

Segmentez des clients d'un site e-commerce

OPENCLASSROOMS

olist



ENJEUX

Mise en Place Ecosysteme Data

Construction & Maintenance d'un Dashboard

Segmentation Clientèle

SOMMAIRE

- **1. Feature Engineering & Analyse Exploratoire**
- **2. Développement & Simulation de Modèles**
- **3. Description Actionnable des Groupes de Clients (Clusters)**
- **4. Proposition de Contrat de Maintenance**

PREPARATION DES DONNEES

OLIST de 2016/09 à 2018/19

93.358 clients uniques

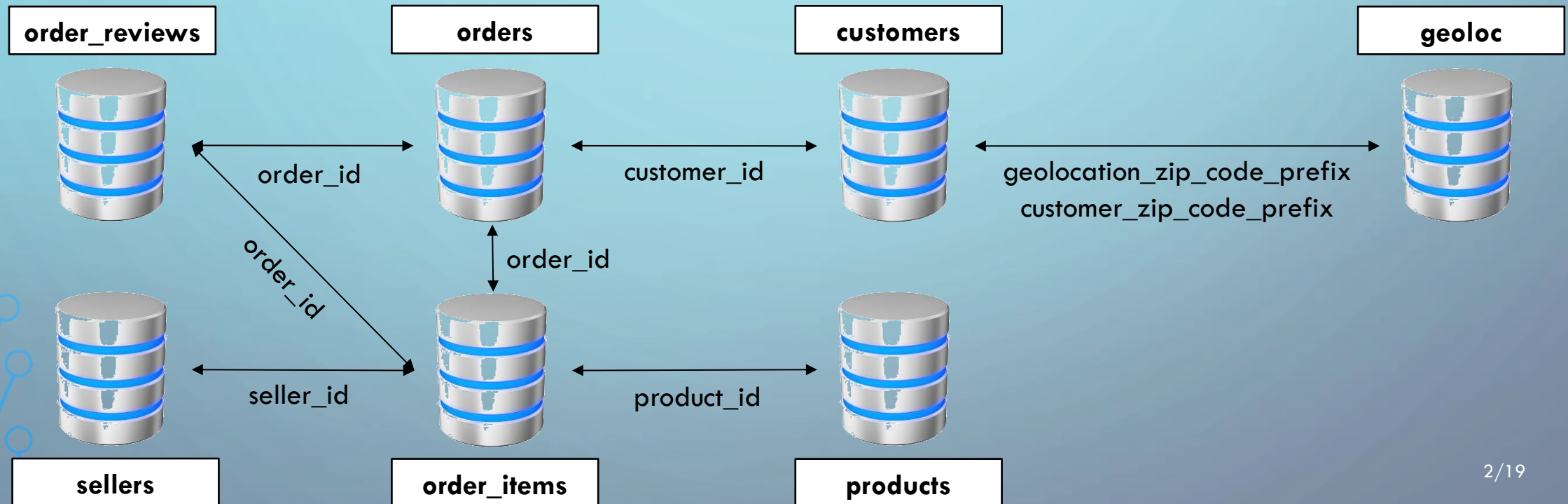
99.441 opérations d'achat

3.905 vendeurs agréés

97% opérations d'achat unique

+75% livraisons en avance

32.951 produits en catalogue

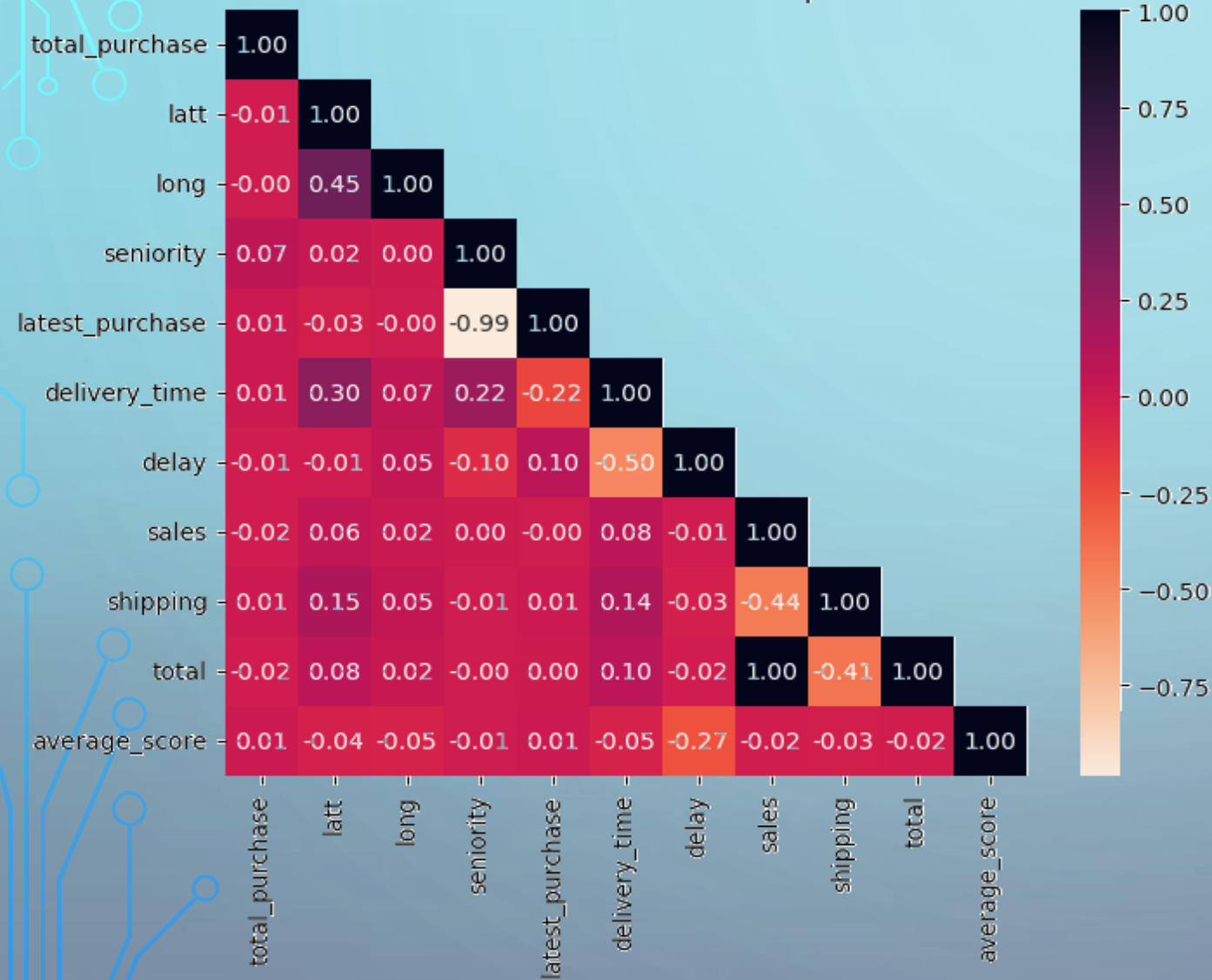


FEATURE ENGINEERING

TYPE VARIABLE	NOM	DESCRIPTION	MOYENNE
Identifiant	customer_unique_id	Identifiant client unique	
Géographique	latt	Localisation Géographique du Client	
	long		
	customer_state	Région Administrative du Client	
Comportemental	total_purchase	Nombre total d'achats effectués depuis 2016-09-15	1,04
	latest_purchase	Date de la dernière opération d'achat du client	2018-01-04
	average_score	Niveau de satisfaction du client (/5)	4,14
Logistique	delivery_time	Date de livraison prévisionnelle – Date de l'Achat (jours)	23,72
	delay	Date de livraison réelle – Date de livraison prévisionnelle (jours)	-11,15
Monétaire	sales	Coût de l'opération d'achat	128,56 R\$
	shipping	Part des frais de fret dans le total de l'opération d'achat (%)	20,86%
	total	Coût total de l'opération d'achat	149.29 R\$

ANALYSE EXPLORATOIRE : FEATURES

Matrice de corrélation des variables quantitatives



Choix des Variables pour nos Modèles de Clustering:

❖ Variables Explicatives :

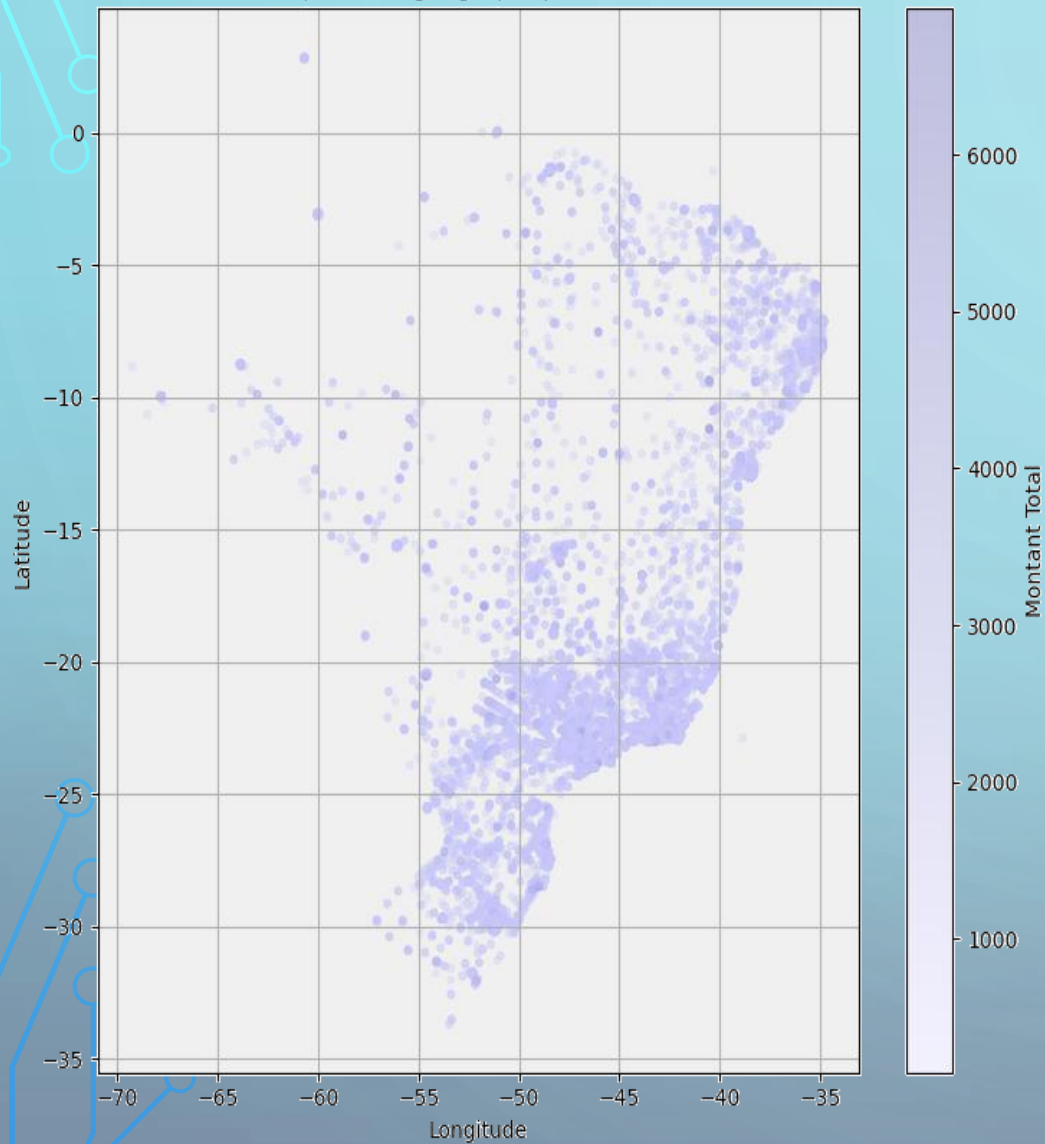
- latest_purchase_numeric
- total_purchase
- total
- average_score

❖ Variables Descriptives :

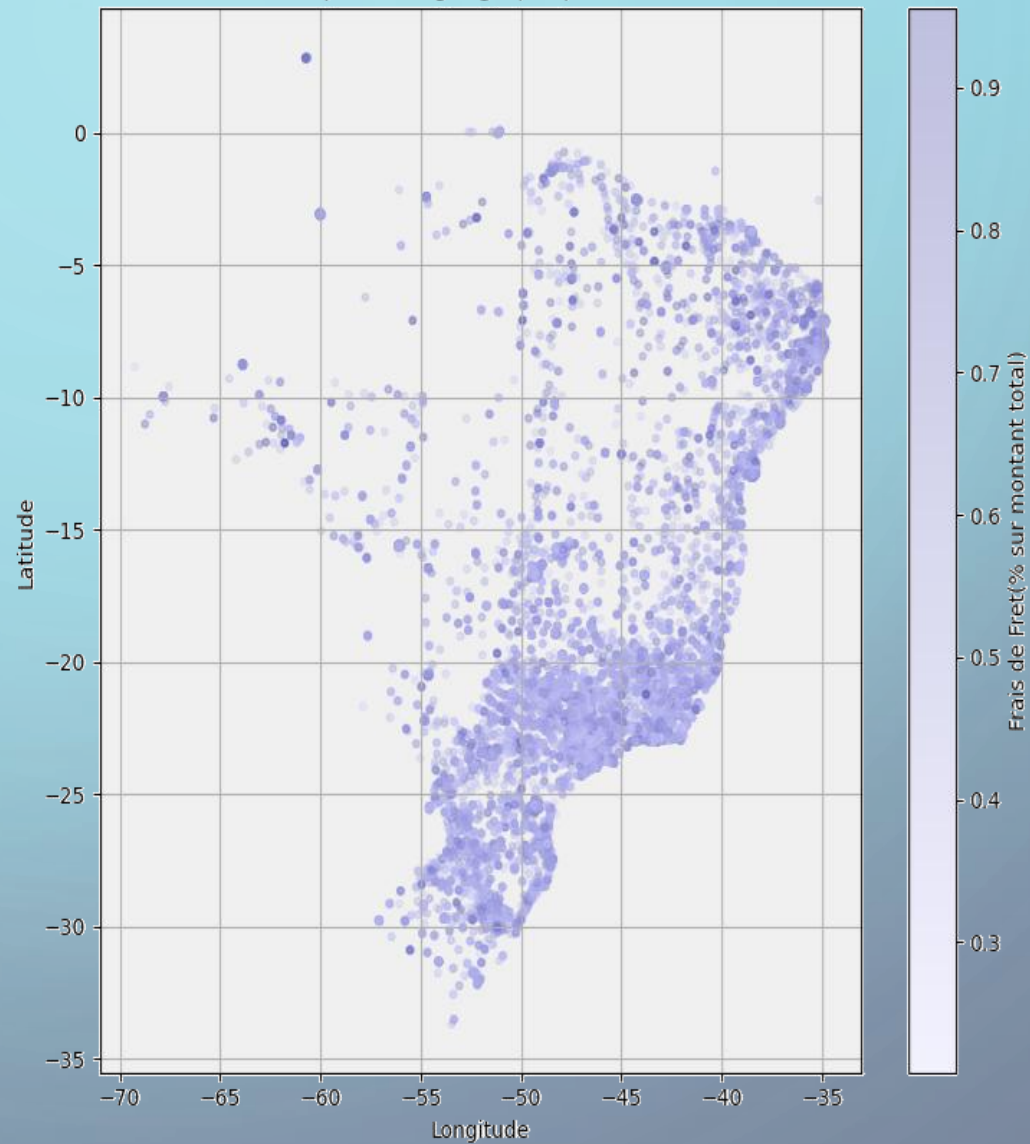
- latt
- long
- delivery_time
- delay
- shipping

ANALYSE EXPLORATOIRE : REPARTITION GEOGRAPHIQUE

Répartition géographique / Ventes

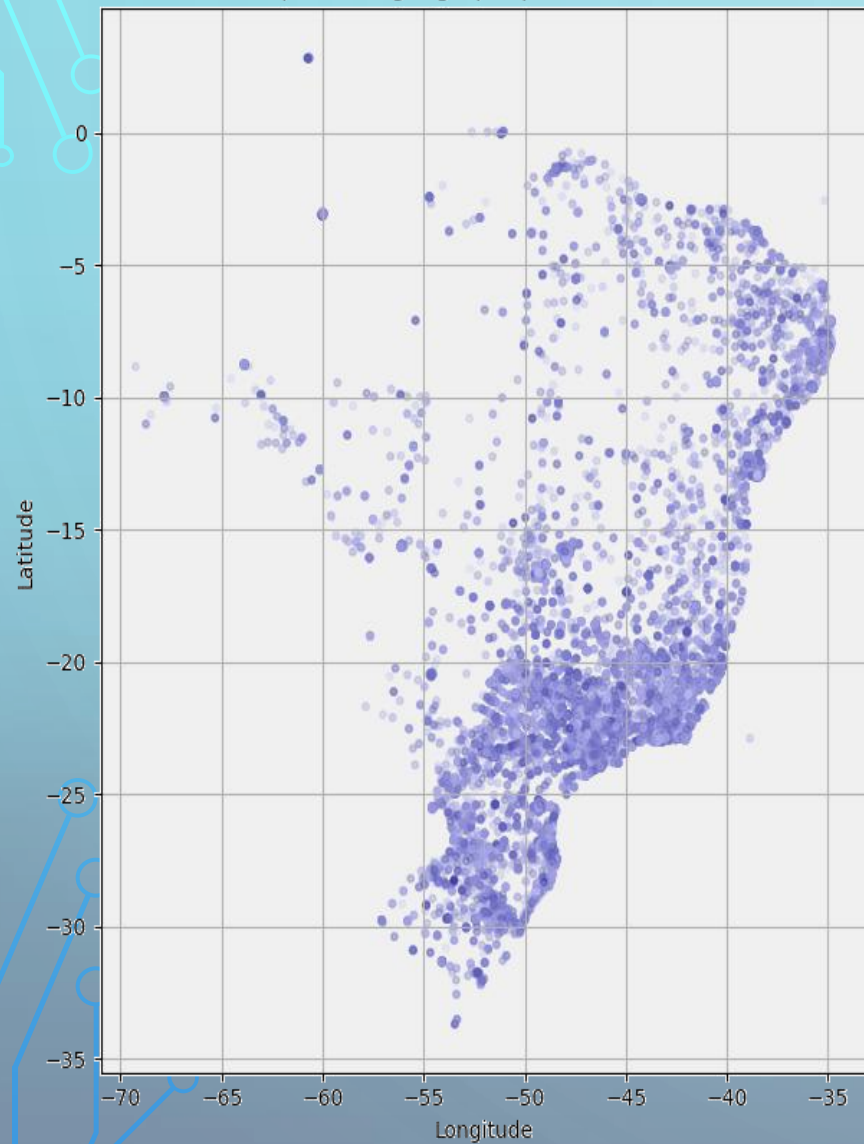


Répartition géographique / Fret

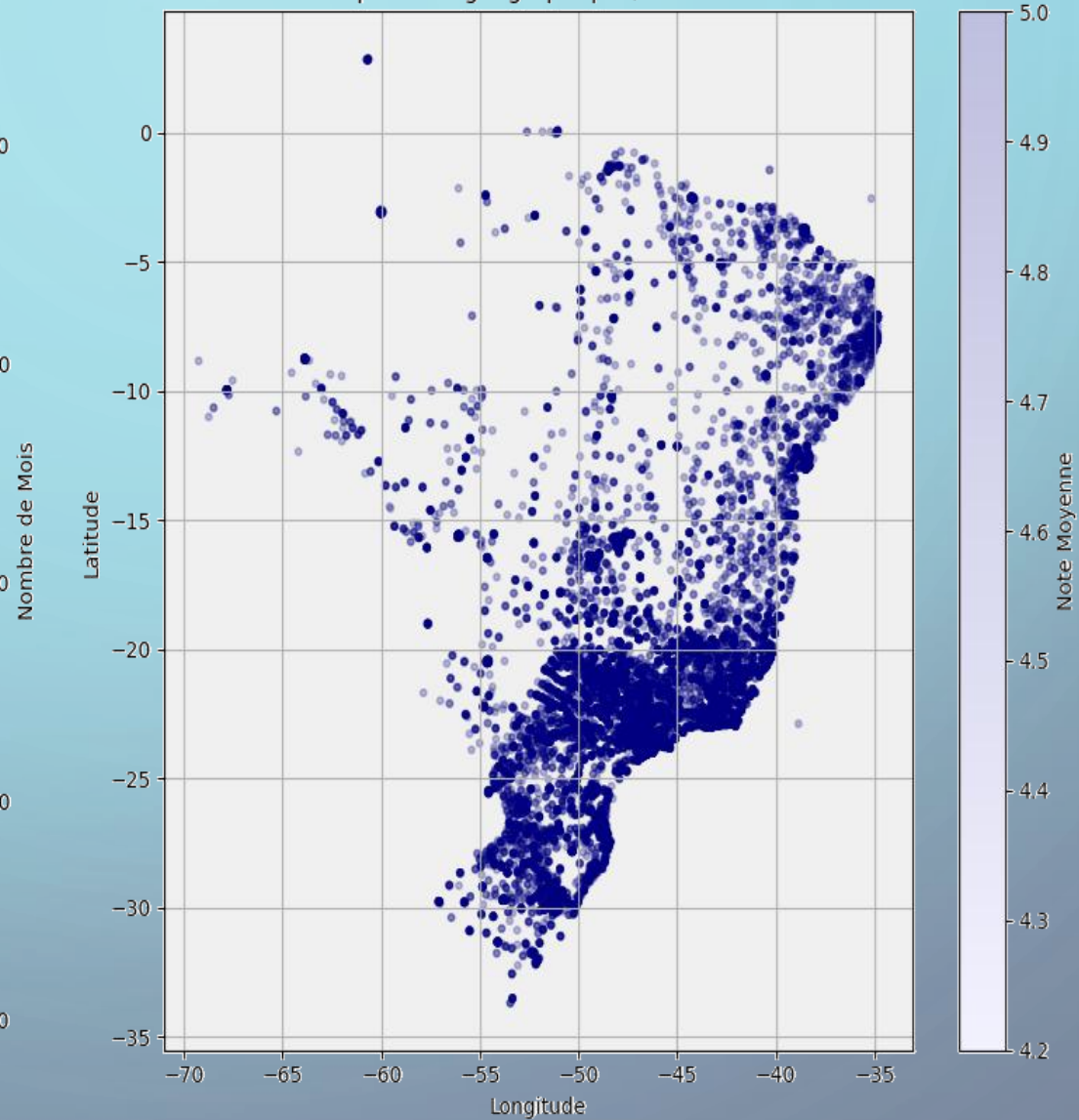


ANALYSE EXPLORATOIRE : REPARTITION GEOGRAPHIQUE

Répartition géographique / Ancienneté

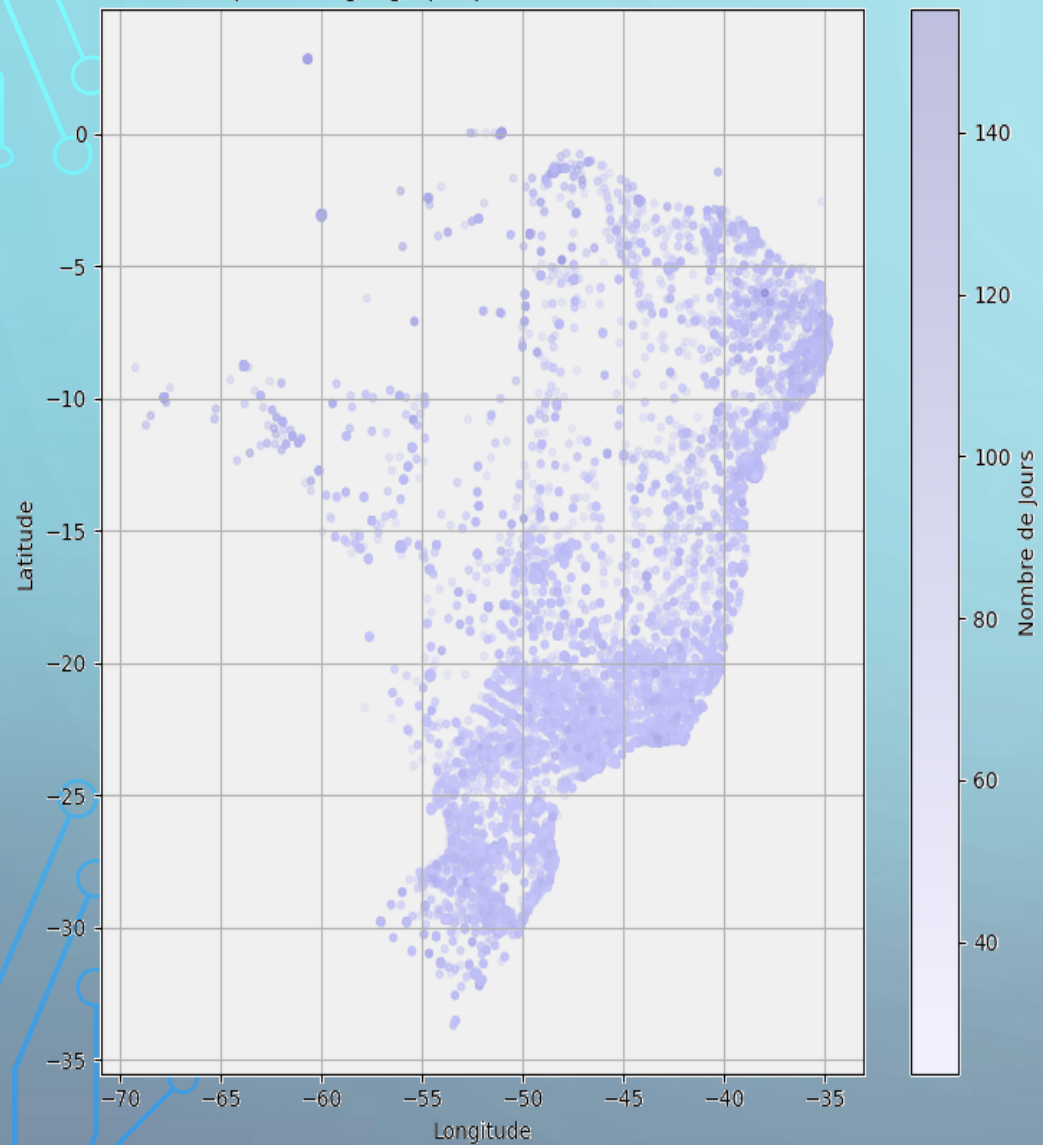


Répartition géographique / Notation

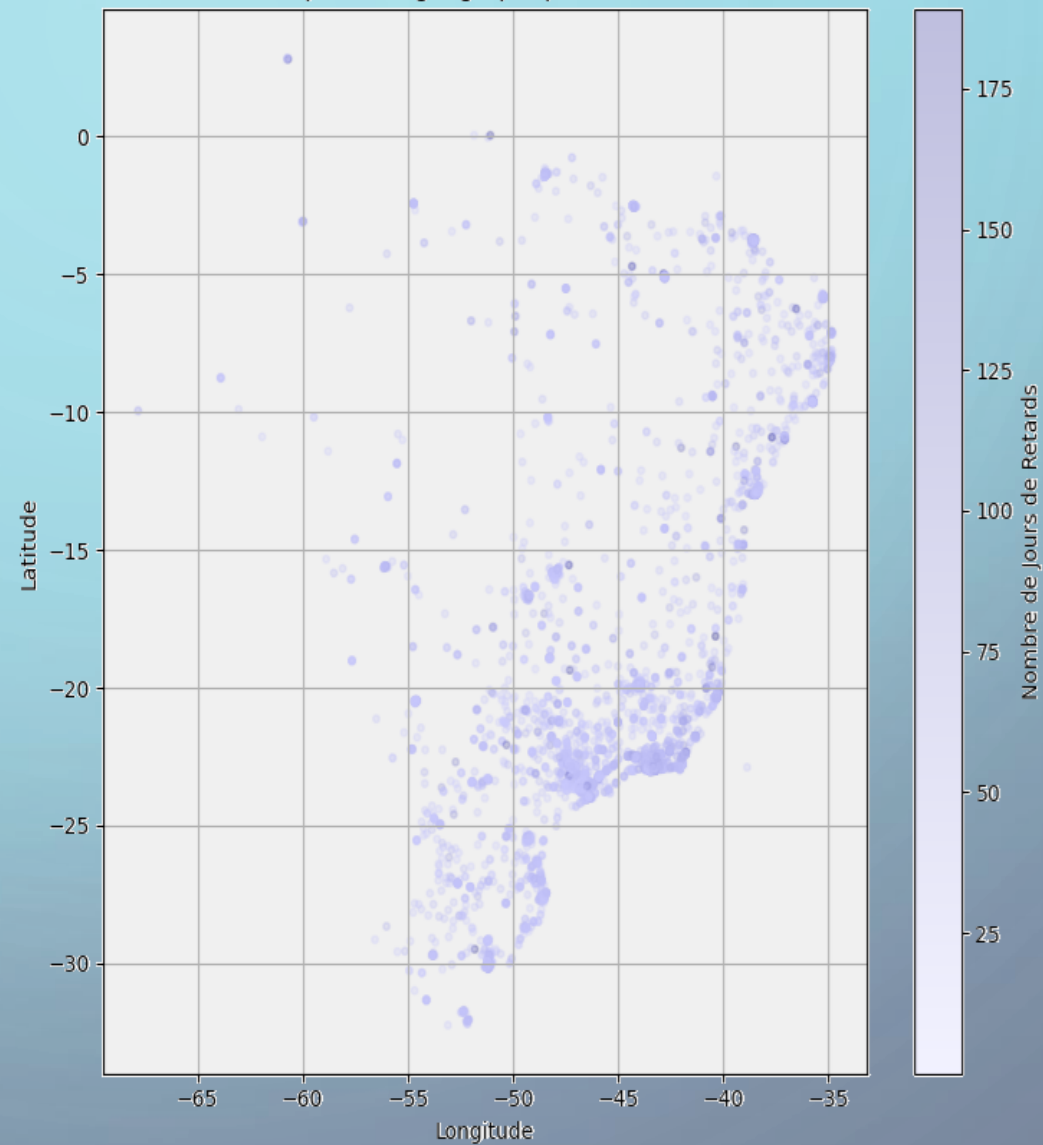


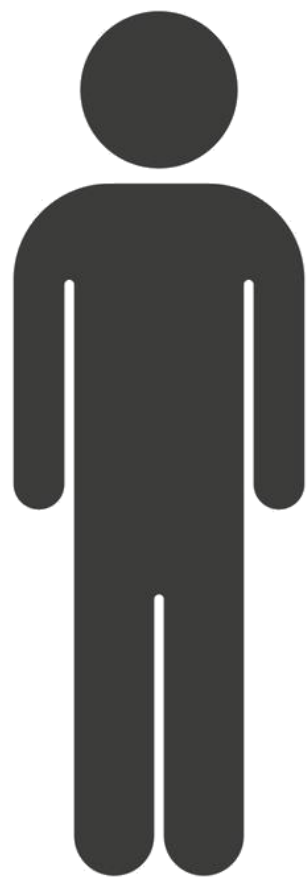
ANALYSE EXPLORATOIRE : REPARTITION GEOGRAPHIQUE

Répartition géographique / Délais de Livraison



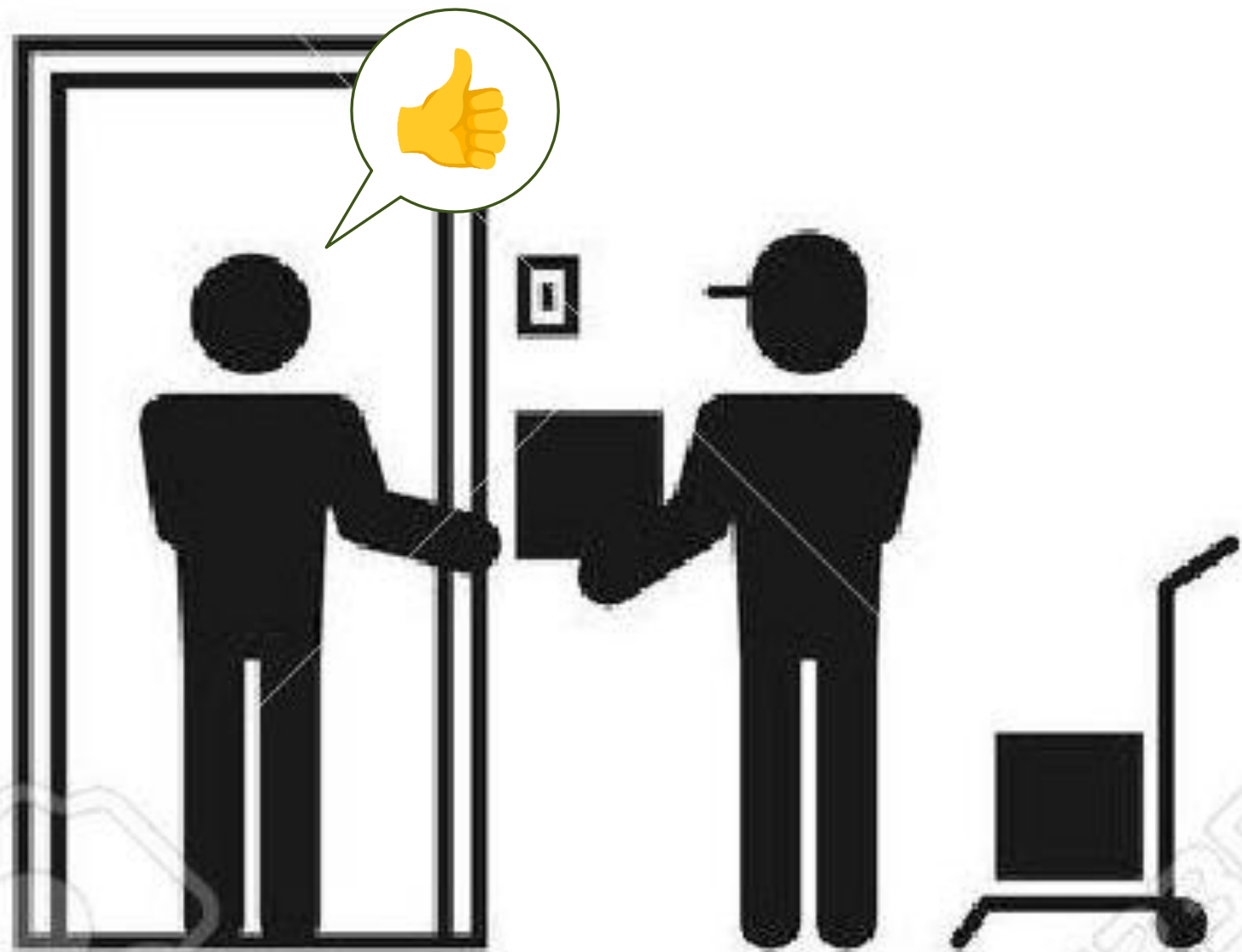
Répartition géographique / Retards

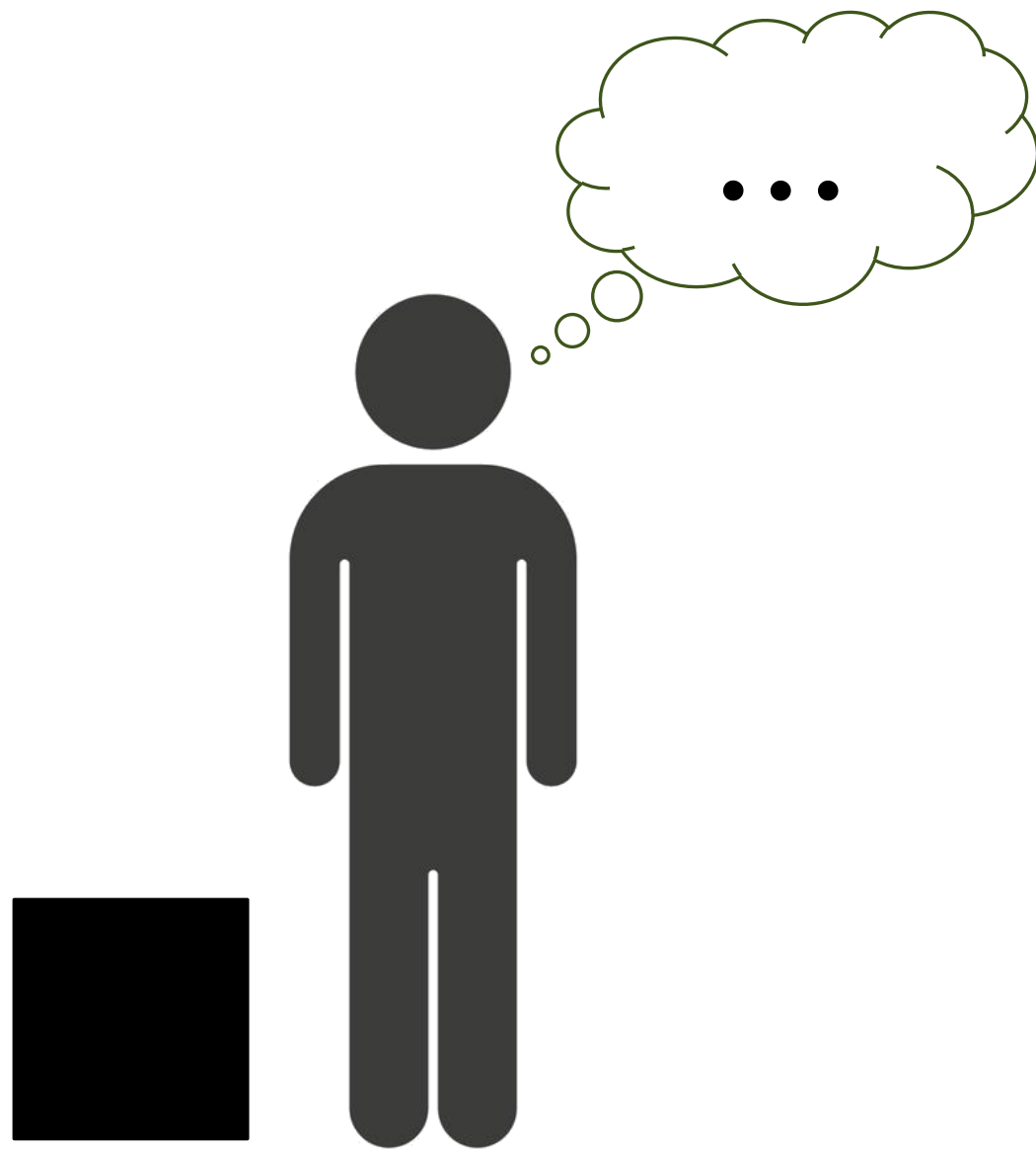






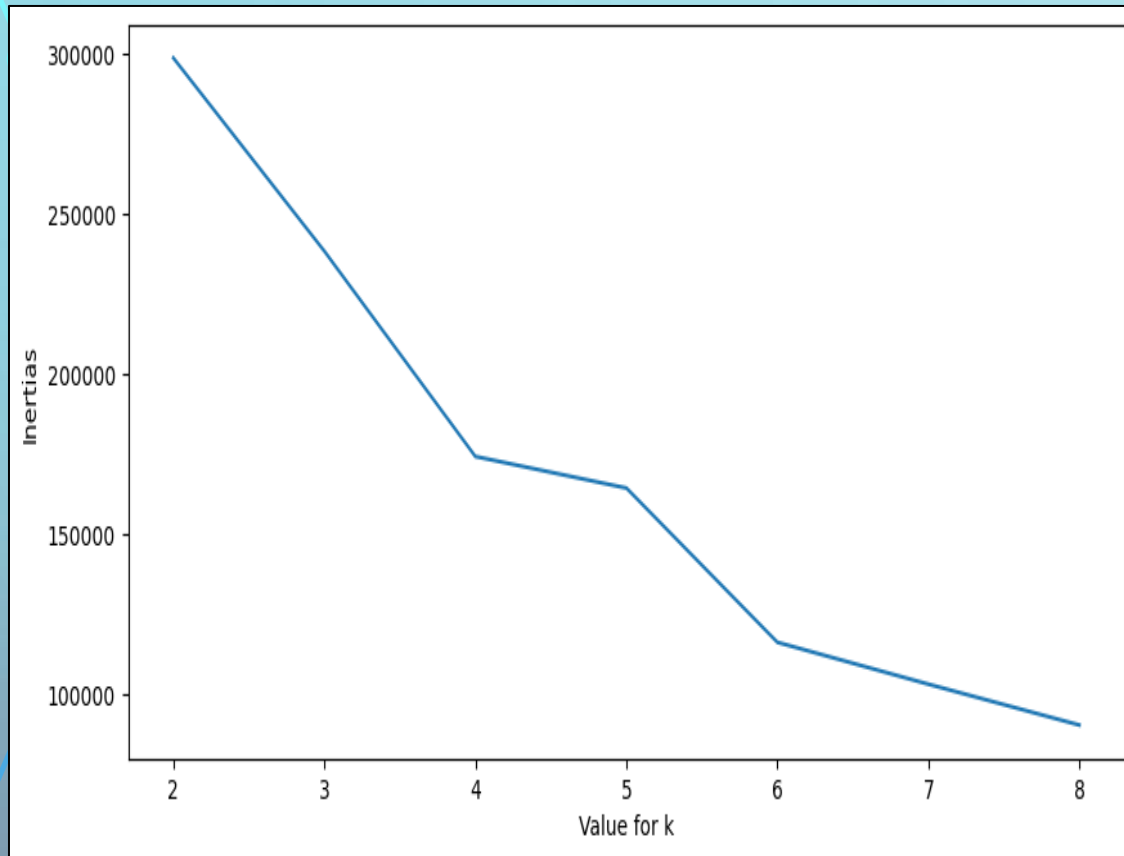
Toc! Toc!



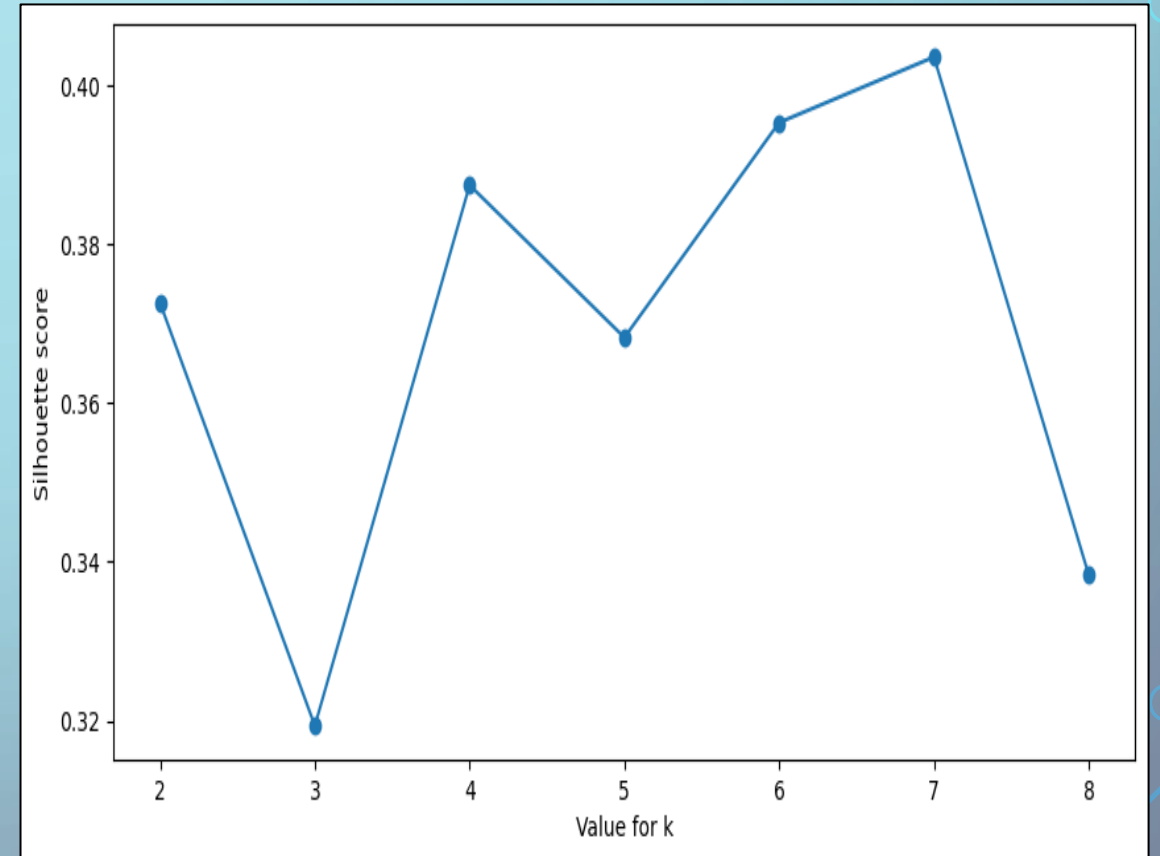


DEVELOPPEMENT DE MODELES : KMEANS

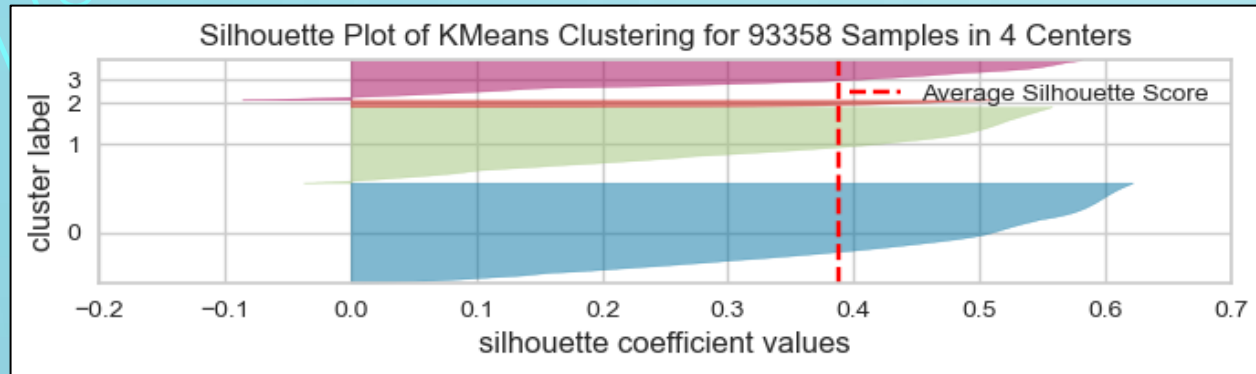
Méthode du Coude



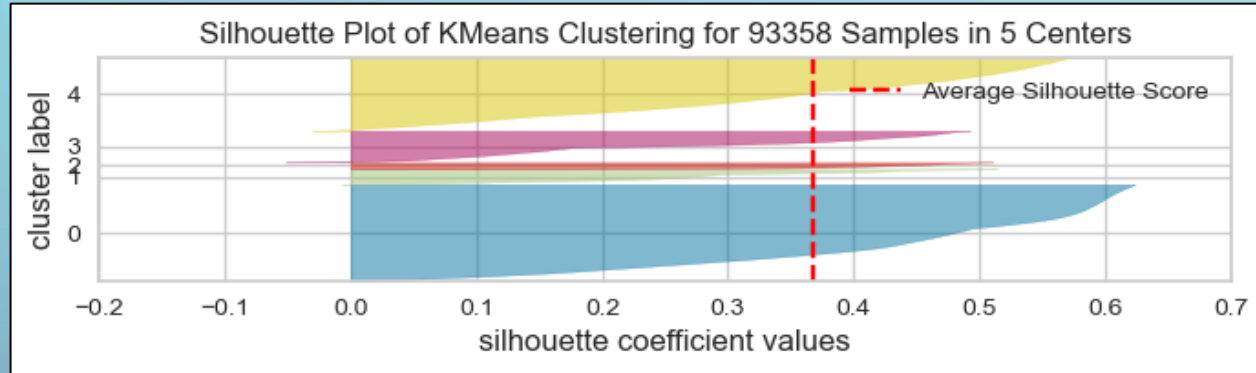
Silhouette Score par Nombre de Clusters



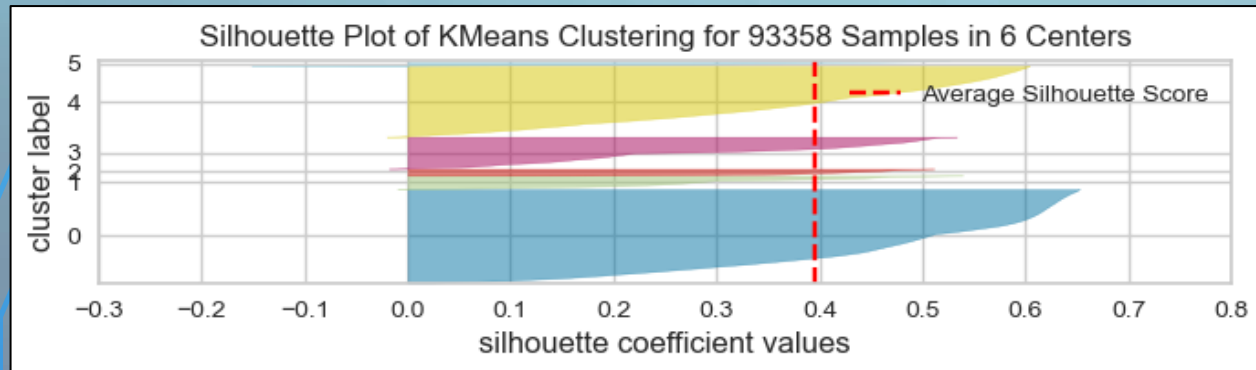
DEVELOPPEMENT DE MODELES : KMEANS



ARI Score Mediane (4 clusters) = 0.90 Moyenne = 0.82



ARI Score Mediane (5 clusters) = 0.95 Moyenne = 0.93

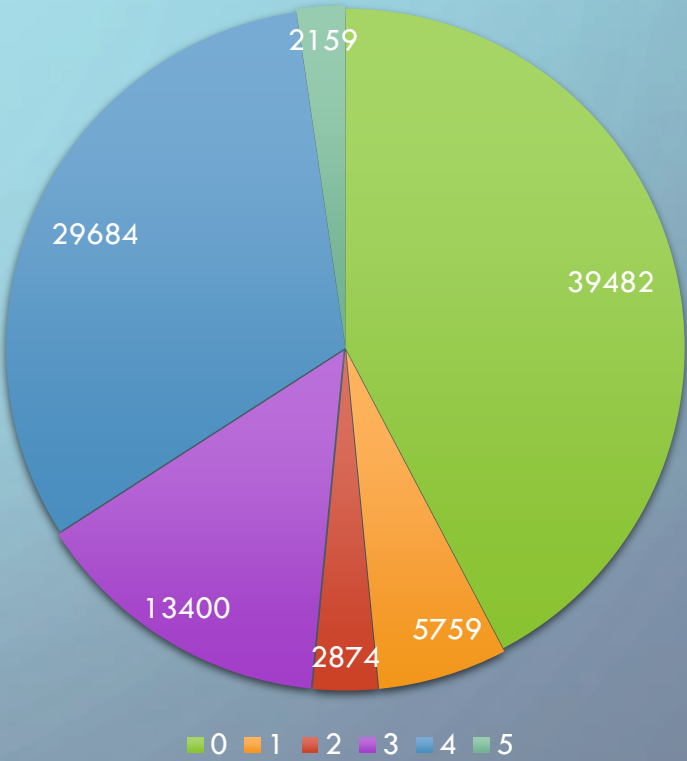


ARI Score Mediane (6 clusters) = 0.86 Moyenne = 0.78

SIMULATION DE MODELES : KMEANS

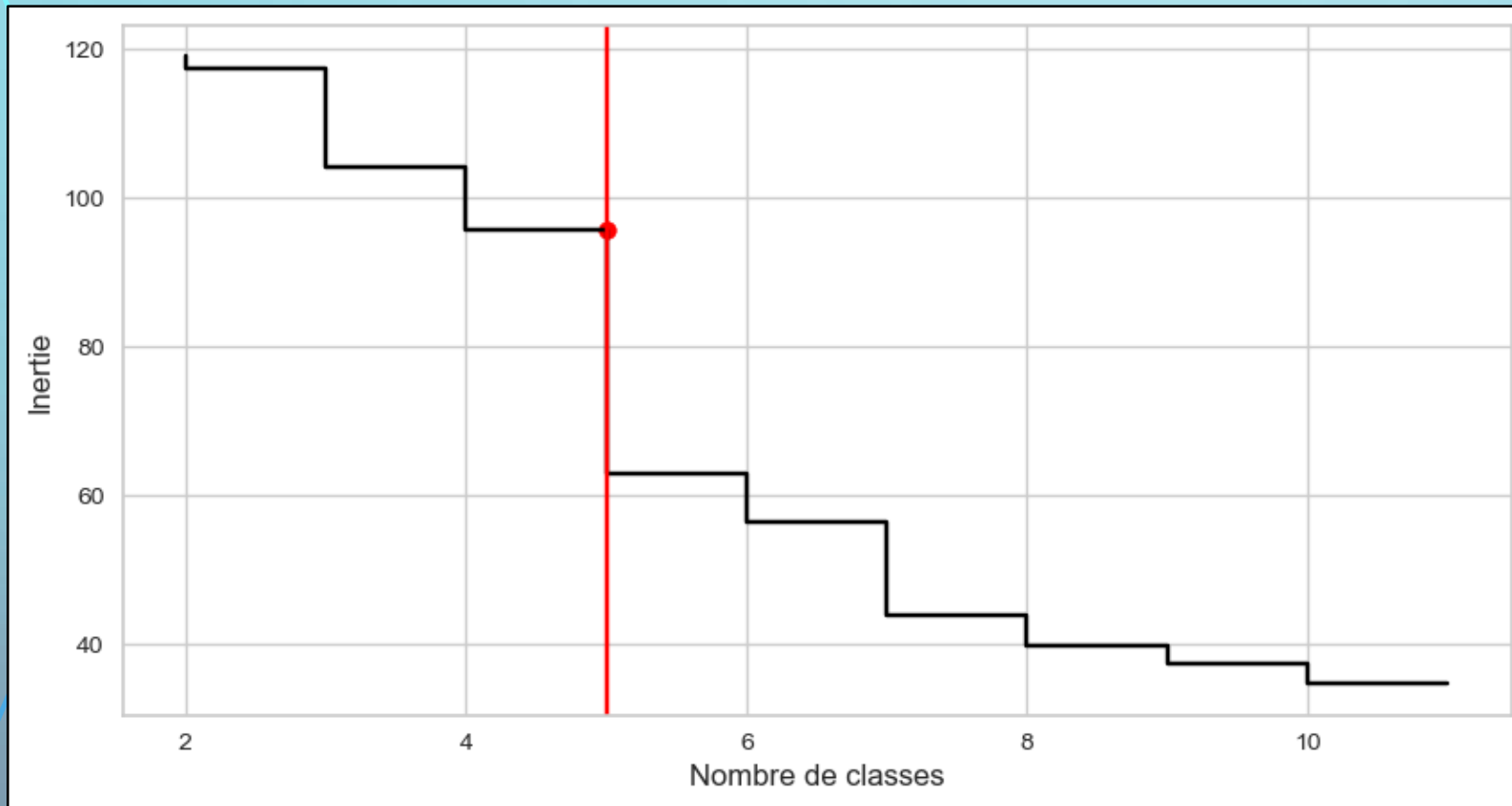
clusters	total purchase	delivery time	delay	shipping	total	average score	Latest purchase numeric
0	1	23	-12	21,12%	125,01BRL	4,75	172
1	1	25	-10	21,97%	126,45BRL	2,11	472
2	2	24	-12	21,59%	124,71BRL	4,19	270
3	1	24	-5	21,79%	134,91BRL	1,88	211
4	1	24	-13	20,91%	124,26BRL	4,74	436
5	1	26	-12	5,52%	1120,15BRL	4,19	289

REPARTITION DES CLUSTERS KMEANS



DEVELOPPEMENT DE MODELES : AGGLOMERATIVE CLUSTERING

Diagramme d'Inertie



Appliqué sur un sample de 10.000 clients

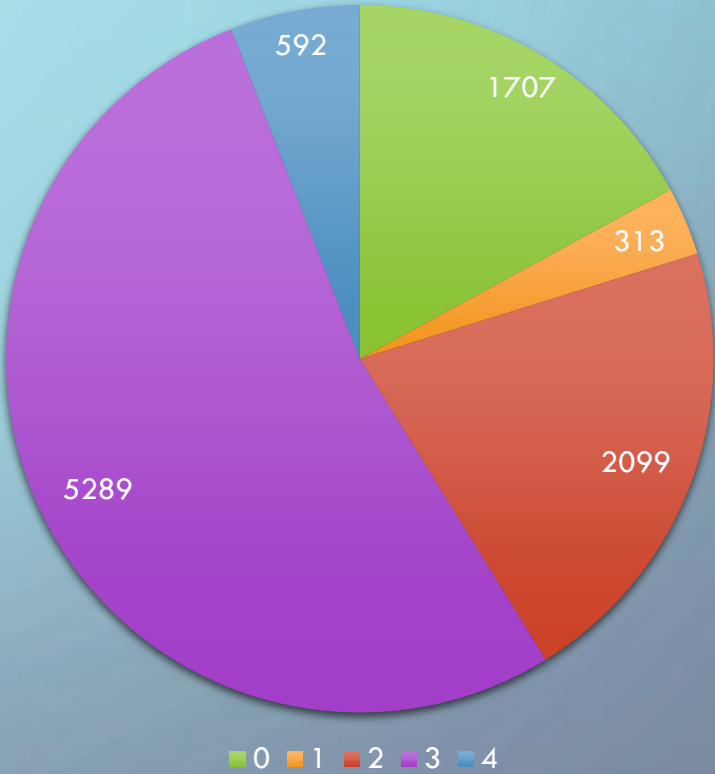
linkage = ward

→ minimise la variance totale au sein des clusters.

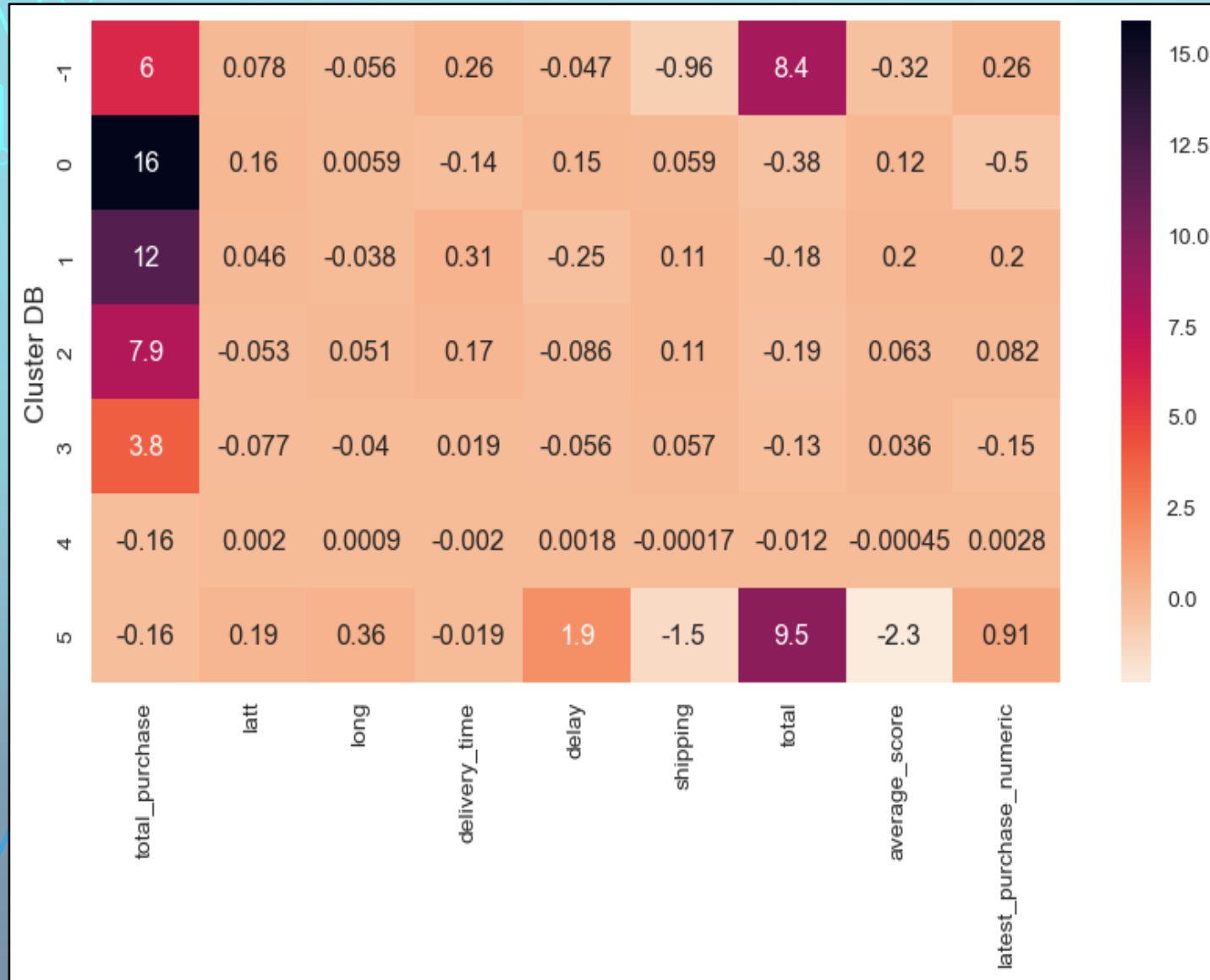
SIMULATION DE MODELES : AGGLOMERATIVE CLUSTERING

clusters	total purchase	delivery time	delay	shipping	total	average score	Latest purchase numeric
0	1	24	-6	22,41%	116,90BRL	1,79	257
1	2	24	-11	21,33%	128,54BRL	4,14	273
2	1	25	-14	21,93%	107,71BRL	4,53	498
3	1	23	-12	21,42%	114,49BRL	4,72	215
4	1	24	-12	7,53%	699,99BRL	4,41	235

REPARTITION DES CLUSTERS
AGGLOMERATIVE CLUSTERING



DEVELOPPEMENT & SIMULATION DE MODELES : DBSCAN



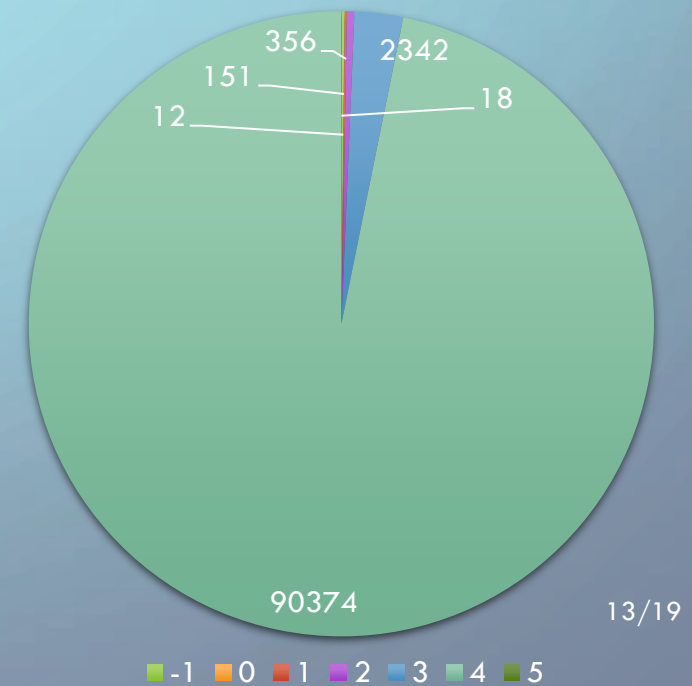
Paramétrage optimal :

Nb. Epsilon = 0,9 & minimum samples = 9

→ nb. Clusters = 6

→ nb. Outliers = 154

REPARTITION DES CLUSTERS DBSCAN



COMPARAISON DES MODELES & CHOIX DU MODELE FINAL

Modèle	+	-
K-Means	Simplicité Scalabilité Segmentation Claire	Sensibilité aux Outliers Forme des Clusters
Agglomerative Clustering	Visualisation Flexibilité	Complexité Pas de Réaffectation
DBSCAN	Robustesse 0 Pré-Requis	Paramétrage Complexe Scalabilité Limitée

Choix Final : K-Means

- **Efficacité sur Données Standardisées**
- **Simplicité & Performance** : Fournit une segmentation claire avec une bonne réactivité aux changements de comportement client.
- **Réplicabilité** : Facilite la mise à jour régulière des clusters, conforme aux besoins d'adaptation du modèle.

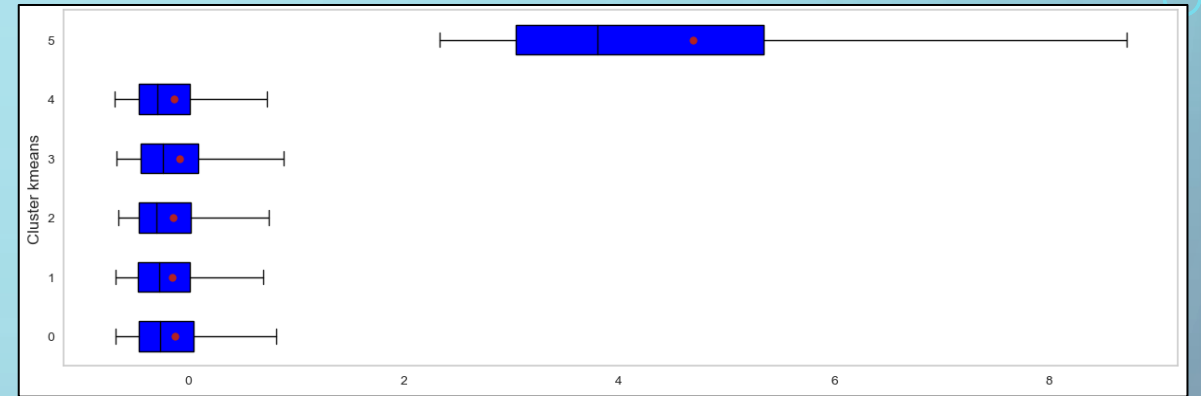
CARACTERISTIQUES DES CLUSTERS

Distribution des Valeurs par Clusters K-Means

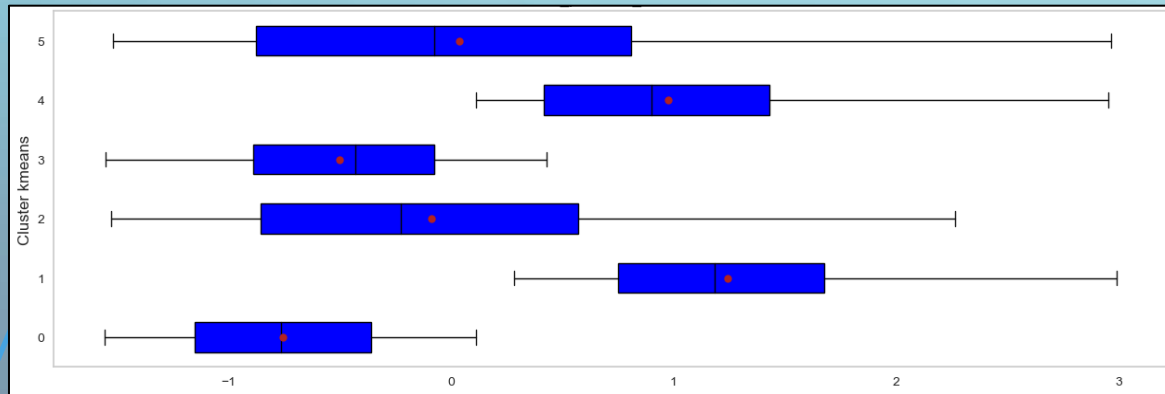
Nombre total d'achats/client



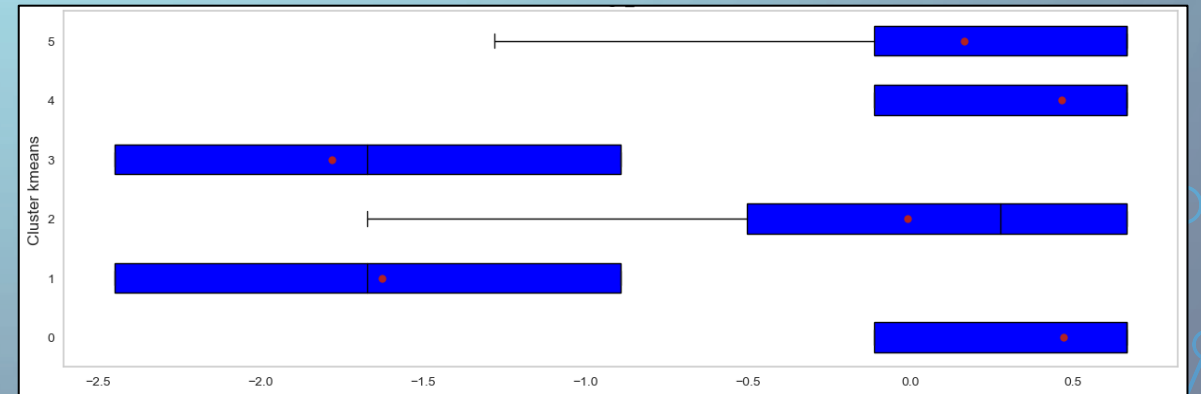
Montant total des achats/client



























Niveau d'ancienneté/client



Score attribué/client

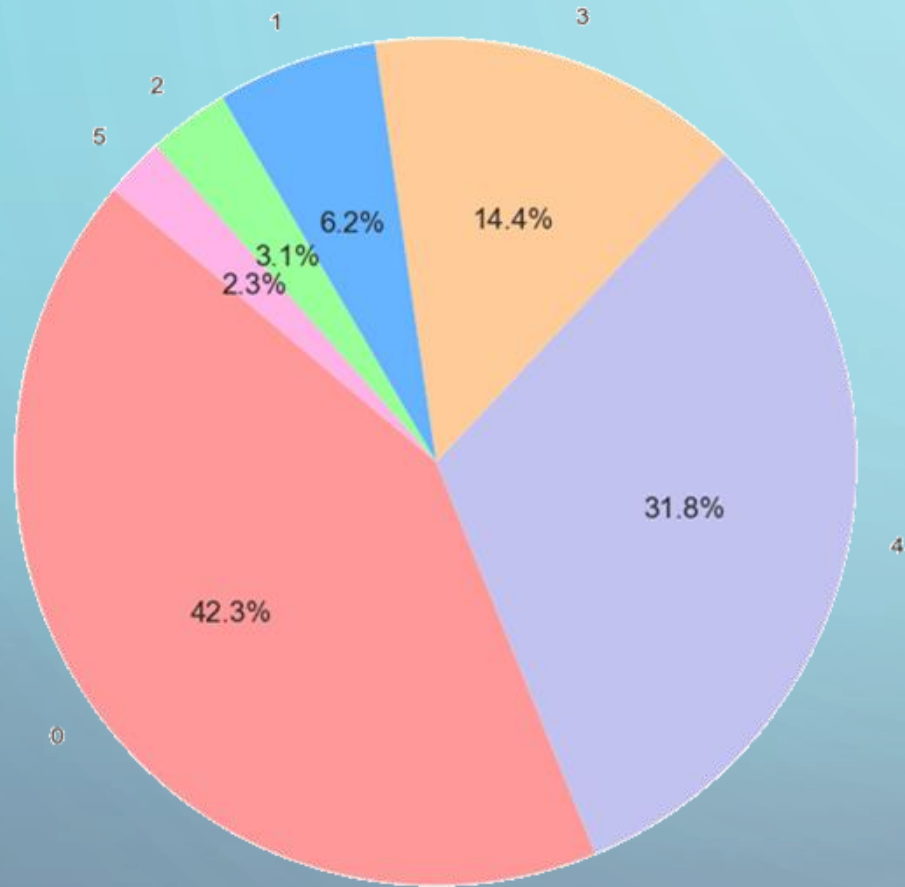


DESCRIPTION ACTIONNABLE DES CLUSTERS

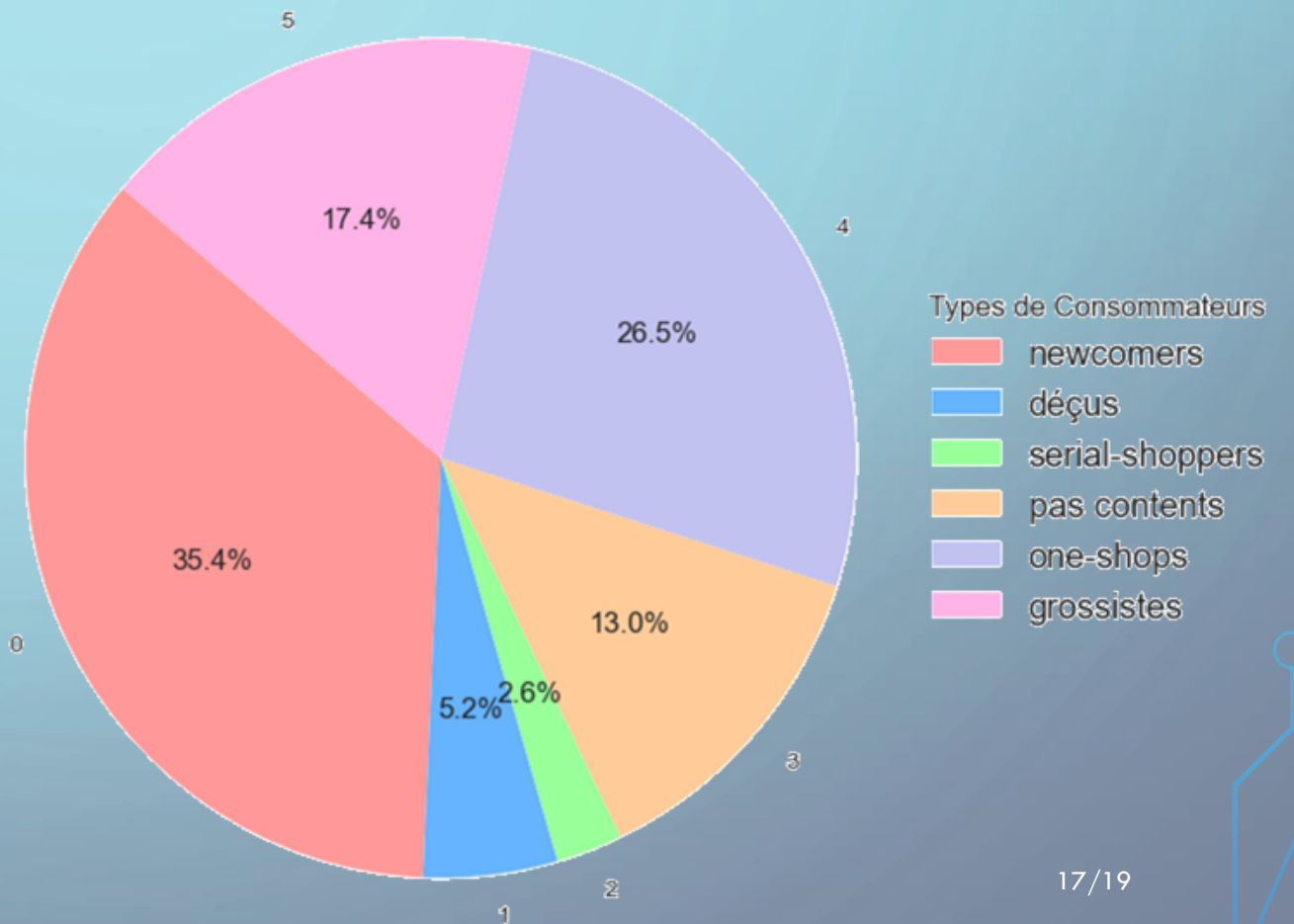
CLUSTERS	Nombre d'Achats	Montant Total	Ancienneté	Note Attribuée
0. Les Newcomers				
1. Les Déçus				
2. Les Serial-Shoppers				
3. Les Pas Contents				
4. Les One-Shop				
5. Les Grossistes				

REPARTITION DES CLUSTERS

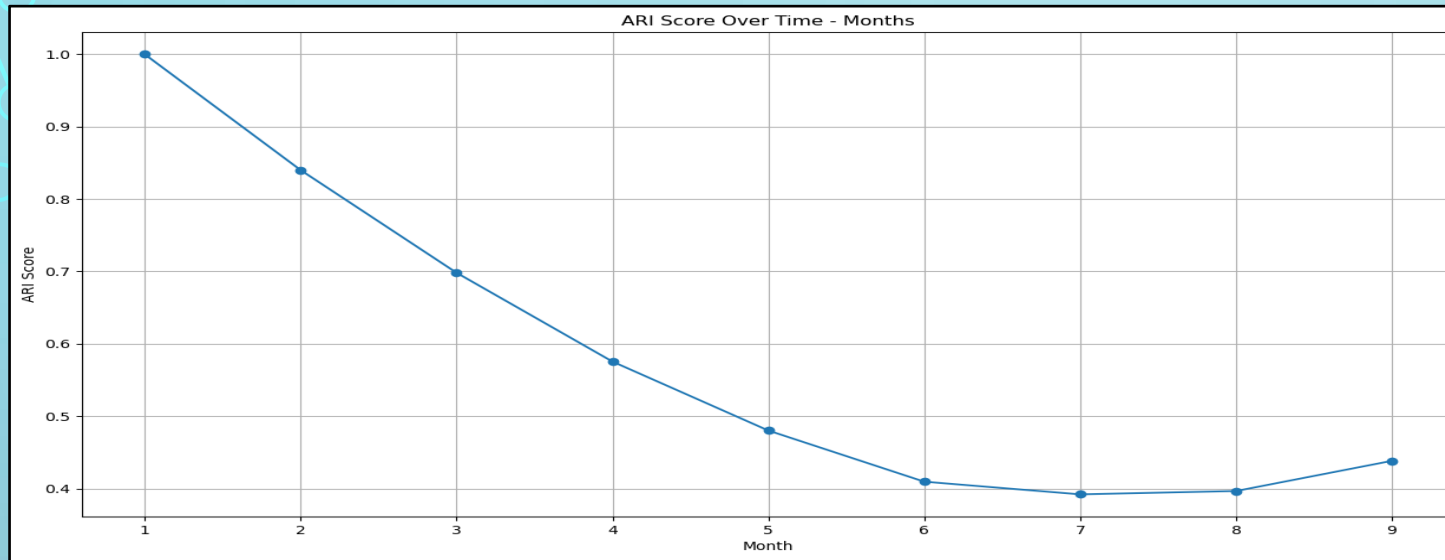
Répartition des Clients par Catégorie de Consommateur - KMeans



Répartition des Clients par Catégorie de Consommateur - Chiffre d'Affaires

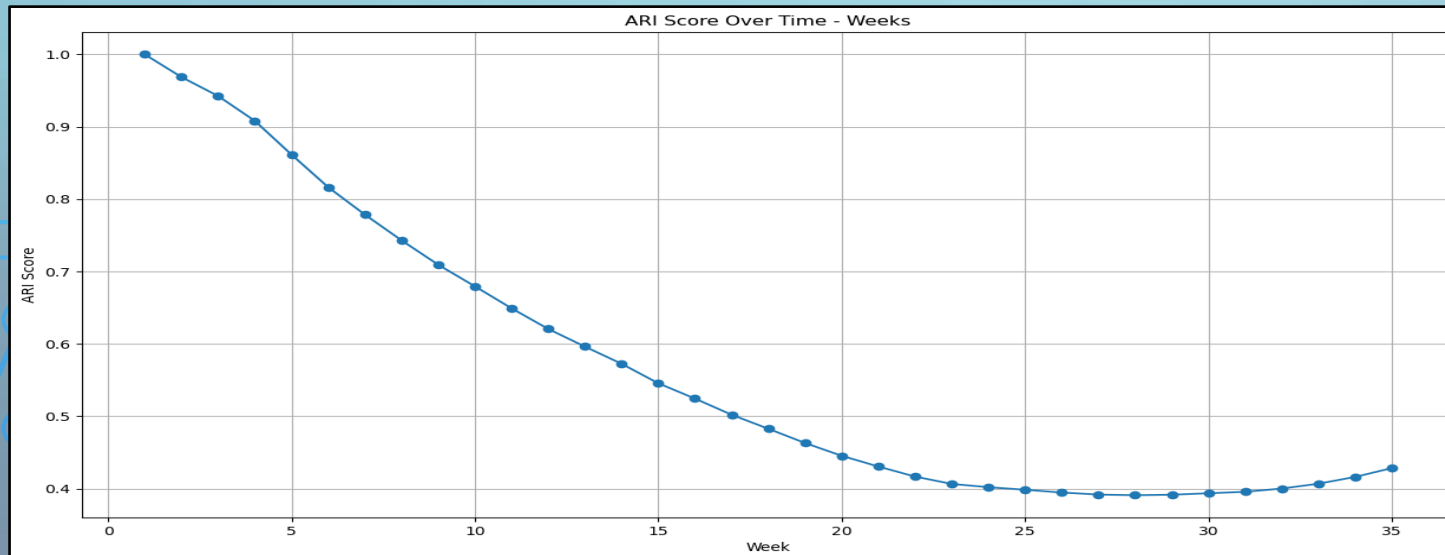


PROPOSITION DE CONTRAT DE MAINTENANCE



Comparaison Mensuelle & Hebdomadaire : Resegmentation basé sur les mêmes critères.

Suivi Chronologique : Indicateur de la stabilité des clusters via l'ARI Score



Finalité : détecter les changements de comportements & mesurer le rythme de ces changements

Résultat : Mise à jour tous les 1 mois et demi recommandée.

CONCLUSION


Segmentation Kmeans

6 Clusters → 6 Stratégies Marketing

- Newcomers & Grossistes : fidélisation & rétention
- One-Shop : recherche de l'origine des non-renouvellements d'achats
- Serial-Shoppers : fidélisation
- Déçus : recherche de l'origine des insatisfactions de la clientèle
- Pas-Contents : priviligélisation & exclusivité

Pistes d'évolution : caractéristiques géographiques



The background is a light blue gradient. In the corners, there are decorative circuit-like patterns made of thin blue lines and small circles, resembling a stylized electronic board.

Merci pour votre attention