

Projet 4 : Segmenter les clients d'un site de e-commerce

Antoine Maby - /09/2021

Problématique

Aider les équipes d'Olist à comprendre les différents types d'utilisateurs

Regrouper ensemble des clients de profils similaires

La segmentation proposée doit être exploitable et facile d'utilisation pour l'équipe marketing.

Problématique

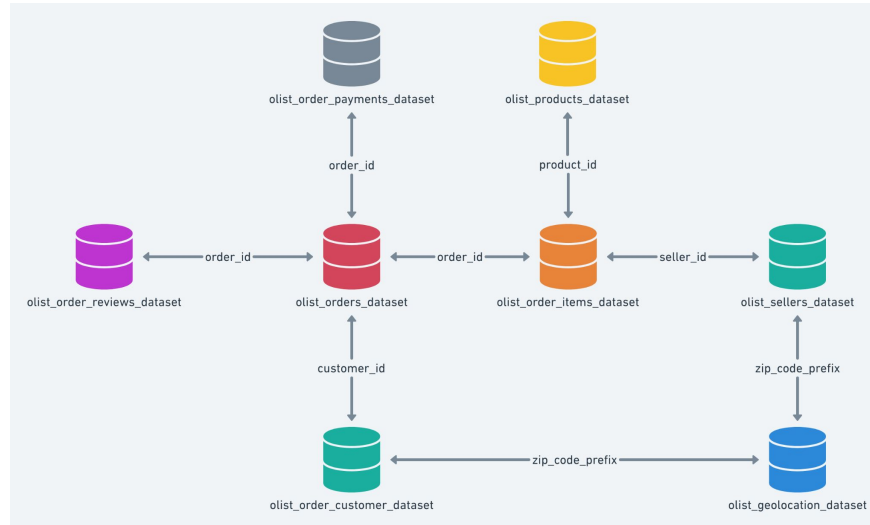
Pour des raisons de confidentialité, Olist ne fournit pas beaucoup de données

Données anonymisée comportant des informations sur l'historique de commandes, les produits achetés, les commentaires de satisfaction, et la localisation des clients depuis janvier 2017

Fréquence à laquelle la segmentation doit être mise à jour, afin de pouvoir effectuer un devis de contrat de maintenance

Nettoyage et Feature Engineering

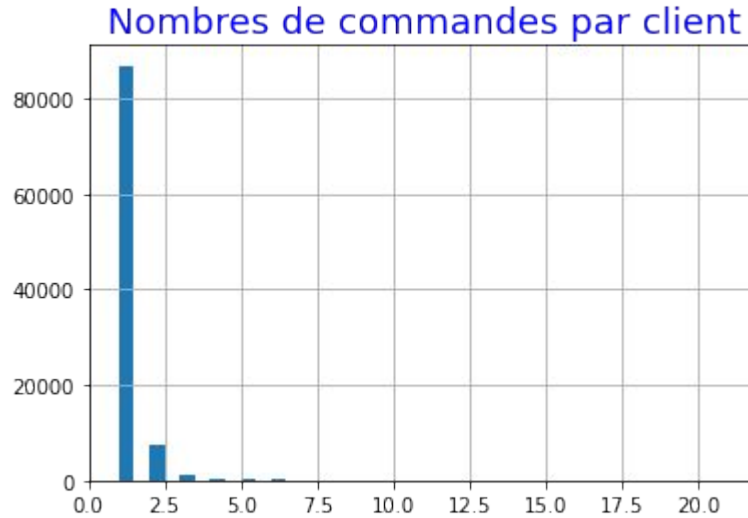
- Fusion des différents dataset



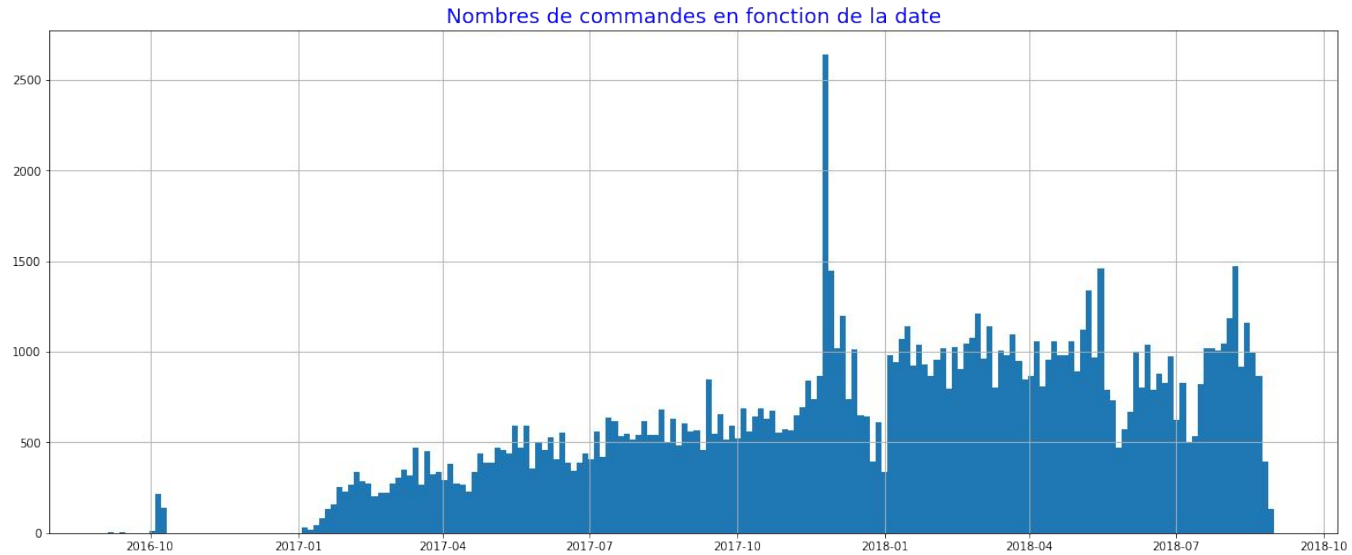
Nettoyage et Feature Engineering

- Fusion des différents dataset
- Création de nouvelles variables : (Temps de Livraison, Position, Écart entre deux commandes, ...)
- Regroupement des catégories de produits

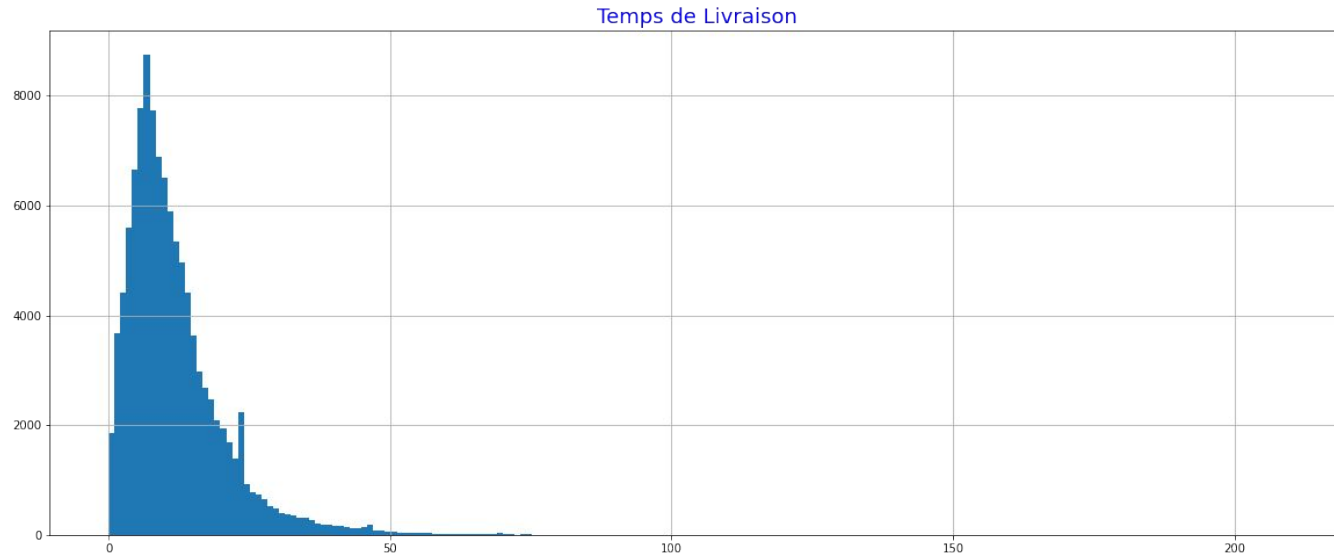
Exploration



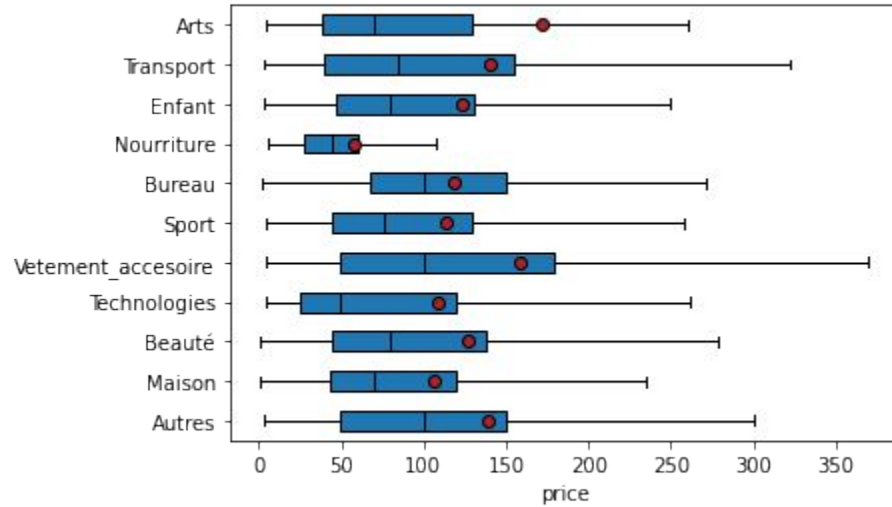
Exploration



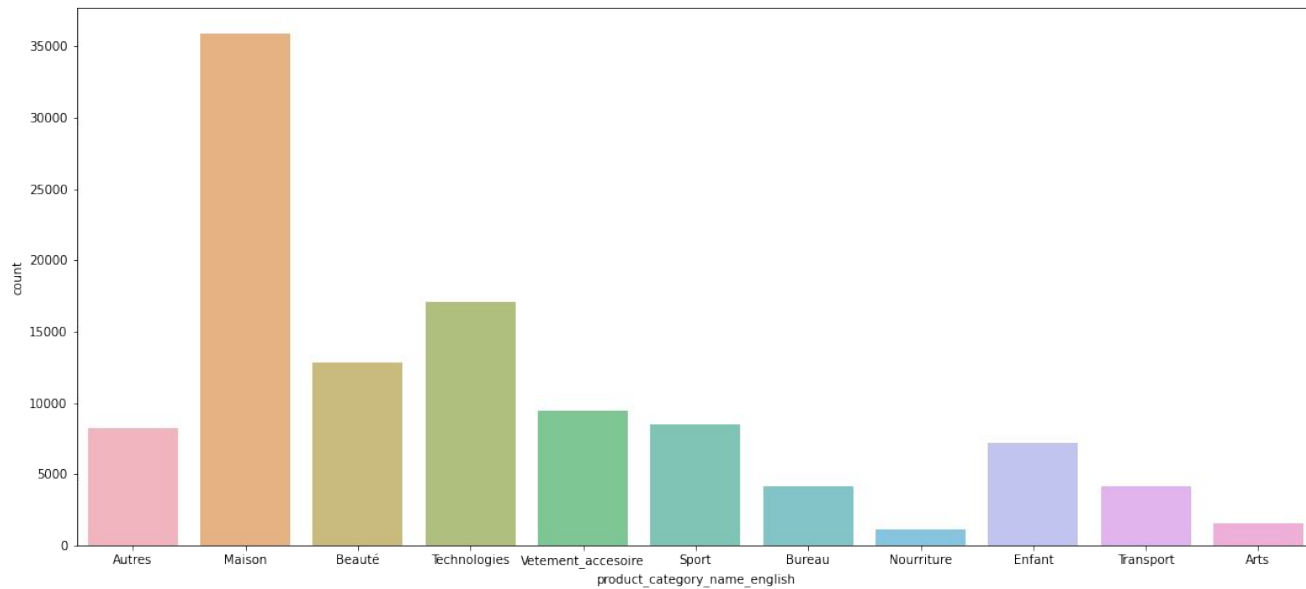
Exploration



Exploration

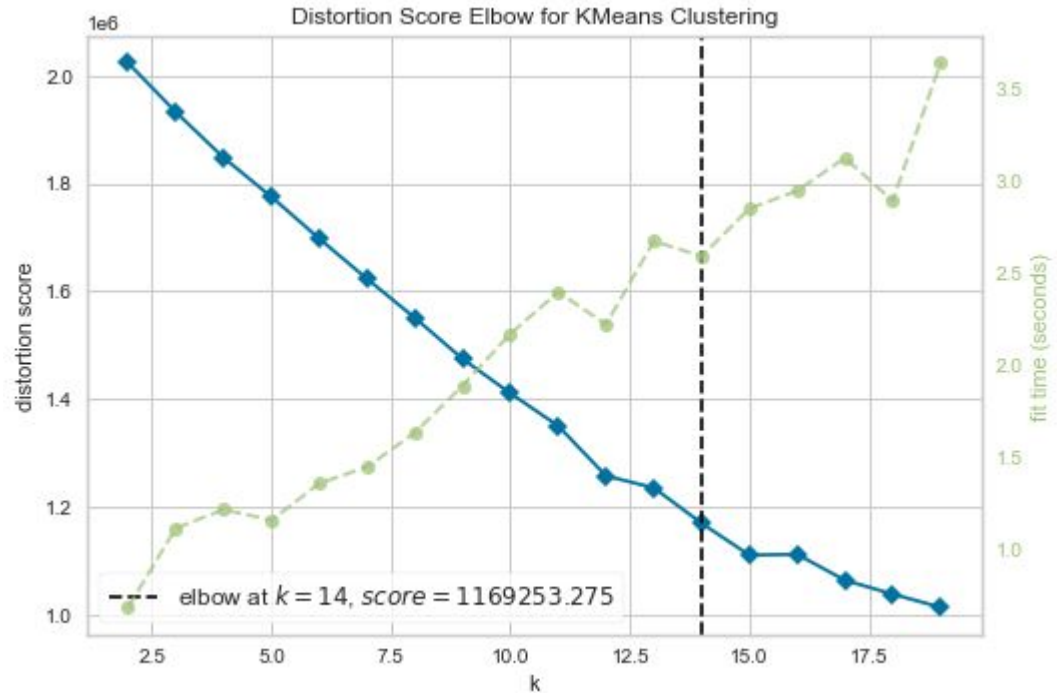


Exploration

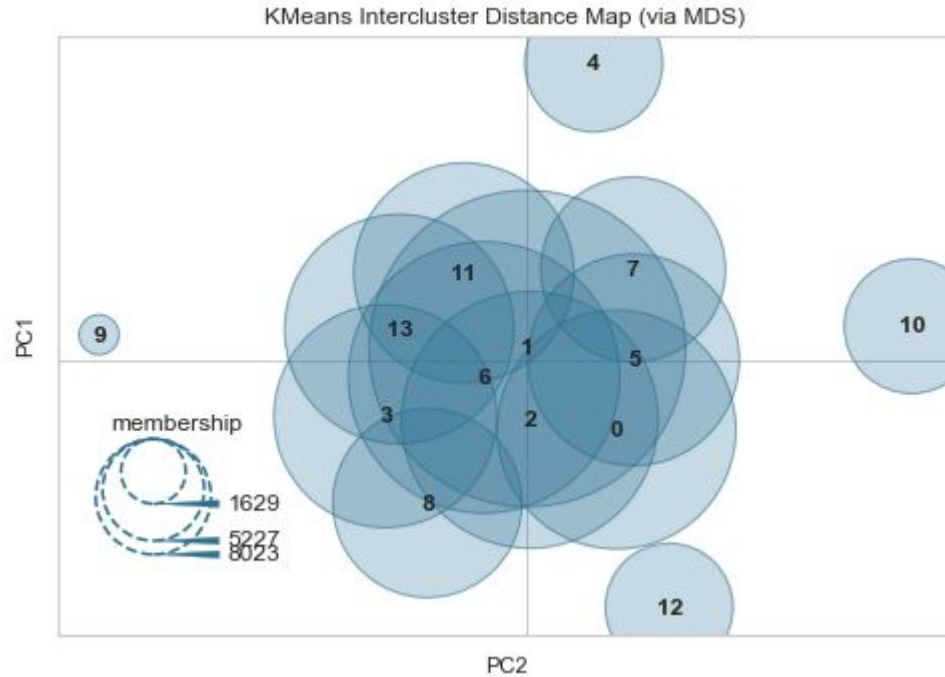


Clustering

- KMeans
- StandardScaler
- Avec les catégories



Clustering

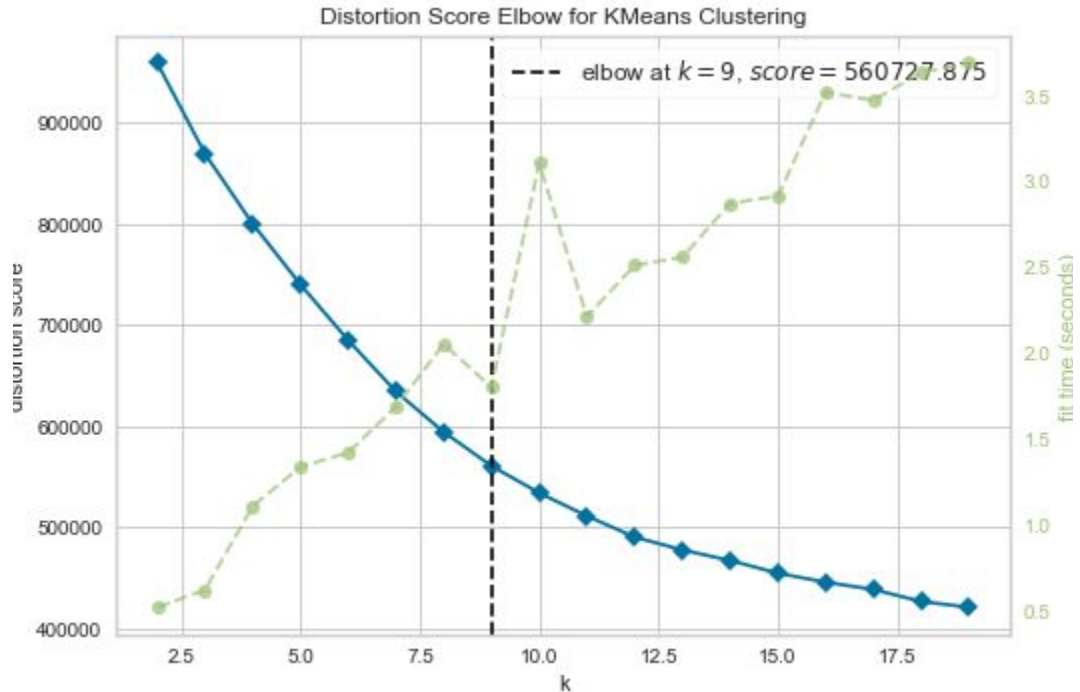


Clustering

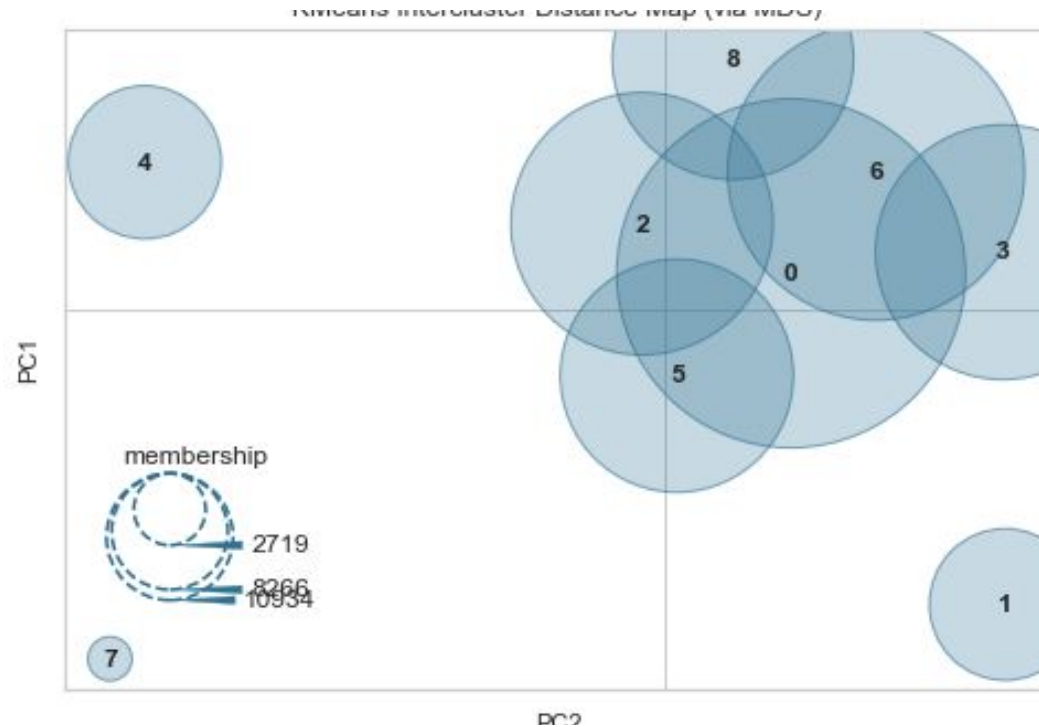
- Les clusters sont trop rapprochés les uns des autres
- Le poids des variables catégories de produits est trop important
- On va supprimer ces variables pour le clustering

Clustering

- KMeans
- Standard Scaler
- Sans les catégories



Clustering

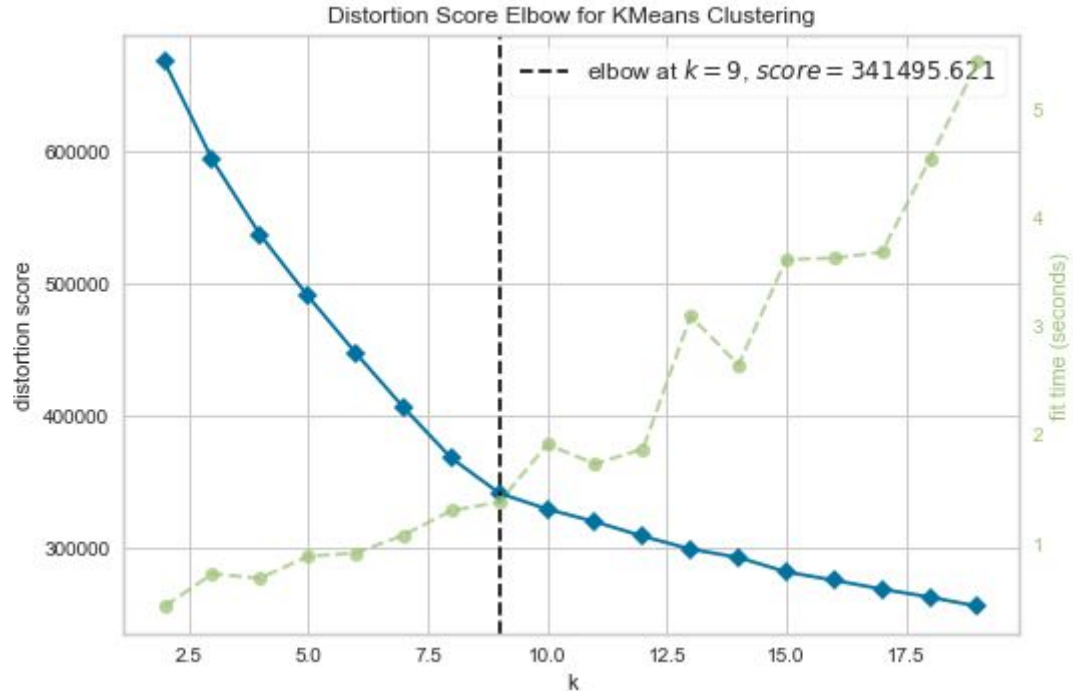


Clustering

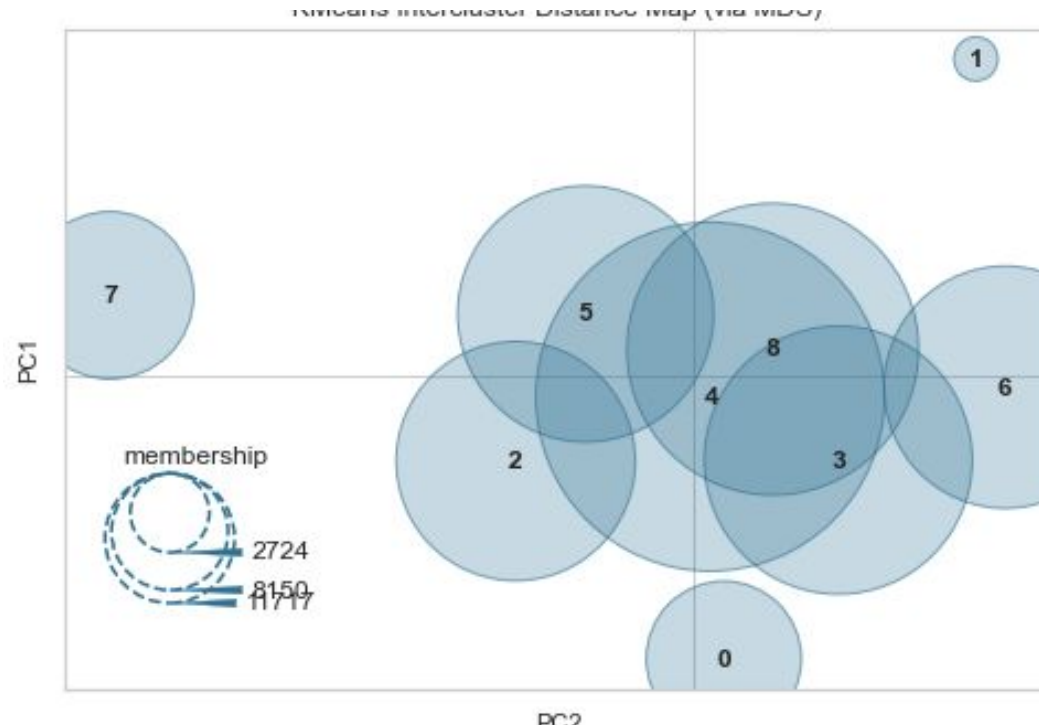
- Les clusters sont toujours rapprochés les uns des autres
- Mais nous avons une meilleure séparation des catégories de clients
- On va essayer d'essayer de faire une PCA pour séparer nos clusters

Clustering

- KMeans
- Standard Scaler
- Sans les catégories
- Avec PCA



Clustering

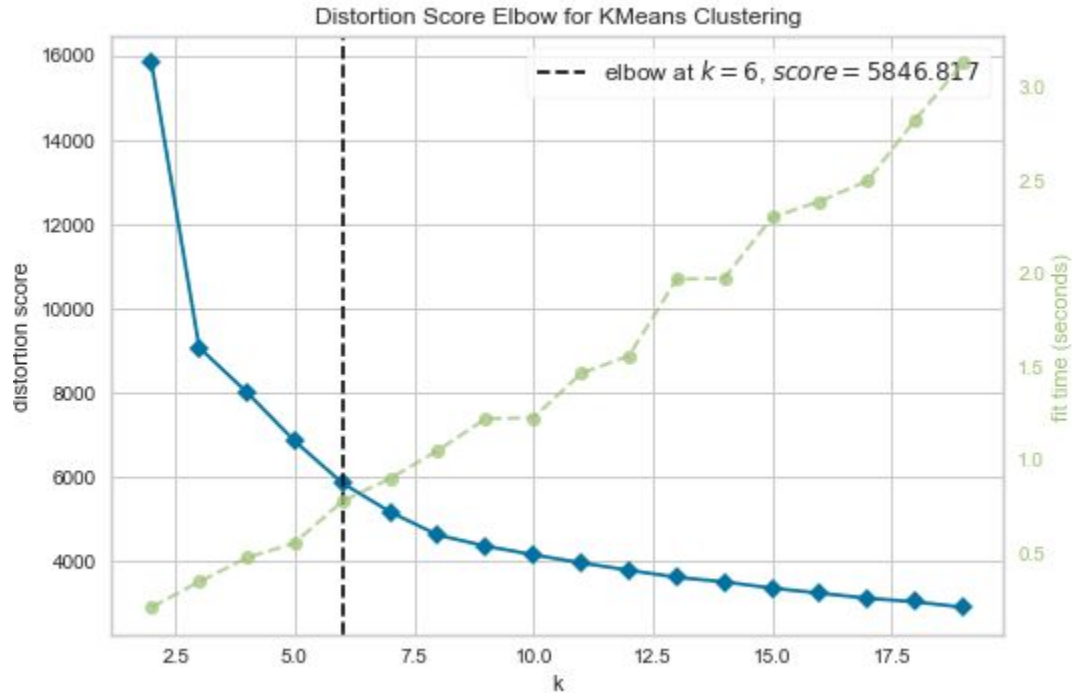


Clustering

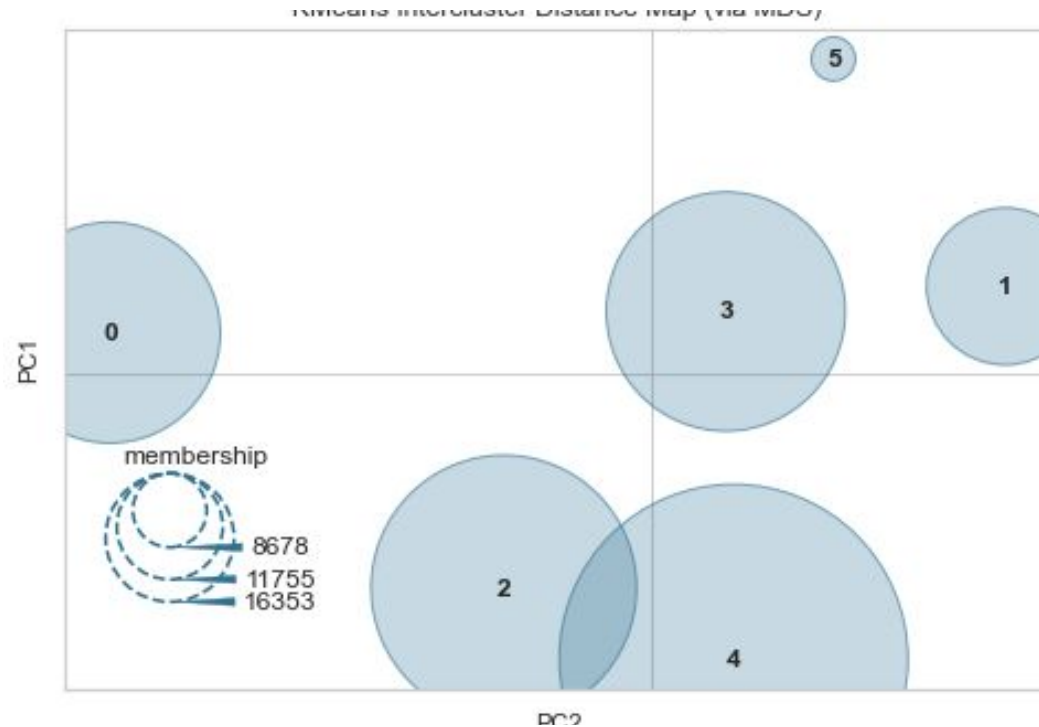
- Les clusters sont toujours rapprochés les uns des autres
- Les séparations entre les clients ont été conservé par la PCA
- Le gain est surtout en de temps de calcul

Clustering

- KMeans
- MinMax Scaler
- Sans les catégories



Clustering



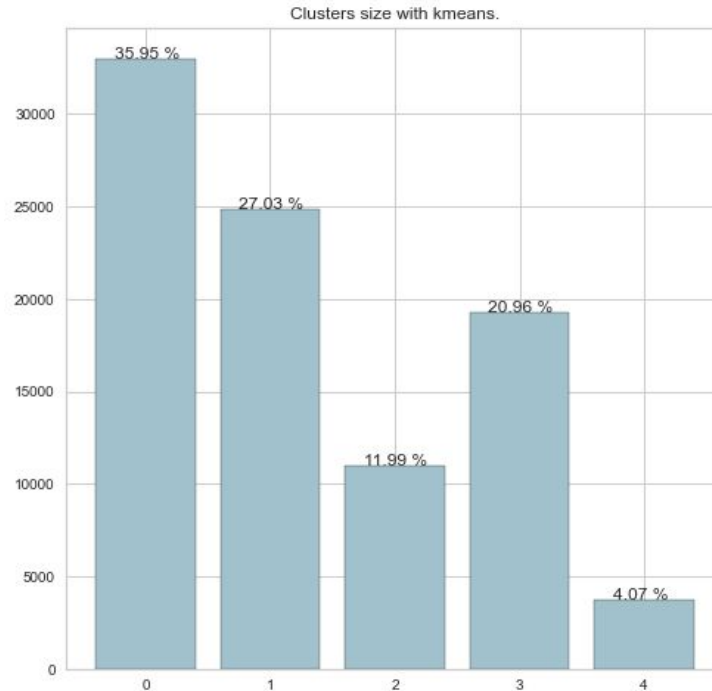
Clustering

- Les clusters sont moins rapprochés les uns des autres
- Les séparations entre les clients ont été conservé
- Il est plus difficile de différencier les clusters de clients, certaines variables sont trop importantes

Clustering

- **Modèle choisi : KMeans avec 5 Cluster sans les catégories et StandardScaler**
- On remettra les catégories dans les données après le clustering.
- Passons au Personae des différents clients

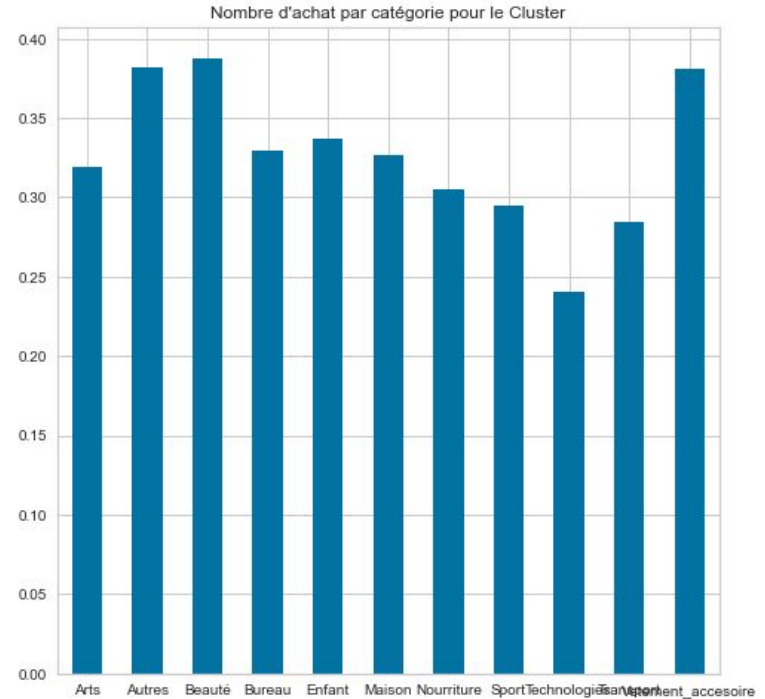
Personae des clients



Personae des clients



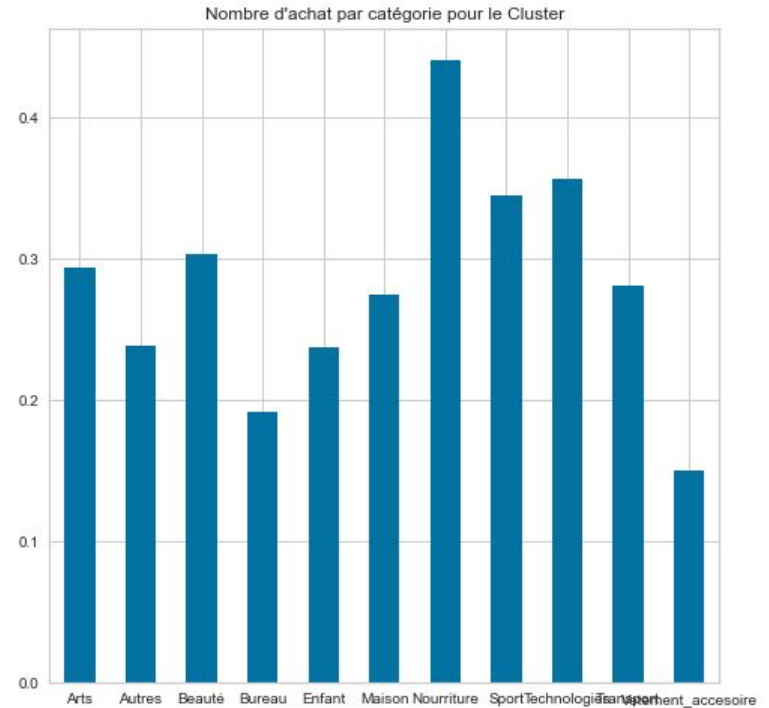
- Prix moyen
- Bonne notation
- Temps de livraison court
- Un seul moyen de paiement



Personae des clients



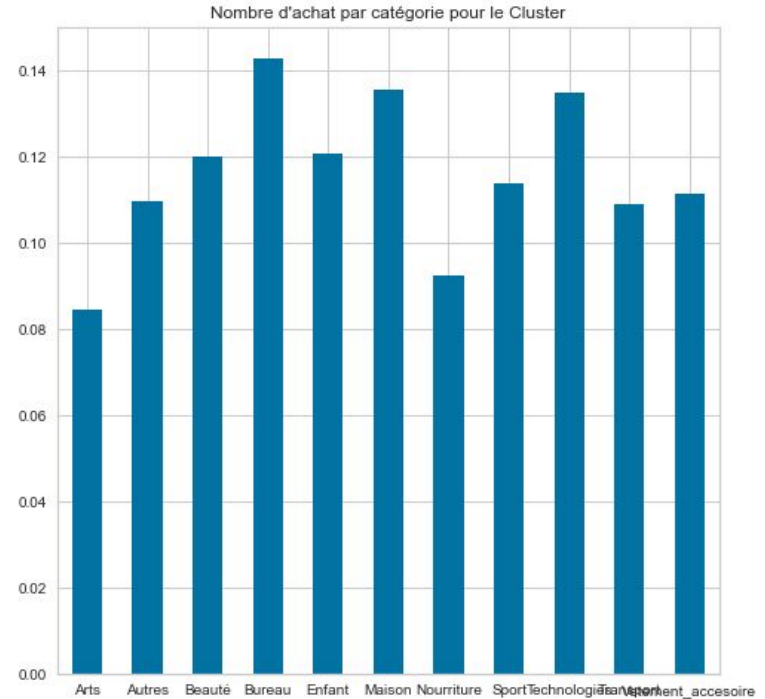
- Prix moyen
- Bonne notation
- Temps de livraison court
- Description des produits longues
- Un seul moyen de paiement



Personae des clients



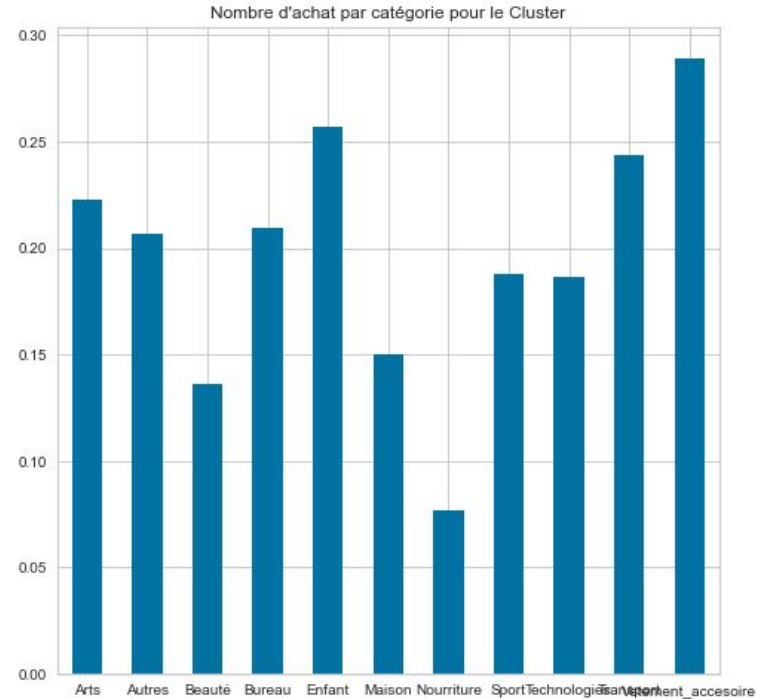
- Prix moyen
- Mauvaise Notation
- Temps de livraison long
- Un seul moyen de paiement



Personae des clients



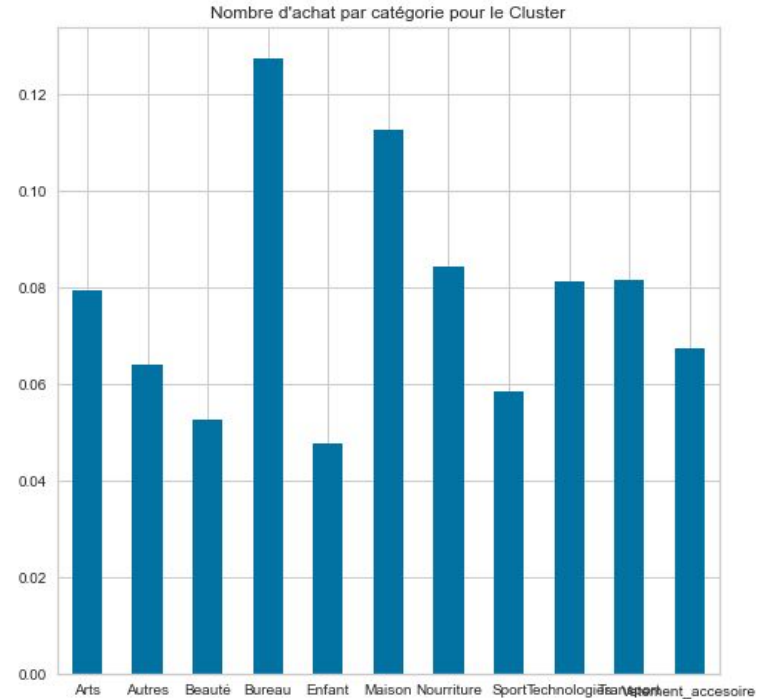
- Prix moyen
- Livraison moyenne
- Plus de photos sur les produits
- Un seul moyen de paiement



Personae des clients



- Prix cher
- Livraison moyenne
- Plus de moyens de paiements
- Description des produits plus longues
- Notes un peu moins bonnes



Personae des clients



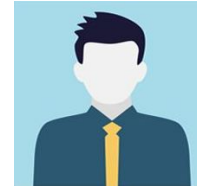
- Prix moyen
- Bonne notation
- Livraison courte
- Un type de paiement



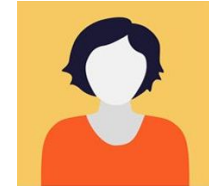
- Prix moyen
- Bonne notation
- Livraison courte
- Description longues
- Un type de paiement



- Prix moyen
- Mauvaise Notation
- Livraison longue
- Un type de paiement



- Prix moyen
- Livraison moyenne
- Plus de photos
- Un type de paiement

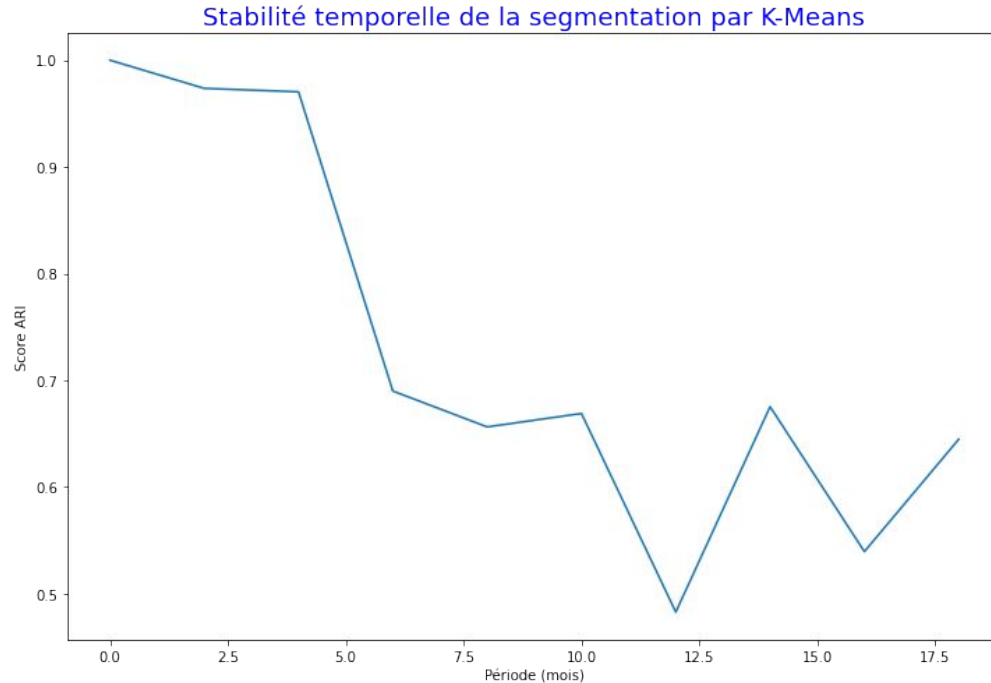


- Prix cher
- Livraison moyenne
- Plus type de paiements
- Description longues
- Notes moins bonnes

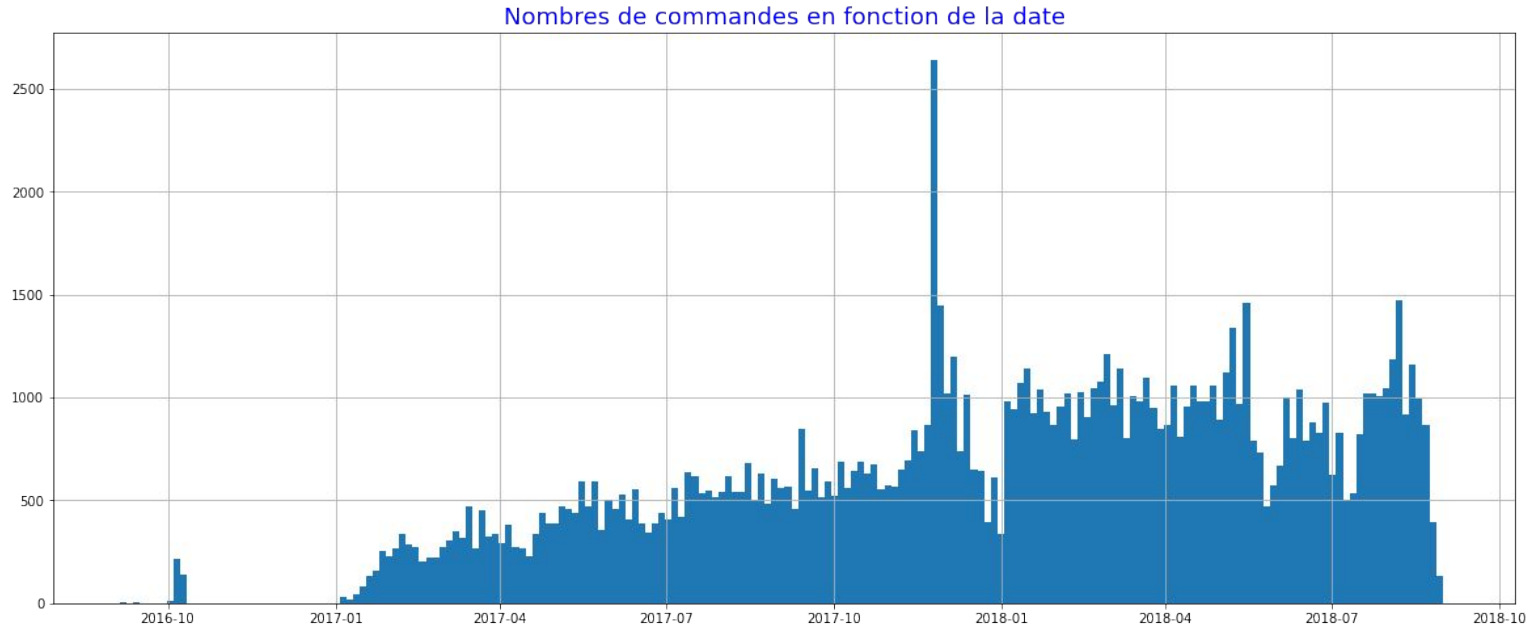
Stabilité Temporelle du KMeans

- Est-ce que les clients changent de Cluster et au bout de combien de temps pour re segmenter
- Période Initiale de 6 mois pour sélectionner des clients
- Calcul du score ARI tous les deux mois

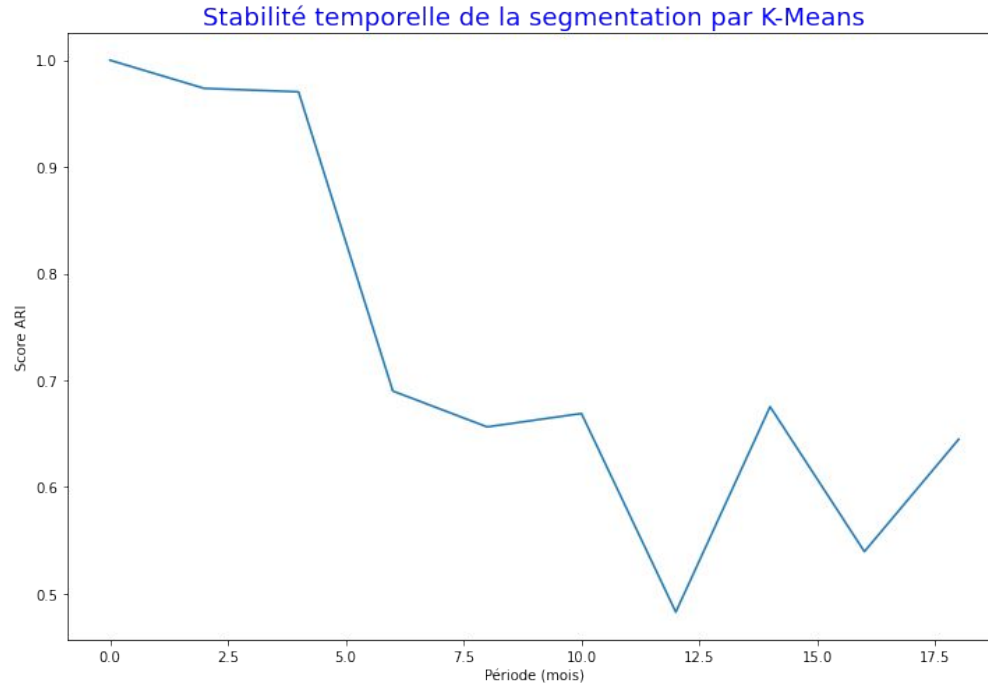
Stabilité Temporelle du KMeans



Stabilité Temporelle du KMeans



Stabilité Temporelle du KMeans





Merci pour votre
attention