## E-COMMERCE CUSTOMER SEGMENTATION

Submitted in partial fulfillment of the requirements of the course

**Business Intelligence Lab (ITL601)**

In

**T. E. Information Technology**

By

| **DURVA KADAM** | **221051** |
| --- | --- |
| **SHRUTI KADAM** | **221052** |
| **ANISH KALBHOR** | **221053** |
| **SHUBHANG KOLAPATE** | **221057** |

Supervisor:

**Ms. Shree Jaswal**

**Assistant Professor**



Department of Information Technology St. Francis Institute of Technology (Engineering College)

University of Mumbai 2024-2025

**CERTIFICATE**

This is to certify that the project entitled **“E-commerce Customer Segmentation”** is a Bonafide work of **Durva Kadam (45), Shruti Kadam (46), Anish Kalbhor (47) and Shubhang Kolapate (50)** submitted in partial fulfillment of the requirements of the course **Business Intelligence Lab (ITL601)**

##### Ms. Shree Jaswal

**Supervisor/Guide**

St. Francis Institute of Technology, Mumbai-400 103

**Department Of Information Technology**

A.Y. 2024-2025

Class: TE-ITA/B, Semester: VI

Subject: **Business Intelligence Lab**

**Business Intelligence Mini Project**

1. **Aim:** Develop a Business IntelligenceMini Project for a particular case study

**2. Objectives:** After study of this experiment, the students will be able to develop mini project

**3.Outcomes:**

**CO6:** Apply BI to solve practical problems: Analyze the problem domain, use the data collected in enterprise apply the appropriate data mining technique, interpret and visualize the results and provide decision support

**4. Prerequisite:** Study of DM&BI Tools.

**5. Requirements:** Personal Computer, Windows XP operating system/Windows 7, Internet Connection, Microsoft Word, WEKA tool, Orange tool, BI Tool.

**6. Theory: Nil**

**7. Laboratory Exercise:**

Each group select one case study for this. A BI report (Initial pages shared separately) must be prepared outlining the following steps:

1. Write the problem statement for your case study
2. Draw star schema and snowflake schema
3. Give dataset details, identifying which data mining task is needed
4. Download and use a standard data mining dataset available for the problem. Some links for data mining datasets are WEKA site, UCI Machine Learning Repository, KDD site, KDD Cup etc.
5. Implement the data mining algorithm using Weka and Orange
6. Interpret and visualize the results using BI tool like Qlikview & Tableau
7. After interpretation clearly provide the BI decision that is to be taken

**8. Post-Experiment Exercise:**

1. **Conclusion:**
   * Summary of mini project

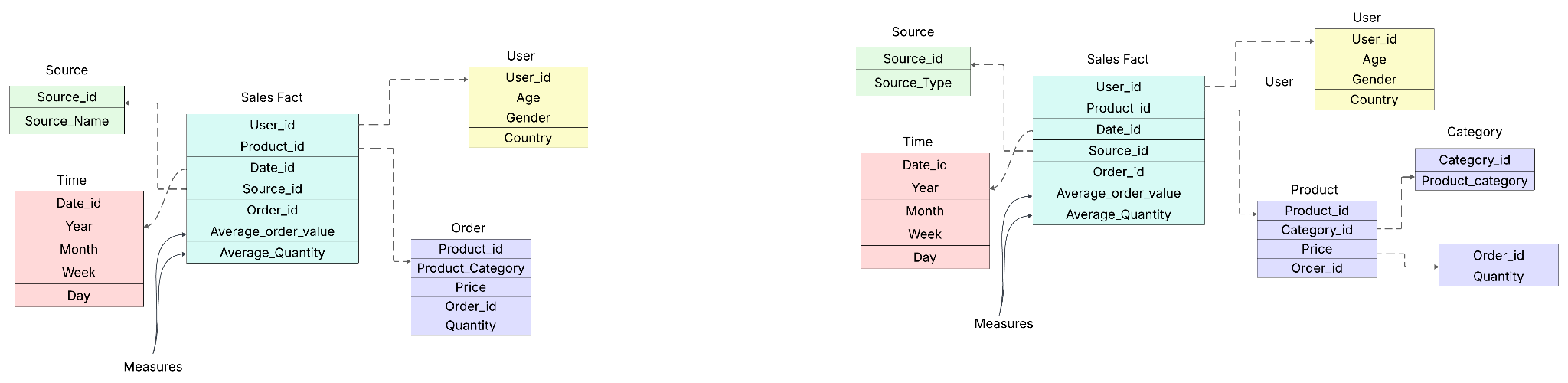
**9. Reference:** Business Intelligence: Data Mining and Optimization for Decision Making by Carlo Vercellis, Wiley India Publication

**PROBLEM STATEMENT:**

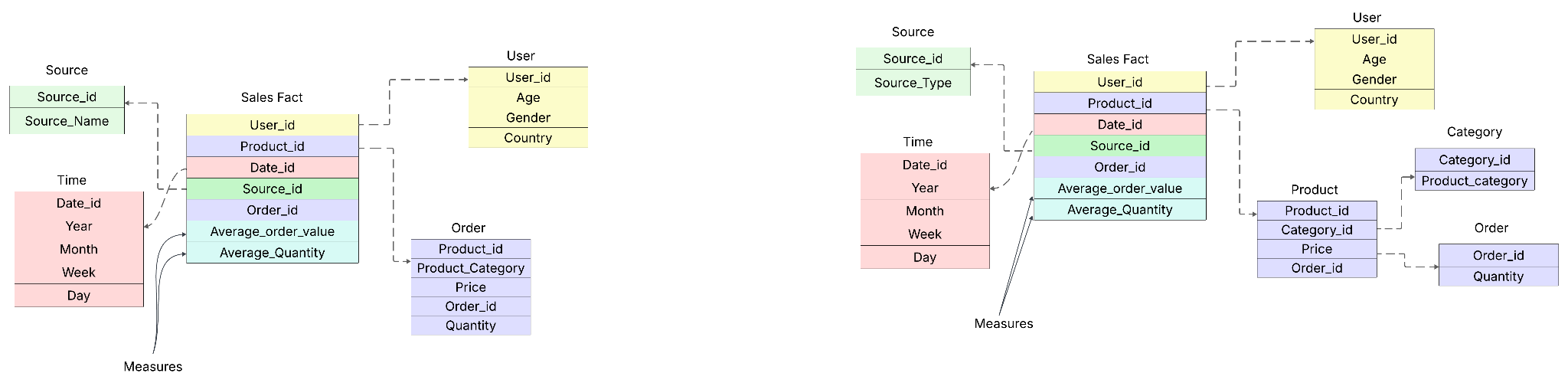
With the rise of e-commerce platforms, understanding customer behavior and optimizing sales strategies are critical to success. This dataset provides synthetic but realistic data about customer demographics, purchasing patterns, and product preferences on an eCommerce clothing platform. The task is to analyze this data to identify key trends, such as the most lucrative customer segments, high-performing product categories, and seasonal sales patterns. Such insights can help design targeted marketing campaigns, improve inventory management, and enhance customer experience.

Additionally, the data offers an opportunity to predict future customer behavior based on historical patterns. By building predictive models, the platform can recommend products, anticipate inventory needs, and identify potential churn risks. The ultimate goal is to leverage these insights to improve operational efficiency, maximize revenue, and create a personalized shopping experience for customers.

**STAR SCHEMA:**



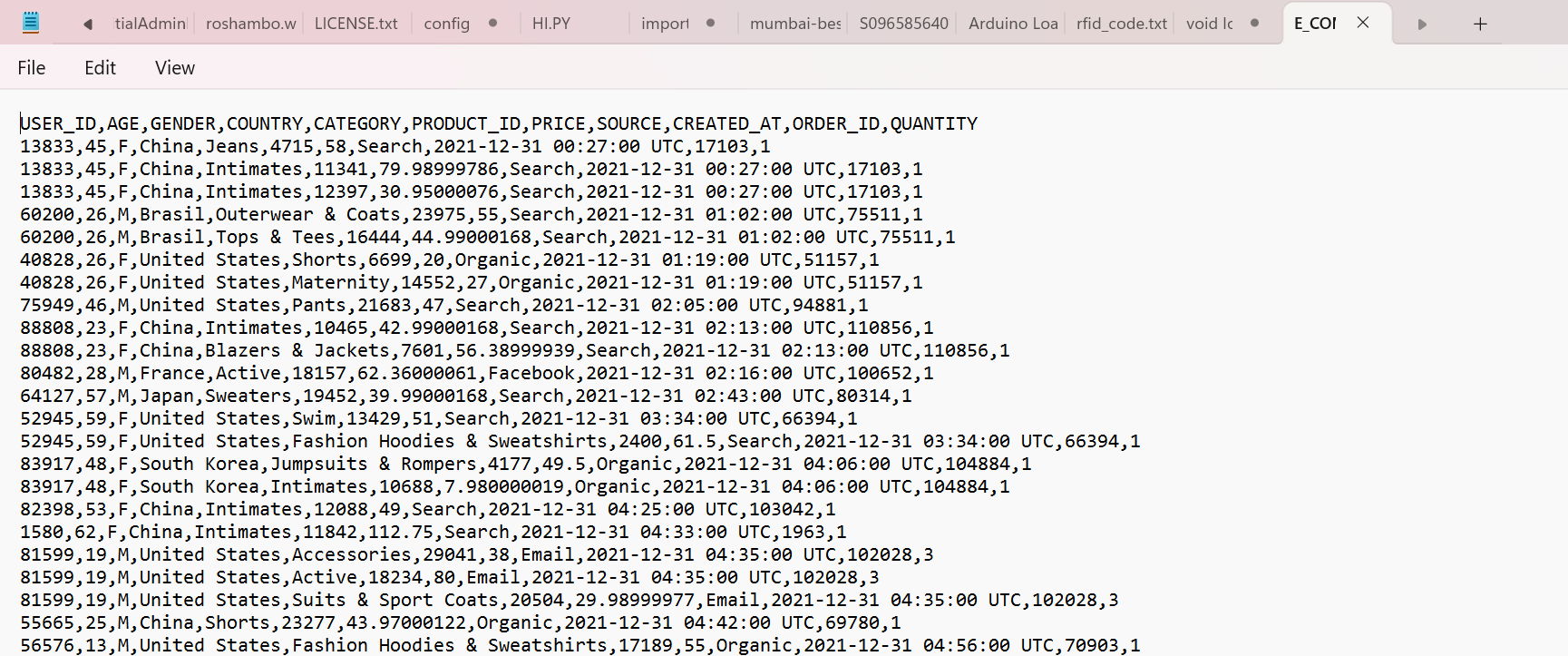
**\\**

**SNOWFLAKE SCHEMA:**

**Dataset Details and Data Mining Task:**

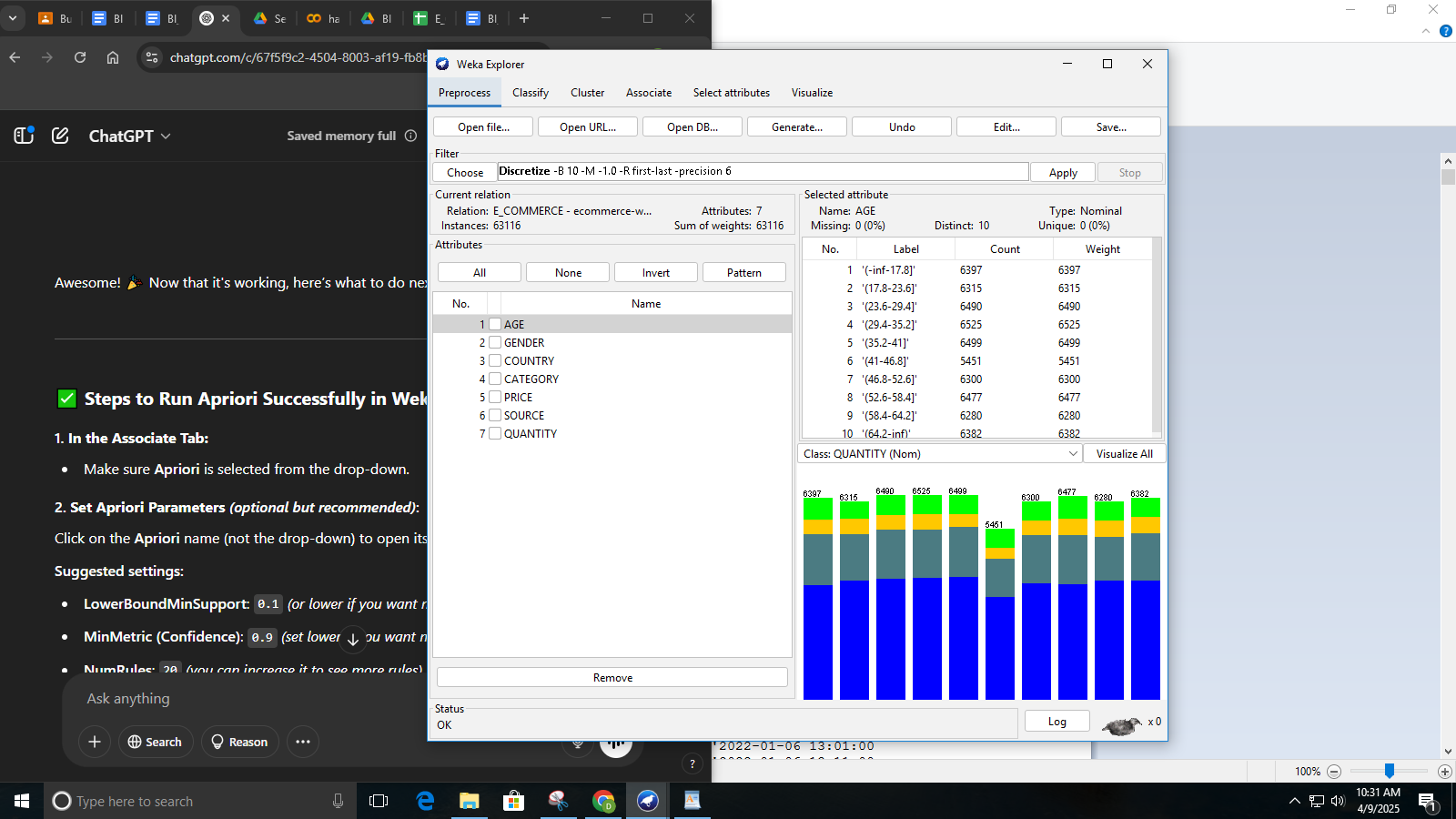
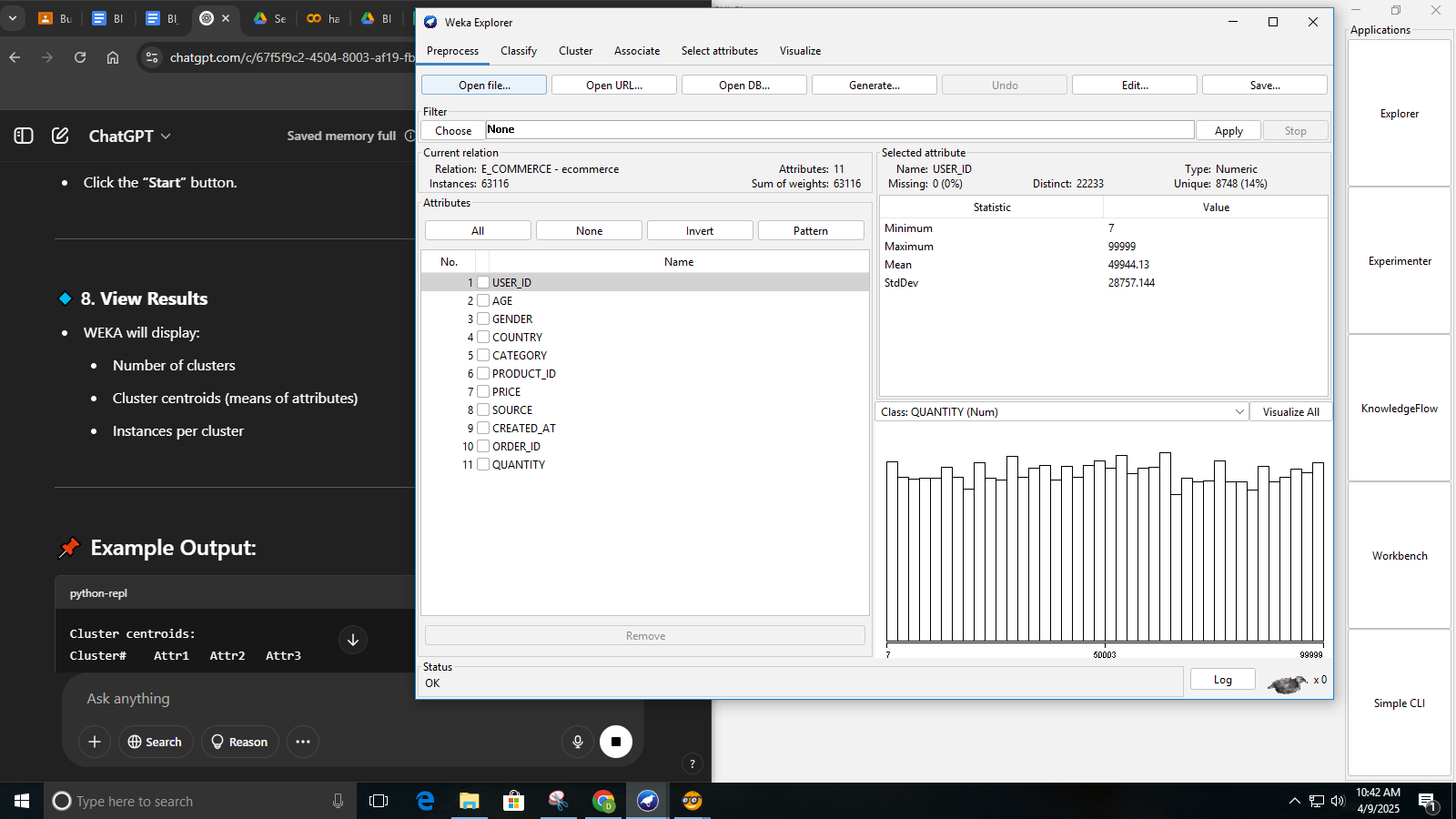
The dataset used is an E-commerce customer transactional dataset sourced from Kaggle, containing details such as User\_ID, Gender, Age, Product\_ID, Category, Quantity, Price, and Purchase Date. It reflects real-world shopping behavior across various product categories and user demographics.

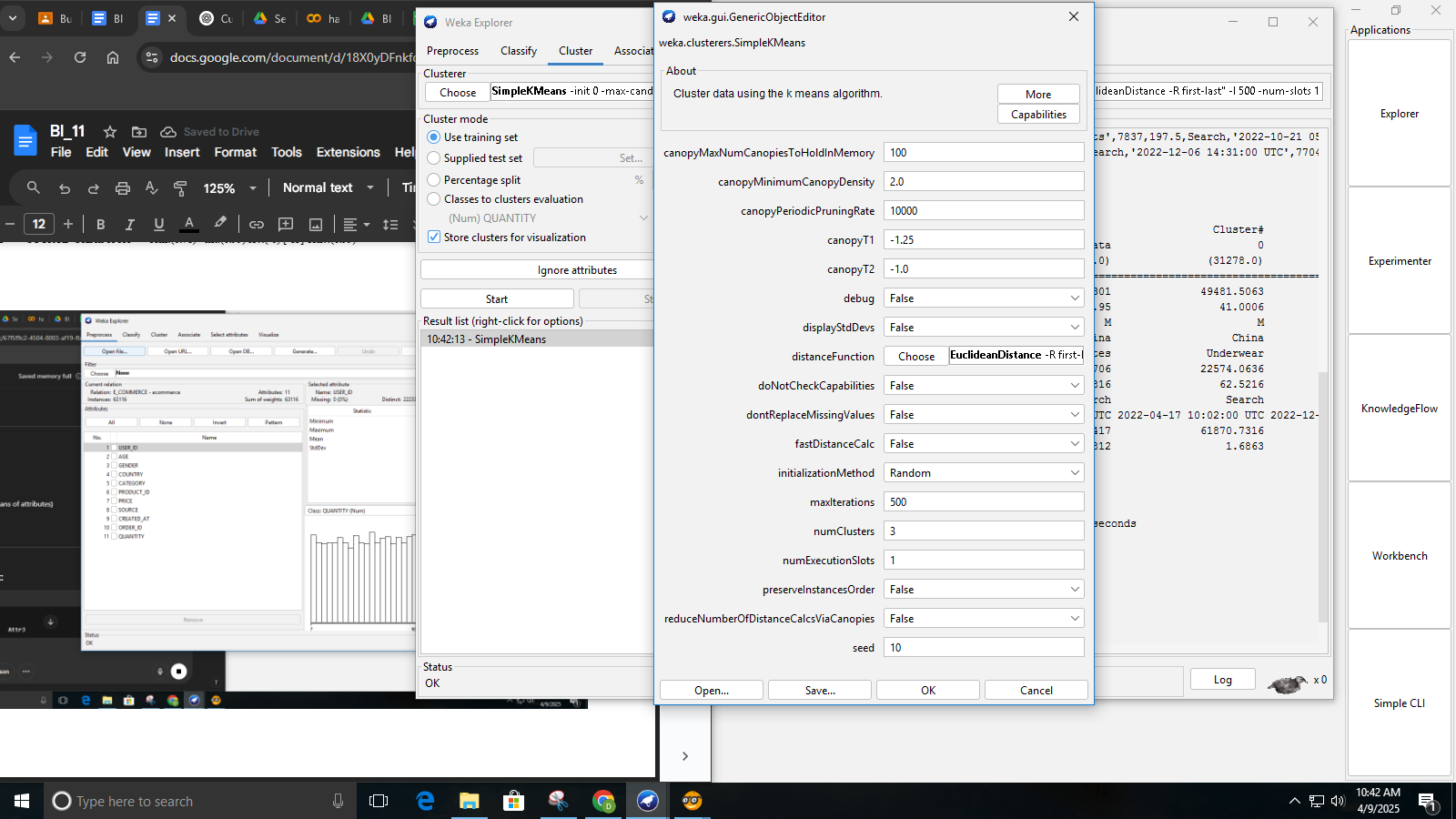
The appropriate data mining task for this dataset is Clustering, specifically Customer Segmentation using the K-Means algorithm. This task helps group users based on similar purchase behavior and demographic features, allowing the business to identify key customer segments for personalized marketing and decision-making.

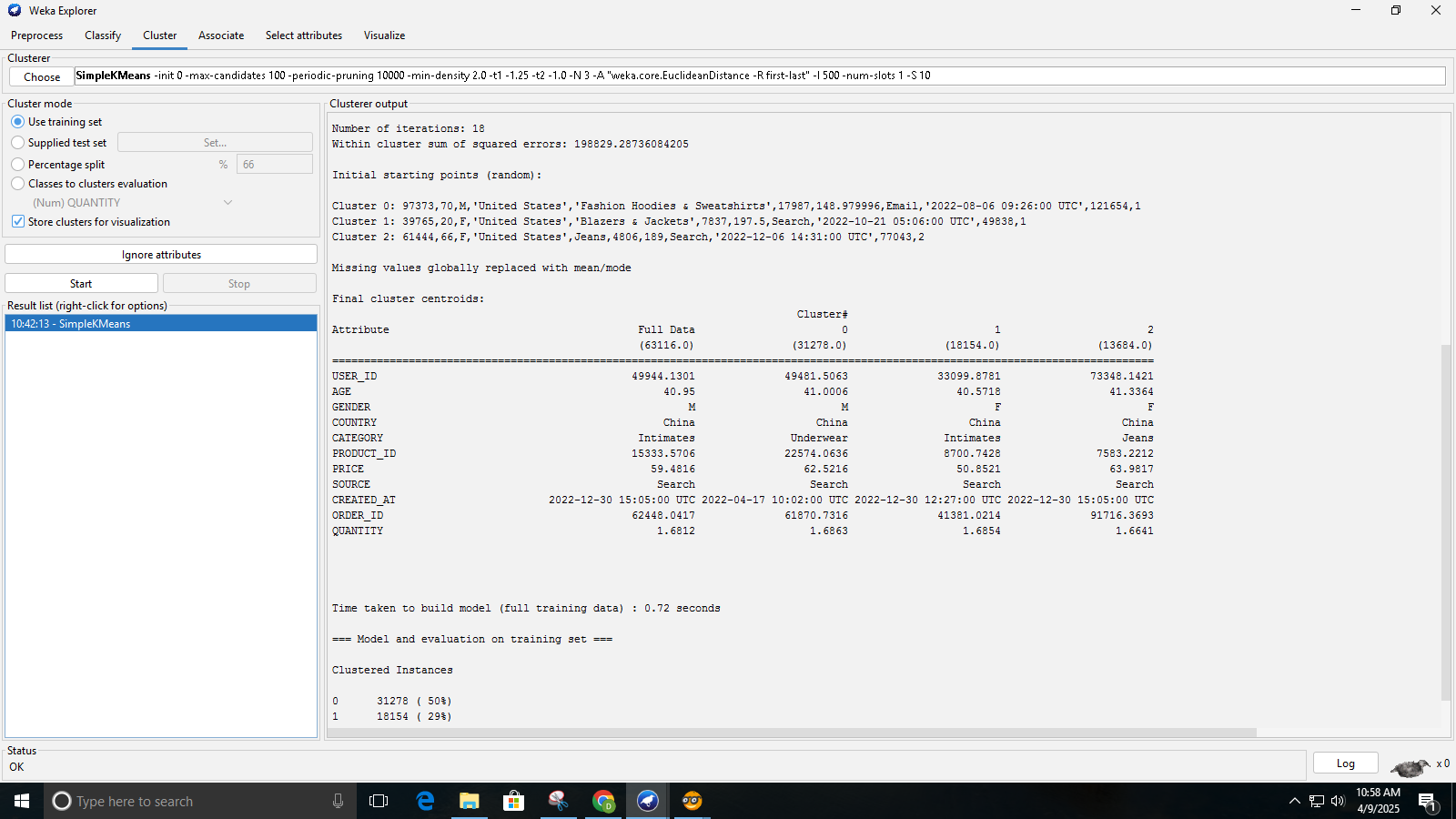
**Downloading the dataset from Kaggle and opening it in Notepad.**

**WEKA:**

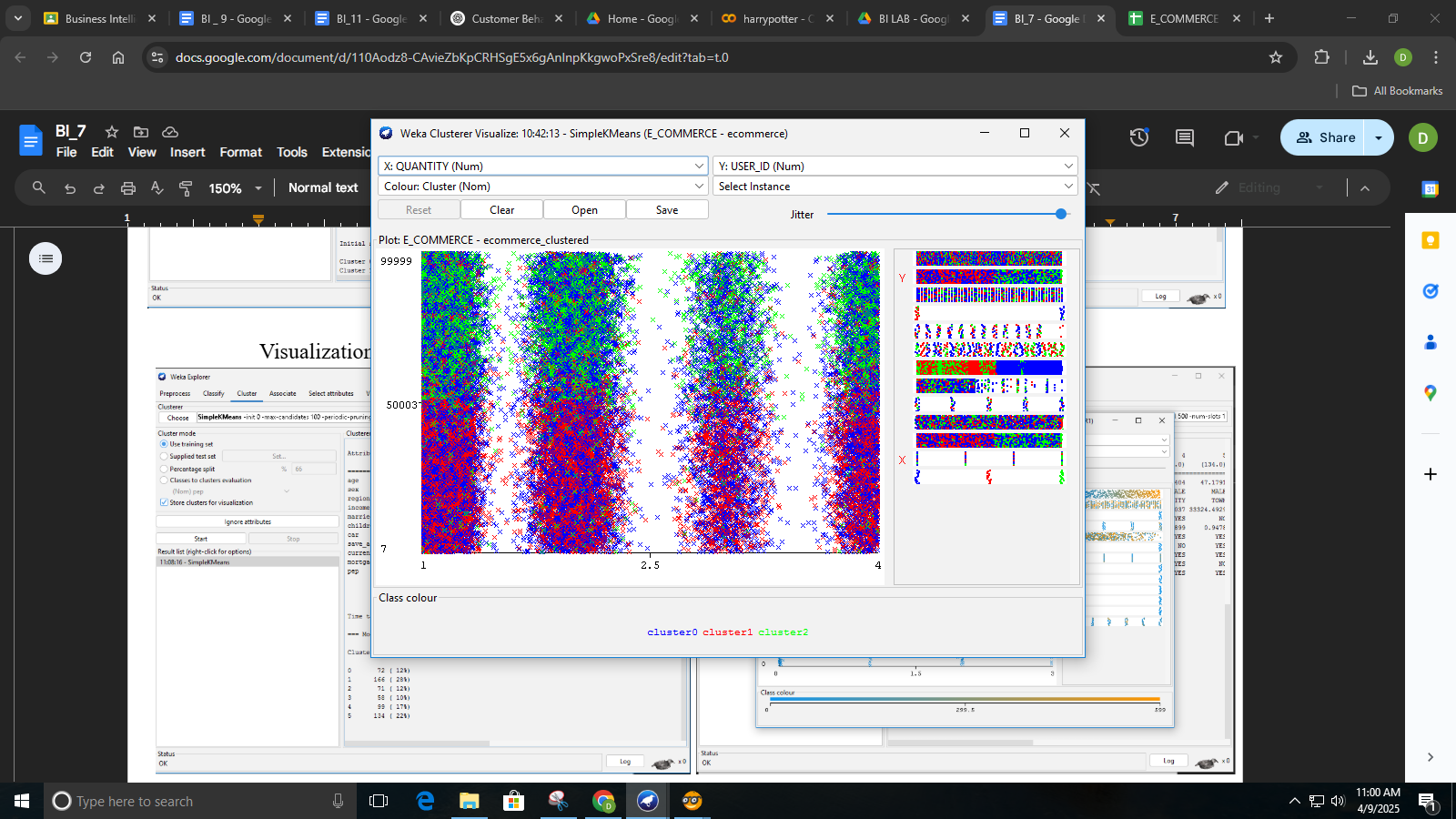
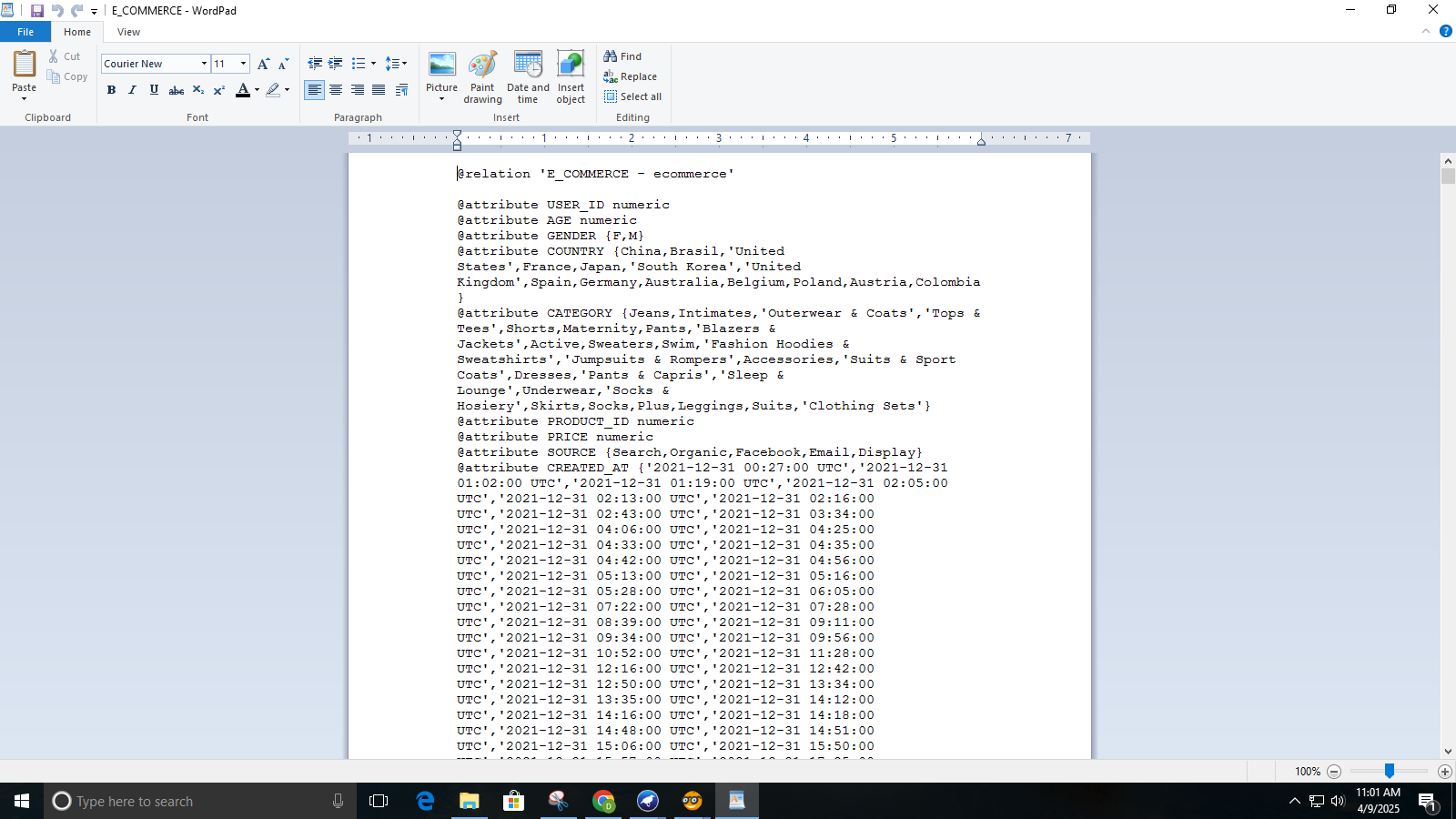
Loading dataset in WEKA

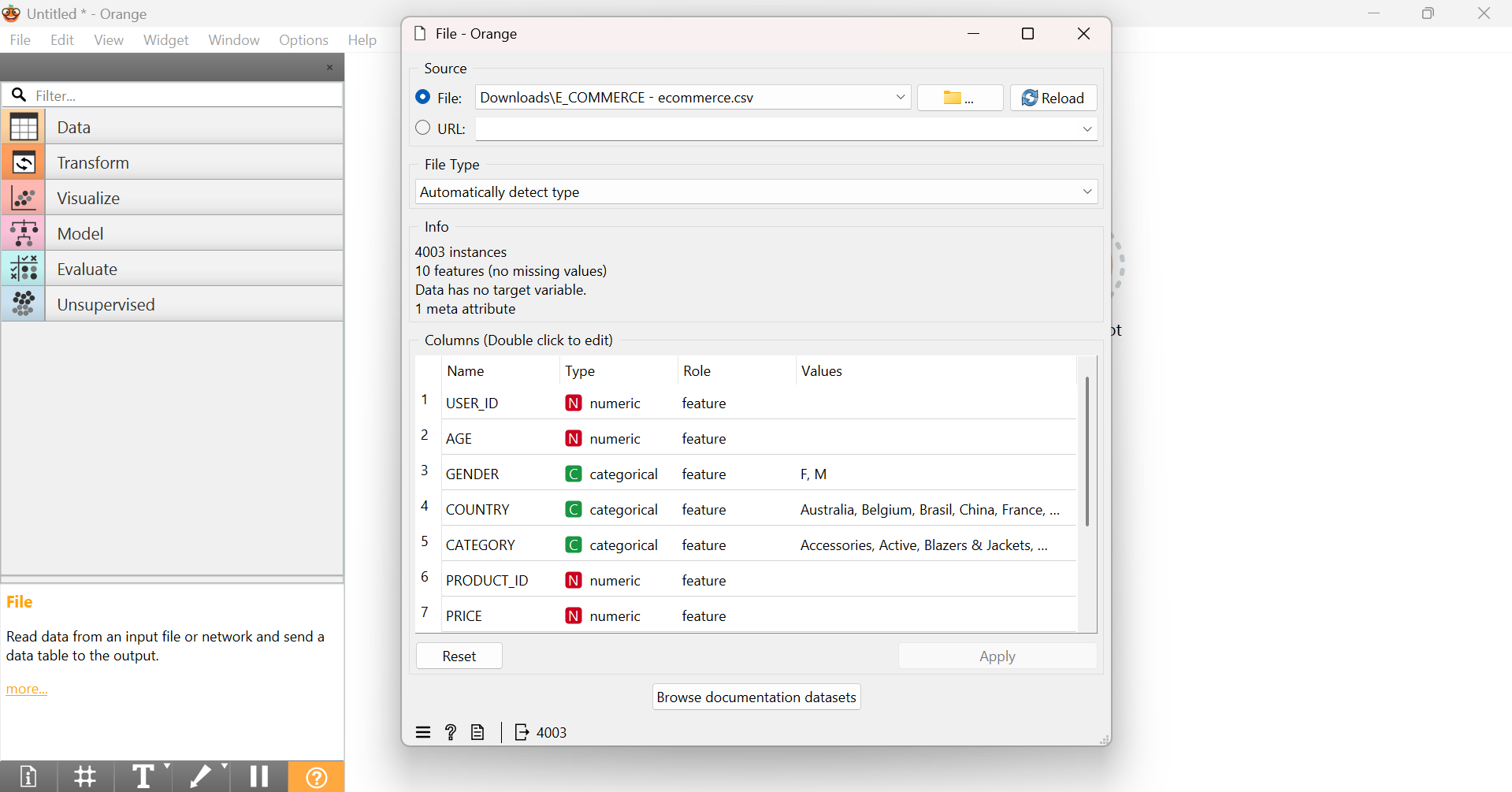
Applying K-means algorithm with 3 clusters:

We see that it ran for 18 iterations and created 3 clusters and their summaries, displaying how data points were grouped based on selected attributes like GENDER, PRODUCT\_ID, and CREATED\_AT.

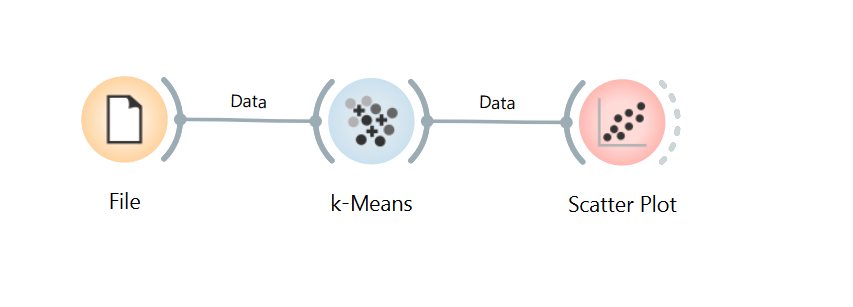
We see the result of KMeans clustering on an e-commerce dataset, with clusters. The X-axis represents QUANTITY and the Y-axis shows USER\_ID, revealing how different users are grouped based on purchase quantity. The dense overlapping areas suggest that the clusters are not well-separated, possibly indicating similar user behavior across clusters.

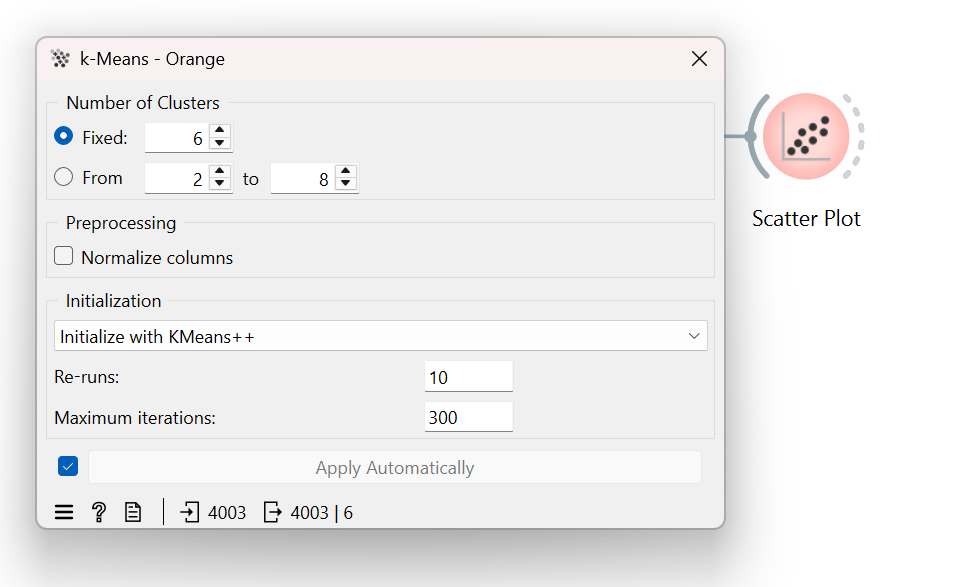
**ORANGE:**

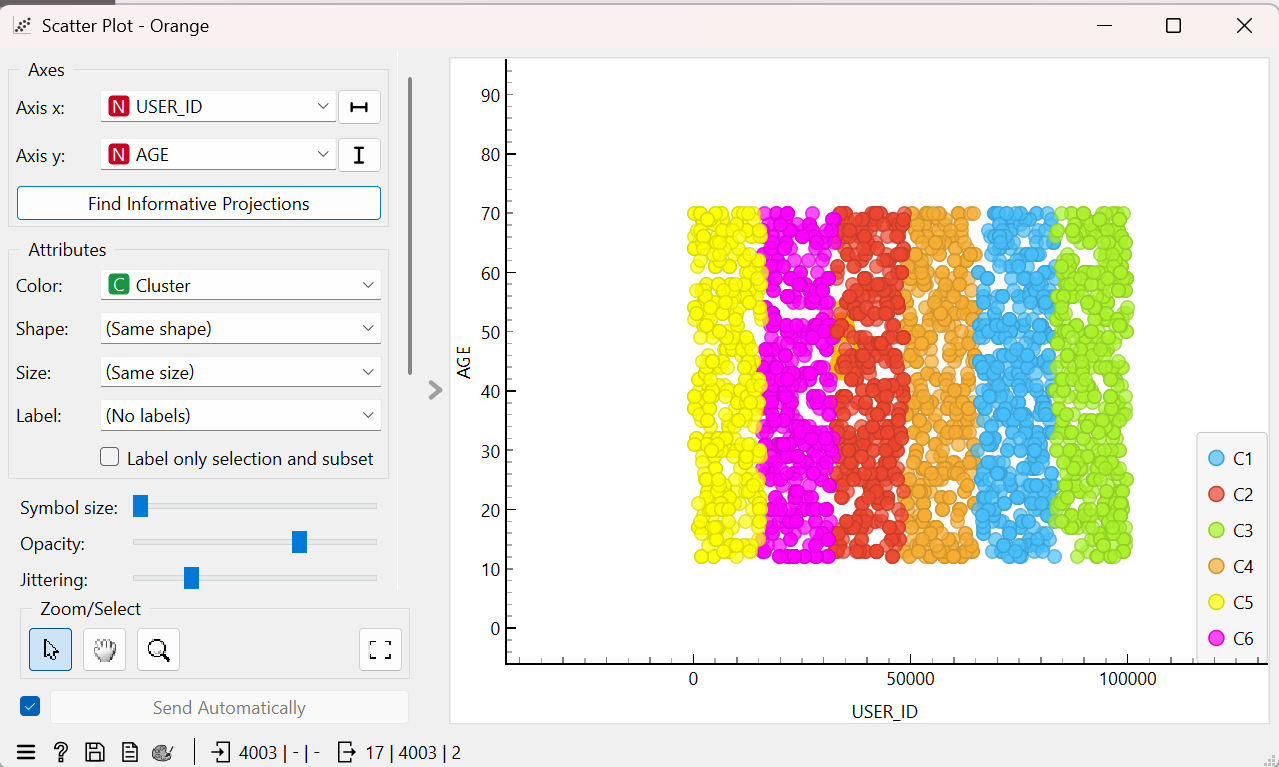
Loading the dataset in Orange:  


Customer Segmentation Workflow Using K-Means Clustering



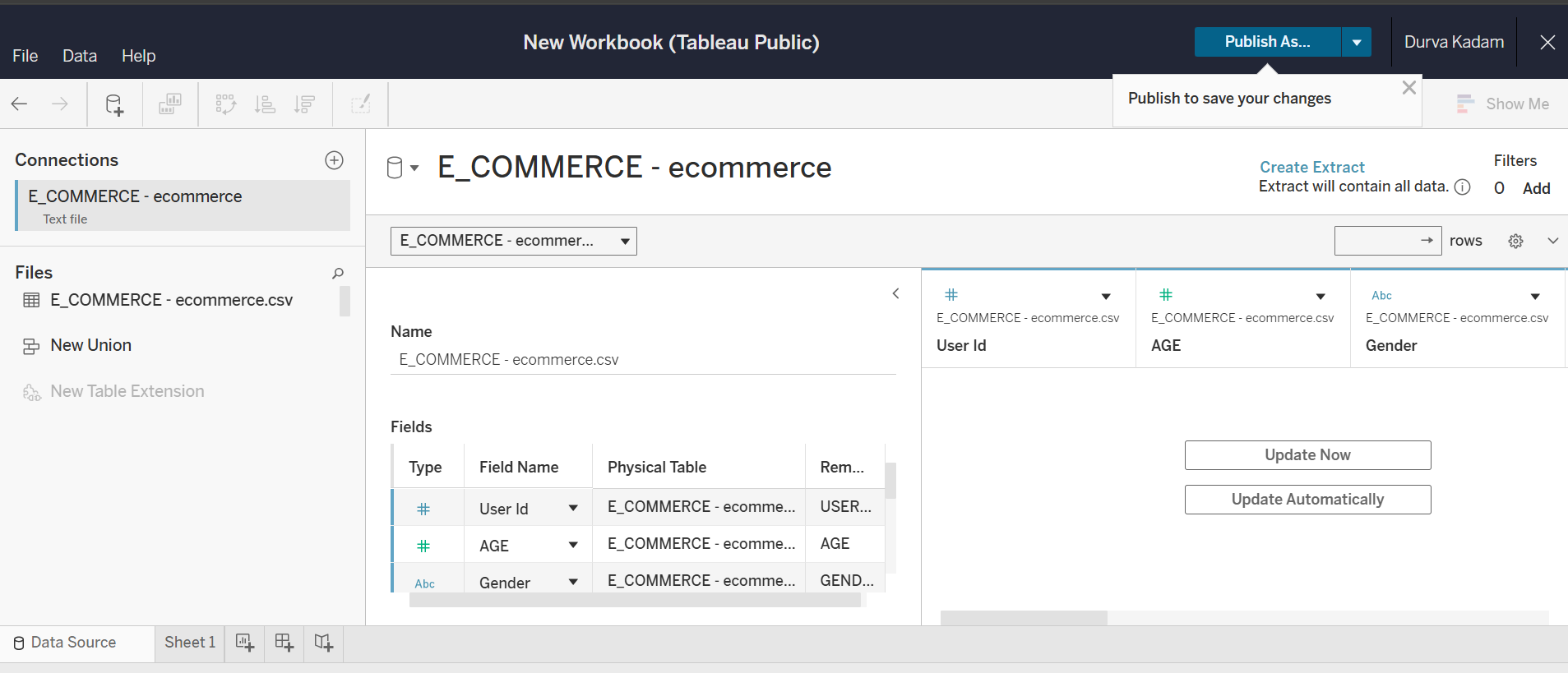
Applying K-means algorithm:

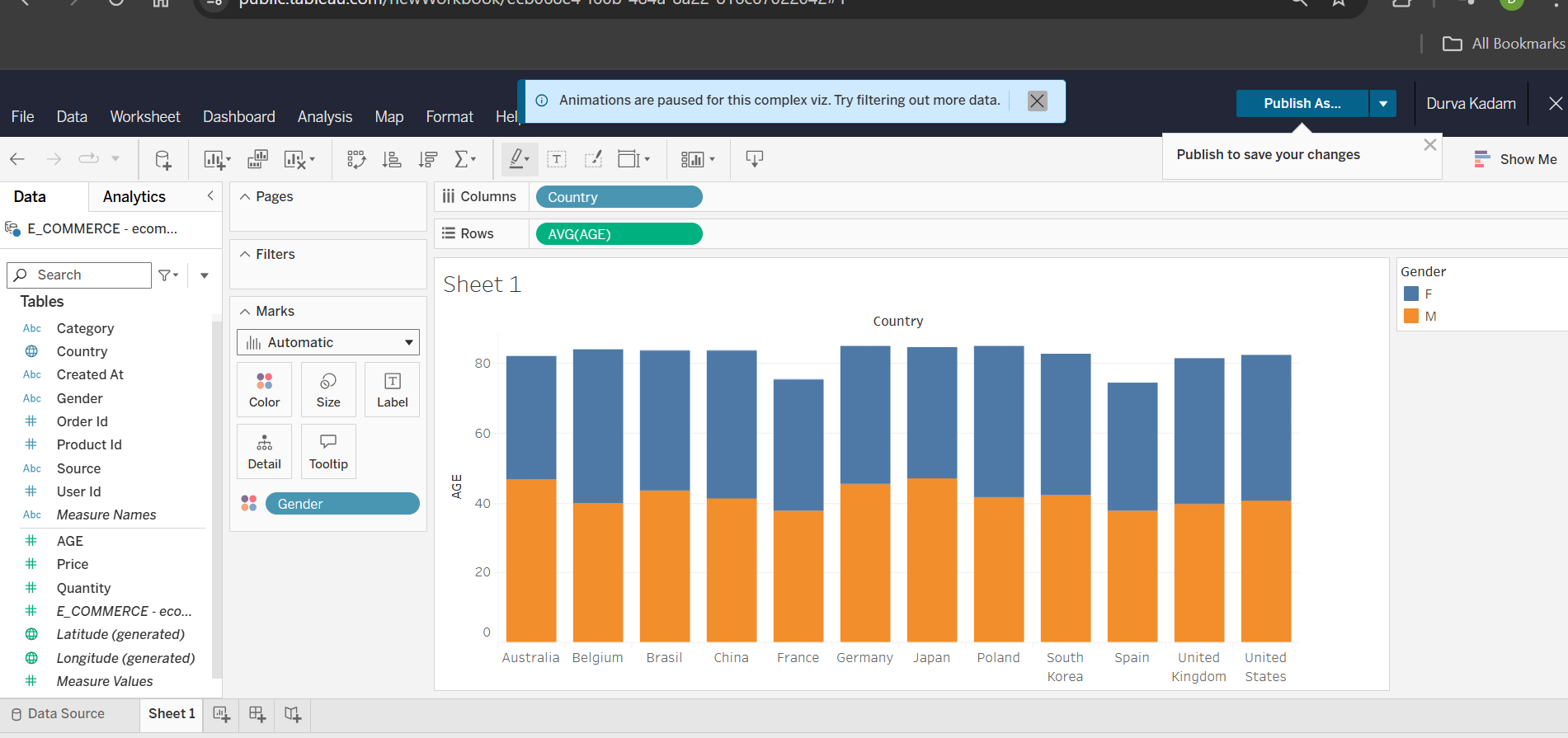


The scatter plot shows users grouped into six different clusters based on their age and user ID. Each color represents a cluster, and we can see that users are somewhat grouped together but with some overlap. This means users share some common traits, but the separation isn’t very clear-cut.  
  


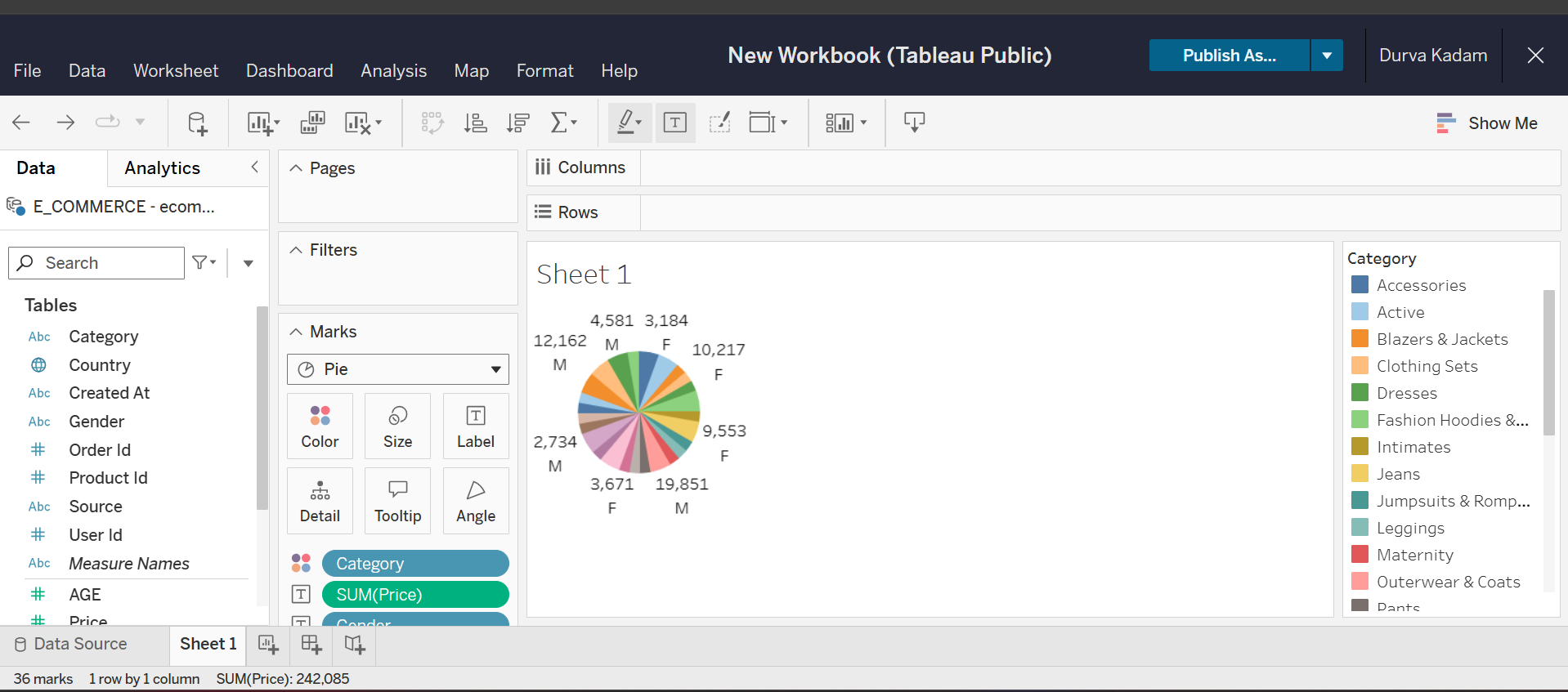
**TABLEAU:**

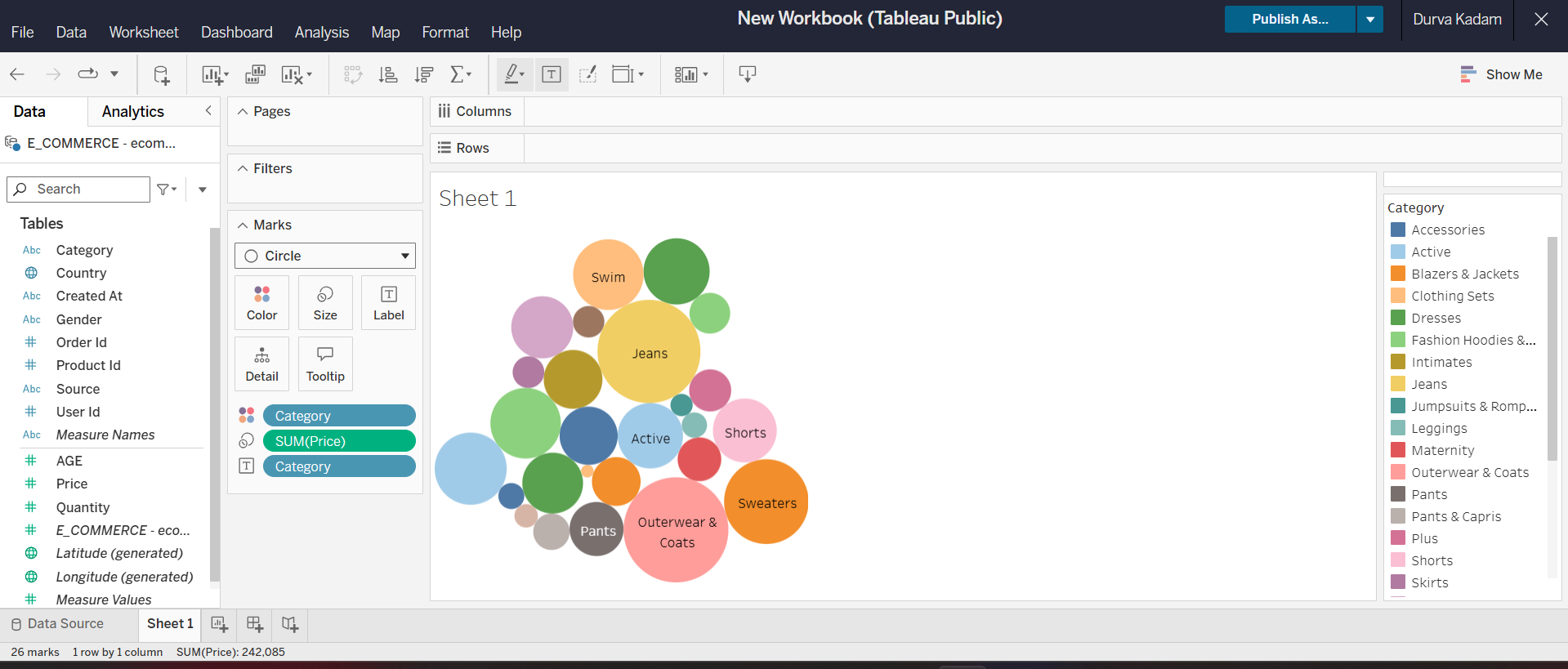
Loading dataset in tableau:



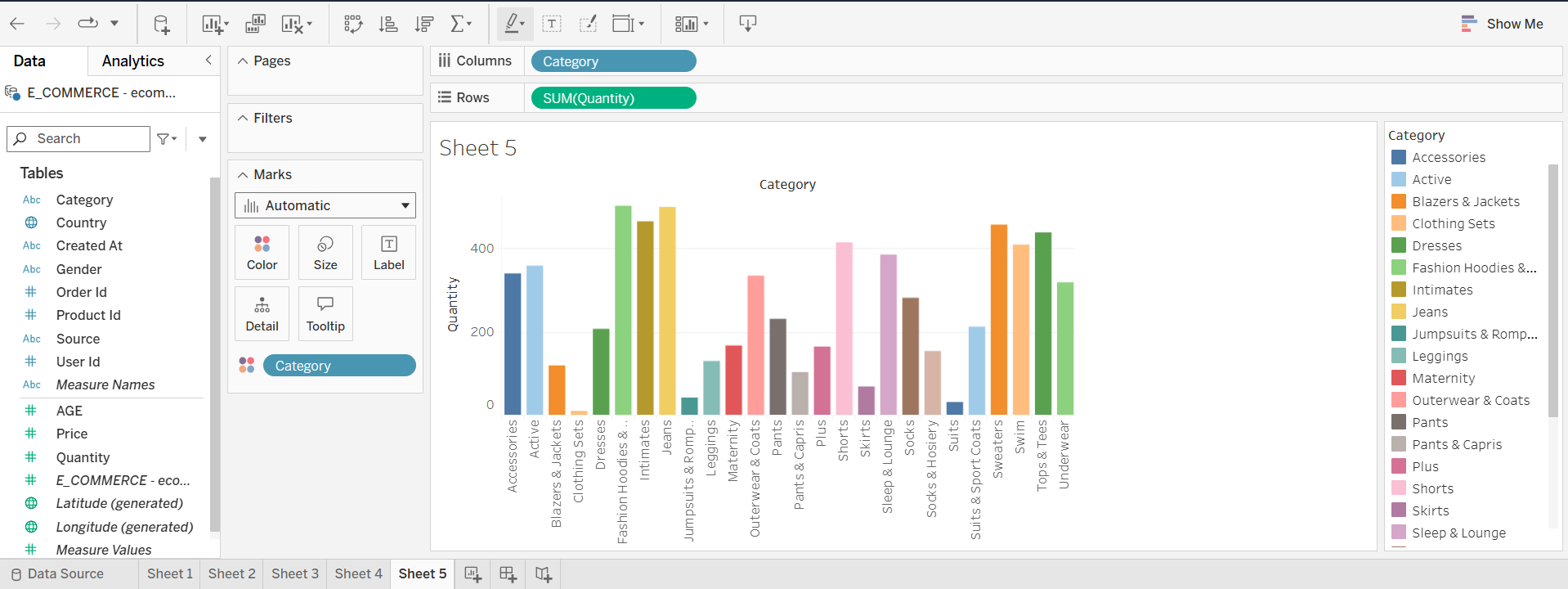
This chart highlights age and gender differences across countries, showing that females generally have a higher average age than males. Countries like Germany and Japan have older users, while France and Spain have younger ones. These insights help in tailoring products and marketing strategies to match each country's demographic profile.

This pie chart shows the distribution of total price across different product categories, segmented by gender. The highest contribution comes from males in one dominant category, while other combinations contribute smaller portions. These insights help identify the most valuable customer segments for focused marketing and sales efforts.

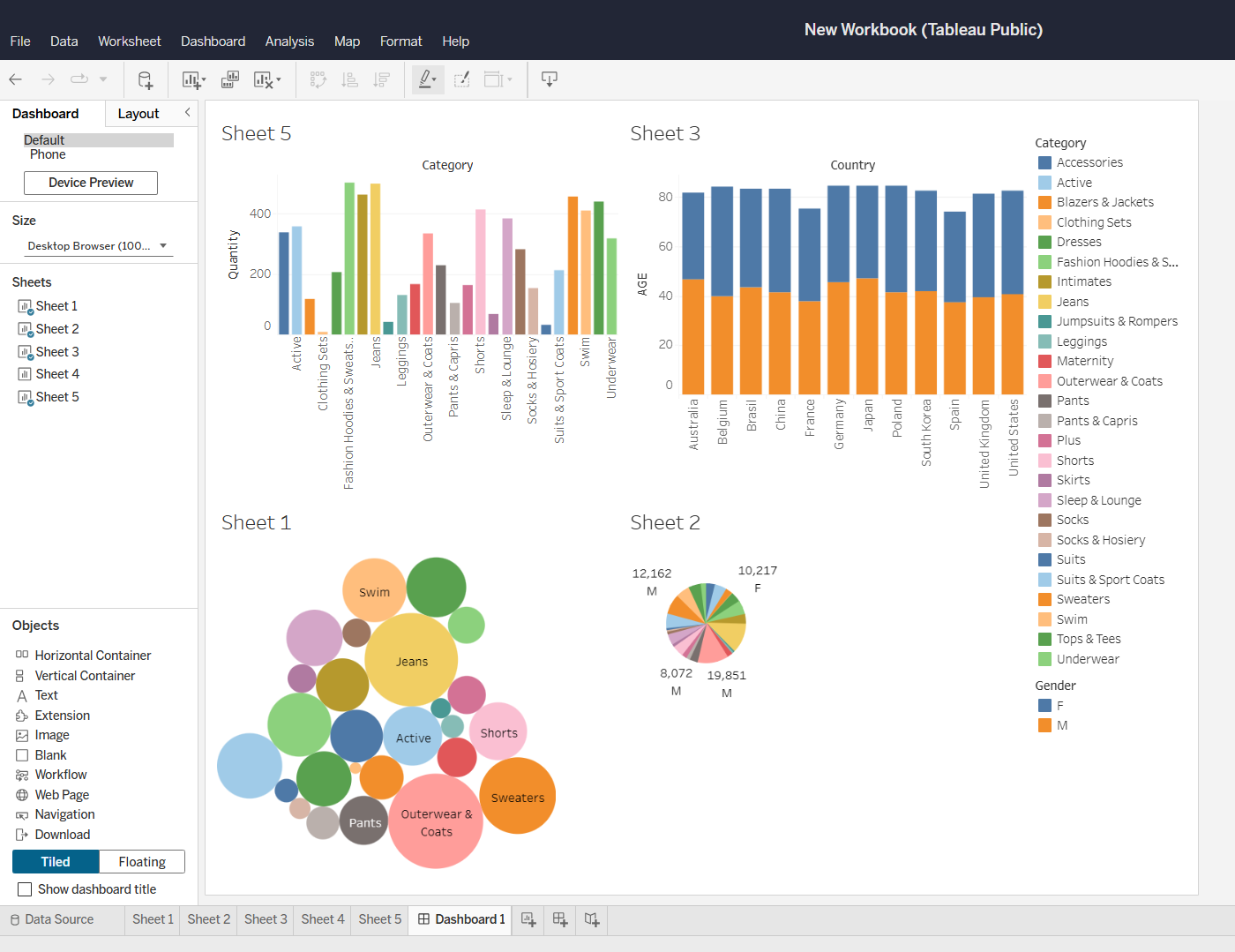


The packed bubble chart shows that Jeans and Outerwear & Coats generate the most revenue. Smaller bubbles indicate less popular or lower-priced categories, highlighting key focus areas for sales and marketing.

The bar chart reveals that Jeans and Intimates have the highest quantities sold, reflecting strong customer demand and popularity in these segments. On the other hand, categories like Clothing Sets and Suits show much lower sales, which could indicate lower consumer interest or possible inventory limitations.

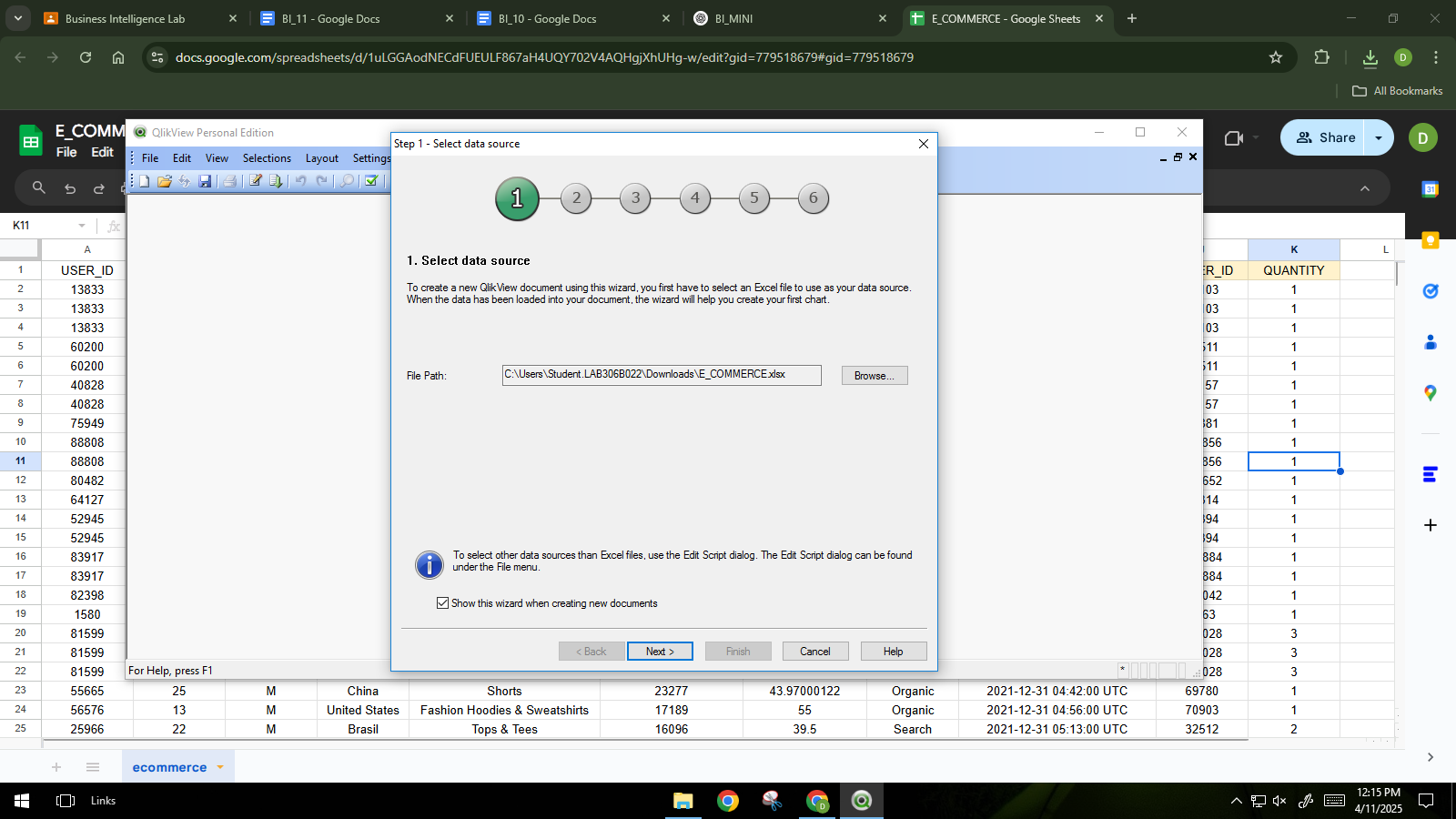


Dashboard:

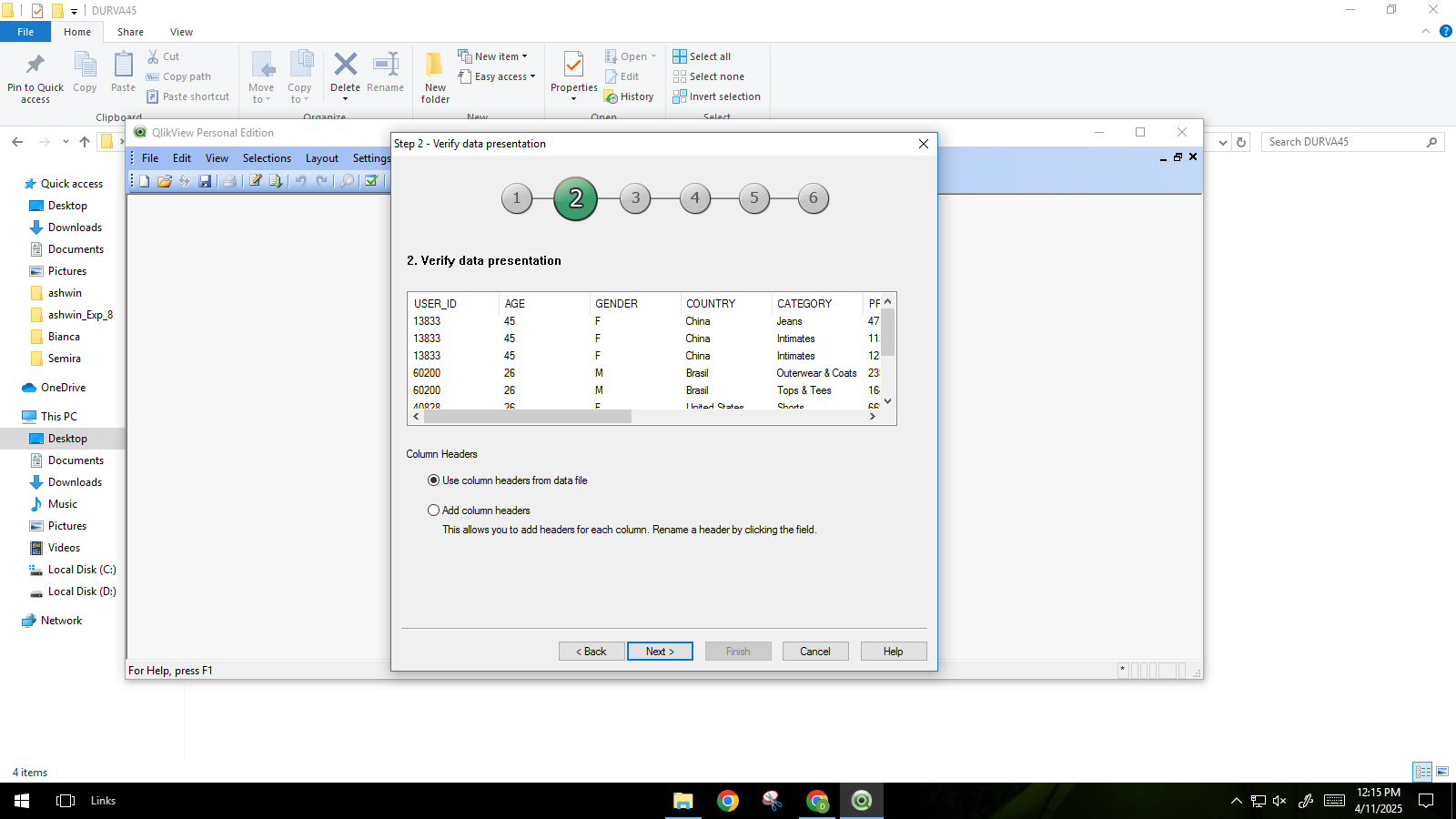


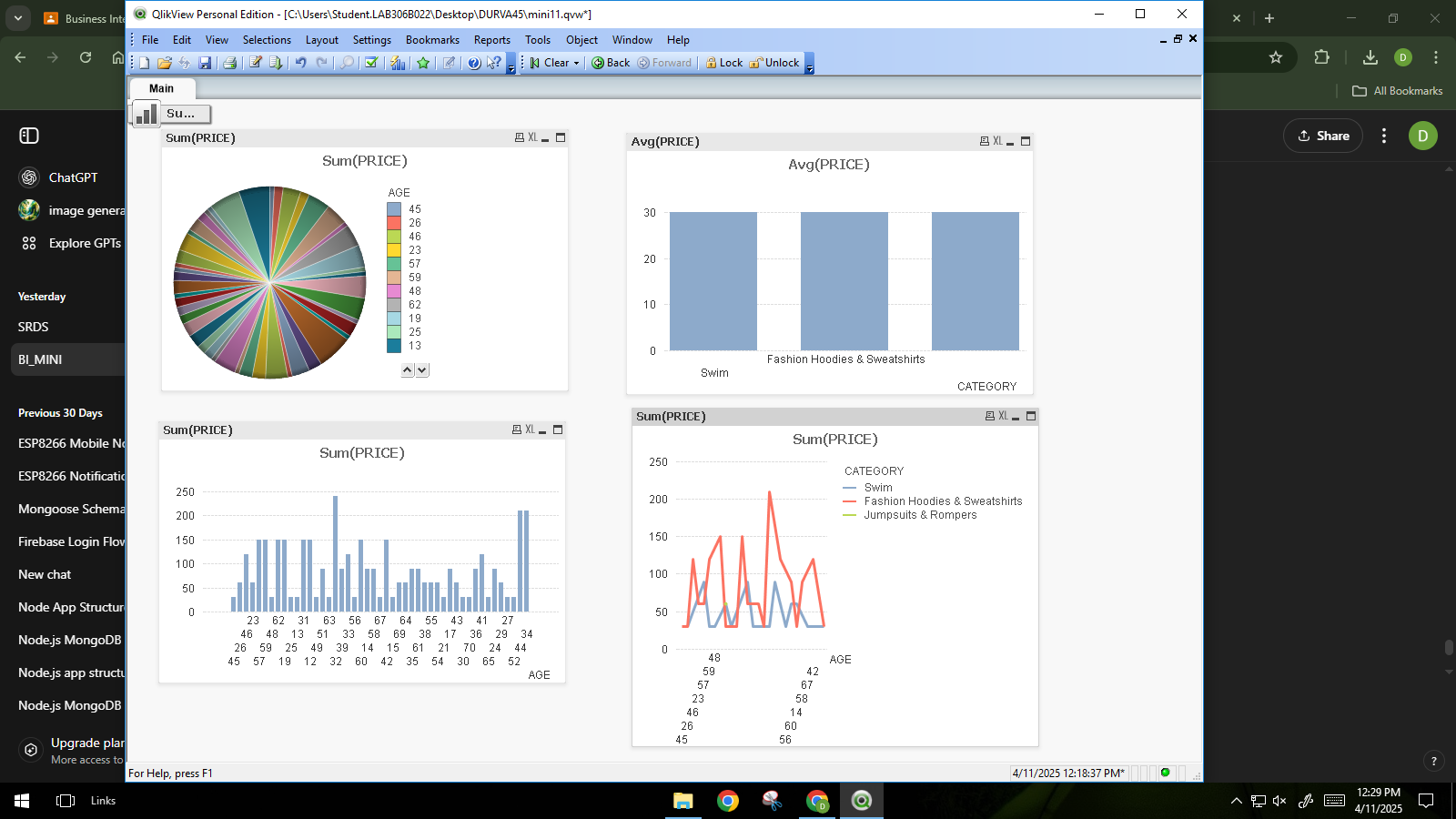
**QLIKVIEW:**

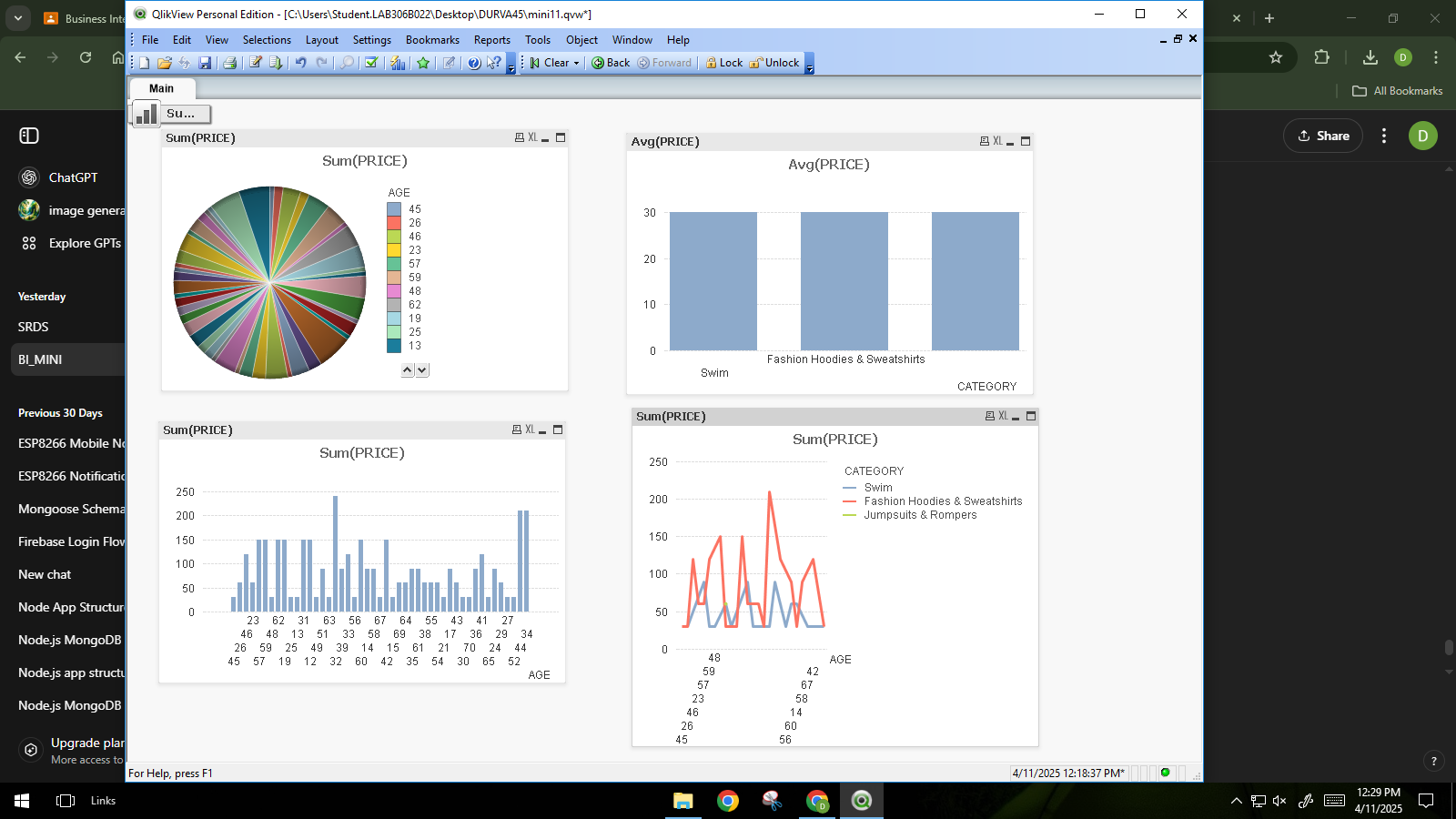
Loading the dataset in Qlikview



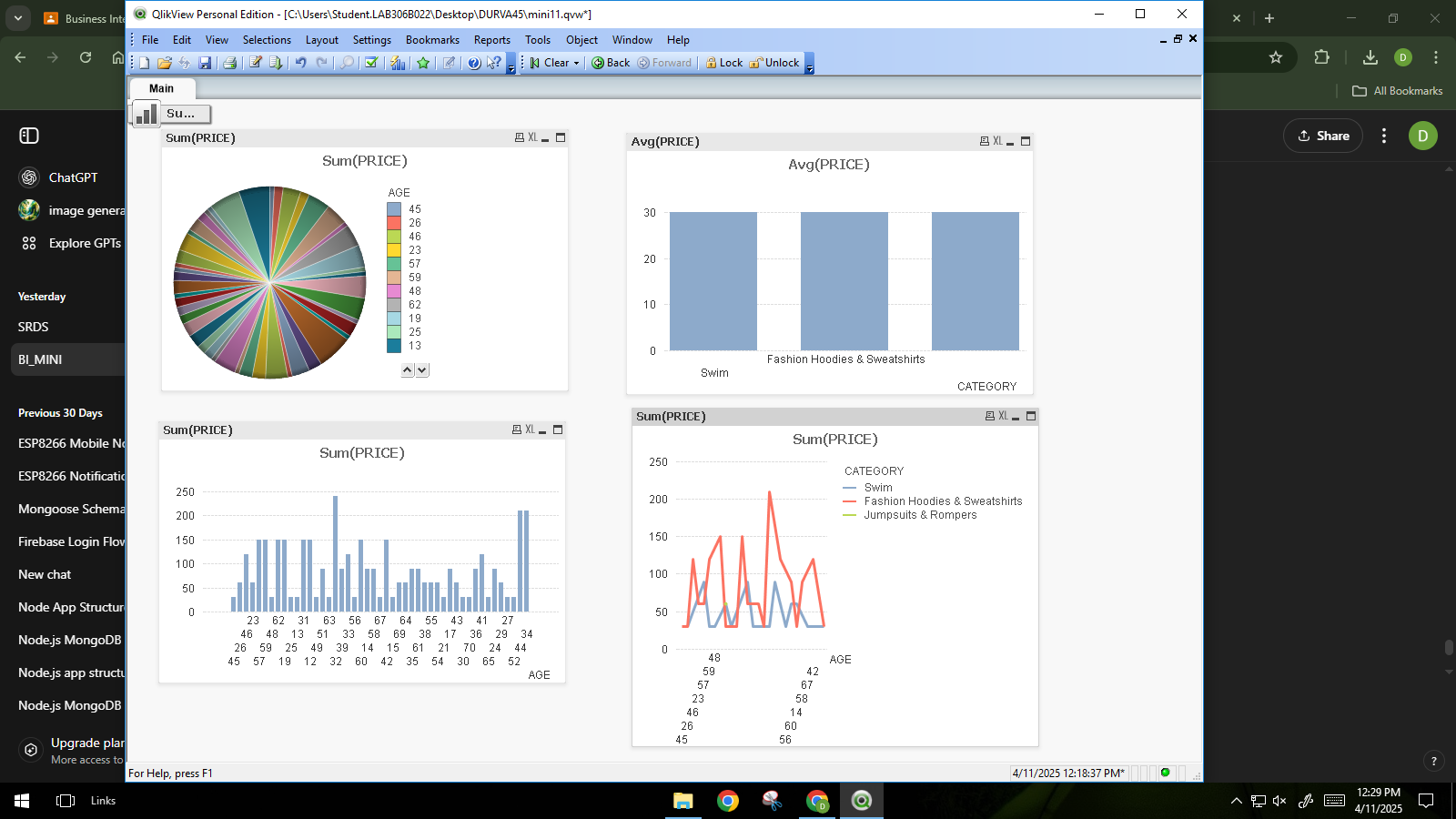
Reviewing imported dataset



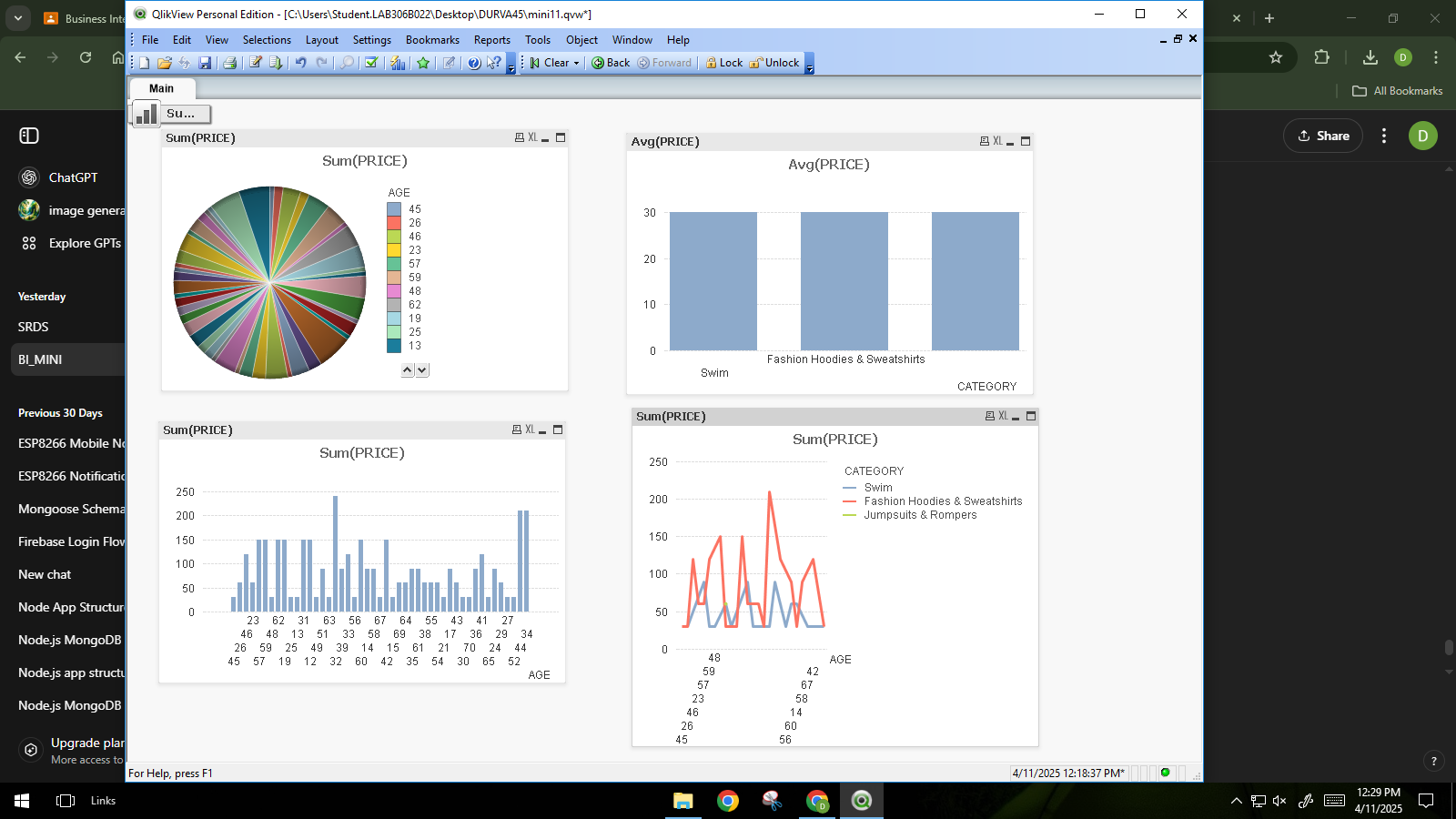
The pie chart displays the total revenue segmented by customer age. While the chart is visually diverse with many age groups represented, a few age segments like 45, 46, and 26 appear to dominate, suggesting these age groups contribute more significantly to overall sales. 

The bar chart shows that the average price is consistent across the categories Swim, Fashion, and Hoodies & Sweatshirts, suggesting uniform pricing strategies in these segments.

The bar chart shows total revenue (Sum of PRICE) by age group, with ages 56 and 27 contributing the highest. This indicates key age segments driving sales and potential focus areas.

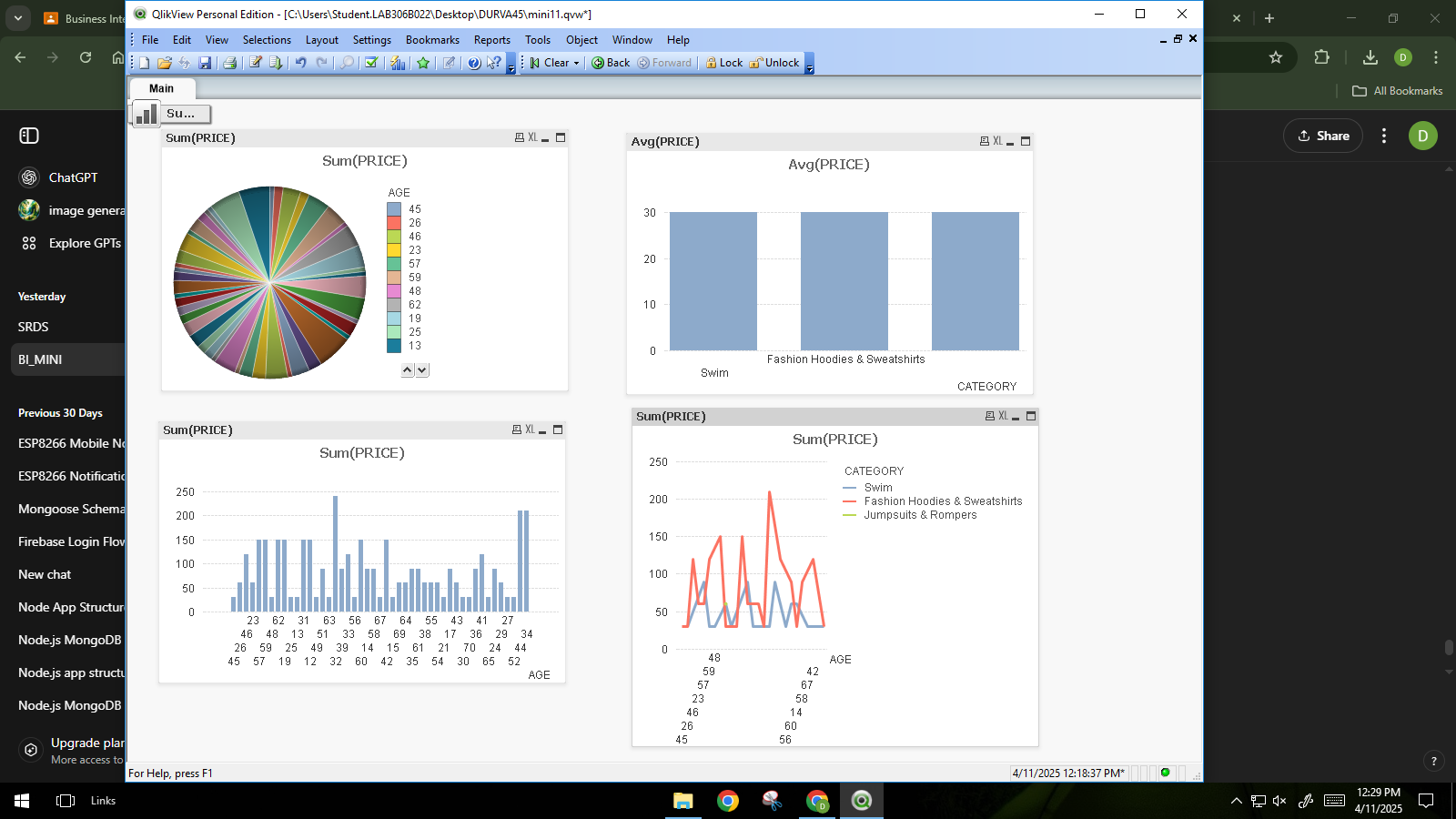


The chart shows that “Fashion Hoodies & Sweatshirts” lead in revenue across most age groups, especially around age 58, while “Swim” and “Jumpsuits & Rompers” generate less.



The QlikView dashboard provides a multi-angle view of customer spending patterns:

1. **Age-based Revenue**: The pie and bar charts show that age groups like 45, 46, and 27 contribute significantly to total sales.
2. **Category Performance**: “Fashion Hoodies & Sweatshirts” consistently lead in both total and average prices, indicating strong customer interest.
3. **Trends Across Age & Categories**: The line chart reveals that revenue from "Fashion Hoodies & Sweatshirts" peaks around age 58, while other categories like "Swim" and "Jumpsuits & Rompers" show lower performance.



**DECISION:**

Based on the interpretation of clustering and visual analytics, the following Business Intelligence decisions can be made:

1. **Target high-value customer segments** (users aged 26, 45, 46) with personalized promotions, as they contribute significantly to total revenue.
2. **Focus marketing efforts** on product categories like Jeans, Intimates, and Fashion Hoodies & Sweatshirts, which show high sales volume and revenue.
3. **Revise inventory and pricing strategies** for underperforming categories such as Clothing Sets, Suits, and Swimwear, either by bundling, discounting, or rebranding.
4. **Create country-specific campaigns** based on age and gender trends, tailoring products to the demographic profiles of regions like Germany (older users) vs. Spain (younger users).

**CONCLUSION:**

Through this Business Intelligence project on E-Commerce Customer Segmentation, we applied K-Means clustering and data visualization tools (WEKA, Orange, Tableau, QlikView) to analyze customer demographics, purchase behavior, and product category trends. The segmentation helped identify key customer groups based on age, gender, and buying preferences. These insights can drive personalized marketing, optimize inventory for high-demand categories like Jeans and Intimates, and tailor pricing strategies to specific age groups that contribute the most revenue. Overall, customer segmentation enables smarter business decisions, improves customer targeting, and enhances profitability in e-commerce.