Part B – Reflection Paper

1. What I Implemented

In this project, I built a **customer spending segmentation model** using **K-Means clustering**. I selected **Income_$** and **SpendingScore** as the two features for clustering. First, I checked for missing numeric values and filled them using the **median** to avoid skewing the data. Then I scaled the features with **StandardScaler** so that income (measured in dollars) and spending score were on the same scale. Next, I ran a loop for **k = 1 to 10**, fitting a K-Means model each time and printing the **Sum of Squared Errors (SSE)** to perform an elbow check. Based on the SSE results, I chose the best value of **K**, trained the final K-Means model, and added a new **Cluster** column to the dataset. Finally, I evaluated the clustering quality using the **Silhouette Score** and the **Davies–Bouldin Index (DBI)**, and I inverse-transformed the cluster centers back to their original units to interpret the results in real-world terms.

2. Choosing K

The printed SSE values decreased rapidly up to **k = 3** and then started to level off, creating a clear elbow at **K = 3**. At this value, the **Silhouette Score** was high and the **Davies–Bouldin Index** was low, confirming that the clusters were both well-separated and compact. Because K = 3 offered the best trade-off between simplicity and performance, I selected **3 clusters** for the final model.

3. Cluster Interpretation

The three clusters can be described as follows:

Cluster	Description	Business Action
0 – Low Income / High Spending	Customers with modest income but high discretionary spending.	Offer **loyalty rewards or discounts** to retain these enthusiastic buyers and encourage repeat purchases.
1 – Mid Income / Moderate Spending	Customers with average income and balanced spending habits.	Use **seasonal promotions** or personalized recommendations to maintain engagement and gently increase spending.
2 – High Income / Low Spending	High earners who currently spend cautiously.	Provide **premium upsell opportunities** (VIP programs, exclusive products) to capture more of their potential spending power.

These insights can guide marketing strategies such as targeted offers and customer retention plans.

4. Limitations & Next Steps

This analysis only used **Income_$** and **SpendingScore**, which limits the depth of customer profiles.

Adding other features such as **Age**, **VisitsPerMonth**, or **OnlinePurchases** could reveal more detailed spending behaviors and improve segmentation. As a next step, I could expand the model to include **three or more features**, experiment with **DBSCAN or Hierarchical Clustering**, or compare K-Means results with other unsupervised techniques to discover more complex customer patterns.