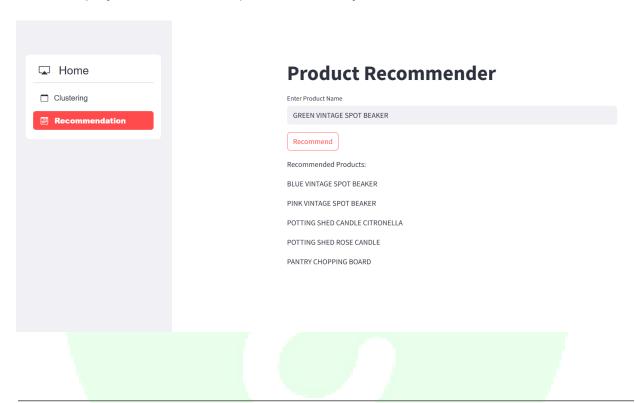
Objective:

When a user inputs a product name, the app recommends **5 similar products** based on collaborative filtering.

Functionality:



- Text input box for Product Name
- Button: Get Recommendations
- Display 5 recommended products as a styled list or card view

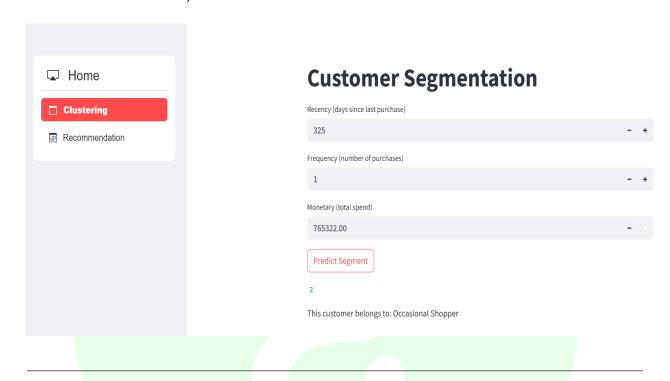


@2 Customer Segmentation Module

- Functionality:
 - 3 number inputs for:
 - Recency (in days)
 - Frequency (number of purchases)
 - Monetary (total spend)



- Button: Predict Cluster
- Display: Cluster label (e.g., High-Value, Regular, Occasional, At-Risk)



X Technical Tags

Pandas, Numpy, DataCleaning, FeatureEngineering, EDA, RFMAnalysis, CustomerSegmentation, KMeansClustering, CollaborativeFiltering, CosineSimilarity, ProductRecommendation, ScikitLearn, StandardScaler, StreamlitApp, MachineLearning, DataVisualization, PivotTables, DataTransformation, RealTimePrediction



roject Deliverables:

- Python Notebook with:
 - Clean, well-documented code with comments
 - Visualizations for EDA and clustering insights
 - o RFM-based customer segmentation and product similarity analysis
 - Model evaluations for clustering (like inertia, silhouette score)
- III Streamlit Web Application:
 - User input for a product name → recommends 5 similar products
 - Customer behavior input (Recency, Frequency, Monetary) → predicts cluster segment
 - Clean, interactive UI with real-time outputs

Timeline

The project should be completed and submitted **within 8 days** from the date it is assigned.

Created By	Verified By	Approved By
Nilofer Mubeen	Shadiya	Nehlath Harmain



References

Streamlit recording (English)	Special session for STREAMLIT(11/08/2024)
Streamlit Reference doc	Streamlit API reference
Project Live Evaluation	Project Live Evaluation
Capstone Explanation Guideline	Capstone Explanation Guideline
GitHub Reference	How to Use GitHub.pptx
Machine Learning(Eng) Recommendation systems	Project Excellence Series: Guided Learning & Problem Solving [Machine Learning USVL](English)
Machine Learning(Tam) Recommendation systems	Project Excellence Series: Guided Learning & Problem Solving [Machine Learning USVL](TAMIL)
Project Orientation	xit-tvjx-xth (2025-06-16 11:53 GMT+5:30)



PROJECT DOUBT CLARIFICATION SESSION (PROJECT AND CLASS DOUBTS)

About Session: The Project Doubt Clarification Session is a helpful resource for resolving questions and concerns about projects and class topics. It provides support in understanding project requirements, addressing code issues, and clarifying class concepts. The session aims to enhance comprehension and provide guidance to overcome challenges effectively.

Note: Book the slot at least before 12:00 Pm on the same day

Timing: Monday-Saturday (4:00PM to 5:00PM)

Booking link: https://forms.gle/XC553oSbMJ2Gcfug9

LIVE EVALUATION SESSION (CAPSTONE AND FINAL PROJECT)

About Session: The Live Evaluation Session for Capstone and Final Projects allows participants to showcase their projects and receive real-time feedback for improvement. It assesses project quality and provides an opportunity for discussion and evaluation.

Note: This form will Open only on Saturday (after 2 PM) and Sunday on Every Week

Timing: Monday-Saturday (05:30PM to 07:00PM)

Booking link: https://forms.gle/1m2Gsro41fLtZurRA