# 

NAME : **GANESH S**

REGISTER NO : **727723EUCS059**

EMAIL ID : [**727723eucs059@skcet.ac.in**](mailto:727723eucs059@skcet.ac.in)

PROJECT NAME : **Customer Segmentation & Market Basket Intelligence Platform**

GITHUB LINK : **https://github.com/GaneshLathin/Customer-Segmentation-Market-Basket-Intelligence-Platform**

**Customer Segmentation & Market Basket Intelligence Platform**

# Problem Statement

An e-commerce business generates hundreds of thousands of transaction records annually across thousands of customers and products. However, without structured analysis, it becomes extremely difficult to identify high-value customers, detect at-risk customers, understand purchasing patterns, or create targeted marketing strategies.

Traditional rule-based segmentation techniques fail to capture complex behavioral patterns hidden in large transactional datasets. Manual analysis is inefficient and prone to oversight, especially when dealing with multi-dimensional customer features such as recency, frequency, and monetary value.

Therefore, there is a need for an intelligent analytics platform that applies unsupervised machine learning algorithms to automatically discover meaningful customer segments, reduce data complexity, uncover product associations, and generate actionable marketing personas. This project addresses that need through the development of SegmentIQ.

# Objectives

The main objectives of this project are:

1. To ingest and clean raw transactional data into a structured customer-level dataset.
2. To engineer RFM (Recency, Frequency, Monetary) features along with additional behavioral metrics.
3. To apply multiple unsupervised clustering algorithms for automated customer segmentation.
4. To reduce high-dimensional customer data using dimensionality reduction techniques.
5. To perform market basket analysis to identify product association patterns.
6. To generate interpretable marketing personas based on cluster characteristics.
7. To build an interactive full-stack web dashboard for real-time analytics visualization.

**Model Implementation :**

**Unsupervised Machine Learning for Customer Segmentation:**

SegmentIQ applies multiple clustering and analytical techniques:

1. **K-Means Clustering:**Uses centroid-based partitioning to divide customers into optimal segments based on behavioral similarity.
2. **Agglomerative Hierarchical Clustering:**Builds a bottom-up dendrogram structure to represent hierarchical merging of customers.
3. **DBSCAN (Density-Based Spatial Clustering):**Identifies arbitrarily shaped clusters and detects anomalous customers as noise points.

**Steps involved:**

• Perform exploratory data analysis (EDA)

• Aggregate invoice-level data into customer-level features

• Engineer RFM and derived behavioral metrics

• Apply log transformation (log1p) to reduce skewness

• Standardize features using z-score normalization

• Determine optimal cluster parameters using silhouette score

• Train clustering models

• Interpret cluster characteristics

### Feature Engineering:

**Customer-level features include:**

• Recency (Days since last purchase)

• Frequency (Number of unique invoices)

• Monetary Value (Total spending)

• Average Order Value

• Total Items Purchased

• Unique Products Bought

• Average Basket Size

## Dimensionality Reduction:

To visualize high-dimensional customer data:

1. **PCA (Principal Component Analysis):**  
   Reduces multi-dimensional feature space into principal components while retaining maximum variance.
2. **LDA (Linear Discriminant Analysis):**  
   Uses cluster labels as targets to maximize separation between discovered segments.

**Techniques used:**  
 • Explained variance ratio analysis

• Component loading interpretation

• 2D and 3D visualization of clusters

## Market Basket Analysis:

## The Apriori algorithm is applied to extract frequent itemsets and association rules from transaction data.

## • Identify frequently co-purchased products

## • Generate rules using support, confidence, and lift

## • Rank product relationships by lift ratio

## • Visualize product co-occurrence using heatmaps

# Hyperparameter Tuning:

# The following parameters are optimized:

# • k (Number of clusters in K-Means) • eps (Neighborhood radius in DBSCAN) • min\_samples (Minimum points for density cluster) • min\_support (Apriori support threshold)

# Techniques used:

# • Elbow method for inertia analysis

# • Silhouette score comparison

# • Dendrogram analysis

# • Lift-based ranking of association rules

# Performance Evaluation:

# Clustering models are evaluated using unsupervised performance metrics.

# The following metrics are used:

# • Silhouette Score

# • Inertia (Within-cluster sum of squares)

# • Davies-Bouldin Index

# • Explained Variance Ratio (PCA)

# • Support, Confidence, and Lift (Apriori)

# System Architecture:

# Backend: • Python 3.11 • FastAPI framework • scikit-learn, SciPy, mlxtend • Automatic dataset download from UCI repository • REST API endpoints for clustering and analytics

# Frontend: • React 18 with Vite • Tailwind CSS styling • Recharts for data visualization • Three.js for 3D background • Framer Motion & GSAP for animations • Axios for API communication

# Results and Outcomes:

# • Successful segmentation of customers into meaningful behavioral groups

# • Silhouette score of approximately 0.42 using K-Means

# • Identification of high-value “Champion” customers

# • Detection of 5–12% anomalous customers using DBSCAN

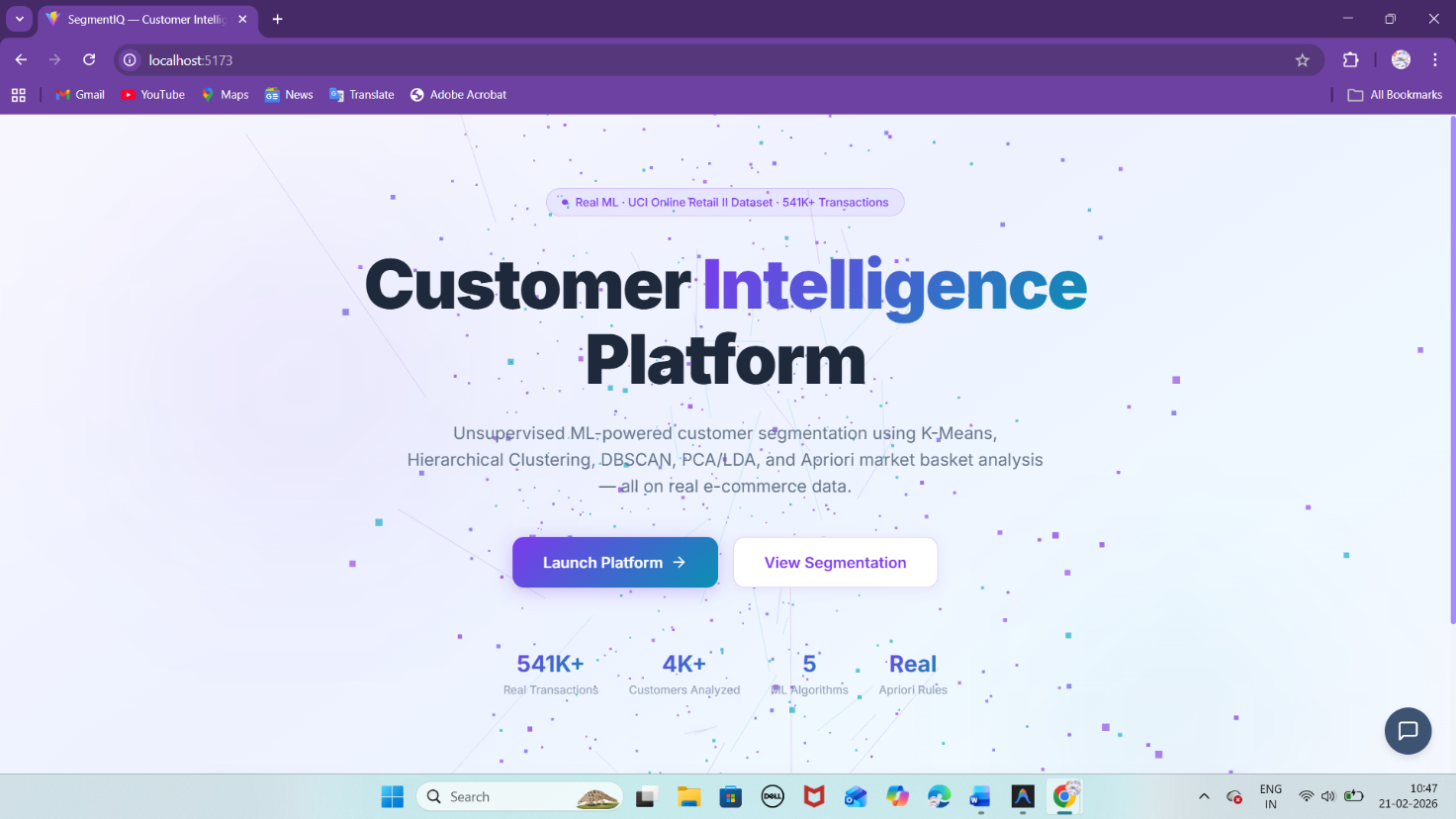
# • PCA explaining approximately 65–70% variance in first two components

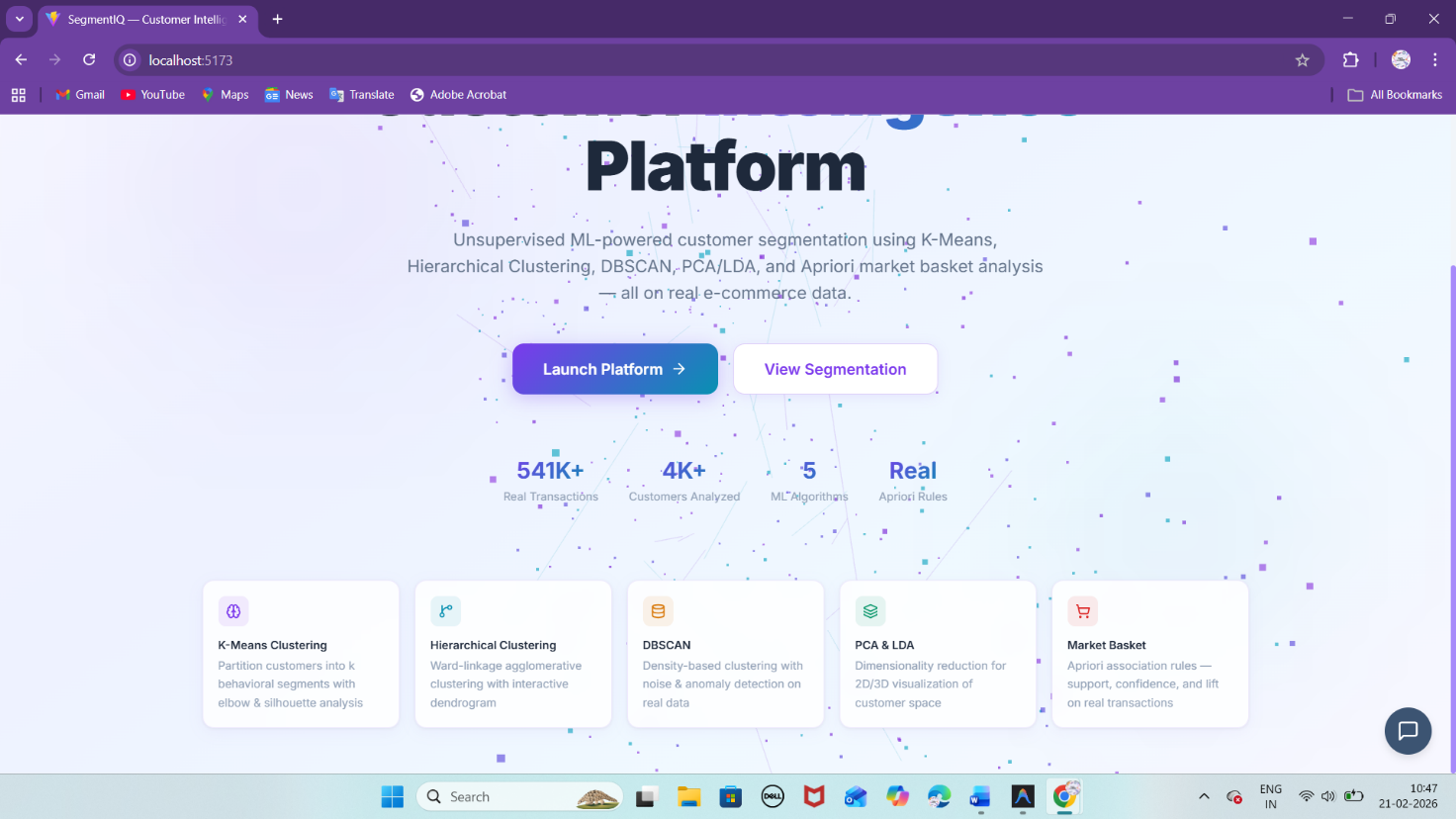
# • Discovery of high-lift product association rules

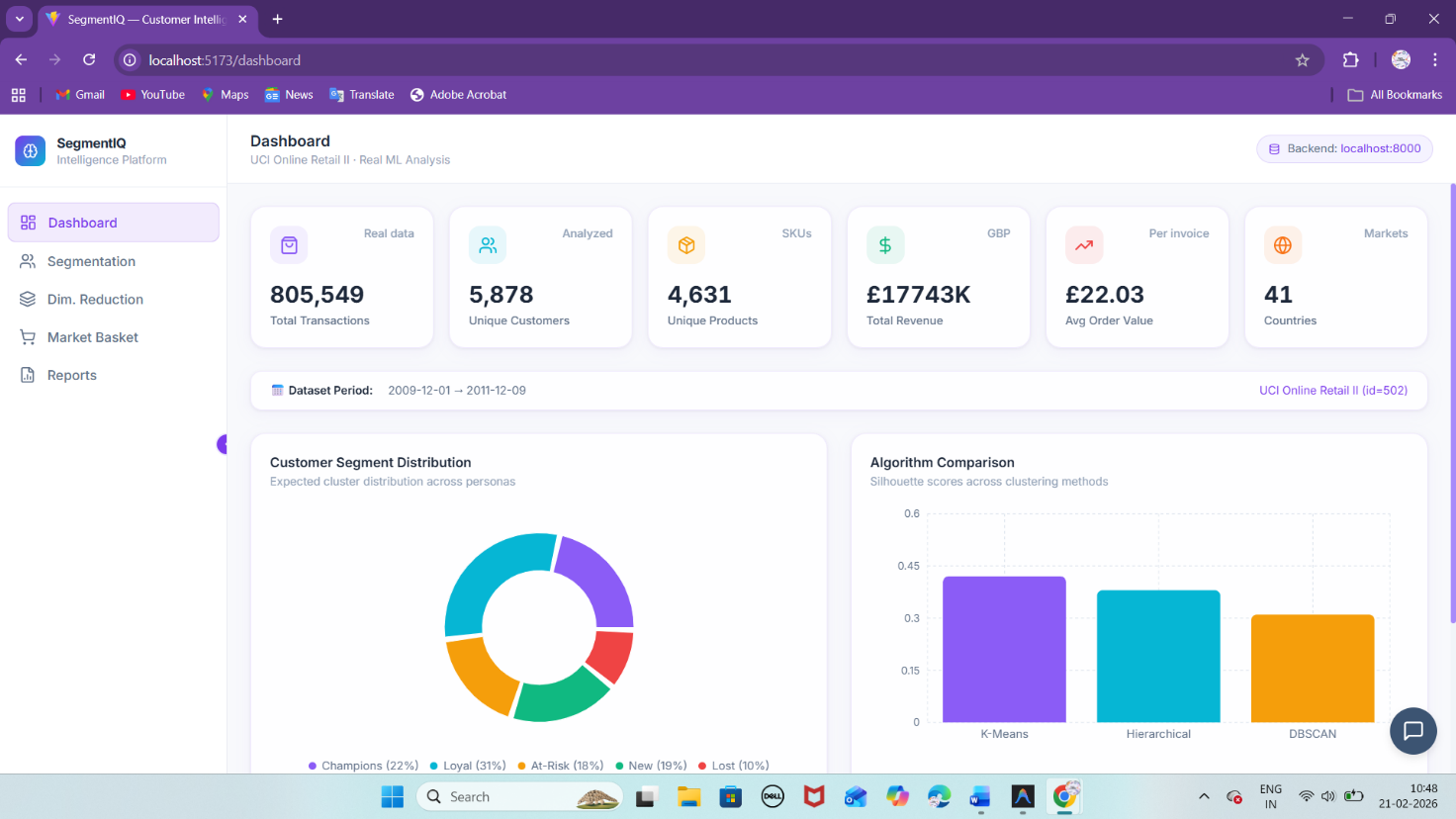
# • Generation of actionable marketing personas with campaign recommendations

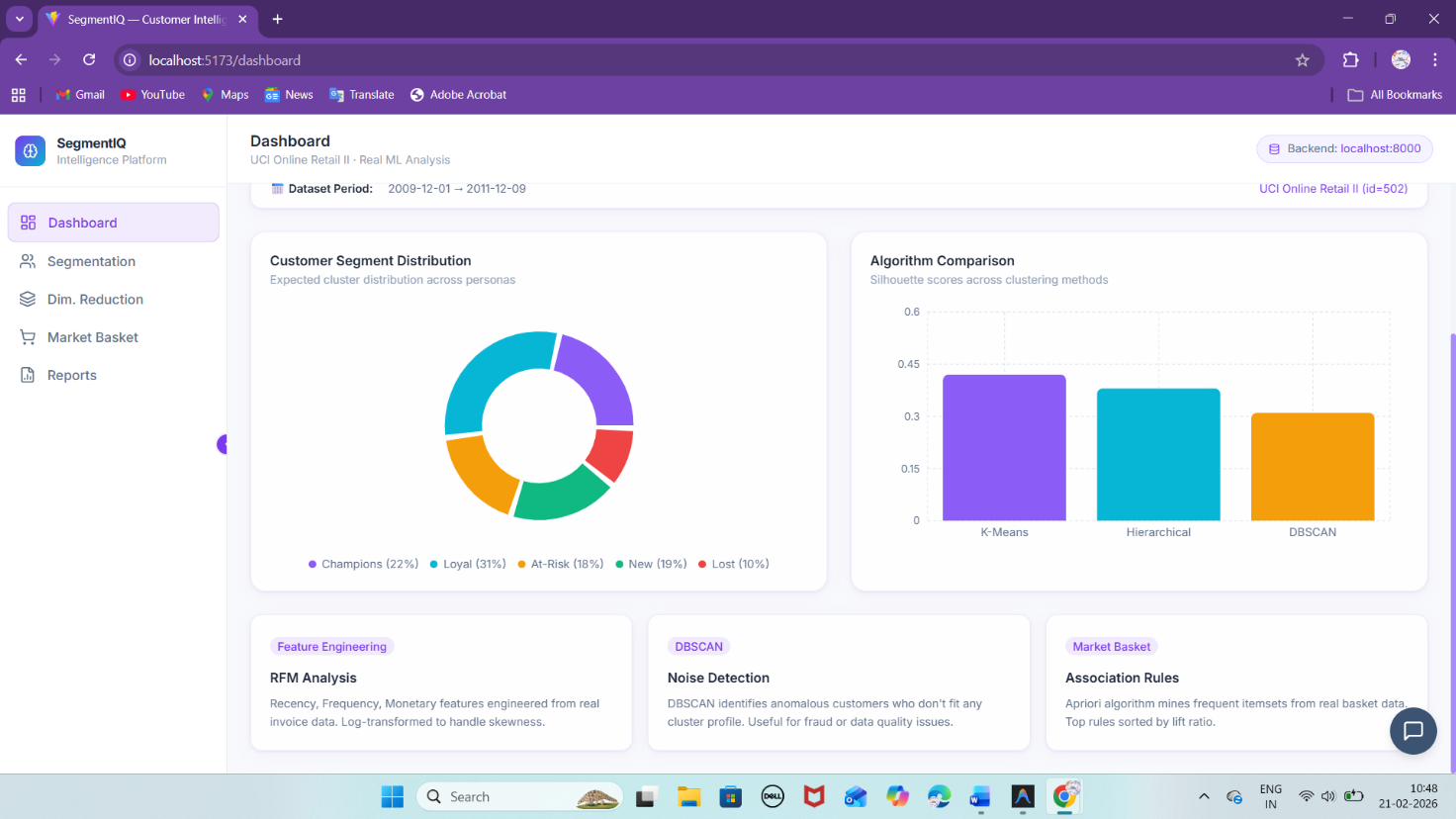
# • Interactive real-time dashboard for business decision support

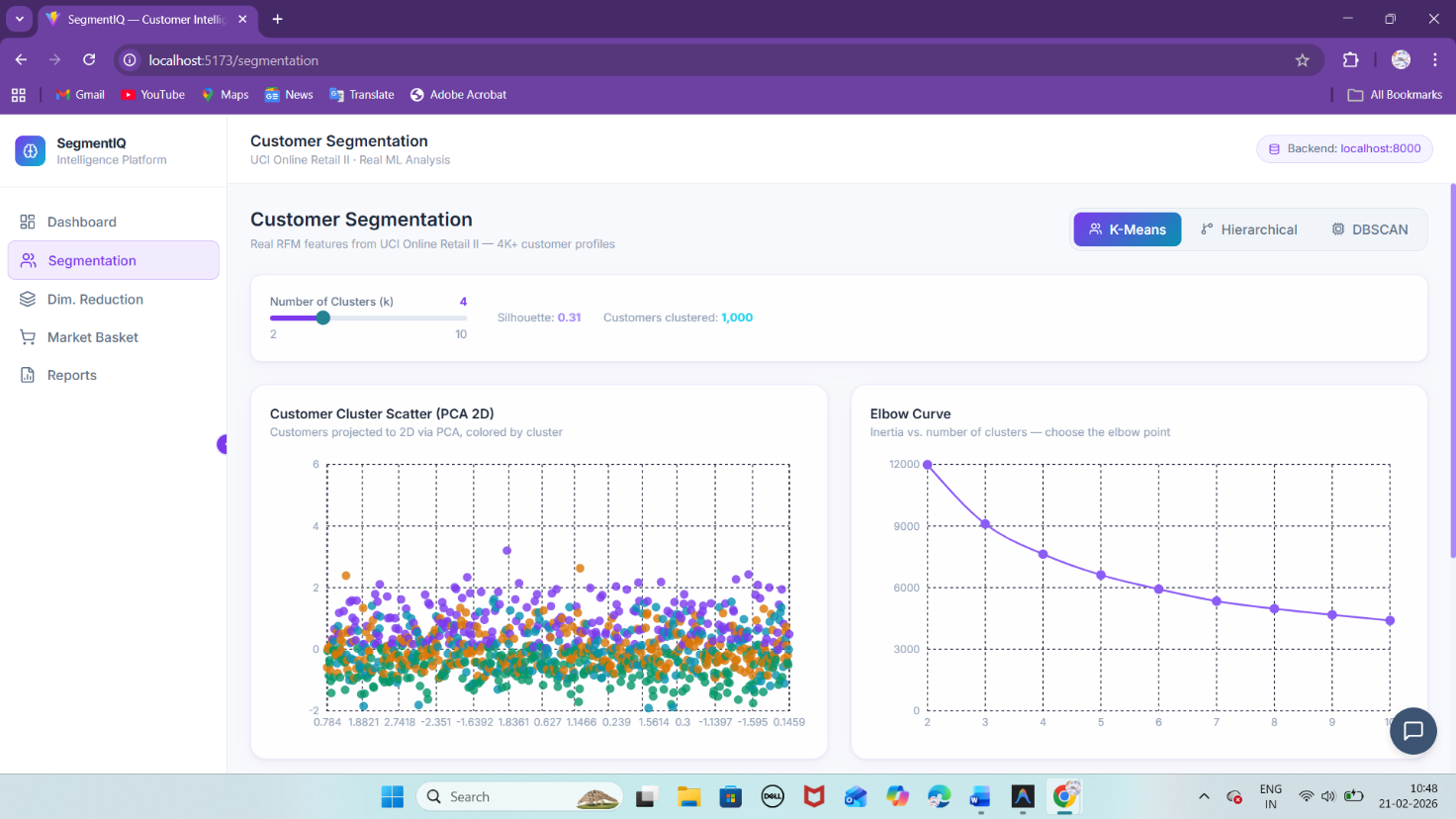
# Output:

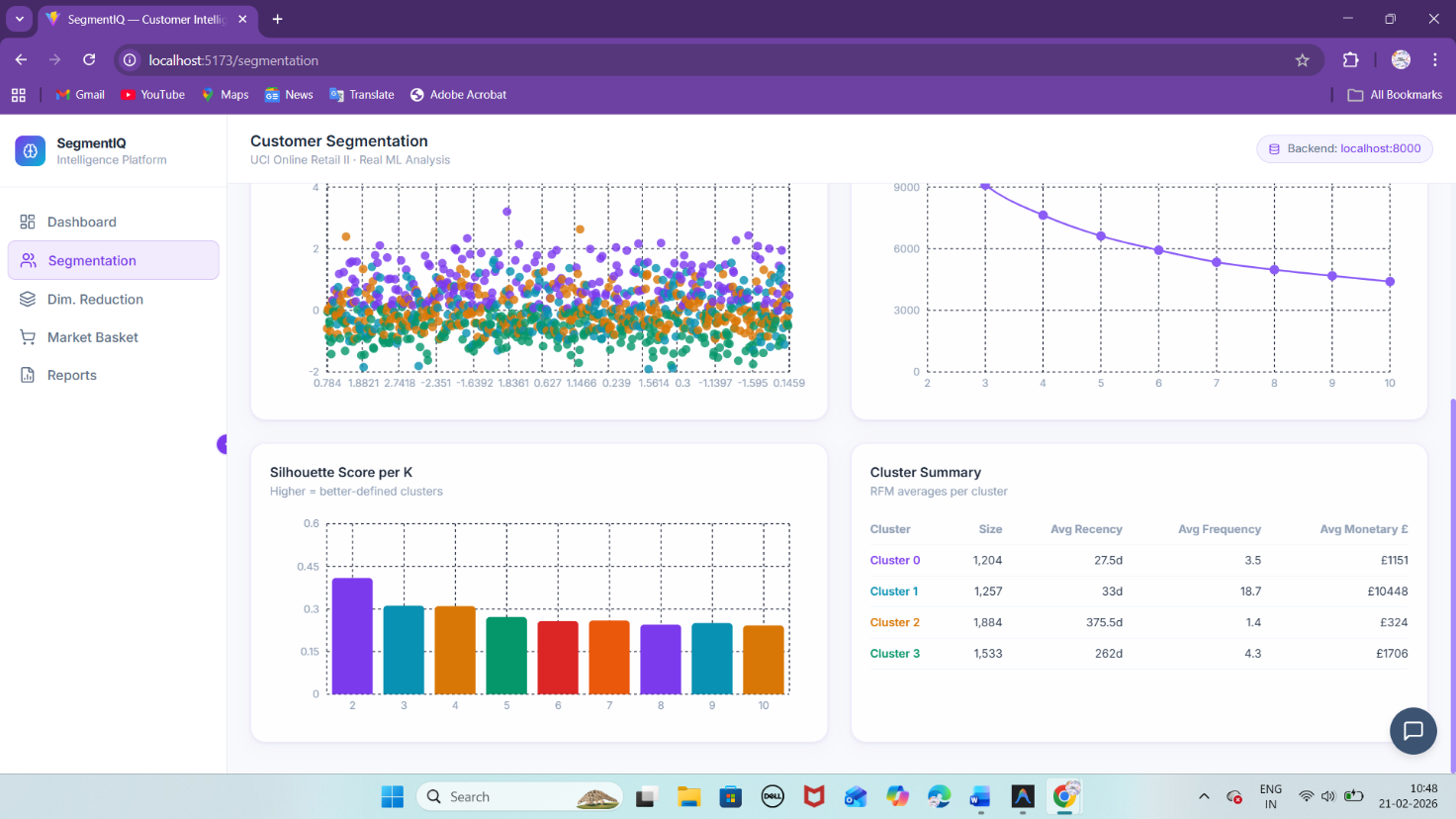
****

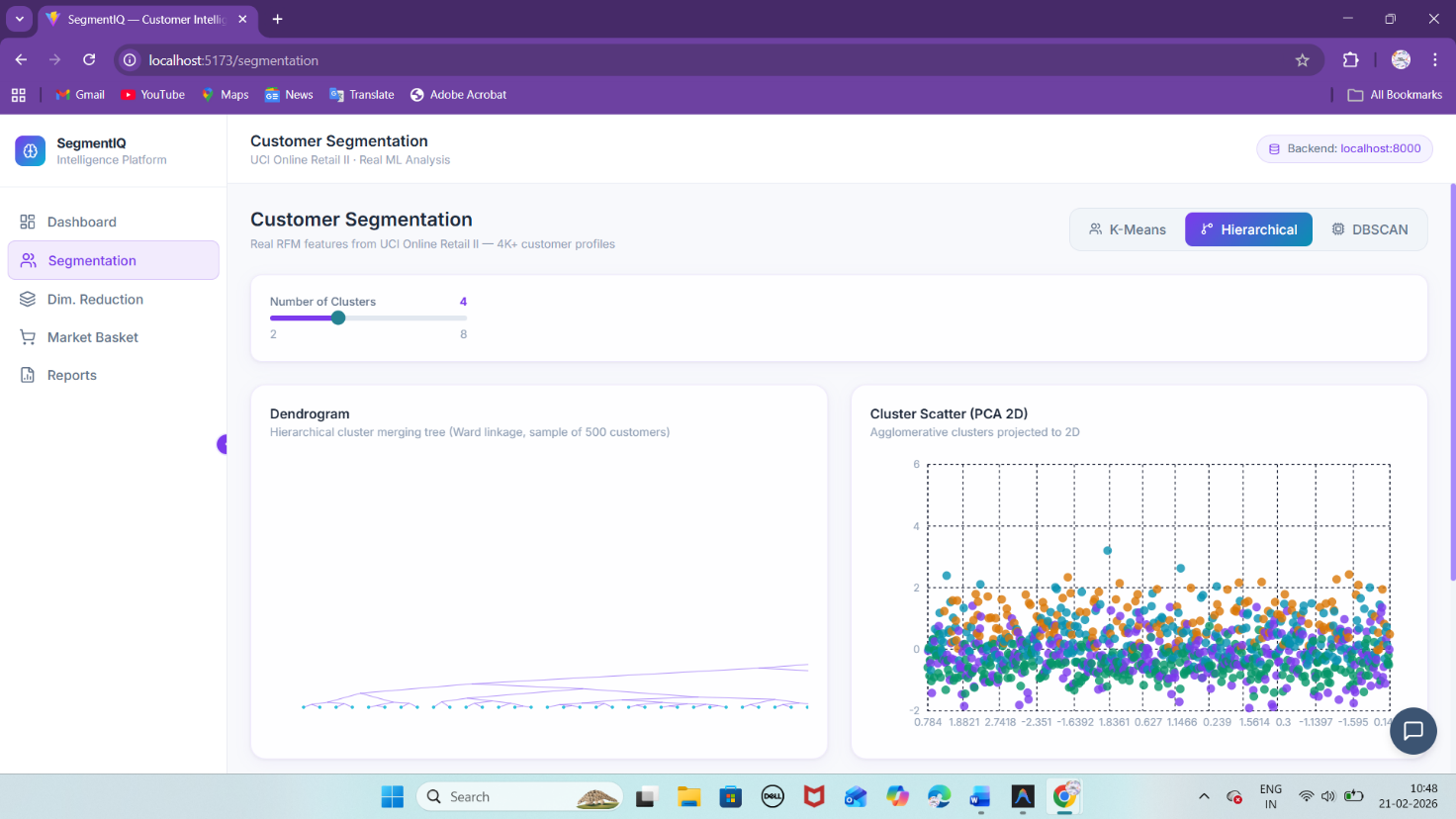
****

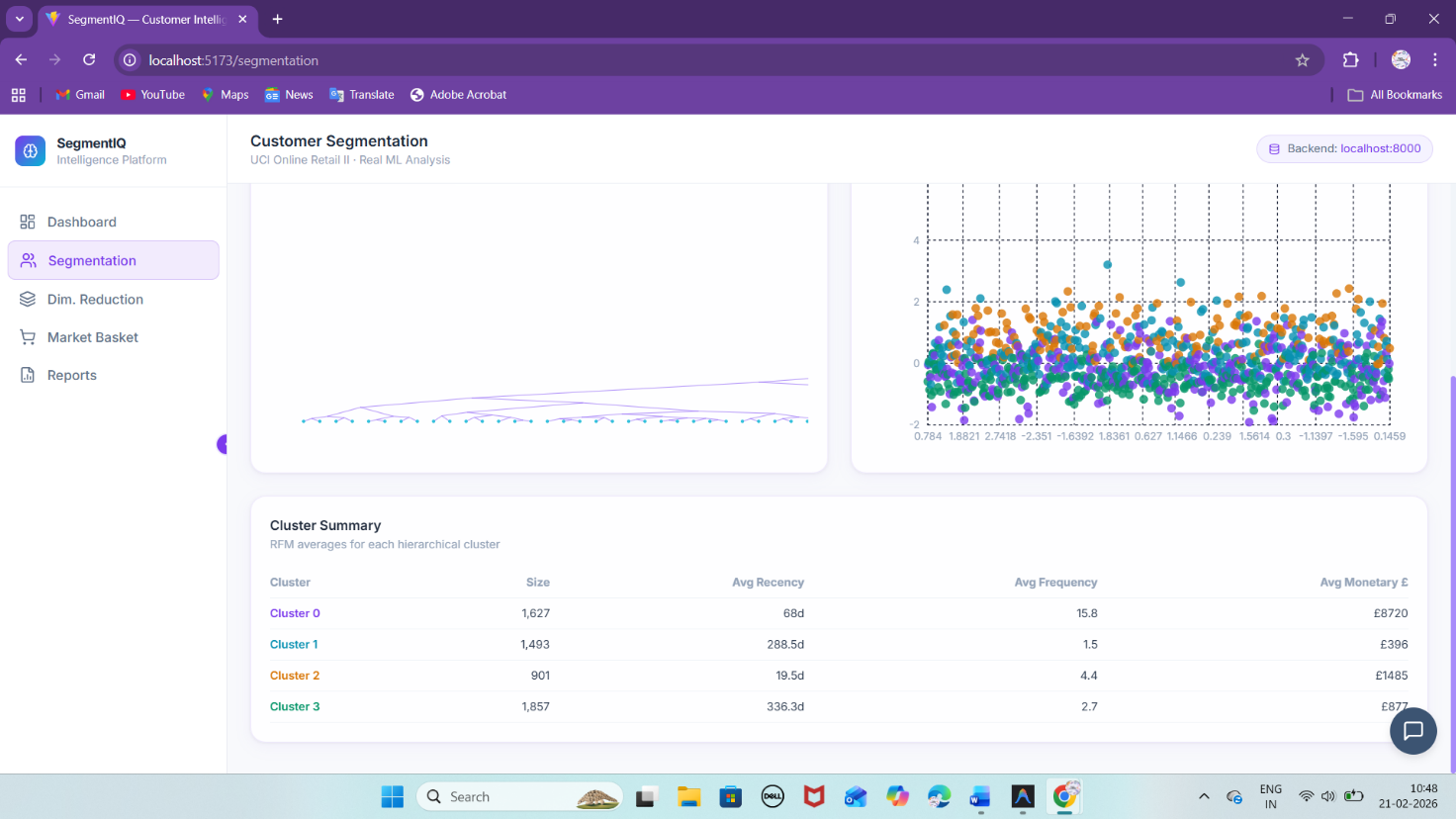
****

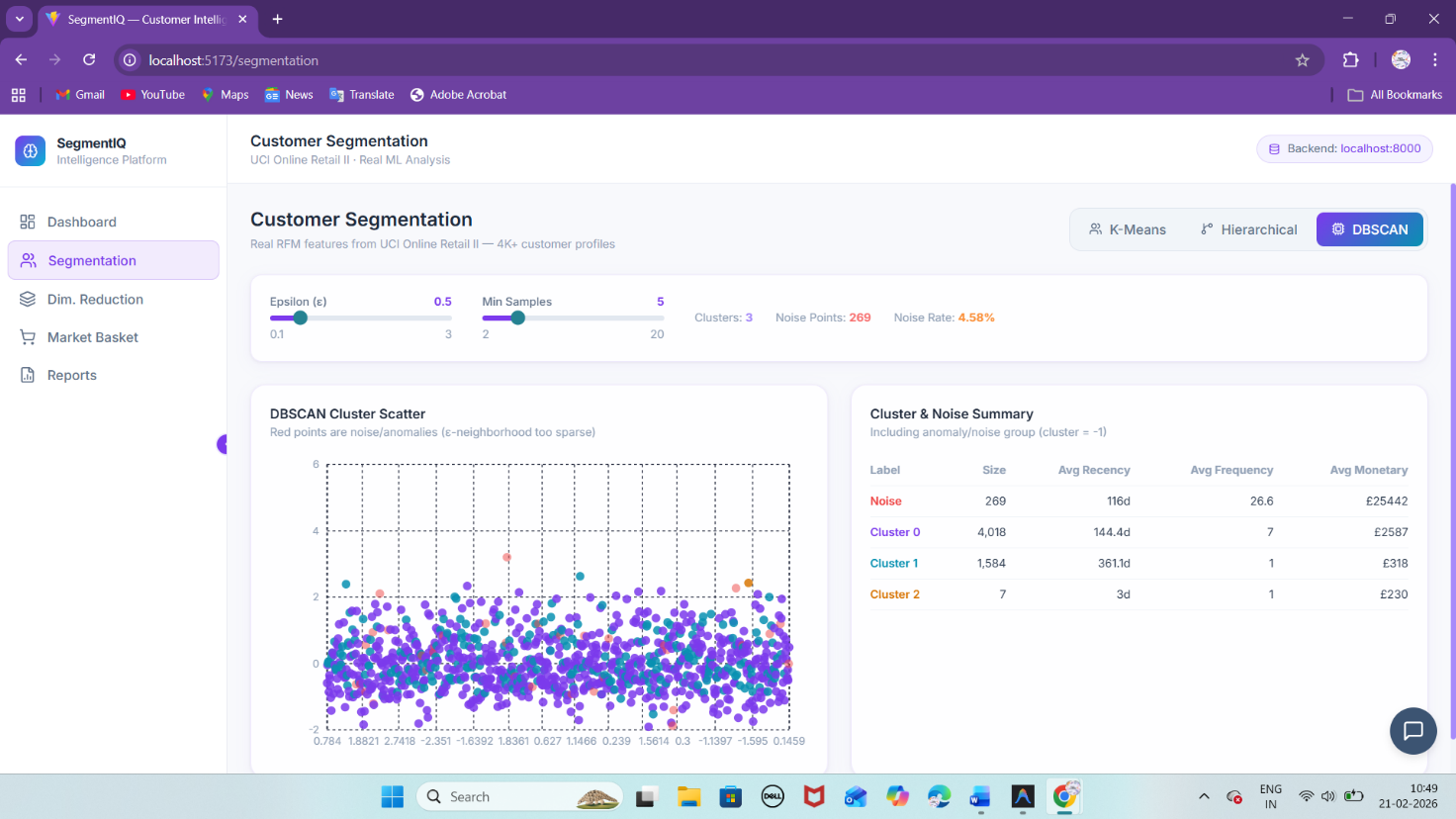
****

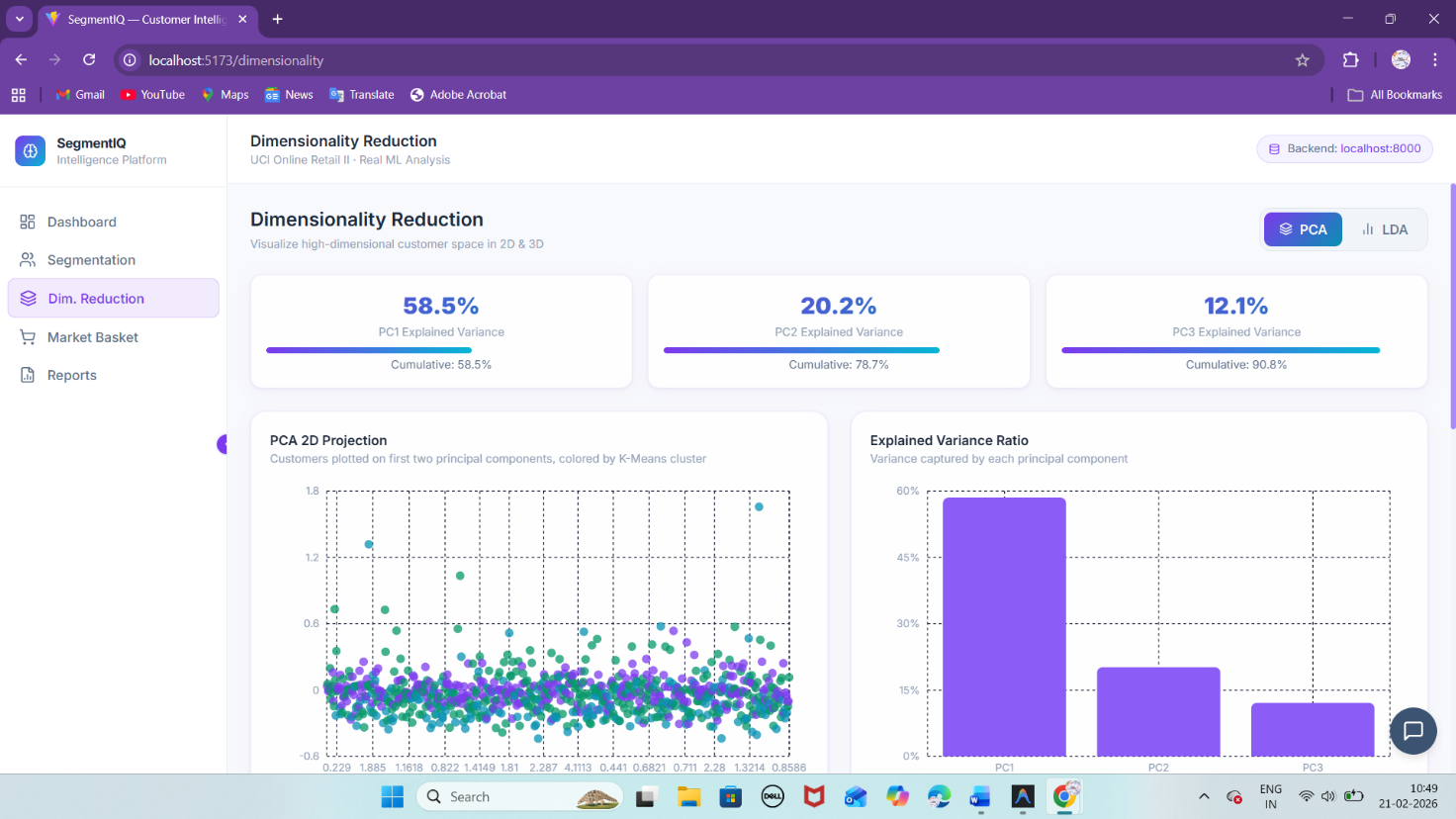
****

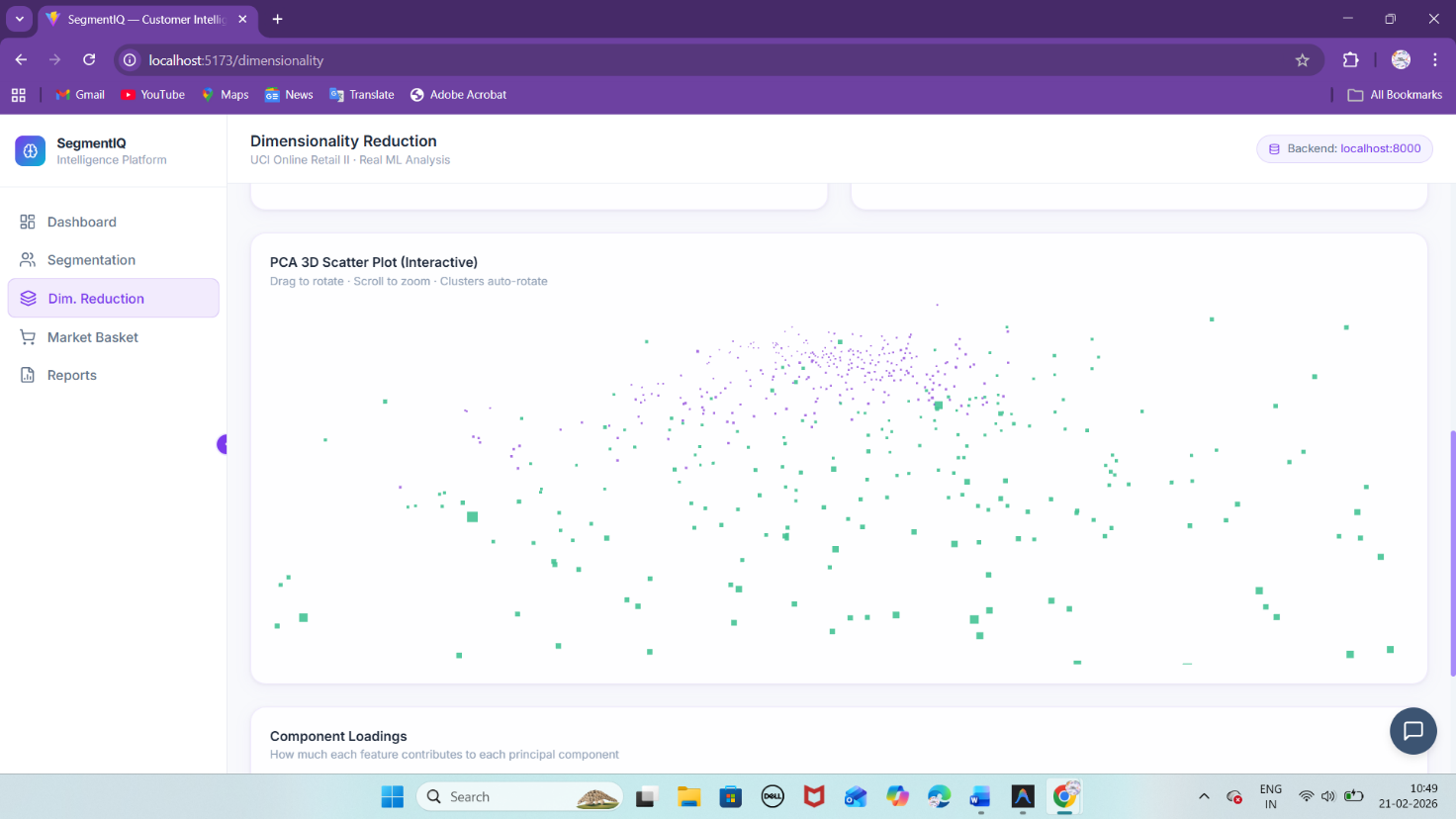
****

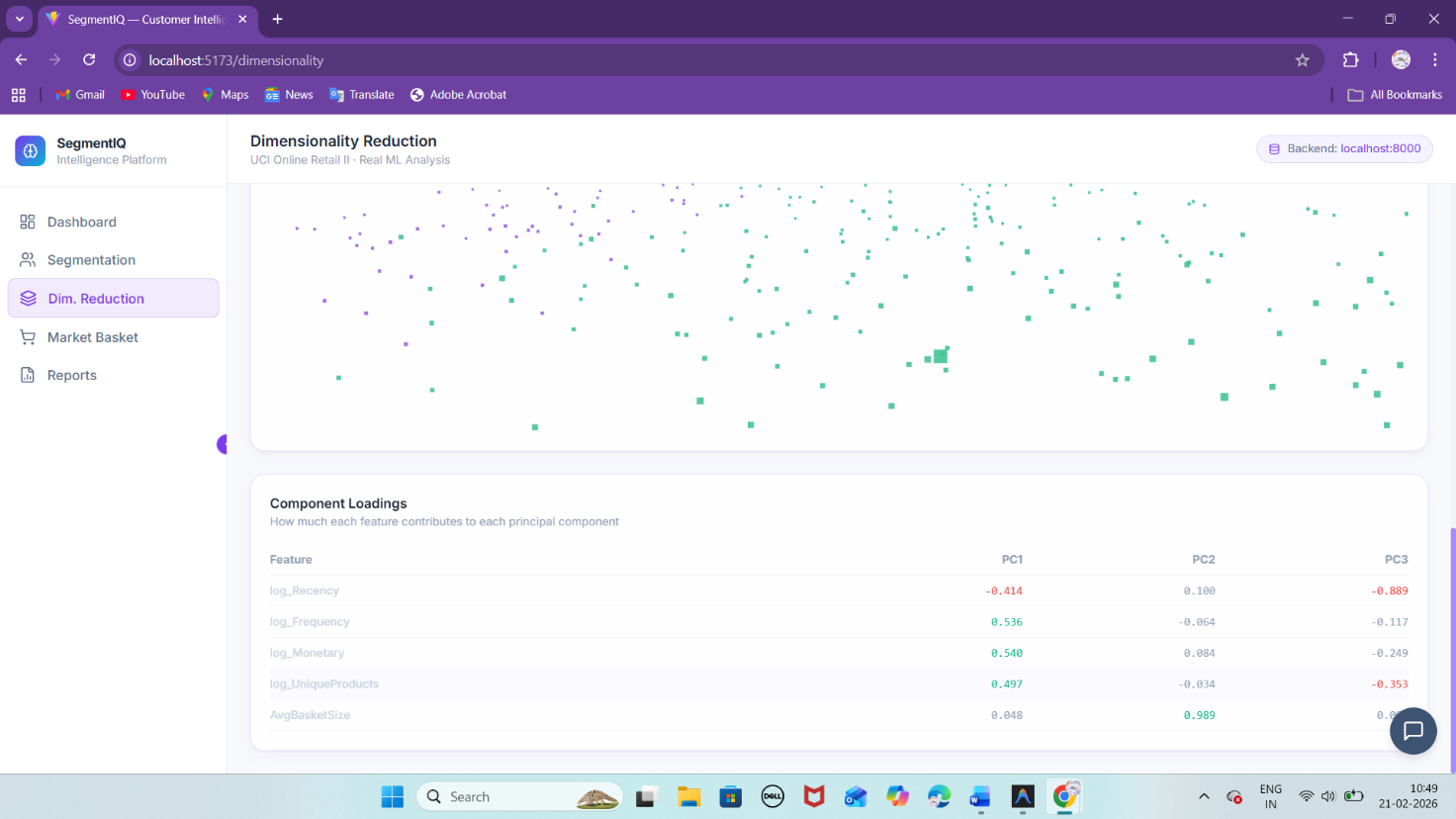
****

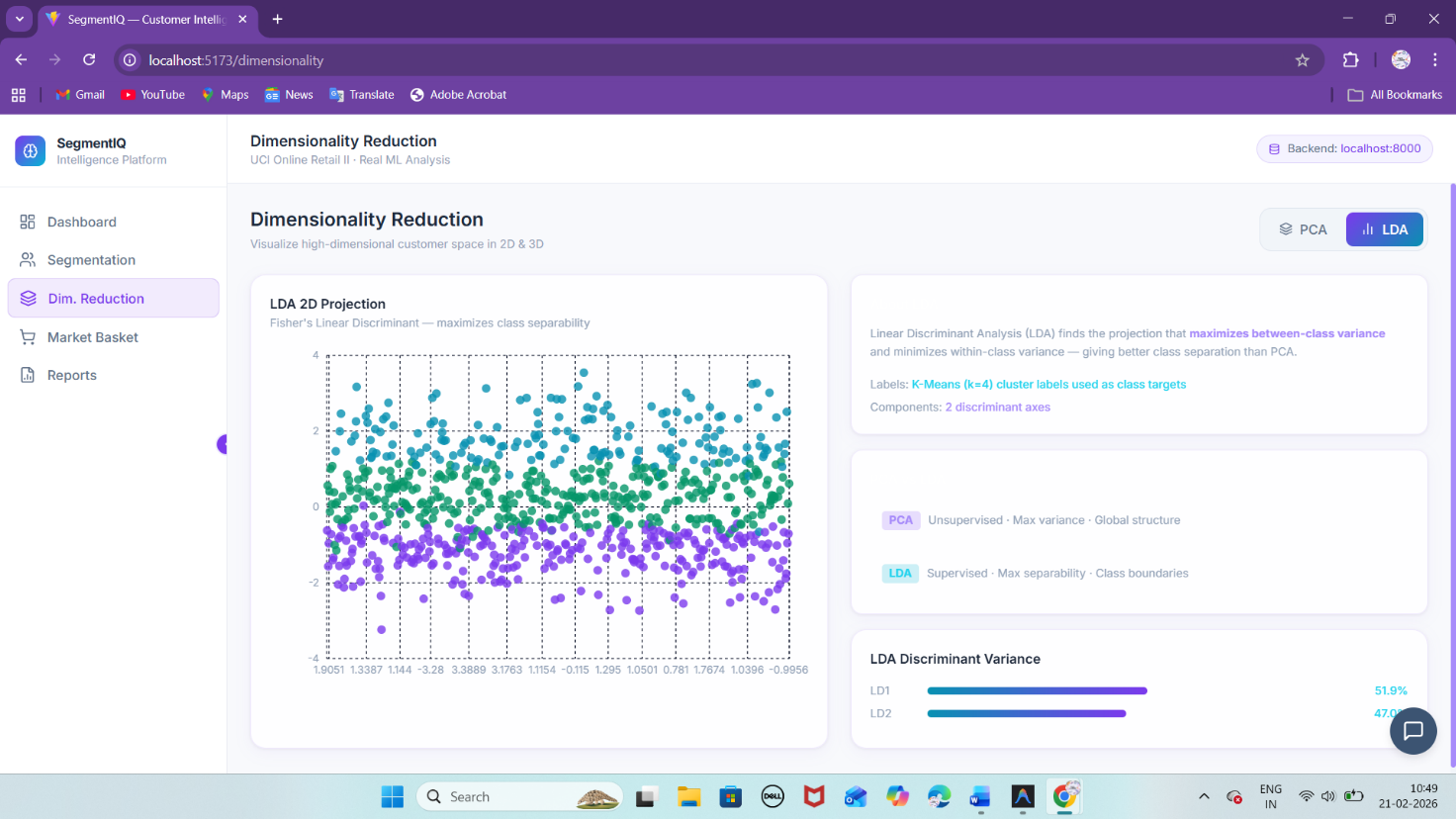
****

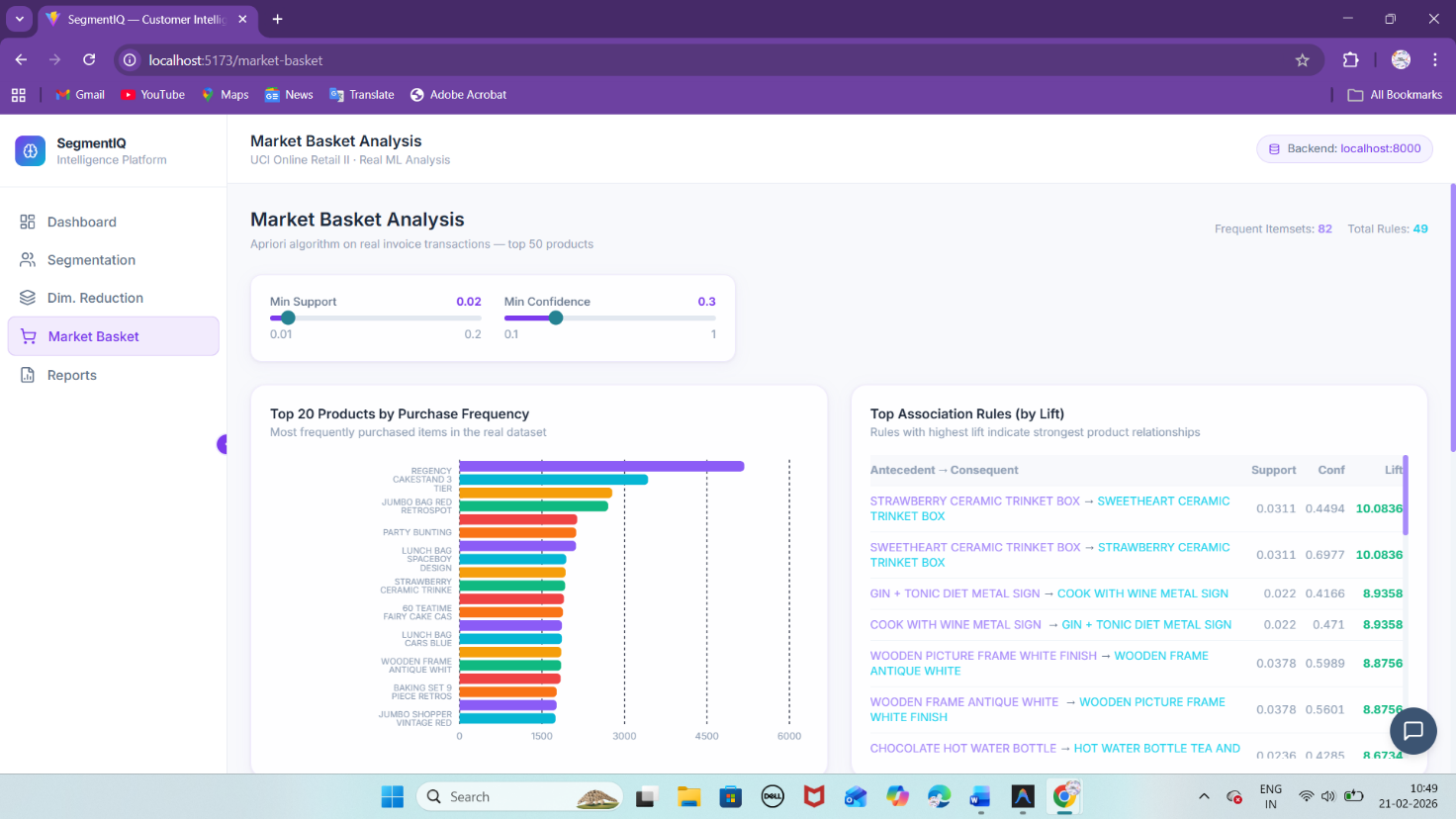
****

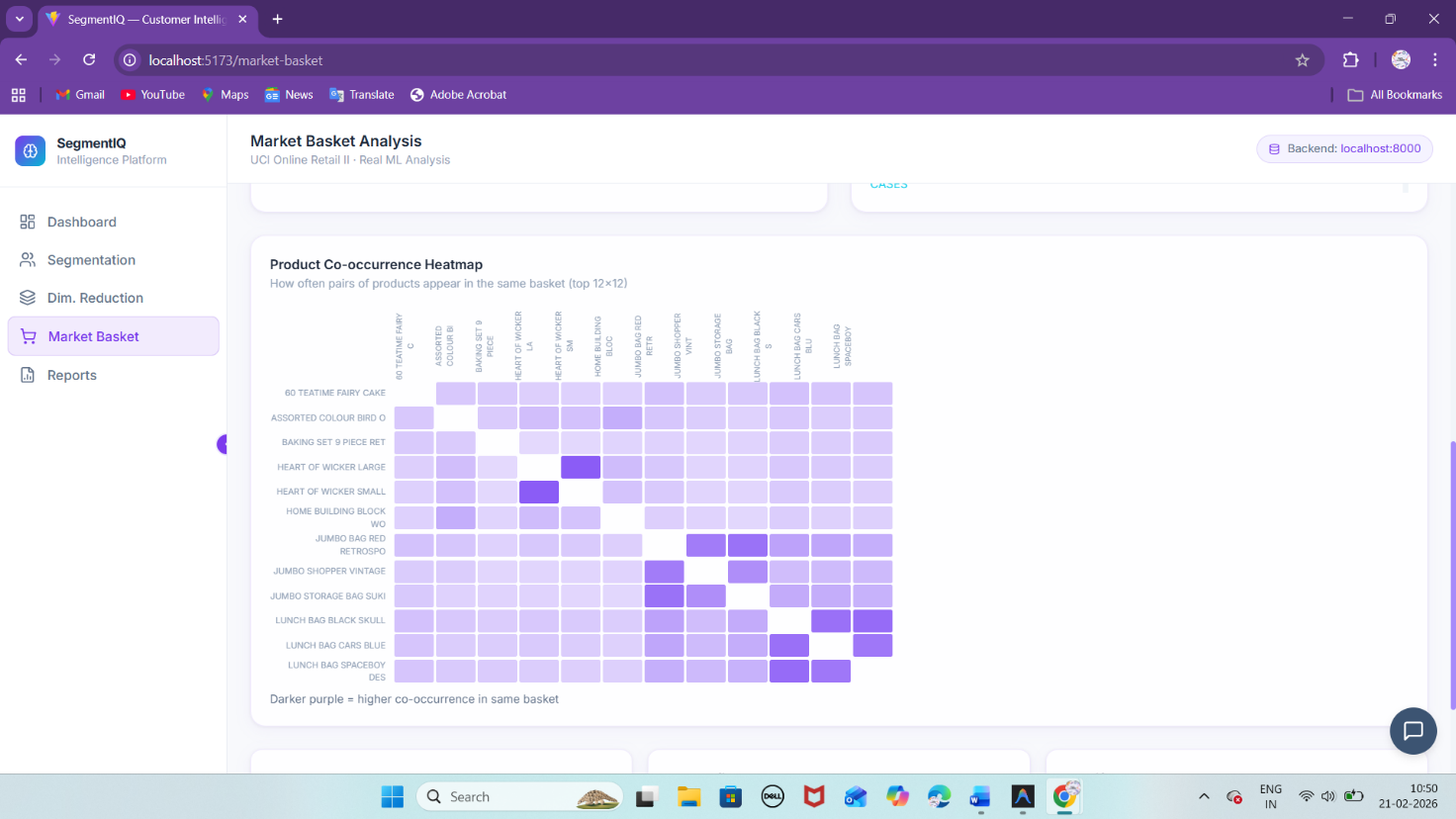
****

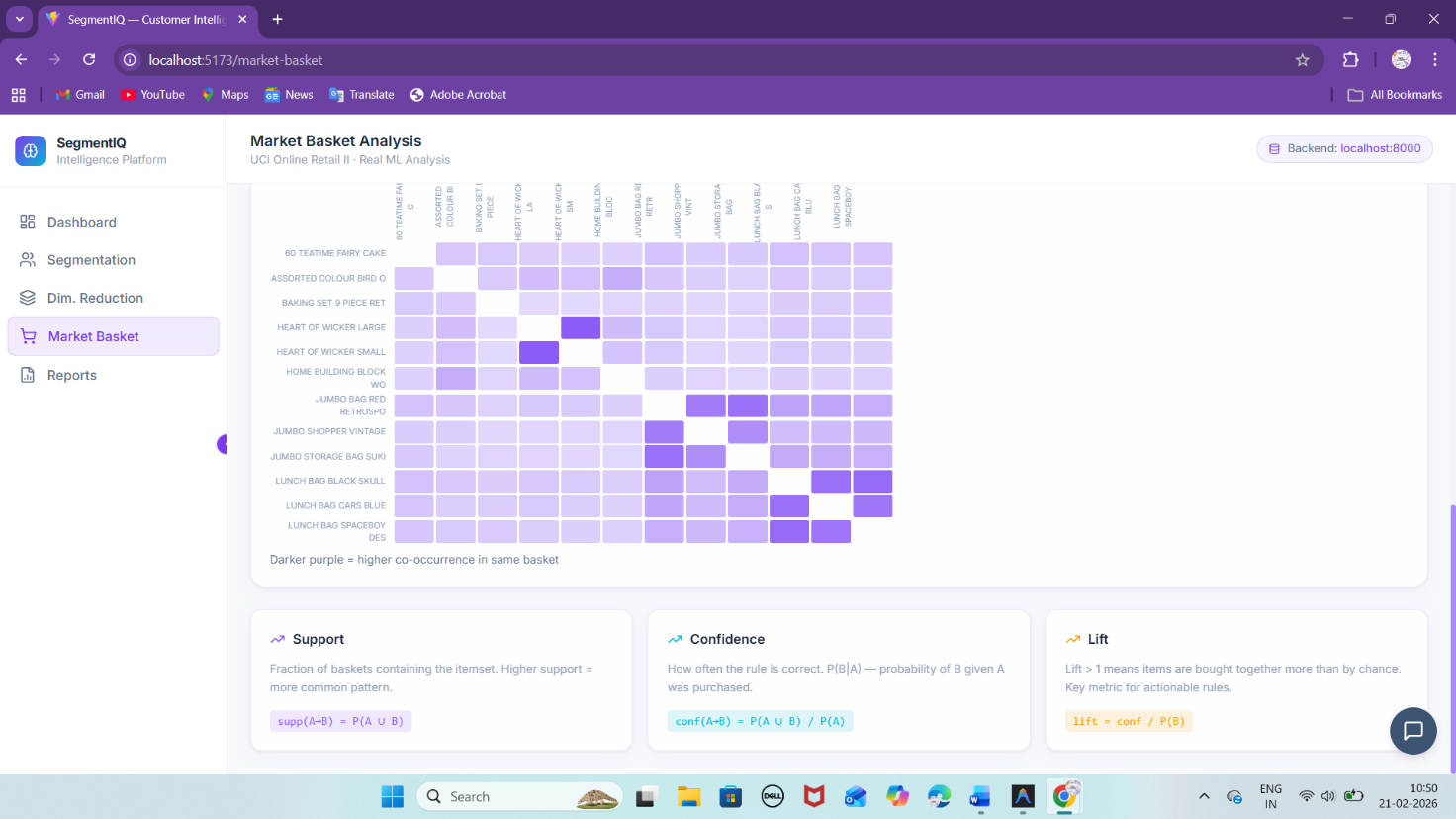
****

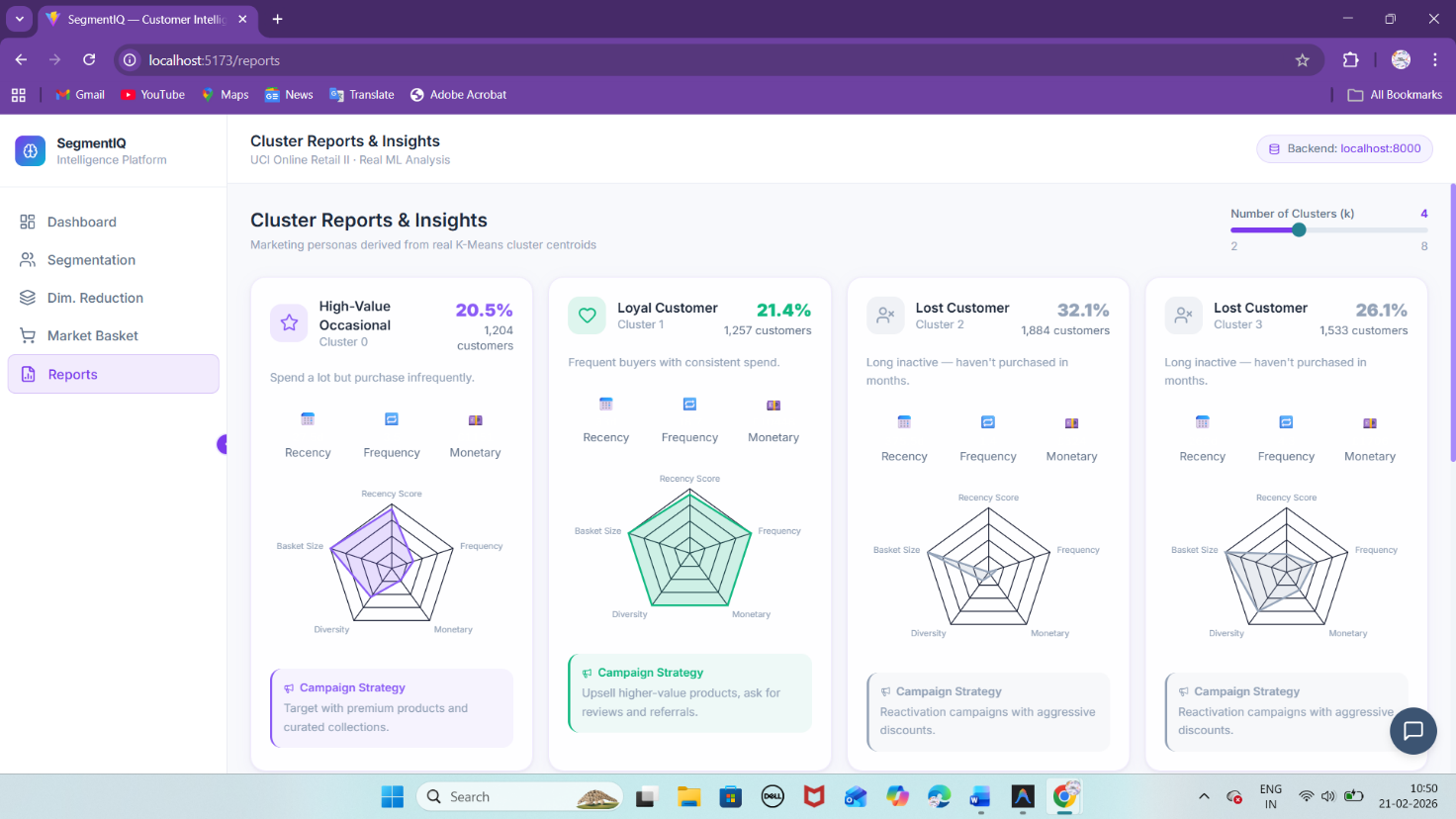
****

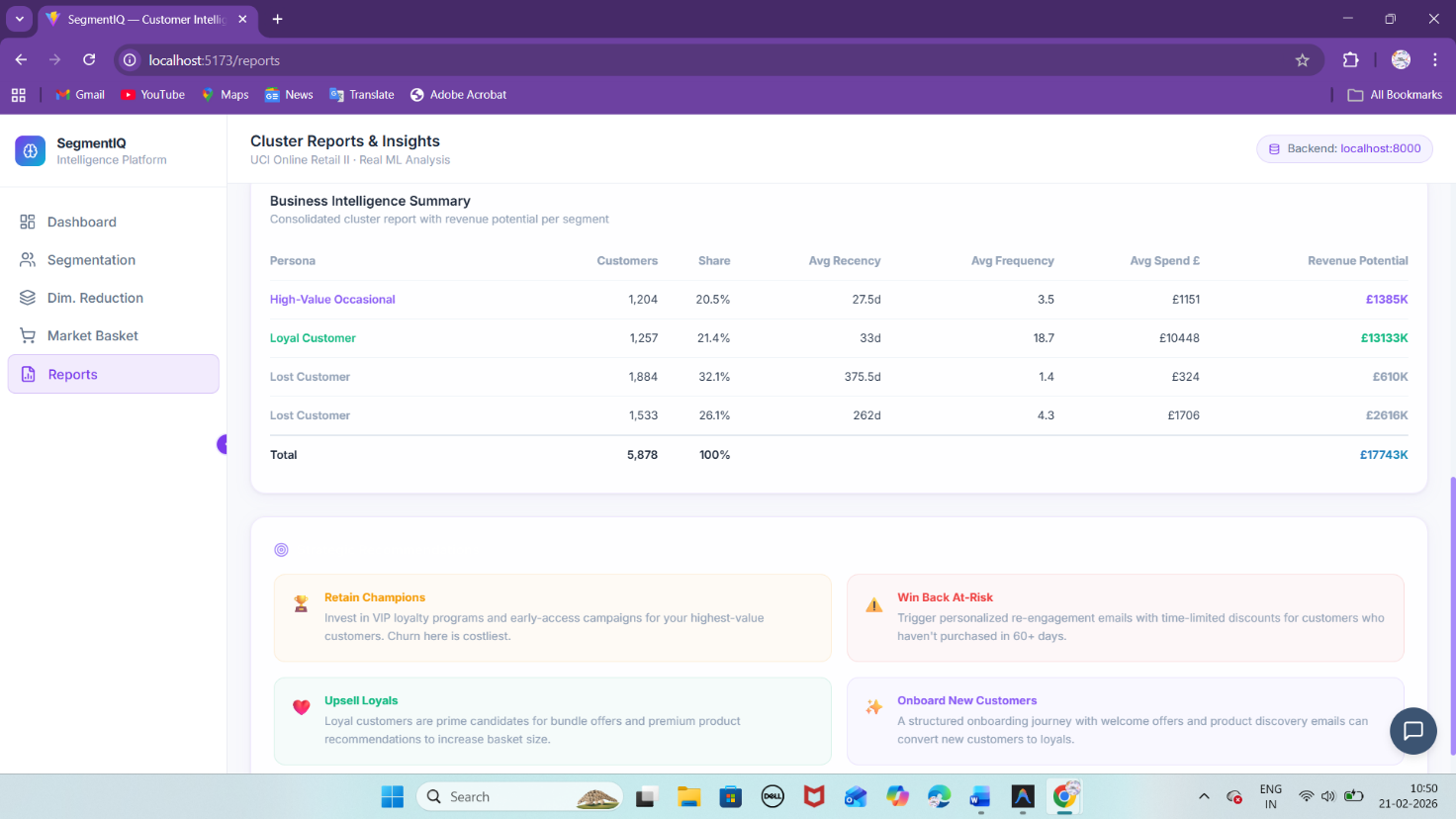
****

****

****

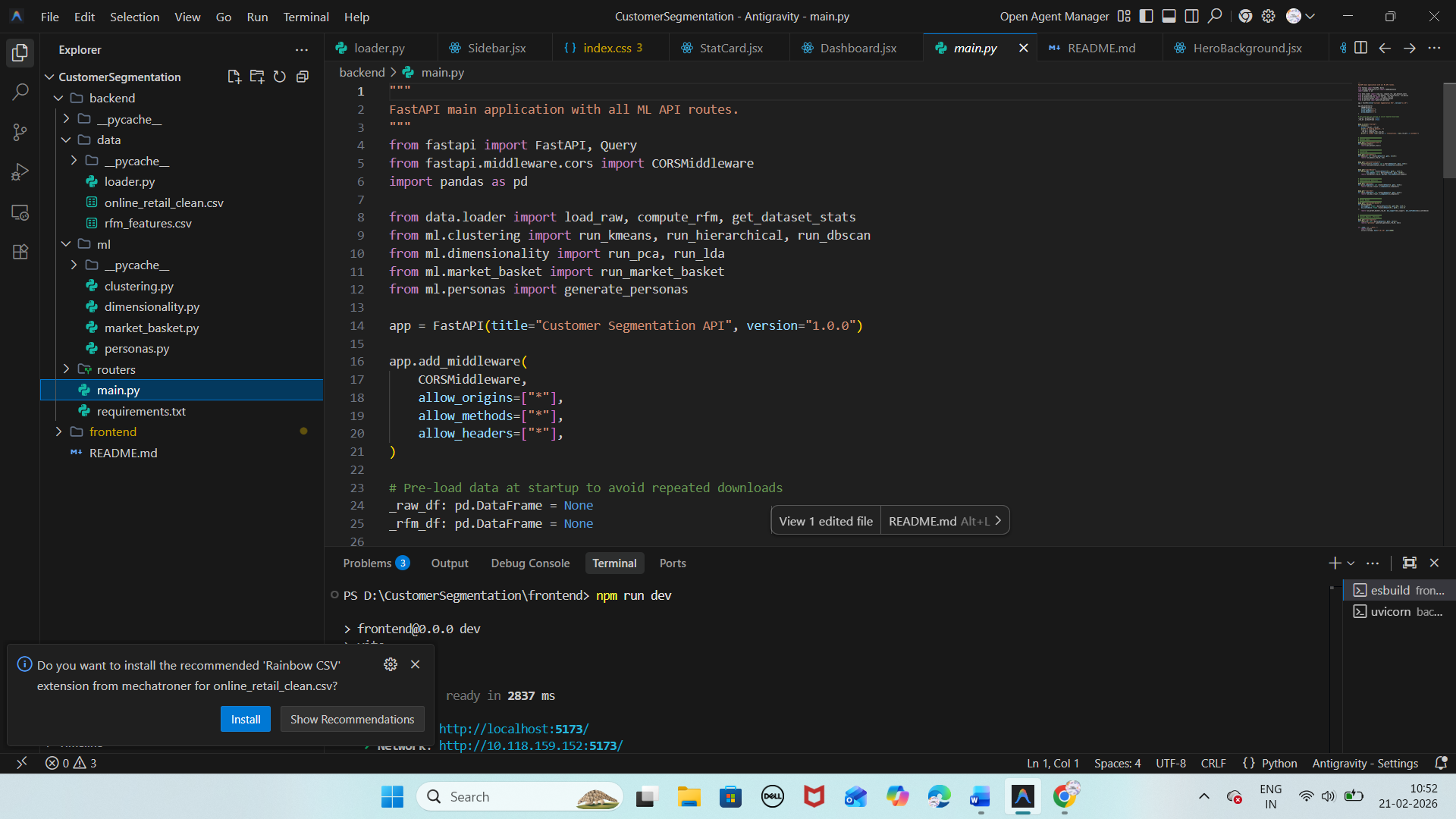
****

****

****

**FILE STRUCTURE:**

****

****

**Conclusion:**

SegmentIQ demonstrates how unsupervised machine learning can transform raw e-commerce transaction data into structured, actionable business intelligence. By combining clustering algorithms, dimensionality reduction, and association rule mining within a full-stack interactive dashboard, the platform enables automated customer segmentation and strategic marketing decision-making.

The modular and scalable architecture allows integration with new datasets, additional machine learning models, and cloud deployment, making SegmentIQ a practical and extensible analytics solution for modern e-commerce businesses.