ZEOTAP DATA SCIENCE ASSESSMENT

Task 3: Customer Segmentation / Clustering

Perform **customer segmentation** using clustering techniques. Use both **profile information** (from `Customers.csv`) and **transaction information** (from `Transactions.csv`).

- You have the flexibility to choose any clustering algorithm and any number of clusters in between(2 and 10)
- Calculate clustering metrics, including the **DB Index(Evaluation will be done on this)**.
- Visualise your clusters using relevant plots.

Deliverables:

- A report on your clustering results, including:
  - The number of clusters formed.
  - DB Index value.
  - Other relevant clustering metrics.
- A Jupyter Notebook/Python script containing your clustering code.

For this task, I have done customer segmentation using K-means Clustering Algorithm as it is simple and effective for our purpose.

First, I have loaded the data and imported the necessary libraries. Then we move on to the preprocessing of the data as per our requirements. I have converted the SignupDate and TransactionDate columns into a datetime format to work with them effectively. I scaled all numerical features using the StandardScaler.

Metrics Calculated:

- **Davies-Bouldin Index (DBI):** Lower values indicate better-defined clusters.
- **Silhouette Score:** Measures how well-separated clusters are (values closer to 1 are better).

I tested cluster sizes from 2 to 10 and plotted the DBI and Silhouette scores.

The optimal cluster count was identified as **3**, based on the lowest DBI value and highest Silhouette Score.

To understand the clusters better, I visualized the results:

1. **Scatter Plot with PCA:** I reduced the dimensions of the dataset to two using PCA and plotted the clusters in 2D space. This provided a high-level view of how the clusters are separated.
2. **Cluster Centers:** I visualized the feature values of each cluster's center to understand what differentiates one group from another.

Thus I was able to successfully complete this task of customer segmentation to the best of my ability and knowledge.

THANK YOU,

SUSHRITH REDDY