Customer Segmentation using K-Means Clustering

🎯 Objective

The main objective of this project is to segment customers into distinct groups based on their purchasing behavior using the K-Means clustering algorithm. This enables businesses to target each customer group more effectively, personalize marketing strategies, and improve customer satisfaction.

📂 Dataset Used

The dataset used is the **Mall Customer Segmentation Data**, which contains the following key features:

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

This dataset is available on Kaggle:

https://www.kaggle.com/datasets/vjchoudhary7/customer-segmentation-tutorial-in-python

---

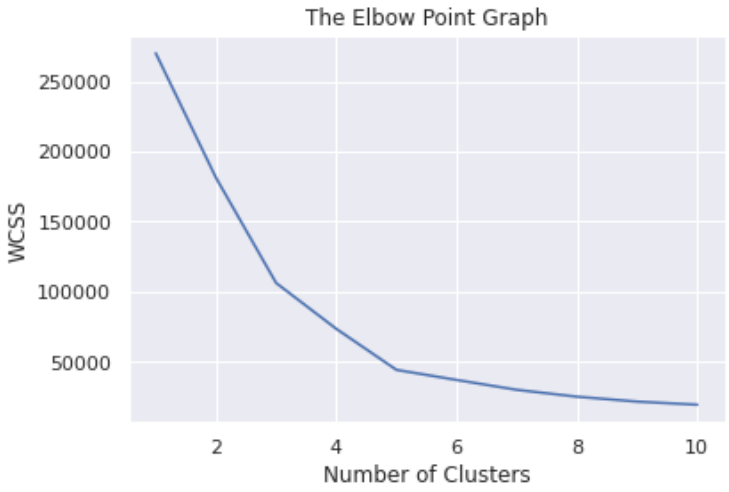
🧠 Model Chosen

The model used in this project is the **K-Means Clustering Algorithm**, an unsupervised machine learning technique. It groups data points into k clusters based on similarity, with each data point belonging to the cluster with the nearest mean.

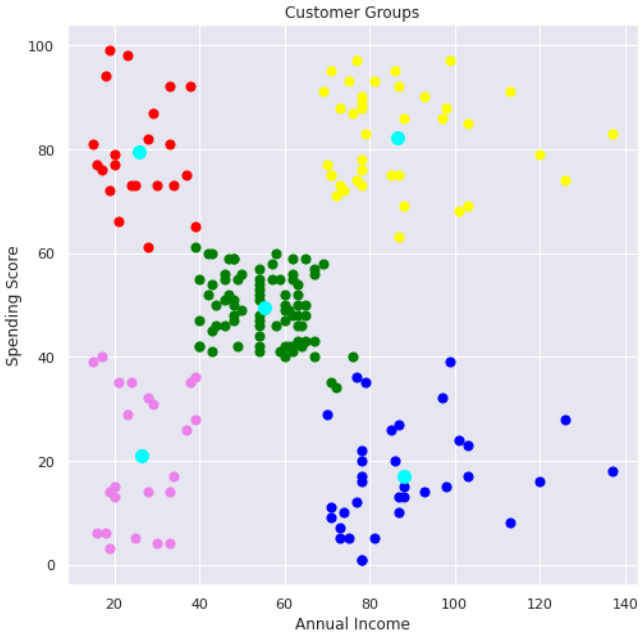
📈 Performance Metrics

As K-Means is an unsupervised algorithm, we use the following metrics and methods to evaluate its effectiveness:

- **Elbow Method**: To determine the optimal number of clusters.



- **Visualization**: Scatter plots are used to visually assess cluster separations.



⚙️ Challenges & Learnings

Challenges:

- Selecting the optimal value of k (number of clusters).
- Ensuring data is well-scaled and normalized before applying K-Means.
- Handling and preprocessing data efficiently to improve clustering quality.
- Interpreting results when clusters overlap or are not linearly separable.

Learnings:

- Gained hands-on experience with unsupervised machine learning.
- Understood the importance of feature scaling and dimensionality in clustering.
- Learned how to use visualization techniques like elbow plots and scatter plots to interpret clusters.
- Developed skills in using Python libraries such as scikit-learn, matplotlib, and seaborn for data science tasks.