Week 9 - Deliverables

Team member's details

Member 1

Name: **Rayan Yassminh**

Email: ryassminh@yahoo.com

Country: USA

Member 2

Name: **Kavinilavan Muthukumar**

Email: mukavin@gmail.com

Country: United Kingdom

College: Teesside University

Group Name   : STEM Group

Specialization : Data Science

Project Name : Customer Segmentation

Problem description

Customer Segmentation approach-
Separate the customers into 5 different groups based on their behaviors and Lifetime values.

Data cleansing and transformation

1.Some Data quality issues. Same subsets of the group have different columns (one hot encoding format). For instance term deposit and account types.problems are difficult to identify the customers which have no term deposit and number of accounts types they have while doing exploratory data analysis. So merging the different columns into single columns which belong to the same categorical group.

2.Features names are inappropriate: all the names of the columns are in spanish, we convert them into english with rename methods in pandas.

3. More than 80% of records are missing particular features: simply drop such columns because those columns are not very crucial for model development.

4. Since we have two members in our groups, we come up with two different methods of handling null values.

1) Simple Imputer
2) IterativeImputer

Github Repo link

Link
https://github.com/kavinilavanM/CustomerSegmentat
iondataglacier.git