**DOCUMENTATION**

**Project Title: “Customer Segmentation”**

**Domain: “Data Science”**

**Batch 2**

**Problem Statement:**

Organizations tend to struggle with determining the correct target audience because of inadequate structured segmentation. Consequences include poor marketing plans, low customer interaction, and lost revenue opportunities.

We apply unsupervised machine learning algorithm to segment customers according to their annual income and spending score , giving insights to optimize business strategies. Hence, we overcome these problems .

**Objective:**

The Project aims to perform exploratory data analysis (EDA) on "Mall\_customer.csv", merge K-Means clustering with identification of unique customer segments, compare with DBSCAN (Density-Based Clustering), customer segments are visualized by scatter plots and to calculate average expenditure per cluster for enhanced business insights and to improve profit.

**Tools Used:**

* Vs code
* Jupyter notebook
* Python

**Libraries Used:**

* Pandas: Used for data handling
* Matplotlib
* Seaborn : Both used for data visualization
* **Scikit-learn** – Clustering algorithms (K-Means, DBSCAN) and scaling

**DataSet Used:**

Dataset used is “Mall\_customer.csv(Kaggle)”

• Attributes:

Customer ID → Distinct ID assigned to each customer

Gender → Male / Female

Customer age

Annual Income (k$) → Annual income in thousands

Spending Score (1-100) → Score rely on spending\_behavior

In this project, we utilize **Annual Income** and **Spending Score** for clustering.

**Key Features:**

• **Exploratory Data Analysis (EDA):**

Identify dataset structure and summary statistics.

Visualize Correlation Heatmap to show relationships.

• **Feature Selection & Scaling:**

Detect or Identify useful attributes (Annual Income & Spending Score) to be clustered.

Apply StandardScaler() to scale data.

**• Customer Segmentation using K-Means:**

Apply Elbow\_Method to calculate the optimal number of clusters (6).

Divided customers into meaningful, recognizable groups.

Scatter plots for cluster visualization for easy manipulation of dataset.

**• Customer Segmentation using DBSCAN:**

Used density-based clustering to identify non-linear patterns.

Identified outliers in customer data.

Compared with K-Means segmentation results.

**Cluster Analysis & Insights:**

Detection of average income per cluster.

Identified unique customer segments (e.g., high-income high-spending, low-income low-spending).

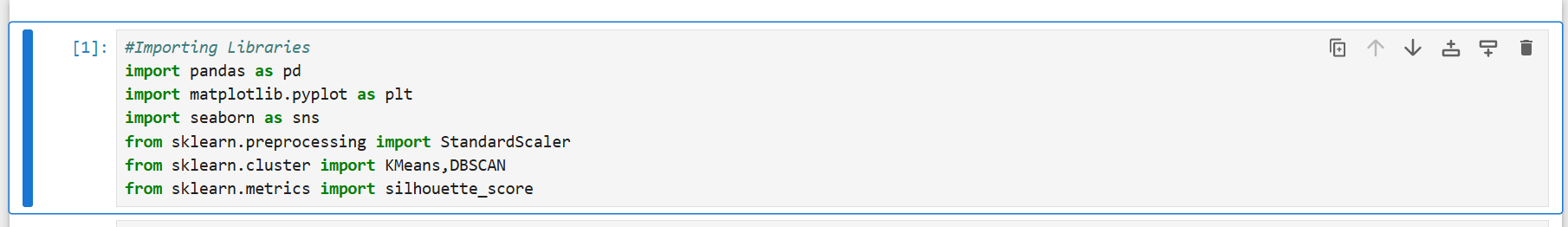
• **Visualization:**

Visualization applied through Heatmaps, pair plots and scatter plot to segment customer behavior based on Annual income and Spending score.

**Code and Output**:

**Step 1**:

**Importing Libraries**

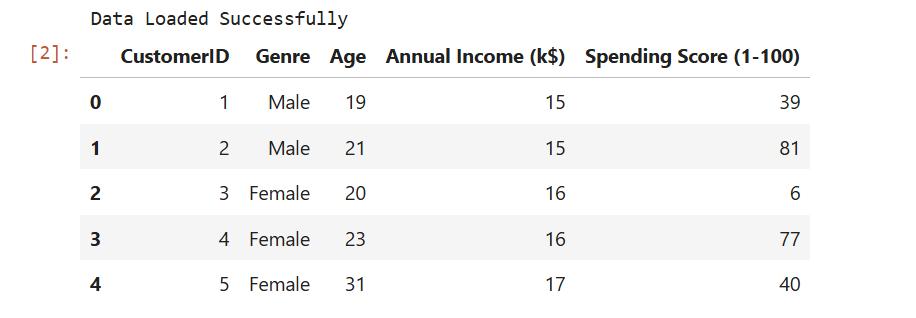


**Step 2:**

**Dataset Loading**

****

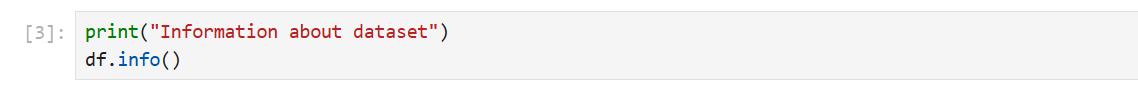
**Output:**

****

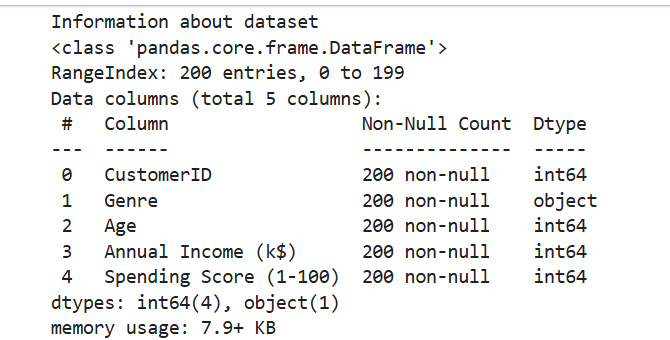
**Step 3:**

**Expolatory Data Analysis:**

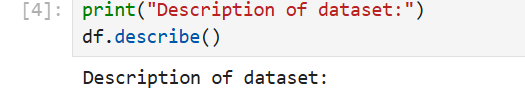
* **Information of dataset**



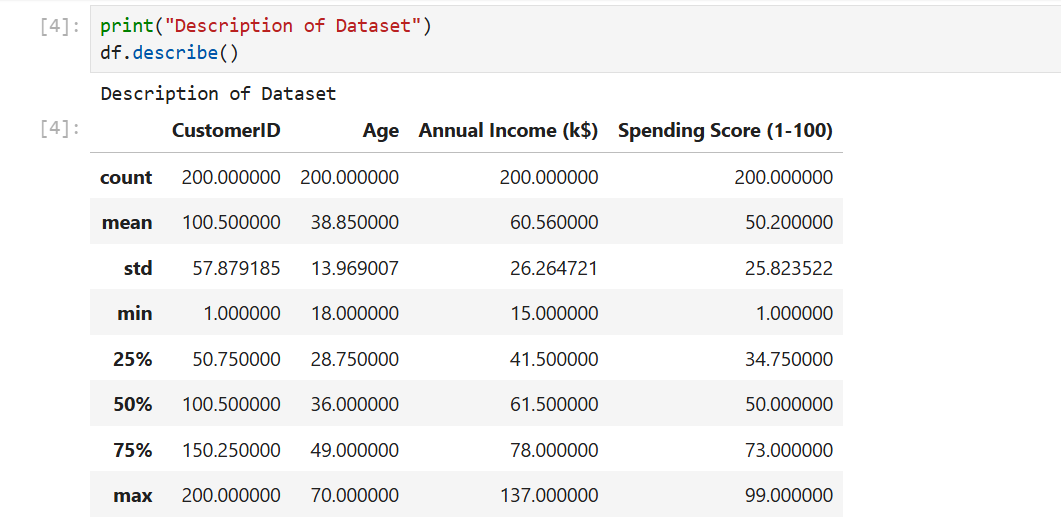
**Output:**

****

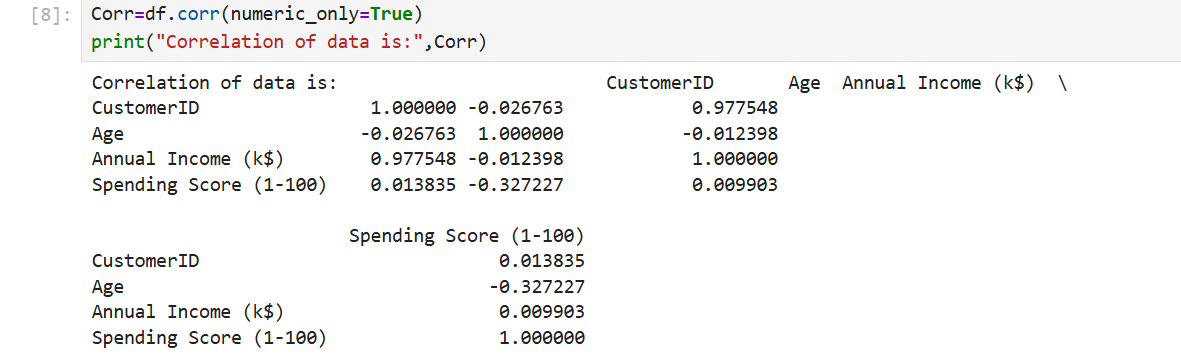
* **Statistic of Dataset:**



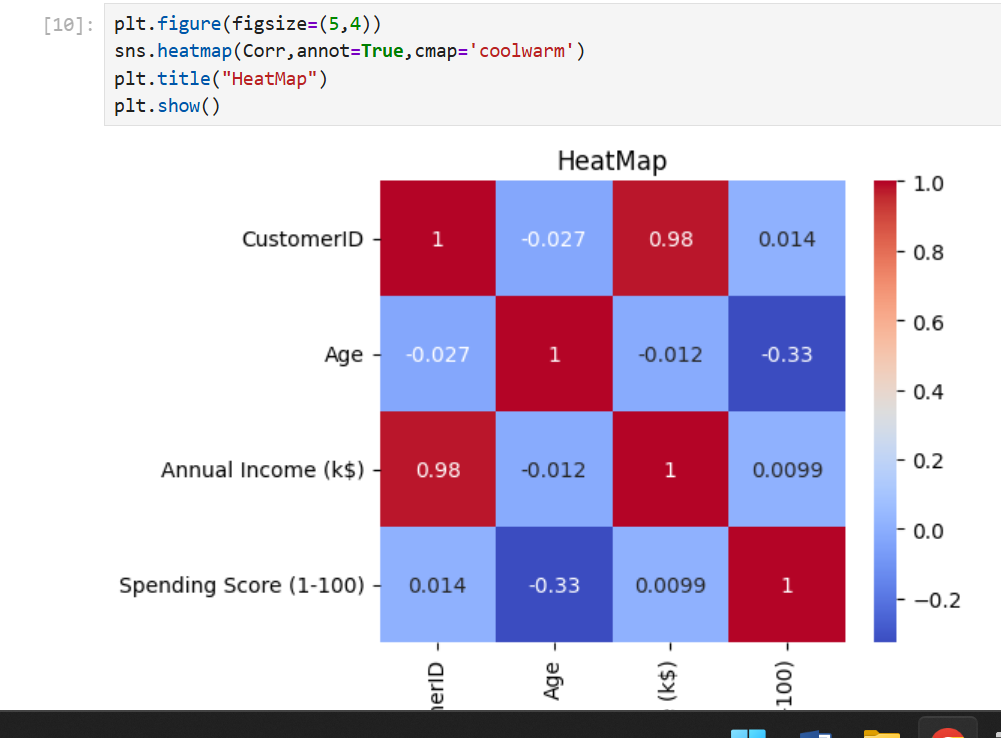
**Output:**

****

* **Correlation :**

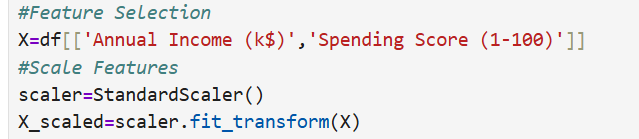
****

* **Correlation Heatmap:**

****

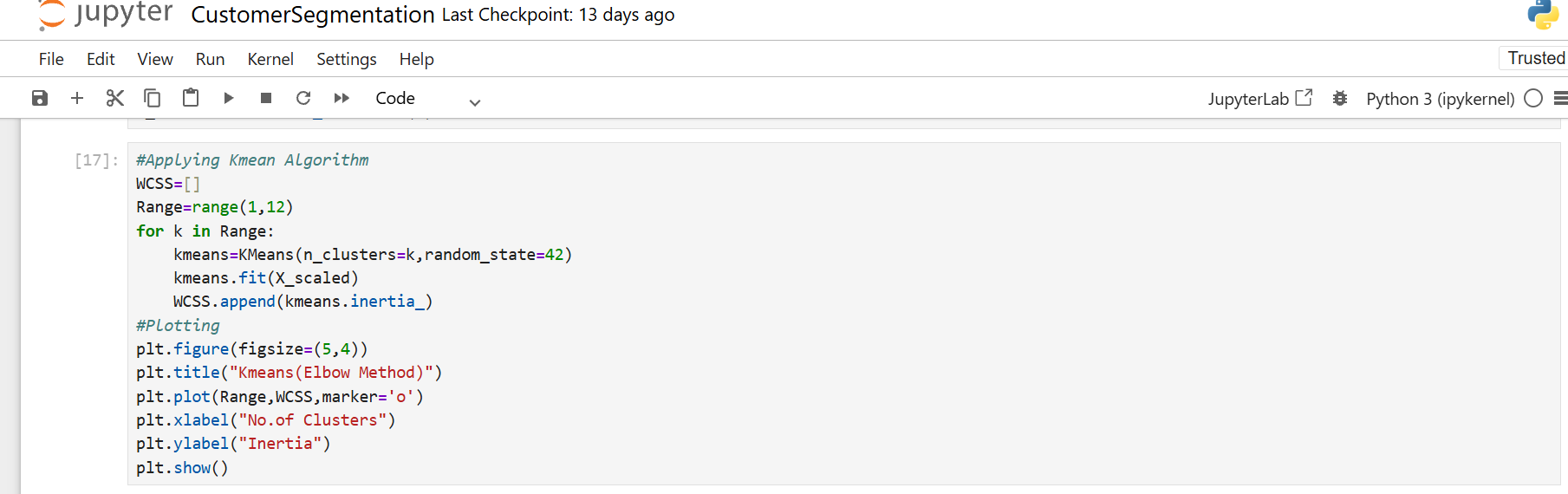
**Step 4:**

**Feature Selection and apply scaling:**

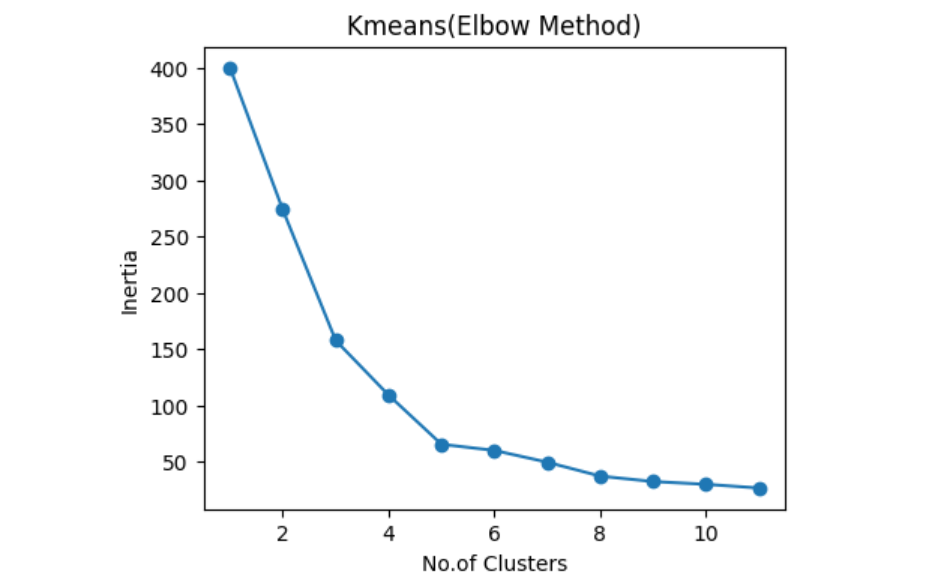
****

**Step 5:**

**Applying Kmean’s Algorithm and its plotting:**

****

**Plotting:**

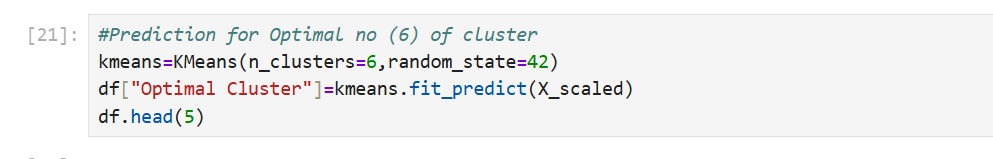
****

**Step 6:**

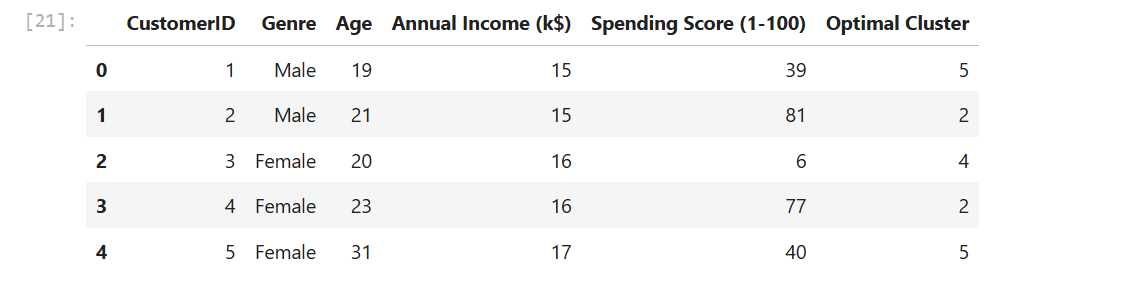
**Prediction for optimal no of clusters:**

* **Create Column named ‘Optimal Cluster’:**

**Code:**

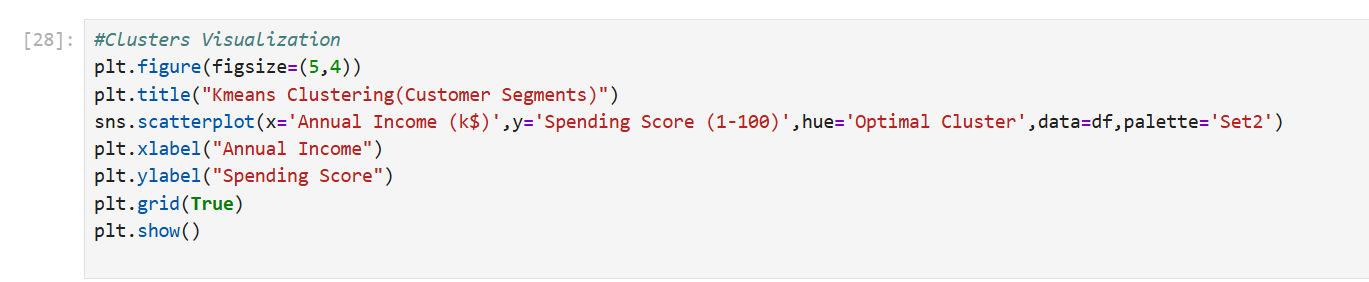
****

**Output:**

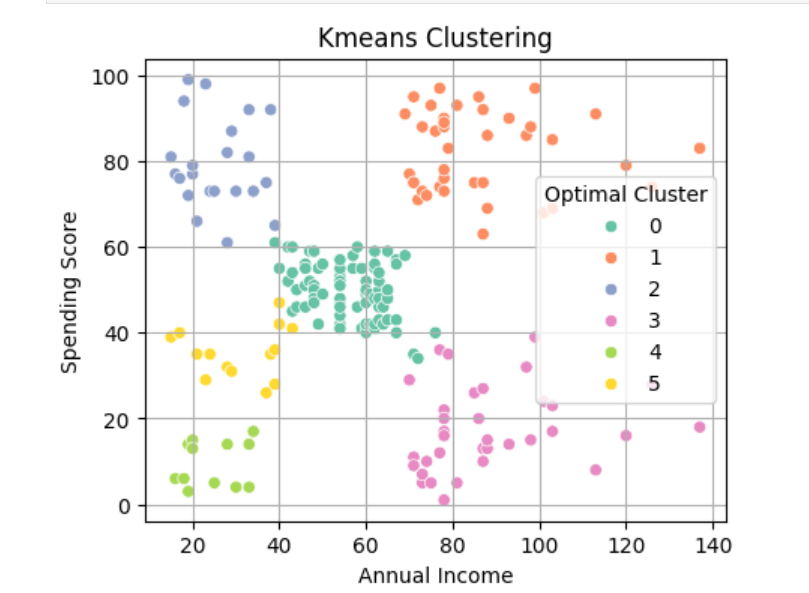
****

* **Optimal Cluster Visualization:**

**Code:**

****

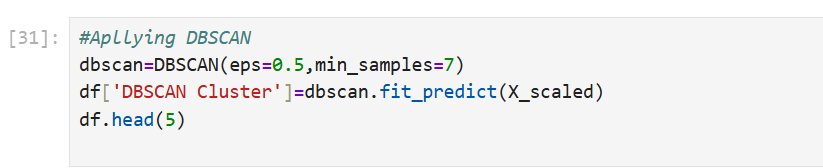
**Output:**

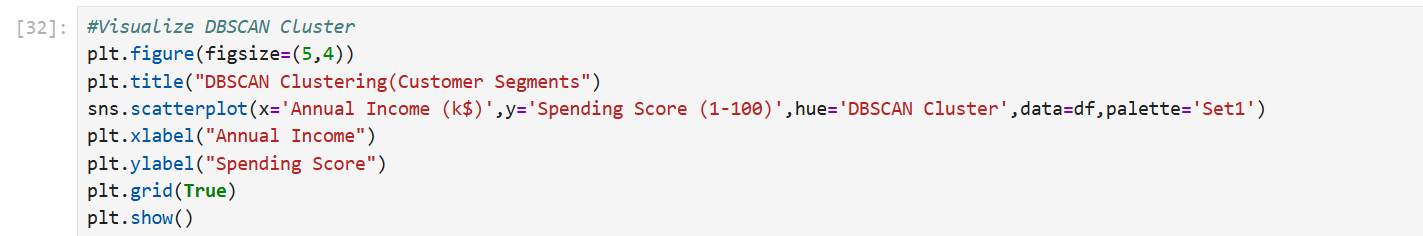
****

**Step 7:**

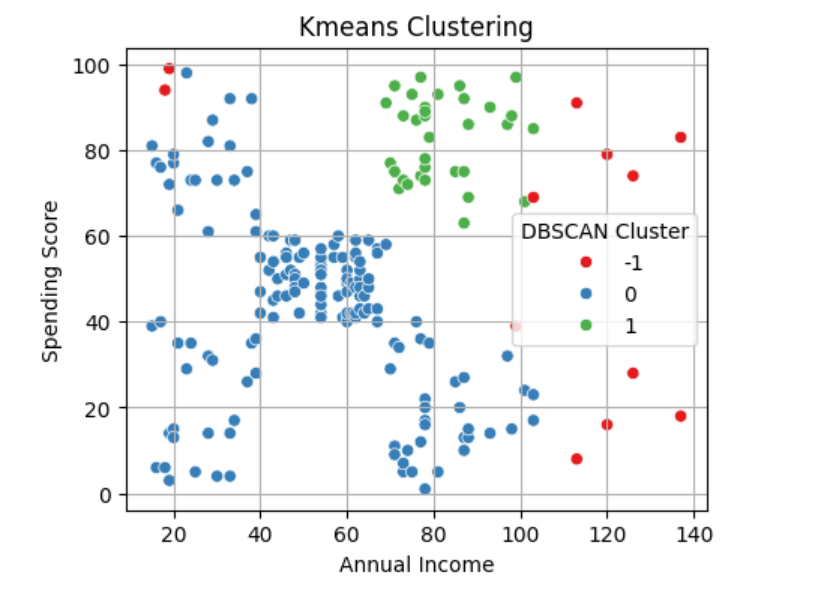
**Applying DBSCAN Algorithm:**

**Code:**

****

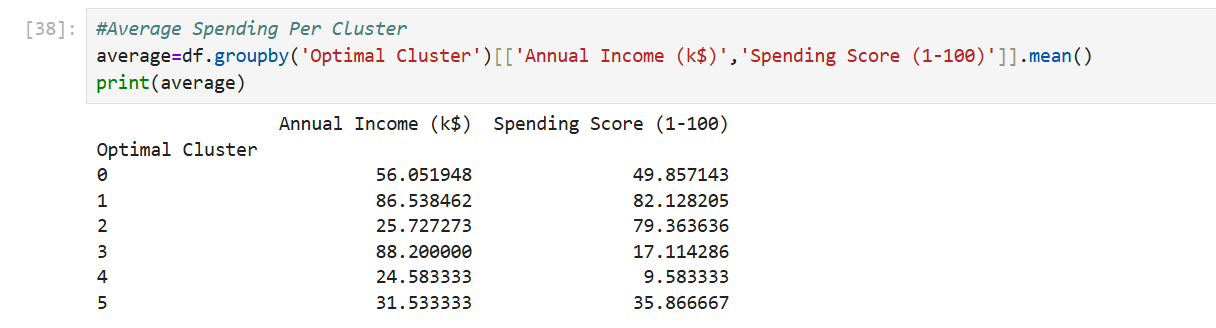
****

**Output:**

****

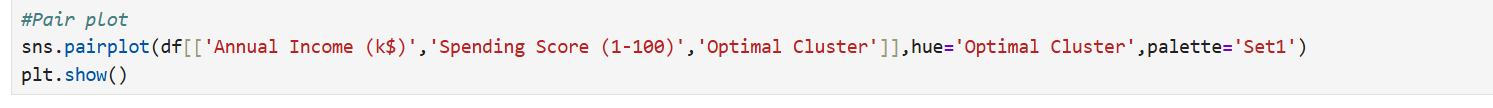
**Step 8:**

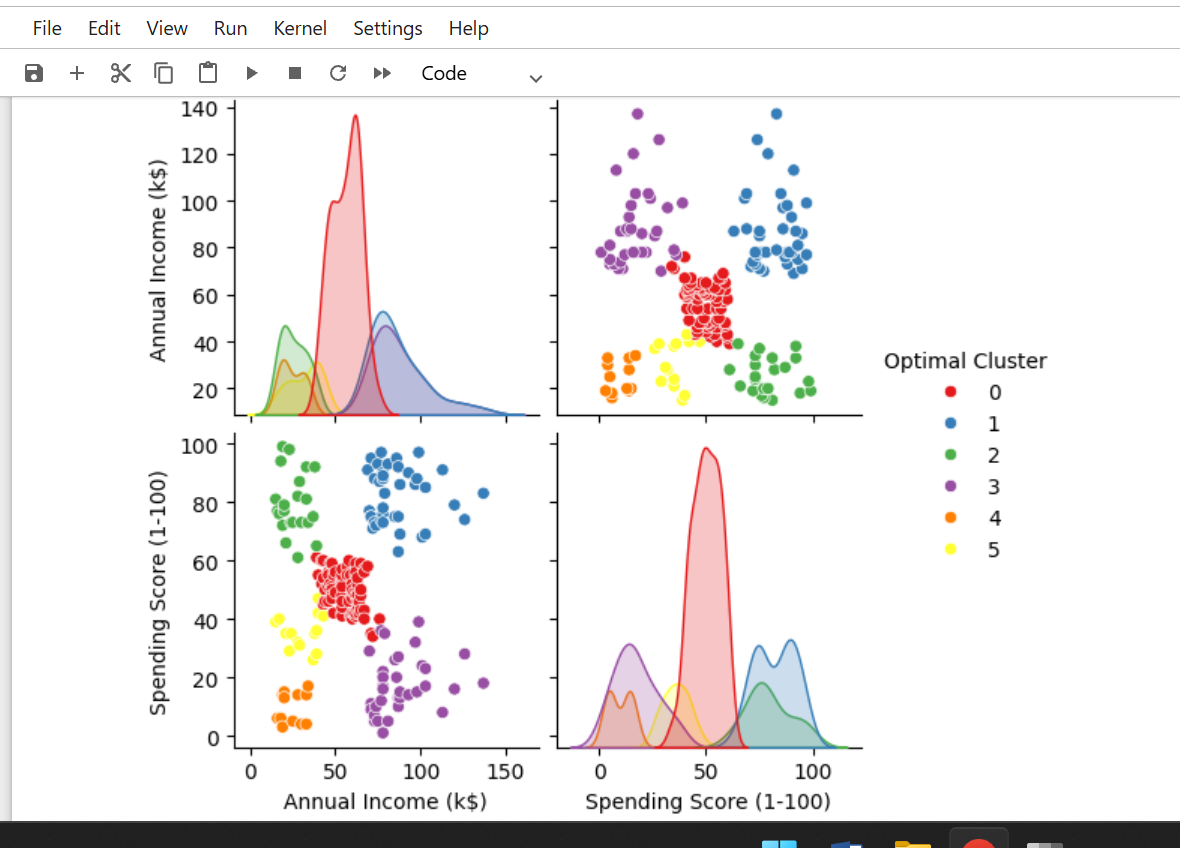
**Find Average per Cluster:**

****

**Step 9:**

**Visualization through Pair Plot:**

****

****

**-----------------------------------------------------------------------------------------------**