**Project Title: Customer Segmentation Using Clustering (Python)**

**Company: Infotech Solution**

**Tools Used: Jupyter Notebook, VS Code**

**Technologies: Data preprocessing, ML models (K-Means), sentiment analysis**

**Libraries: pandas, scikit-learn, matplotlib, seaborn**

**Created By: Sakshi Desai**

* **Problem Statement :**
* The objective of this project is to **segment customers** based on demographic and behavioral data using **K-Means clustering**.
* This segmentation helps businesses design **targeted marketing strategies**, improve customer engagement, and optimize sales efforts.
* **Project Overview**
* **Objective :**

 Collect and clean customer dataset.

 Perform **Exploratory Data Analysis (EDA)** to understand customer demographics.

 Normalize and scale data for clustering.

 Apply **K-Means clustering** to segment customers.

 Determine the optimal number of clusters using **Elbow Method** and **Silhouette Score**.

 Profile customer clusters and provide **marketing recommendations**.

 Visualize results with charts and PCA projections.

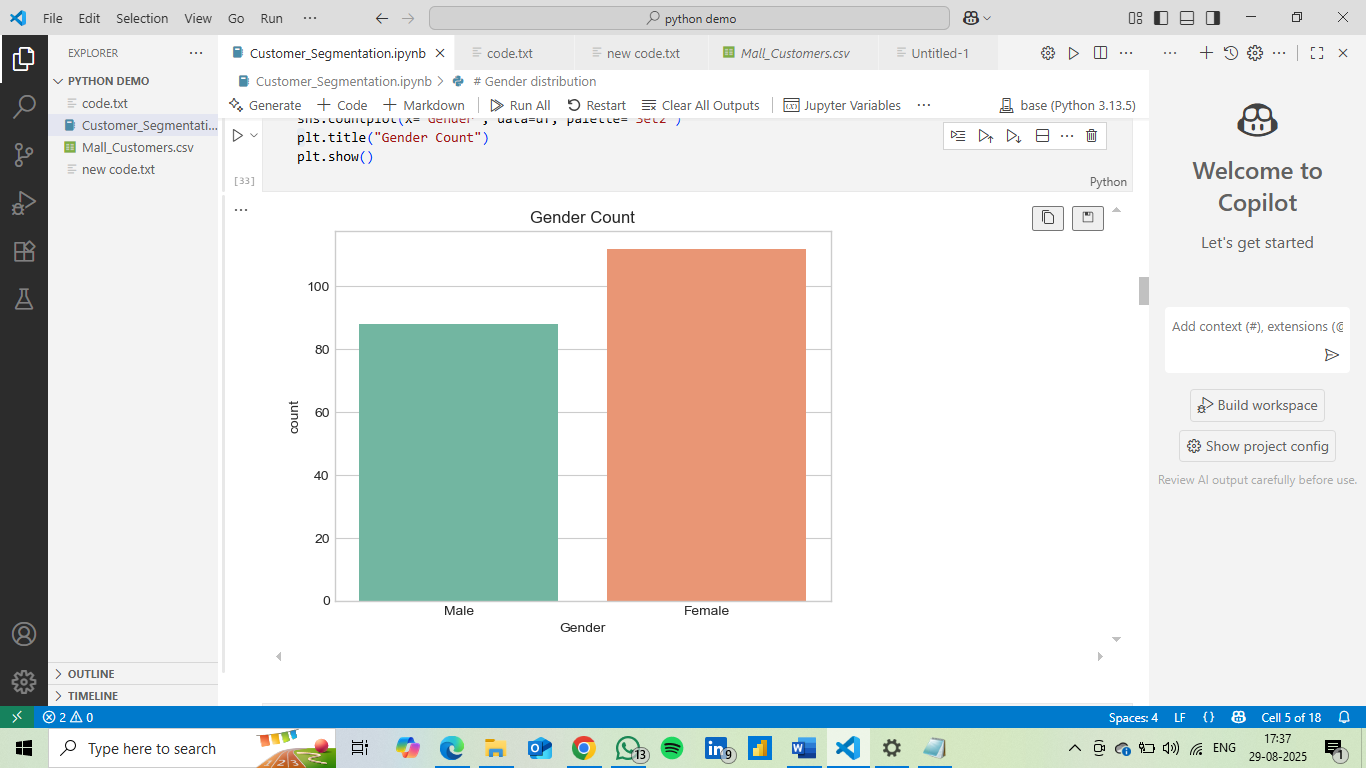
* **Methodology :**

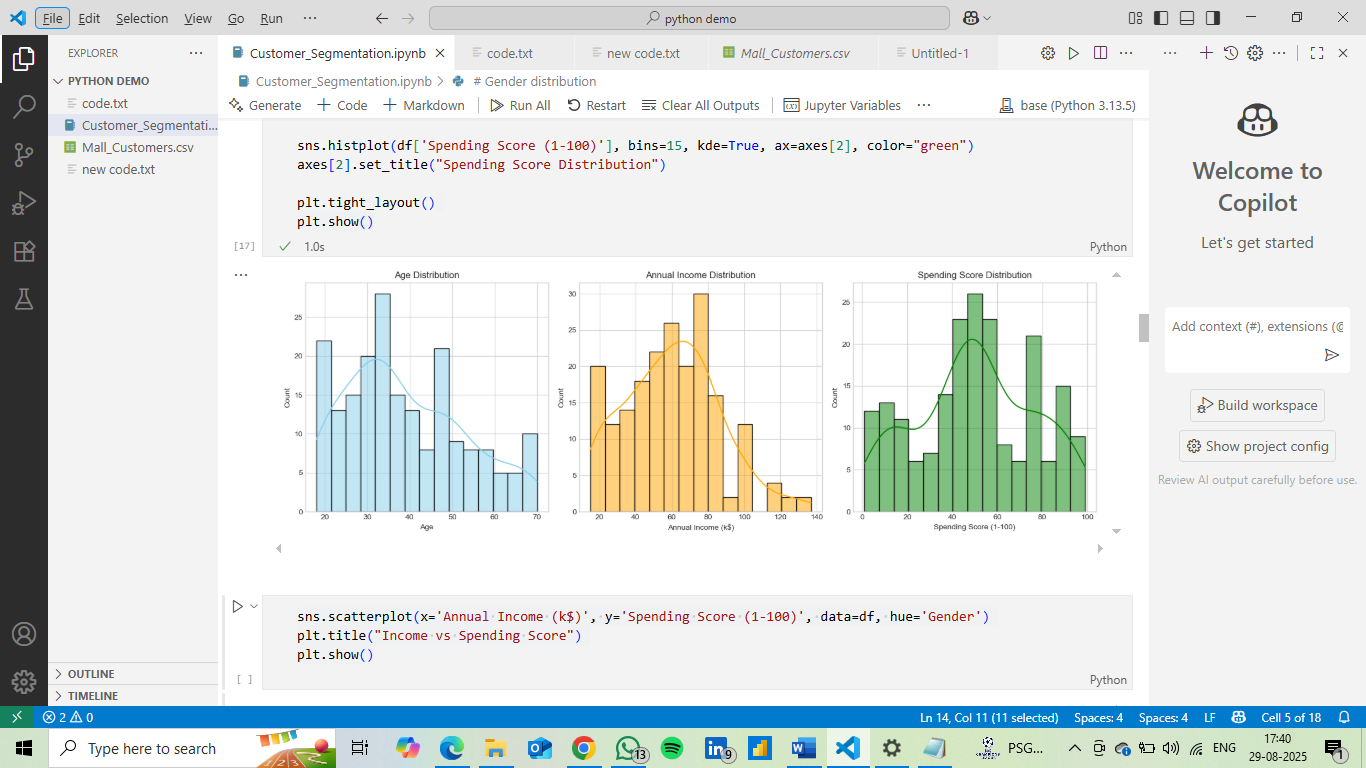
1. **Data Preprocessing: Cleaning, handling duplicates, encoding categorical data.**
2. **EDA: Studying age, income, and spending score distributions.**
3. **Feature Scaling: Standardization for better clustering performance.**
4. **K-Means Clustering: Applying algorithm to group customers.**
5. **Model Evaluation: Elbow method and silhouette score to select best k. Cluster Profiling: Analysing traits of each cluster.**
6. **Visualization: Scatter plots, PCA projection, bar charts, and radar charts.**

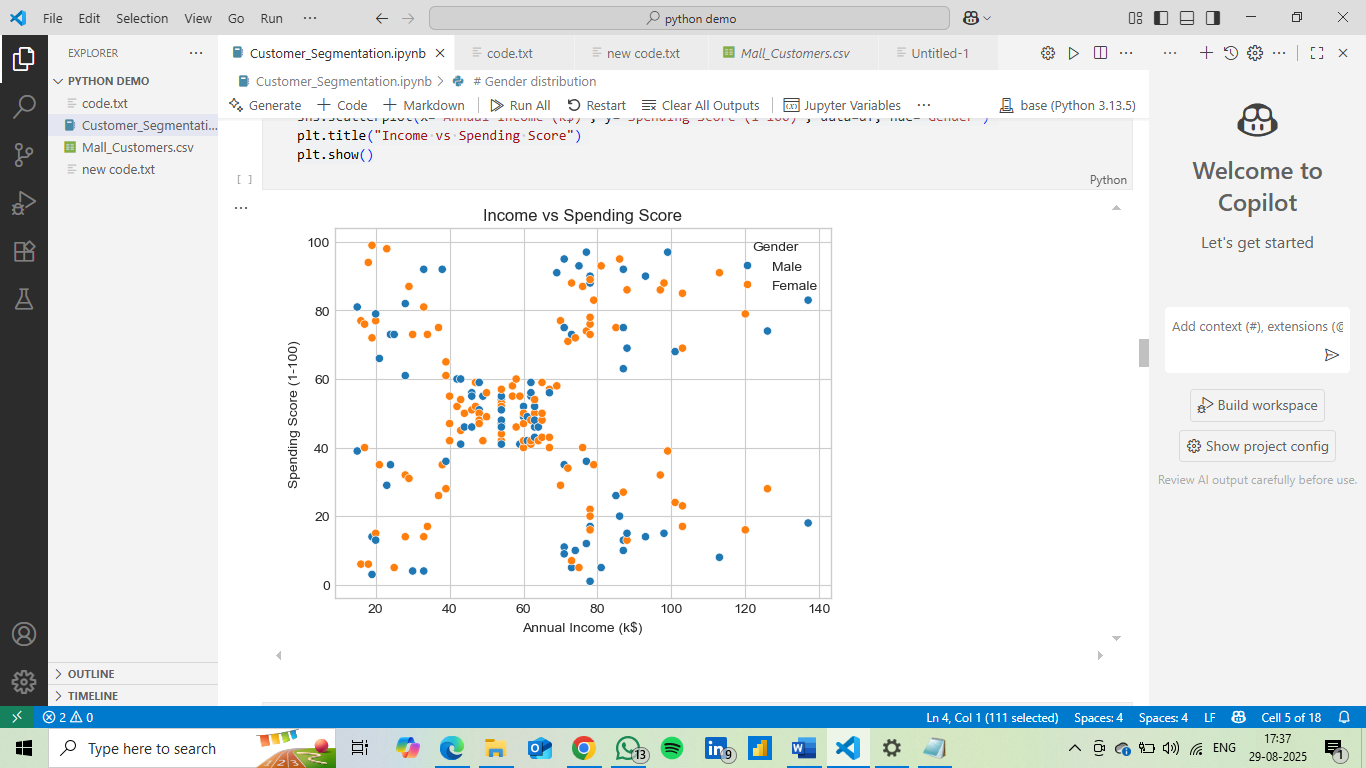
* **Tools & Libraries**
* **Python: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn**
* **Clustering: K-Means**
* **Visualization: PCA plots, histograms, scatter plots**
* **Week-wise Work Summary**

**Week 1:** Data Collection & Preparation

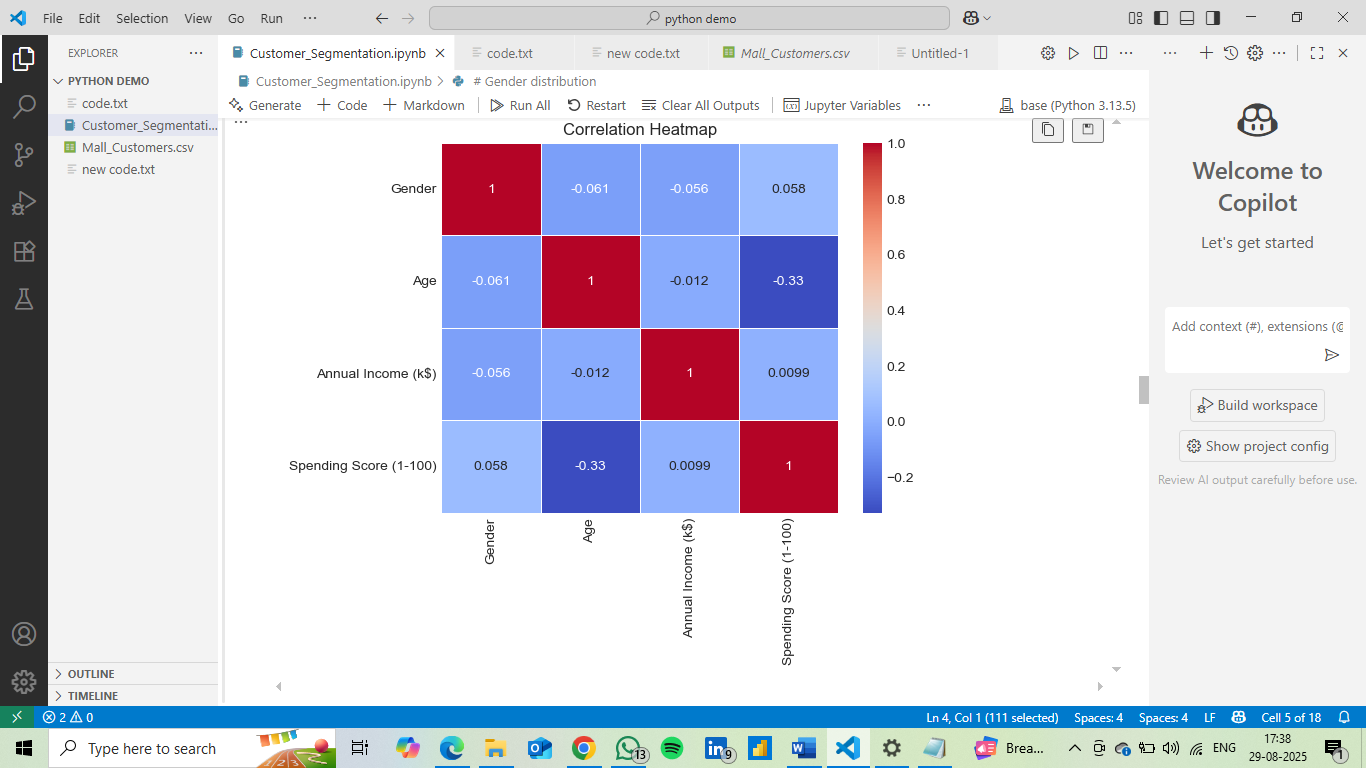
* Collected dataset: *Mall Customers dataset* from Kaggle (200 records).
* Cleaned dataset by removing unnecessary columns (e.g., CustomerID).
* Encoded categorical variables (Gender → Male = 0, Female = 1).
*  Checked for missing values & duplicates (none found).
*  Performed **EDA**:
* Age distribution: Most customers between **25–40 years**.
*  Annual Income: Spread across low, medium, and high income levels.
* Spending Score: Customers distributed across low, medium, and high spending categories.
* Normalized and scaled numerical features (Age, Annual Income, Spending Score) using StandardScaler.
* Gender Distribution:



* Age, Income, Spending Score distributions
* Income vs Spending Score



* Correlation Heatmap

****

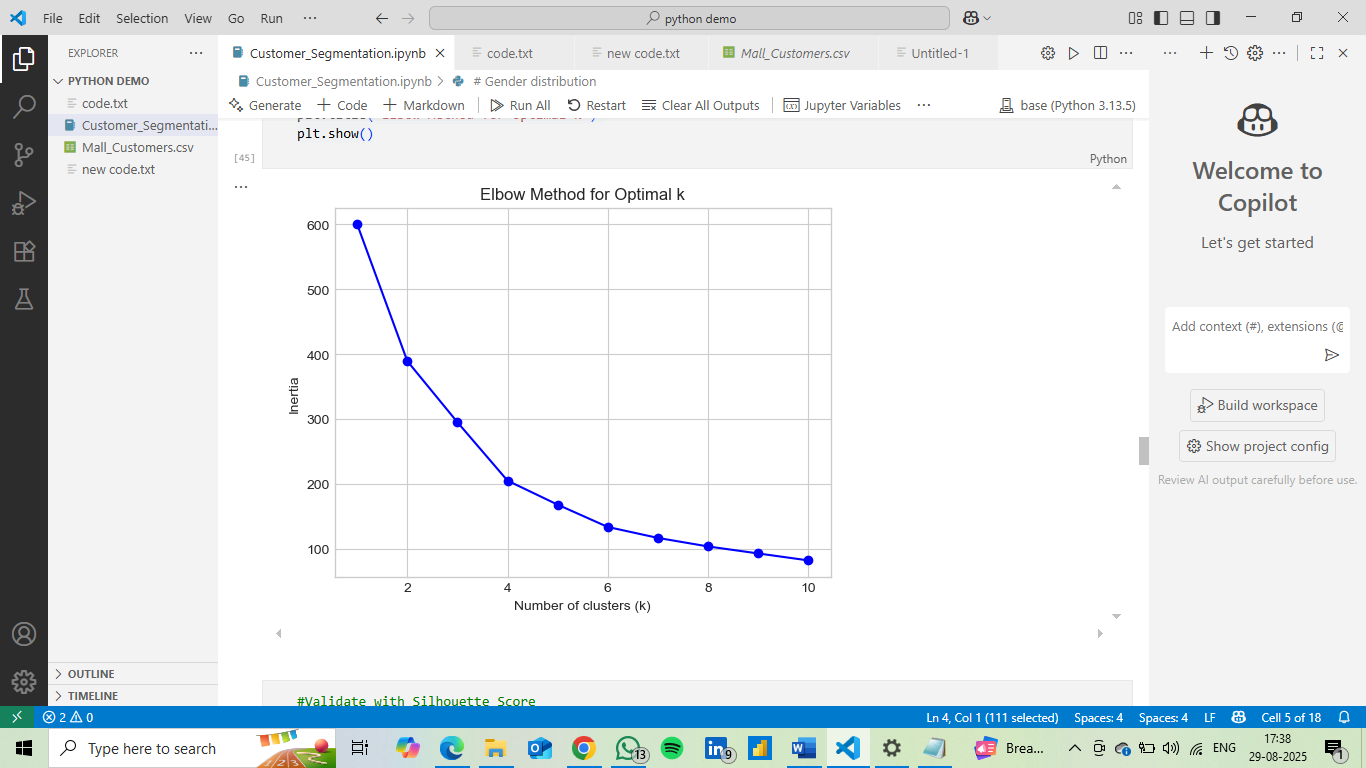
**Week 2** : Clustering Model Development

* Applied K-Means clustering on scaled features.
* Determined optimal clusters:

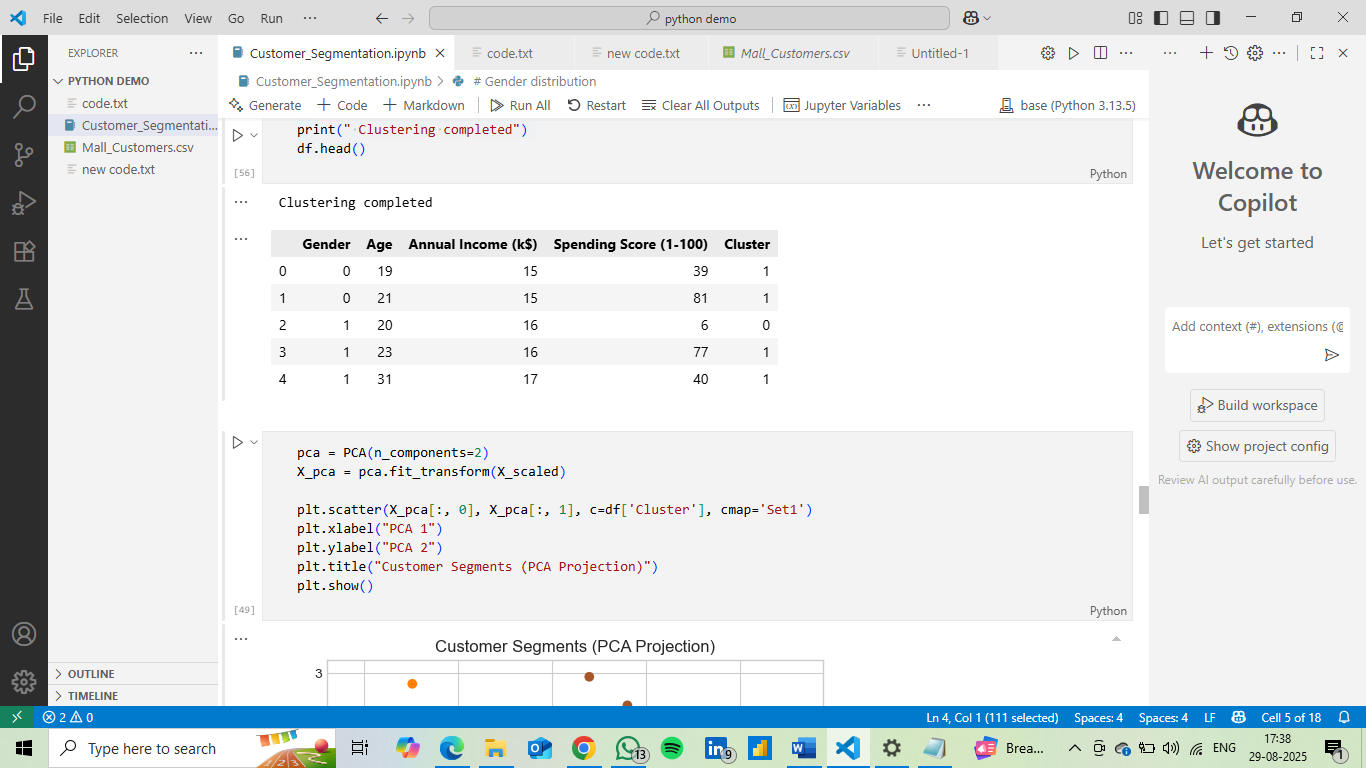
 Elbow Method suggested **4 clusters**.

 **Silhouette Score** confirmed good cluster separation.

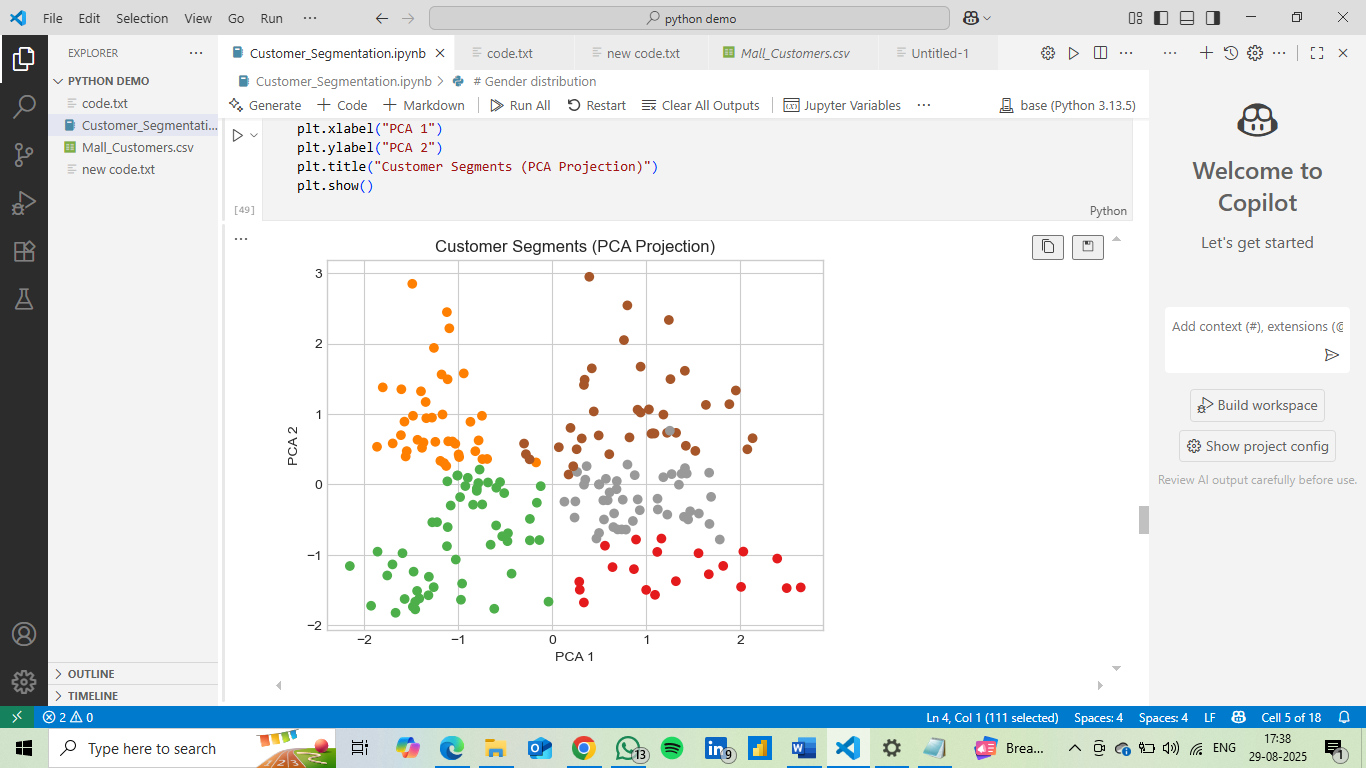
* Visualized clusters using scatter plots and **PCA (2D projection)**.
* Elbow Method For Optimal K



* Clustering Completed

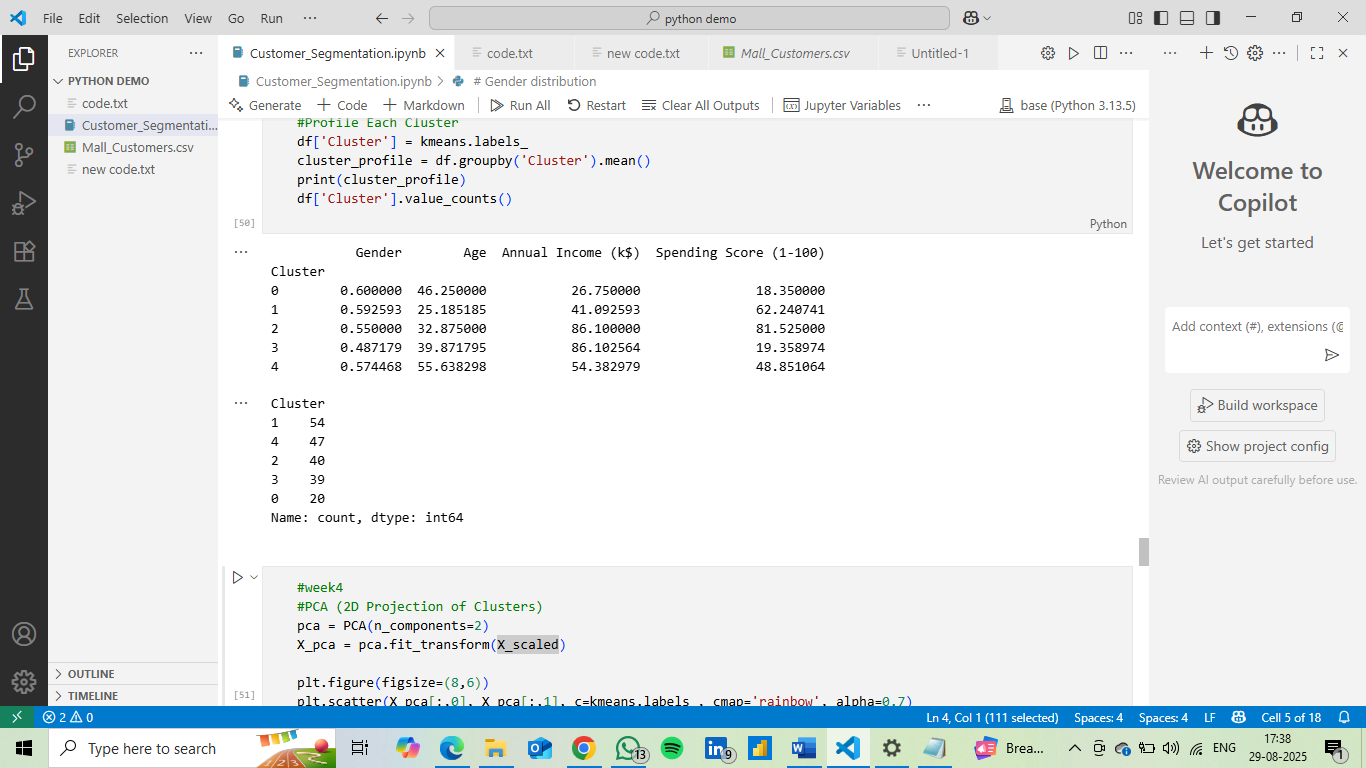


* Customer Segments (PCA Projection)



**Week 3:** Cluster Profiling & Insights

* **Cluster 1 (Young Spenders)**: Younger customers, medium income, **high spending** → Target with luxury, fashion, lifestyle promotions.
* **Cluster 2 (High Income, Low Spending)**: Wealthy but less engaged → Offer loyalty programs, personalized offers to boost spending.
* **Cluster 3 (Budget Customers)**: Low income, low spending → Focus on discounts, essential product campaigns.
* **Cluster 4 (Balanced Middle Group)**: Middle-aged, moderate income and spending → Upsell with family-oriented deals, value bundles.
* Profile Each Cluster



* Marketing Strategies for Each Cluster

|  |  |  |
| --- | --- | --- |
| Cluster | Traits | Marketing Strategy |
| Cluster 0 – Young Impulsive Spenders | Younger customers, low income, but high spending score | - Attract with **discount offers** and **seasonal sales**.  - Use **social media campaigns** to target trends.  - Introduce **loyalty programs** (cashback, points). |
| Cluster 1 – Wealthy but Conservative | High income but low spending | - Focus on **personalized marketing** (exclusive offers).  - Promote **premium products/services**.  - Use **email campaigns & VIP invitations**. |
| Cluster 2 – Balanced Shoppers | Middle income, moderate spenders | -Retain with **reward points** and **festive promotions**.  - Suggest **bundle deals** to increase basket size.  - Target with **mainstream campaigns**. |
| Cluster 3 – Premium Customers | High income & high spending, luxury seekers | - Position as **priority customers** with **VIP programs**.  - Offer **exclusive launches** and **luxury collections**.  - Provide **priority support & concierge services**. |
| Cluster 4 – Budget-Conscious Customers | Low income, low spending, least engaged | - Keep them engaged with **low-cost campaigns**.  - Send **newsletter/email updates**.  - Promote **affordable/basic products**. |

* Draft cluster summary report

**Cluster 0 – Young Impulsive Spenders**

* **Traits:** Younger age, relatively low income, but **high spending score**.
* **Behavior:** Spend disproportionately to their income, enjoy lifestyle shopping.
* **Marketing Strategy:**
  + Attract with **discounts and seasonal offers**.
  + Promote through **social media and influencer campaigns**.
  + Introduce **loyalty programs** to retain them.

**Cluster 1 – Wealthy but Conservative**

* **Traits:** Middle-aged, **high income**, but **low spending score**.
* **Behavior:** Financially capable but prefer saving or cautious spending.
* **Marketing Strategy:**
  + Provide **exclusive offers** and **premium memberships**.
  + Encourage with **VIP events** and **personalized email campaigns**.
  + Market **luxury and high-value products**.

**Cluster 2 – Balanced Shoppers**

* **Traits:** Average age, **medium income and spending score**.
* **Behavior:** Neither extravagant nor conservative; steady buyers.
* **Marketing Strategy:**
  + Retain with **reward points** and **festive promotions**.
  + Encourage with **bundle deals** and **cross-selling**.
  + Target using **mainstream campaigns**.

**Cluster 3 – Premium Customers**

* **Traits:** Mostly young-to-middle age, **high income and high spending score**.
* **Behavior:** **Most valuable group**, brand-conscious, frequent luxury buyers.
* **Marketing Strategy:**
  + Offer **exclusive launches and luxury collections**.
  + Provide **priority support, concierge services, and VIP clubs**.
  + Maintain loyalty with **personalized rewards**.

**Cluster 4 – Budget-Conscious Customers**

* **Traits:** Higher age group, **low income, low spending score**.
* **Behavior:** Least engaged, purchase only essentials.
* **Marketing Strategy:**
  + Keep engaged with **low-cost campaigns** and **basic product promotions**.
  + Use **email/newsletter marketing** instead of costly campaigns.
  + Promote **affordable product lines**.

**Key Insights**

* Customers are effectively grouped into **5 segments** with clear differences in **income and spending behavior**.
* **Clusters 3 & 0** (Premium + Young Spenders) are the **highest potential revenue drivers**.
* **Clusters 1 & 4** require **specialized approaches** to increase engagement.
* **Balanced shoppers (Cluster 2)** represent a **stable middle group** for long-term retention.

**Week 4 :** **Finalize report with supporting charts and tables**

**Tasks Completed**

* Finalized Report: Combined data cleaning, EDA, clustering, profiling, and strategies into one document.
* Supporting Charts & Tables: Added visual evidence (Elbow Plot, Silhouette Scores, PCA scatter, Cluster averages).
* Prepared Presentation: A short summary deck highlighting problem, methodology, findings, and recommendations**.**
  1. **Clustering Summary :**

 Optimal clusters: **5 (using Elbow Method + Silhouette Score)**.

 Features used: **Age, Annual Income, Spending Score**.

 Clusters created with **K-Means** and visualized using **PCA**.

* 1. **Supporting Charts & Tables :**

 **Elbow Plot:** Confirmed 5 clusters.

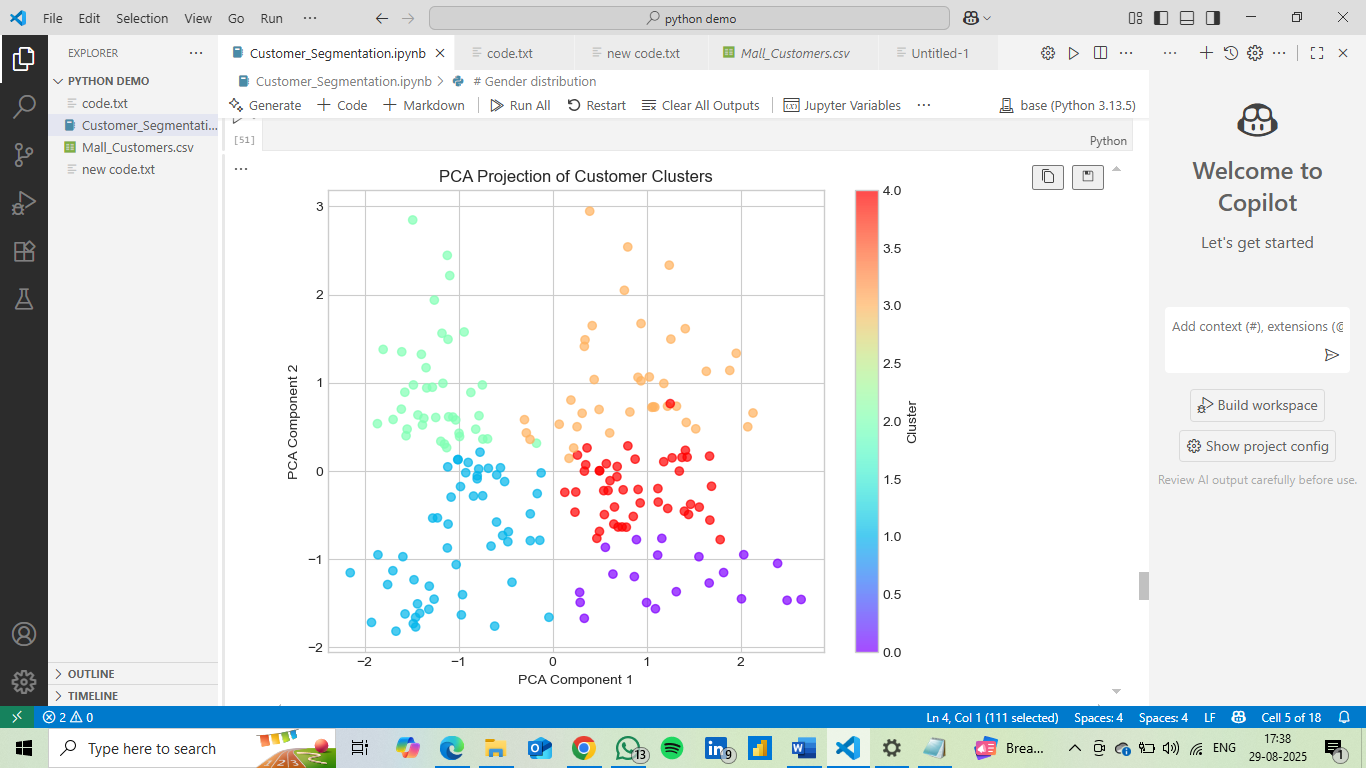
 **Silhouette Score Table:** Best score near k=5.

 **Scatter Plot (Income vs Spending Score):** Clear customer segmentation.

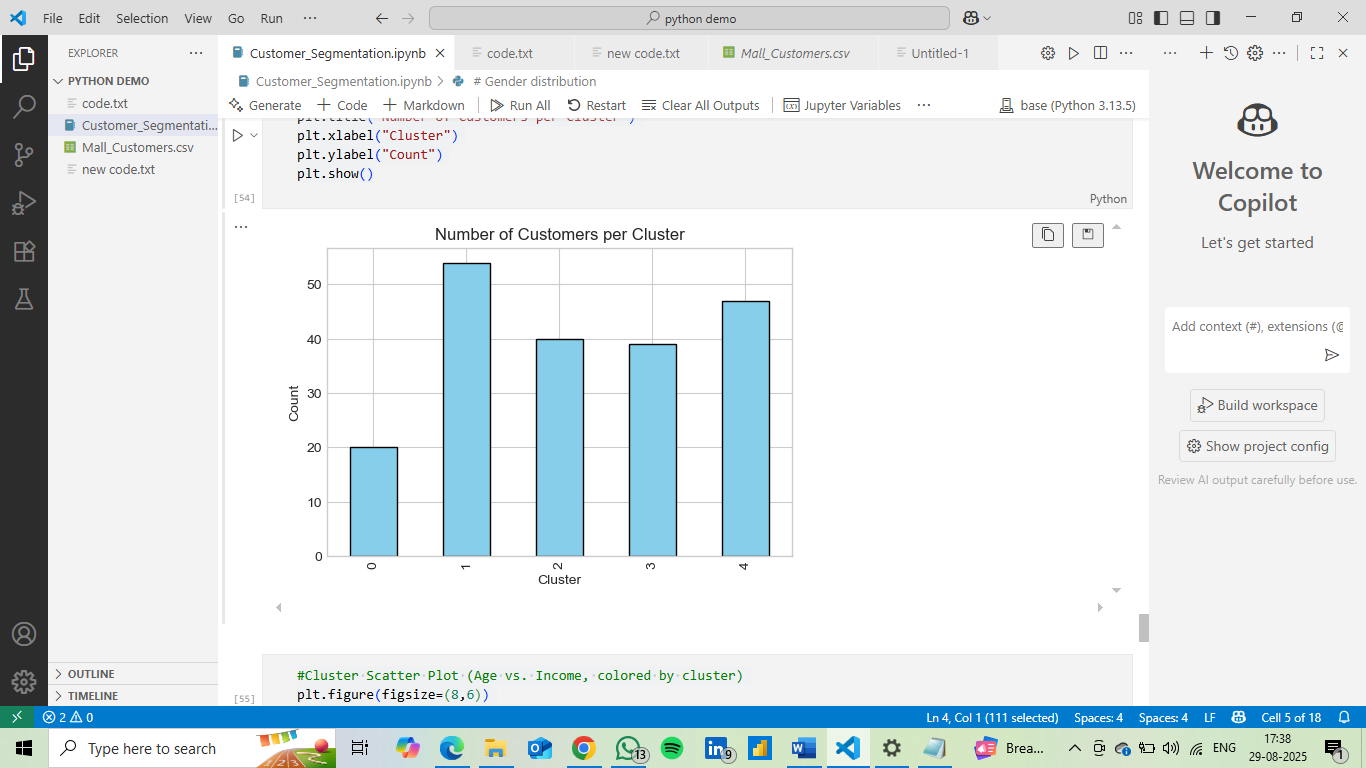
 **PCA 2D Plot:** Dimension reduction preserved separation.

 **Cluster Profile Table:** Average age, income, spending score for each group.

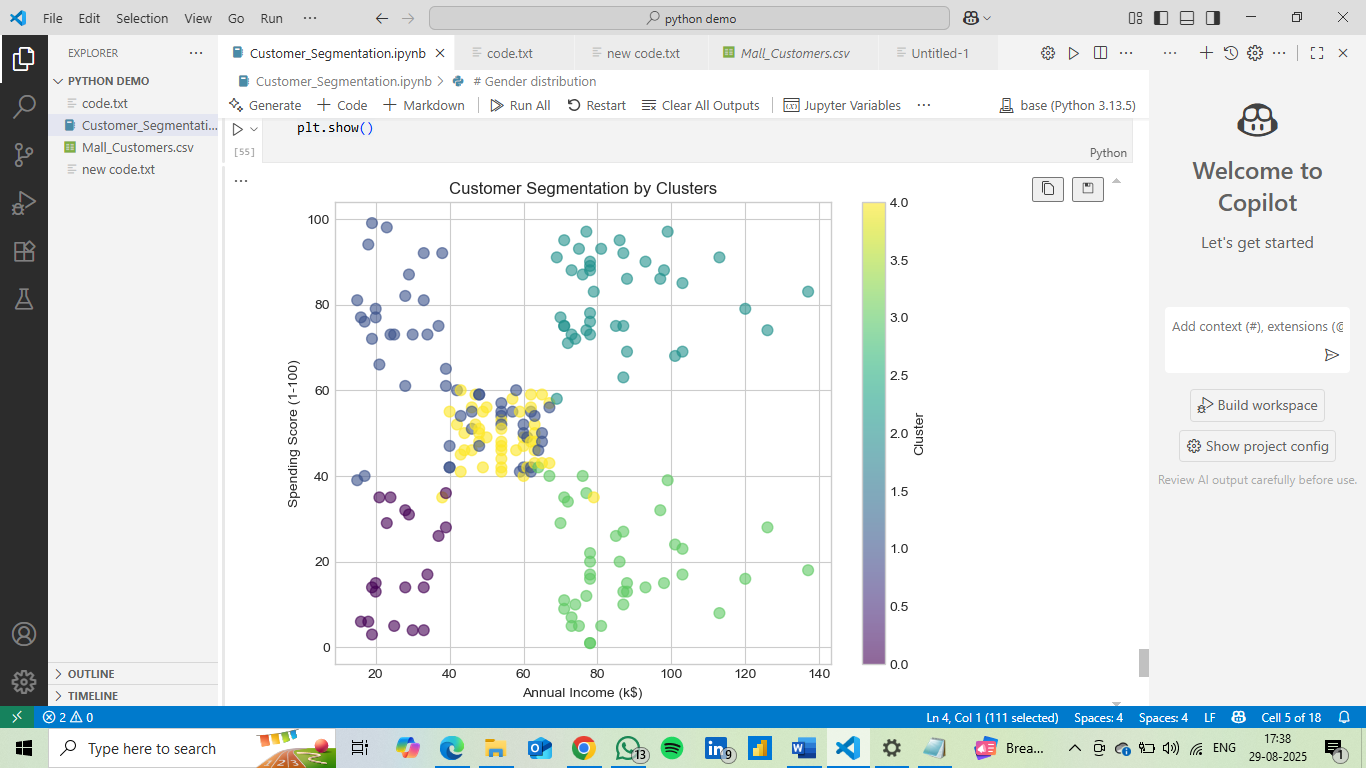
* PCA Projection of Customer Clusters



* Number of Customers per Cluster



* Customer Segmentation by Clusters



**Key Insights**

* **Cluster 3 (Premium Shoppers)** → High income & high spending → *Best target for luxury marketing*.
* **Cluster 0 (Young Impulsive Spenders)** → Low income but high spending → *Engage via social media + discounts*.
* **Cluster 1 (Wealthy but Conservative)** → High income but cautious → *Personalized VIP offers*.
* **Cluster 2 (Balanced Shoppers)** → Average group → *Festive promotions, bundles*.
* **Cluster 4 (Budget-Conscious)** → Low engagement → *Basic, affordable product marketing*.

Conclusion:

* The customer segmentation analysis successfully grouped mall customers into **five distinct clusters** based on **Age, Annual Income, and Spending Score**.
* **Cluster 0**: Older customers with **low income and low spending** – least profitable segment.
* **Cluster 1**: Young adults with **moderate income but high spending** – potential high-value segment.
* **Cluster 2**: Middle-aged customers with **high income and high spending** – premium customers.
* **Cluster 3**: Professionals with **high income but low spending** – price-conscious or less engaged.
* **Cluster 4**: Mature customers with **average income and balanced spending** – steady but moderate contributors.
* Using **PCA**, we reduced dimensions and visualized how clearly separated these clusters are.

**Key Takeaways:**

* Mall customers are **not homogeneous**; they have diverse income and spending patterns.
* High-income & high-spending clusters should be targeted with **premium offers, loyalty programs, and exclusive experiences**.
* Young, moderate-income but high-spending clusters are **growth opportunities**; they respond well to **trendy promotions and discounts**.
* Low-spending clusters require **cost-effective strategies** or can be deprioritized in marketing.
* Overall, this segmentation enables **data-driven marketing**, better **customer engagement**, and more **personalized strategies** to maximize profitability.

THANK YOU