



Université Sultan Moulay Slimane
Ecole Supérieure de Technologie
Fkih Ben Salah



Projet

Module : Programmation Python

Filière Licence- BIG DATA, Semestre : 5, AU : 2024-2025

Travail N°1 :

Partie 2

1. Comment interpréter la valeur 0 dans la colonne *Inter_achat* ? Remplacer cette valeur par la valeur moyenne de cette colonne (en prenant en compte que les valeurs différentes à zéro).
2. Sauvegarder le data frame final sous le nom "LRFMP1P2IT.xlsx" avec comme première colonne *Customer_ID*.

Après le prétraitement de données (nettoyage, transformation, normalisation), un modèle de clustering (segmentation) a été appliqué sur les données de notre DataFrame final qui contient les colonnes suivantes : *Customer_ID*, *Frequency*, *Recency*, *Length*, *Monetary*, *NbrP1*, *NbrP2* et *Inter_achat*).

Le modèle de clustering de client a donné comme résultat de regroupement sept clusters. Les résultats sont enregistrés dans le fichier json nommé "Clustering_Model_result.json"

3. Charger le fichier généré par le modèle de clustering ("*Clustering_Model_result.json*") sous forme d'un DataFrame.
4. Faire les traitements nécessaires pour ajouter au dernier DataFrame généré dans la partie 1, une nouvelle colonne nommée *Cluster*, qui indique le N° du Cluster de chaque client.
5. Pour chaque groupe de client, calculer le nombre correspondant de clients, la longueur moyenne (L), la récurrence moyenne (R), la fréquence moyenne (F), le montant moyen (M).
6. Ajouter une colonne nommée *LRFM-Pattern* qui va indiquer le pattern LRFM de chaque cluster (Comparer chaque valeur moyenne des clusters avec celle de la valeur moyenne total) (Voir tableau ci-dessous).

Cluster	Count	Mean(L)	Mean(R)	Mean(F)	Mean(M)	Pattern
Cluster_0	332	282,31	211,26	12,41	8204,28	$\underline{L} \overline{R} \overline{F} \overline{M}$
Cluster_1	760	752,87	48,31	4,47	857,62	$\overline{L} \underline{R} \underline{F} \underline{M}$
Cluster_2	375	707,71	206,59	13,34	2187,08	$\overline{L} \overline{R} \overline{F} \underline{M}$
Cluster_3	509	742,81	210,91	14,65	10817,28	$\overline{L} \overline{R} \overline{F} \overline{M}$
Cluster_4	210	741,10	209,95	6,97	8266,72	$\overline{L} \overline{R} \underline{F} \overline{M}$
Cluster_5	428	699,51	212,89	3,93	1063,69	$\overline{L} \overline{R} \underline{F} \underline{M}$

7. Interpréter les résultats trouvés (Essayer d'expliquer les habitudes ou bien les comportements d'achats de chaque cluster en comparant les valeurs moyennes de chaque attribut de chaque cluster avec les valeurs moyenne totales).

Exemple : Si la valeur moyenne de R d'un cluster est supérieure à celle de la moyenne totale, on dit que la valeur de R est élevée, ce qui signifie que les clients appartiennent à ce cluster ont récemment fait un achat.

8. Réaliser les visualisations suivantes :
- Nombre de clients par Cluster
 - Total dépensé par Cluster
 - Nombre de transaction par Cluster
 - Panier moyen par Cluster
 - Pour le cluster qui représente les nouveaux clients, réaliser un graphe qui montre la comparaison entre le NbrP1 et NbrP2