


<b>Project Title</b>	 <b>Shopper Spectrum: Customer Segmentation and Product Recommendations in E-Commerce</b>
<b>Skills take away From This Project</b>	<ul style="list-style-type: none"> <li>• Public Dataset Exploration and Preprocessing</li> <li>• Data Cleaning and Feature Engineering</li> <li>• Exploratory Data Analysis(EDA)</li> <li>• Clustering Techniques</li> <li>• Collaborative Filtering-based Product Recommendation</li> <li>• Model Evaluation and Customer Segmentation Interpretation</li> <li>• Streamlit</li> </ul>
<b>Domain</b>	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
<b>InvoiceNo</b>	Transaction number
<b>StockCode</b>	Unique product/item code
<b>Description</b>	Name of the product
<b>Quantity</b>	Number of products purchased
<b>InvoiceDate</b>	Date and time of transaction (2022–2023)
<b>UnitPrice</b>	Price per product
<b>CustomerID</b>	Unique identifier for each customer
<b>Country</b>	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, Hierarchical 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	<b>High-Value</b>
Medium F, Medium M	Steady purchasers but not premium	<b>Regular</b>
Low F, Low M, older R	Rare, occasional purchases	<b>Occasional</b>

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

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

7 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

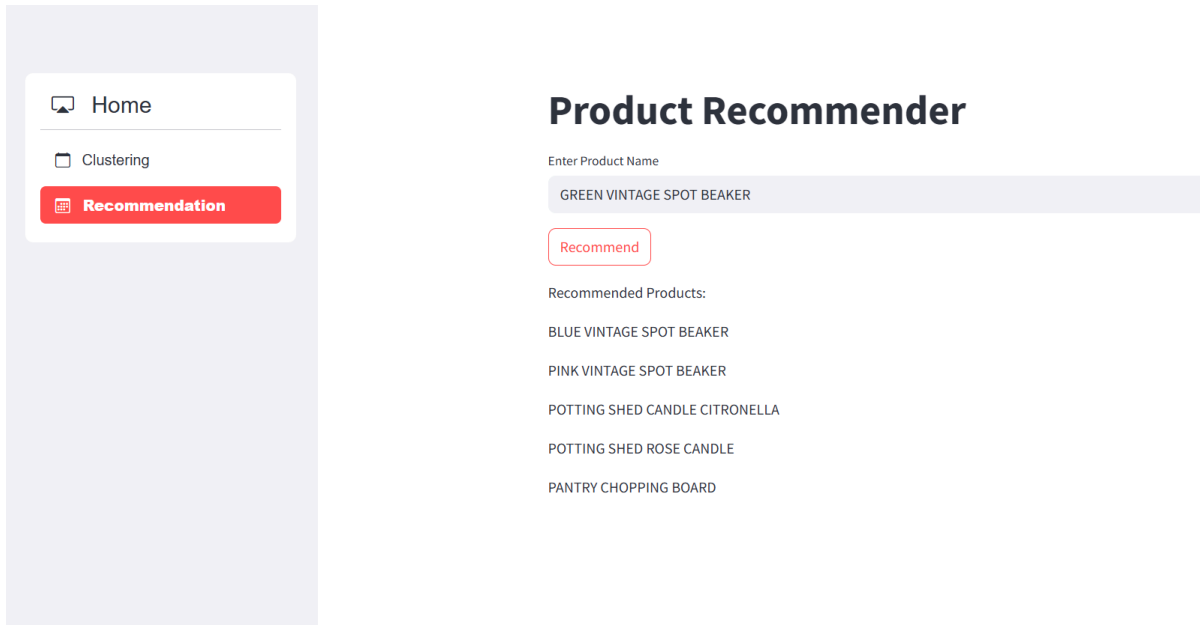
### 1 Product Recommendation Module

#### 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



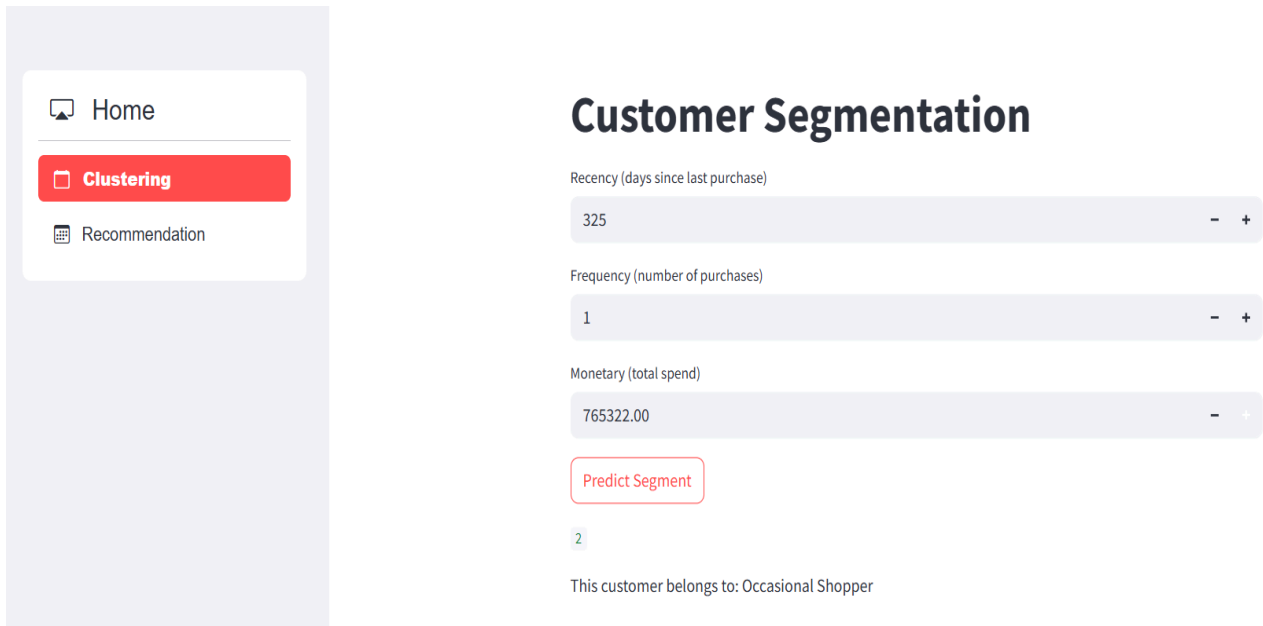
The mockup shows a web interface for a 'Product Recommender'. On the left is a sidebar with three menu items: 'Home' (selected), 'Clustering', and 'Recommendation' (highlighted with a red background). The main content area has the title 'Product Recommender'. Below the title is a text input field labeled 'Enter Product Name' containing the text 'GREEN VINTAGE SPOT BEAKER'. A red 'Recommend' button is positioned below the input field. Underneath the button, the text 'Recommended Products:' is followed by a list of five items: 'BLUE VINTAGE SPOT BEAKER', 'PINK VINTAGE SPOT BEAKER', 'POTTING SHED CANDLE CITRONELLA', 'POTTING SHED ROSE CANDLE', and 'PANTRY CHOPPING BOARD'.

## 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**)



The screenshot shows a web application for Customer Segmentation. On the left is a sidebar with three menu items: 'Home' (with a house icon), 'Clustering' (highlighted in red with a square icon), and 'Recommendation' (with a calendar icon). The main content area is titled 'Customer Segmentation'. It contains three input fields with labels and values: 'Recency (days since last purchase)' with value '325', 'Frequency (number of purchases)' with value '1', and 'Monetary (total spend)' with value '765322.00'. Each input field has minus and plus buttons on the right. Below these fields is a red 'Predict Segment' button. Underneath the button, the number '2' is displayed. At the bottom of the main area, the text 'This customer belongs to: Occasional Shopper' is shown.

---



## Technical Tags

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

---



## Project Deliverables:

-  Python Notebook with:
  - Clean, well-documented code with comments
  - Visualizations for EDA and clustering insights
  - RFM-based customer segmentation and product similarity analysis
  - Model evaluations for clustering (like inertia, silhouette score)
-  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)	<a href="#">Special session for STREAMLIT(11/08/2024)</a>
Streamlit Reference doc	<a href="#">Streamlit API reference</a>
Project Live Evaluation	<a href="#">Project Live Evaluation</a>
Capstone Explanation Guideline	<a href="#">Capstone Explanation Guideline</a>
GitHub Reference	<a href="#">How to Use GitHub.pptx</a>
Machine Learning(Eng) Recommendation systems	<a href="#">Project Excellence Series: Guided Learning &amp; Problem Solving [Machine Learning USVL](English)</a>
Machine Learning(Tam) Recommendation systems	<a href="#">Project Excellence Series: Guided Learning &amp; Problem Solving [Machine Learning USVL](TAMIL)</a>
Project Orientation	<a href="#">xit-tvjx-xth (2025-06-16 11:53 GMT+5:30)</a>

## 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>

---