﻿**DSAA 5002 - Data Mining and Knowledge Discovery in Data Science**

**Final Exam Report – Q6 ﻿Bank Customer Clustering**

**50015940 Jiaxiang Gao**

1. **Visualization and Business Insights**
   1. **Distribution of Transaction Amount (INR)**
   2. **Distribution of Customer Account Balances**
   3. **Distribution of Customer Gender**
   4. **Distribution of Customer Age**
   5. **Customer Count and Total Account Balance in Top 50 Locations by Customer Count**
   6. **Transaction Volume by Hour of Day**
   7. **Relationship Between Transaction Amount and Account Balance by Gender**
   8. **Relationship Between Transaction Amount and Customer Age by Gender**
   9. **Relationship Between Account Balance and Customer Age by Gender**
   10. **Recency Distribution from RFM Analysis**
   11. **Frequency Distribution from RFM Analysis**
   12. **Monetary Distribution from RFM Analysis**
2. **﻿** **Customer Clustering Using Three Distinct Algorithms**

For this task, I use three features: ‘CustomerAge’, ‘TransactionAmount (INR)’ and ‘CustAccountBalance’.

For K-Means and Hierarchical clustering, we used sample data to calculate the silhouette score, and then use the silhouette score to determine the number of clusters.

For K-Means, We Use the entire dataset to calculate the clusters.

For Hierarchical and DBSCAN, we use the sample data to calculate the clusters, since the entire dataset is too large to calculate.

* 1. **K-Means (Full Dataset)**
  2. **Hierarchical** **(10000 Sample Dataset)**
  3. **DBSCAN (10000 Sample Dataset)**

1. **Understanding Customer Clusters: Common Characteristics and Differences**

**Reference**

https://connectif.ai/en/blog/what-are-rfm-scores-and-how-to-calculate-them/