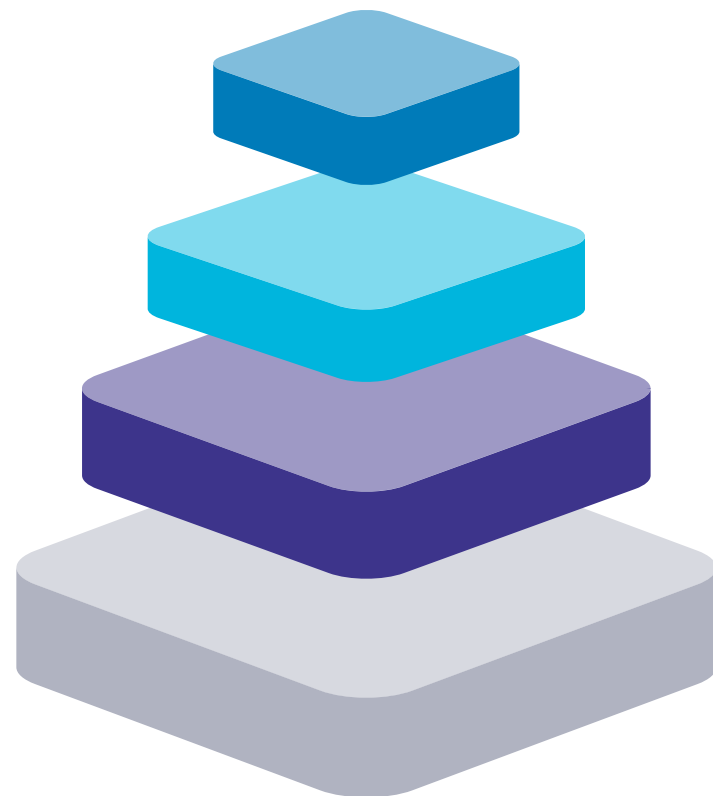


Segmentez des clients d'un site e-commerce

Formation Data scientist : Projet n°5

Bayram DONAT



Sommaire

- Problématique
- Cleaning
- Feature engineering
- Exploration
- Différentes modélisations effectuées
- Délai de maintenance du modèle



Problématique

- Entreprise : Olist
- Activité : solution de vente sur les marketplaces en ligne au Brésil
- Besoins :
 - segmentation des clients pour des campagnes de communication adaptées.
 - compréhension des différents types des clients en fonction
 - Des comportements
 - Des données personnelles.
 - liste d'actions ciblées par segment
 - contrat de maintenance basé sur une analyse de la stabilité des segments au cours du temps.

Cleaning



- 9 fichiers CSV

- olist_customers_dataset : données des clients : numéro d'identification, **n° commande client**, ville, état, **code postal**
- olist_geolocation_dataset : données de géolocalisation : **code postal**, ville, état, latitude, longitude
- olist_order_items_dataset : contenus de commande : **n° commande, produits**, vendeur, prix, fret, ...
- olist_order_payments_dataset : données sur les paiements : **n° commande**, type, fois, montant,...
- olist_order_reviews_dataset : données sur l'avis : **n° commande**, avis, commentaires...
- olist_orders_dataset : données sur la livraison : **n° commande**, date commande, date transporteur, date de livraison estimée, date de livraison, état de commande, **n° commande client**,
- olist_products_dataset : données sur les produits en vente : **produit, catégorie**, poids, dimensions...
- olist_sellers_dataset : vendeur, ville, état, **code postal**
- product_category_name_translation : **catégorie**, catégorie en anglais

Cleaning



- Traitements effectués
 - Correction des types de données par colonne
 - Traitement des valeurs manquantes ou erronées
- Fusion en un dataframe
- Aggrégation
 - customer_unique_id,
 - category_name_english (la plus achetée par commande)

Featuring engineering

olist

- Filtrage

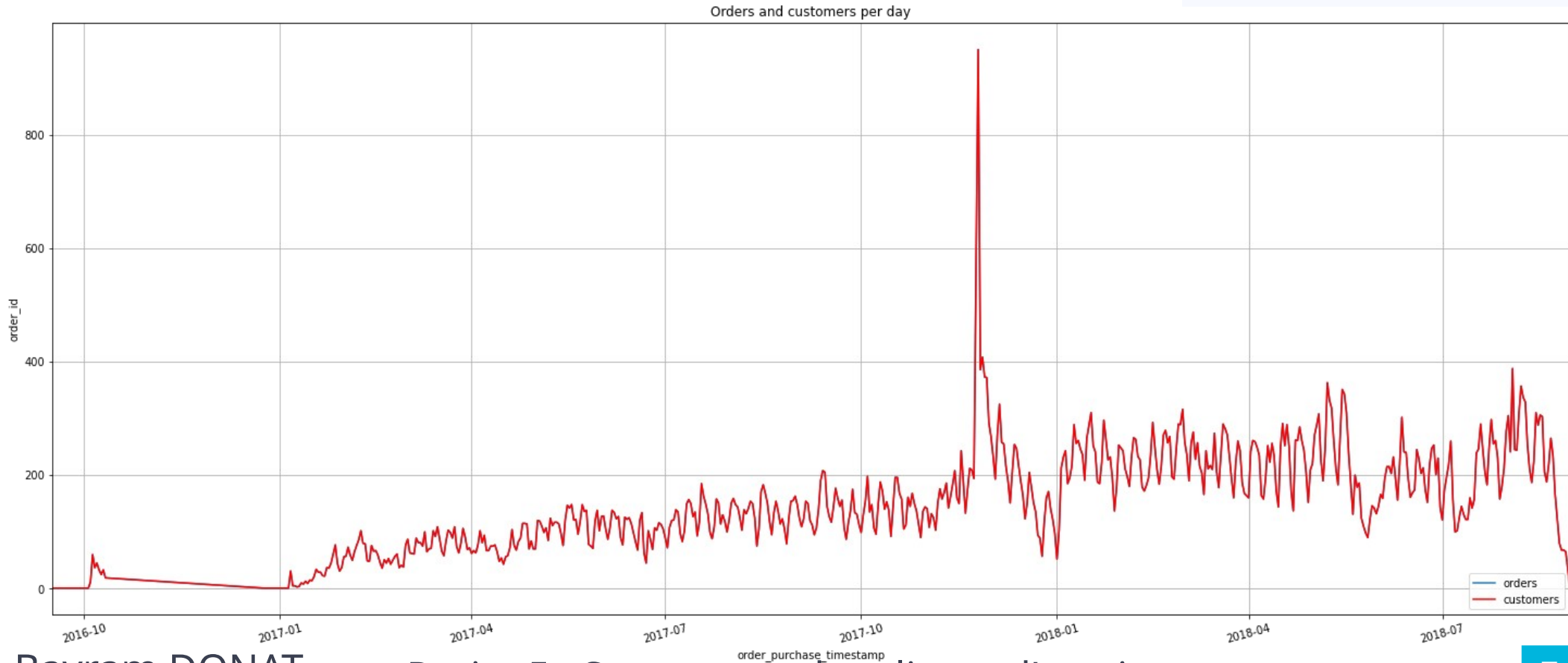
- Etat de commande : livrée
- Commande avec un avis

- Variables créés

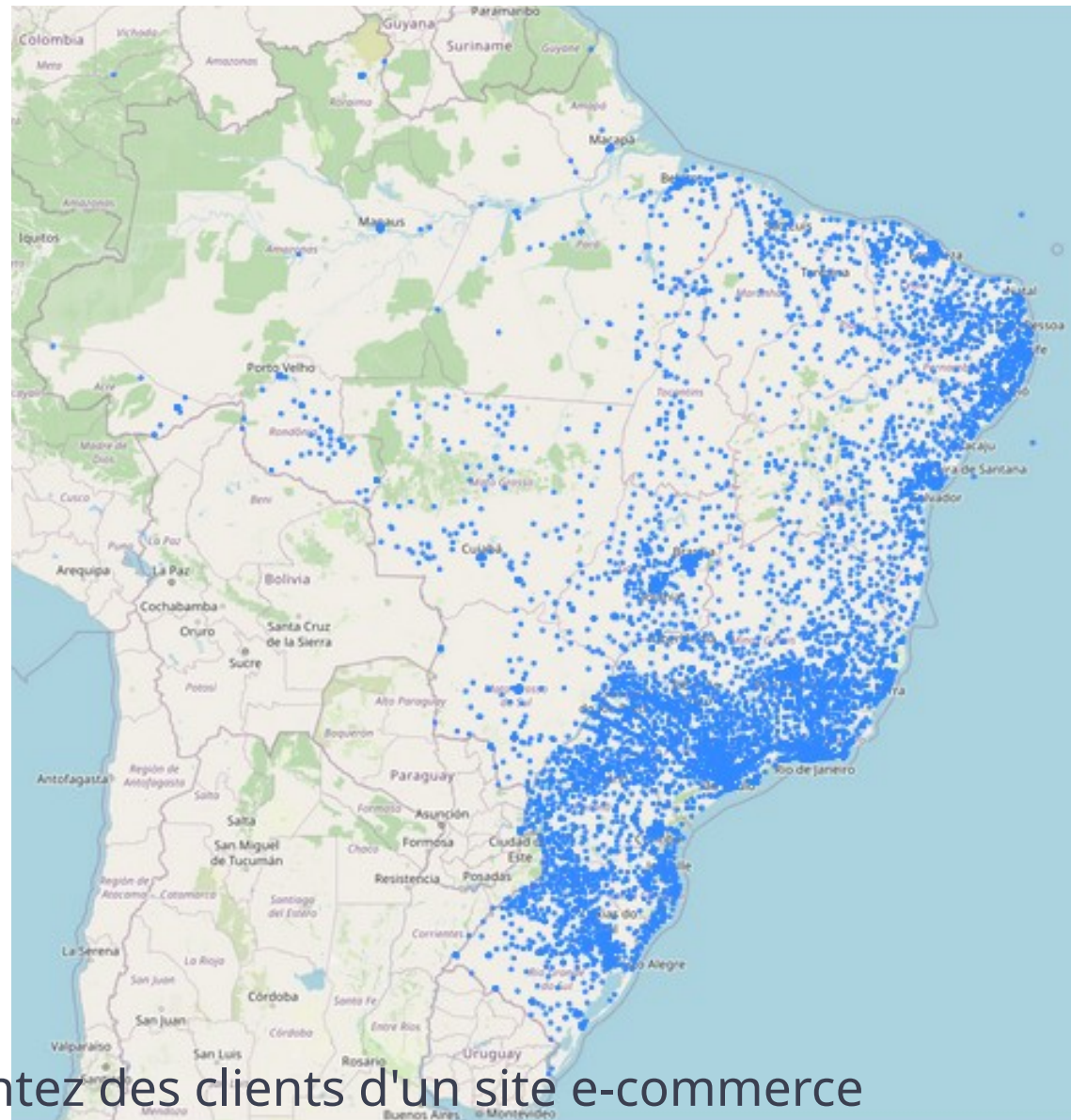
- Total_price : montant d'une commande
- items_count : nombre de produits d'une commande
- sellers_count : nombre de vendeurs par commande
- volume_L : volume total par commande
- delivery_time : temps de livraison réel en jours
- estimated_delivery_time : temps de livraison estimé en jours
- delivery_time_error : erreur sur le temps de livraison en jours (estimé - réel)
- days_since_last_order : temps depuis la dernière commande en jours
- first_order_date, last_order_date : dates de la première et de la dernière commande

Exploration

olist



Exploration

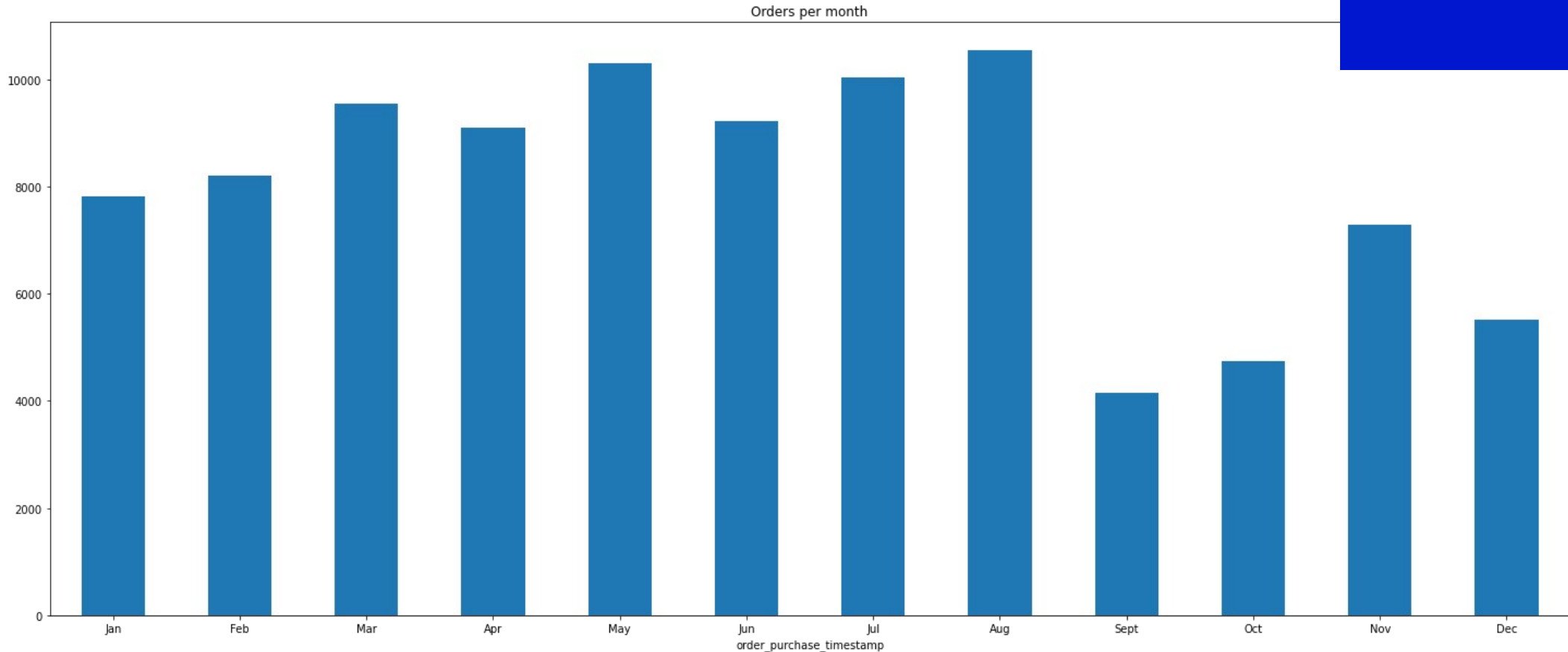


Bayram DONAT

Projet 5 : Segmentez des clients d'un site e-commerce

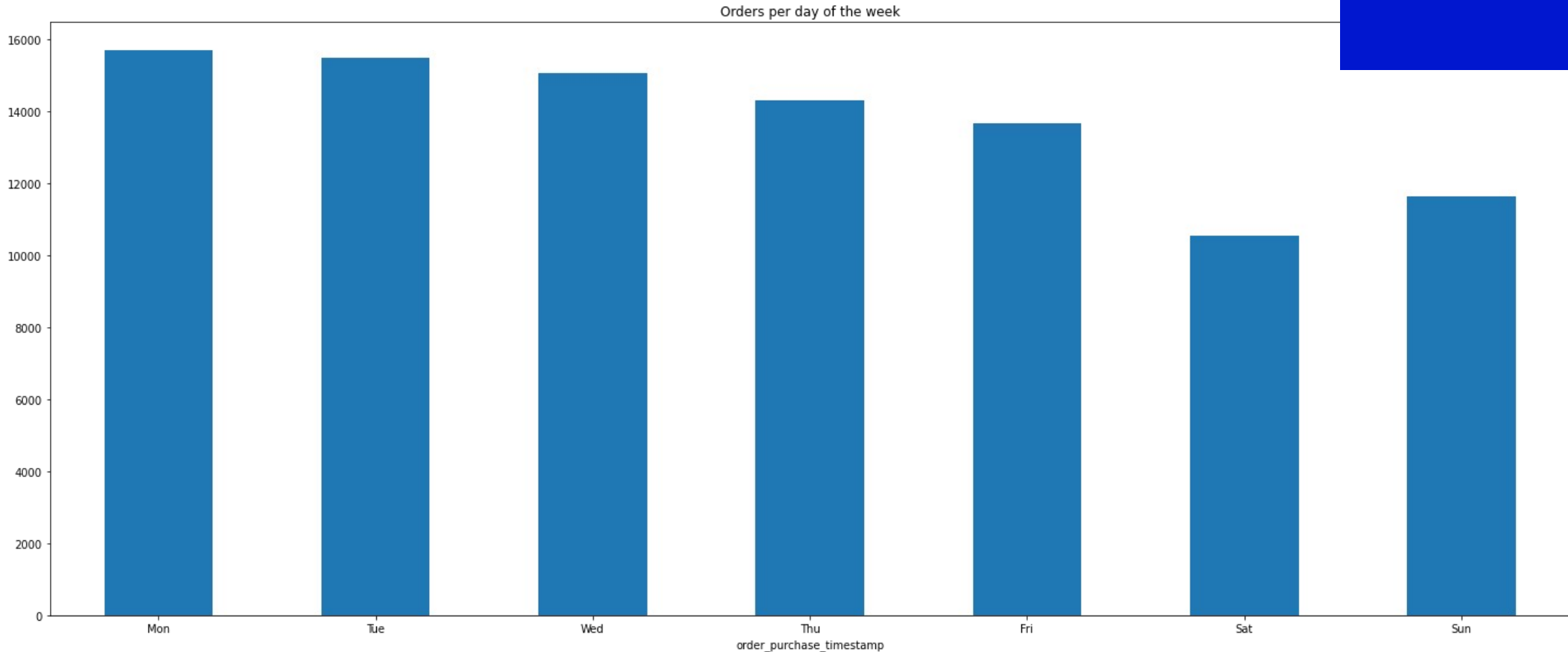
Exploration

olist

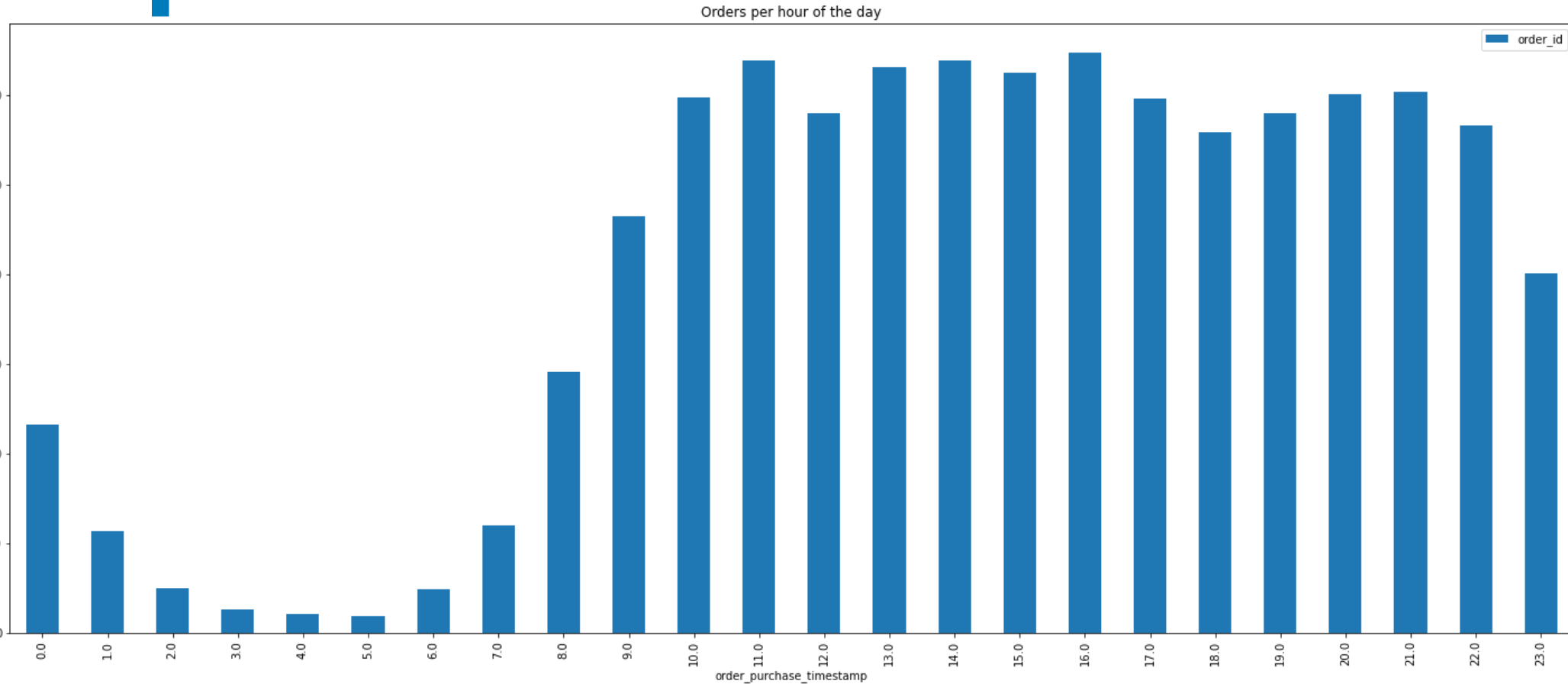


Exploration

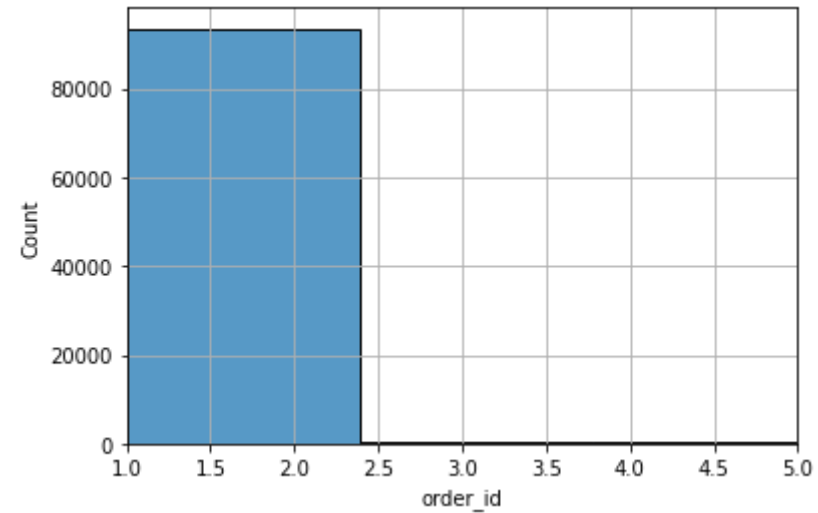
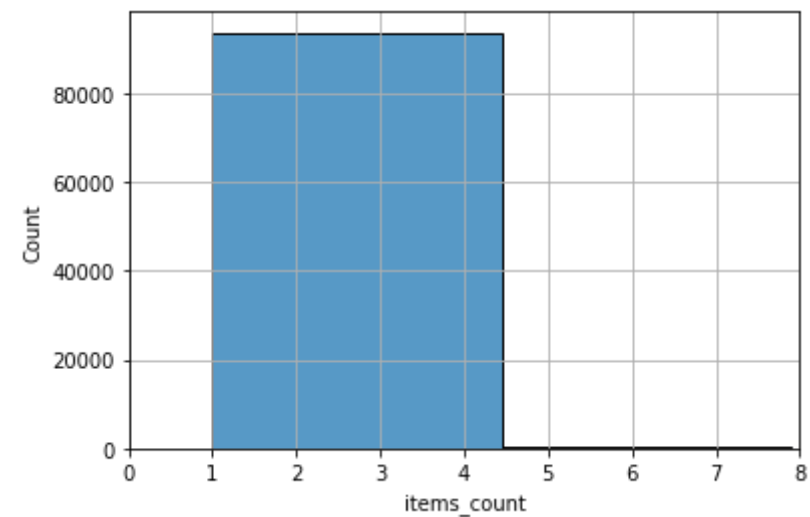
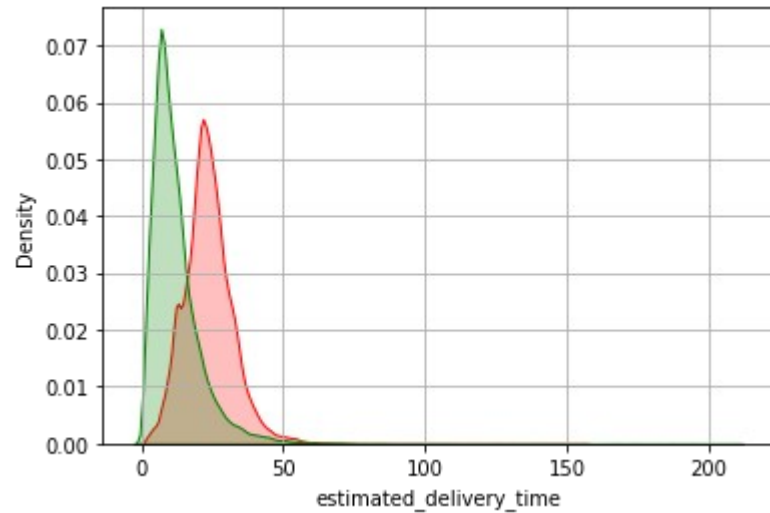
olist

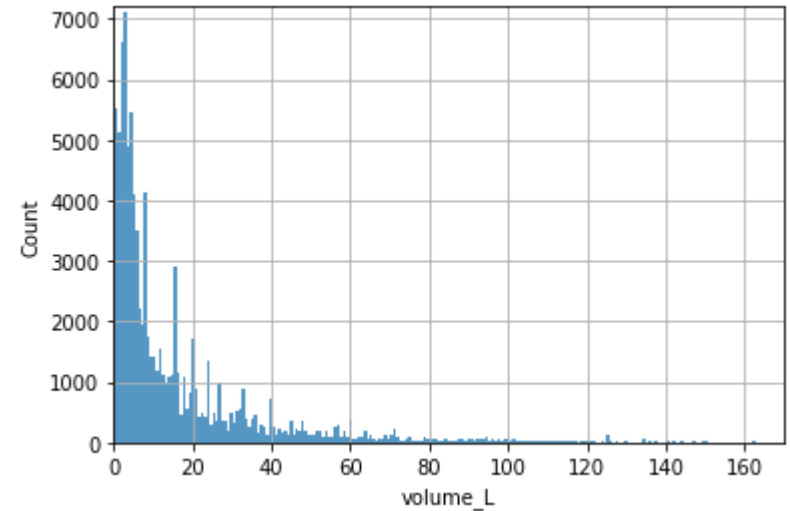
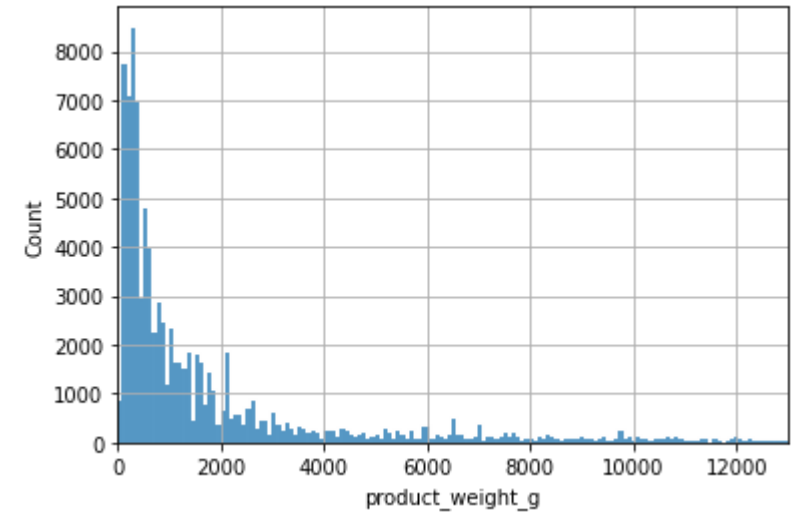
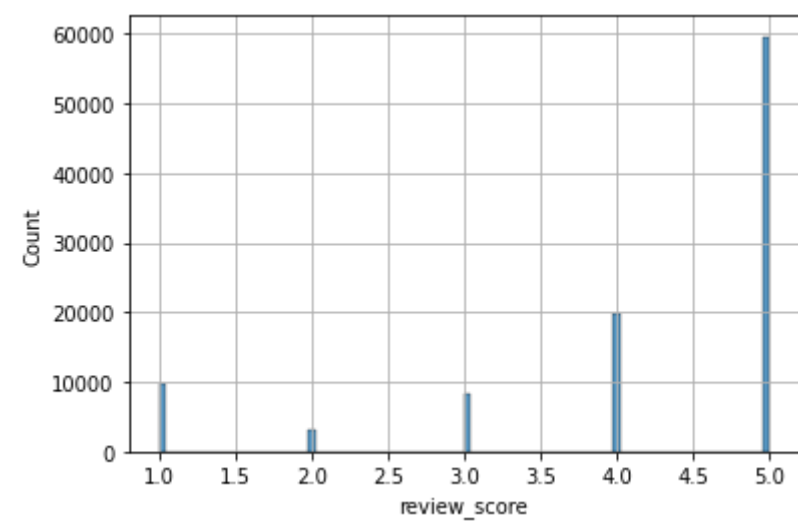
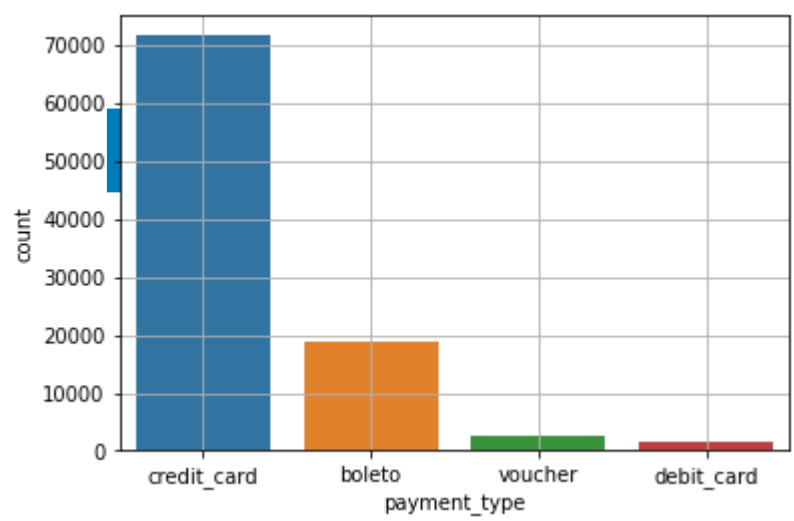


Exploration

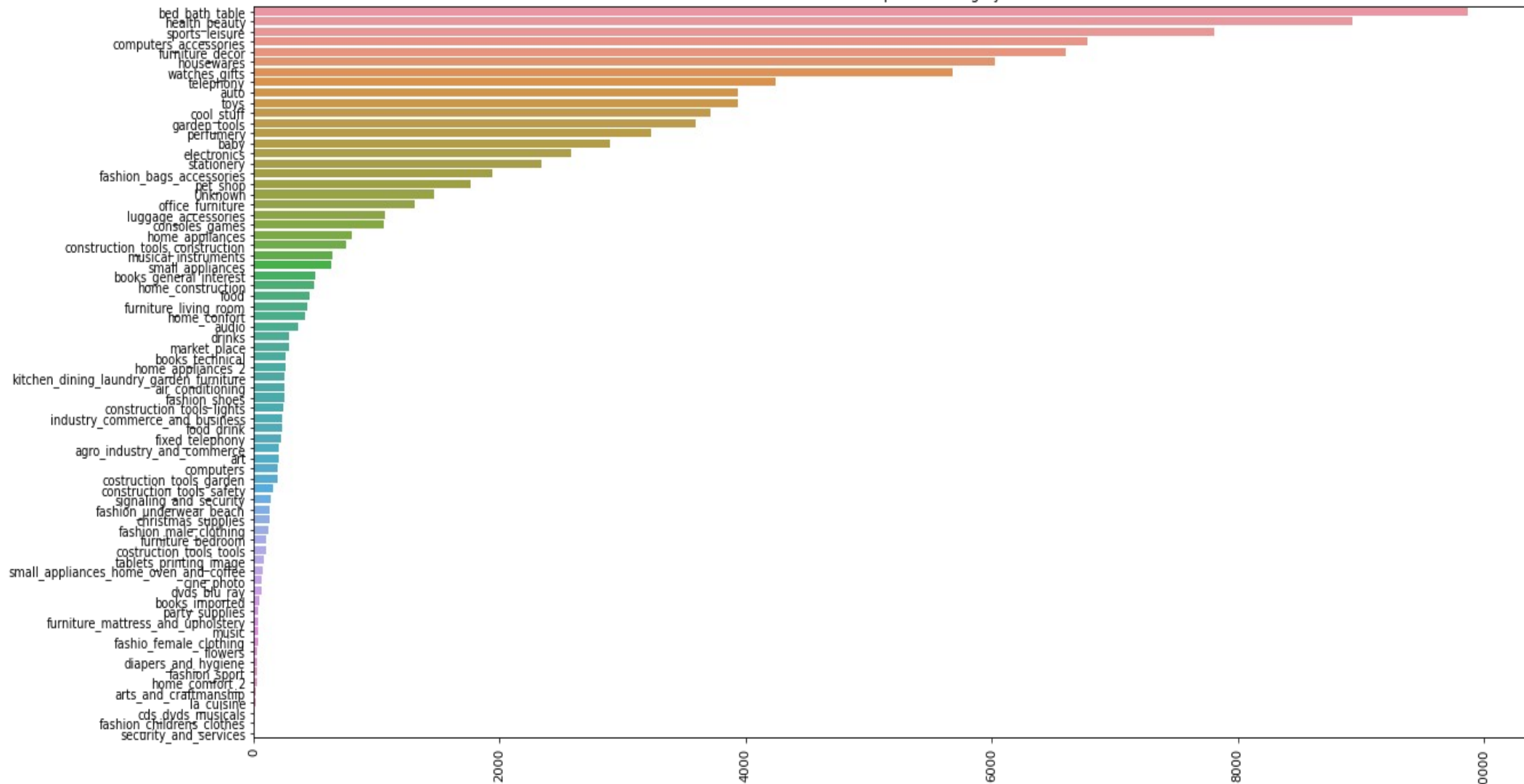


Exploration





Distribution of product category

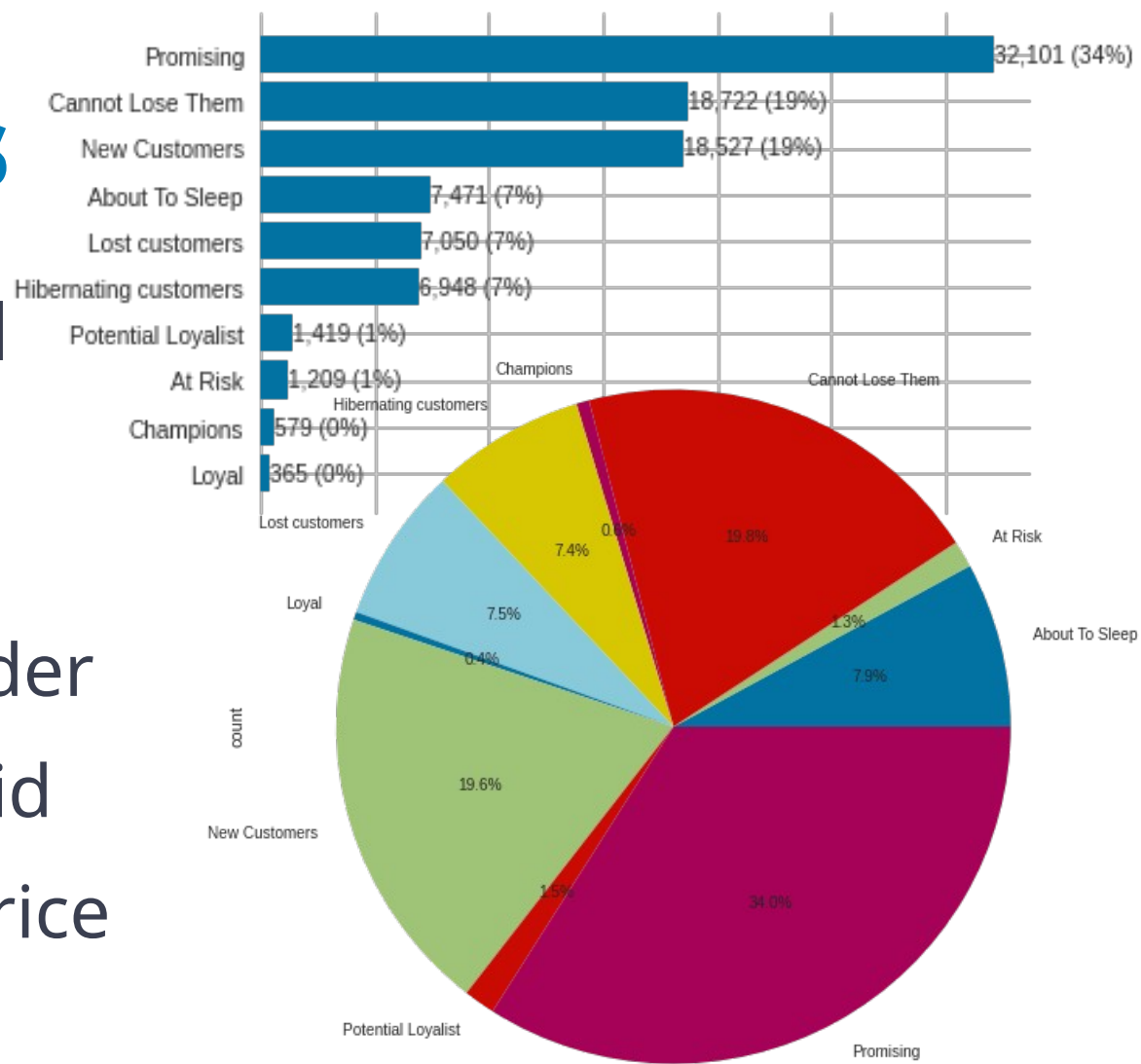


Modélisations

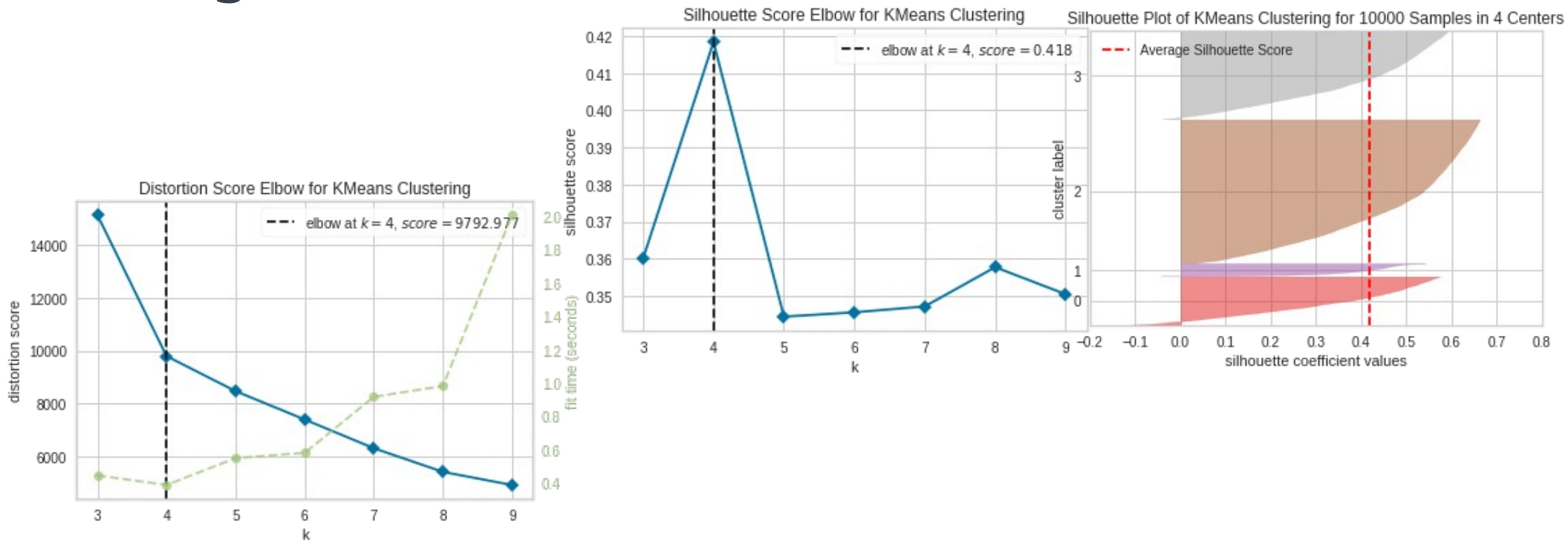
- Segmentation RFM manuelle
- Segmentation RFM évoluée
 - Kmeans
 - Agglomerative clustering
- Segmentation RFM incluant d'autres variables
 - Kmeans
 - Agglomerative clustering

Modélisations

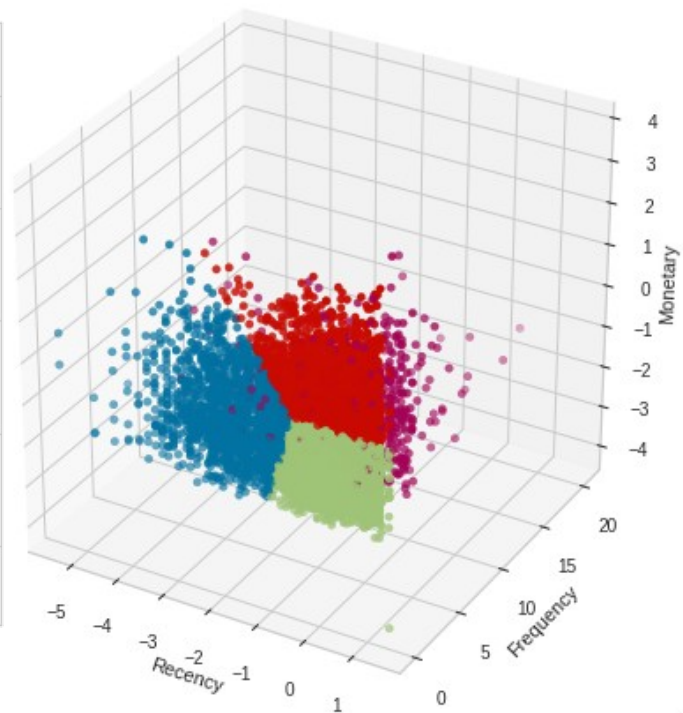
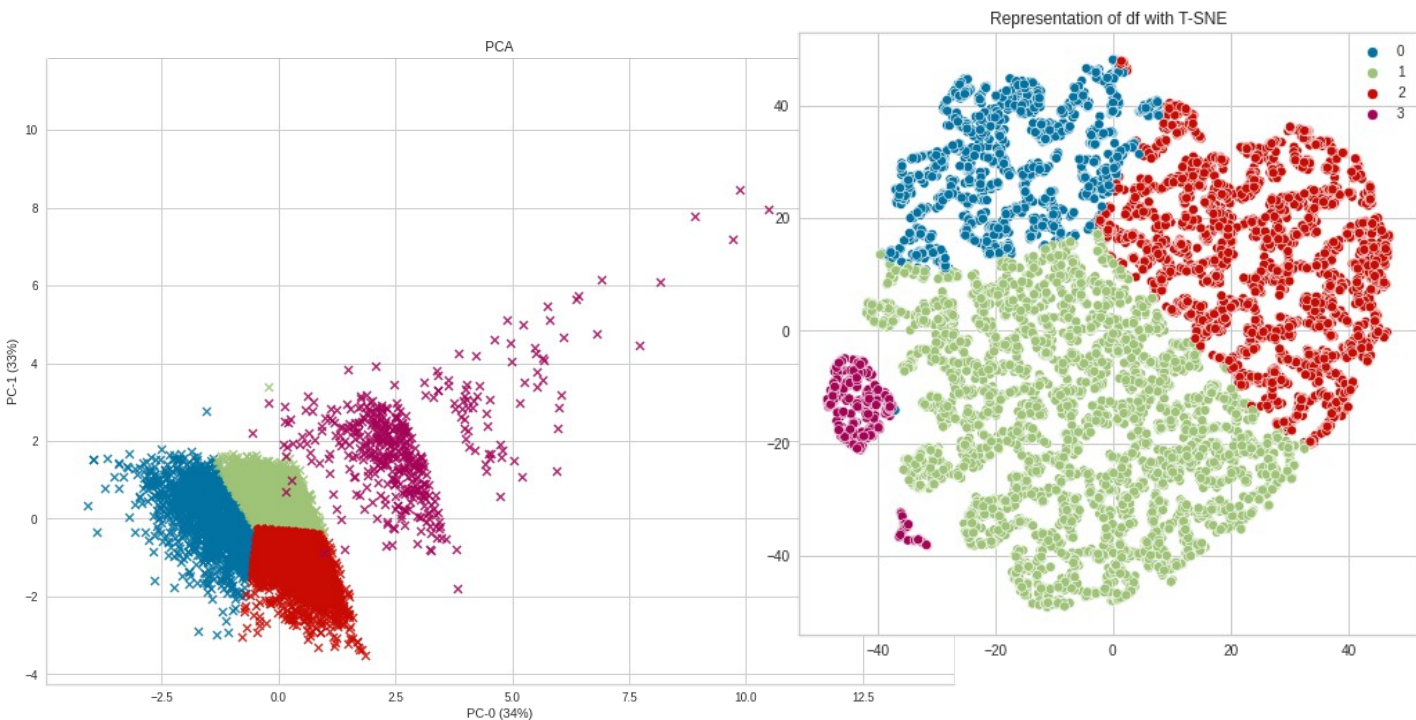
- Segmentation RFM manuelle
 - Recency : `days_since_last_order`
 - Frequency : `order_id`
 - Monetary : `Total_price`



● Segmentation RFM évoluée : Kmeans

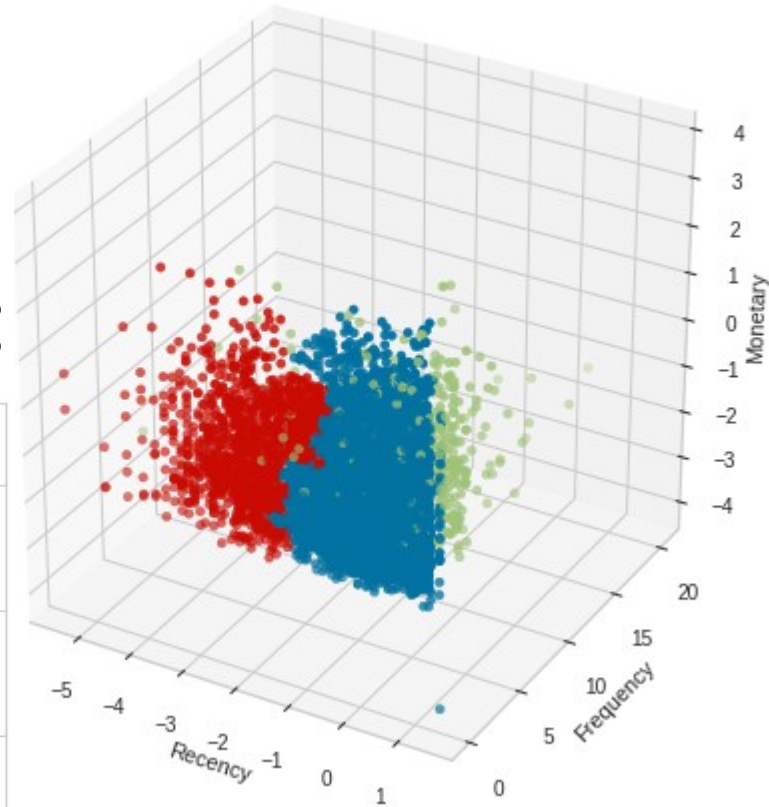
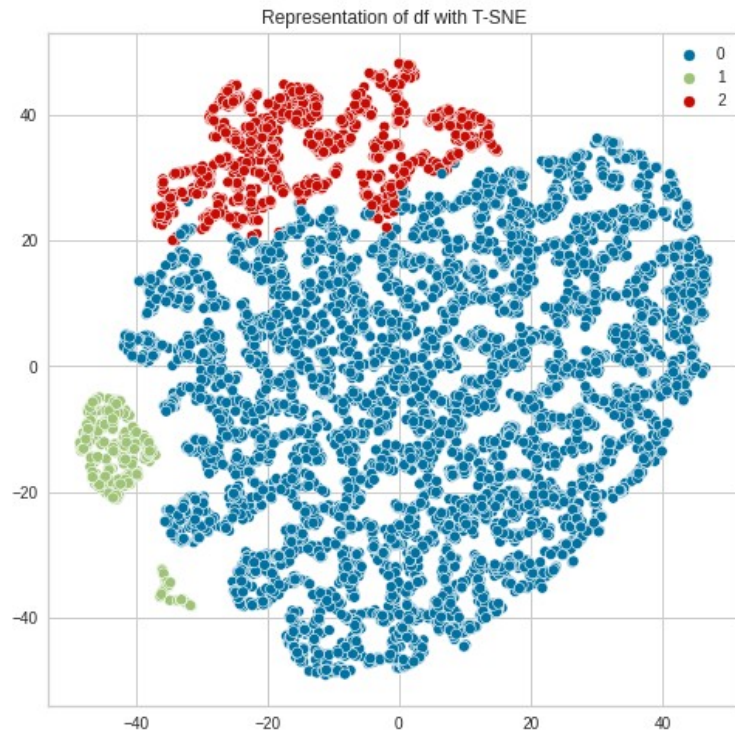
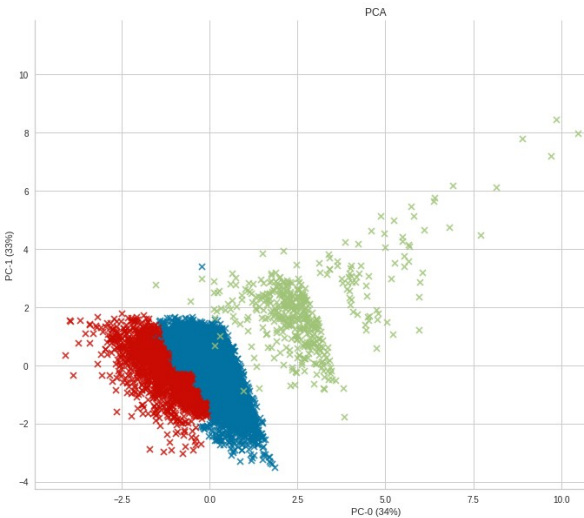


Modélisations



Modélisations

- Segmentation RFM évoluée : clustering



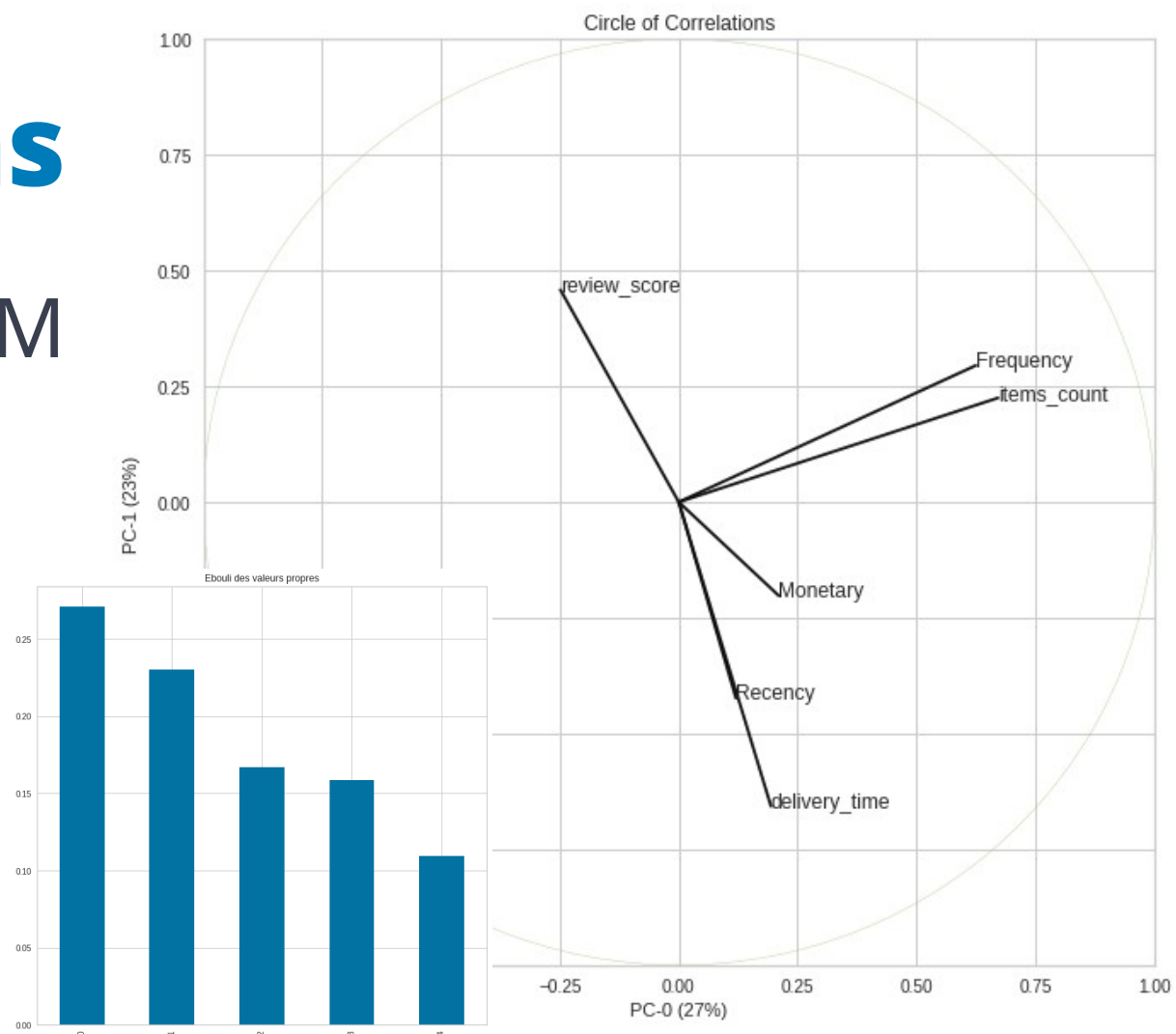
● Segmentation RFM évoluée : Analyse kmeans

Segment	Description	Actions possibles
0 New customer	A acheté les plus récemment	fournir une assistance de premier abord. lui donner un accès anticipé. Commencer à établir une relation
1 Loyal	A acheté au moins deux fois dans la période	Lui proposer des produits de plus grande valeur. Lui demander des avis.
2 Small	A acheté avec les plus petites dépenses et n'est pas revenu depuis un bon moment	lui fournir de l'aide personnalisé. Bien personnaliser ses mails.
3 Big	A acheté avec les plus grandes dépenses et n'est pas revenu depuis un bon moment.	le regagner via des renouvellements ou de nouveaux produits. Rentrer en contact si nécessaire. Bien personnaliser les mails

Modélisations

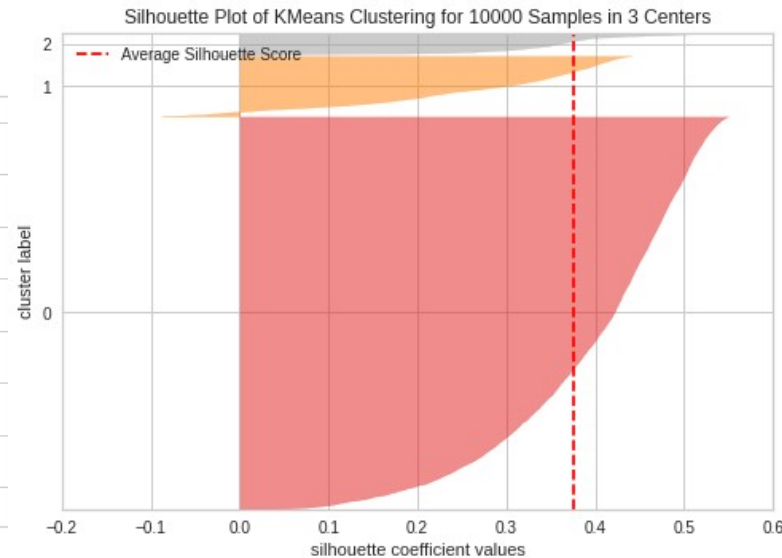
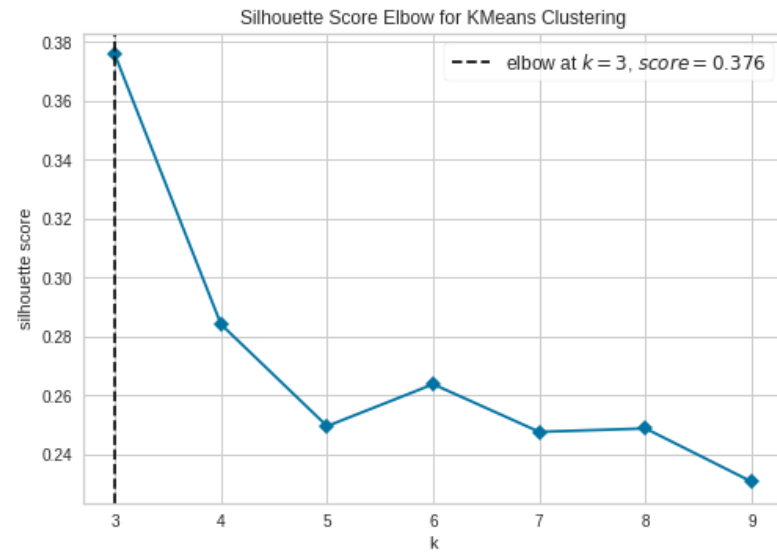
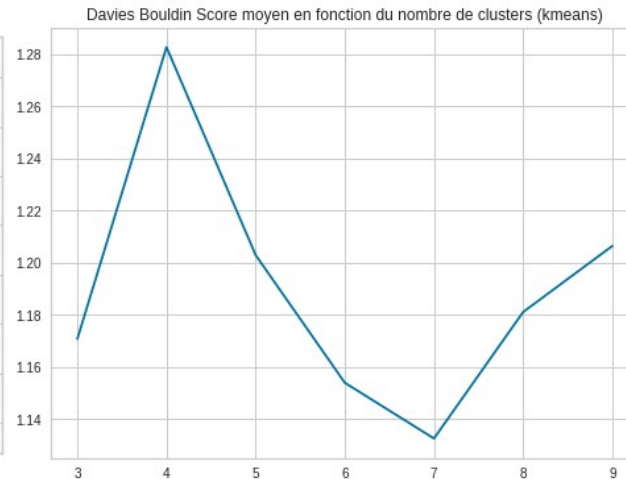
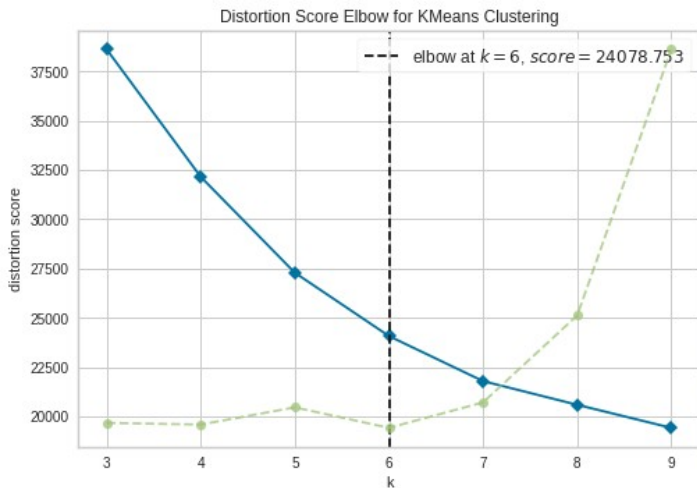
- Segmentation RFM incluant d'autres variables

- 'delivery_time'
- 'review_score'
- 'items_count'

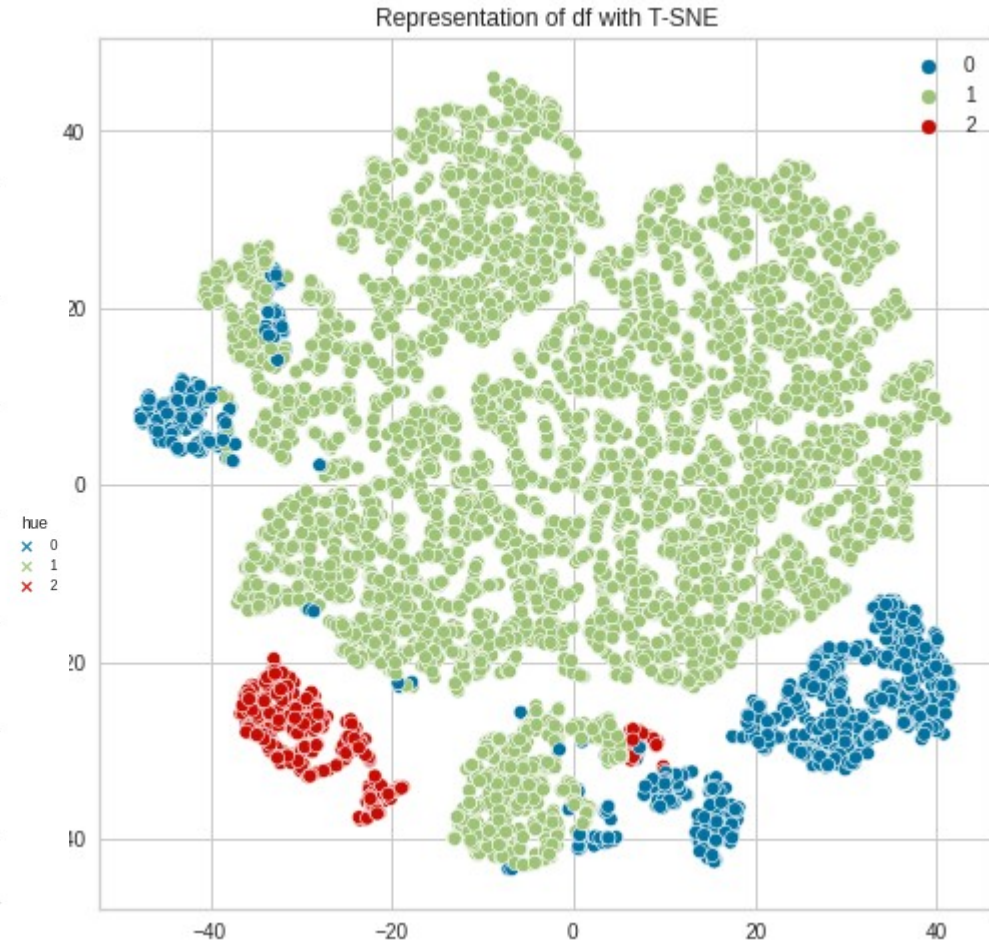
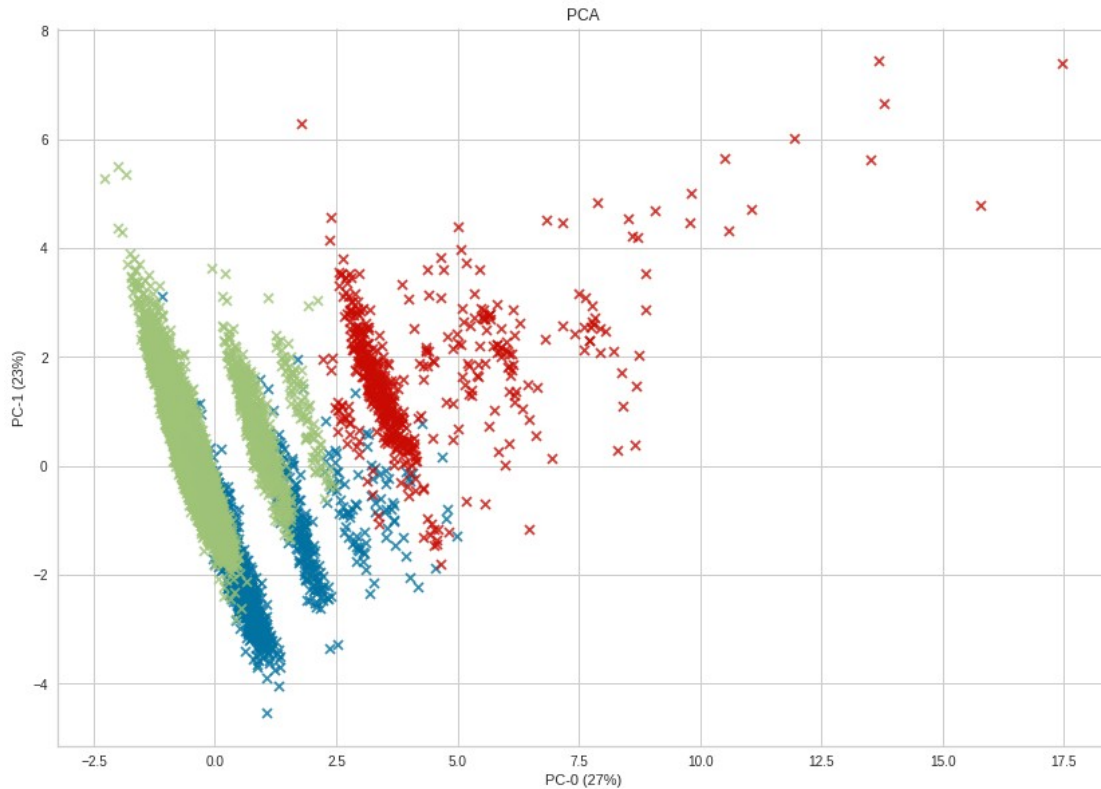


Modélisations

- Segmentation RFM incluant d'autres variables : kmeans

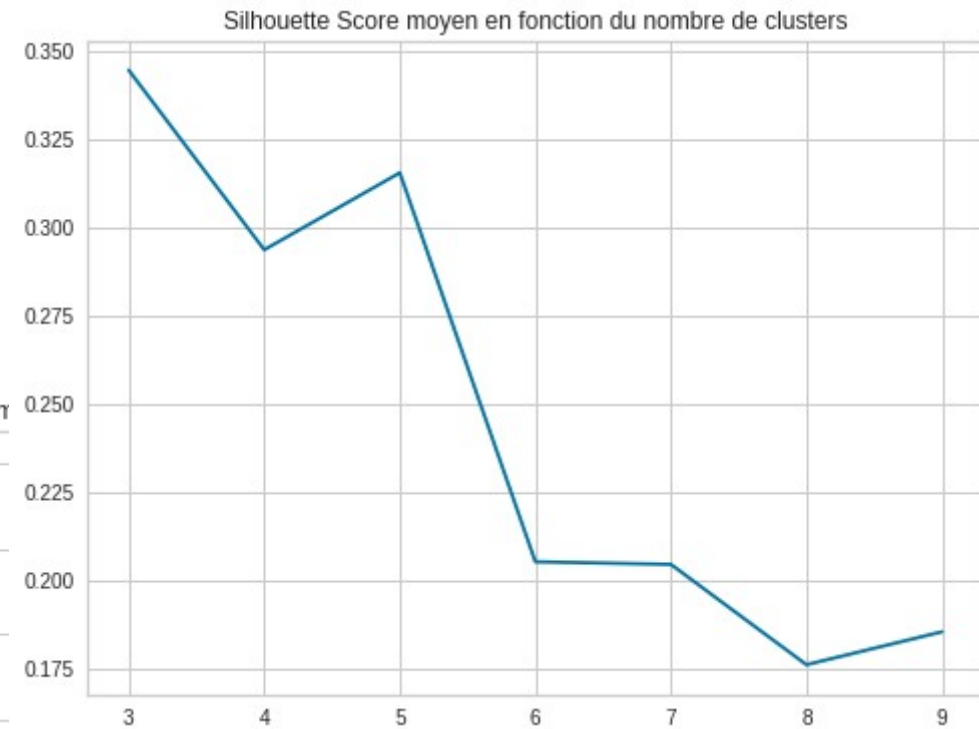
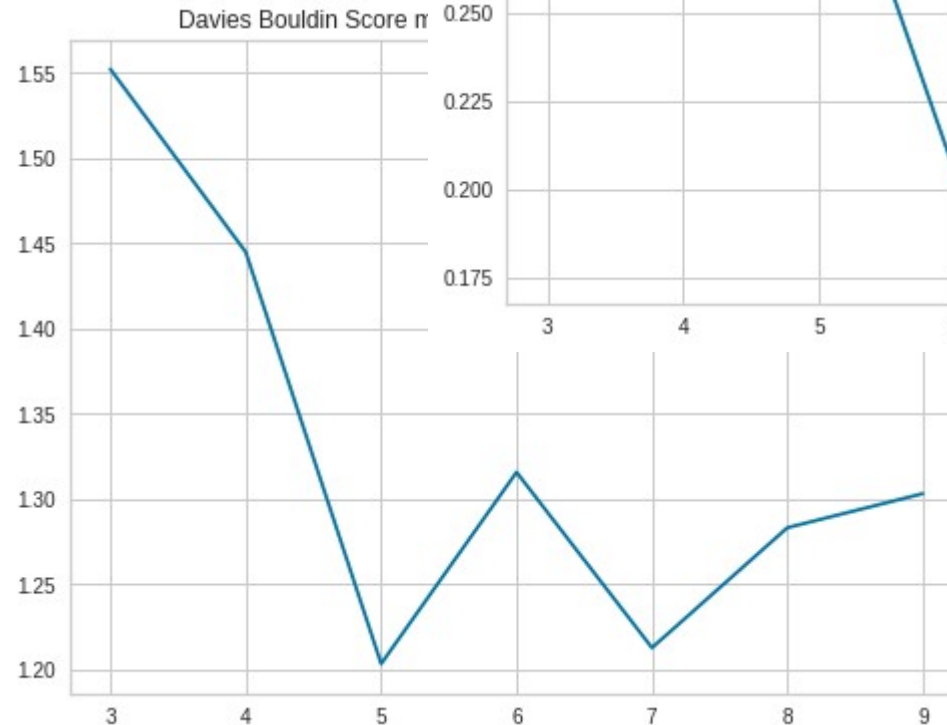


Modélisations

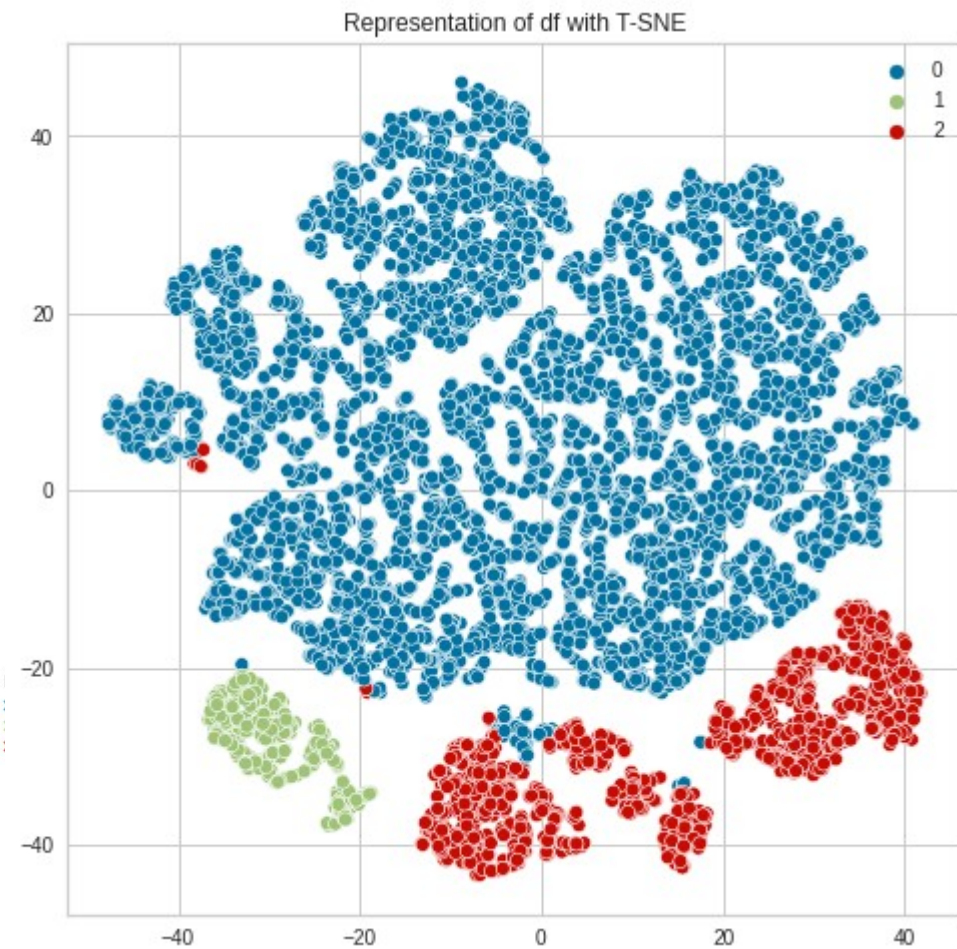
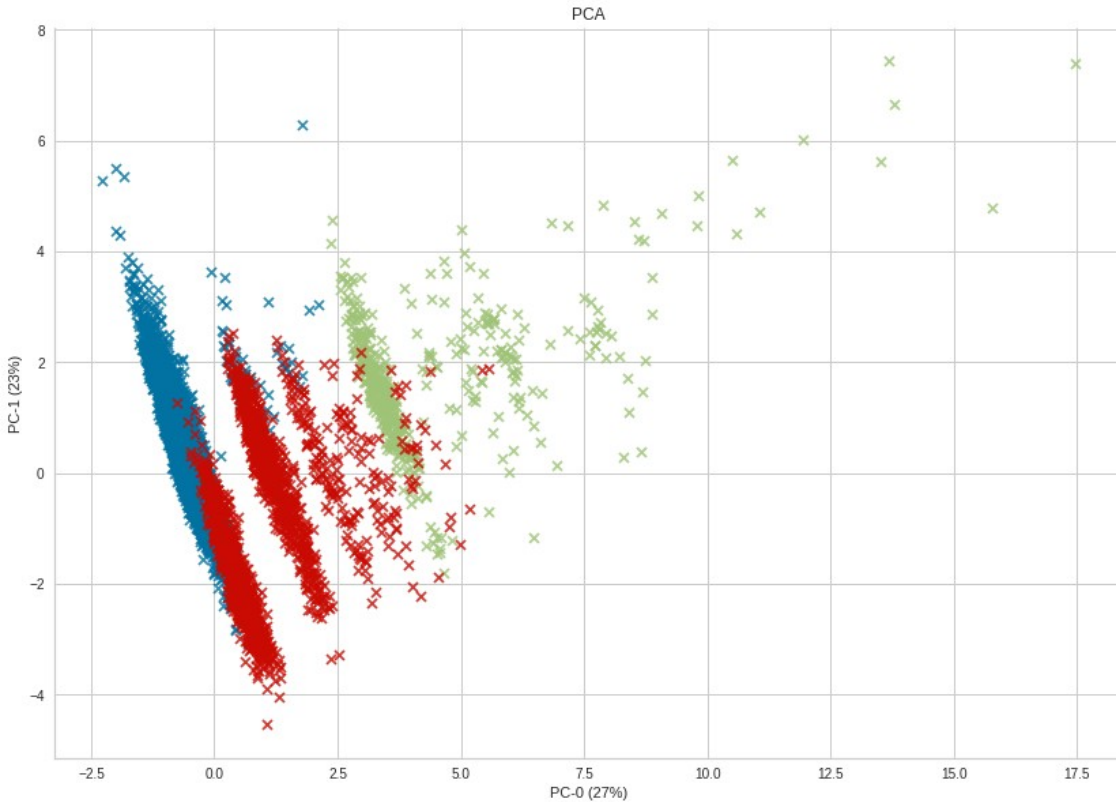


Modélisations

- Segmentation RFM incluant d'autres variables : Agglomerative Clustering



Modélisations



Modélisations

The logo for 'olist' is displayed in a bold, blue, sans-serif font. The letters are lowercase, and the 'o' is a simple circle. The logo is positioned in the top right corner of the slide.

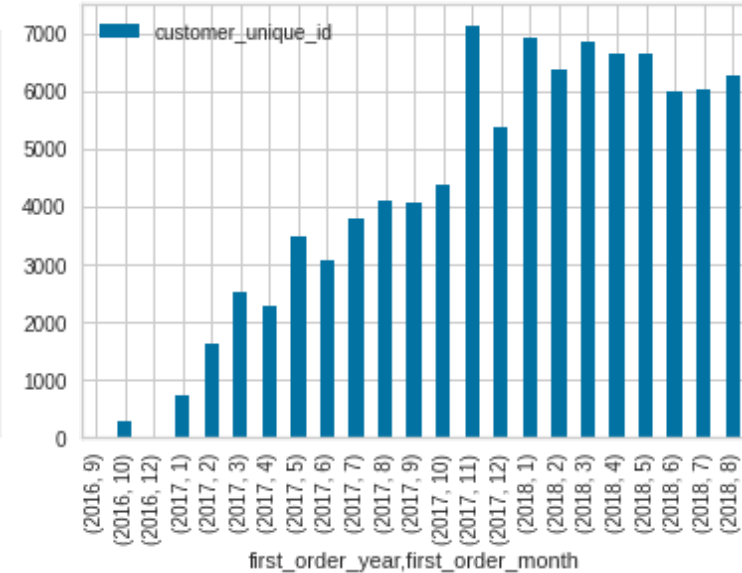
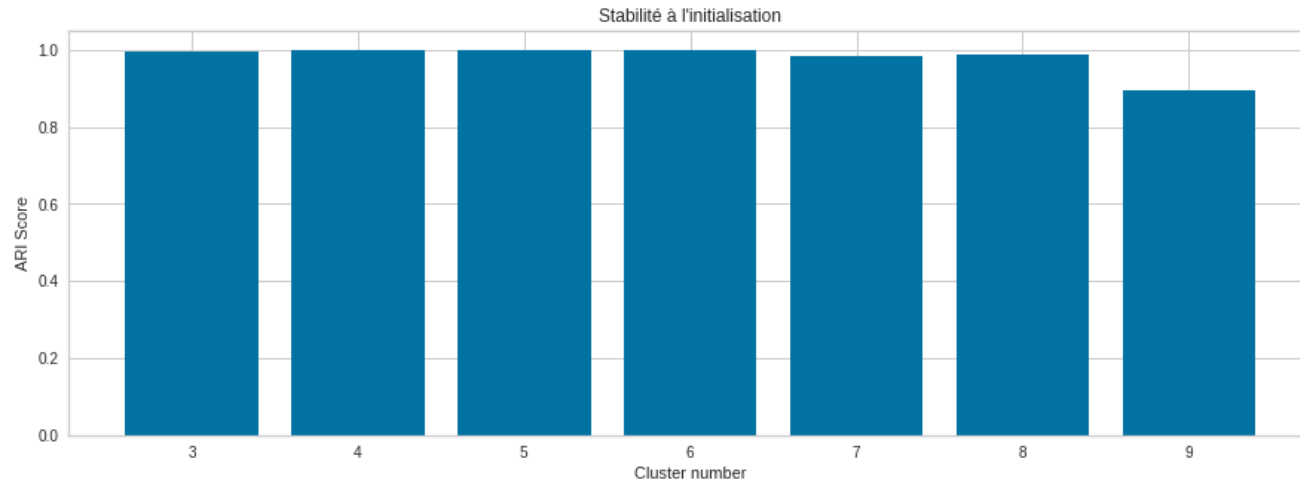
Segment	Description	Actions possibles
0 Fidèles	A acheté à moins de vendeurs différents. avec le plus d'articles dans chaque commande, avec moins de frais, plus léger, moins encombrant, avec un temps de livraison très rapide.	Lui proposer des produits de plus grande valeur. Lui demander des avis.
1 Corps	Le temps de livraison moyen, volume le plus grand, poids le plus lourd	Faire des offres limitées dans le temps.
2 Mécontents	Une mauvaise staisfaction. Le temps de livraison le plus long et pas revenu depuis longtemps	Retravailler son intérêt s'il répond à la campagne de sensibilisation.

Délai de maintenance

- Variables
 - Recency : `days_since_last_order`
 - Frequency : `order_id`
 - Monetary : `Total_price`
 - First_order_date : date de la première commande
 - Last_order_date : date de la dernière commande
- Fonctions : `sklearn.metrics.adjusted_rand_score`

Délai de maintenance

- `sklearn.metrics.adjusted_rand_score=1` pour $k=4$



Délai de maintenance

- Sklearn
 - `metrics.adjusted_rand_score=1` pour `k=4`
 - `decomposition.pca.explained_variance_ratio_`

Délai de maintenance

- RFM avec kmeans

- $ARI < 0.5$

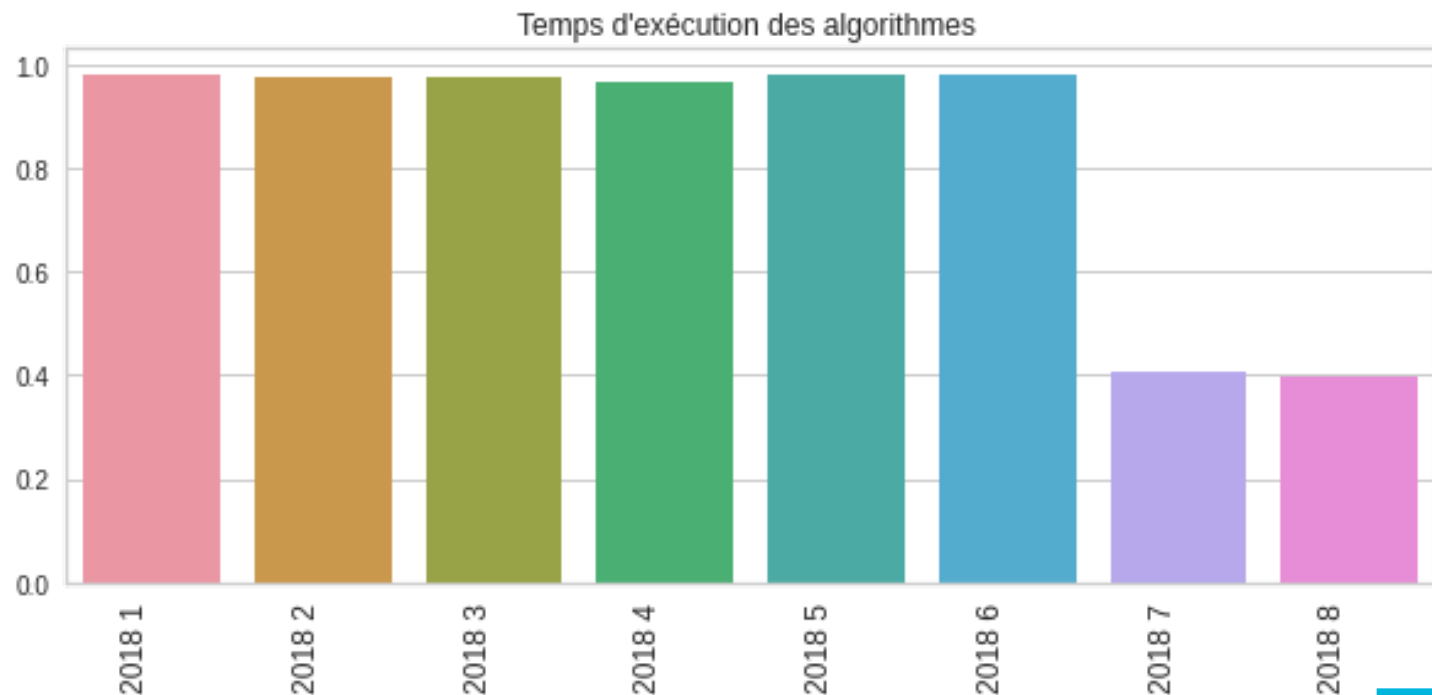
- 07-2018

- 08-2018

- Maintenance

à faire tous

les 6 mois



Conclusion

- Une bonne segmentation (modèle) dépend
 - Des variables à prendre en compte
 - De l'évolution de la base de données
 - De l'interprétation
- Améliorations
 - Utilisation des variables catégorielles
 - Autres algorithmes comme DBSCAN