**K-Means Clustering**

* **K-Means** clustering is an **unsupervised** machine learning algorithm used to group similar data points into distinct clusters.
* In the context of retail analytics, it helps segment customers based on their shopping behaviour, enabling businesses to tailor their marketing strategies.
* In this implementation, we cluster customers based on Age, Annual Income, and Spending Score to identify different customer groups. Each group may represent a unique customer type, such as:

1. High-income, high-spending customers (Premium Shoppers)
2. Low-income, high-spending customers (Budget-Conscious but Frequent Shoppers)
3. Low-income, low-spending customers (Occasional Shoppers)
4. Young customers with varying spending habits (Students, Young Professionals)

* This helps businesses design targeted marketing campaigns, loyalty programs, and personalized promotions.
* Selecting the correct number of clusters is crucial for meaningful segmentation. Two common methods used:
* **ELBOW METHOD:** Measures the inertia (within-cluster sum of squares, WCSS) for different values of K.
* The optimal K is chosen where the curve bends (elbow point), indicating diminishing returns in reducing WCSS.
* Performance Evaluation Metrics, to assess the quality of clustering, we use three metrics:

1. **Silhouette Score:** Measures cohesion within a cluster and separation between clusters. Ranges from -1 to 1, where a higher value indicates better clustering.
2. **Davies-Bouldin Score:** Measures the similarity between clusters. A lower value indicates well-separated clusters.
3. **Calinski-Harabasz Score:** Measures the ratio of between-cluster dispersion to within-cluster dispersion. A higher score indicates well-defined clusters.