

# Segmentation des clients du site Olist

olist



Moussa KIBALY



# SOMMAIRE

1. CONTEXTE DU PROJET
2. ANALYSE EXPLORATOIRE ET TRANSFORMATION DES DONNÉES
3. SYNTHÈSE DES DIFFÉRENTS MODÈLES
4. CONCLUSION

# 1. CONTEXTE DU PROJET

Le société de e-commerce **Olist** souhaite connaître les différents types de client de leur site ainsi que leur comportement et habitudes afin d'effectuer des campagnes marketing ciblées.

Pour cela, le jeu de données comprenant les commandes effectuées sur le site ainsi que le fichier des clients et des paiements serviront à effectuer une segmentation des clients.







## 2. ANALYSE EXPLORATOIRE ET TRANSFORMATION DES DONNÉES

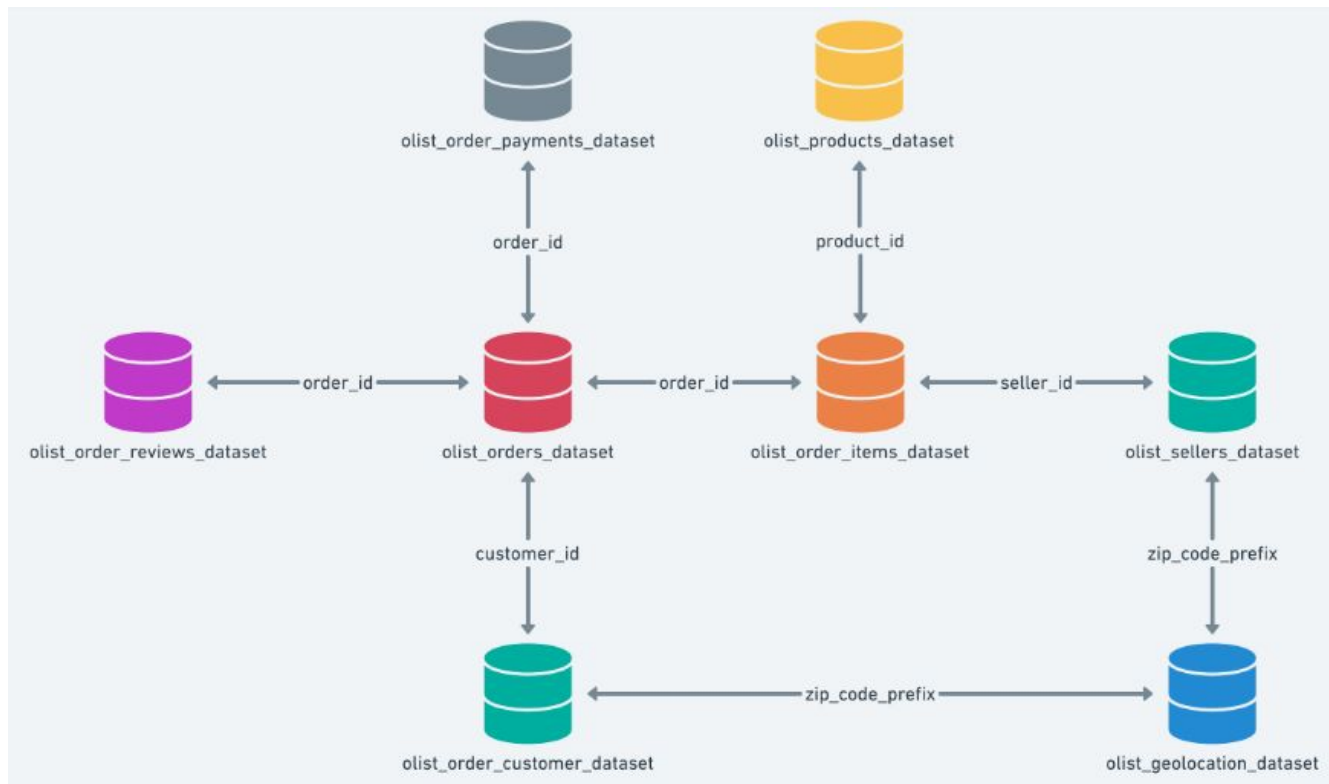
Le jeu de données comprend 8 fichiers à savoir le fichier des clients, des commandes, des paiements, du détail des commandes, des commentaires client, des vendeurs, un fichier sur la géolocalisation et un fichier sur les produits vendus sur le site.

Le fichier qui a servi à la segmentation vient de la fusion des 6 fichiers suivants :

fichier des commandes, des paiements, des clients, des détails des commandes, des commentaires, des produits et des vendeurs.

Les dates du fichier fusionnée sont transformées en type Datetime, les variables catégorielles en variables numériques à l'aide du One hot encoding ou Label encoding.

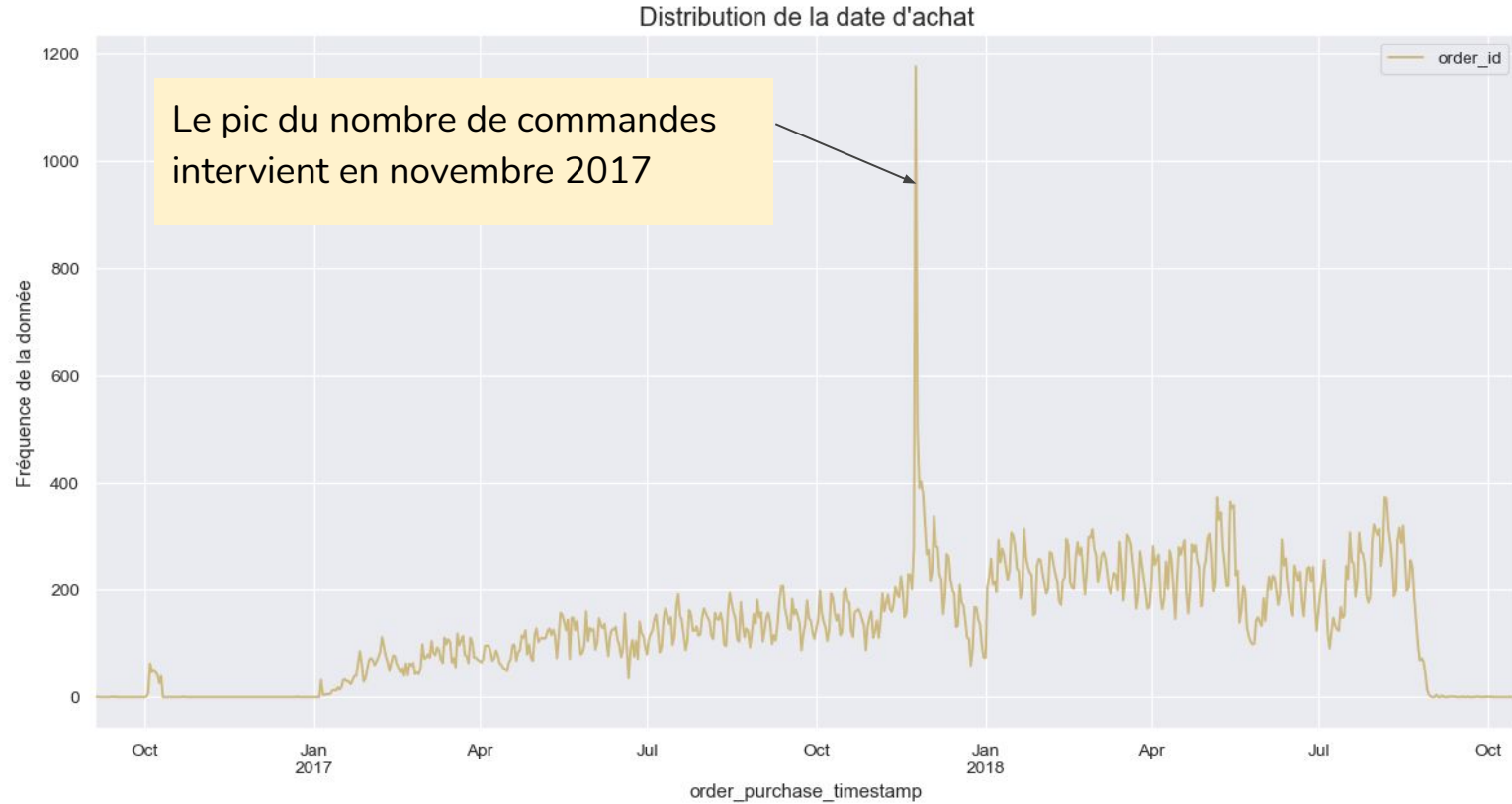
## Base de données du site Olist



## Répartition de l'état de la commande

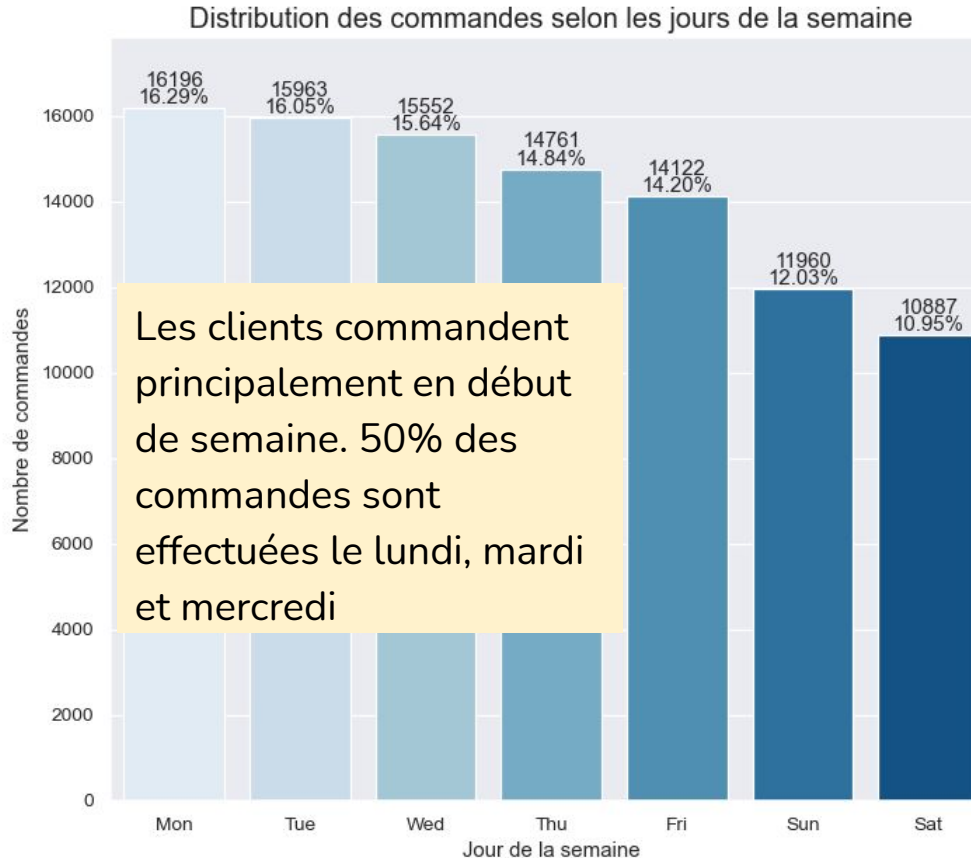


## Fréquence d'achat en fonction du temps



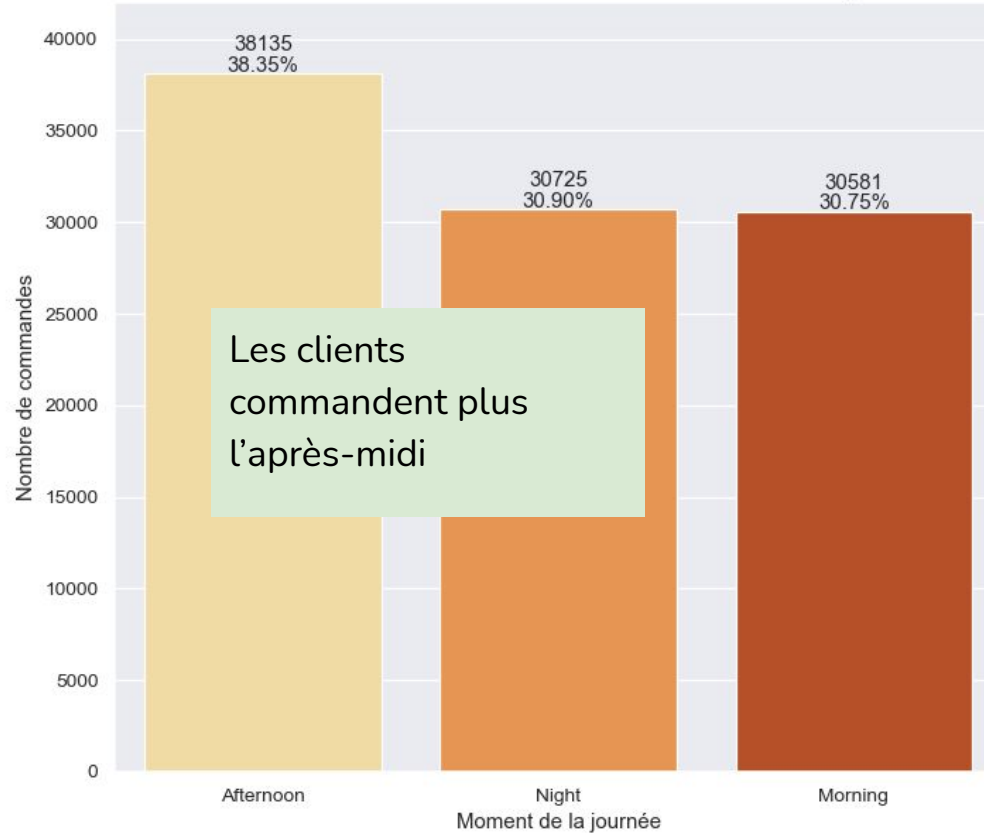


## Nombre de commandes au cours de la semaine

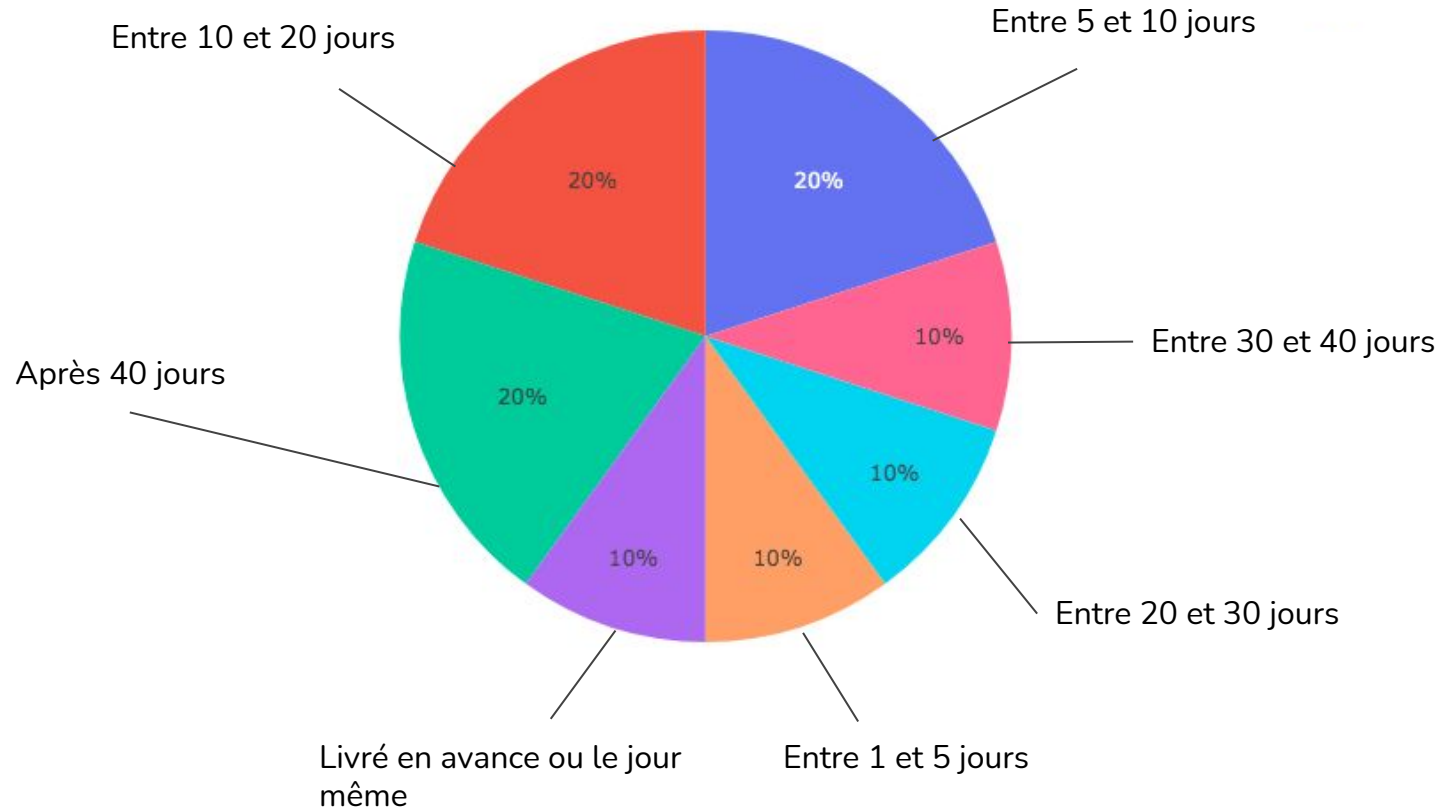


# Nombre de commandes au cours de la journée

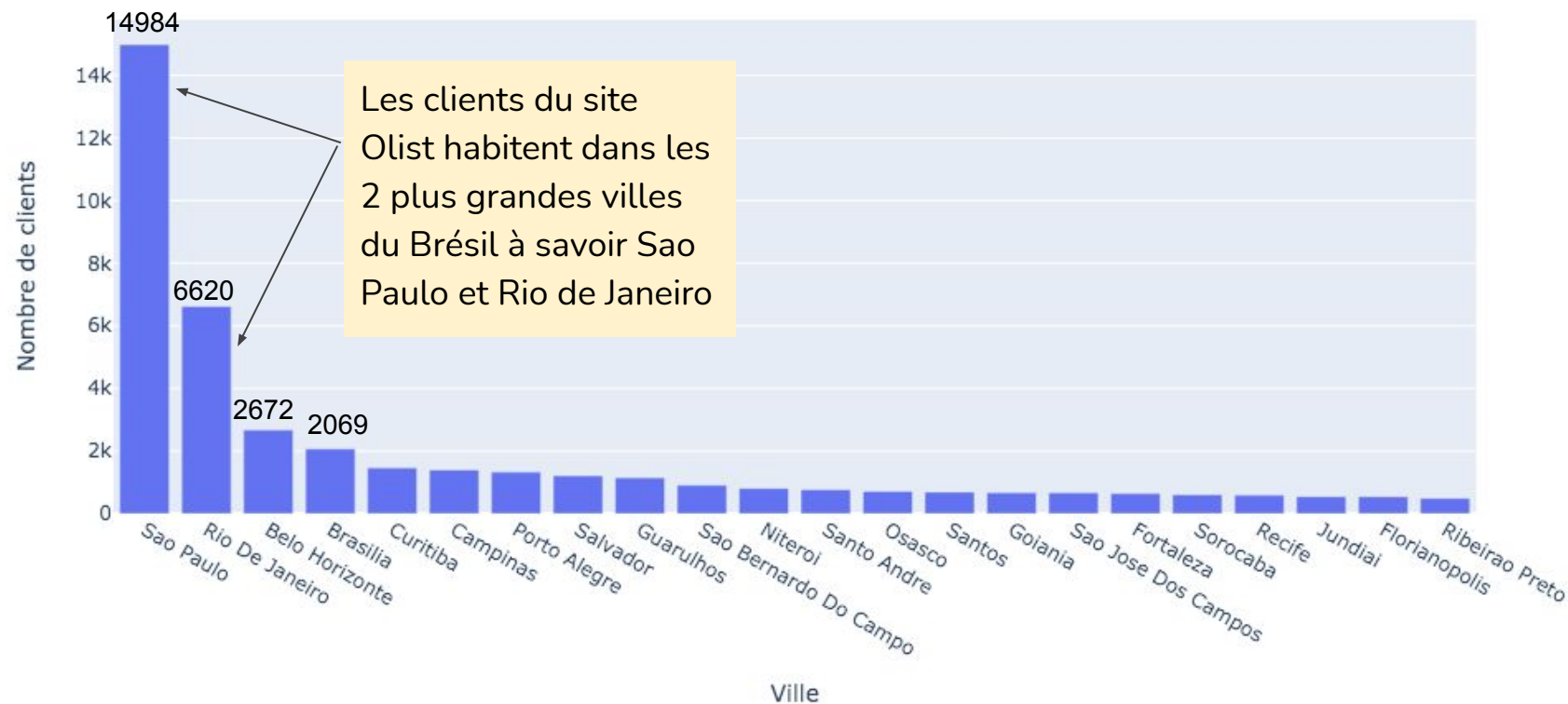
Distribution des commandes selon les moments de la journée



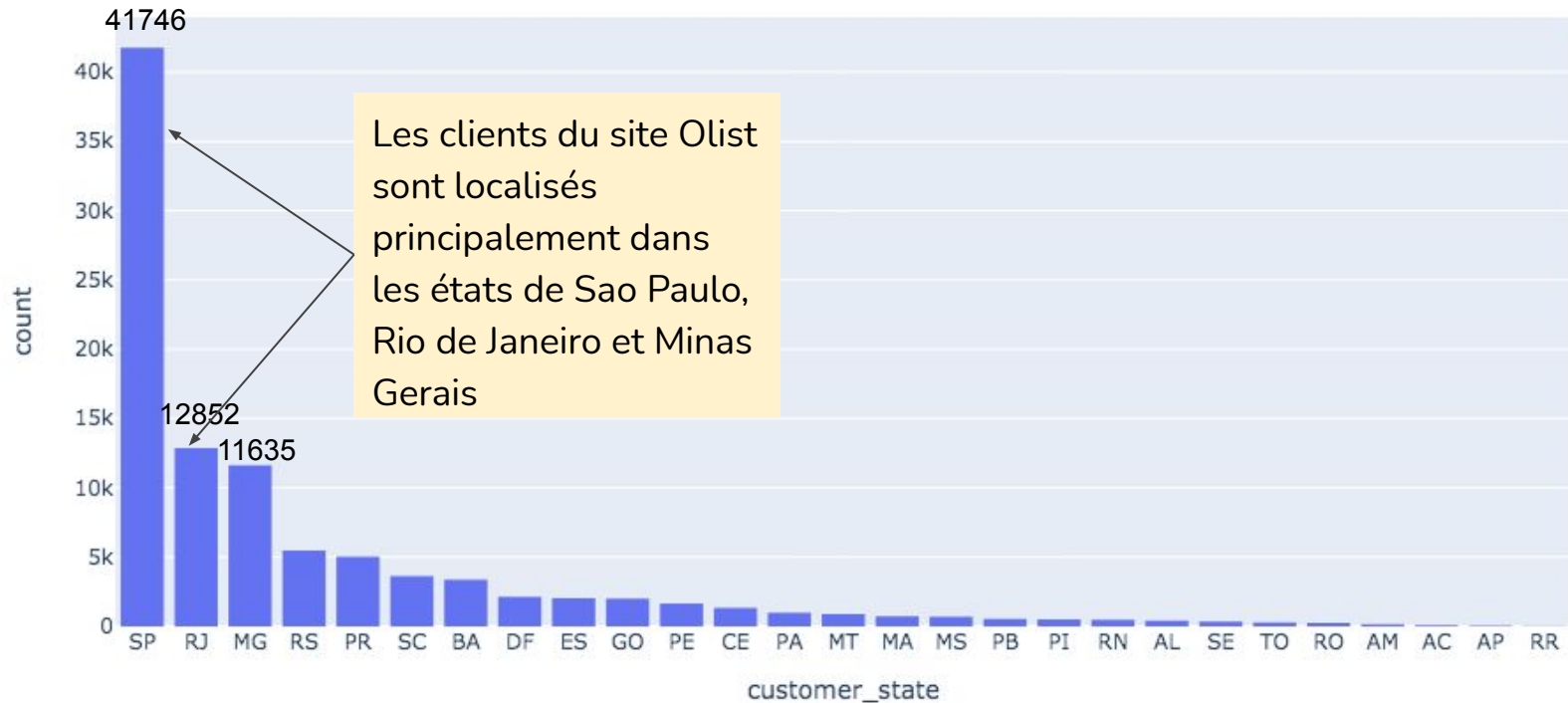
## Durée de livraison en nombre de jours



## Répartition des clients par ville du Brésil

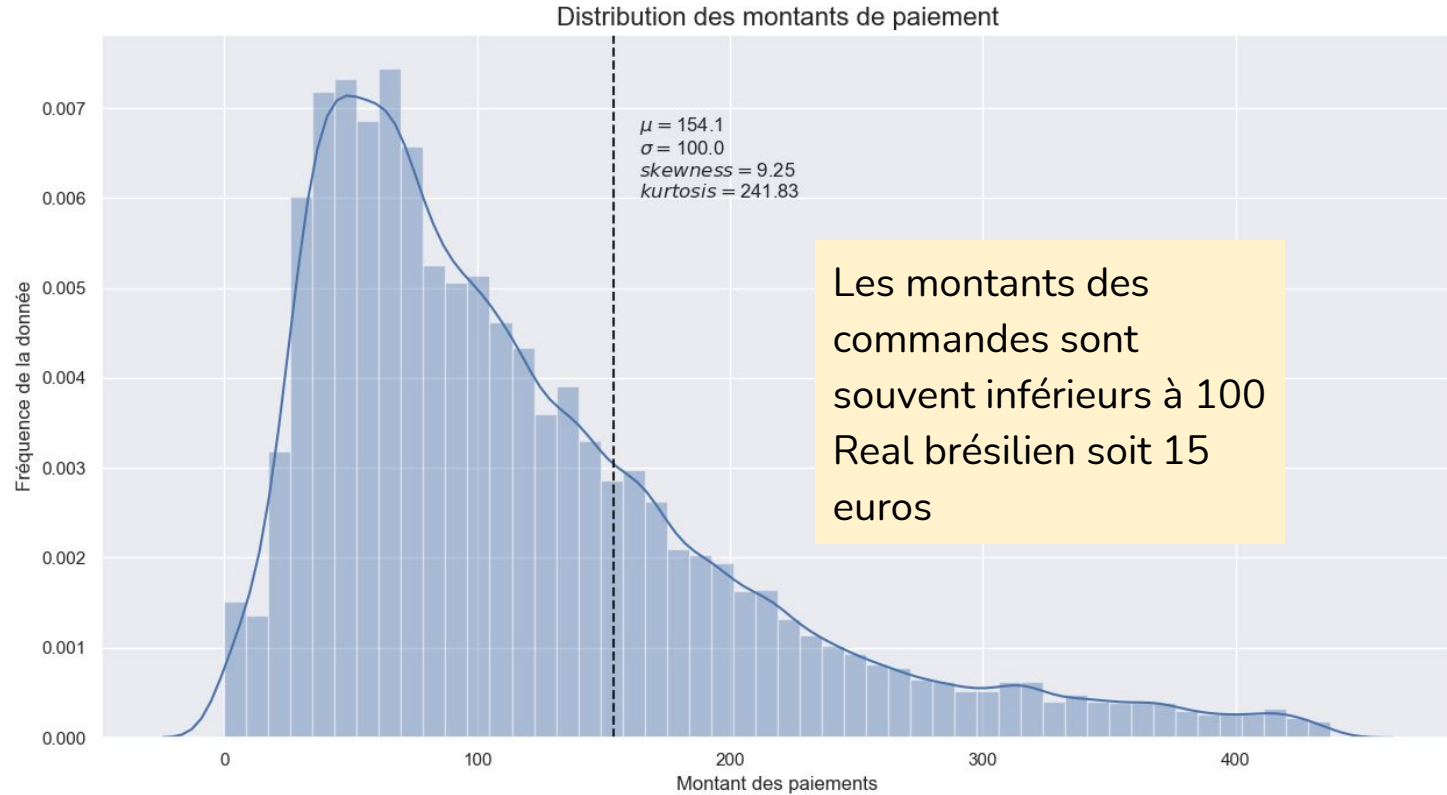


## Répartition des commandes par état du Brésil

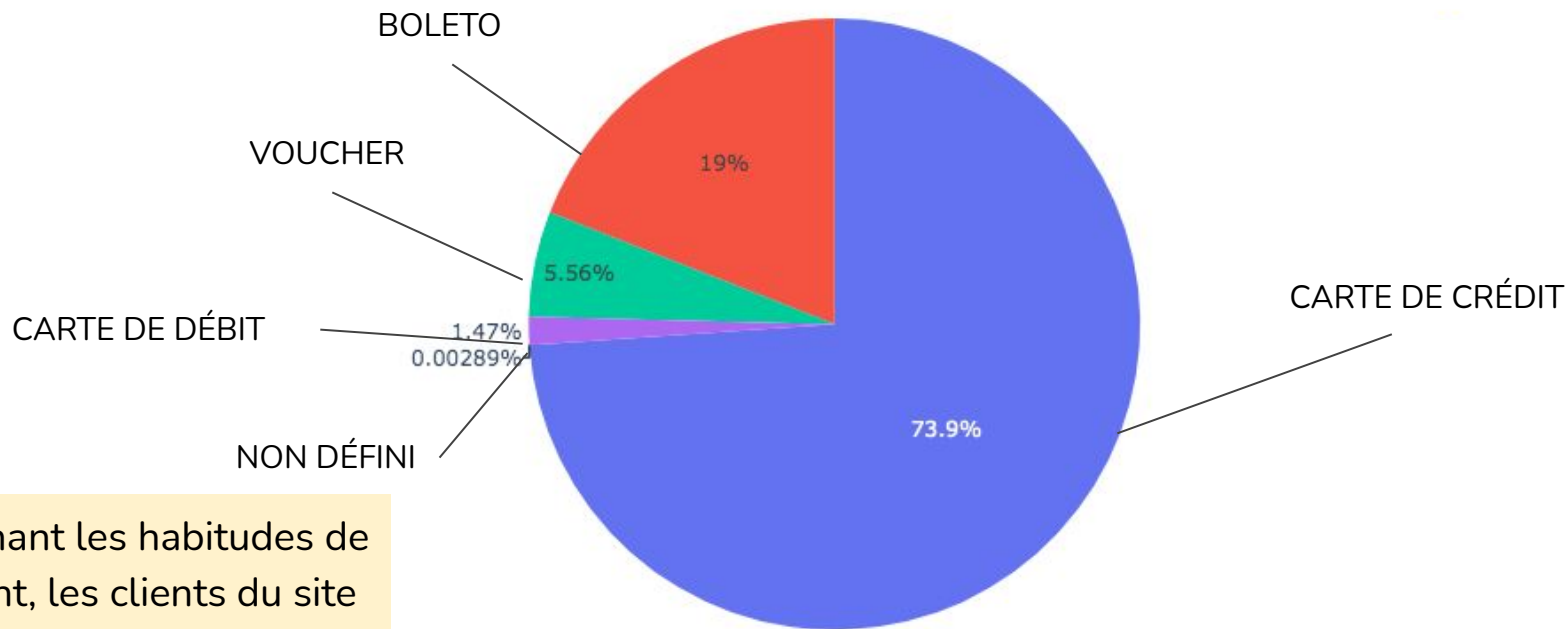




## Répartition du montant de paiement sur le site

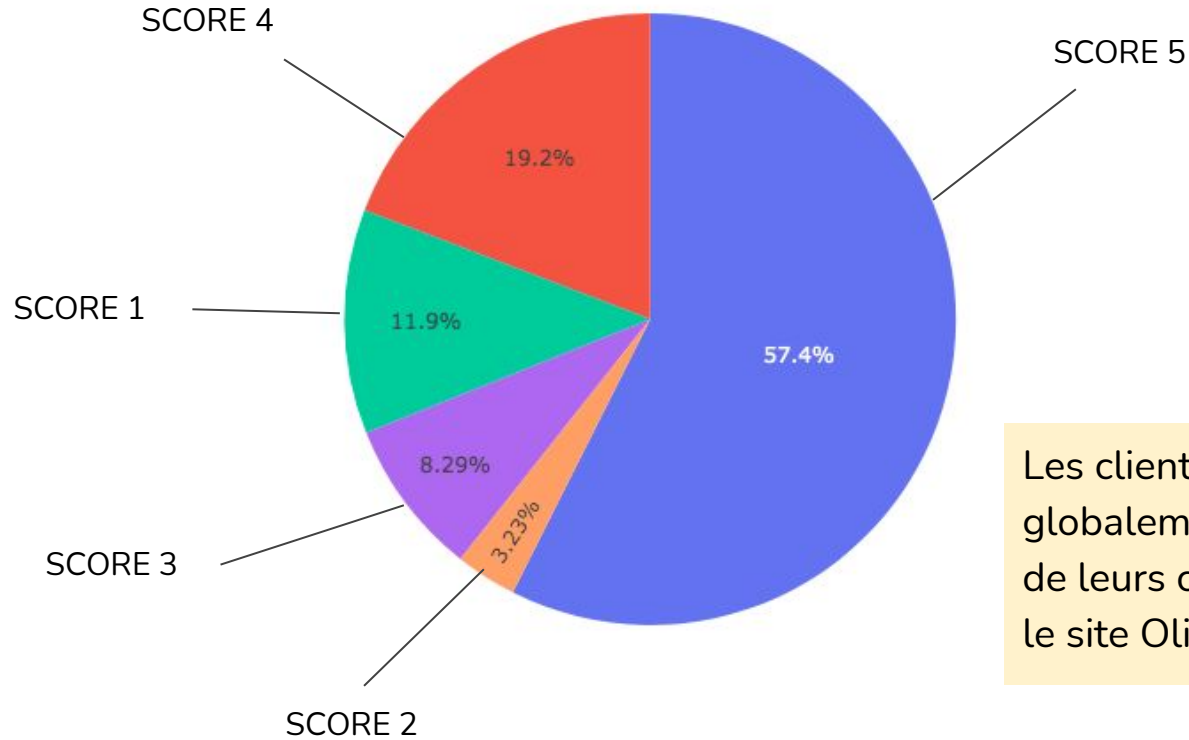


Distribution des types de paiement



Concernant les habitudes de paiement, les clients du site paient avec une carte de crédit ou le boleto qui est un moyen de paiement en espèces

## Distribution des scores des commandes

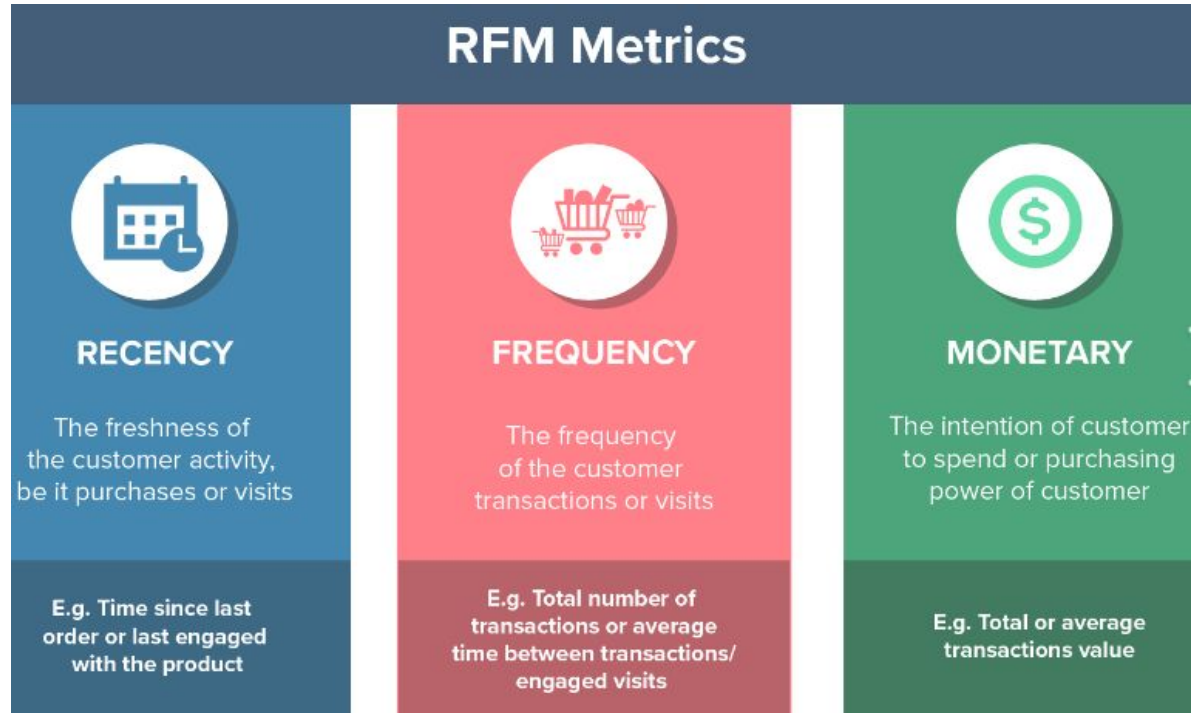


Les clients sont globalement satisfaits de leurs commandes sur le site Olist (score 4 et 5)

### 3. SYNTHÈSE DES DIFFÉRENTS MODÈLES

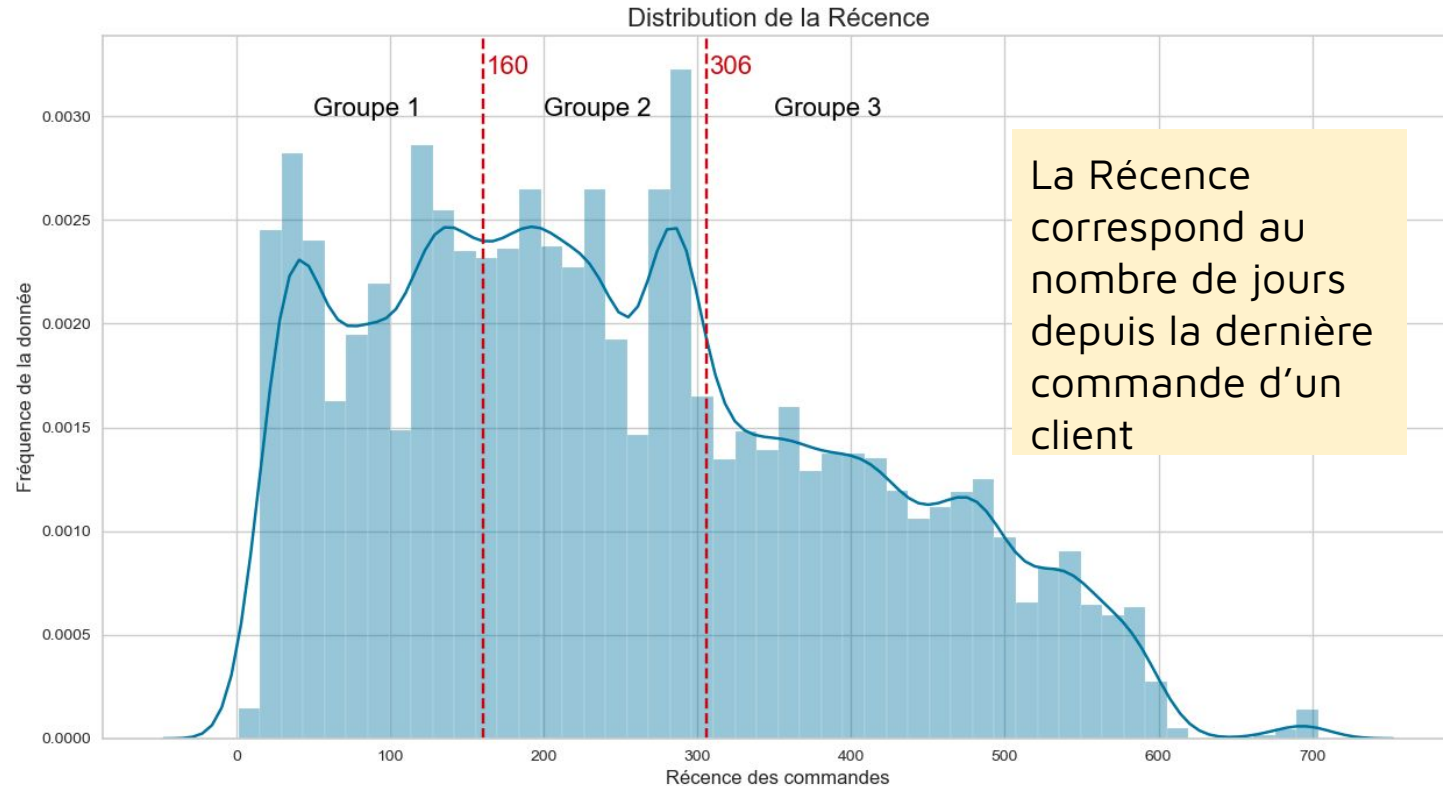


# SEGMENTATION MÉTIER RFM





## Représentation graphique de la Récence

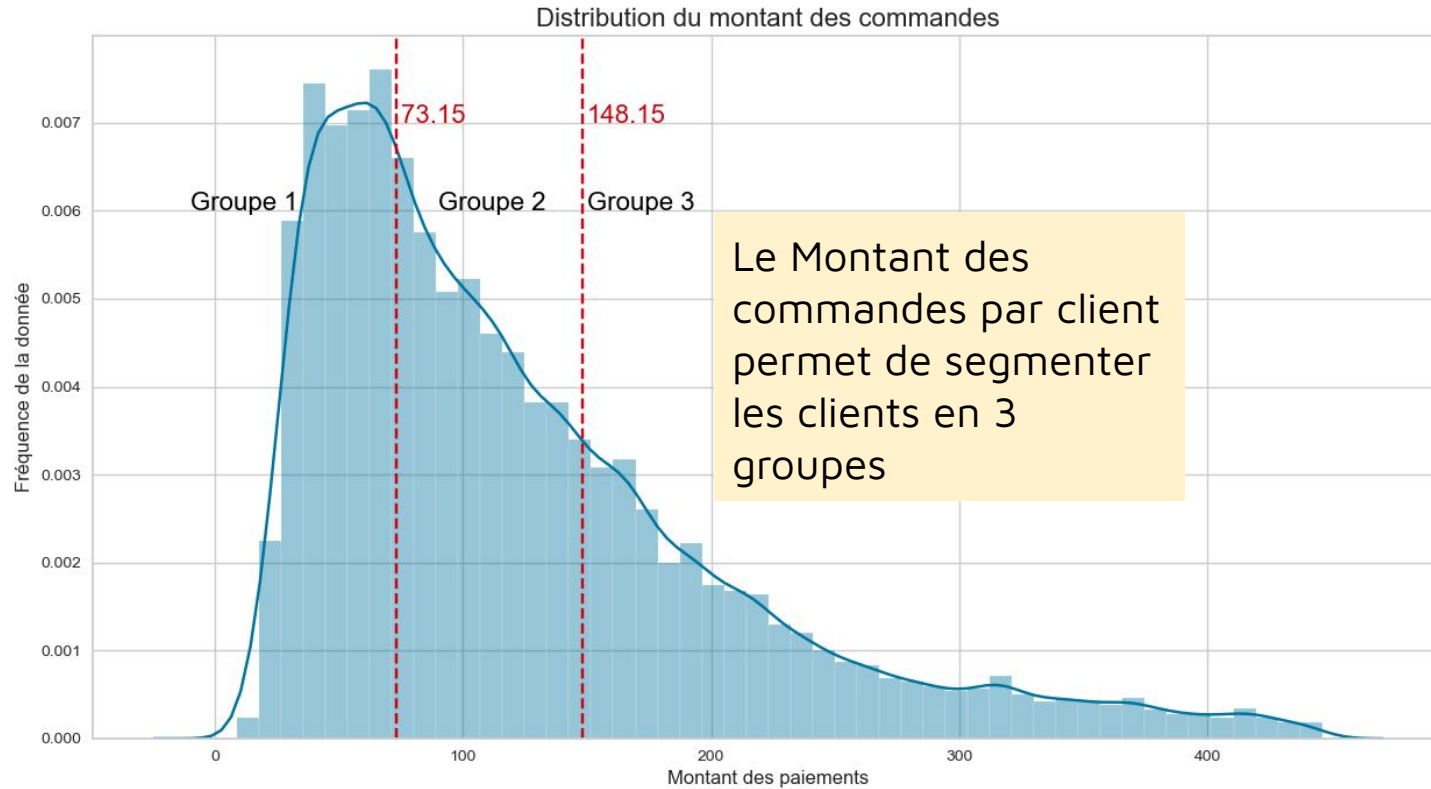


## Fréquence des commandes par client

Nombre d'achats	Effectif	Pourcentage
1	94753	97.13
2	2339	2.39
3	296	0.3
4	106	0.1
5	52	0.05

99% de la Fréquence des commandes est inférieure à 5 dont 97% correspond à un achat

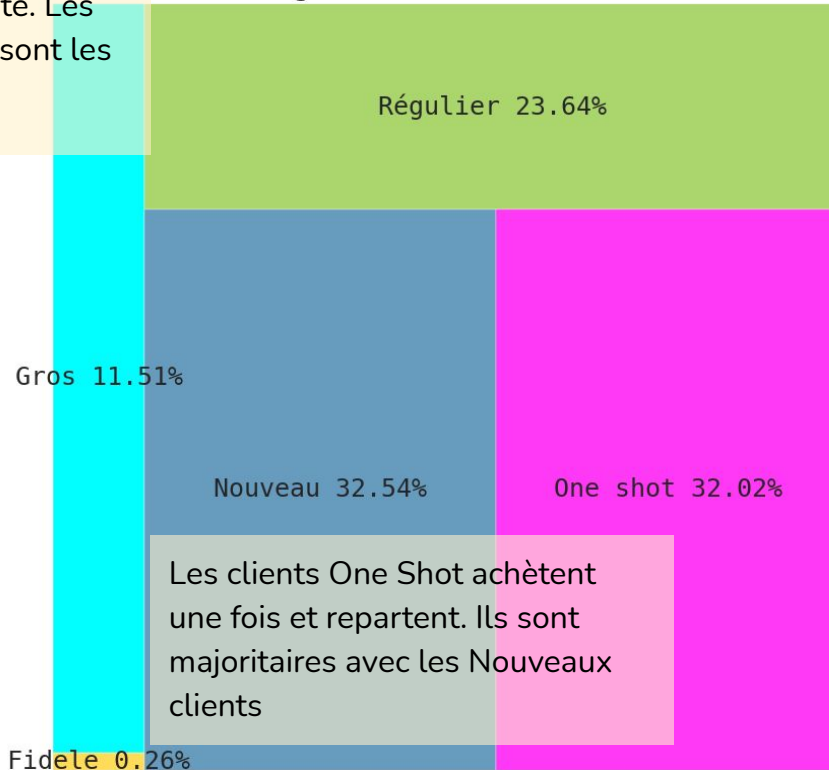
## Représentation graphique du Montant



## Définition des différents types de client Olist

Les Gros clients sont ceux générant le plus de revenus sur le site. Les clients Régulier sont les plus loyaux

### Segments RFM des clients



Les clients One Shot achètent une fois et repartent. Ils sont majoritaires avec les Nouveaux clients

### Définition de différents segments clients :

Fidèle - '123'

One shot - '311', '312', '313'

Nouveau - '111', '112', '113'

Gros - '323', '213', '223'

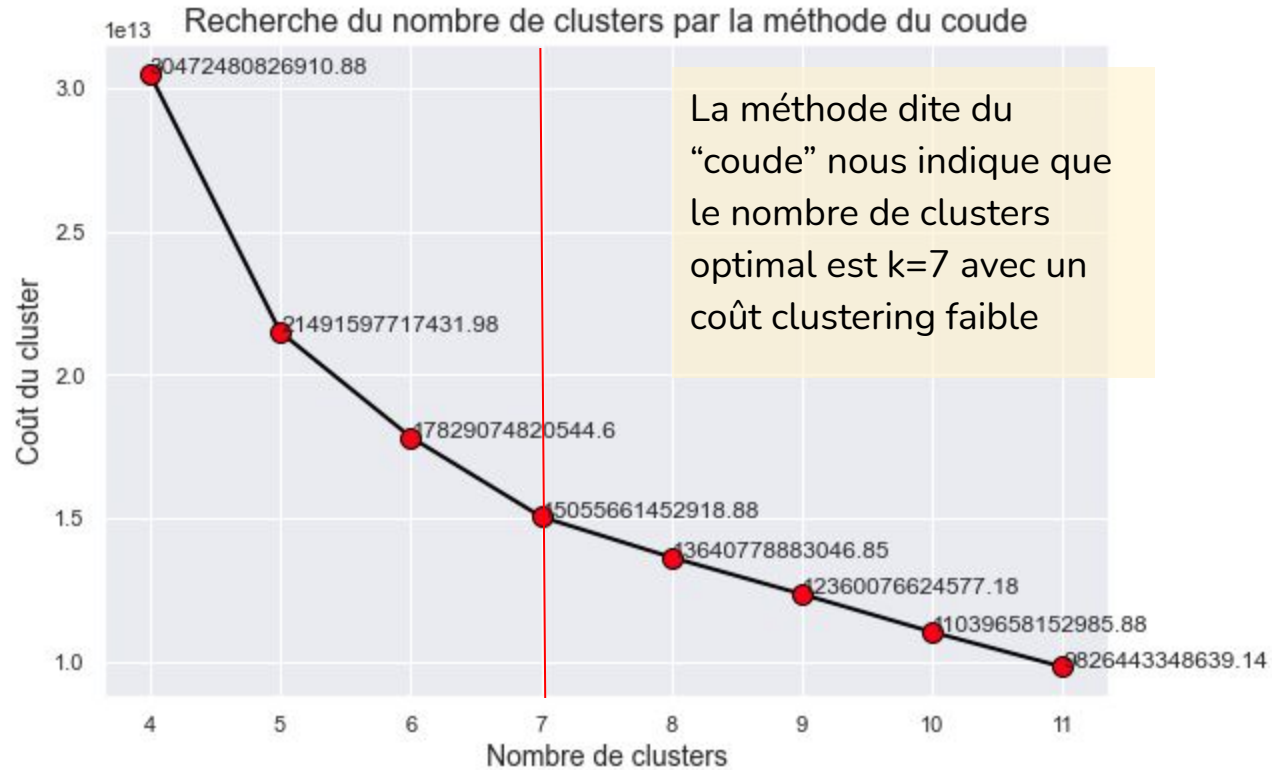
Régulier - '221', '222', '321', '322', '121', '122', '211', '212'

Les clients fidèles achètent régulièrement et pour des montants importants

# **CLUSTERING DES DONNÉES BRUTES AVEC K PROTOTYPE**



## Evolution du coût du clustering en fonction du nombre de cluster

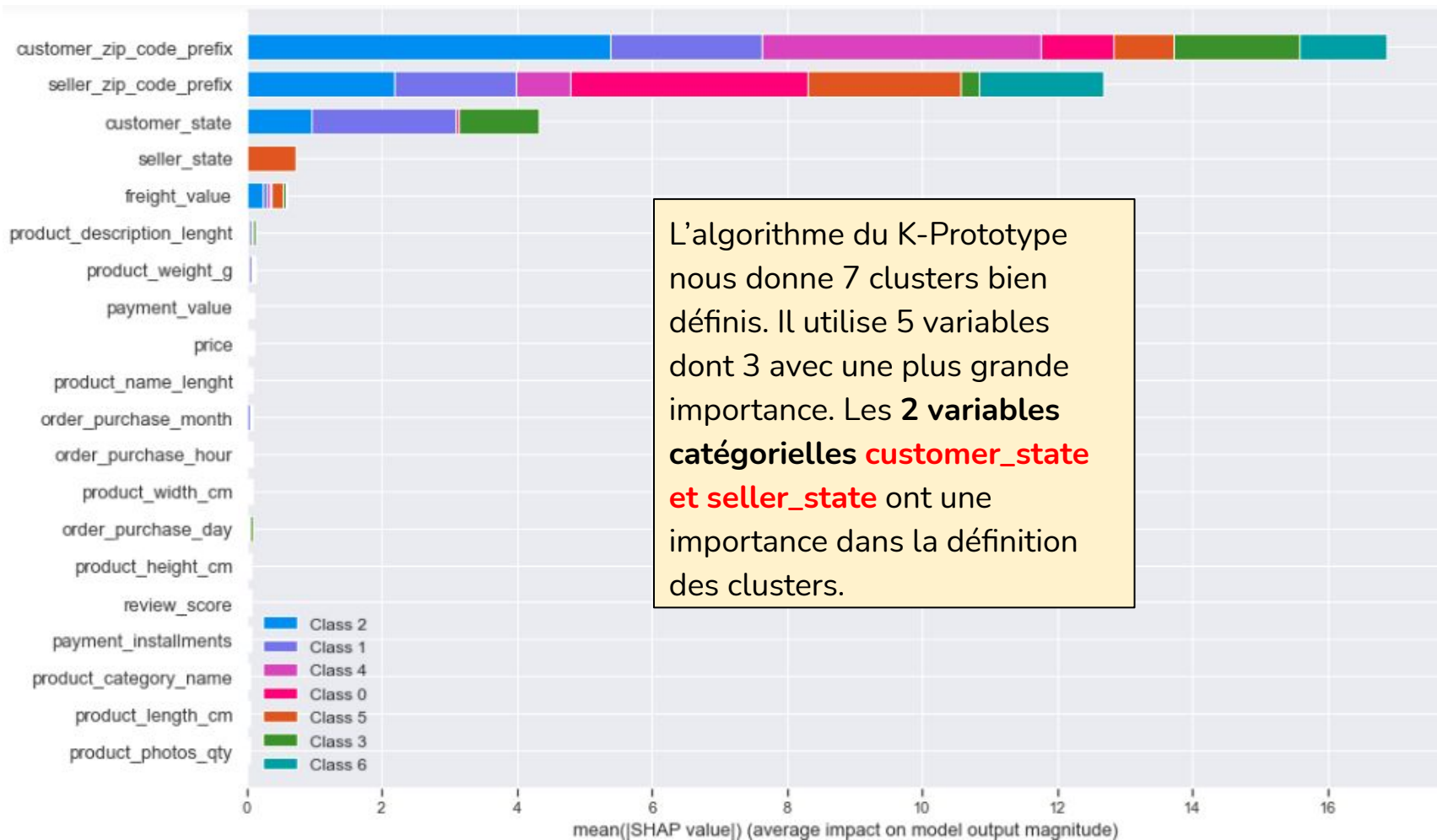


## Caractéristiques des 7 types de clients

Segment	payment_value_mean	review_score_mode	price_mean	order_status_mode	pay_inst_mode	purchase_time_day	purchase_dayofweek	product_category
Cluster 1	196.844499	5.0	143.537914	delivered	1.0	Afternoon	Mon	informatica ace
Cluster 2	168.232581	5.0	113.814975	delivered	1.0	Afternoon	Tue	cama mesa
Cluster 3	144.042377	5.0	101.681568	delivered	1.0	Afternoon	Tue	cama mesa
Cluster 4	205.722824	5.0	139.467668	delivered	1.0	Afternoon	Mon	beleza
Cluster 5	179.523710	5.0	119.278593	delivered	1.0	Afternoon	Wed	cama mesa
Cluster 6	191.282884	5.0	143.420538	delivered	1.0	Afternoon	Tue	informatica ace
Cluster 7	219.829522	5.0	156.642450	delivered	1.0	Afternoon	Mon	informatica ace

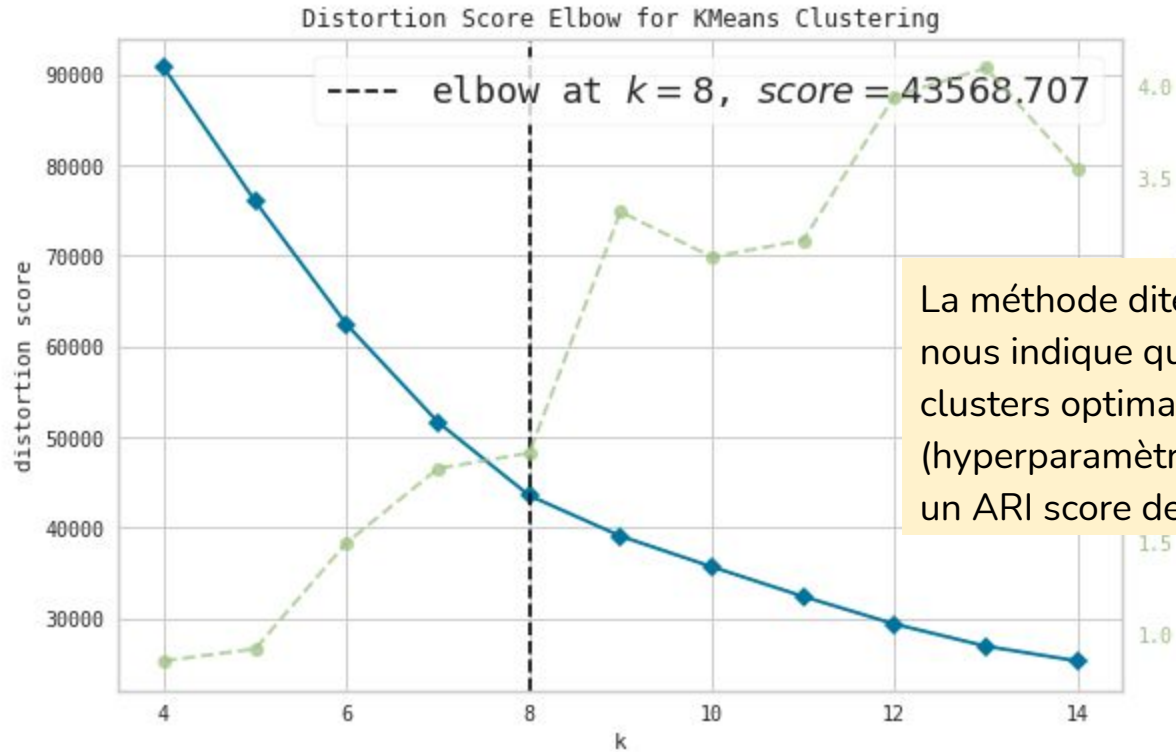
purchase_time_day	purchase_dayofweek	product_category_name	customer_state	seller_state_mode	ord_purch_year_mode	ord_purch_month_mode	Total
Afternoon	Mon	informatica acessorios	SP	PR	2018.0	7.0	11831
Afternoon	Tue	cama mesa banho	RJ	SP	2018.0	5.0	23953
Afternoon	Tue	cama mesa banho	SP	SP	2018.0	8.0	38412
Afternoon	Mon	beleza saude	DF	SP	2018.0	7.0	10932
Afternoon	Wed	cama mesa banho	RS	SP	2018.0	5.0	15441
Afternoon	Tue	informatica acessorios	SP	MG	2017.0	8.0	9626
Afternoon	Mon	informatica acessorios	RS	PR	2018.0	8.0	6407

## Interprétabilité des 10 clusters formés par K-prototypé



# **CLUSTERING DES DONNÉES RFM AVEC K MEANS**

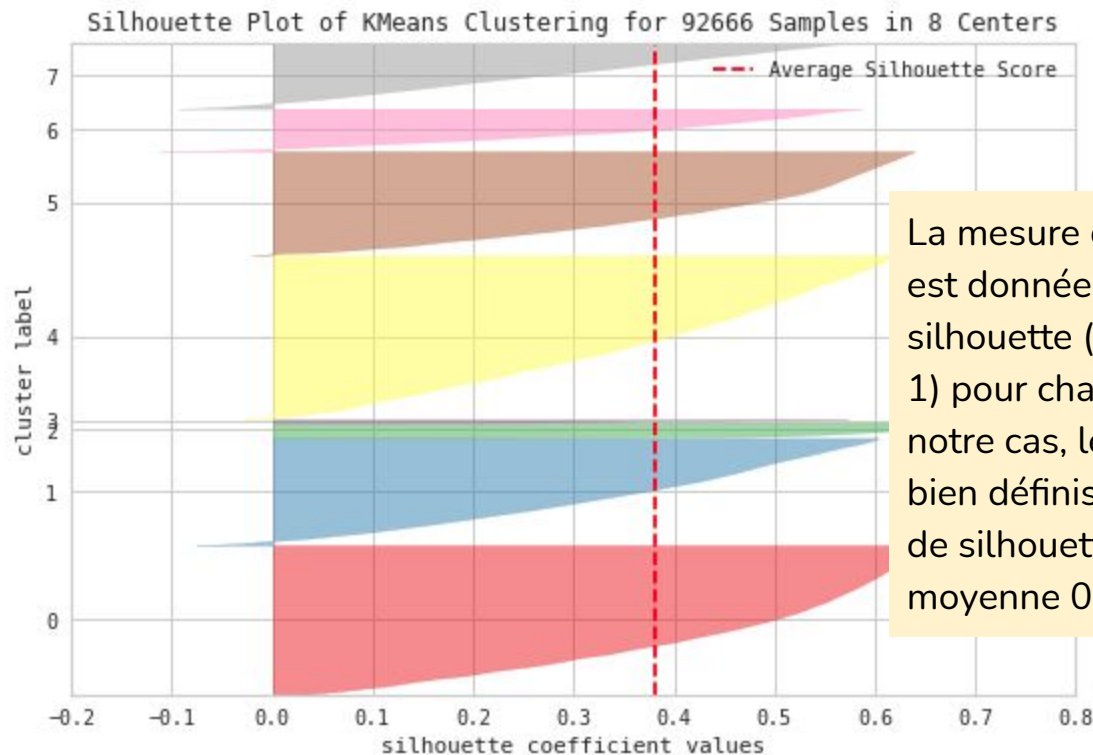
## Recherche du nombre optimal de clusters k



La méthode dite du “coude” nous indique que le nombre de clusters optimal est  $k=8$  (hyperparamètre du K Means) et un ARI score de 0.97



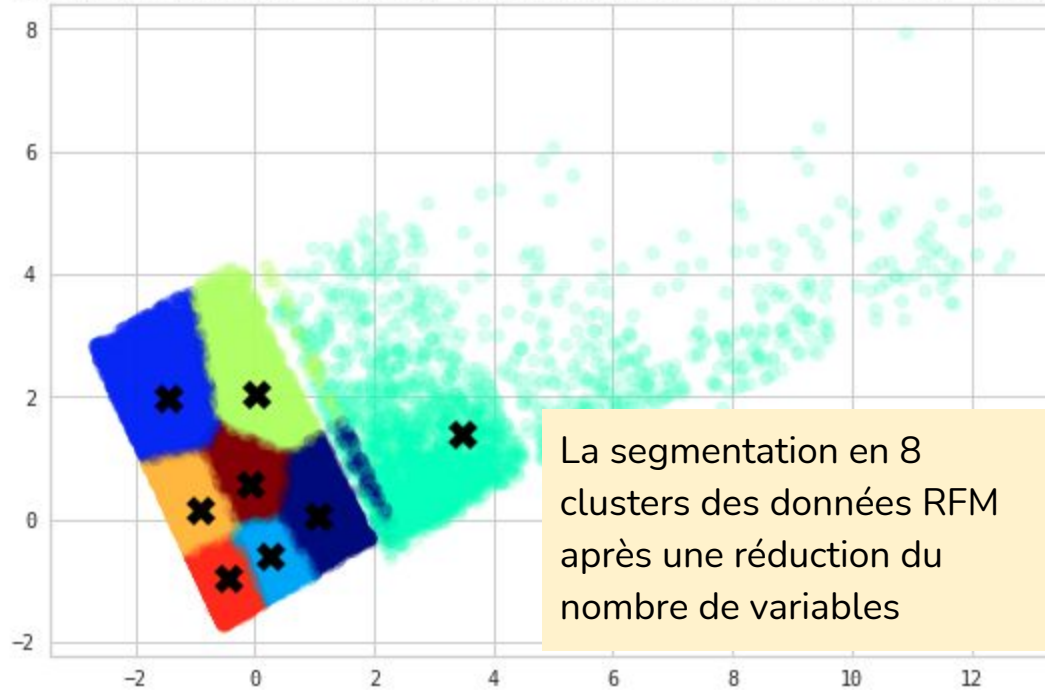
## Affichage du coefficient de silhouette pour chaque cluster



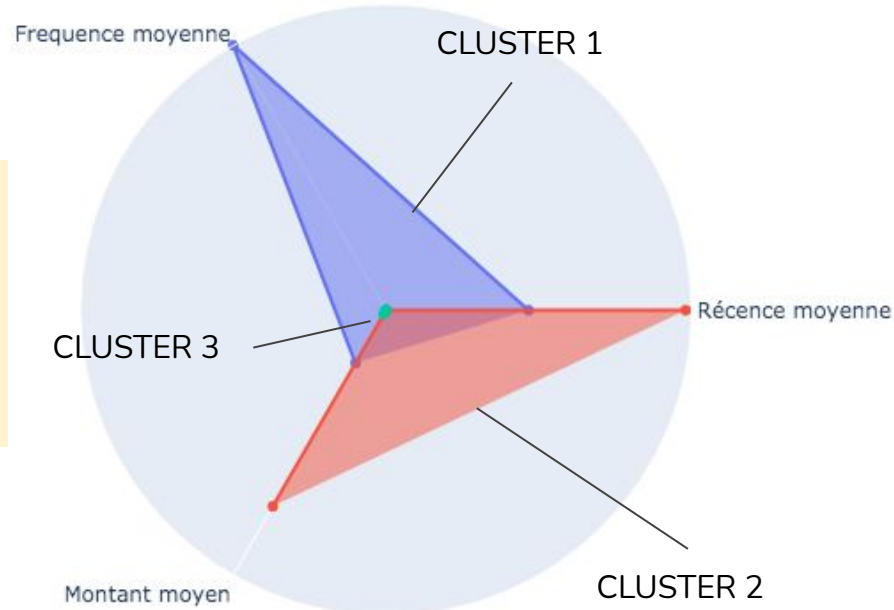
La mesure de la performance est donnée par le coefficient de silhouette (compris entre -1 et 1) pour chaque cluster. Dans notre cas, les clusters sont bien définis et leur coefficient de silhouette est supérieur à la moyenne 0.38

## Affichage des 8 clusters ainsi que les centroïdes

Projection des 92666 individus et des 8 centroïdes sur le plan factoriel



## Radar Chart des clusters 1, 2 et 3



Définition des différents clusters :

One shot - Clusters 1 et 3

Gros clients - Cluster 2

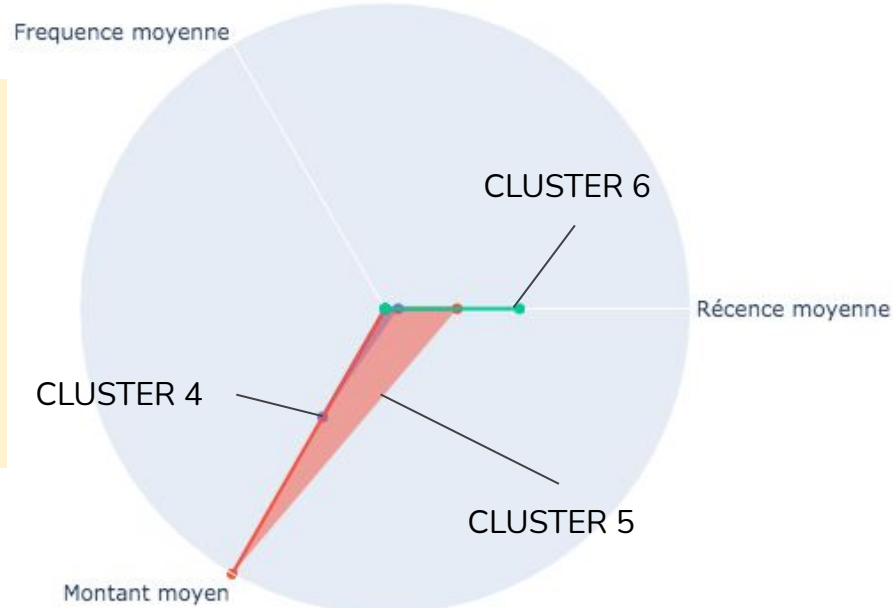
## Radar Chart des clusters 4, 5 et 6

Définition des différents clusters :

Gros One shot - Cluster 5

Très Réguliers - Cluster 4

Nouveaux - Cluster 6

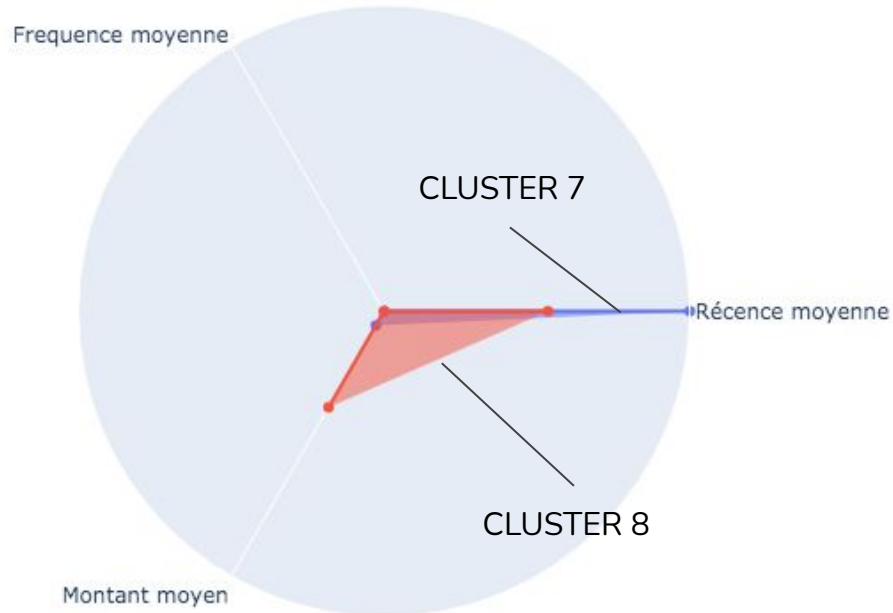


## Radar Chart des clusters 7 et 8

Définition des différents clusters :

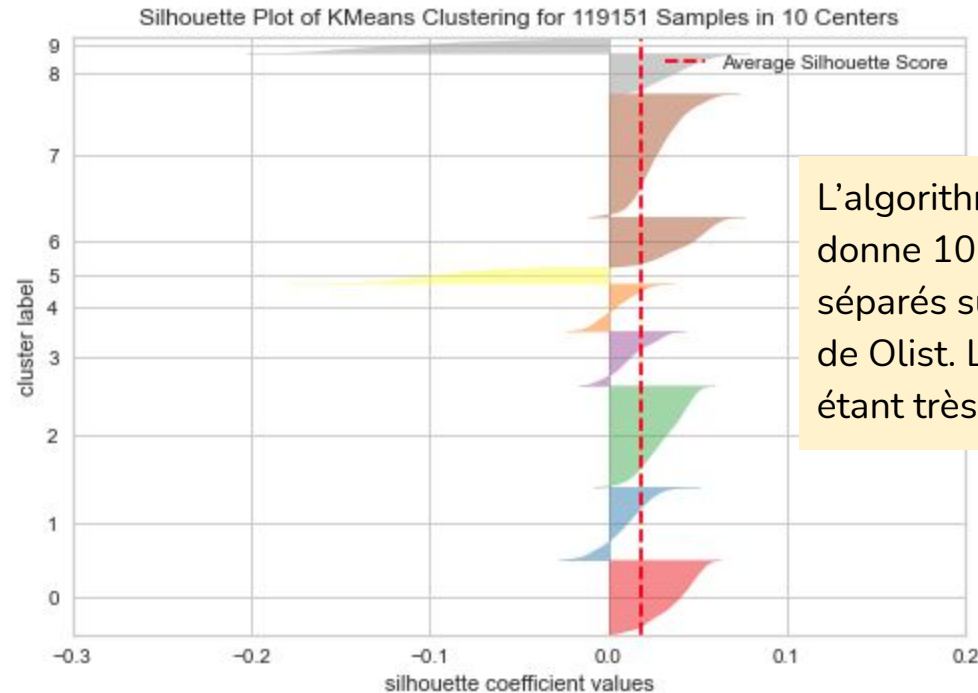
Fidèles - Cluster 7

Moins Réguliers - Cluster 8



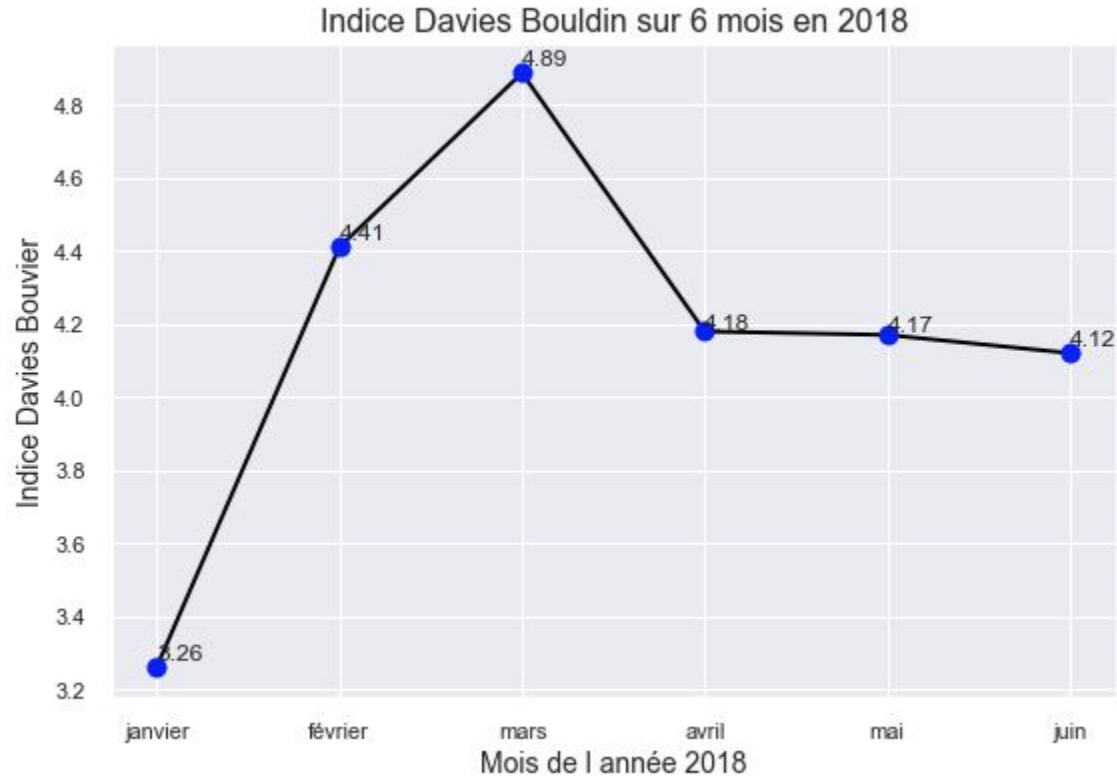
# **CLUSTERING DES DONNÉES BRUTES OLIST AVEC K MEANS**

## Affichage du coefficient de silhouette pour chaque cluster



L'algorithme du KMeans nous donne 10 clusters bien définis et séparés sur les données brutes de Olist. Les clusters 2, 3, 4 et 5 étant très proches

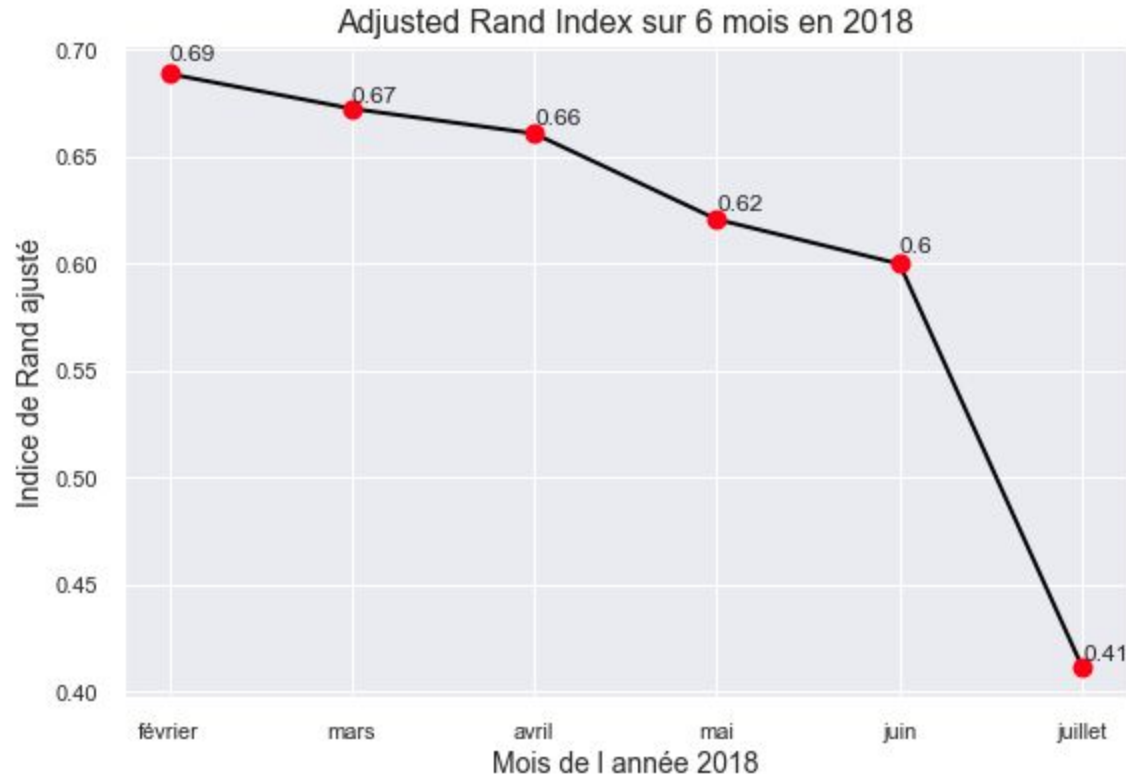
## Mesure de la performance à l'aide de l'indice Davies Bouldin



La mesure de la performance par l'indice Davies Bouldin nous montre une valeur stable dès avril 2018 indiquant une meilleure segmentation des clients en 10 clusters.

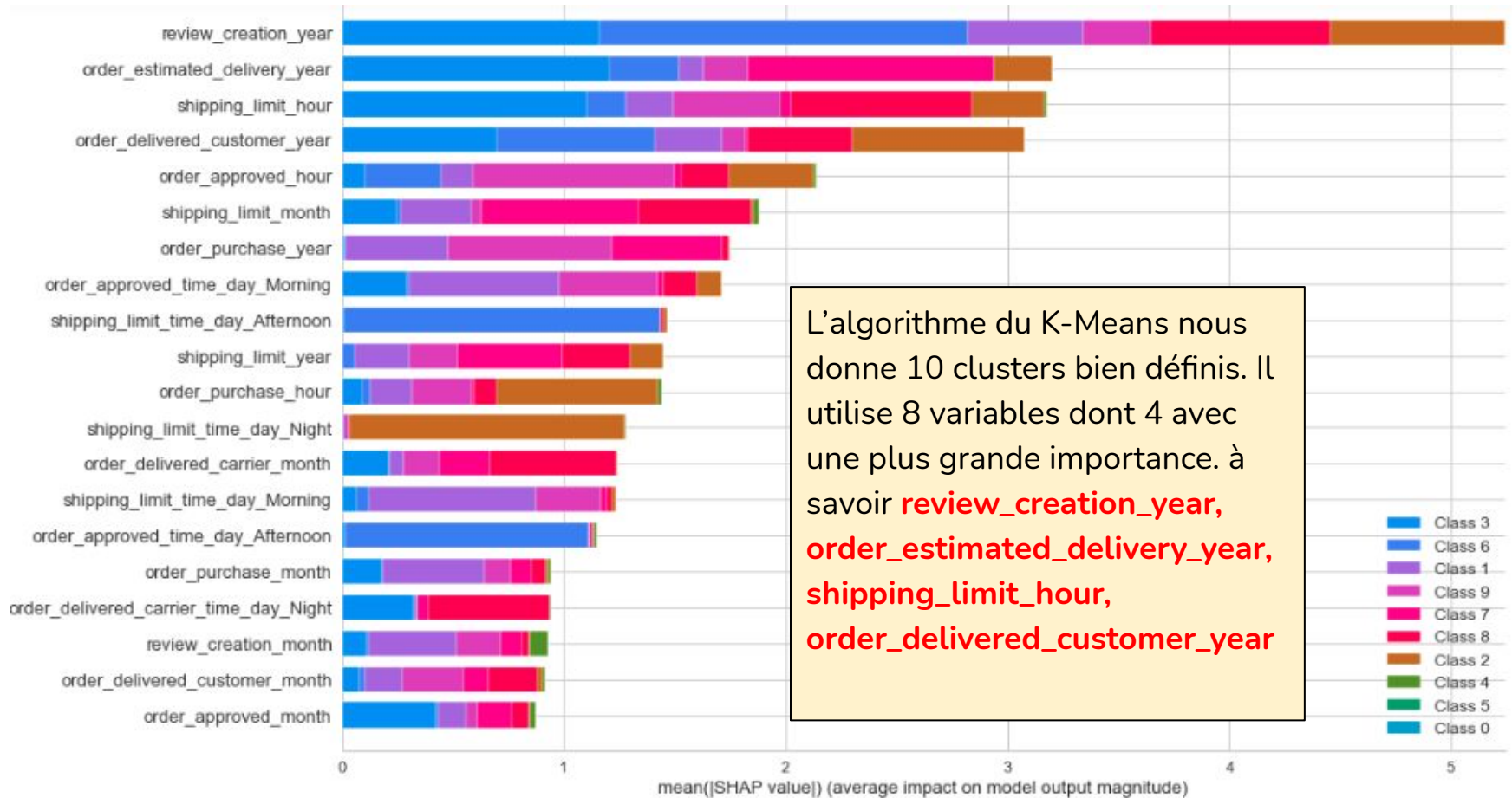


## Mesure de la similarité des clusters au fil des mois



Une périodicité de 5 mois est mise en évidence sur la stabilité du score ARI qui mesure la variabilité du clustering. Un contrat de maintenance de la segmentation des clients de Olist tous les 5 mois est réaliste afin d'opérer un nouveau clustering

## Interprétabilité des 10 clusters



## 4. CONCLUSION



## Segmentation métier RFM

Définition de 5 types de client :

Fidèle - '123'

One shot - '311', '312', '313'

Nouveau - '111', '112', '113'

Gros - '323', '213', '223'

Régulier - '221', '222', '321', '322',  
'121', '122', '211', '212'

## Segmentation par K Means

Définition de 8 types de client à partir  
des données RFM ::

One shot - **Clusters 1 et 3**

Gros clients - **Cluster 2**

Gros One shot - **Cluster 5**

Très Réguliers - **Cluster 4**

Nouveaux - **Cluster 6**

Fidèles - **Cluster 7**

Moins Réguliers (montants  
importants) - **Cluster 8**

## Segmentation par K Prototype

Définition de 7 types de client :

Achetant du matériel informatique et  
habitant à Sao Paulo - **Clusters 1, 6  
et 7 (26000 clients)**

Achetant des couvertures de lit et le  
vendeur est à Sao Paulo. Les clients  
habitent à Sao Paulo, Rio et Rio  
Grande do Sul - **Clusters 2, 3 et 5  
avec le plus de clients (75000  
clients)**

Achetant des produits de beauté et  
habitant le **District Fédéral** -  
**Clusters 4 (11000 clients)**

# RECOMMANDATIONS POUR OLIST

Pour augmenter les revenus, Olist doit convertir les **Gros clients** et les clients **One Shot et Nouveaux** en clients fidèles qui représentent moins de 1% de la clientèle. Cela passe par des envois réguliers d'offres promotionnels par mail ou sms à ces clients afin d'augmenter leurs fréquentations sur le site,

On peut augmenter le montant d'achat des **Nouveaux clients** ainsi que **les réguliers** en attribuant une réduction à partir d'un certain montant d'achat.

Les clients du site Olist achètent moins le week-end. Afin d'augmenter les commandes en fin de semaine, Olist pourrait lancer **des offres promotionnelles spéciales week-end**.

Les clients du site Olist achètent moins le matin et la nuit. Afin d'augmenter les commandes durant ces périodes, Olist pourrait lancer **des offres promotionnelles spéciales nuit et matin**.

