**Customer Segmentation for Credit Card**

**In the being time, the importance of data becomes high priority at a lot of decision maker as well as the data exponentially grows, so it must invest the precious treasure or as said row oil, so we can’t achieve any benefits without mining or analysing the data.**

**In this report, we will introduce by using ML models specifically in Financial field ( Credit card ) some conclusions from analysing the behavior of customers during using the credit card to enable us direct suitable campaigns for the suitable customers in suitable time and in suitable way.**

**The data we given us is about some conducts of client during use the credit card as : balance, credit limit, how much pay, how much money use to purchase ..etc, it is a structured data with 18 columns and 8950 customers as well as there is no missing values except in min\_payment with 0.013 of total values also most of features are skewed to right .Furthermore, there is good correlation from moderate to strong at the majority of attributes with 0.27 of cell are zero values.**

**Some problems and suggested solution:**

* **Problem1 : Missing values**
* **Sol : imputing the null values with the median of feature due to the distribution is skewed to right.**
* **Problem2 : skewed data or existence of outlier :**
* **Sol : handled with using exponential and square transformation, the number of outlier decreases but does not on the conclusion of sillhoutte score, so we leave them.**
* **Problem : some of feature has different Range.**
* **Sol : using standardScaler it is good due to being less sensitive to outlier.**
* **Problem : Reducing from probability of showing the suboptimal solution .**
* **Sol : k-mean ++ initialization used and depending on sillhoutte score and sillhoutte diagram to determine optimal value of k.**

**Some Conclusions about EDA :**

* **Correlation of most features are very good.**
* **Number of zeros values are 27%.**
* **No duplicated rows.**
* **17 feature numerical and 1 is object.**
* **Big number of outliers.**
* **Most of customers prefer high Tenure (12 month) , this can increase utilization or increase the credit limit.**
* **Some try to make their balance near to 0 to improve the credit score or to reduce from interest.**
* **For the most less balance , more Percent of full payment paid by user.**
* **These feature produce highest sillhoutte score : ['BALANCE', 'CREDIT\_LIMIT', 'PURCHASES'].**
* **The elbow value at k = 2.**
* **Highest sillhoutte at k = 2.**
* **Closest clusters at sillhoutte score mean was at k = 2,4.**

**Some conclusions about Clusters at k = 2 :**

* **For the first cluster : they have balance from low to high the most are range (0- 1000) and some extend to 15000, For the credit limit from average to high and for purchases (0- 12000) and some of them tend to high purchases.**
* **For the second cluster : they have balance less than cluster 1 (0 -4000) therefore credit limit (0-1000), for purchase is low and most of them is near to zero.**