Module Project.

Title: Customer Segmentation for Marketing.

**Members: Suyog Yadav, Sonal Pohare**

**1. Introduction**

Customer segmentation is a pivotal strategy in marketing that involves dividing customers into distinct groups based on shared characteristics such as demographics, behaviour, or purchasing patterns. This approach allows businesses to tailor their marketing efforts more effectively, offering personalized experiences that resonate with different customer segments. The Mall Customer Segmentation Dataset on Kaggle provides an opportunity to explore and apply clustering techniques to identify meaningful customer segments based on their shopping behaviours.

**2. Objective**

The primary objective of this project is to conduct clustering analysis on customer data from the Mall Customer Segmentation Dataset. By applying machine learning algorithms, specifically clustering techniques, we aim to uncover hidden patterns within the data and segment customers into distinct groups. This segmentation will enable us to better understand customer preferences and behaviours, ultimately leading to more targeted marketing strategies.

**3. Literature Survey**

* **Overview:** Customer segmentation has been widely studied in marketing literature for its effectiveness in enhancing marketing strategies. It involves partitioning customers into groups with similar needs, behaviours, or characteristics.
* **Key Approaches:**
  + **Clustering Algorithms:** Commonly used algorithms include K-means, hierarchical clustering, and DBSCAN, each offering unique advantages depending on data characteristics.
  + **Behavioural Segmentation:** This approach focuses on customer actions and purchase behaviours to define segments, allowing for personalized marketing strategies.
  + **Demographic Segmentation:** Based on customer demographics such as age, income, gender, this method helps in targeting specific groups.
* **Applications in Marketing:** Successful case studies demonstrate how segmentation improves customer retention, acquisition, and overall profitability.
* **Research Papers:**
* Jain, A., Murty, M., & Flynn, P. (1999). Data clustering: A review. ACM Computing Surveys, 31(3), 264-323.  
  This seminal paper provides an extensive review of clustering techniques, highlighting their applications and algorithms used in various domains, including customer segmentation.
* Verhoef, P. C., & Donkers, B. (2005). The effect of acquisition channels on customer loyalty and cross-buying. Journal of Interactive Marketing, 19(2), 31-43.  
  This study explores how different acquisition channels influence customer loyalty and cross-buying behaviour, highlighting the relevance of segmentation in personalized marketing efforts.
* Rust, R. T., & Huang, M. H. (2014). Optimizing service productivity: How companies can achieve profitable growth. Journal of Marketing, 78(2), 100-114.  
  This paper discusses optimization strategies in service industries through customer segmentation and personalized service delivery.

**4. Goal**

Our overarching goal is to enhance marketing effectiveness by developing tailored strategies for each identified customer segment. By leveraging insights gained from the clustering analysis, we intend to optimize marketing campaigns, improve customer engagement, and drive business growth. Through personalized marketing efforts, we aim to increase customer satisfaction and loyalty.

**5. Scope**

The scope of this project includes:

* Utilizing machine learning algorithms, particularly clustering methods (e.g., K-means, hierarchical clustering), to segment customers based on attributes such as age, income, and spending behaviour.
* Analysing and interpreting customer demographics and purchasing patterns to derive actionable insights.
* Implementing and evaluating the effectiveness of segmentation strategies through targeted marketing campaigns.
* Exploring additional features or data sources to enrich customer segmentation results, if applicable.

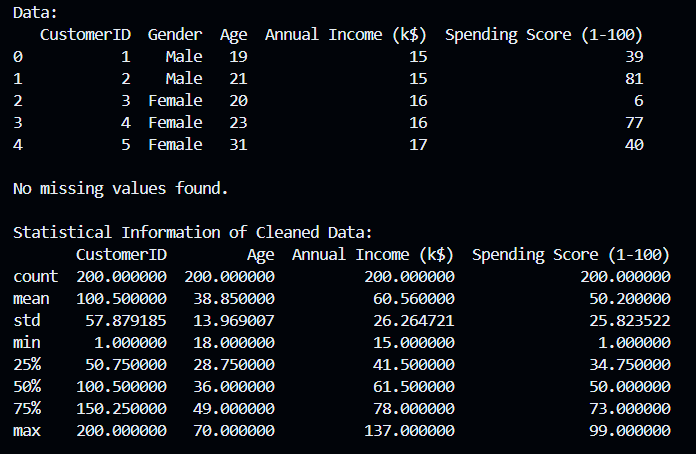
**6. Data Collection and Preparation**

**Data Source:** The Mall Customer Segmentation Dataset, sourced from Kaggle ([Dataset Link](https://www.kaggle.com/datasets/vjchoudhary7/customer-segmentation-tutorial-in-python)).

**Data Description:** The dataset comprises various customer attributes, including Customer ID, Gender, Age, Annual Income, and Spending Score. These attributes provide valuable insights into customer behaviour and preferences.

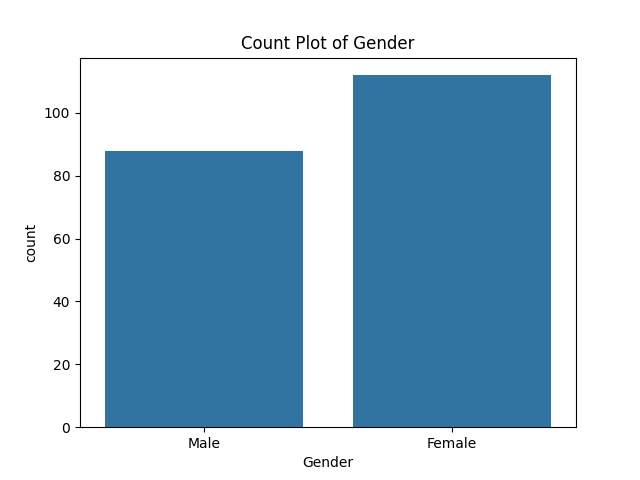
**Data Preparation Steps:**

* **Data Cleaning:** Addressing missing values and outliers to ensure data quality.

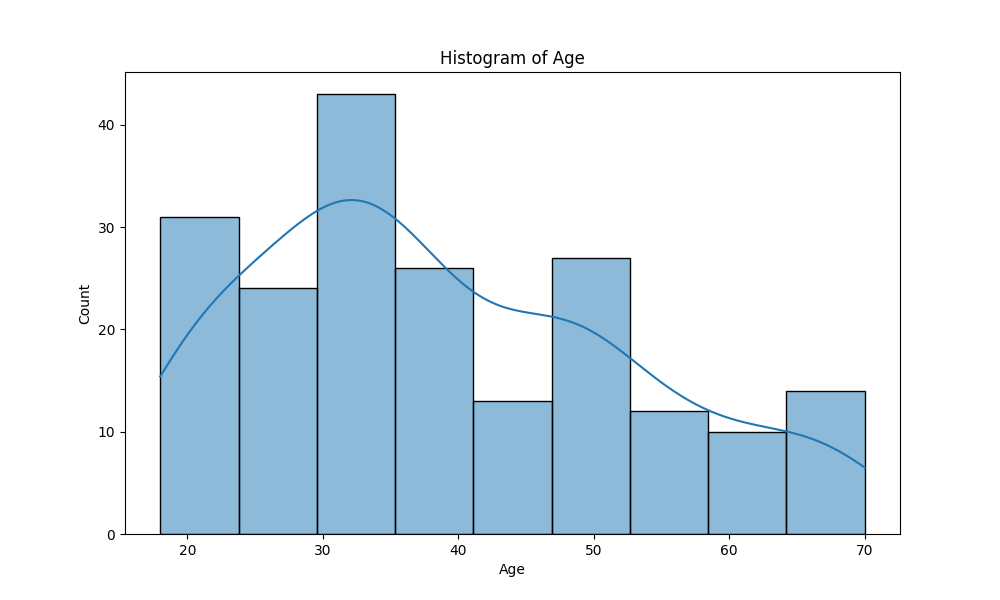
****

* **Exploratory Data Analysis (EDA):** Exploring distributions, correlations, and patterns within the dataset to gain initial insights.

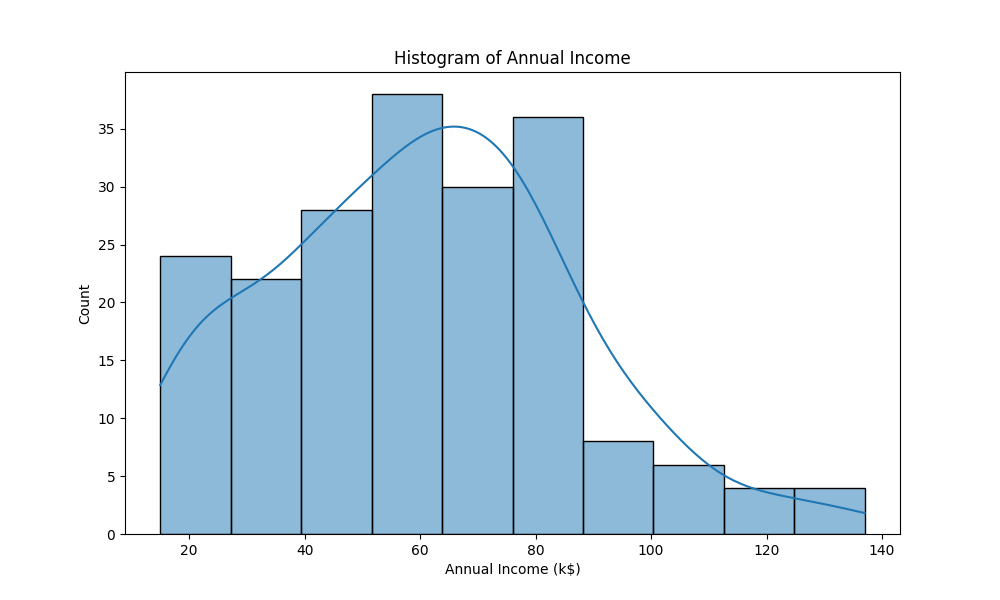
**.** Number of Customers w.r.t Genders (Male, Female).



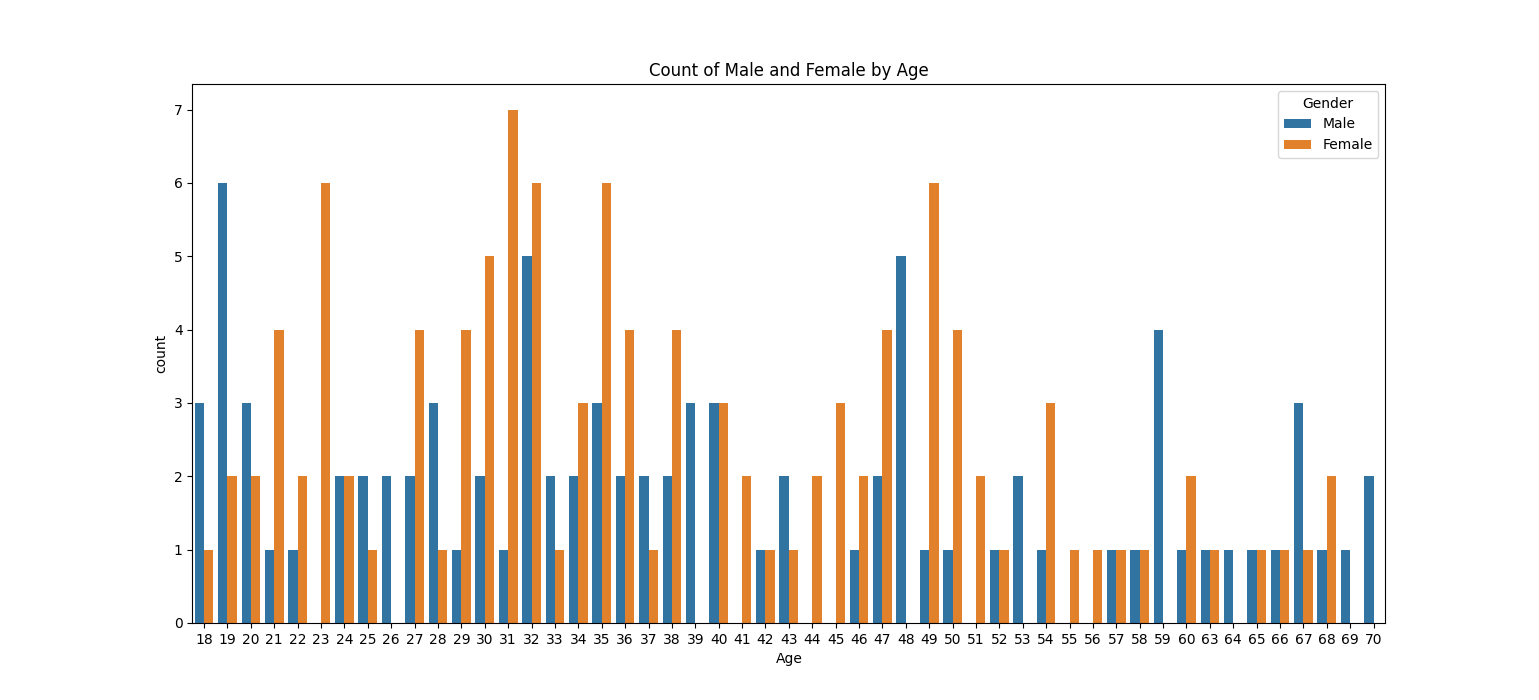
**.** Count of Customers w.r.t Age.



. Count of Customers w.r.t Annual Income (k$)



. Count of Male & Female w.r.t Age



* **Normalization:** Standardizing numerical features to ensure consistency across different scales, if required for clustering algorithms.

**7. Model Selection and Development: -**

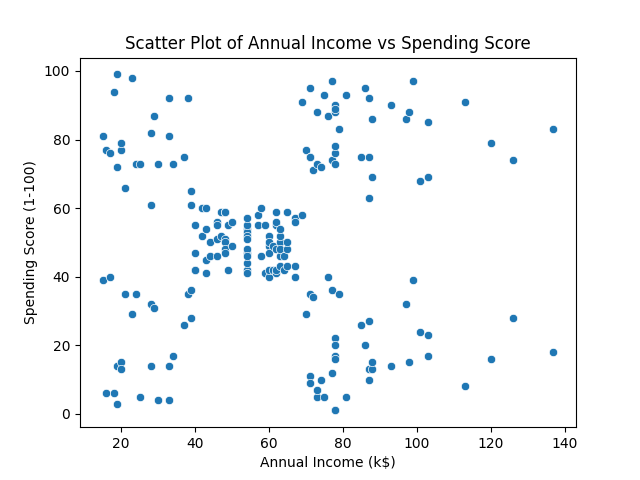
The nature of customer segmentation tasks involves clustering based on features like demographics, behaviour, etc:

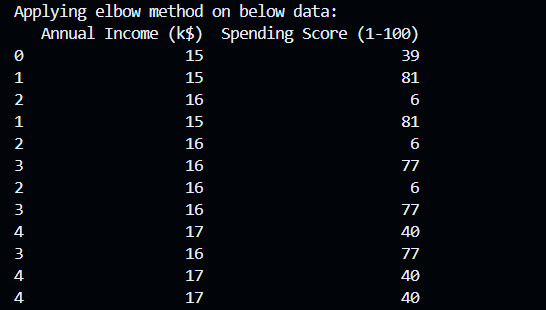
### K-Means Clustering

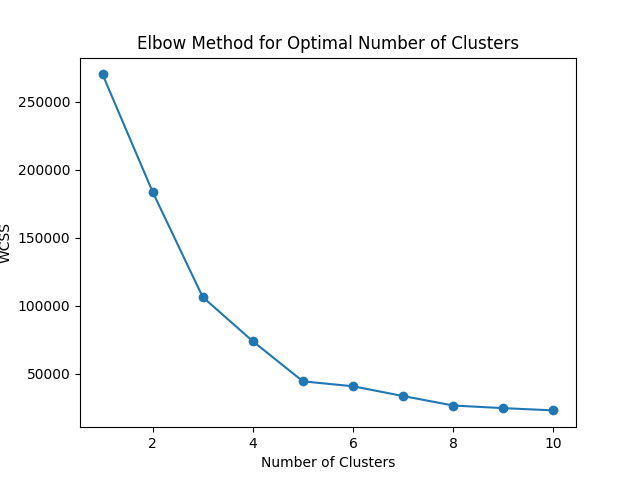
K-means clustering is a popular unsupervised machine learning algorithm used to partition a dataset into a set number of clusters. Each cluster is defined by its centroid, which is the mean of the points in the cluster.

* **Centroids**: The centre of each cluster, representing the average position of all points in the cluster.
* **Inertia**: The sum of squared distances from each point to its assigned centroid, which the algorithm aims to minimize.

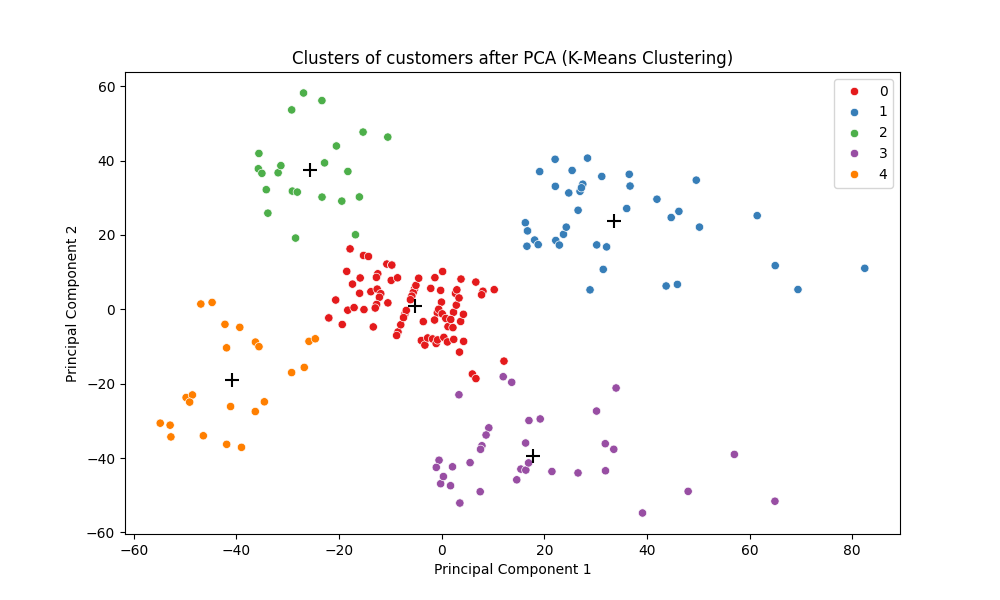
1. **K-mean Clustering for Annual Income & Spending Score.**

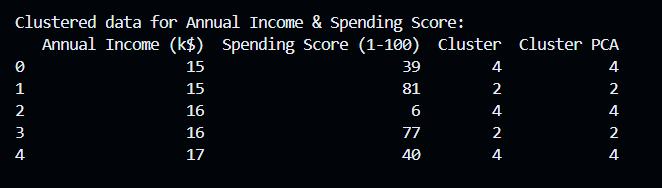


****

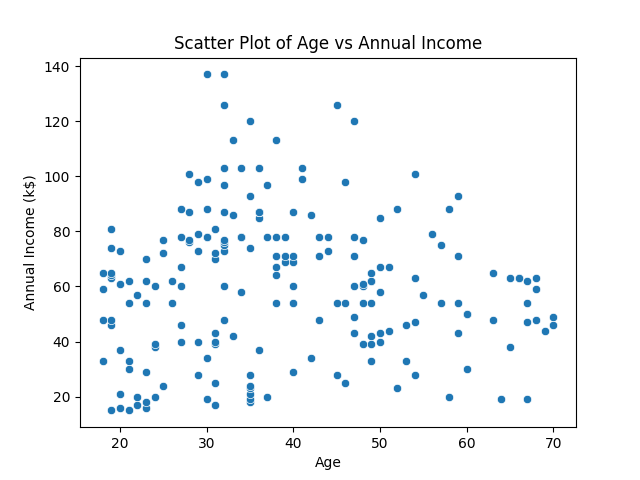


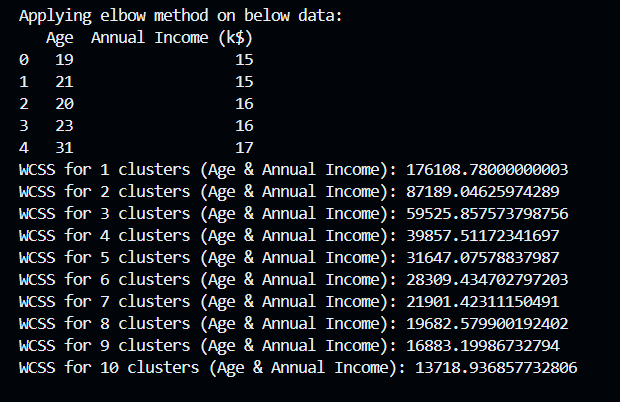


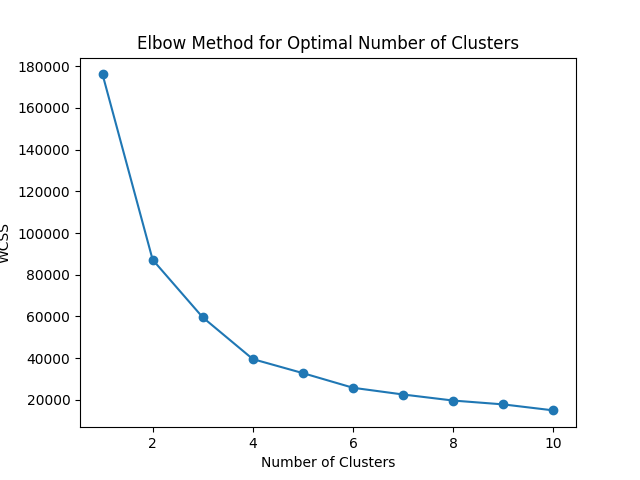


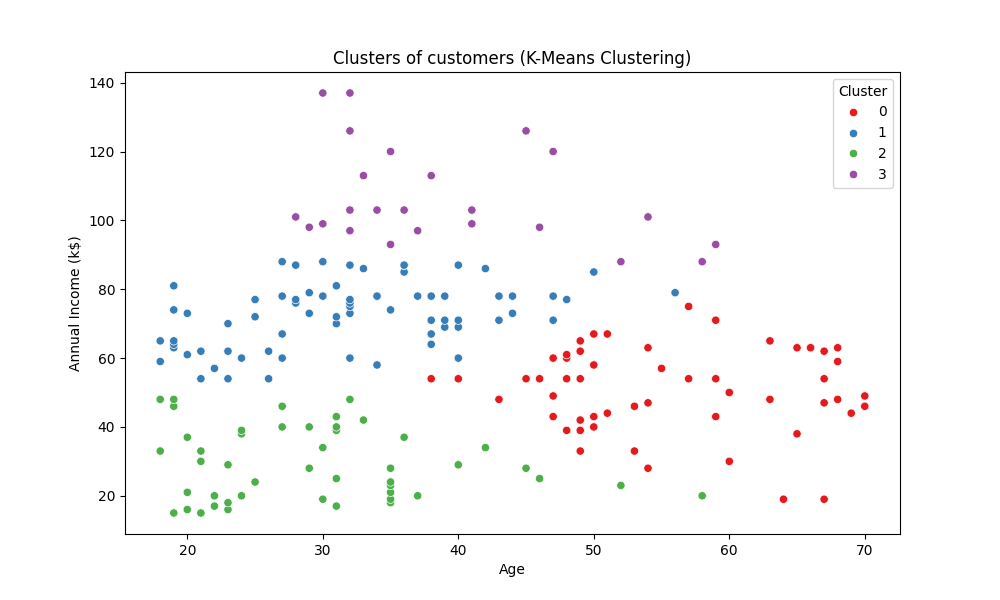


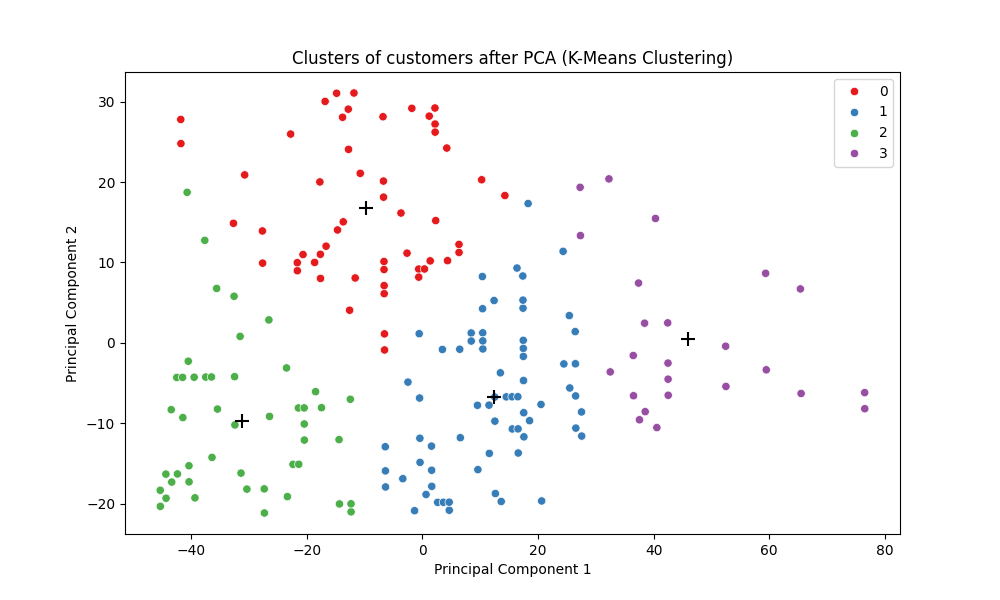
**2.K-mean Clustering for Age & Annual Income.**

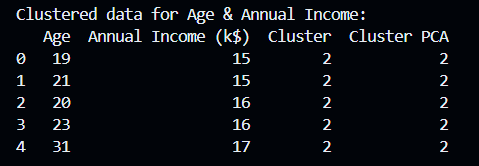




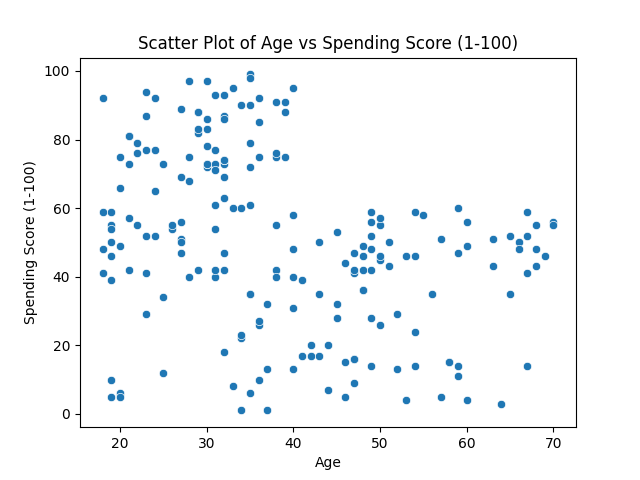


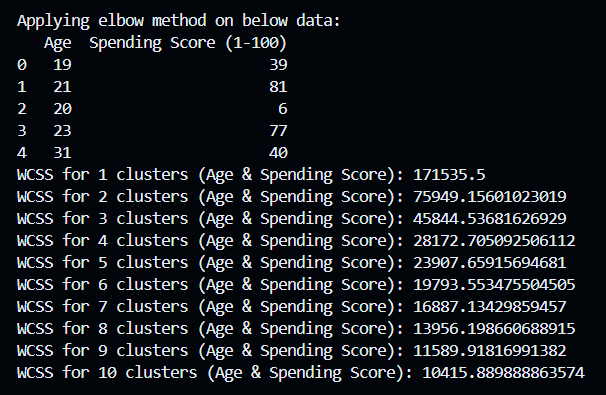


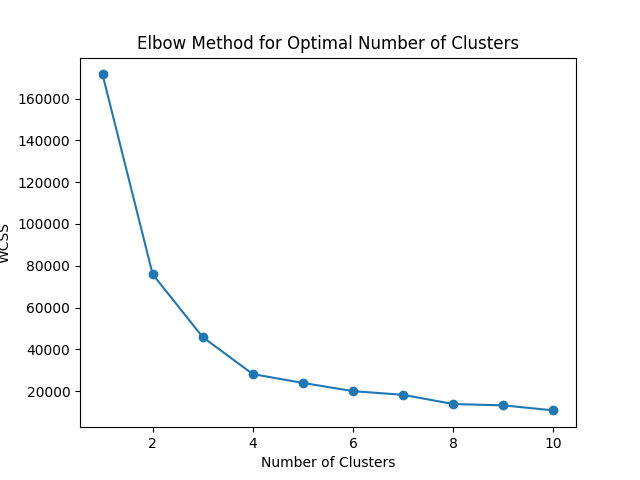


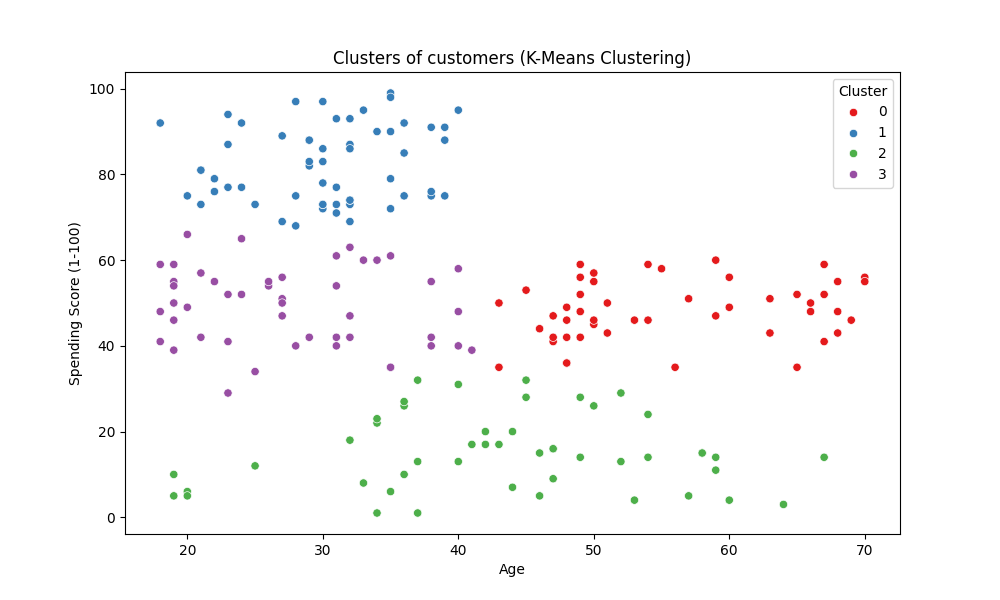


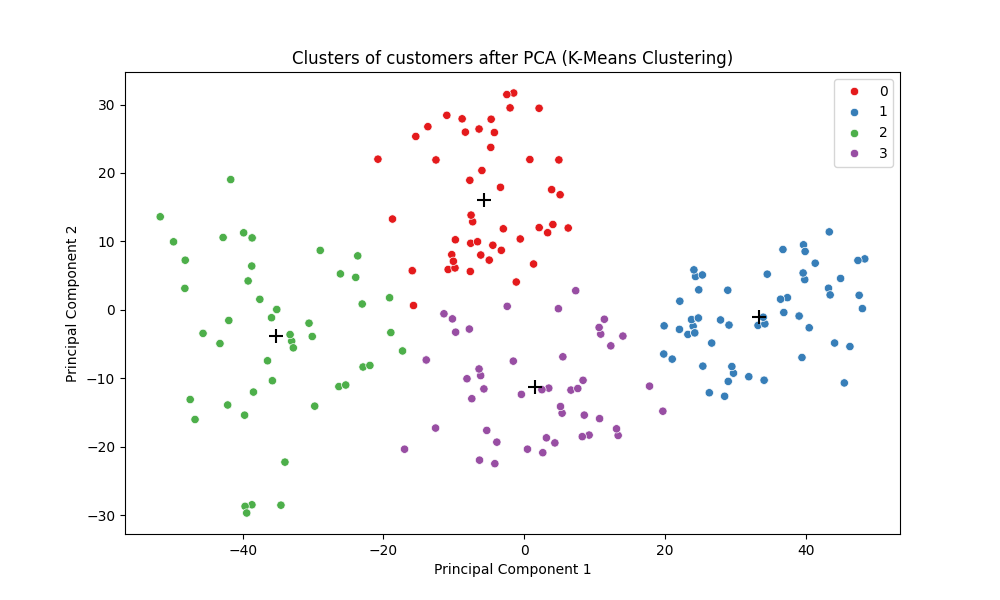
**3.K-mean Clustering for Age & Spending Score.**

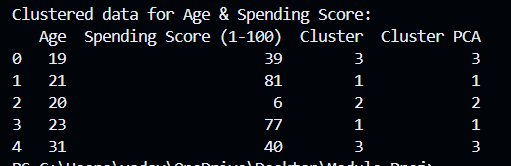












**8. Analysis:**

* Female Customers are more than Male Customers.
* Customers between Age group (30-40) are higher.
* Customers with Annual Income between (50k$ - 70k$) are higher.
* Young Customers (18 age – 22 age) with less Annual Income have high Spending Score.
* Customers (30 age – 40 age) have High Income & Spending Score
* Customers above 50 Age have High income but low Spending Score.