Customer Segmentation using K-Means Clustering

This Python script performs customer segmentation using the **K-Means clustering algorithm**. It clusters customers based on two numerical attributes from the dataset.

---

1. Import Required Libraries

- pandas is used to load and process CSV files.
- numpy provides support for numerical computations.
- matplotlib.pyplot and seaborn are used for data visualization.
- KMeans from sklearn.cluster is used for clustering.

---

2. Load the Data

- **Customers.csv**: Contains customer profile information (e.g., ID, Name, Age, Region).
- **Transactions.csv**: Contains purchase history (e.g., CustomerID, Transaction Amount, Date).

---

3. Merge the Two Datasets

- **Merging on CustomerID**: Combines profile and transaction data into one dataset.
- **how='inner'**: Ensures that only customers with transactions are included.

---

4. Select Features for Clustering

- Here, **columns 7 and 9** from merged_df are used as clustering features.
- Ensure these columns contain numerical values (e.g., "Total Spend", "Number of Transactions").
- X is a NumPy array containing feature values.

---

5. Determine Optimal Clusters using the Elbow Method

Explanation:

- **WCSS (Within-Cluster Sum of Squares)**: Measures the compactness of clusters.

- **The "Elbow Method"**: Helps determine the optimal number of clusters by looking for a "bend" in the WCSS plot.
- **init='k-means++'**: Smart centroid initialization to improve convergence.
- **random_state=0**: Ensures reproducibility of results.

6. Plot the Elbow Curve

- A **sharp bend (elbow point)** in the graph suggests the optimal number of clusters.

---

7. Apply K-Means Clustering

- Based on the elbow method, 3 clusters are selected.
- : Performs clustering and assigns each data point to a cluster.

---

8. Visualizing the Clusters

- This plots **three clusters** using different colors.
- X[y_kmeans == 0, 0] selects customers assigned to **Cluster 0**.

9. Plot the Centroids

plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s=100, c='magenta', label='Centroids')

- This plots the **cluster centers** in **magenta**.

10. Finalize the Plot

- Adds a title, labels, and a legend to the plot.

Final Output

- **Elbow Curve**: Determines the best number of clusters.
- **Cluster Scatter Plot**: Shows the three customer groups with different colors.
- **Centroids**: Represent the center of each cluster.
