

A photograph of a modern office interior. In the center, the word "olist" is displayed in large, blue, three-dimensional letters on a glass partition. Behind the glass, several employees are seated at desks, working on computers. The office has a high ceiling with exposed pipes and modern lighting fixtures. The floor is polished and reflects the overhead lights.

olist

**SEGMENTATION DE DONNÉES
CLIENTS E-COMMERCE**

INTRODUCTION

_OBJECTIFS

- Comprendre le comportement des clients
- Evaluer la segmentation RFM
- Tester de nouvelle modélisation
- Définir des personas pour optimiser les campagnes de pub

INTRODUCTION

_PLAN

1. Introduction:

- Objectifs
- Données

2. Analyse exploratoire:

- Comportements
- Features

3. Modélisations:

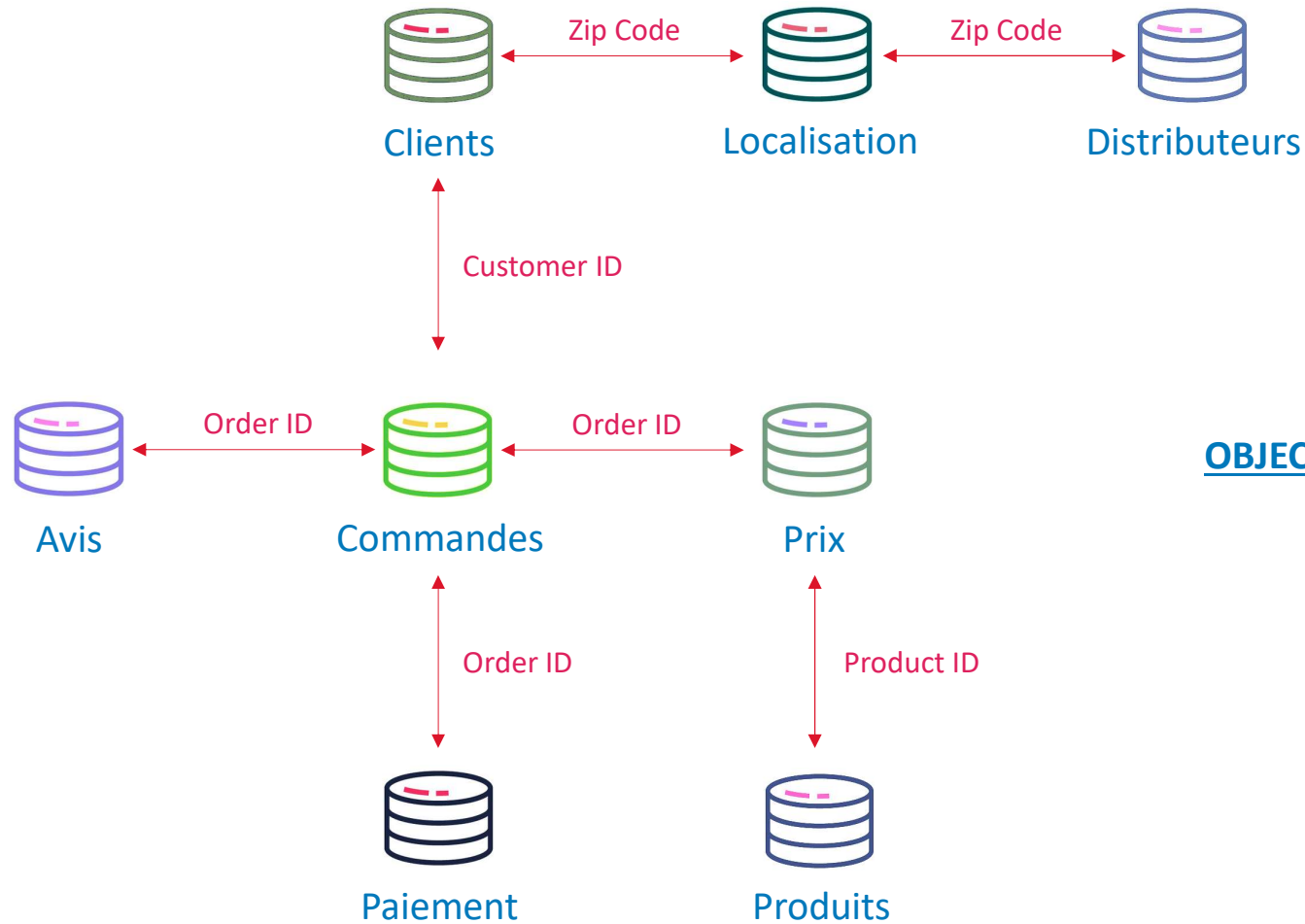
- RFM
- K-Means
- K-Modes

4. Conclusion:

- Personae
- Maintenance

INTRODUCTION

_DONNÉES



OBJECTIF: Réunir les données

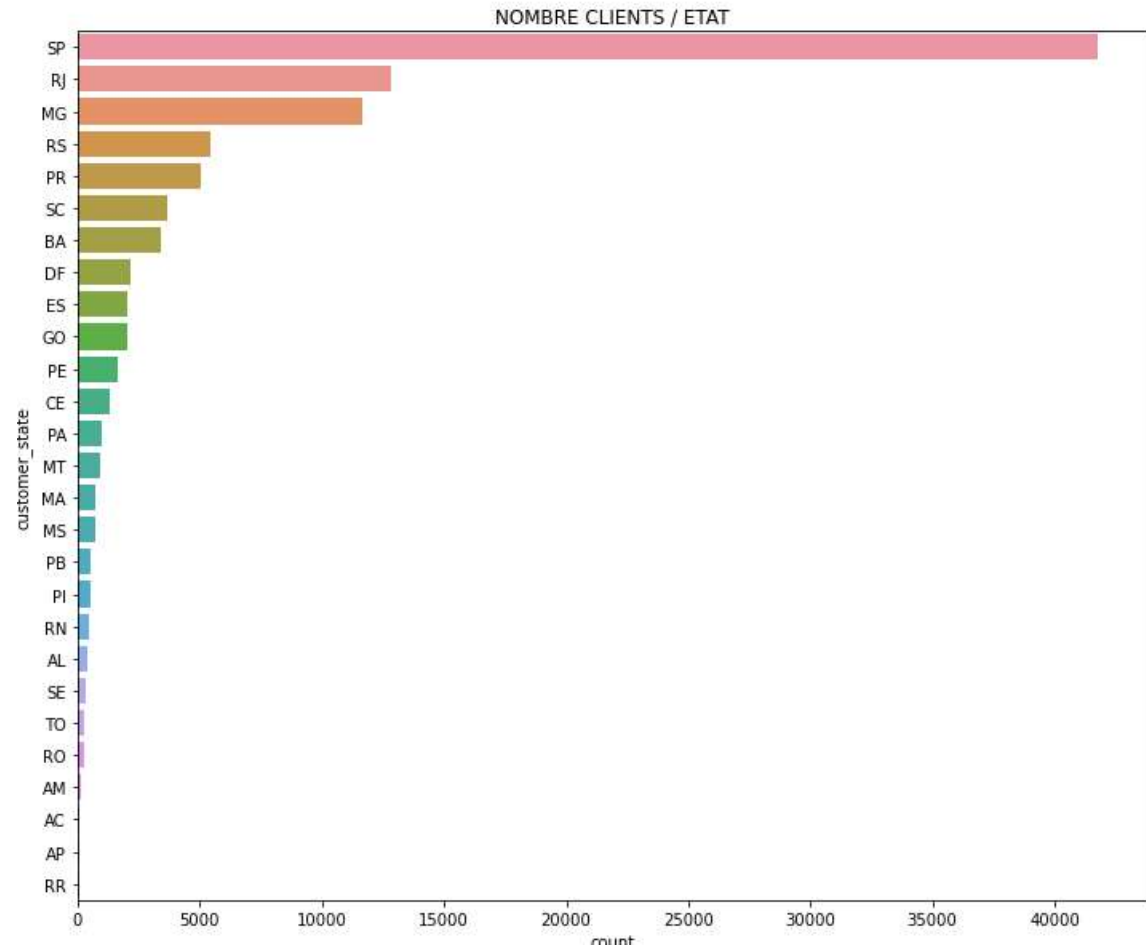
ANALYSE

_TENDANCE

CLIENTS

ID: 99441

ID unique: 96096



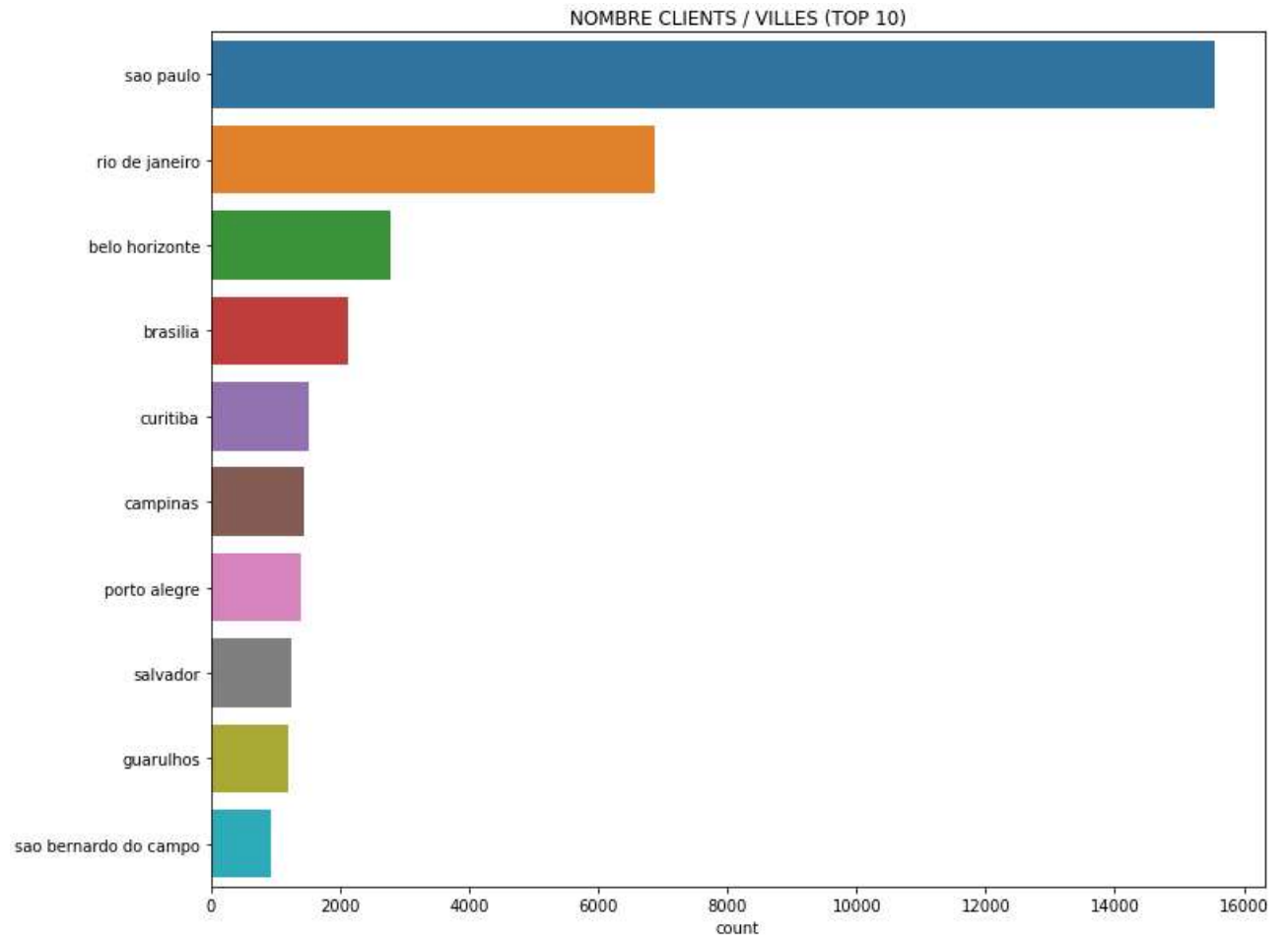
ANALYSE

_TENDANCE

CLIENTS

ID: 99441

ID unique: 96096

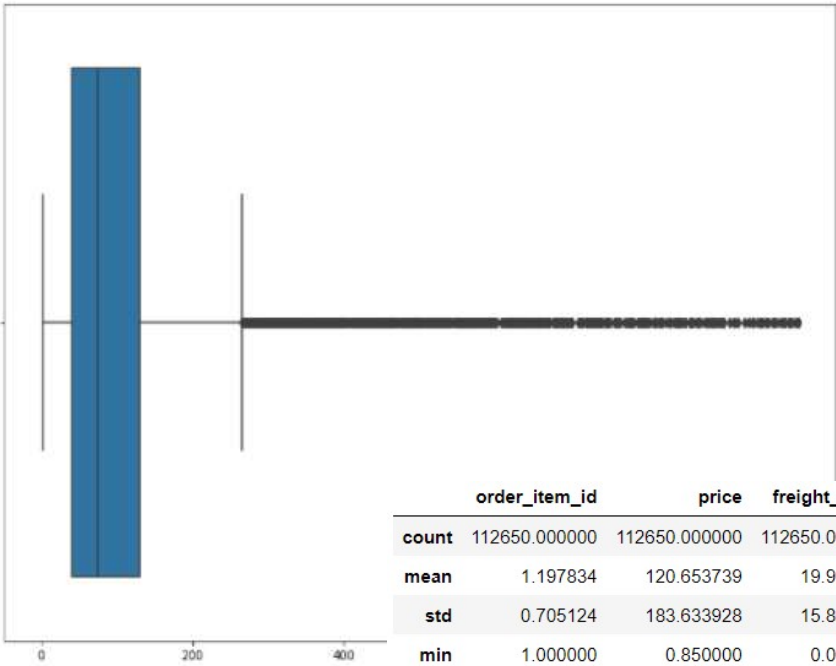
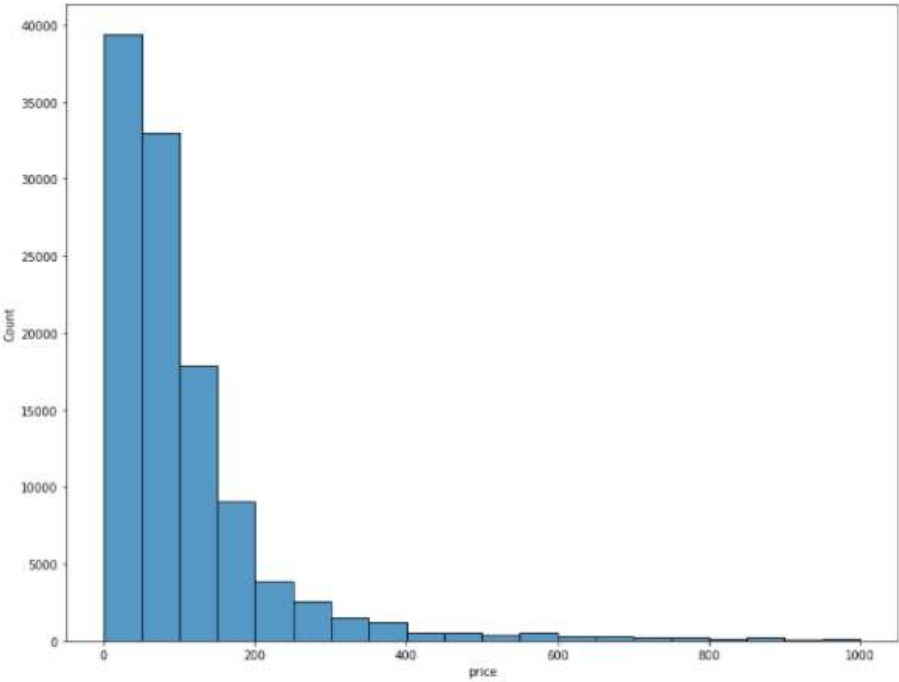


ANALYSE

_TENDANCE

PRIX

REPARTITION VALEURS COMMANDES < 1000

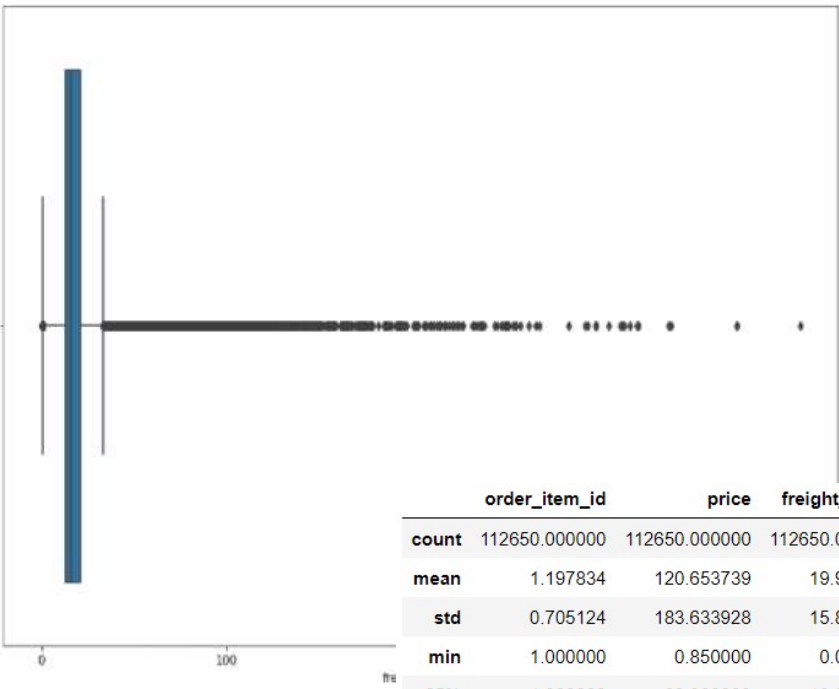
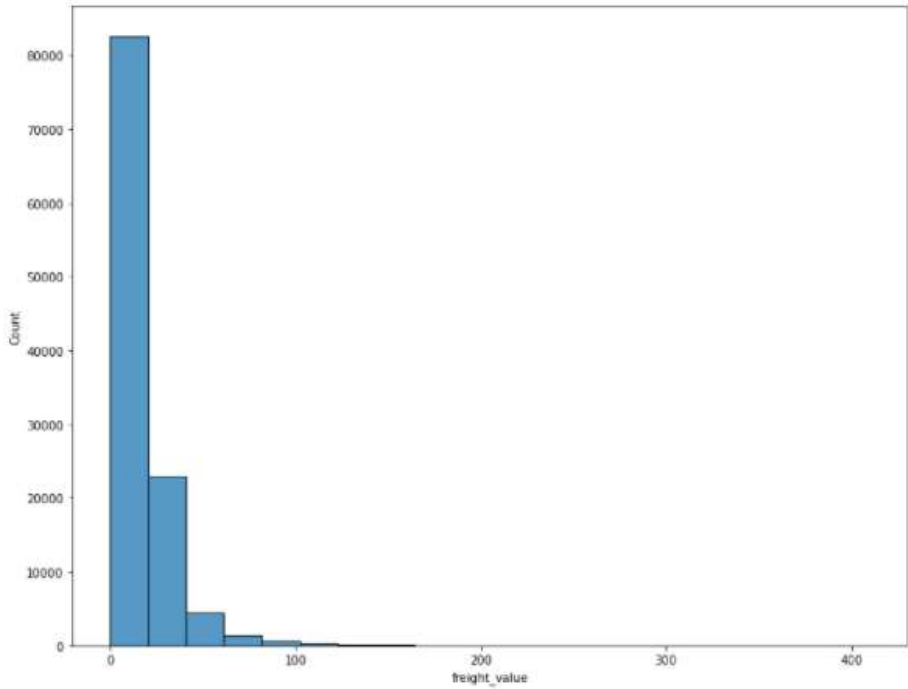


	order_item_id	price	freight_value
count	112650.000000	112650.000000	112650.000000
mean	1.197834	120.653739	19.990320
std	0.705124	183.633928	15.806405
min	1.000000	0.850000	0.000000
25%	1.000000	39.900000	13.080000
50%	1.000000	74.990000	16.260000
75%	1.000000	134.900000	21.150000
max	21.000000	6735.000000	409.680000

ANALYSE
_TENDANCE

PRIX

REPARTITION VALEURS FRAIS PORT



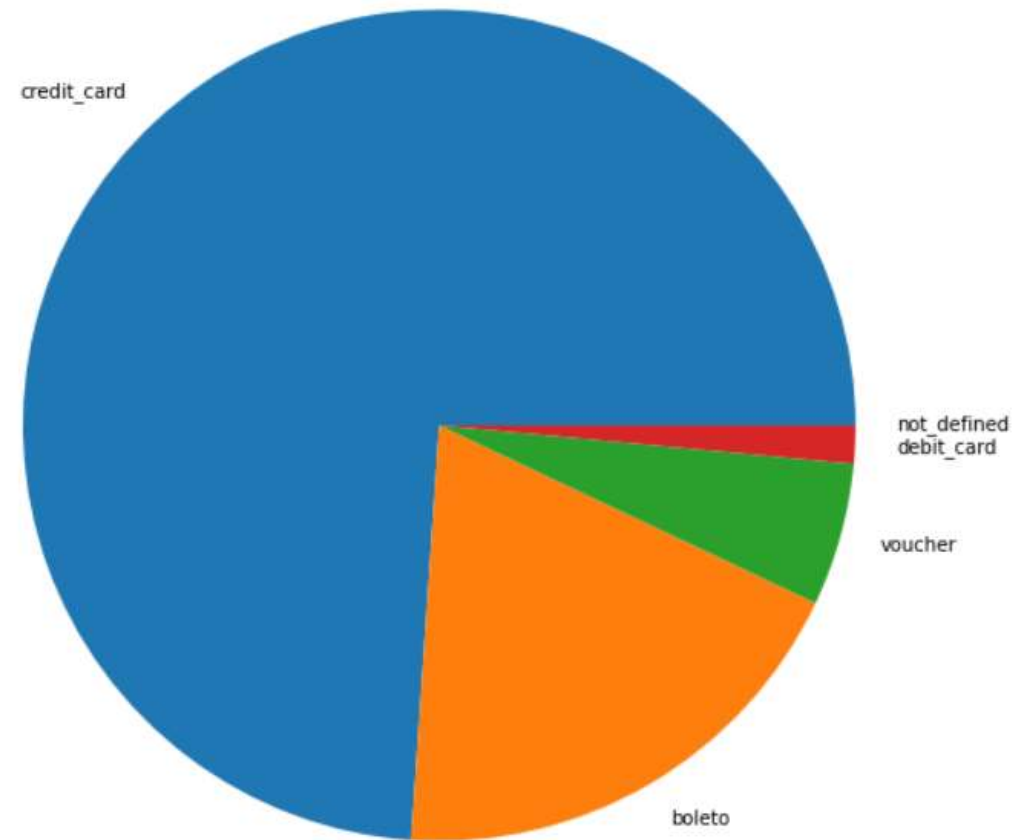
	order_item_id	price	freight_value
count	112650.000000	112650.000000	112650.000000
mean	1.197834	120.653739	19.990320
std	0.705124	183.633928	15.806405
min	1.000000	0.850000	0.000000
25%	1.000000	39.900000	13.080000
50%	1.000000	74.990000	16.260000
75%	1.000000	134.900000	21.150000
max	21.000000	6735.000000	409.680000

ANALYSE

_TENDANCE

PAIEMENT

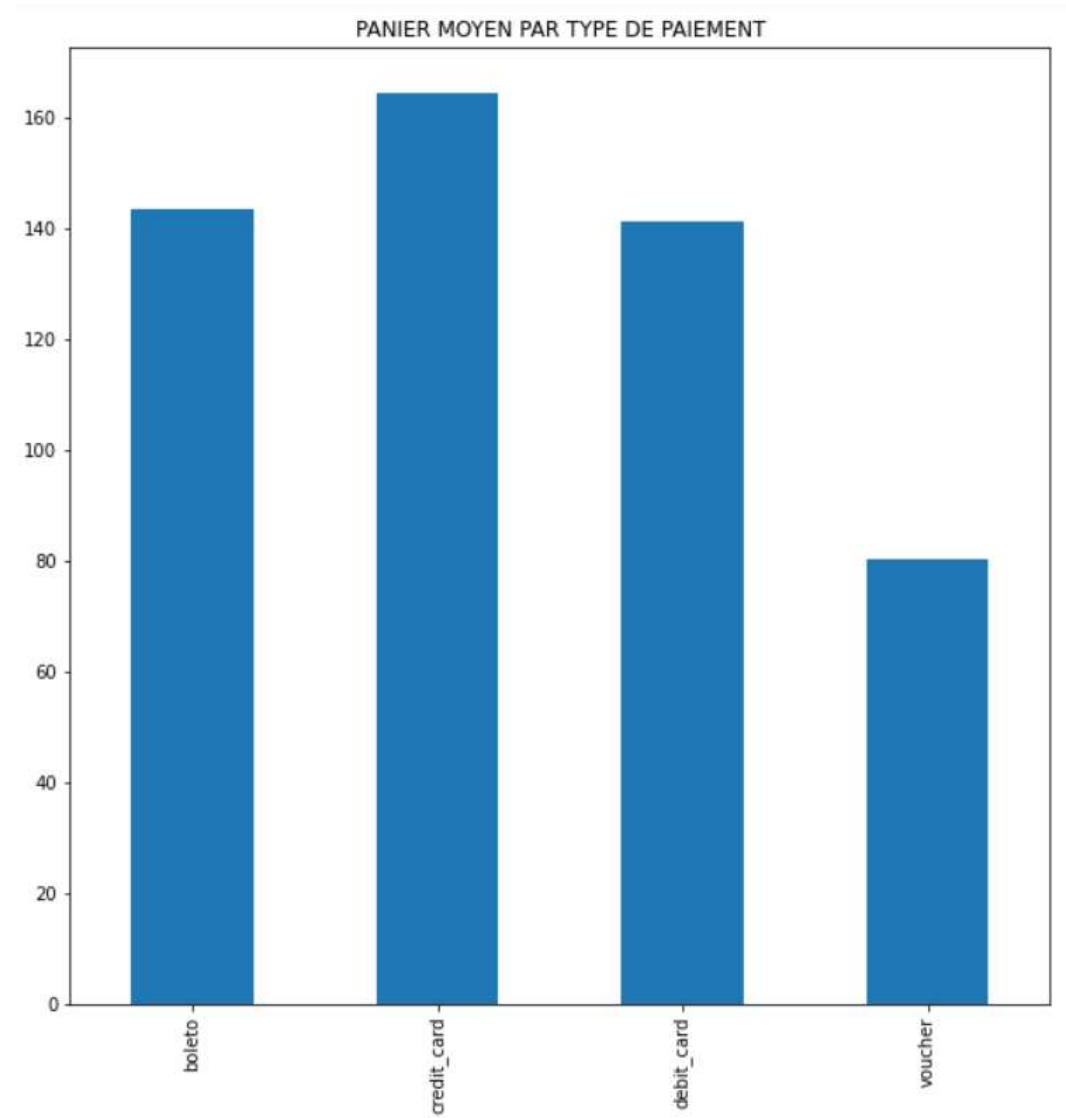
REPARTITION TYPES PAIEMENT



ANALYSE

_TENDANCE

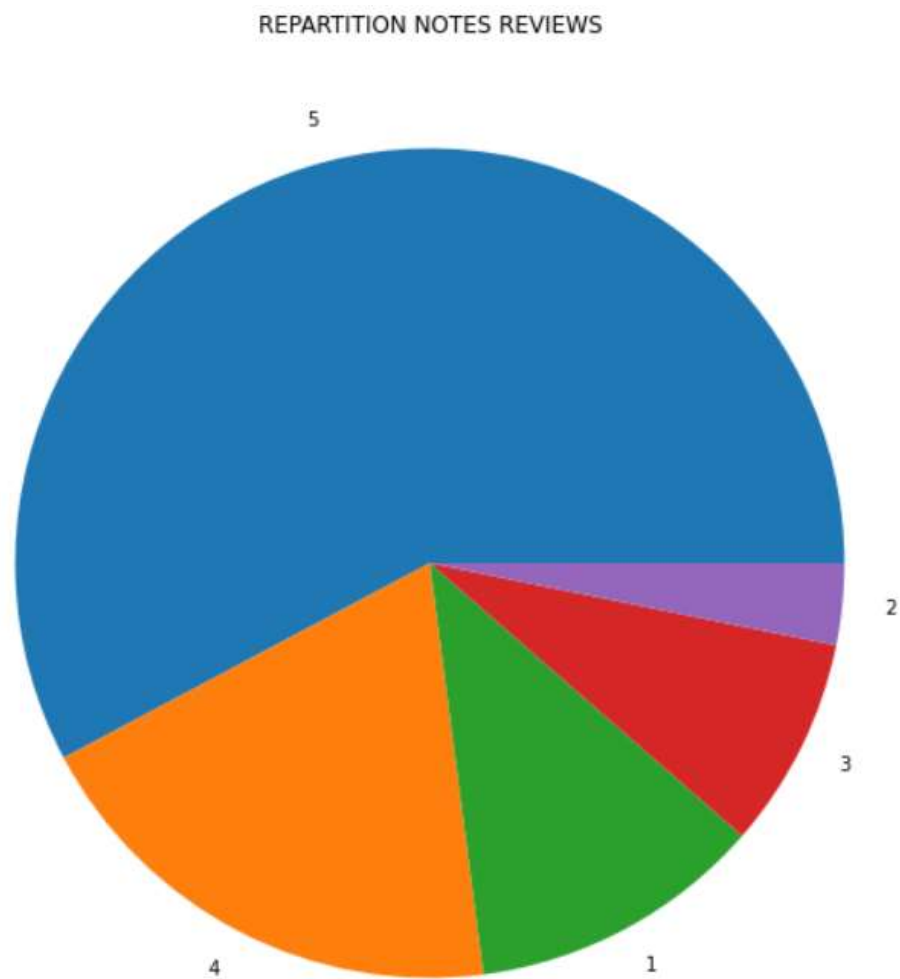
PAIEMENT



ANALYSE

_TENDANCE

AVIS

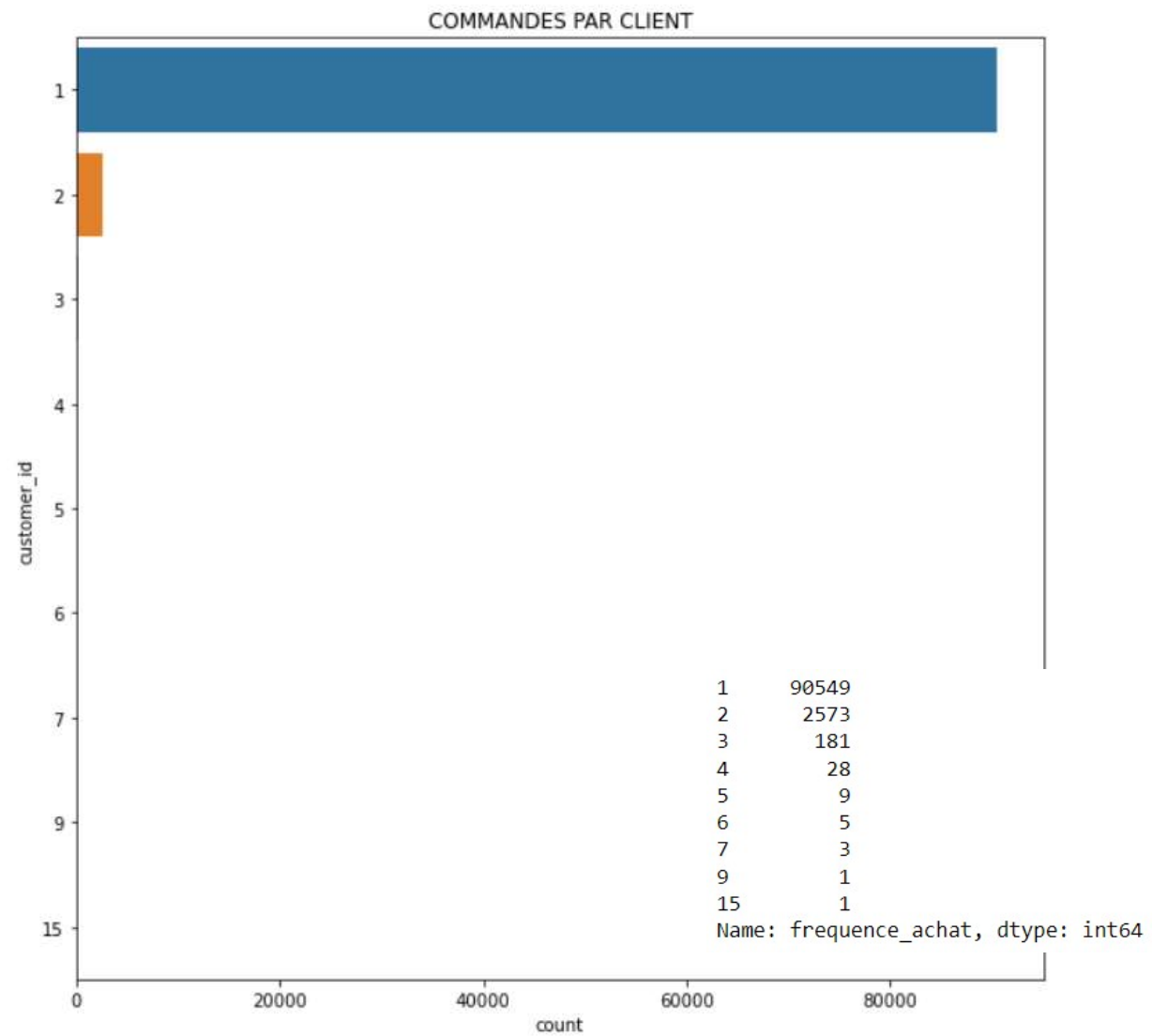


ANALYSE

_TENDANCE

COMMANDES

1 commande: 97%

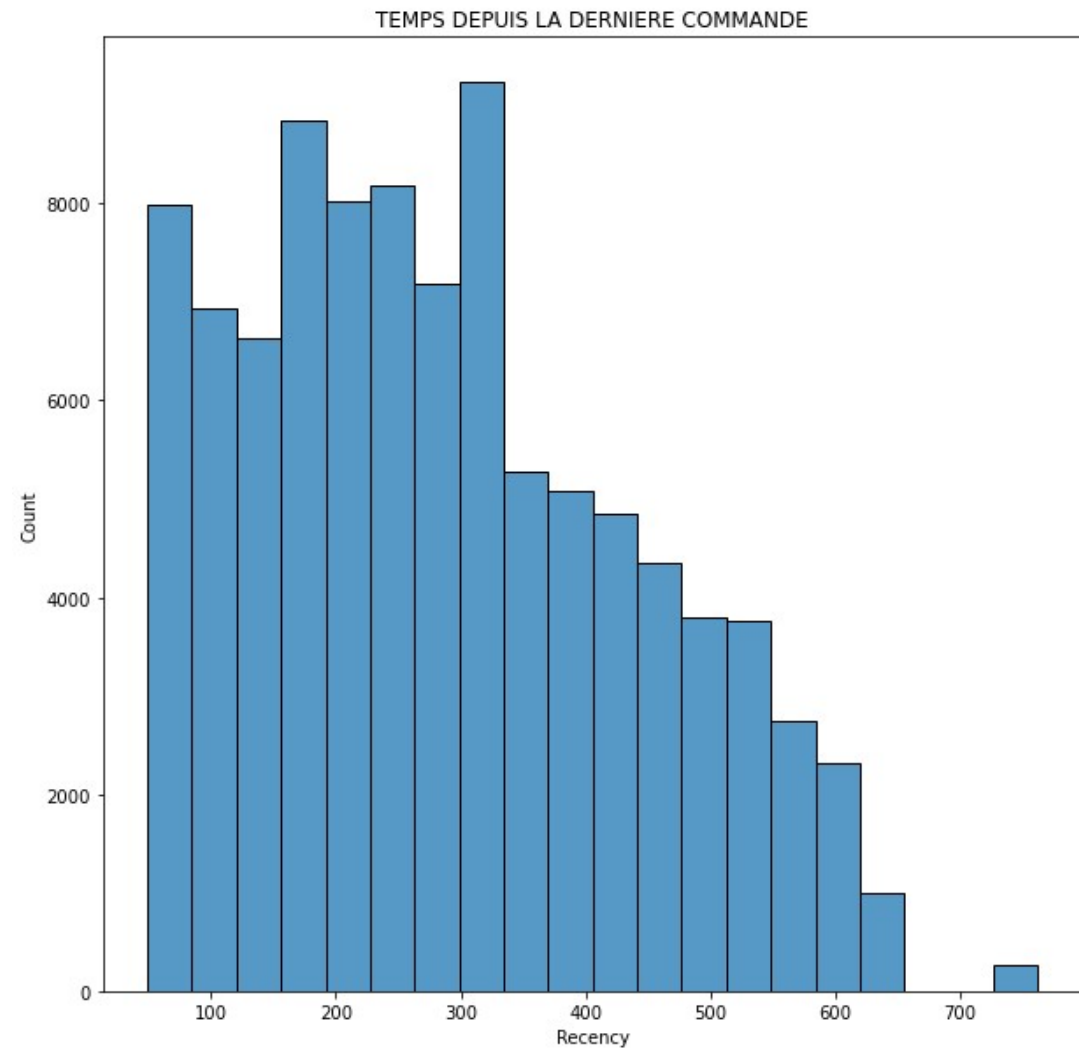


ANALYSE

_TENDANCE

COMMANDES

Recency



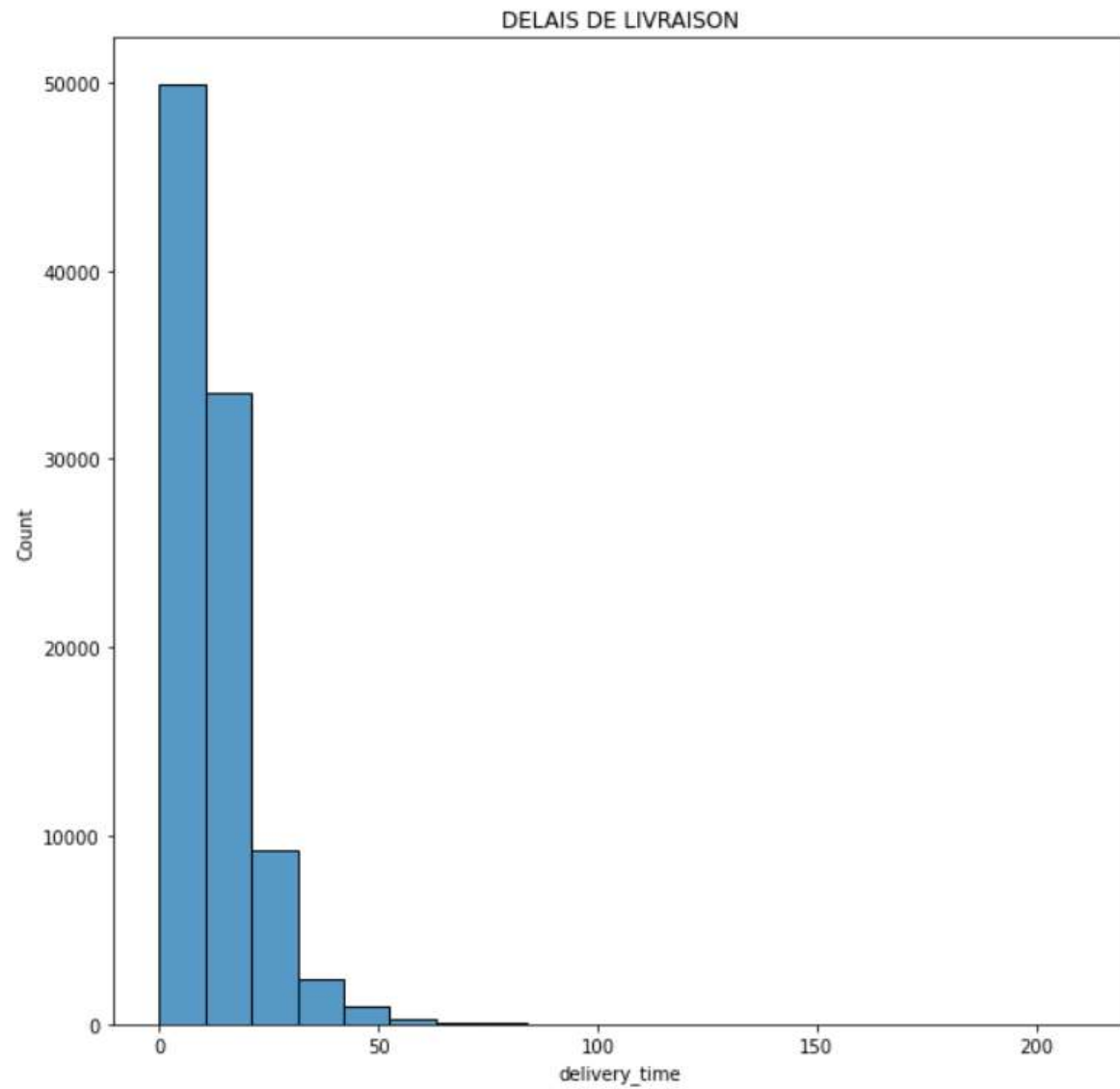
ANALYSE

_TENDANCE

COMMANDES

Livraison:

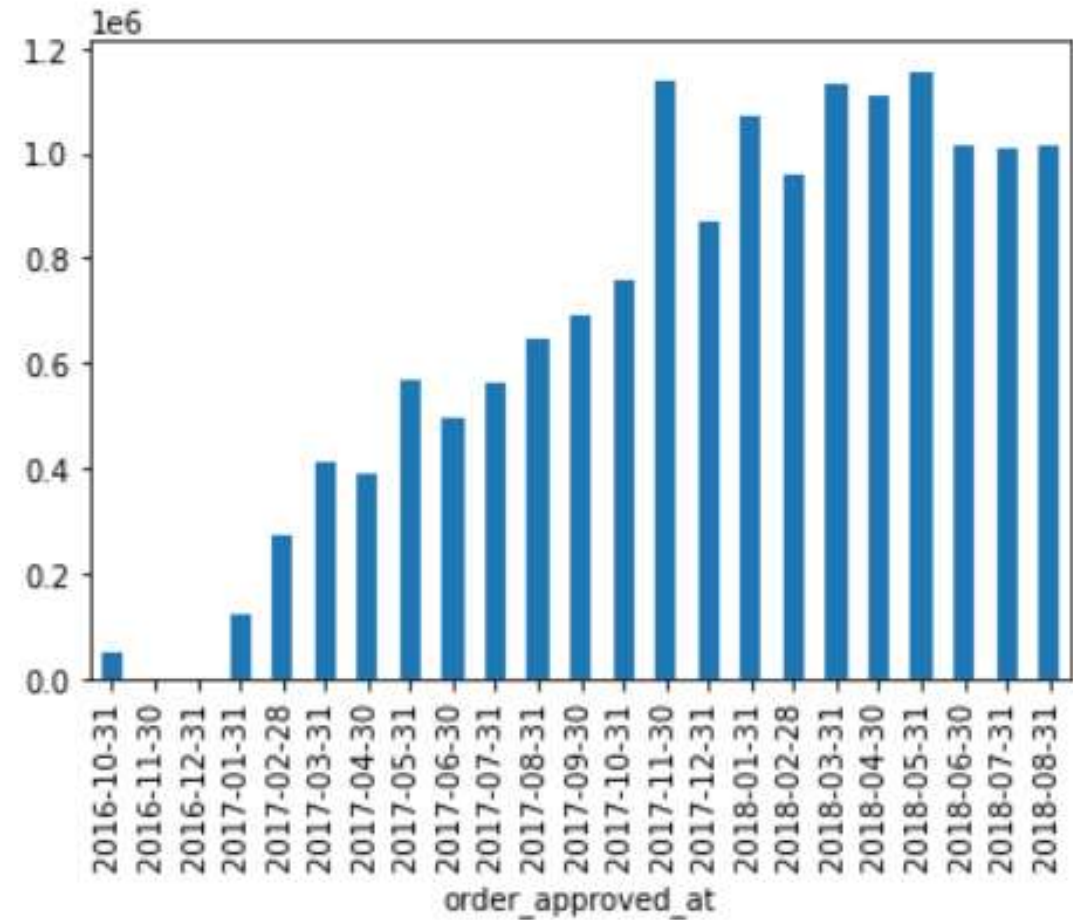
- `pd.to_datetime`
- `df[].dt.date`



ANALYSE

_TENDANCE

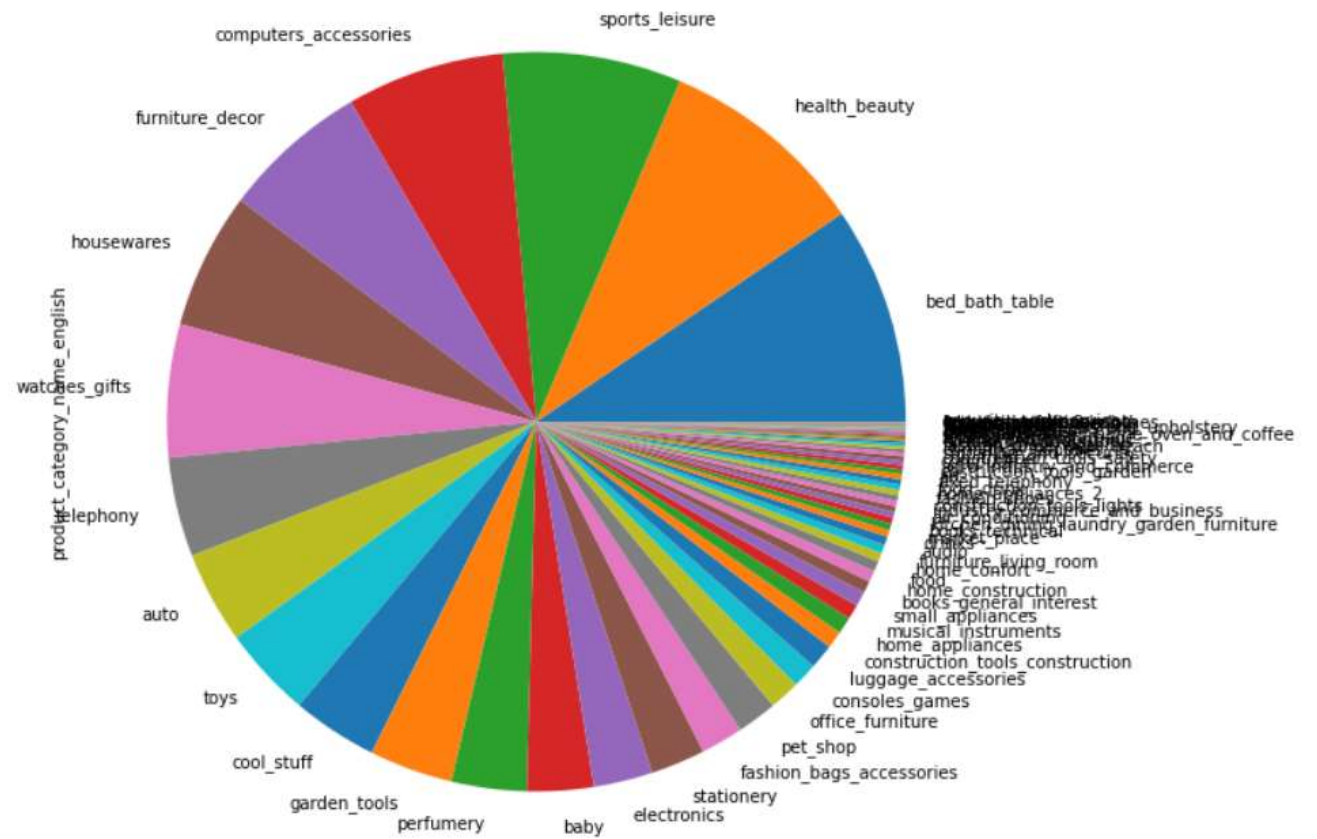
COMMANDES



ANALYSE

_TENDANCE

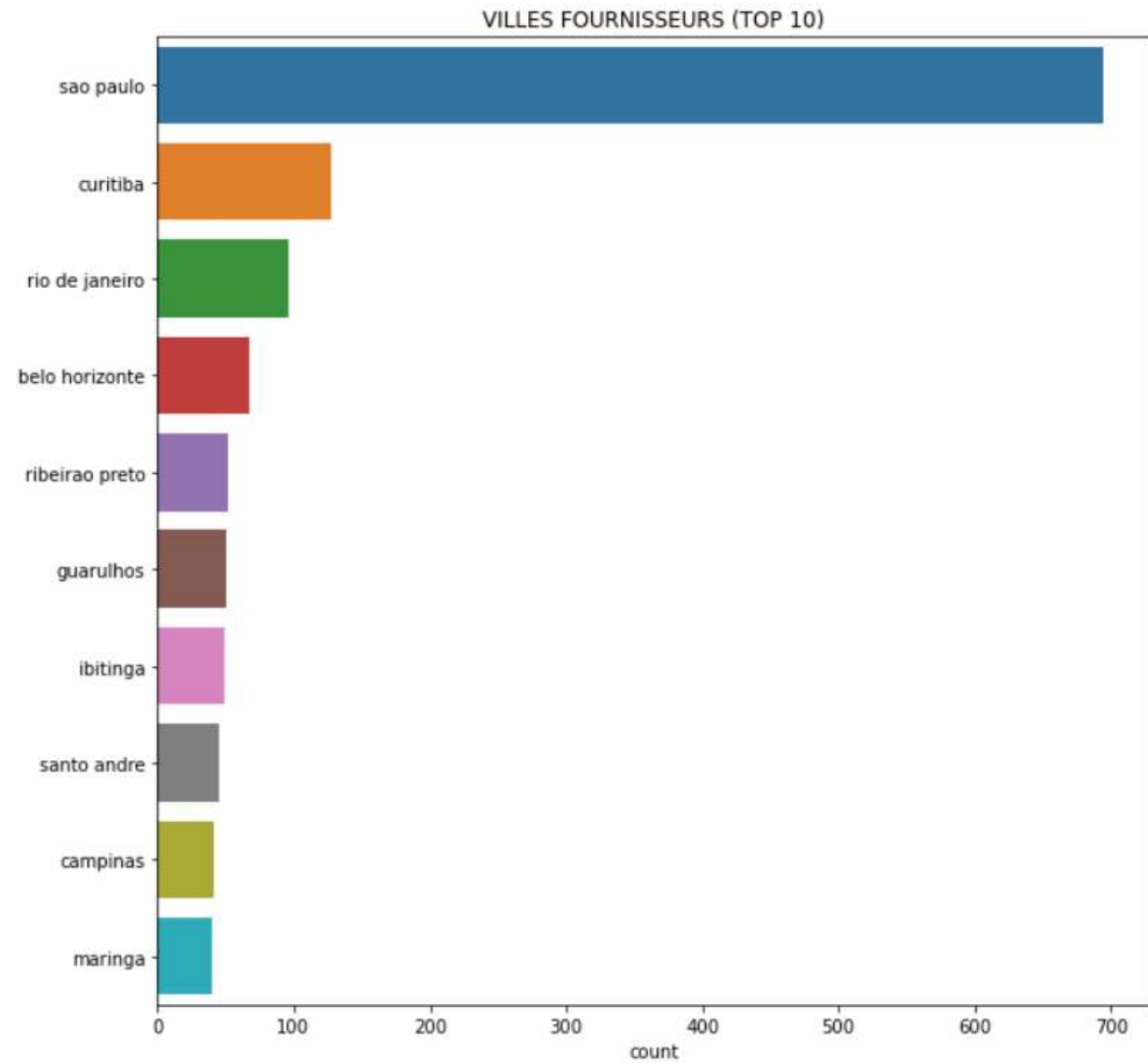
PRODUITS



ANALYSE

_TENDANCE

DISTRIBUTEURS



ANALYSE

_ FEATURES

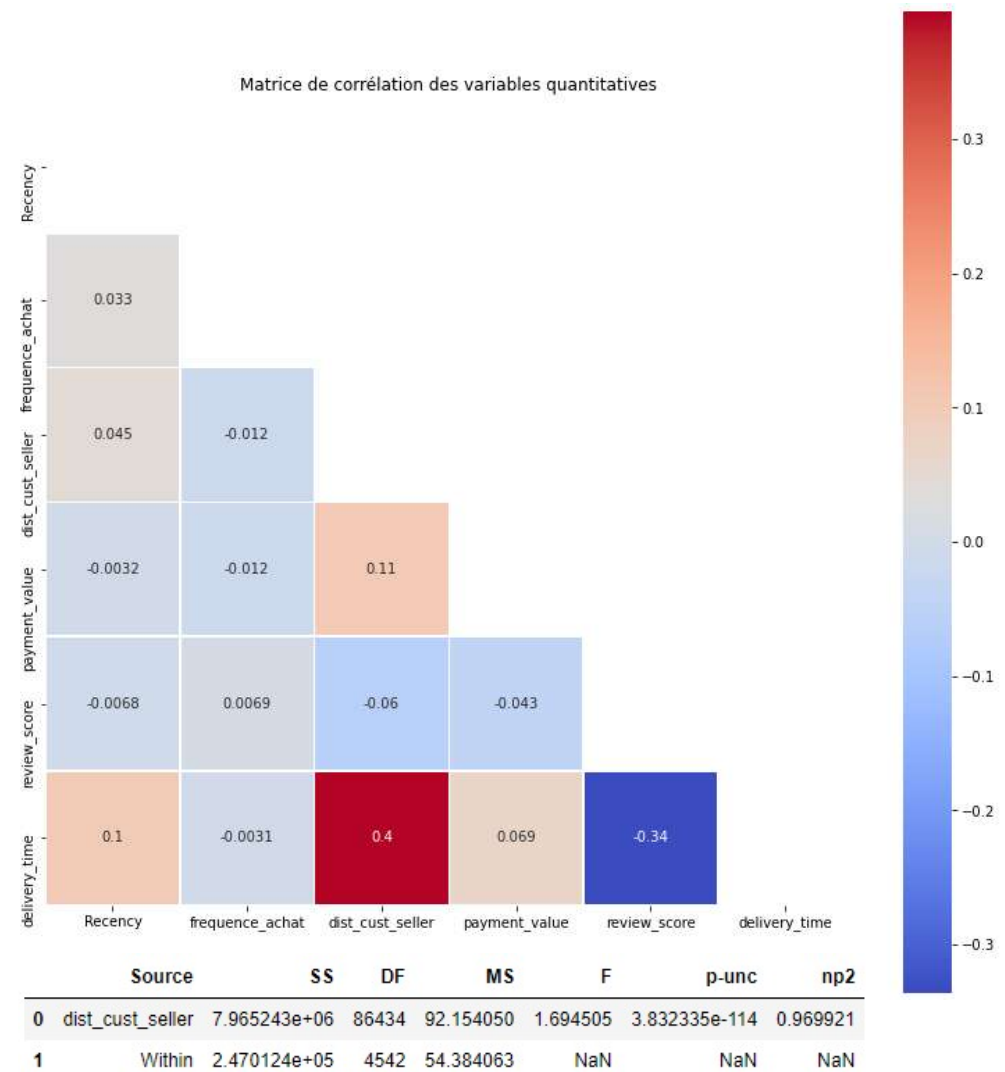
FEATURES

- RFM: Recency et Frequency
- Distance Customer-Seller
- Délais de livraison
- Taille villes
- Premium
- Catégories de produits

ANALYSE

_FEATURES

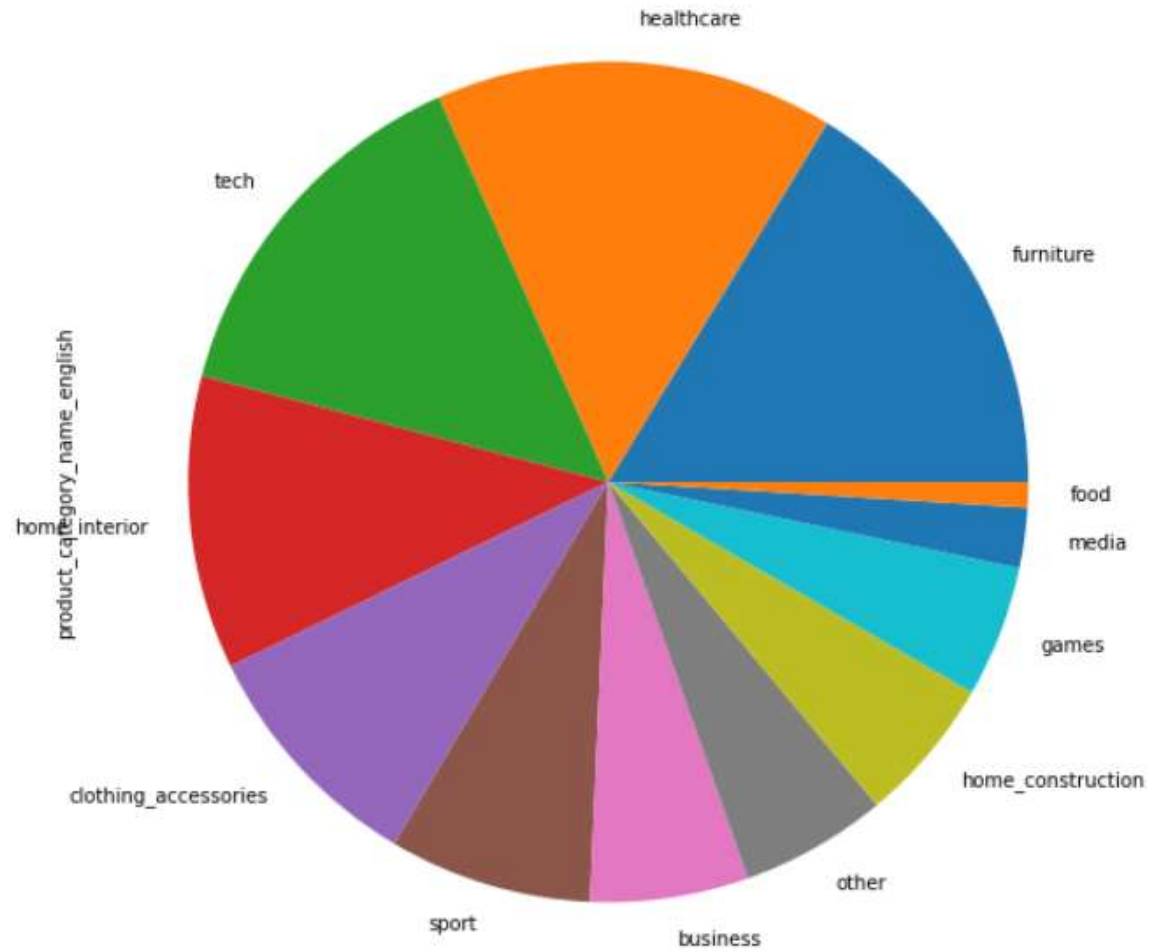
FEATURES



ANALYSE

_FEATURES

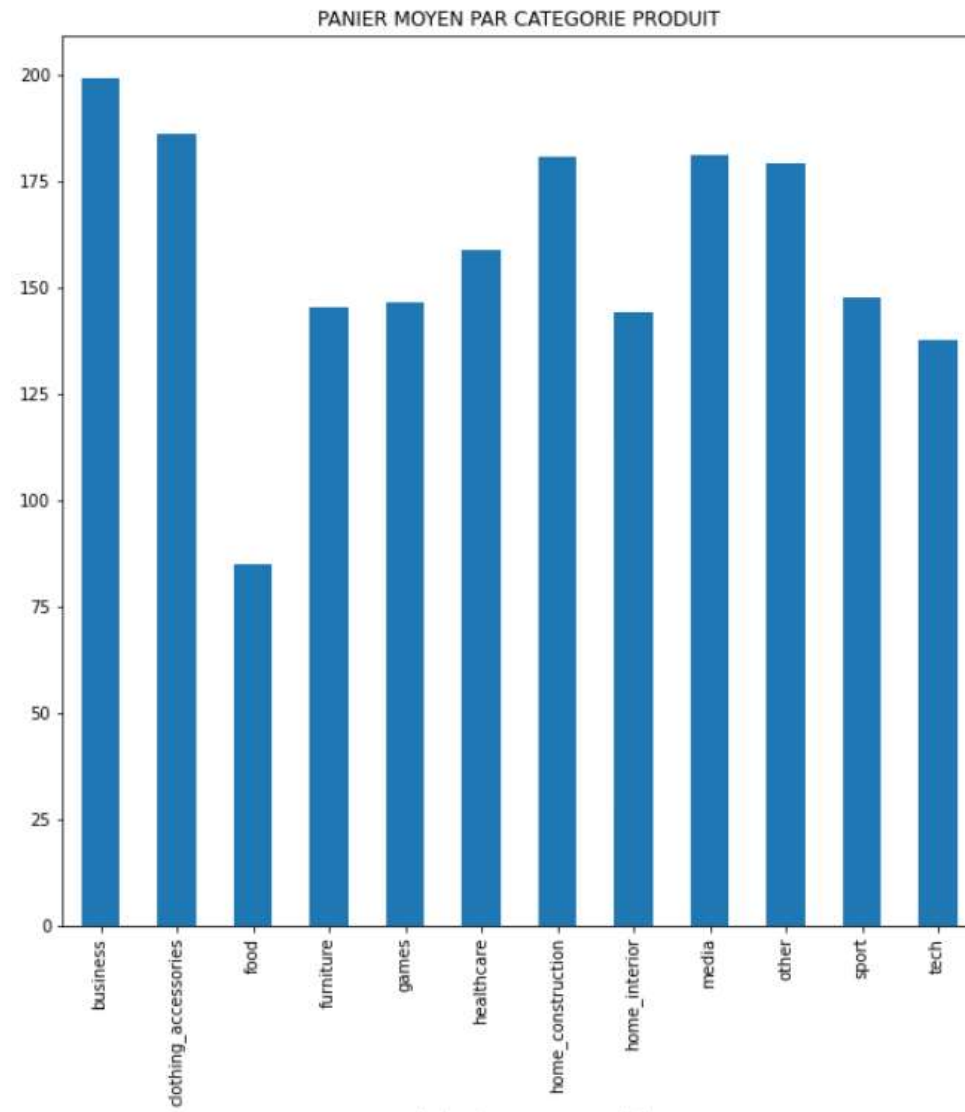
FEATURES



ANALYSE

_ FEATURES

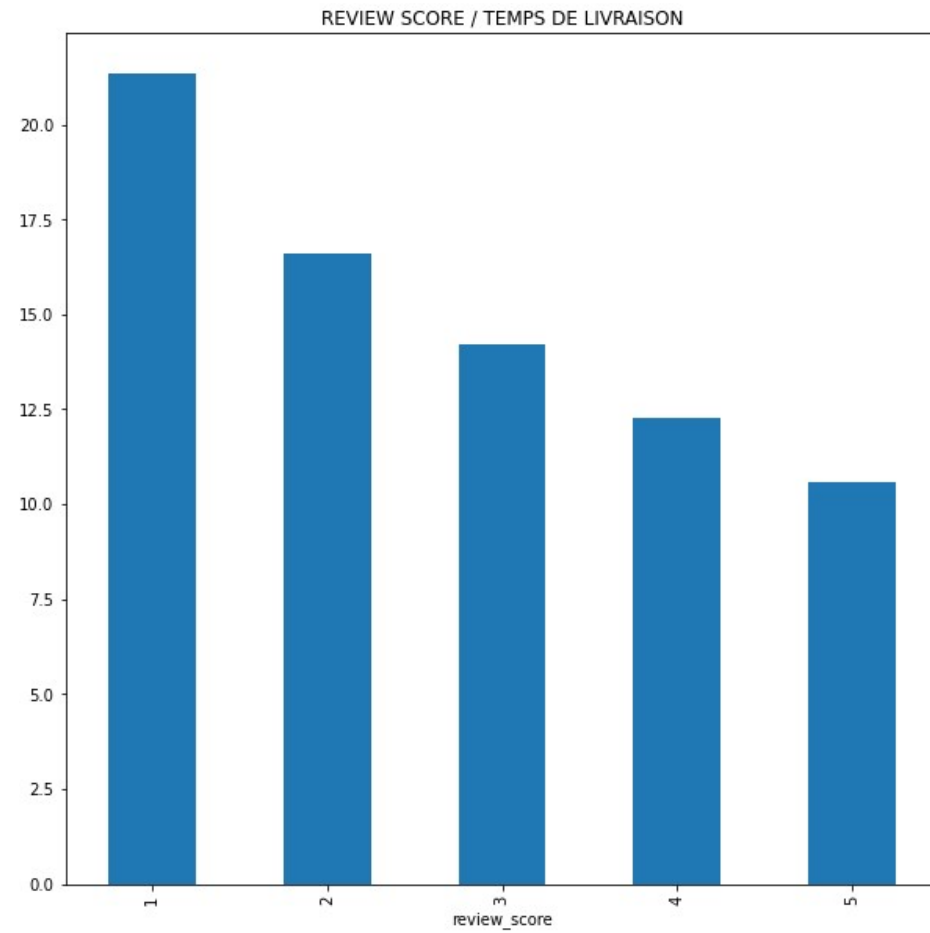
FEATURES



ANALYSE

_FEATURES

FEATURES



	Source	SS	DF	MS	F	p-unc	np2
0	review_score	9.590743e+05	4	239768.567403	3007.263319	0.0	0.116786
1	Within	7.253181e+06	90972	79.729821	NaN	NaN	NaN

MODÉLISATION

_RFM

FEATURES

- Recency
- Frequency
- Monetary Value

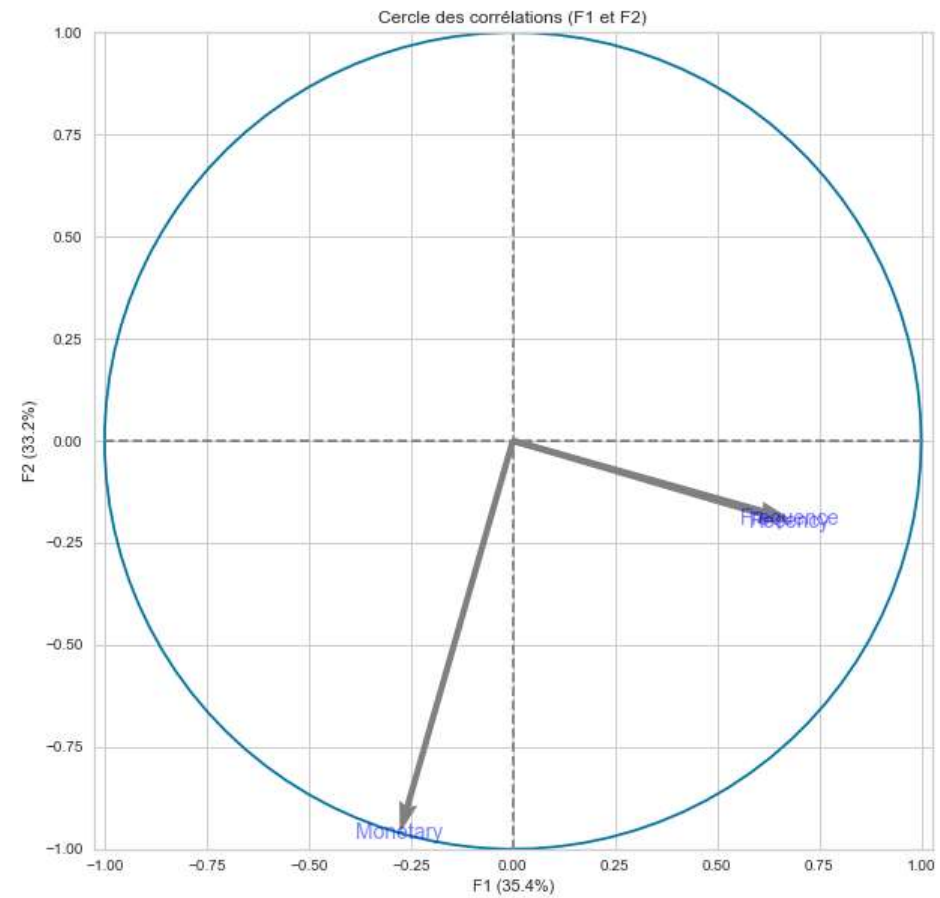
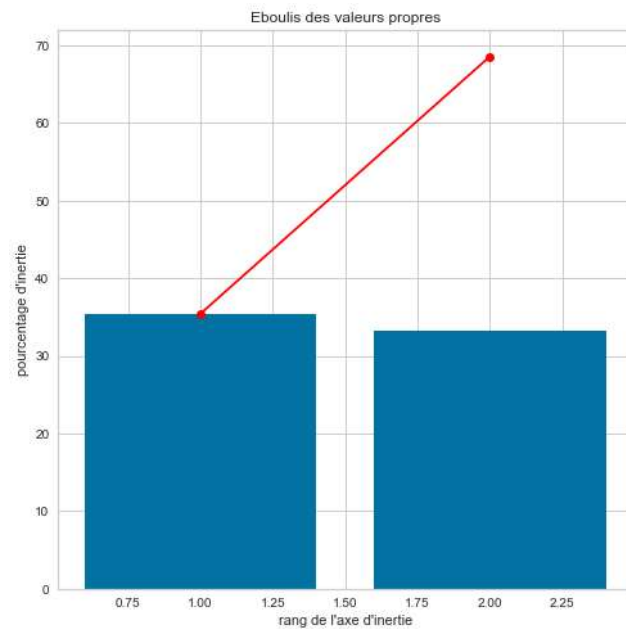
TRANSFORMEURS

- Passage au log
- StandardScaler

MODÉLISATION

_RFM

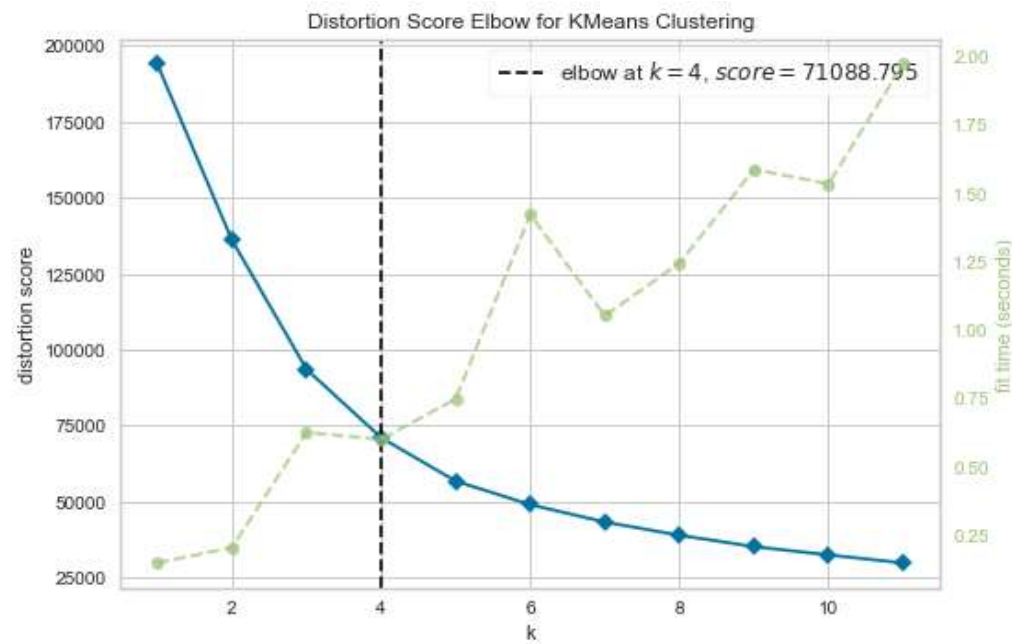
REDUCTION DES DIMENSIONS



Composants PCA

```
[[0.67890559 0.67996018 0.2770223 ]  
 [0.20008858 0.19167933 0.96084525]]
```


ELBOW METHOD

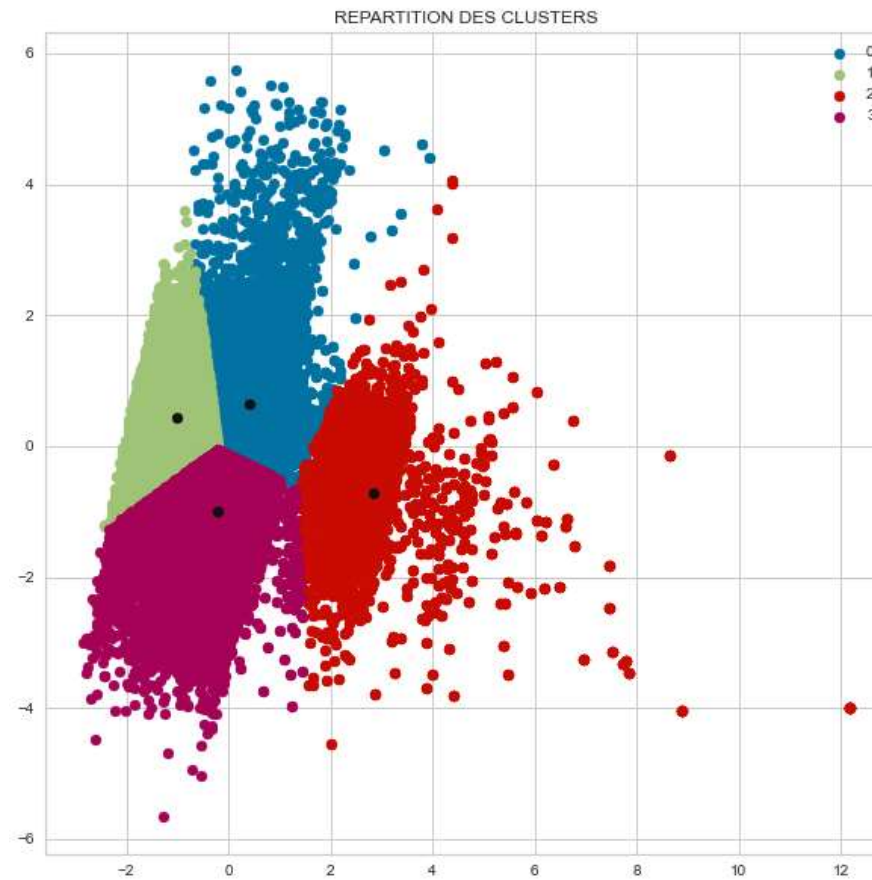


MODÉLISATION

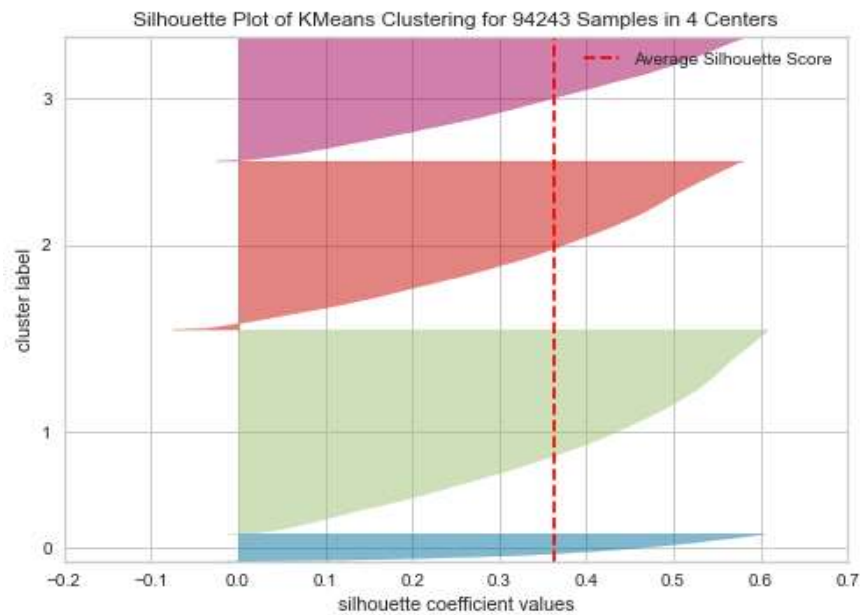
_RFM

MODELE

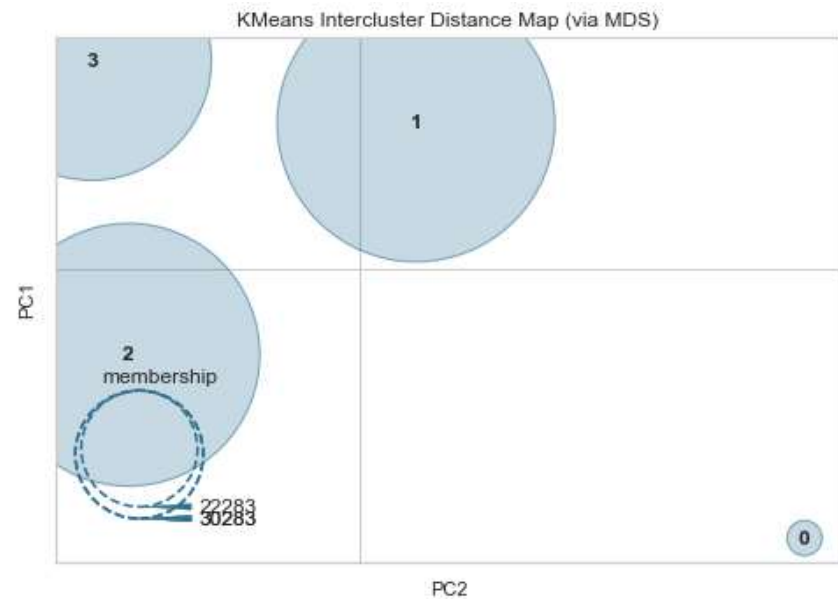
K-Means



EVALUATION DU MODELE



Score de silhouette: 0.36



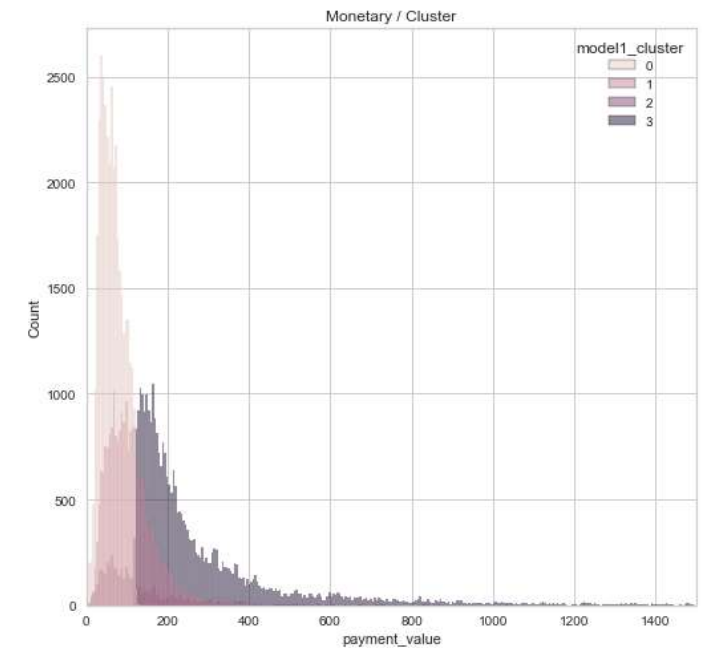
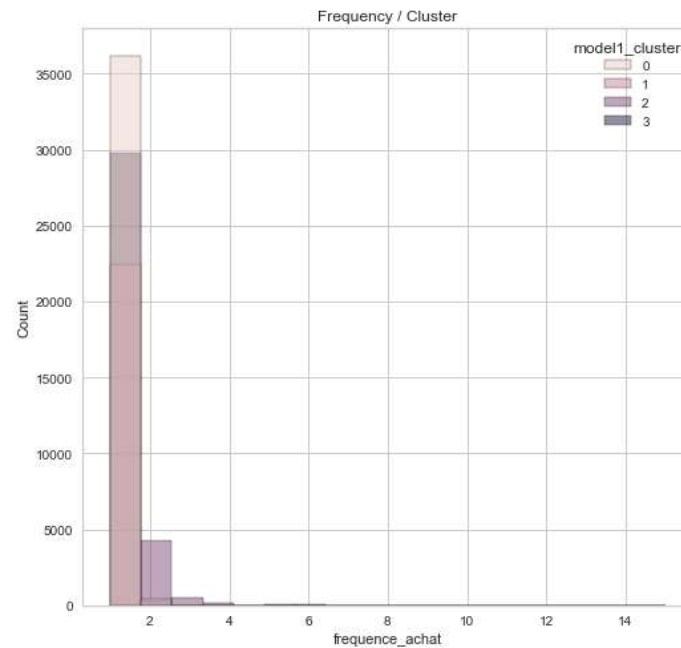
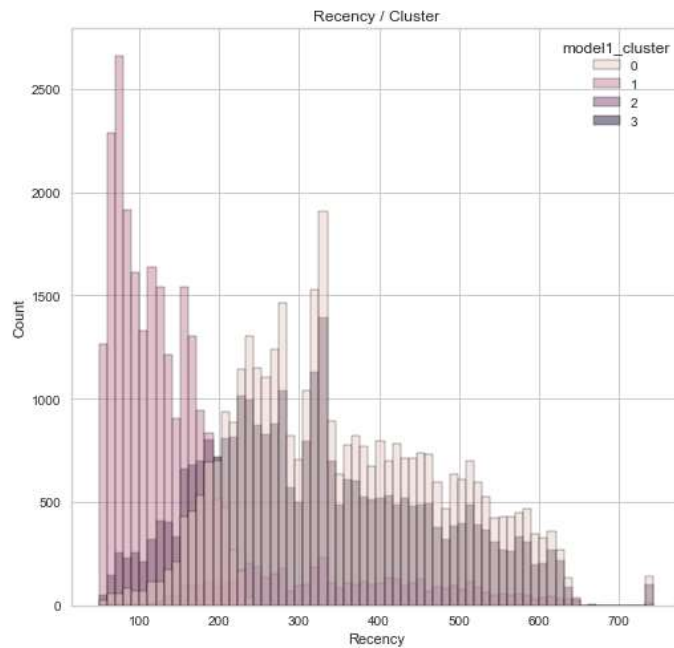
MODÉLISATION

_RFM

EVALUATION DU MODELE

	Recency	frequence_achat	payment_value
model1_cluster			
0	354.563454	1.012516	63.603252
1	117.262180	1.000000	108.726555
2	349.888735	2.279747	135.147765
3	324.055509	1.010670	311.003400

DISTRIBUTION FEATURES / CLUSTER



MODÉLISATION

_K-MEANS 5 FEATURES

FEATURES

- RFM
- Temps de livraison
- Distance Client-Distributeur

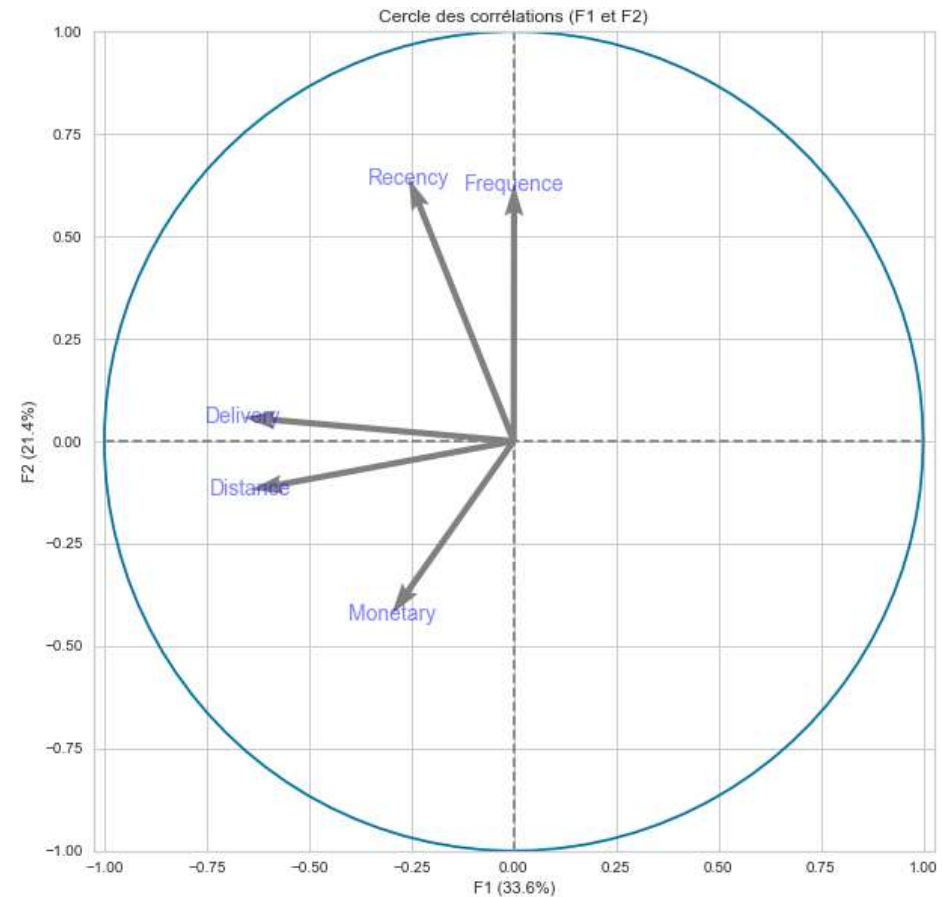
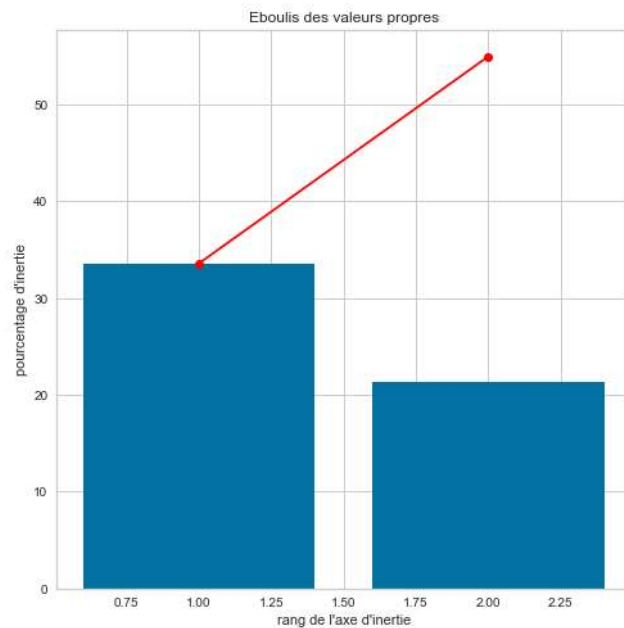
TRANSFORMEURS

- Passage au log
- StandardScaler

MODÉLISATION

_K-MEANS 5 FEATURES

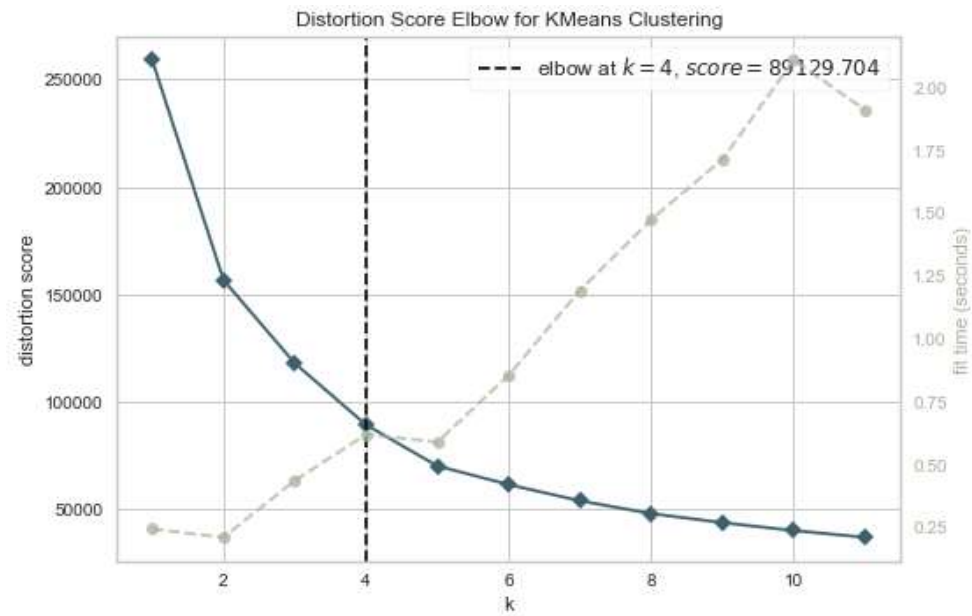
REDUCTION DES DIMENSIONS



Composants PCA

```
[ [0.25624541 0.00097914 0.29778893 0.65897217 0.64141622]  
  [0.64112765 0.62702005 0.422204 0.05822382 0.11897466]]
```

ELBOW METHOD

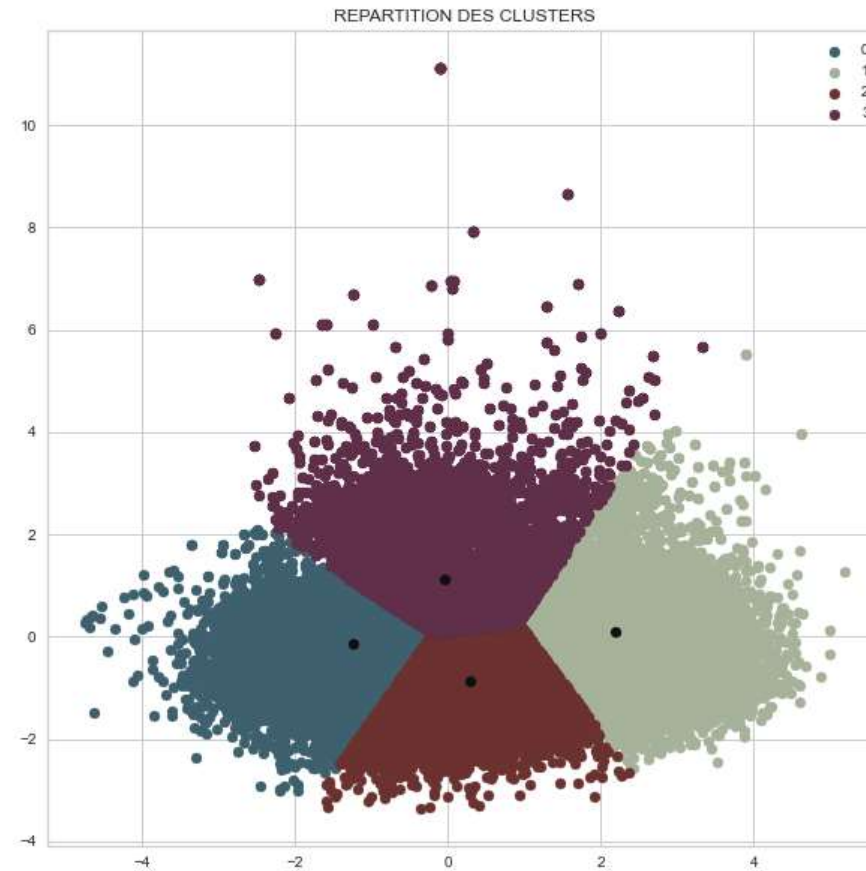


MODÉLISATION

_K-MEANS 5 FEATURES

MODELE

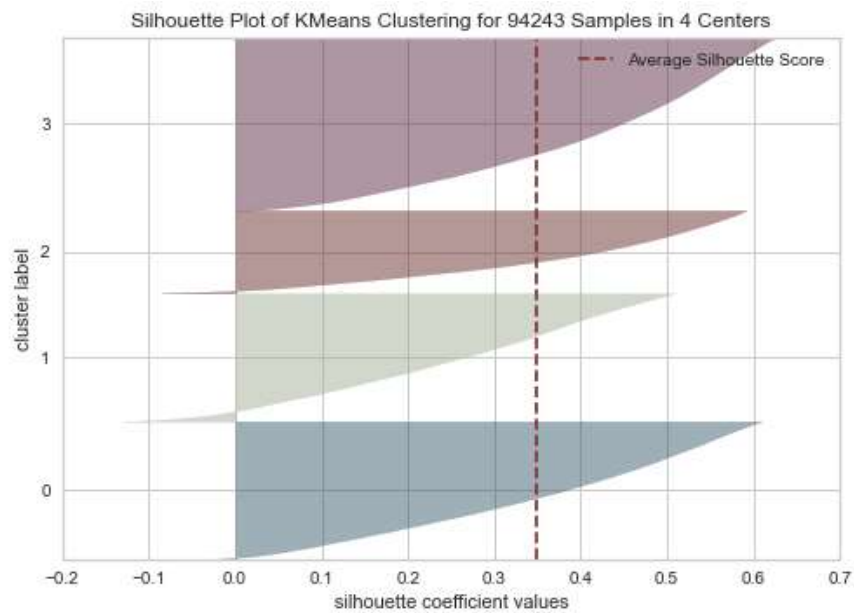
K-Means



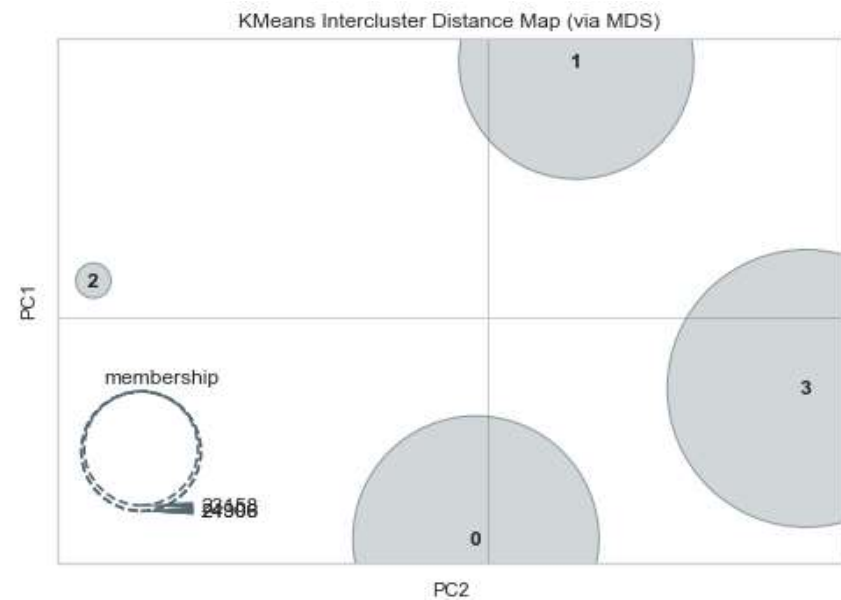
MODÉLISATION

_K-MEANS 5 FEATURES

EVALUATION DU MODELE



Score de silhouette: 0.34



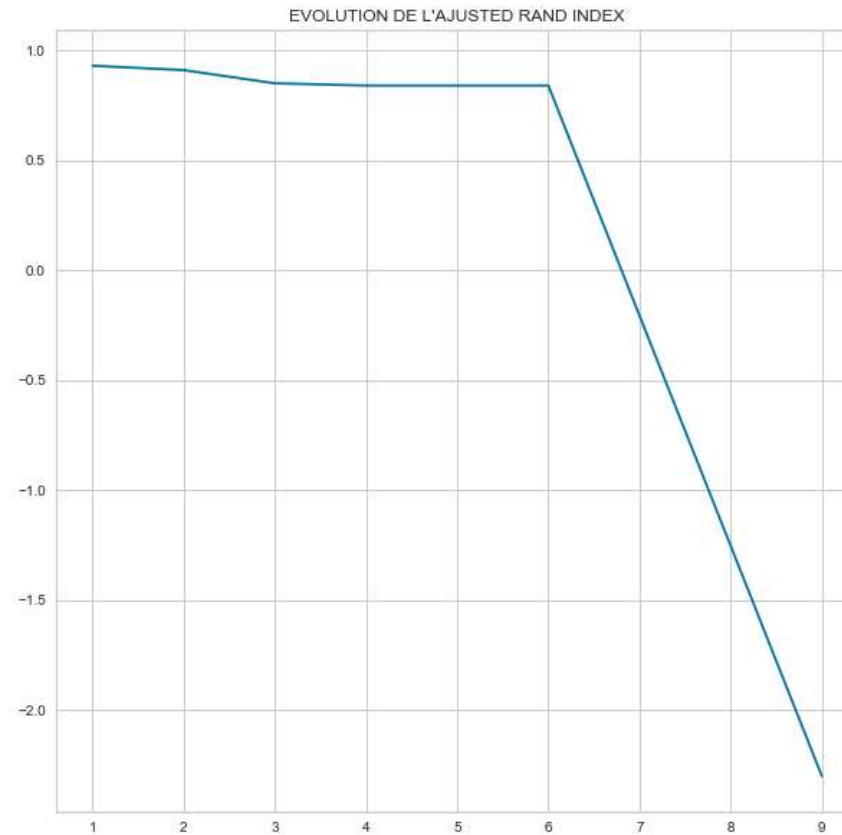
MODÉLISATION

_K-MEANS 5 FEATURES

EVALUATION DU MODELE

	Recency	frequence_achat	payment_value	delivery_time	dist_cust_seller
model2_cluster					
0	328.393701	1.005423	216.107346	19.988303	1031.125790
1	238.348901	1.042013	88.132607	4.465358	55.344151
2	160.369533	1.000237	187.130246	8.796686	506.113109
3	408.762323	1.288537	85.569288	11.218229	461.648487

STABILITÉ DU MODELE



FEATURES

- 5 Dernières features
- Avis
- Premium
- Tailles villes
- Catégories de produits

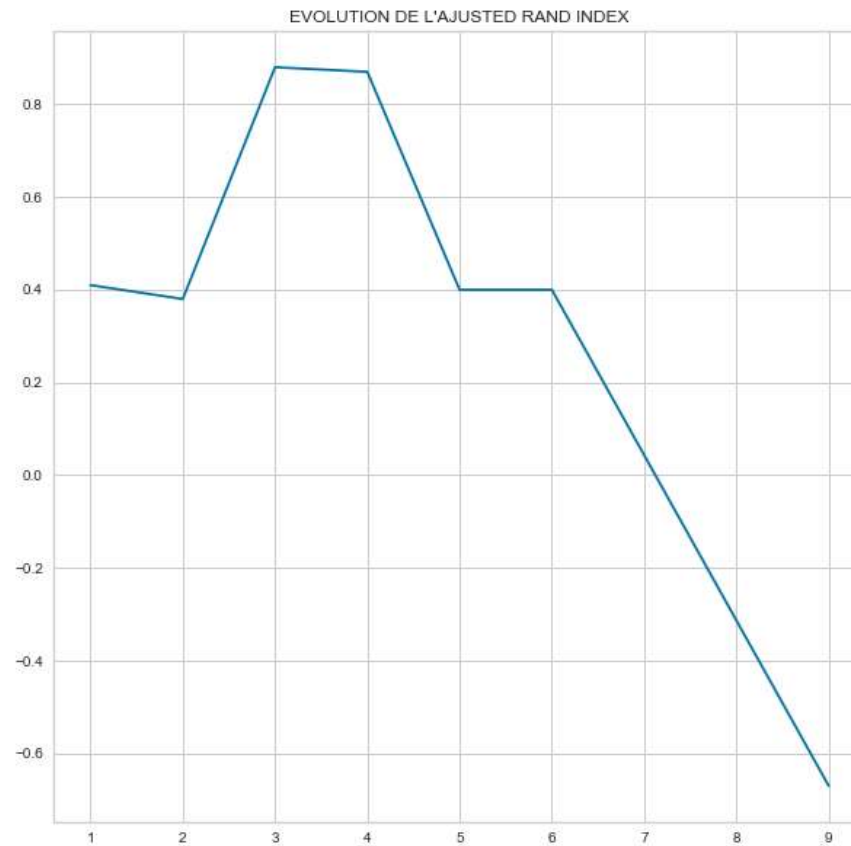
TRANSFORMEURS

- Passage au log
- StandardScaler

EVALUATION DU MODELE

	Recency	frequence_achat	payment_value	delivery_time	dist_cust_seller	review_score	m_quartile	customer_city	product_category_name_english
model3_cluster									
0	292.373417	1.082370	138.320494	12.822972	623.079042	4.163812	2.663867	1.288395	furniture
1	282.484481	1.074054	120.107030	10.525592	406.738241	4.185861	3.038479	3.832471	healthcare
2	282.632392	1.077412	288.876758	13.138745	699.731599	4.063529	1.591765	1.695686	tech
3	287.996199	1.063204	156.406431	13.633587	742.691712	4.164555	1.969595	0.847480	healthcare

STABILITÉ DU MODELE



CONCLUSION

_PERSONAS DU K-MODES



Le/La Geek:

- Prix élevé: 290 Réaux
- Exigeant: note 4.0
- Villes moyennes
- Intérêt: Tech



Le/La Citadin(e):

- Prix bas: 120 Réaux
- Conciliant: note 4.2
- Grandes villes
- Intérêt: Santé
- Proche distributeurs
- Délais cours: 10j



Le/La ménagère(e):

- Prix moyen: 140 Réaux
- Conciliant: note 4.15
- Petites villes
- Intérêt: Fournitures



L'adèpte des produits cosmétiques:

- Prix moyen: 160 Réaux
- Conciliant: note 4.15
- Petites villes
- Intérêt: Cosmétiques

CONCLUSION

_BILAN

- **Modélisation:**
 - K-Modes donne de bonnes idées de segmentations (var. catégorielles)
 - Maintenance tous les 6 mois
- **4 Personas identifiés**
- **Permet de prévoir des campagnes marketing ciblées**
- **Segmentation RFM insuffisante**

CONCLUSION

_BILAN

AXES D'AMÉLIORATIONS

- **Obtenir plus d'informations discriminantes:**
 - Age, sexe, centres d'intérêt, métier
 - Localisation plus précise (distance au centre ville)
- **Connaitre les comportements de navigation:**
 - Utiliser des cookies
 - Pages web visitées
 - Cheminement jusqu'à l'achat
- **Affiner les avis/commentaires – Renforcer les personas**
 - Envoi d'un questionnaire de satisfaction après achat
 - Coupler avec Facebook pour récupérer des données clients