**SECTION B: Data Analytics - ML & Business Analytics**

1. Q2. With the data above, perform clustering analysis or customer segmentation. Create a Report of your approach and explain the different segments or clusters you have business wise.

Steps taken to get my clusters

1. I downloaded the dataset from the link provided.
2. I imported the necessary libraries needed for Exploratory Data Analysis(EDA)
3. I loaded the dataset into the Jupyter Kernel.
4. I performed Exploratory Data Analysis on the datasets. The goal was to find out more about the data and become a subject matter expert on the dataset.
5. Checked for the missing values and substitute them with the median of the individual column for the integers.
6. Checked for the missing values and substitute them with “Missing” using the fillna method on the object columns.
7. Converted all the categorical variables to integers.
8. I used StandardScaler to re-scale the data to have a mean of 0 and standard deviation of 1(Unit Variance).
9. Now to model the clusters, I used Clustering using K-Prototypes.
10. Wrote the code to get the optimal K using the Elbow method plot with Cost.
11. Using the Elbow method I came into the conclusion that I have 4 Optimal K values.
12. Then I carried out the clustering analysis with the number of clusters set to 4
13. I then merged the original data with clusters.
14. Then used a graph to analysis the clusters with respect to the original datasets
15. Analysis gotten from the 4 clusters can be interpreted below:
16. **Cluster 0:** Customers here mainly use channel 8 and a little bit of channel 6&10. They don't spend much as compared to those in Cluster 1, they have little transaction count and purchase product from reseller within the ID number 0-464 while a little use reseller with the ID number 2379. They purchase product across all service providers which majority use for Telecom providers in the area of Data.
17. **Cluster 1:** Customers here mainly use channel 1 and some use channel 2, heavy spenders as their TotalTransactionAmount is on the high side. They have high transaction count and they are very active customers. They purchase goods and product mainly from reseller with the ID number 0&56 while a few uses reseller with the ID 1130. They purchase product mainly from MTN, 9Mobile, Glo and Airtel which majority use for Airtime while some use it for Data. These set of customers are group as **Elite Customers**.
18. **Cluster 2:** Customers here mainly use channel 2 and a little bit of channel 1,3,4,5,6,8. They don't spend much as compared to those in Cluster 1, they have fair bit of transaction count and purchase product from reseller within the ID number 0-553 while a few use reseller with the ID number 2557. They purchase product mainly from MTN and few purchase from 9Mobile, Glo and Airtel which majority use for Airtime while some use it for Data.
19. **Cluster 3:** Customers here mainly use channel 1 and some use channel 2, they spend little. They have little transaction count, and purchase product from reseller within the ID number 2379-2557. They purchase product mainly from MTN, 9Mobile, Glo and Airtel which majority use for Airtime while some use it for Data