Customer Segmentation

**Phase – 4**

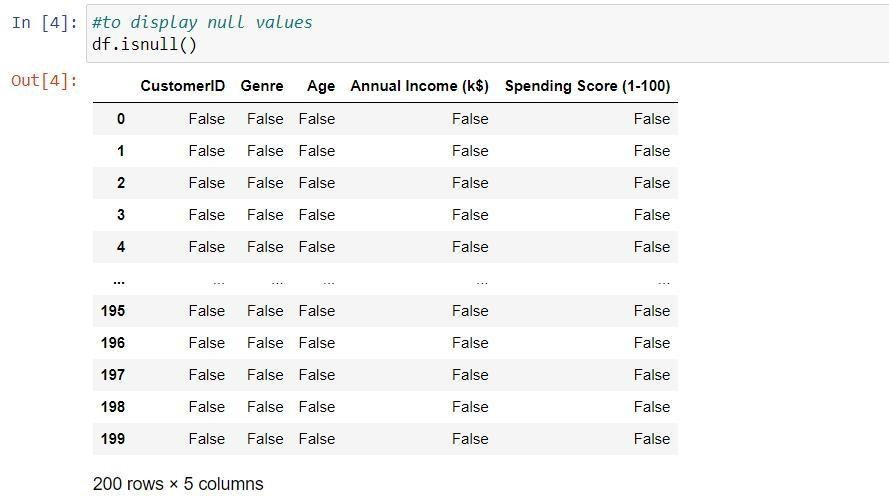
**Development part -2**

**Rino S(Team**

**Member)**

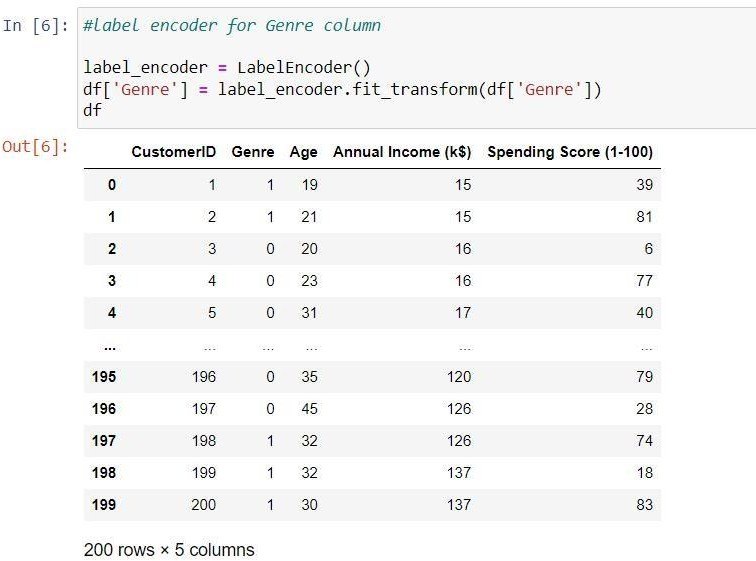
## Explanation for Development part -1:

### **Step 1**: Importing the required libraries and loading the dataset

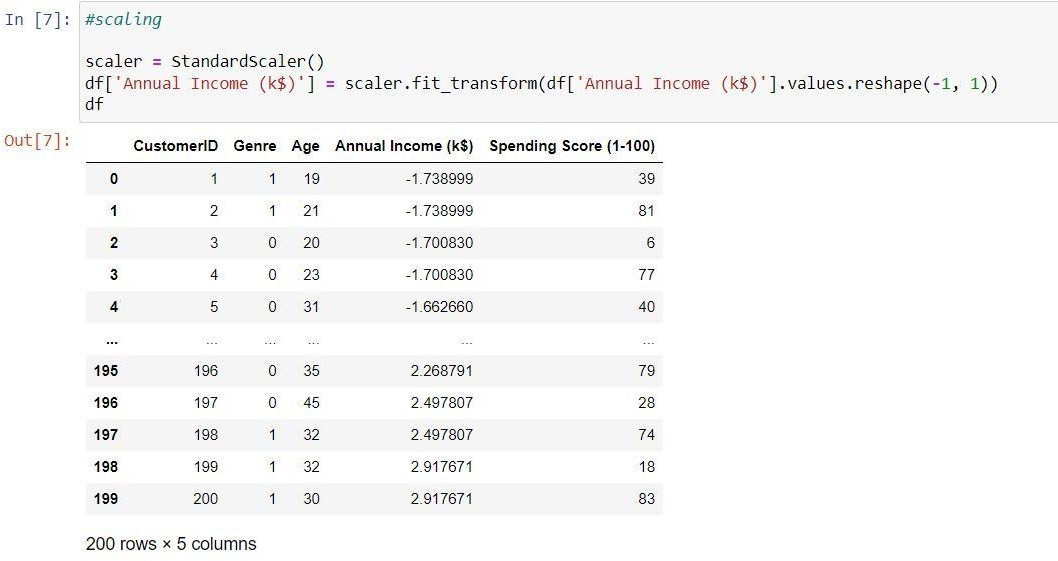
**Step 2**: Handling Missing Data

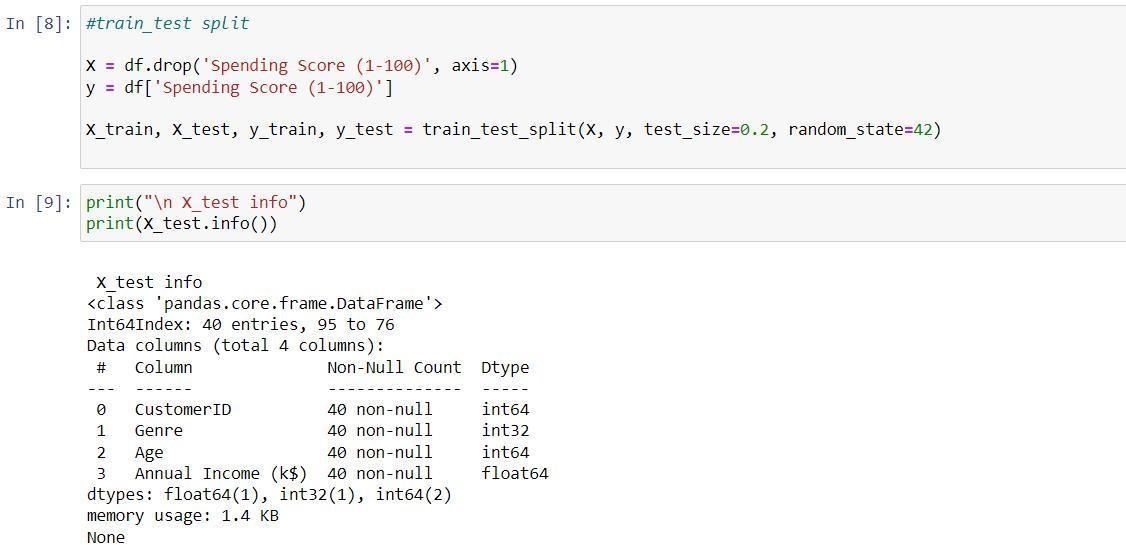
### Handling the missing data

**Step 3:** Label encoder for Genre column



### **Step 4**: Feature Scaling using StandardScaler



**Step 5**: Splitting the data into a training set and a test

**Algorithm for Customer Segmentation:**

**Objective:**

This algorithm aims to guide the development of a customer segmentation using the provided dataset. It covers essential steps, including feature engineering, model training, and evaluation, to ensure accurate predictions.

1. **Import necessary libraries:**
   * **Import essential Python libraries, including pandas, scikit-learn, and matplotlib, for data manipulation, clustering, and visualization.**
2. **Suppress FutureWarnings:**

**- Configure the system to suppress FutureWarnings to prevent unnecessary warning messages.**

1. **Read the dataset:**

* **Load the customer data from a CSV file located at a specified file path.**
* **Use the specified encoding (ISO-8859-1) to read the data.**

1. **Data Exploration:**

* **Display the DataFrame (`df`) to inspect the loaded data.**
* **Check the data's information, including data types and missing values.**
* **Display the first few rows of the dataset for a quick overview.**

1. **Handling Missing Values:**

* **Check for missing values within the dataset.**
* **Fill missing values with the mean of the respective columns.**
* **Drop rows with any remaining missing values.**

1. **Label Encoding:**

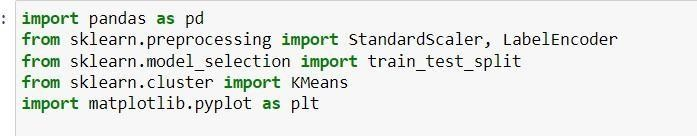
**- Apply label encoding to the 'Genre' column to convert categorical values (e.g., 'Male' and 'Female') into numerical values (e.g., 0 and 1).**

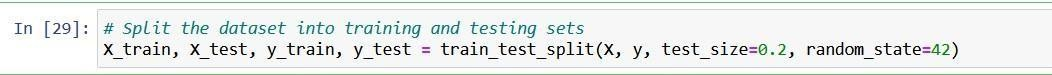
1. **Feature Scaling:**
   * **Use StandardScaler to scale the 'Annual Income (k$)' column to have a mean of 0 and a standard deviation of 1.**
   * **Standardization helps ensure that features with different scales contribute equally to clustering.**
2. **Data Splitting:**
   * **Split the dataset into features (X) and the target variable (y).**
   * **Divide the data into training and testing sets using train\_test\_split, with a specified test size and random seed.**
3. **K-Means Clustering:**

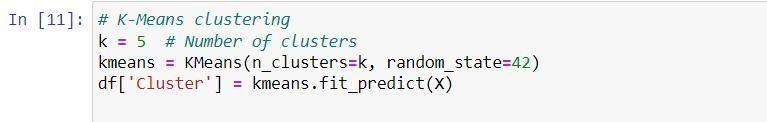
* **Define the number of clusters (k) for K-Means clustering (in this case, k=5).**
* **Apply K-Means clustering to the feature data (X) to segment customers into 'k' clusters.**
* **Assign cluster labels to the data points.**

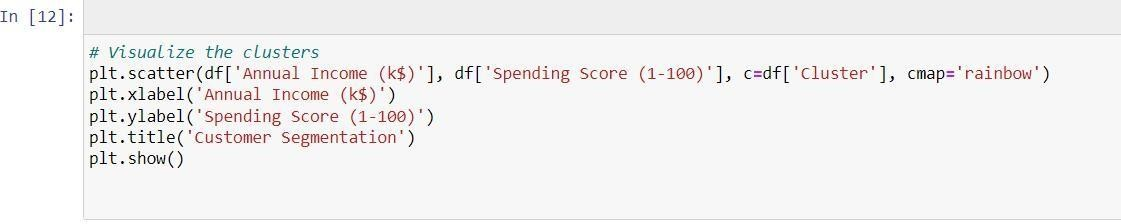
1. **Cluster Visualization:**
   * **Create a scatter plot to visualize the clusters.**
   * **Plot 'Annual Income (k$)' on the x-axis and 'Spending Score (1-100)' on the y-axis.**
   * **Color the data points based on their assigned clusters.**
2. **Cluster Analysis:**
   * **Calculate and display the center points (centroids) of each cluster.**
   * **Interpret the cluster centers' coordinates in terms of 'Annual Income' and 'Spending Score'.**
3. **Cluster Size Analysis:**
   * **Analyze the size (number of data points) in each cluster.**
   * **Print the size of each cluster to understand how customers are distributed among the segments.**
4. **Further Analysis:**
   * **Mention that further in-depth analysis can be performed within each cluster to gain insights into customer demographics, behaviors, and preferences.**

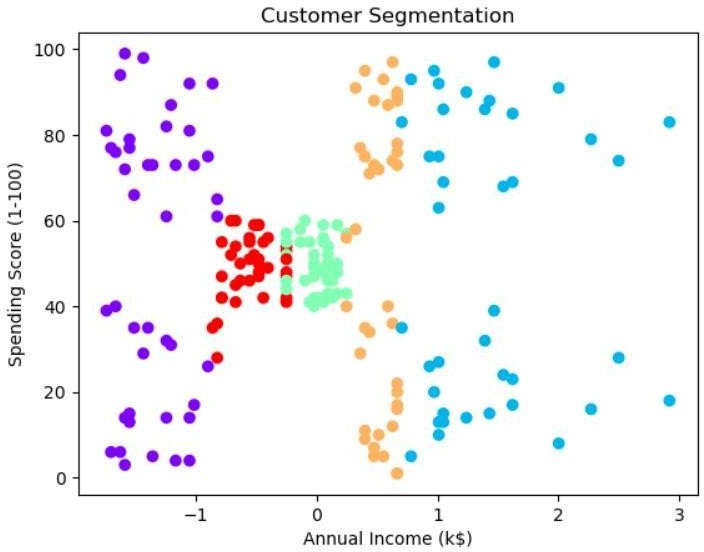
# Execution of the K-Means:

Importing the necessary libraries:

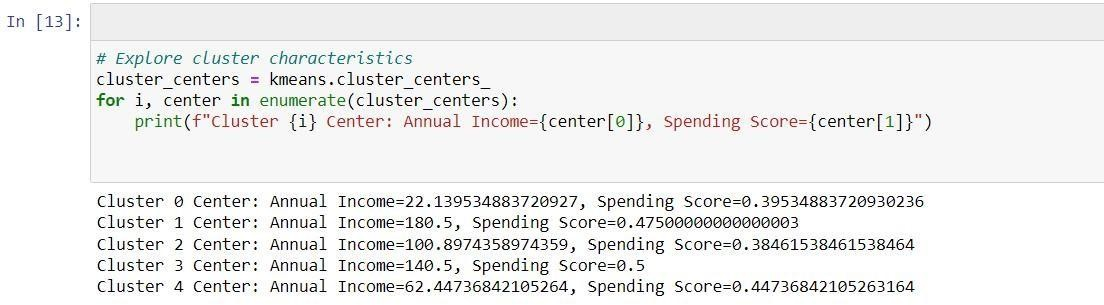
Train test split:

K-Means Clustering:

Visualization of the result:



### Explore the clusters:



Analyze each cluster for insights:



Interpretation:

The code provided offers a complete workflow for customer segmentation using K-Means clustering. It begins with data preprocessing, including handling missing values, label encoding, and feature scaling. It then performs clustering to segment customers into distinct clusters based on their 'Annual Income' and 'Spending Score.' The clusters are visualized, and cluster characteristics are analyzed. The code sets the foundation for understanding customer behavior and tailoring marketing strategies to specific customer segments. Additionally, it highlights the potential for further analysis within each cluster to gain deeper insights and make data-driven decisions.