

# PROJET 5: DATA SCIENCE

Alexandre Jacqueline



# Plan de Soutenance



PRÉSENTATION  
DU PROJET



CIBLAGE ET  
NETTOYAGE DES  
DONNÉES



ANALYSE  
EXPLORATOIRE



SEGMENTATION



QUESTIONS –  
RÉPONSE

# Présentation du projet et objectifs

- Comprendre les différents types d'utilisateurs
- Fournir une description marketing actionnable de la segmentation
- Evaluer la fréquence à laquelle la segmentation doit être mise à jour
- Effectuer un devis de contrat de maintenance



# Workflow

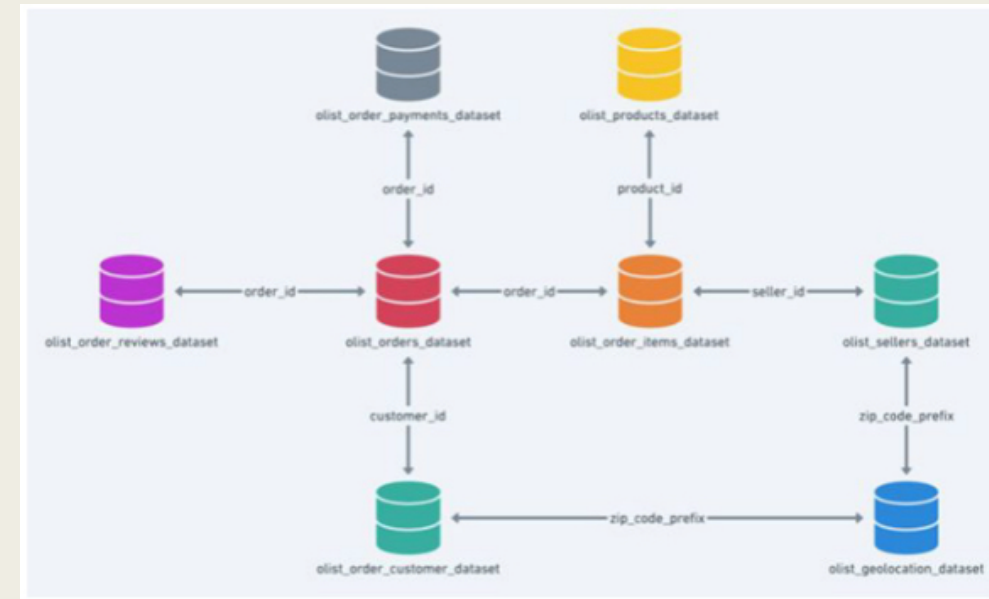
# Présentation du jeu de données

FICHIER (.csv)	NB LIGNES	NB COLONNES	DESCRIPTION
Customers	99441	5	Données clients
Geolocalisation	1000163	5	Données
Order_items	112650	7	Articles commandés
Order_payments	103886	5	Paieement commandes
Order_reviews	99224	7	Avis commandes
Orders	99441	8	Commandes
Products	32951	9	Produits
Sellers	3095	4	Vendeurs
Product_category_name_translation	71	2	Table de traduction des noms de produits

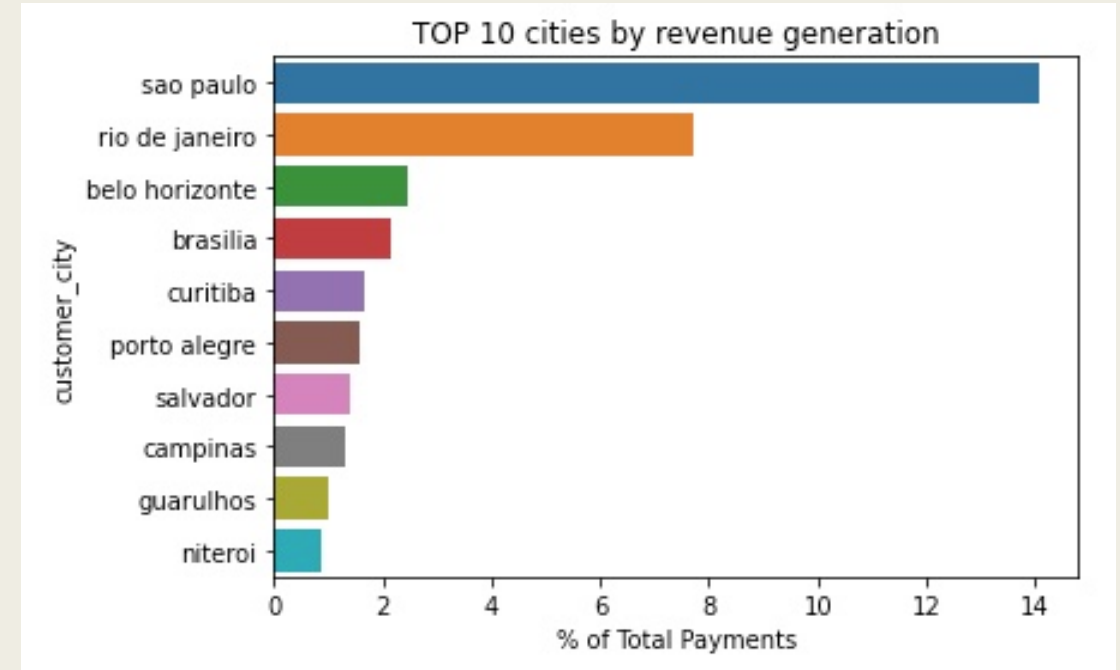
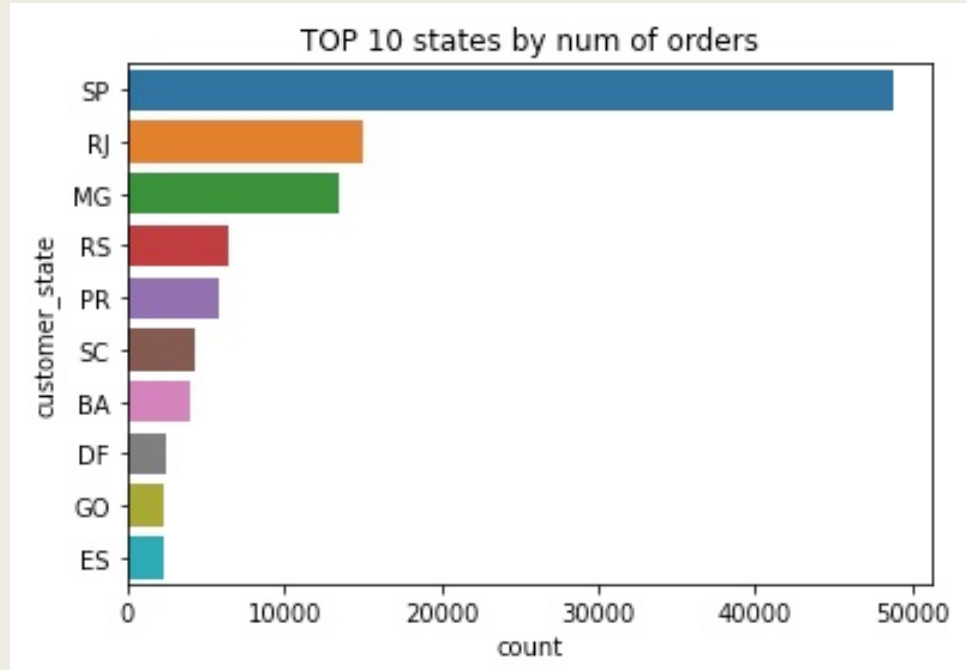
## Données

Base de données anonymisée de 2016 à 2018

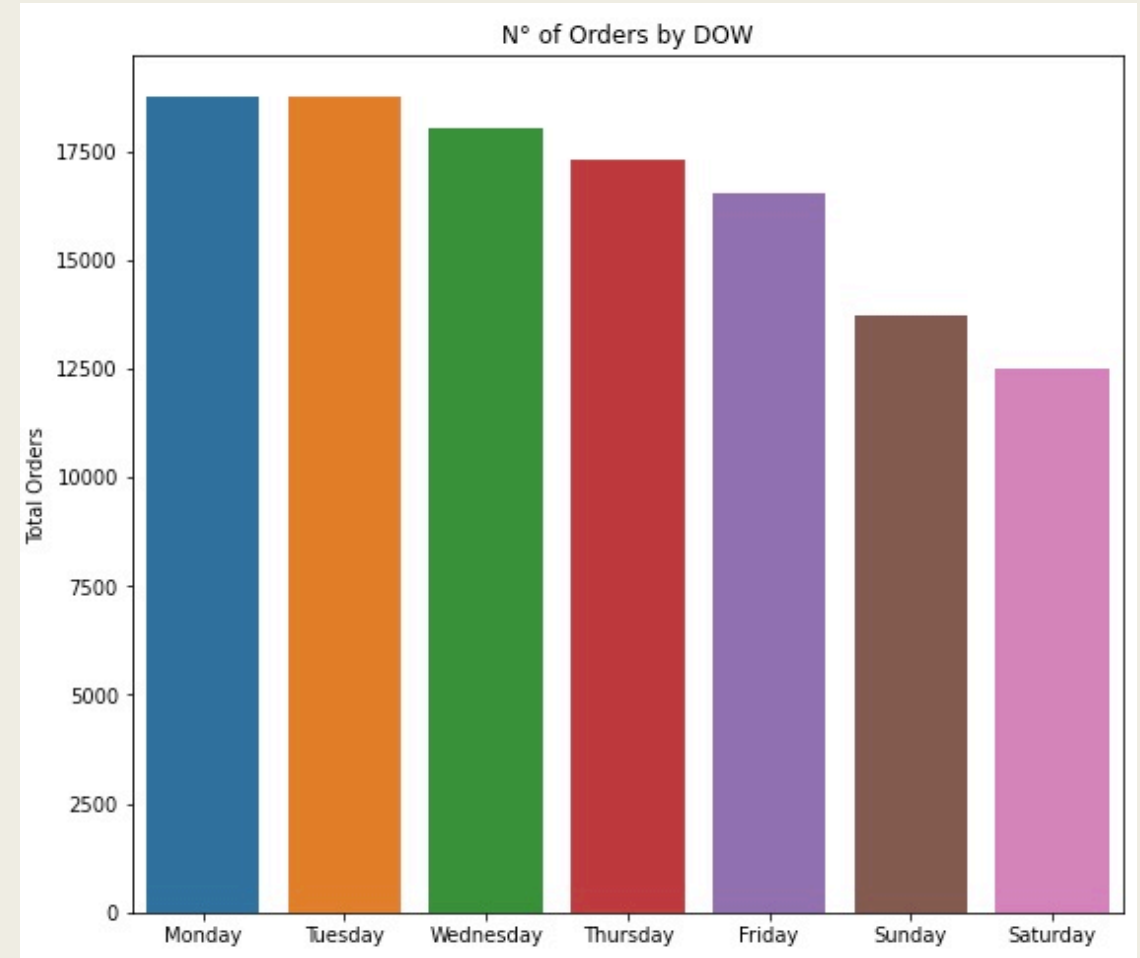
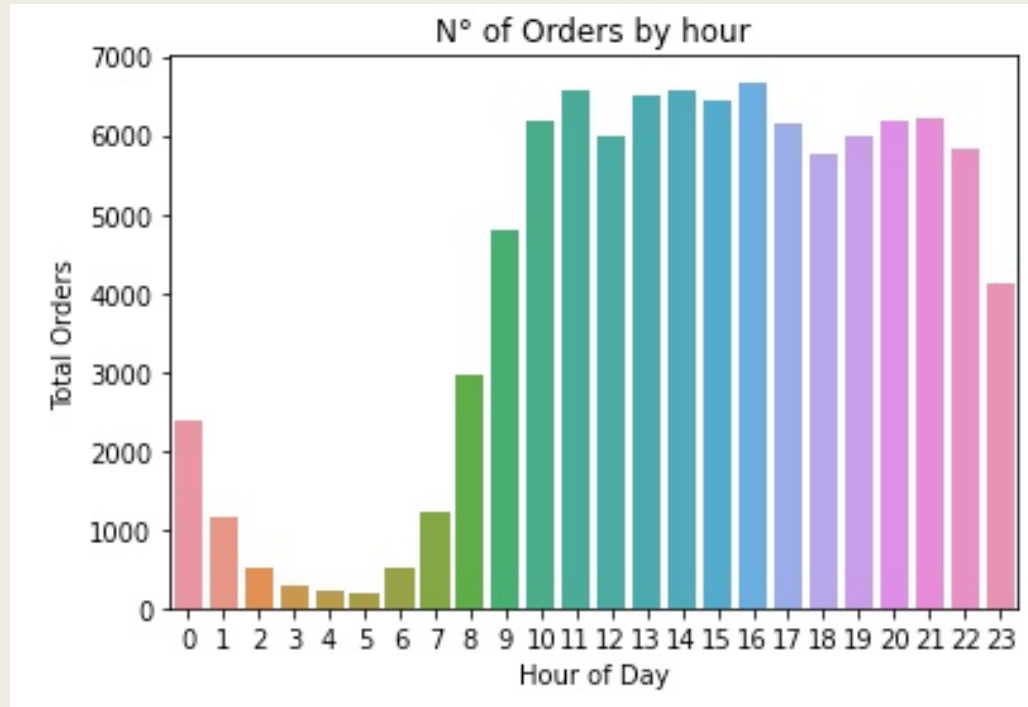
- 9 fichier
- 52 indicateurs
- 3 type d'informations
- Historique de commandes
- Localisation clients et vendeurs
- Satisfaction client
- Conditions de paiement



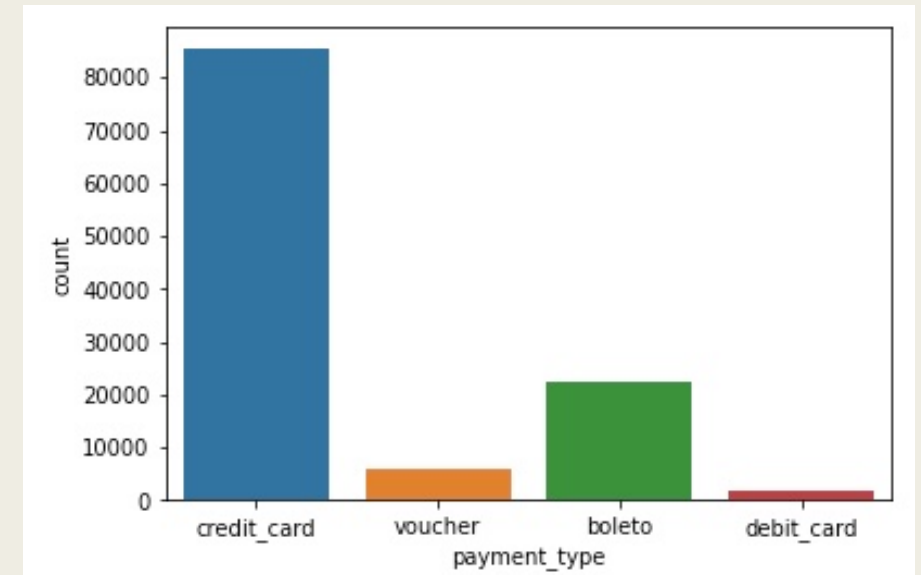
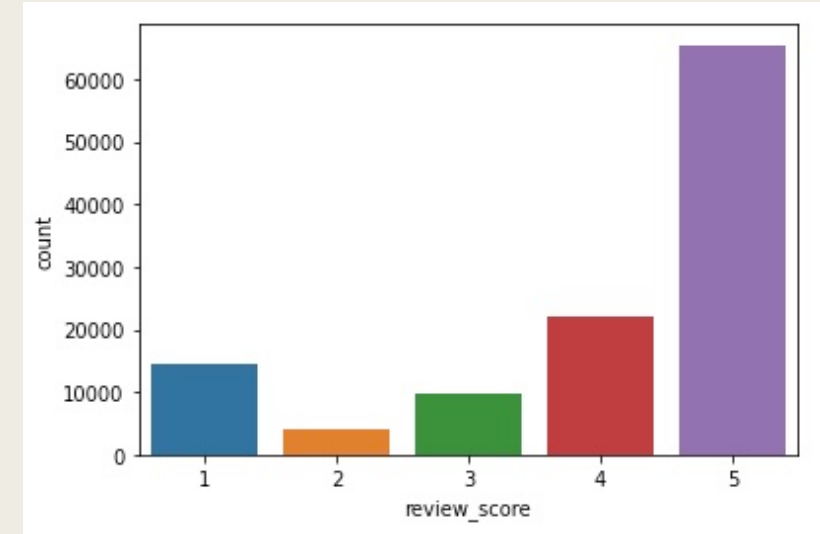
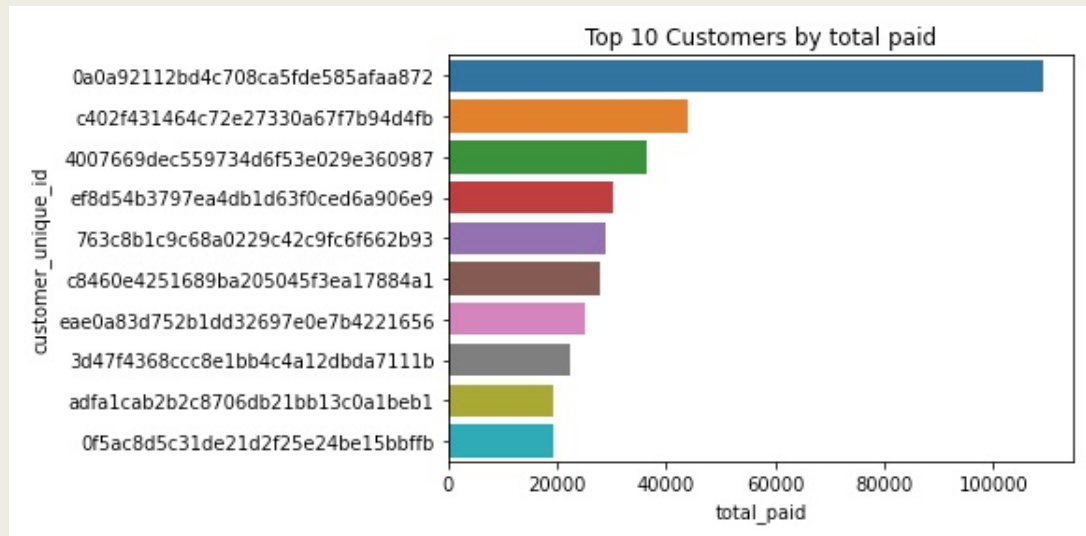
# Analyse exploratoire



# Analyse exploratoire



# Analyse exploratoire





# Modélisations:

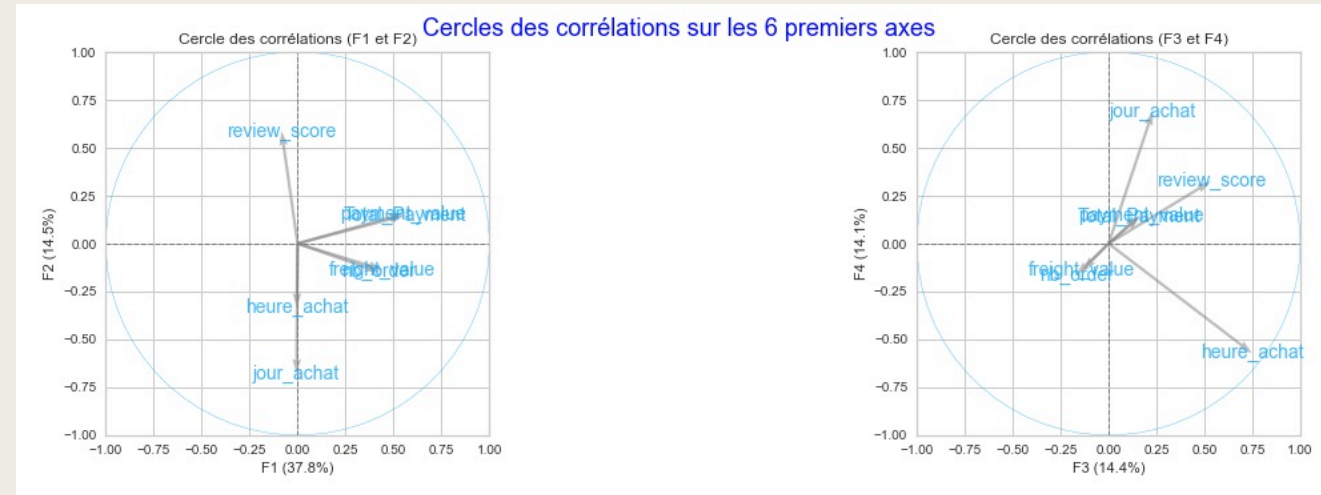
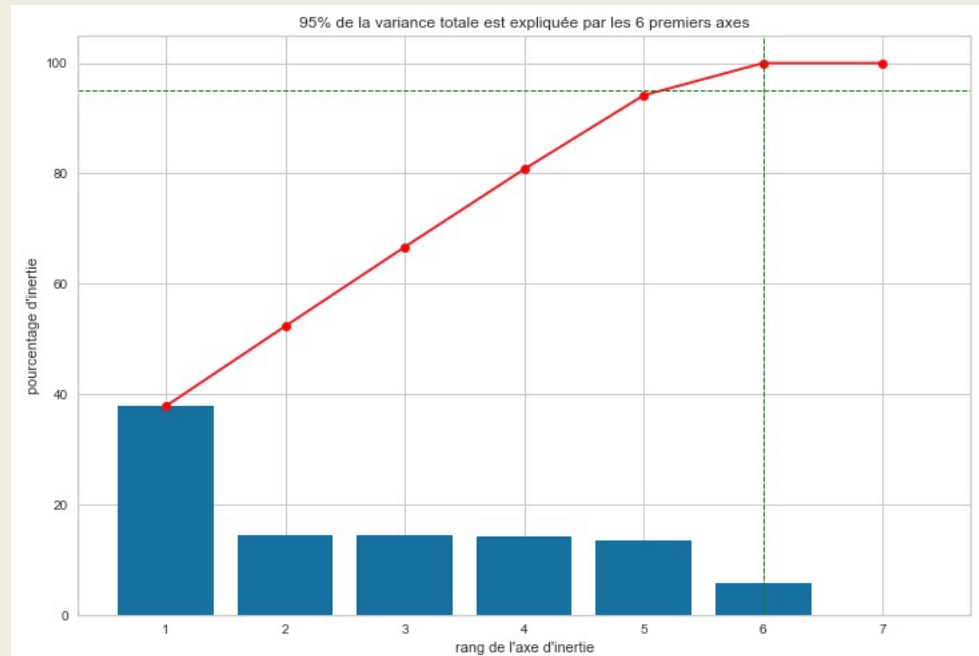
PCA KMEAN

RFM + KMEANS sur customer DF

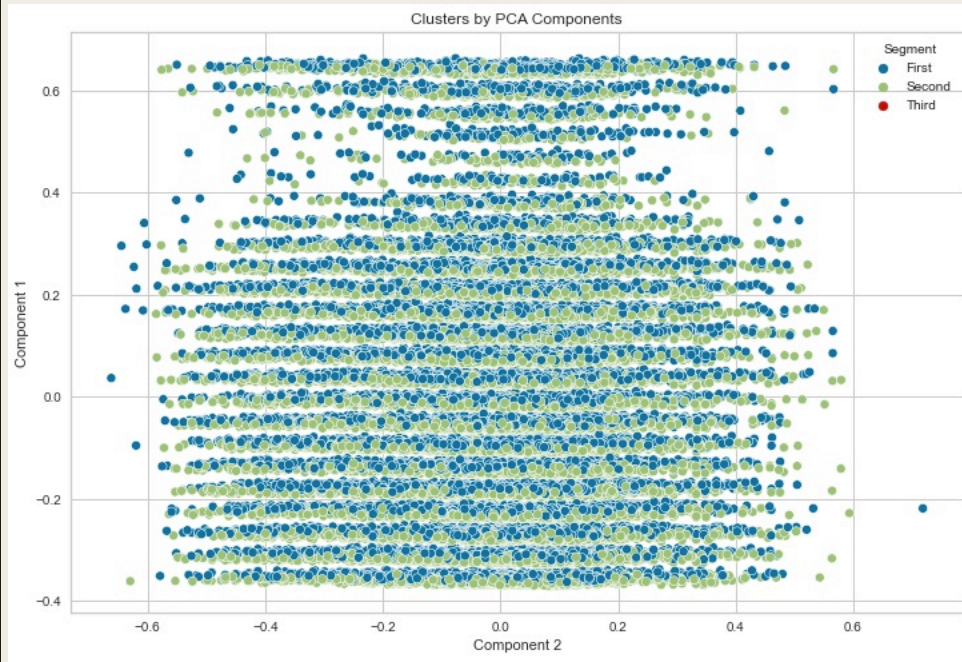
LINKAGE

Dbscan Tsne ???

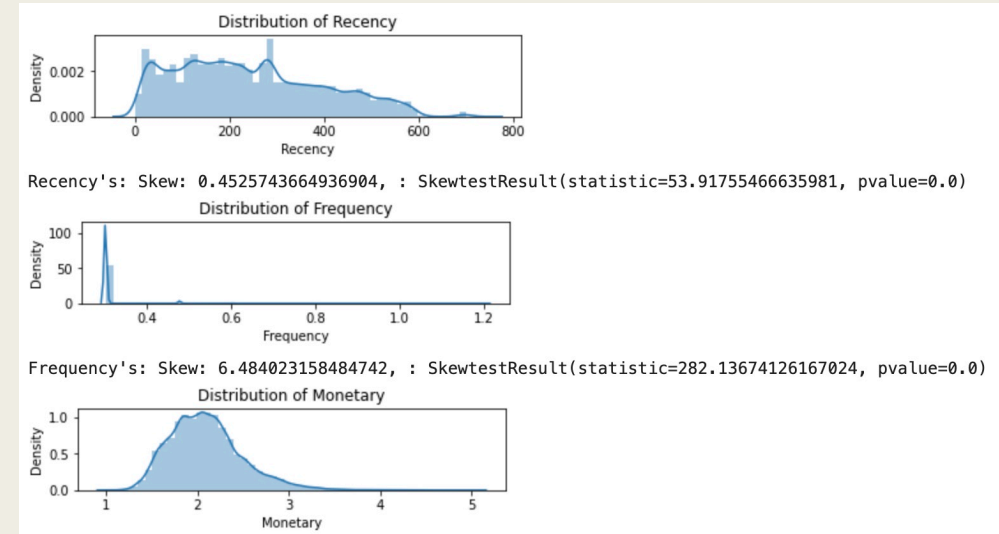
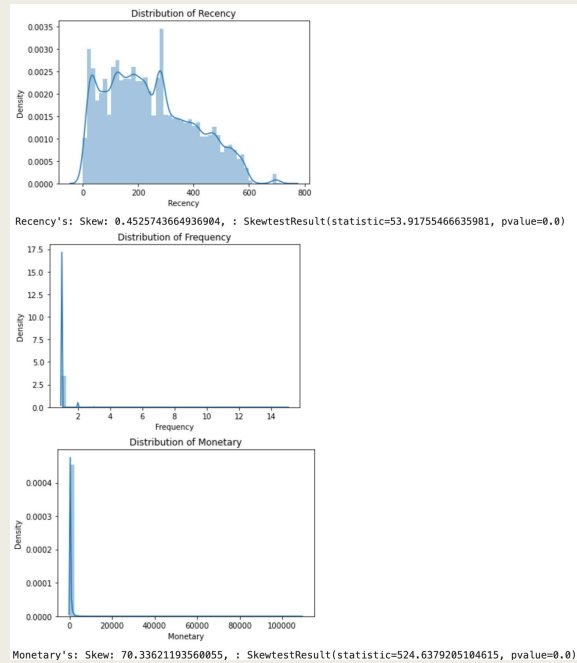
# PCA



# PCA + Kmean

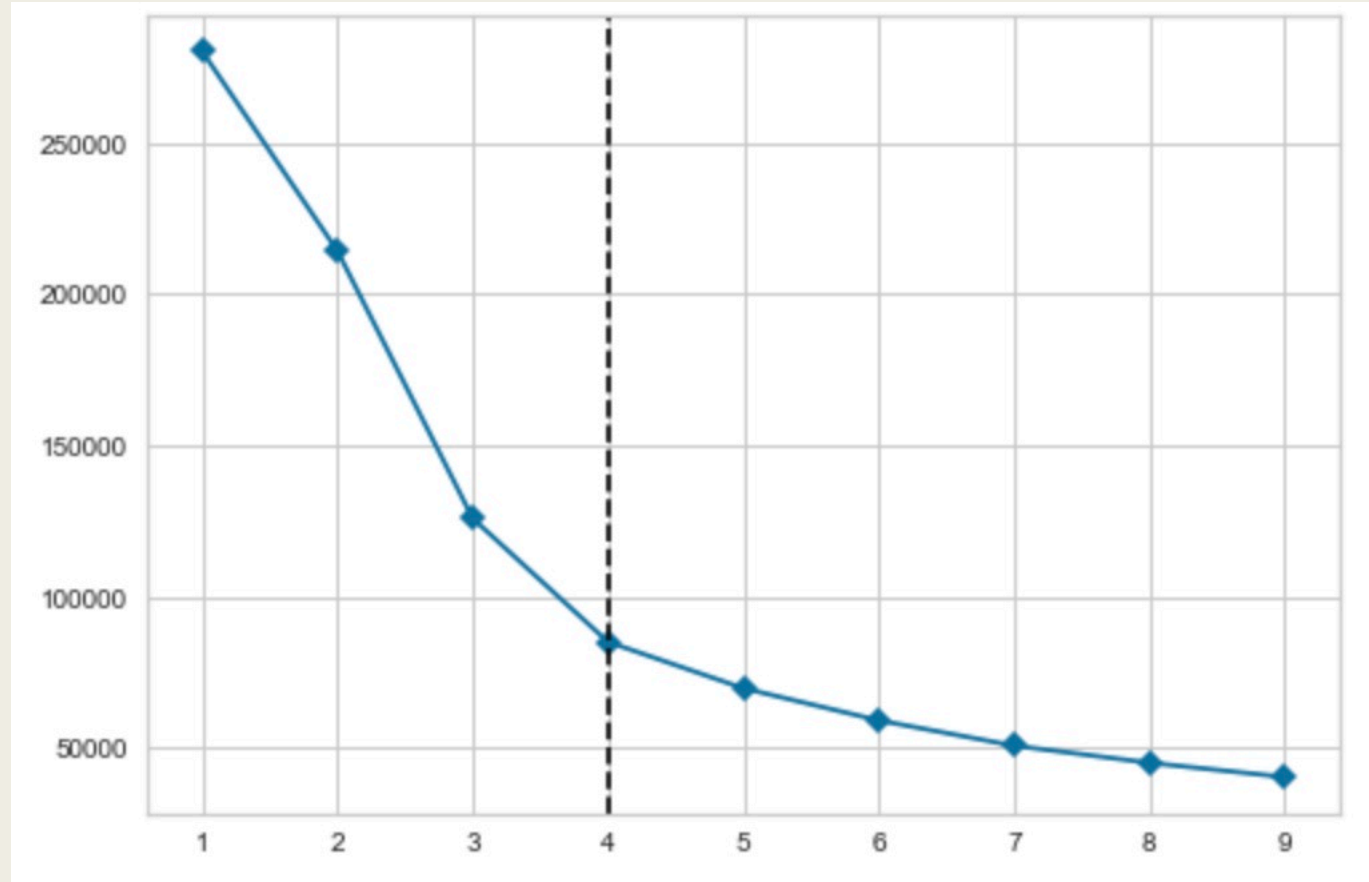


# RFM KMEAN

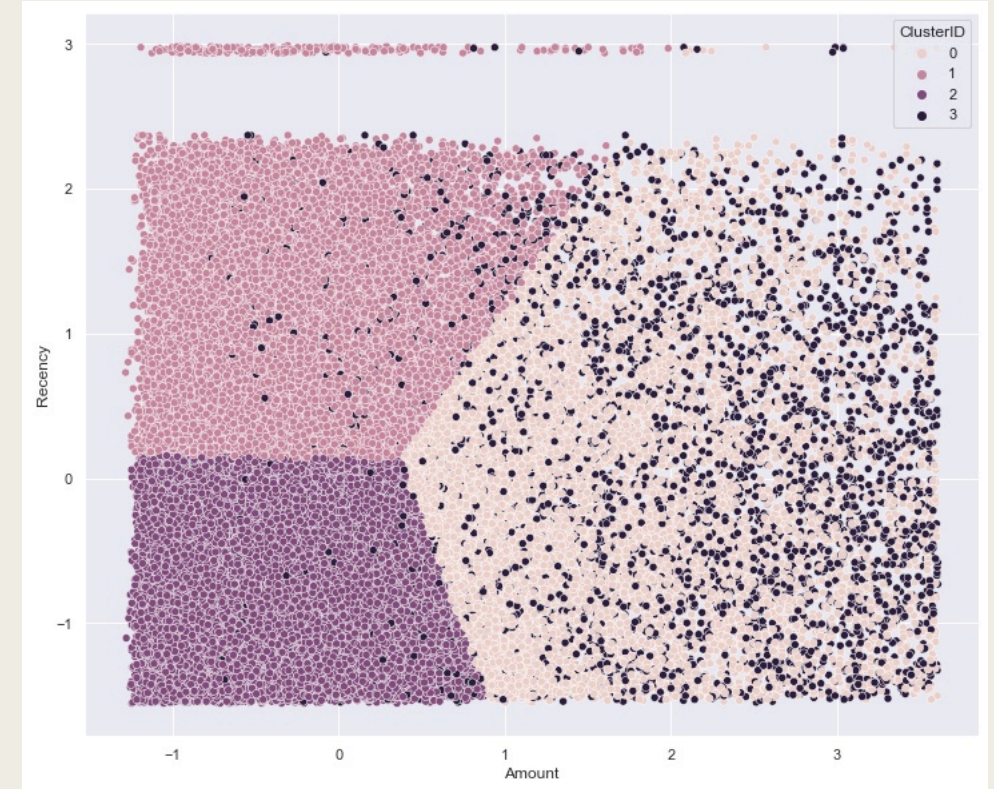
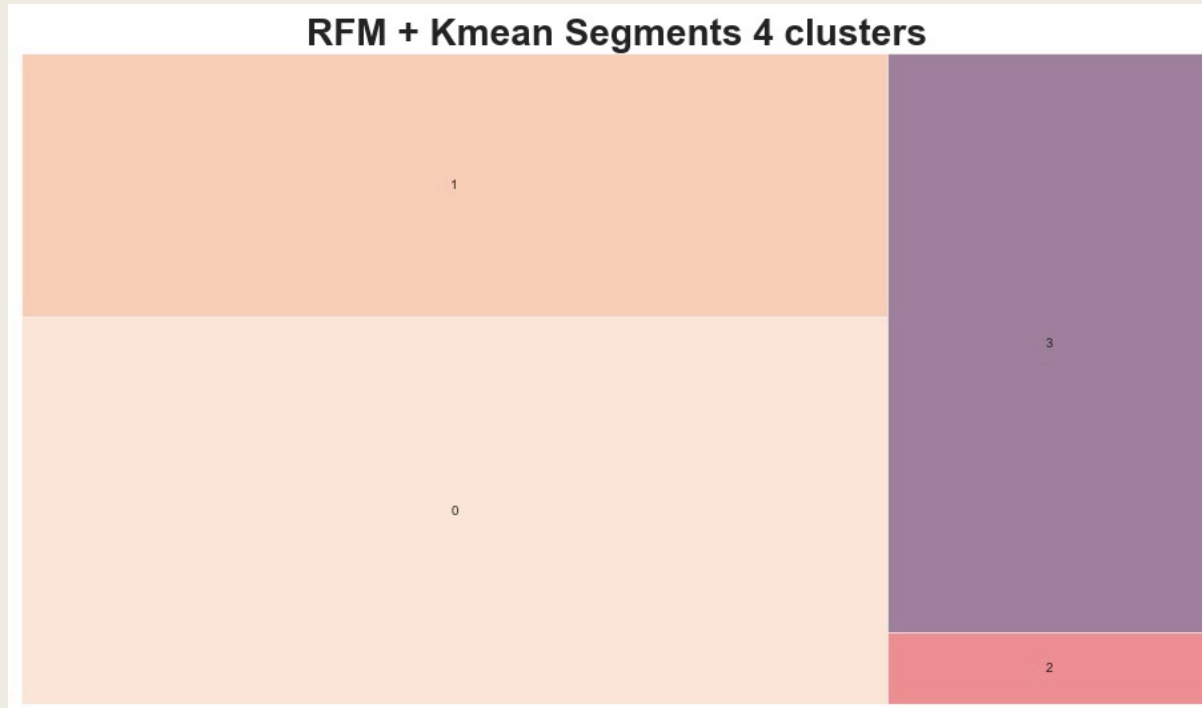


LOG + STANDARSCALER()

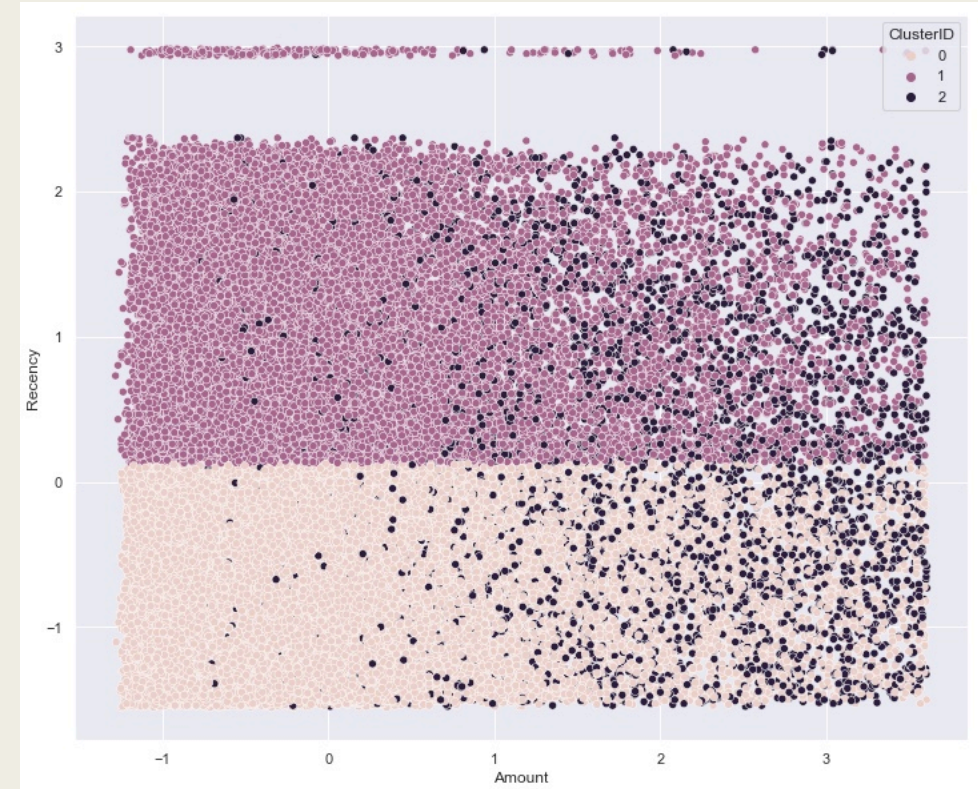
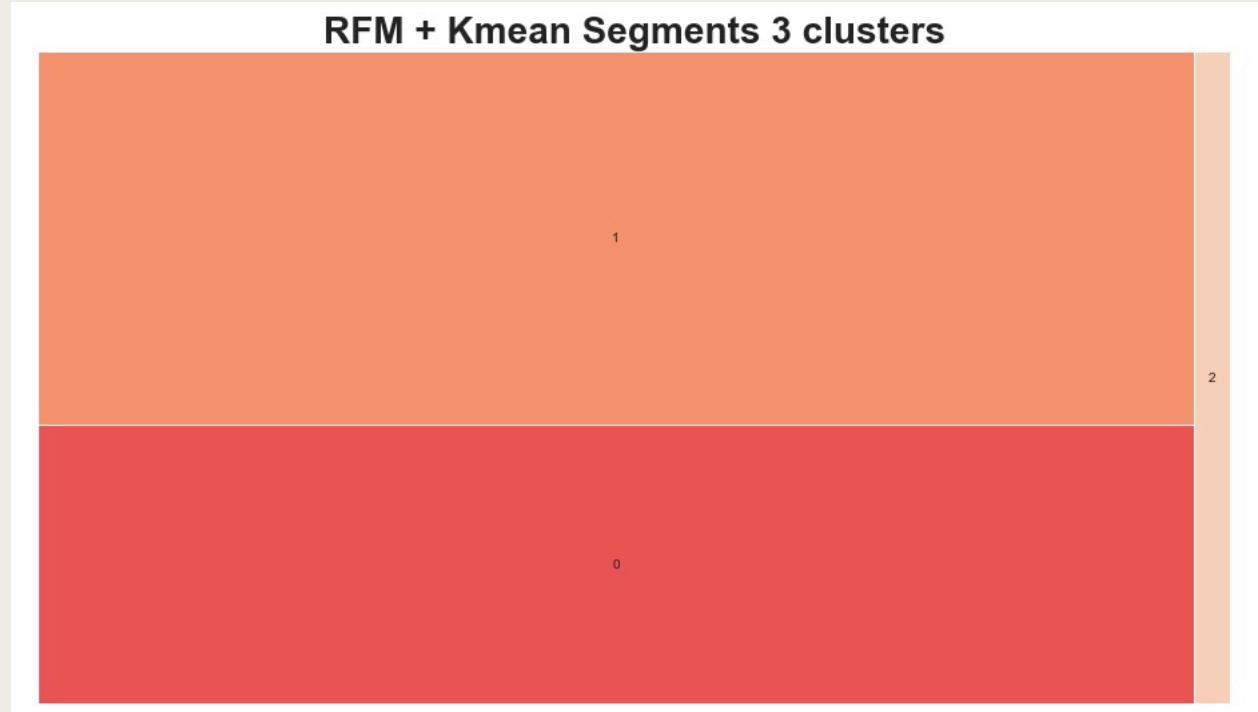
# RFM KMEAN



# RFM KMEAN 4



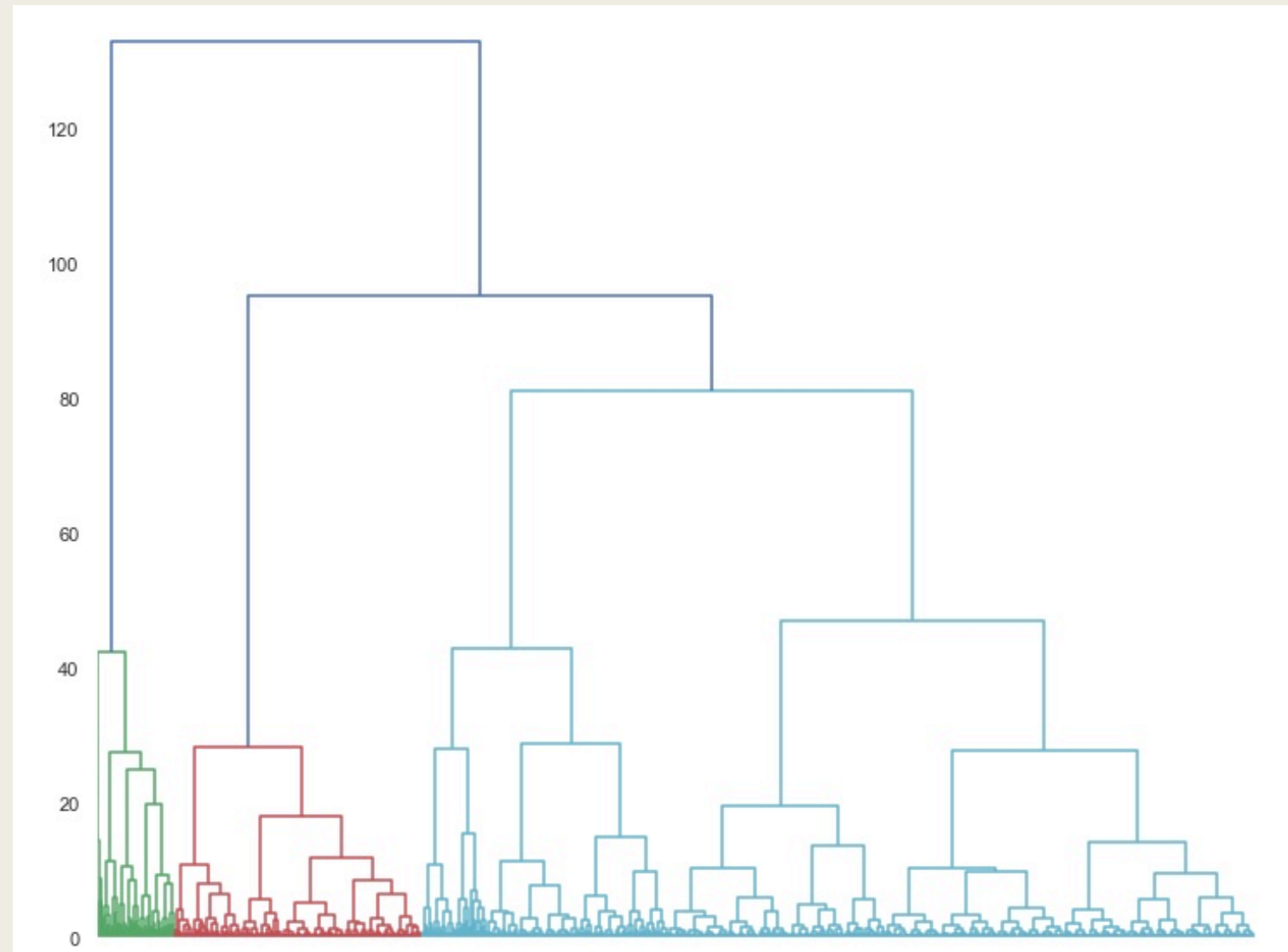
# RFM KMEAN 3

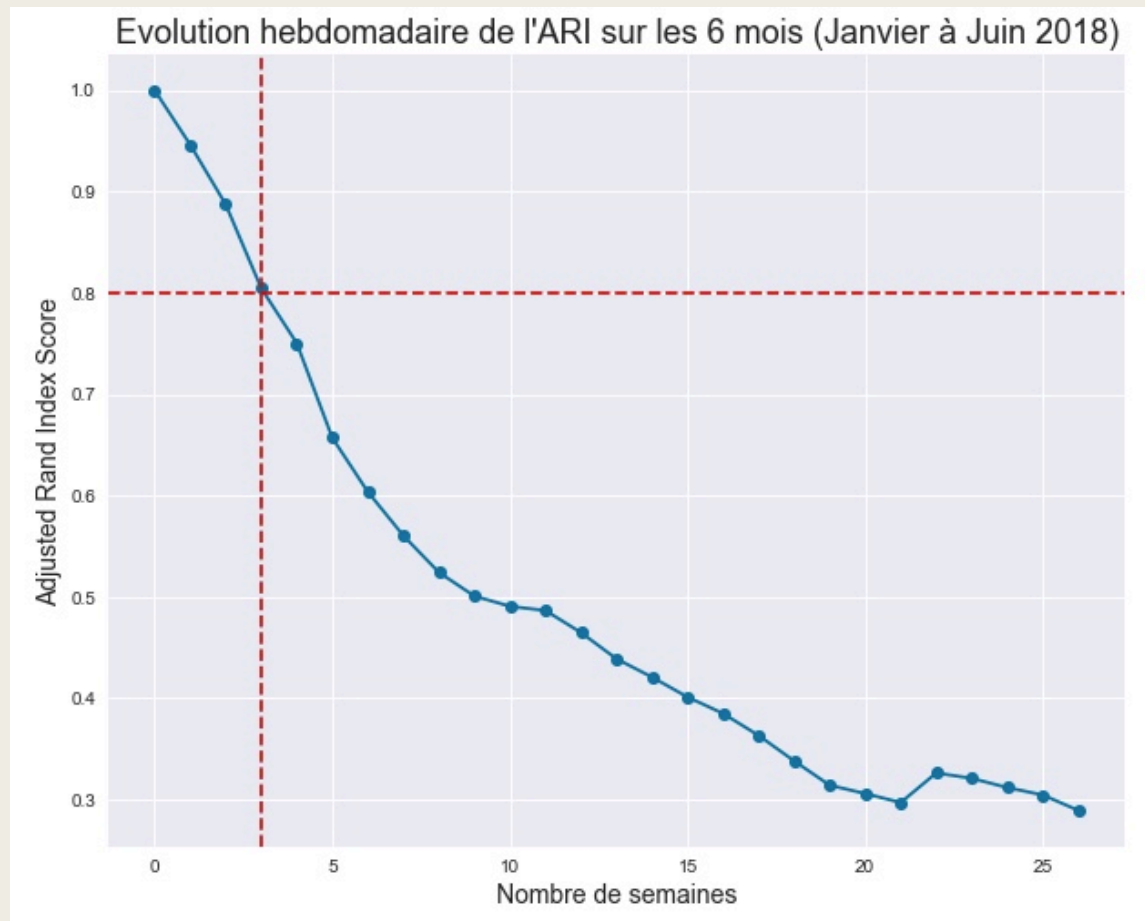


# Analyse des clusters $K=4$



# Linkage





MAINTENANCE