**Name:** Aayush Avinash Bokde   
**Class:** CSE (AI) B div  
**Roll No.:** 282016  
**PRN No.:** 22310316  
**Subject:** Machine Learning

**Assignment 3 – Data Visualization**

**Objective**

The goal of this assignment is to analyze and understand the **Mall\_Customers.csv** dataset through effective data visualization techniques. The visualizations help uncover insights related to customer demographics, income levels, and spending behavior.

**Dataset Overview**

The dataset includes important customer attributes such as:

* **CustomerID**
* **Gender**
* **Age**
* **Annual Income (k$)**
* **Spending Score (1–100)**

These attributes help explore customer segmentation and behavior patterns within a retail mall environment.

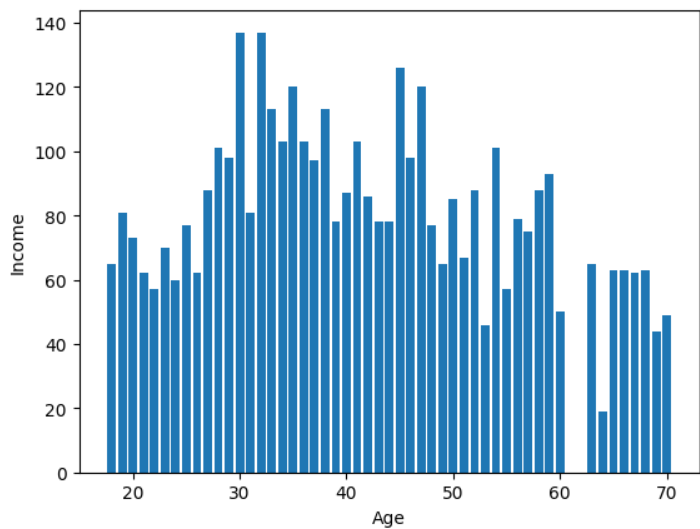
**Data Import & Preparation Steps**

* The dataset was imported using Google Colab’s file upload functionality.
* Essential libraries including **Pandas**, **NumPy**, **Matplotlib**, and **Seaborn** were used for data handling and plotting.
* The dataset was read into a DataFrame and its structure was verified using preview methods.

**Visualizations**

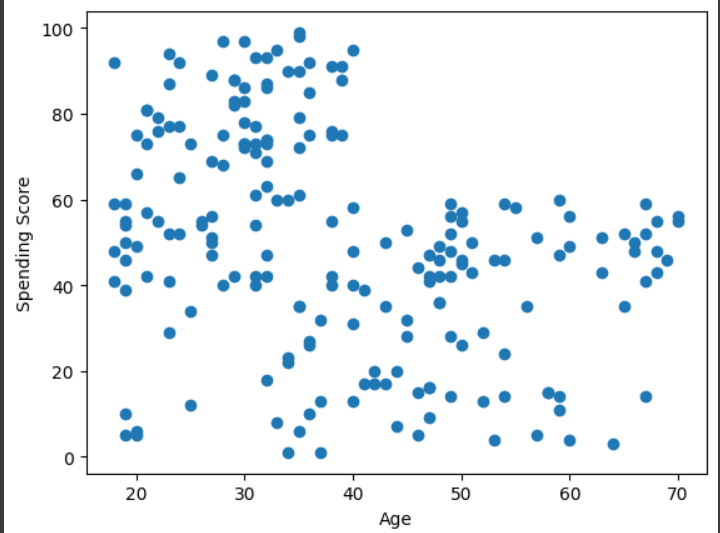
**1. Bar Plot – Age vs Annual Income**

A bar chart was created to visualize the relationship between customer **Age** and **Annual Income**. This visualization helps identify which age groups have higher income levels, giving insight into potential purchasing power.



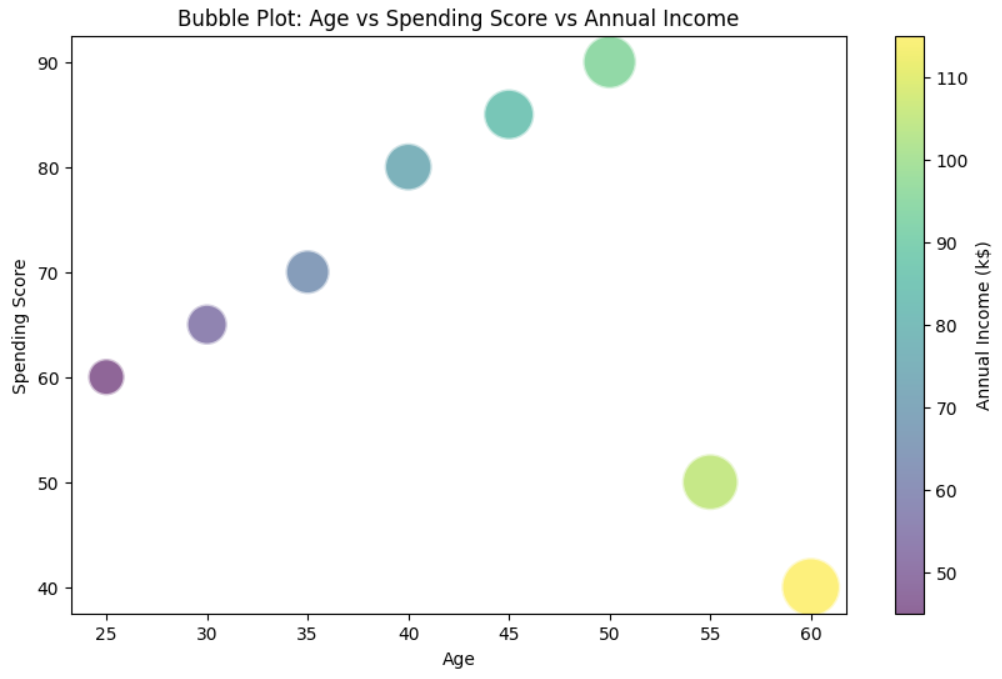
**2. Scatter Plot – Age vs Spending Score**

A scatter plot was used to observe how **Spending Score** varies with **Age**. This plot reveals spending behavior across age demographics, highlighting which age ranges tend to spend more or less within the mall.



**3. Bubble Plot (Sample Data for Reference)**

To understand the concept of bubble plots, a sample dataset was used where **Age** and **Spending Score** were plotted with bubble size and color intensity representing **Annual Income**. This visualization demonstrates how multiple variables can be represented in a single, intuitive chart.

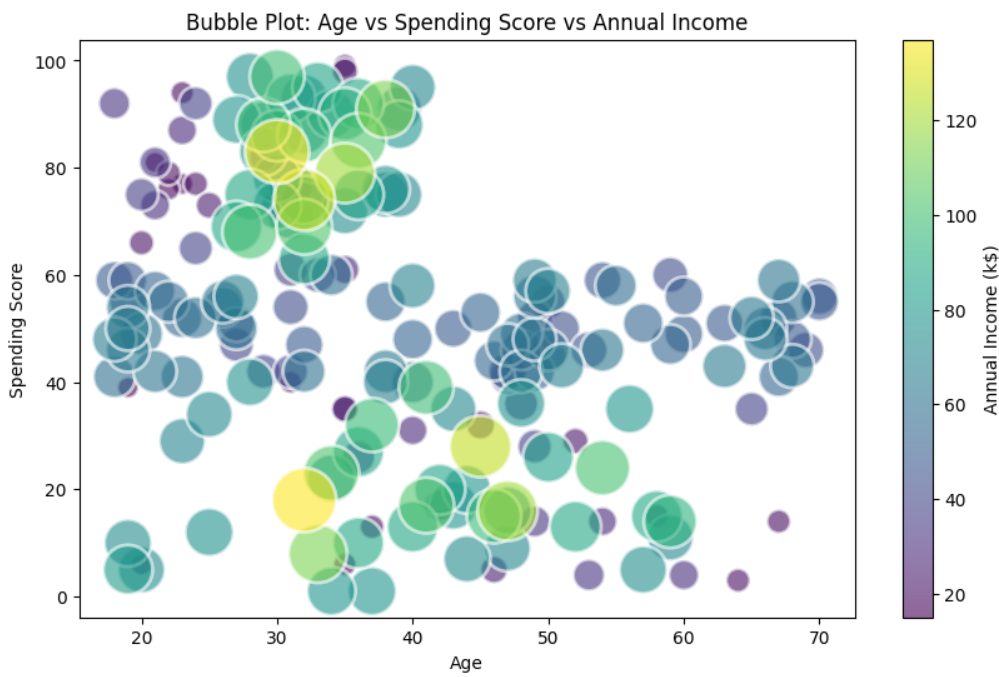


**4. Bubble Plot – Age vs Spending Score vs Annual Income (Real Data)**

Using the actual mall customer dataset, a bubble chart was created where:

* The **X-axis** represented **Age**
* The **Y-axis** represented **Spending Score**
* The **bubble size and color intensity** represented **Annual Income**

This plot provides a deeper understanding of customer clusters. For instance, it shows which customers are young, high-income, and high spenders—ideal for targeting in marketing strategies.



**Conclusion**

This visualization-based assignment provided valuable insights into customer behavior using multiple graphing techniques:

* **Bar plots** revealed the income spread across age groups.
* **Scatter plots** helped track spending behavior by age.
* **Bubble charts** were effective in capturing three variables at once—Age, Spending Score, and Income.

These visualizations play a crucial role in understanding customer segments and support decision-making in retail strategies and personalized marketing efforts.