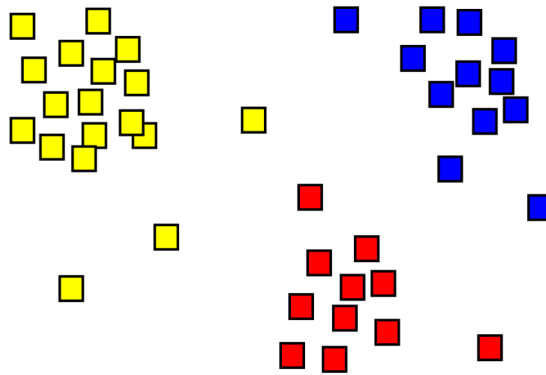


## Atelier 3 «Clustering »

### Objective :

l'objectif principal de cet atelier est de pratiquer les concepts du clustering, en traitant les données d'une Data Sets.



**Outils :** Python , Pandas, Sklearn, matplotlib.

### Data Sets :

Credit Card Data Set : <https://www.kaggle.com/arjunbhasin2013/ccdata>

### Partie 1 (Data Visualisation):

1. En utilisant pandas essayer d'explorer les données du Data set.
2. Afficher le résumer statistique du Data Sets avec une interprétation des résultats obtenues.
3. Afficher les nuages des points du data set selon les propriétés « Features » en utilisant matplotlib et pandas « scatter\_matrix ».



**Université Abdelmalek Essaadi**  
**Faculté des Sciences et techniques de Tanger**  
**Département Génie Informatique**  
Cycle Ingénieur: LSI s4  
Machine Learning  
Pr . EL AACHAK LOTFI  
2019/2020



## **Partie 2 (Clustering ):**

1. Essayer de construire le modèle de clustering en utilisant sklearn ainsi l'algorithme Kmeans,
2. Définir le K nécessaire pour avoir un modèle efficace selon la méthode de Elbow.
3. Présenter les clusters obtenus dans un graphique en utilisant matplotlib.
4. Interpréter les résultats obtenus.
5. refaire la même chose en utilisant l'algorithme fuzzy cmeans « il faut utiliser la bibliothèque skfuzzy ».
6. Comparer les deux algorithmes.

**Note : un rapport bien détaillé doit être rédigé et imprimé.**

### **Références :**

<https://towardsdatascience.com/k-means-clustering-with-scikit-learn-6b47a369a83c>  
[http://eneskemalergin.github.io/blog//blog/Fuzzy\\_Clustering/](http://eneskemalergin.github.io/blog//blog/Fuzzy_Clustering/)