



OLIST

oooo

oooo

olist
store

NOTRE OBJECTIF

L'objectif est de soutenir l'équipe marketing d'Olist pour augmenter l'efficacité de leurs campagnes de communication. Pour y parvenir, il est nécessaire de créer une segmentation des clients en utilisant des techniques de clustering basées sur un jeu de données fourni par Roxanne, qui doit représenter le comportement des clients. Nous allons effectuer une analyse exploratoire pour essayer de trouver les tendances globales.

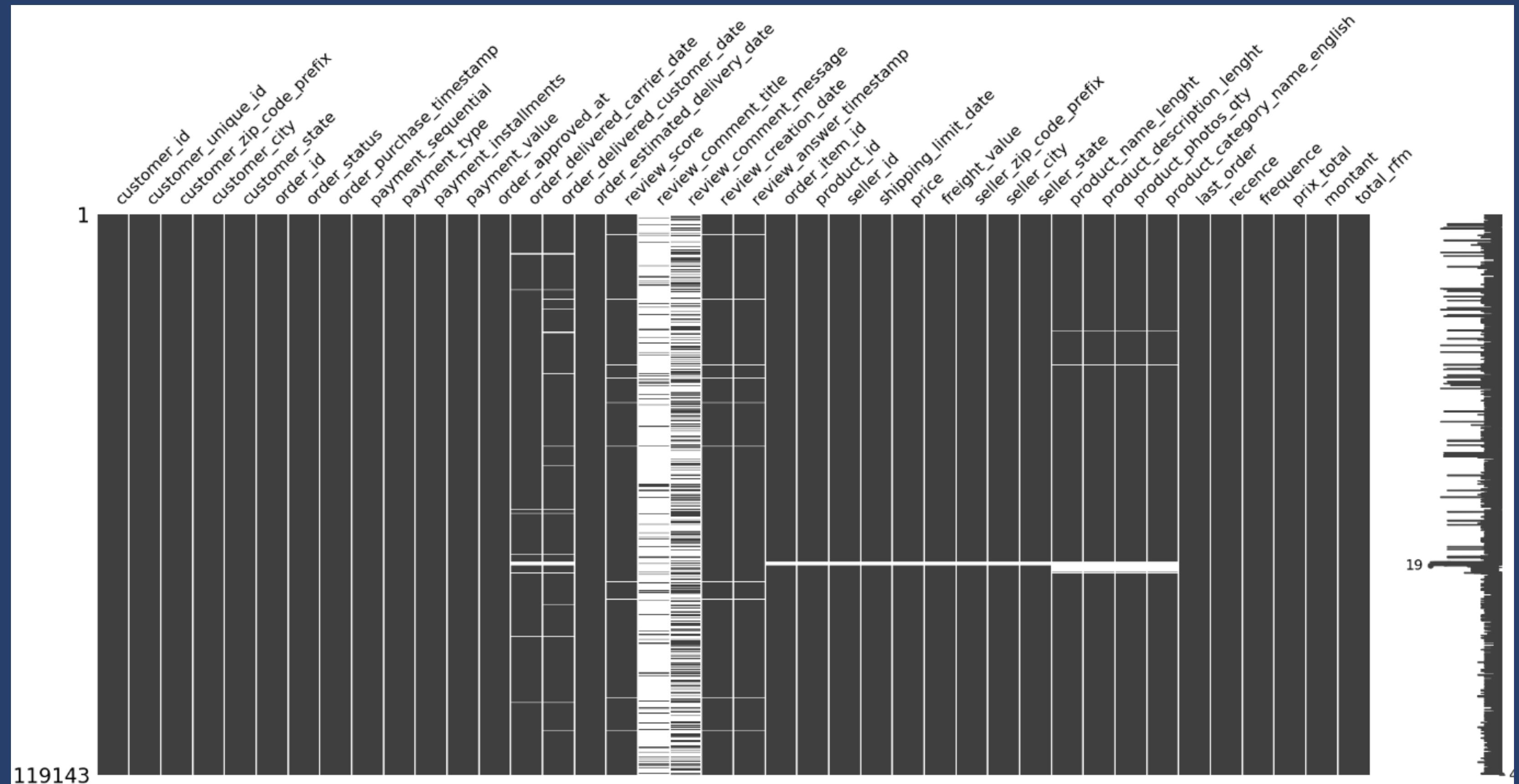


PLAN DE LA PRESENTATION

- Comprendre les objectifs
 - Nettoyage des données
 - Analyse exploratoire
 - Modélisation
 - Code Review
 - Présentation des clusters
 - Recommandations marketing
 - Conclusion
 - Difficultés rencontrées

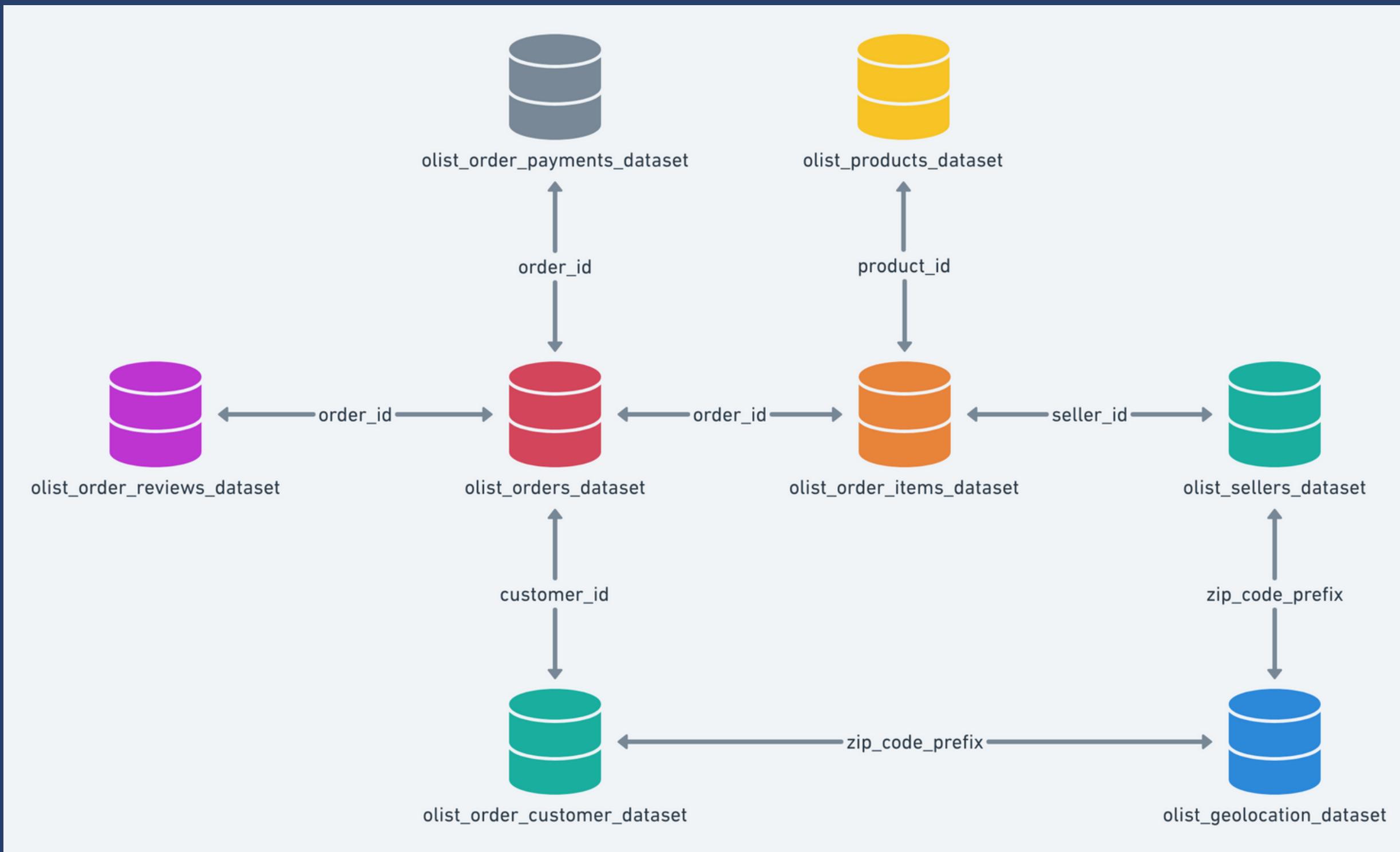


NETTOYAGE DES DONNÉES



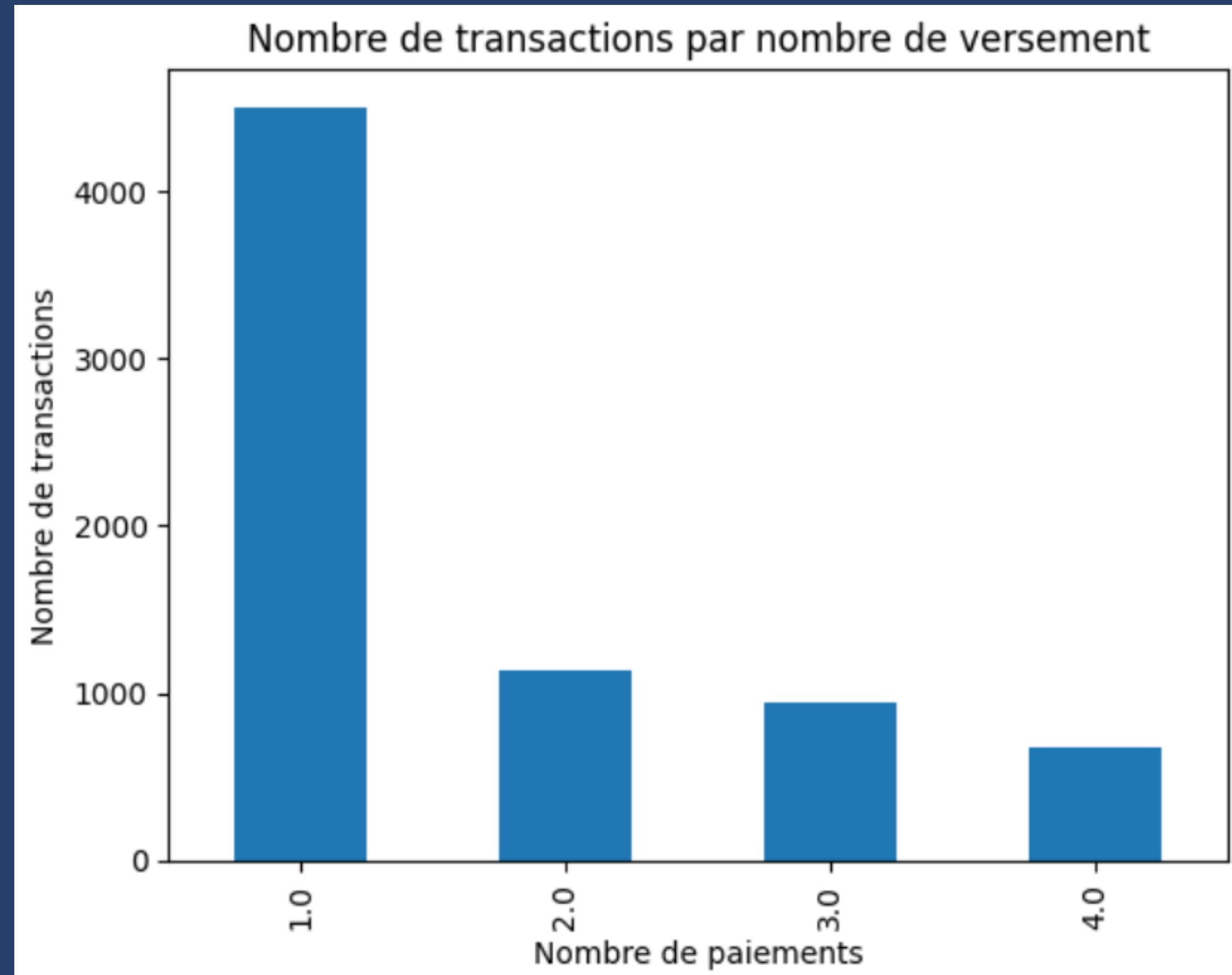
ANALYSE EXPLORATOIRE

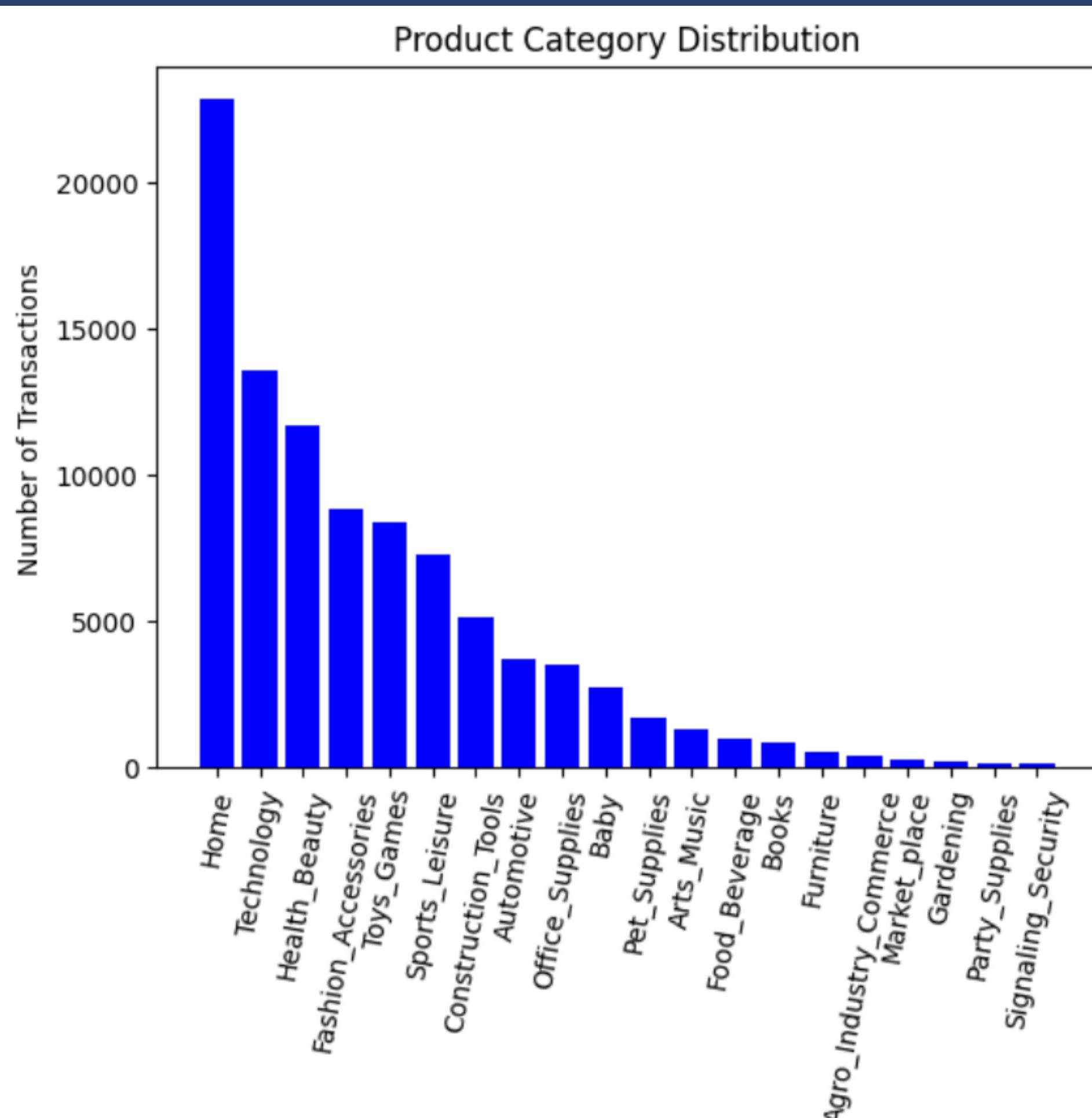




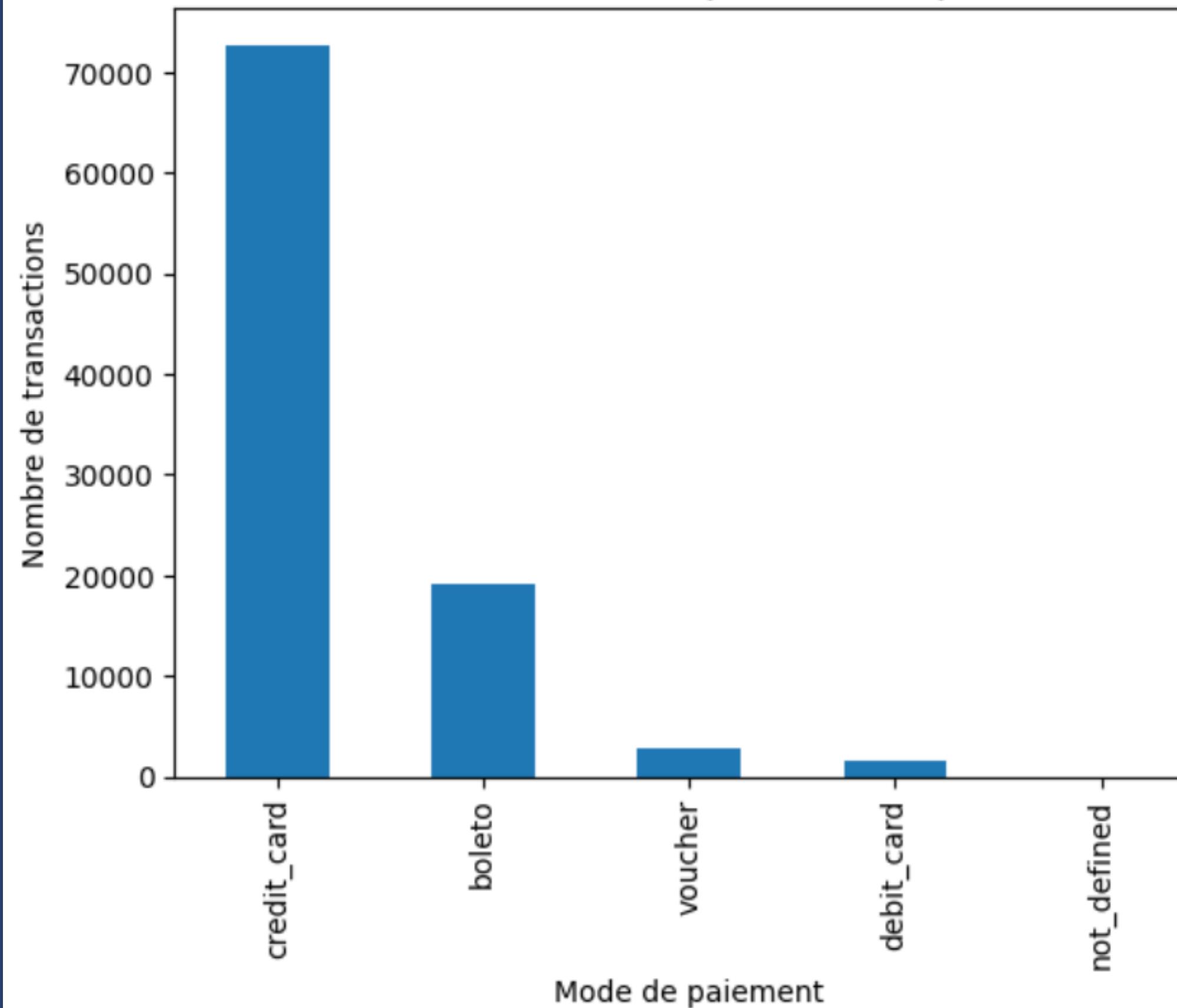
payment_installments

1.0	59446
2.0	13838
3.0	11889
4.0	8072
10.0	6976
5.0	6097
8.0	5120
6.0	4674
7.0	1848
9.0	739
12.0	168
15.0	93
18.0	38
24.0	34
11.0	26
20.0	21
13.0	19
14.0	16
17.0	8
16.0	7
21.0	6
0.0	3
23.0	1
22.0	1

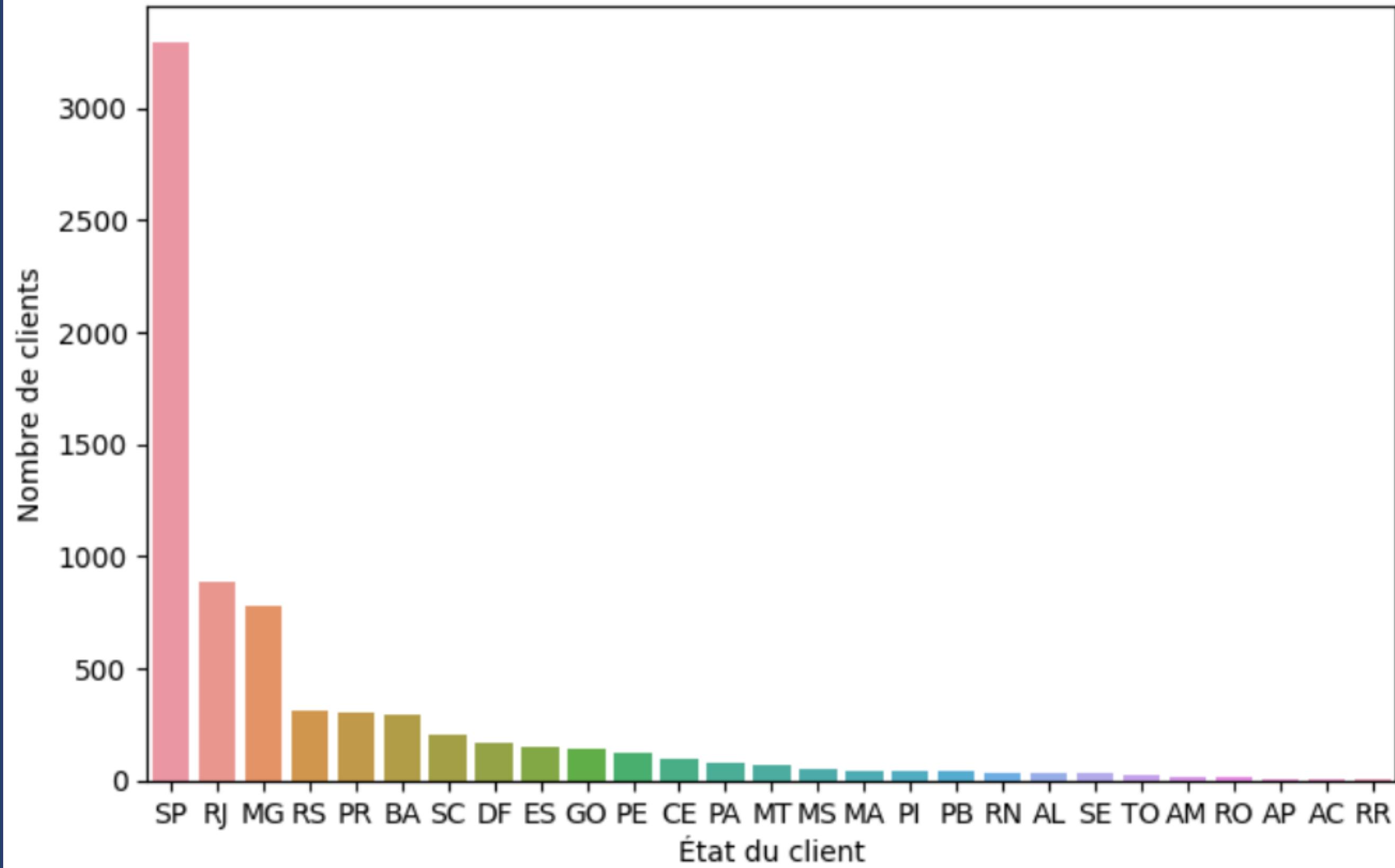




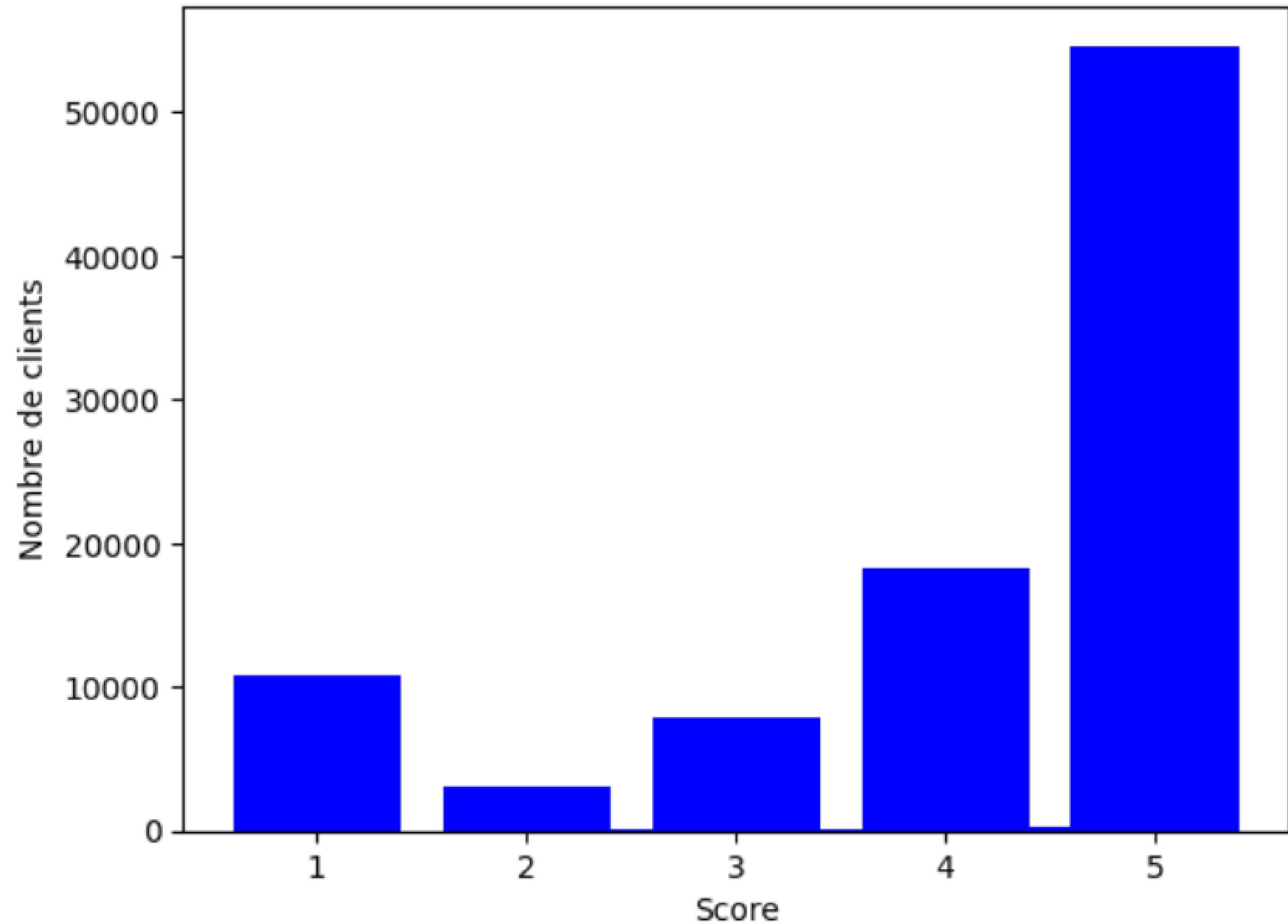
Nombre de transactions par mode de paiement



Nombre de clients par état



Distribution des scores



count	7251.000000
mean	3.927410
std	1.497337
min	1.000000
25%	3.000000
50%	5.000000
75%	5.000000
max	5.000000

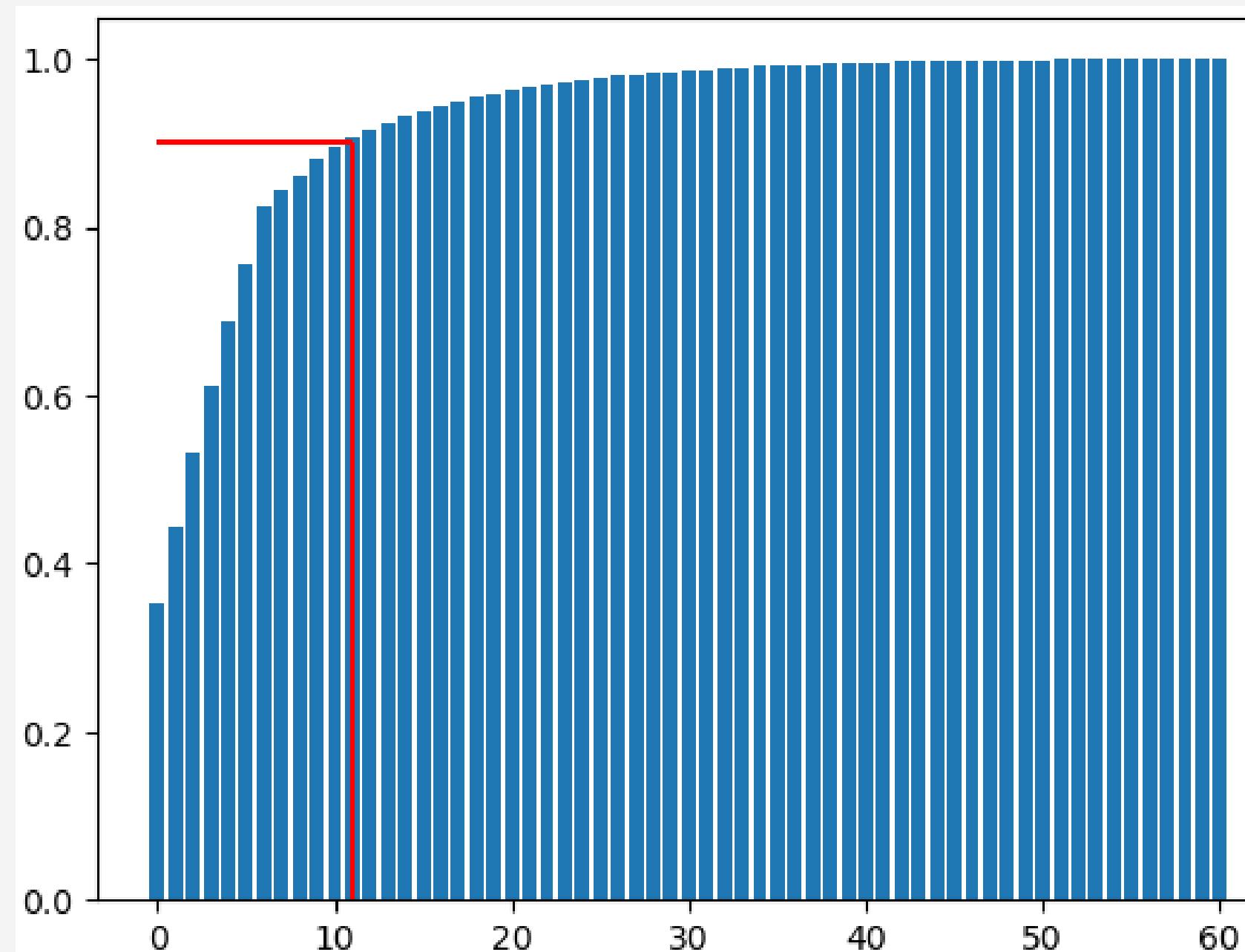
ooo **MODELISATION**

- Nous allons tester l'algorithme de clustering Kprototypes en ajustant le nombre de cluster via la méthode du coude . Pour m'assurer de la qualité de la segmentation, je vais évaluer mon modèle à l'aide de métriques telles que le score silhouette.
- Il évalue à quel point les observations d'un même cluster sont similaires entre elles, et à quel point elles sont différentes de celles des autres clusters.

ooo

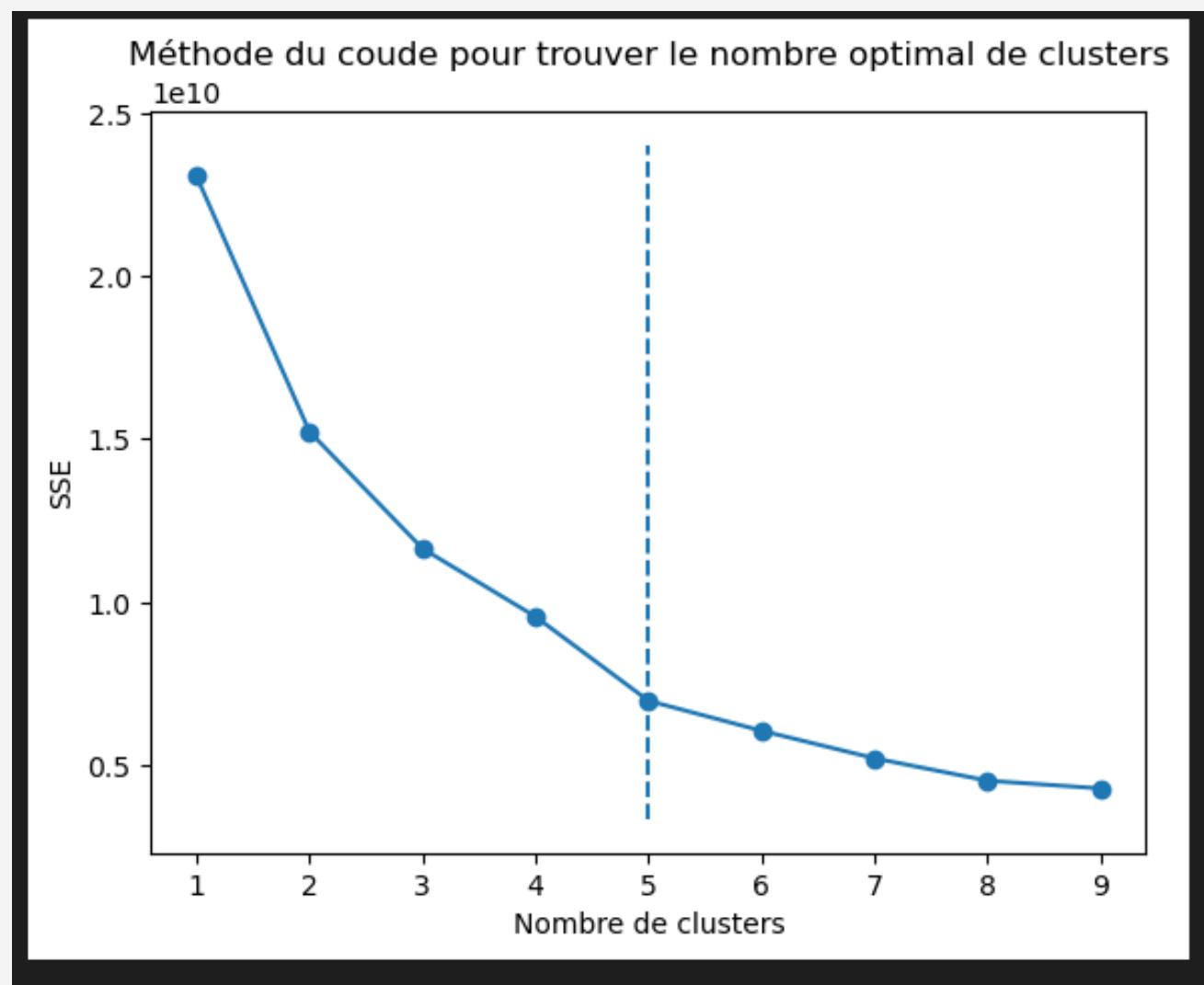
ooo MODELISATION

- Etude de la réduction des dimensions



- Voici le résultat de notre PCA , nous pouvons décider de nous concentrer uniquement sur une dizaine de données pour notre clusterisation.

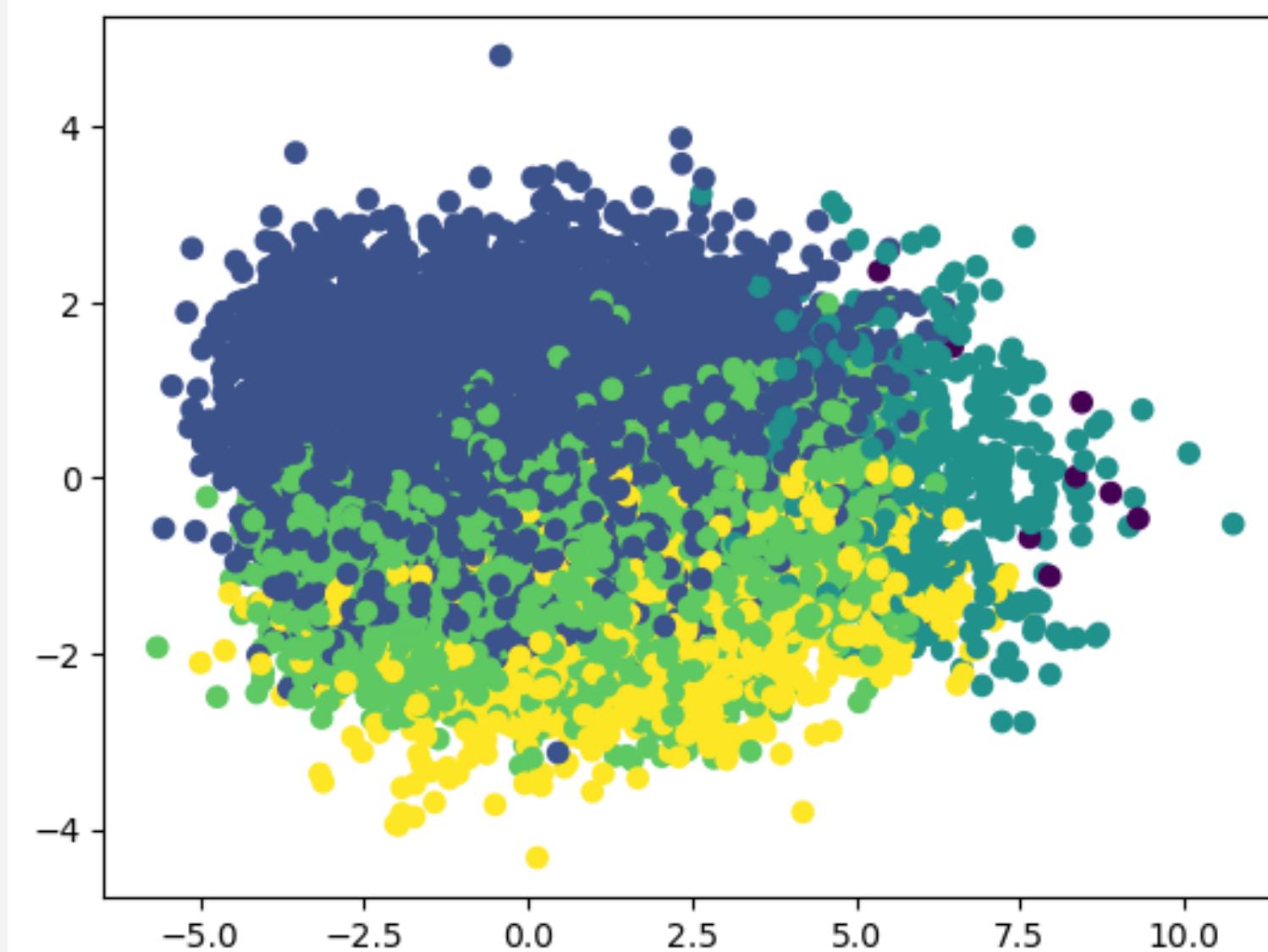
ooo MODELISATION



- En appliquant une méthode du coude , nous arrivons à trouver notre nombre de clusters. Après etude de cette méthode et certaines considérations , nous allons choisir 5 clusters pour notre clustering , après avoir verifier sur notre graphique.



ooo MODELISATION



- Voici une representation des clusters que nous avons obtenu.
- En etudiant le silhouette score , on peut determiner la validité de nos clusters.
- Le silhouette score est positif, donc nos clusters sont relativement séparés. Cependant notre score n'est pas élevé, ce qui revèle que nos clusters ne sont pas très fiable. Il faudra retourner .

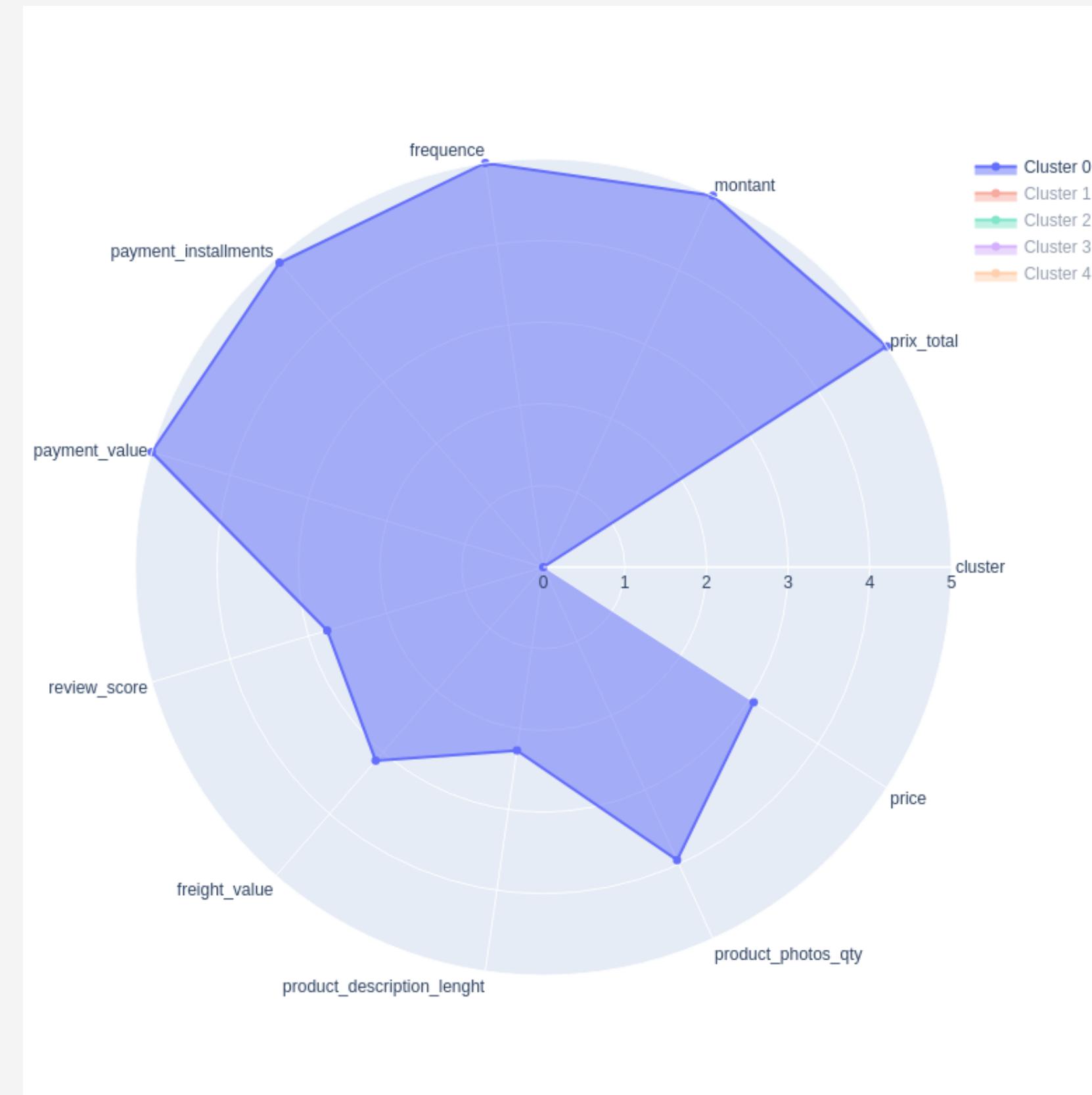
Silhouette score: 0.07439759259782465

	customer_state	payment_type	product_category_name_english	prix_total	montant	frequence	payment_installments	payment_value	review_score	freight_value	product_description_lenght	product_photos_qty	price
cluster													
0	SP	credit_card	Technology	2919.40	58.0	3	5.800000	19380.811000	2.300000	30.883583	1150.800000	2.033333	584.090833
1	SP	credit_card	Home	99.90	2.0	0	2.810180	154.921216	4.112872	18.712897	416.694933	2.095386	96.508393
2	SP	credit_card	Home	1957.72	39.0	3	5.877573	2386.771239	3.747228	49.241493	1055.968781	2.242644	953.399650
3	SP	credit_card	Home	99.00	2.0	3	2.889033	188.563197	4.121845	20.906558	1100.600115	2.483536	128.216743
4	SP	credit_card	Health_Beauty	59.99	1.0	3	3.163755	248.774433	4.167951	23.644580	2536.927799	2.577106	189.309048

3	16871
1	8545
2	2141
0	452
4	10

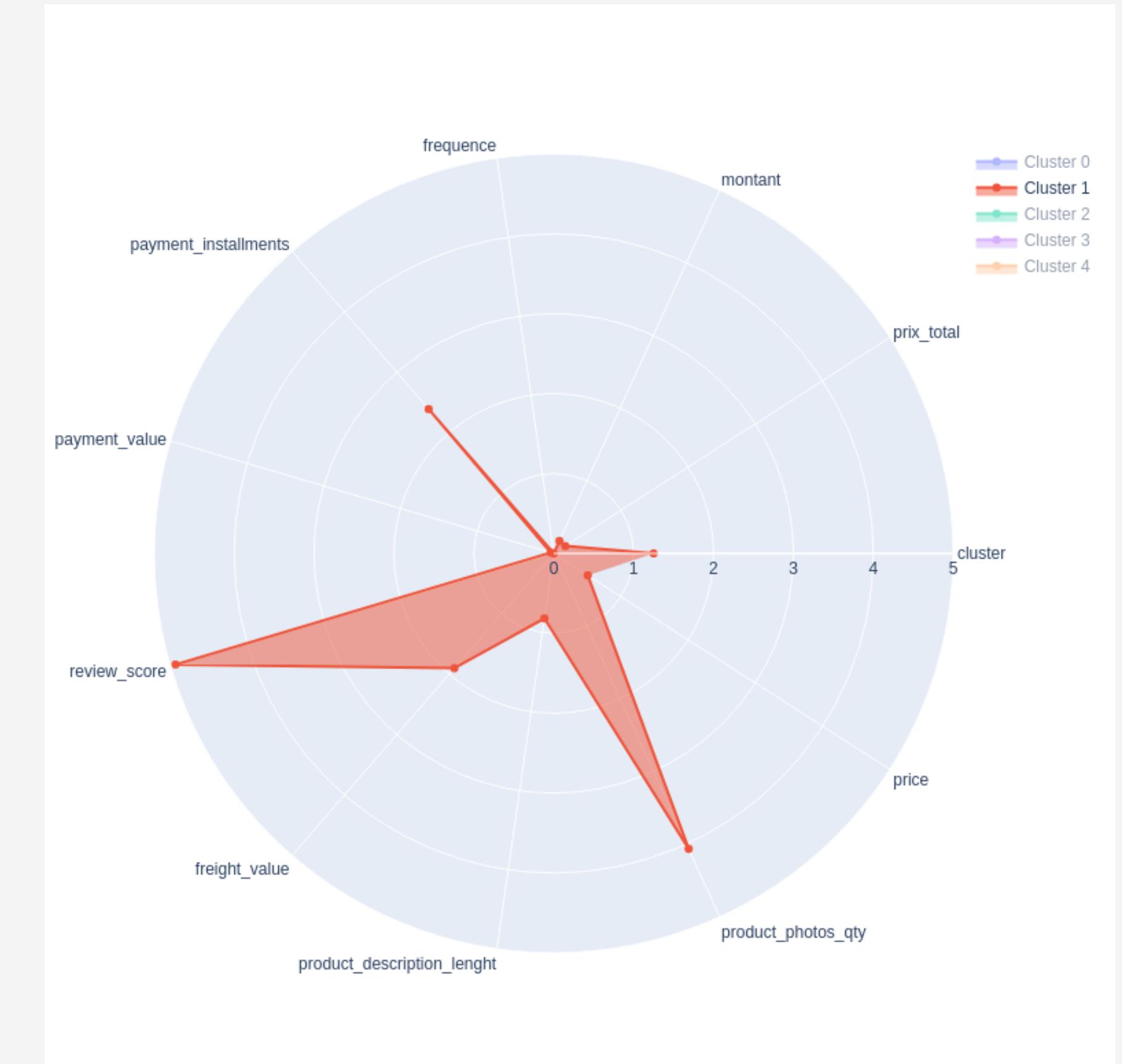
PRESENTATION DES CLUSTERS

0° CLUSTER :



PRESENTATION DES CLUSTER

1^o CLUSTER :



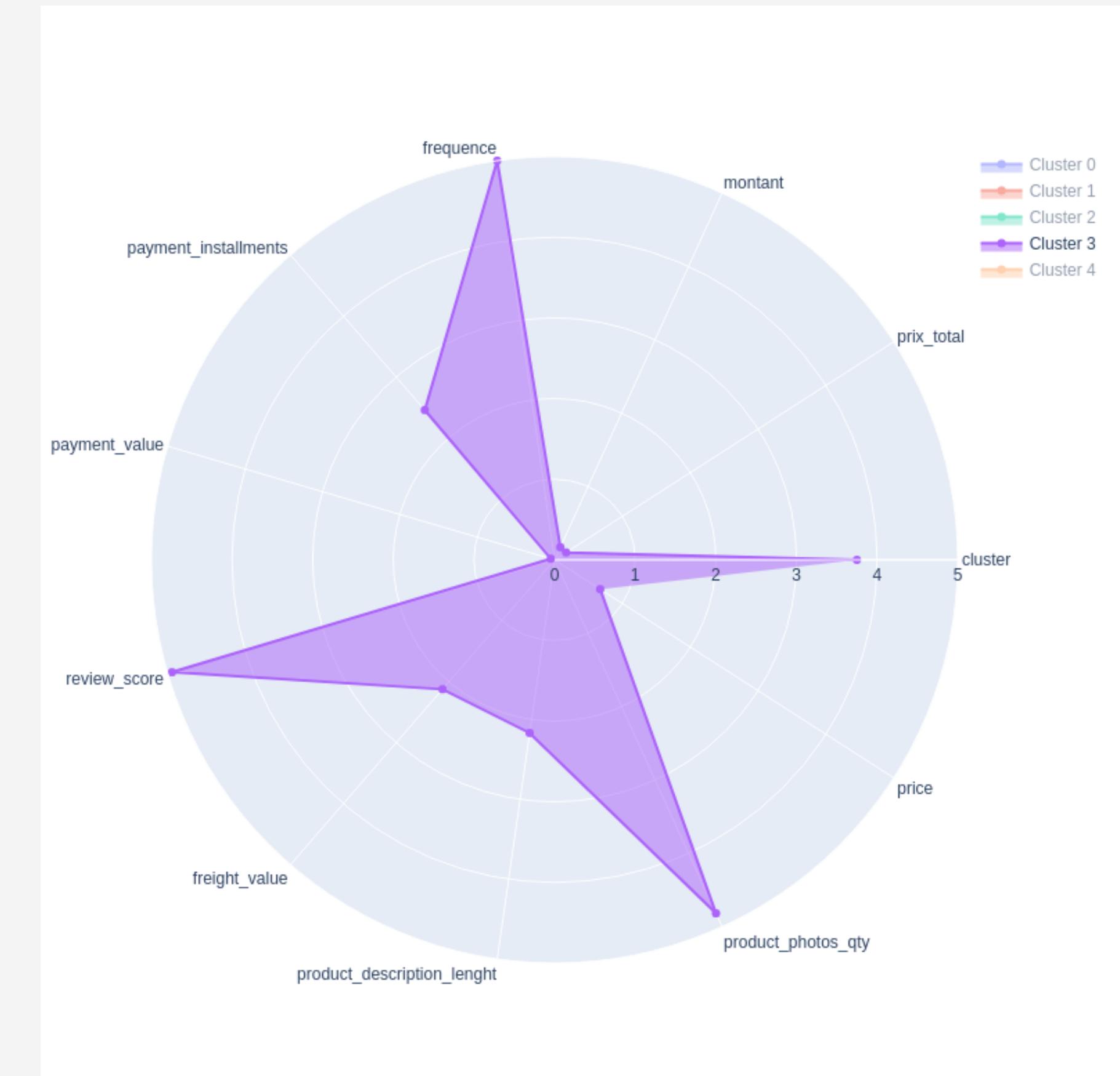
PRESENTATION DES CLUSTER

2° CLUSTER :



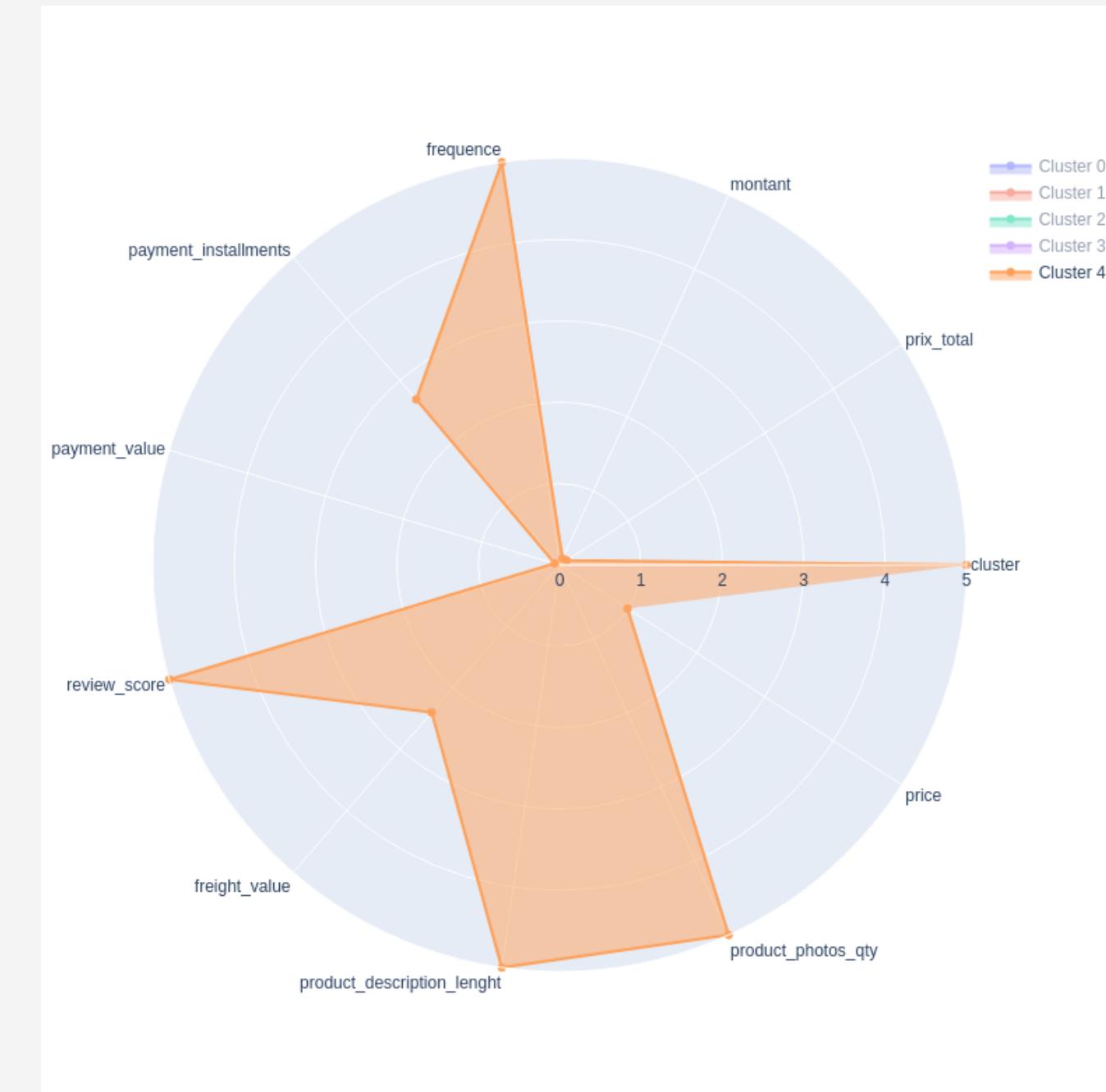
PRESENTATION DES CLUSTER

3° CLUSTER :



PRESENTATION DES CLUSTER

4° CLUSTER :



0 0 0

RECOMMENDATIONS MARKETING

**APRÈS ÉTUDE DES CLUSTERS , JE CONSIDERERAIS REVOIR
NOTRE STRATÉGIE CONCERNANT LE CLUSTER N°1 ,
SACHANT QUE C'EST LA LIGNE PRINCIPAL DE
L'ENTREPRISE , LE REVIEW SCORE EST TROP BAS ET PEUT
IMPACTER NÉGATIVEMENT L'IMAGE DE L'ENTREPRISE. LE
RESTE DES CLUSTERS MONTRENT UNE BONNE IMAGE ,
REPRÉSENTANT LE BON TRAVAIL DE L'ENTREPRISE**

**JE CONSIDERERAI AUSSI UNE ÉTUDE SUR LES MÉNAGES ,
QUI SEMBLE ÊTRE APRÈS ÉTUDES LA PARTIE MAJEUR DE
NOTRE CLIENTÈLE BIEN QUE NE PAYANT PAS AUTANT ,
NOTRE IMAGE DE MARQUE EST PRINCIPALEMENT
APPUYÉ PAR CETTE CATÉGORIE DE LA POPULATION**

CODE REVIEW

	customer_state	payment_type	product_category_name_english	prix_total	montant	frequence	payment_installments	payment_value	review_score	freight_value	product_description_lenght	product_photos_qty	price
cluster													
0	SP	credit_card	Technology	2919.40	58.0	3	5.800000	19380.811000	2.300000	30.883583	1150.800000	2.033333	584.090833
1	SP	credit_card	Home	99.90	2.0	0	2.810180	154.921216	4.112872	18.712897	416.694933	2.095386	96.508393
2	SP	credit_card	Home	1957.72	39.0	3	5.877573	2386.771239	3.747228	49.241493	1055.968781	2.242644	953.399650
3	SP	credit_card	Home	99.00	2.0	3	2.889033	188.563197	4.121845	20.906558	1100.600115	2.483536	128.216743
4	SP	credit_card	Health_Beauty	59.99	1.0	3	3.163755	248.774433	4.167951	23.644580	2536.927799	2.577106	189.309048

3	16871
1	8545
2	2141
0	452
4	10

OOOO

DIFFICULTÉS RENCONTRÉES ET PISTES D'AMÉLIORATION

**NOUS AURIONS PU SE
CONCENTRER DAVANTAGE SUR
DES ALGORITHMES DE REDUCTION
DES DONNÉES AFIN D'AMÉLIORER
LES ALGORITHMES.**

**NOUS POUVONS AMÉLIORDER
NOS CLUSTERS PAR LA
RECHERCHE D'ANOMALIES.**

**NOUS POUVONS ESSAYER DE
VISUALISER SUR UNE CARTE DU
BRESIL LA CONCENTRATION DE
NOS ACHETEURS POUR ADAPTER
NOTRE POLITIQUE.**



OOOO

○ ○ ○ ○

THANK
YOU