## **PROJECT PROPOSAL**

**AML Group 16:** Ziwei Han(zh2484), Anne Lin (al4215), Xingcheng Rong (xr2150), Chandan Suri (CS4090), Yunchen Yao (yy2949)

**BACKGROUND - PROBLEM STATEMENT**

Nowadays, customer segmentation becomes important in both marketing and product strategies because it helps a business to better understand its customers. In marketing, customer segmentation can help businesses attract customers more effectively by suggesting specific types of campaigns for different groups of customers. In product improvement, customer segmentation makes it easier to modify products according to different types of customers’ specific needs, behaviors, and concerns. In this project, we will use the Customer Personality Analysis dataset on Kaggle to analyze and use unsupervised machine learning methods to categorize those customers into specific groups.

**DESCRIPTION OF THE DATASET (RELEVANT DETAILS)**

**PROPOSED ML TECHNIQUES**

As we need to group the data into customer segments based on the customer analysis. The data is very well structured and performs decent knowledge that can be used to cluster the groups based on the high dimensional data. Furthermore, this problem statement is inherently forming groups which we don’t know about as the data is not labeled. This insight from the data gives us the basis that we need to perform some kind of unsupervised learning which involved forming groups. Thus, we would choose some kind of “Clustering” algorithm in order to achieve this task.

Moreover, for deciding which Clustering technique would suit this problem statement the best will require a lot of data analysis and pre-processing. Just by looking at the data, some of the Clustering algorithms that come to my mind are as follows:

1. K-Means Clustering
2. BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)
3. Gaussian Micture Model
4. DBSCAN (Density-Based Spatial Clustering of Applications and Noise)

**REFERENCES:**

1. Dataset: <https://www.kaggle.com/imakash3011/customer-personality-analysis>
2. Clusering Methods: <https://machinelearningmastery.com/clustering-algorithms-with-python/>