

Analyse de la segmentation clients

Présenté par : NEQQAZ Wissal
BELHAID Nassrou-eddine

Filière : Data Science & Cloud Computing

CASA

Plan de présentation



■ I-Introduction

■ II-Processus de préparation
des données

■ III-1er traitement :
segmentation

■ IV-2ème traitement :
recommandation
prédiction

■ Conclusion

I-Introduction

CONTEXTUALISATION

- Définition du segmentation des clients
- Secteurs d'activités
- L'utilité de la data science à ce propos

Problématique

- Comment identifier les caractéristiques clés des clients qui peuvent être utilisées pour segmenter efficacement la clientèle ?

1

2

LA PRESENTATION DES ENJEUX

- Mieux comprendre les besoins de chaque groupe de clients
- Améliorer la satisfaction et fidéliser les clients

3

4

les etapes du segmentations

- Importer la base des données
- Nettoyage des données
- Visualisation des données
- Traitement de la data
- Suggérer une solution fiable

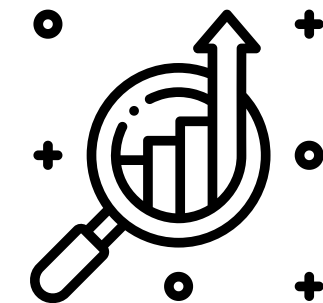
II-Processus de préparation des données



- DataSet importer de kaggle
- Contenne liée à des transactions bancaires avec des informations de base sur les clients



- Importation du DATA
- Nettoyage des données manquantes
- Nettoyage des données dupliquées
- Nettoyage des données aberrantes



- Explorer les données pour mieux les comprendre et identifier les tendances, les schémas et les anomalies.



	TransactionID	CustomerID	CustomerDOB	CustGender	CustLocation	CustAccountBalance	TransactionDate	TransactionTime	TransactionAmount (INR)
0	T1	C5841053	10/1/94	F	JAMSHEDPUR	17819.05	2/8/16	143207	25.00
1	T2	C2142763	4/4/57	M	JHAJJAR	2270.69	2/8/16	141858	27999.00
2	T3	C4417068	26/11/96	F	MUMBAI	17874.44	2/8/16	142712	459.00
3	T4	C5342380	14/9/73	F	MUMBAI	866503.21	2/8/16	142714	2060.00
4	T5	C9031234	24/3/88	F	NAVI MUMBAI	6714.43	2/8/16	181156	1762.50

- **TransactionID**: Identifiant unique pour chaque transaction effectuée par le client.
- **CustomerID**: Identifiant unique pour chaque client.
- **CustomerDOB**: Date de naissance du client.
- **CustGender**: Genre du client (masculin ou féminin).
- **CustLocation**: Emplacement géographique du client.
- **CustAccountBalance**: Solde du compte du client au moment de la transaction.
- **TransactionDate**: Date à laquelle la transaction a été effectuée.
- **TransactionTime**: Heure à laquelle la transaction a été effectuée.
- **TransactionAmount (INR)**: Montant de la transaction effectuée en roupies indiennes.

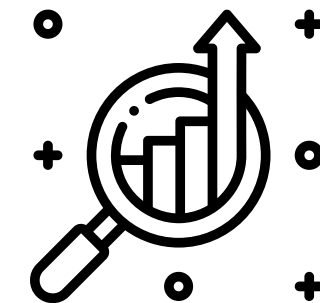
II-Processus de préparation des données



- DataSet importer de kaggle
- Contenne liée à des transactions bancaires avec des informations de base sur les clients



- Importation du DATA
- Nettoyage des données manquantes
- Nettoyage des données dupliquées
- Nettoyage des données aberrantes



- Explorer les données pour mieux les comprendre et identifier les tendances, les schémas et les anomalies.



1.Importation du DATA

```
# chargement et affichage des données
data = pd.read_csv('C:/Users/HP/Downloads/archive/bank_transactions.csv')
data.head(10)
```

	TransactionID	CustomerID	CustomerDOB	CustGender	CustLocation	CustAccountBalance	TransactionDate	TransactionTime	TransactionAmount (INR)
0	T1	C5841053	10/1/94	F	JAMSHEDPUR	17819.05	2/8/16	143207	25.00
1	T2	C2142763	4/4/57	M	JHAJJAR	2270.69	2/8/16	141858	27999.00
2	T3	C4417068	26/11/96	F	MUMBAI	17874.44	2/8/16	142712	459.00
3	T4	C5342380	14/9/73	F	MUMBAI	866503.21	2/8/16	142714	2060.00
4	T5	C9031234	24/3/88	F	NAVI MUMBAI	6714.43	2/8/16	181156	1762.50
5	T6	C1536588	8/10/72	F	ITANAGAR	53609.20	2/8/16	173940	676.00
6	T7	C7126560	26/1/92	F	MUMBAI	973.46	2/8/16	173806	566.00
7	T8	C1220223	27/1/82	M	MUMBAI	95075.54	2/8/16	170537	148.00
8	T9	C8536061	19/4/88	F	GURGAON	14906.96	2/8/16	192825	833.00
9	T10	C6628024	22/6/84	M	MUMBAI	4270.22	2/8/16	102446	280.44



2 .Nettoyage des données manquantes

Avant le nettoyage:

Les Nombres de valeurs
manquantes dans chaque
colonne

TransactionID	0
CustomerID	0
CustomerDOB	3397
CustGender	1100
CustLocation	151
CustAccountBalance	2369
TransactionDate	0
TransactionTime	0
TransactionAmount (INR)	0
dtype: int64	



Après le nettoyage:

Les Nombres de valeurs
manquantes dans chaque
colonne

TransactionID	0
CustomerID	0
CustomerDOB	0
CustGender	0
CustLocation	0
CustAccountBalance	0
TransactionDate	0
TransactionTime	0
TransactionAmount (INR)	0
dtype: int64	



3. Nettoyage des données dupliquées

Dans notre dataset on a pas trouver des valeurs dupliquées

```
0      False
1      False
2      False
3      False
4      False
...
1048562 False
1048563 False
1048564 False
1048565 False
1048566 False
Length: 1041614, dtype: bool
```



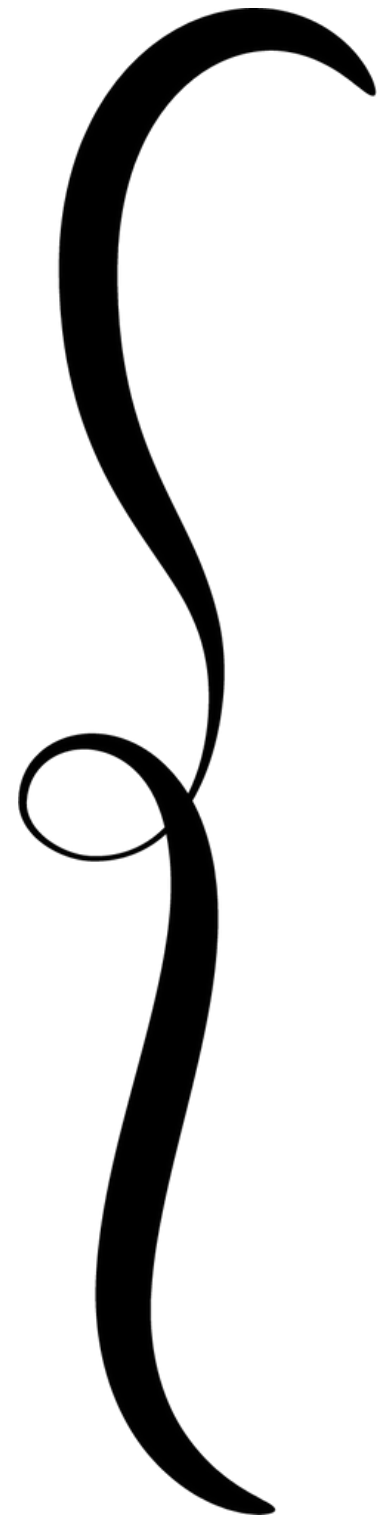
4. Nettoyage des données aberrantes

"Méthode IQR" pour détecter les valeurs aberrantes dans une distribution.

```
#made a copy of data  
df1 = data  
#calcule quartile and IQR  
Q1 = data["CustAccountBalance"].quantile(0.25)  
Q3 = data["CustAccountBalance"].quantile(0.75)  
IQR = Q3 - Q1  
#find upper and lower limits  
upper_limit = Q3 + 1.5 * IQR  
lower_limit = Q1 - 1.5 * IQR  
  
upper_limit , lower_limit
```



5. Les Bibliothèques utilisées

- 
- Pandas
 - Numpy
 - RE
 - Matplotlib
 - Seaborn

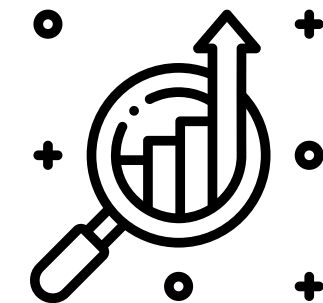
II-Processus de préparation des données



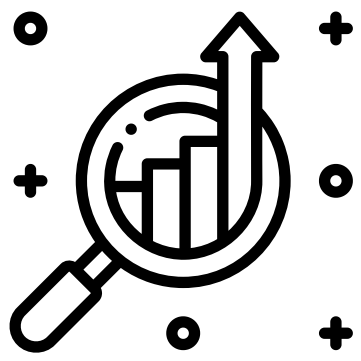
- DataSet importer de kaggle
- Contenne liée à des transactions bancaires avec des informations de base sur les clients



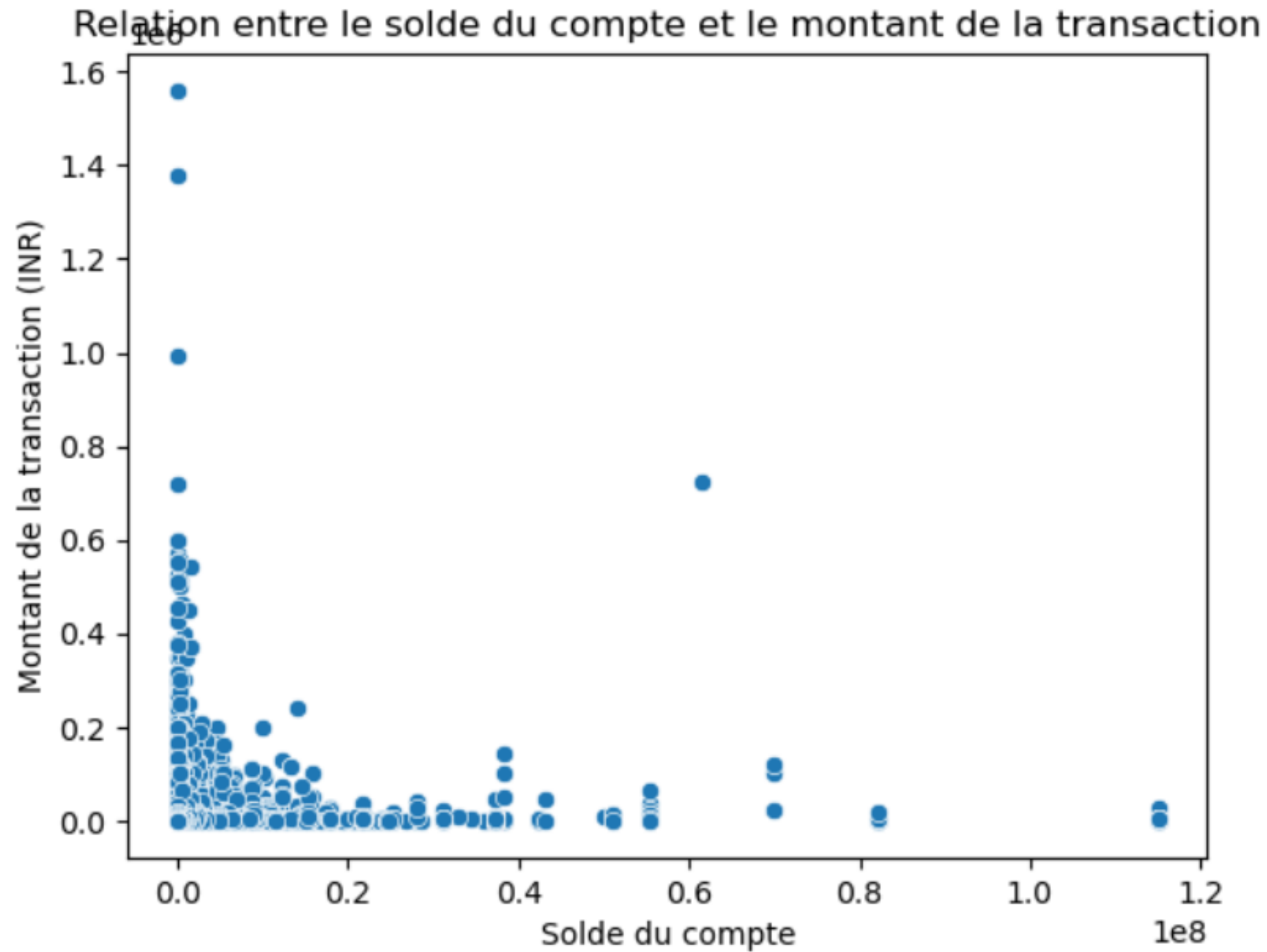
- Importation du DATA
- Nettoyage des données manquantes
- Nettoyage des données dupliquées
- Nettoyage des données aberrantes



- Explorer les données pour mieux les comprendre et identifier les tendances, les schémas et les anomalies.



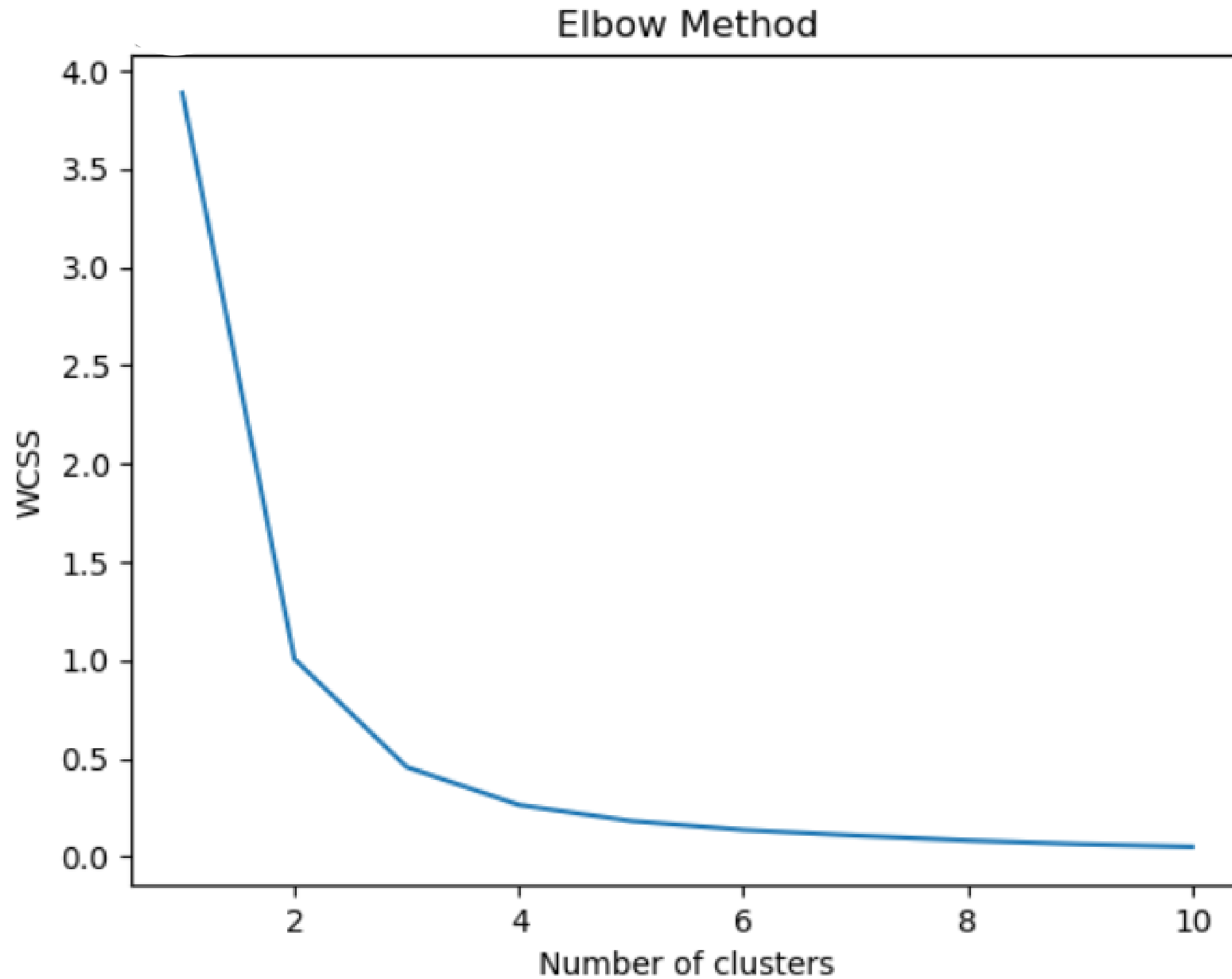
Visualiser notre dataset :



Traitement de la segmentation:

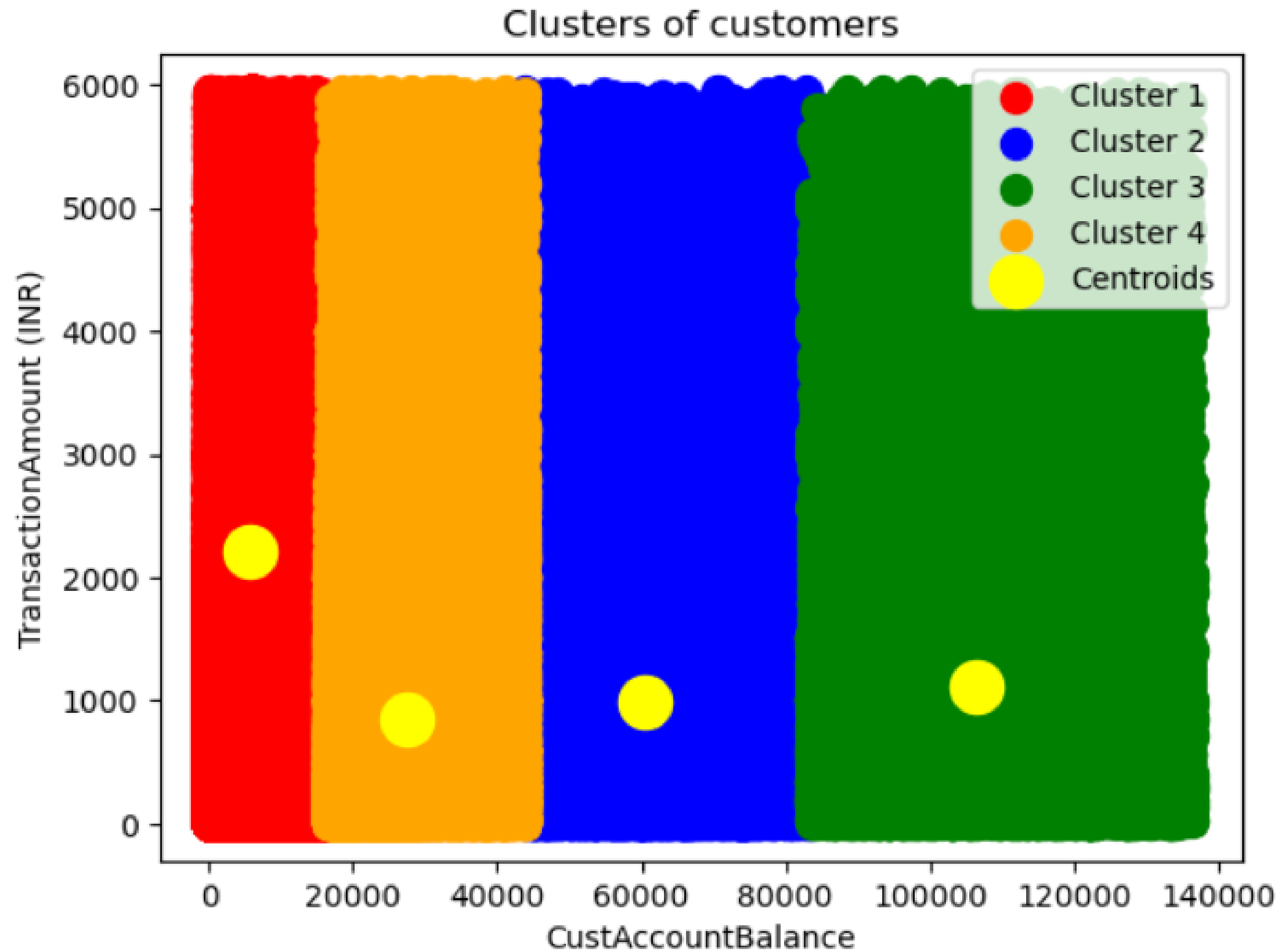
- ➡ 1/ **Séparer** nos données en deux parties en raison de leur volume important :
 - *Utilisé la bibliothèque scikit-learn, et la fonction 'train_test_split'*
- ➡ 2/ Choisir deux colonnes: 'CustAccountBalance' et 'TransactionAmount (INR)' pour la segmentation
 - *Afin d'identifier des groupes de clients similaires en termes de richesse et de dépenses*
- ➡ 3/ **Implémenter** la méthode de **Elbow** pour déterminer le nombre de cluster nécessaire
 - *Le choix d'un nombre prédéfini de clusters permet de simplifier l'analyse de segmentation et de faciliter l'interprétation des résultats obtenus*
 - *Cette méthode consiste à tracer un graphe qui se base sur un point d'inflexion où la courbe commence à s'aplatir*

-le graphe de la méthode d'ELBOW :



➔ Utilisé la méthode de KMEANS

- **Application** de K-means pour segmenter les clients en groupes homogènes en fonction de leur solde de compte et de leur montant de transaction



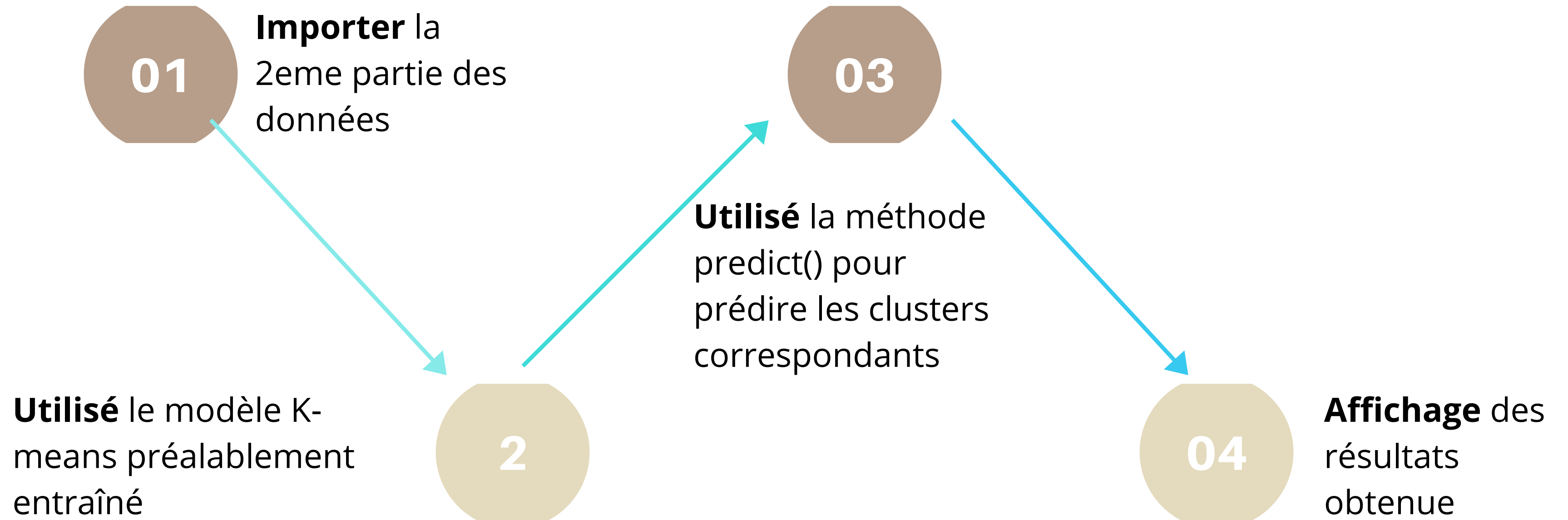
Evaluation de la segmentation :

- le **Davies-Bouldin Index** (DBI) mesure à quel point les clusters sont séparés les uns des autres tout en maintenant une cohésion élevée à l'intérieur des clusters. Le score est calculé en utilisant uniquement les distances euclidiennes entre les points de données.

```
# Évaluation de la qualité de la segmentation  
dbi = davies_bouldin_score(X, y_kmeans)  
print('Davies-Bouldin index:', dbi)
```

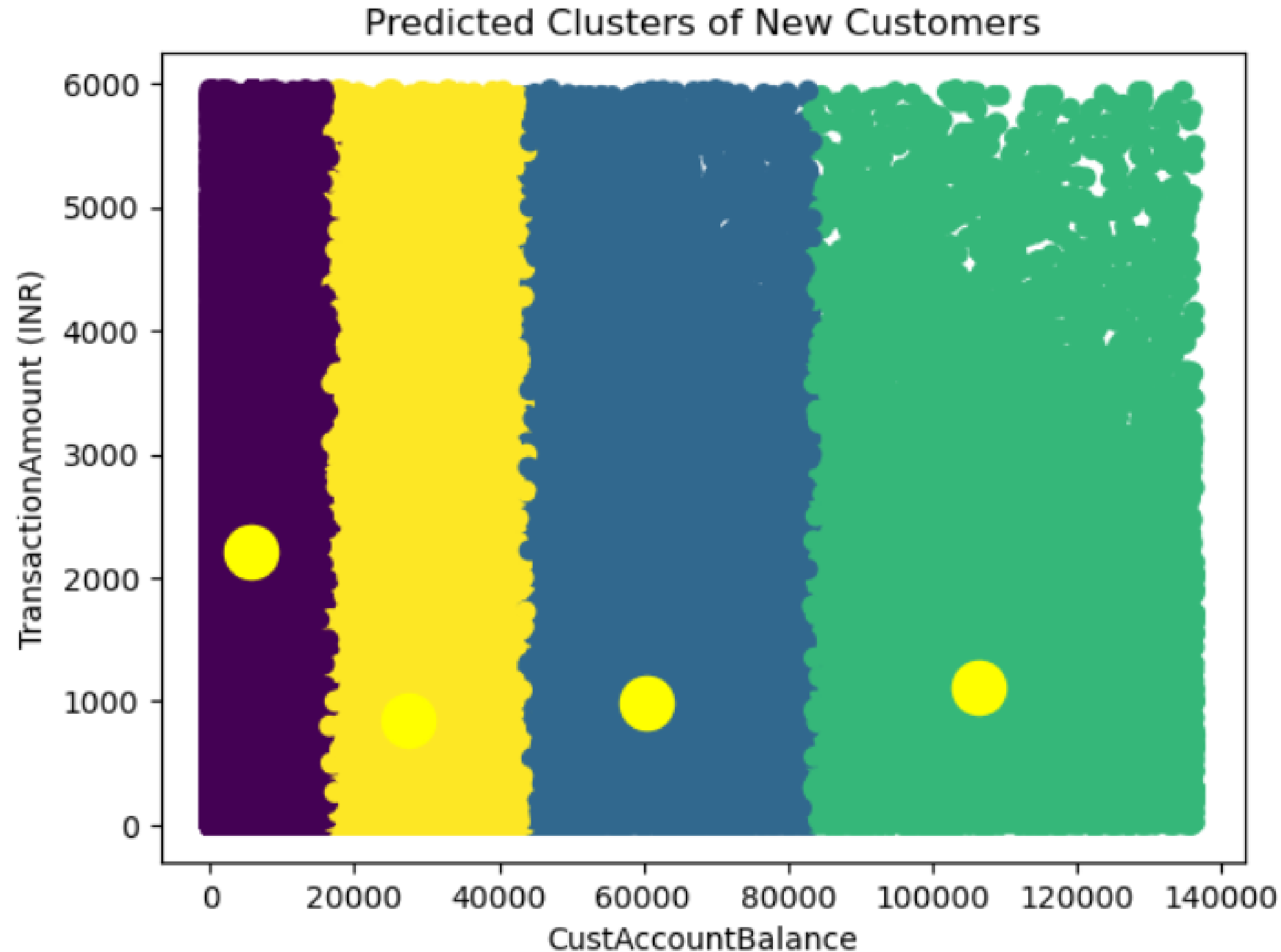
```
Davies-Bouldin index: 0.5018397682923231
```

2ème traitement : I- prédiction



Visualisation

Ce diagramme illustre la prédiction des nouveaux clients présents dans la 2eme partie des données et affecte a chaque client le cluster approprié



2ème traitement :

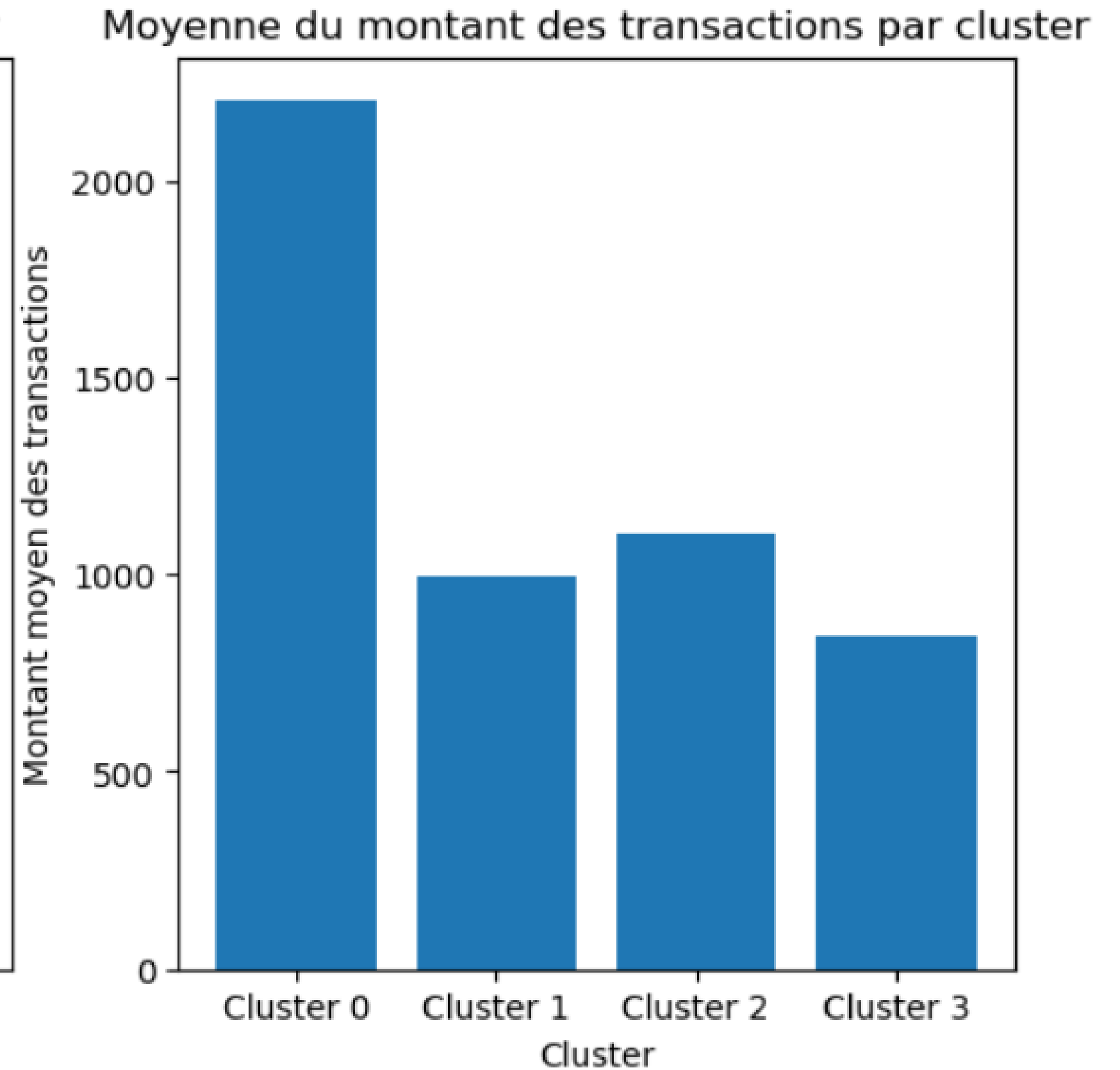
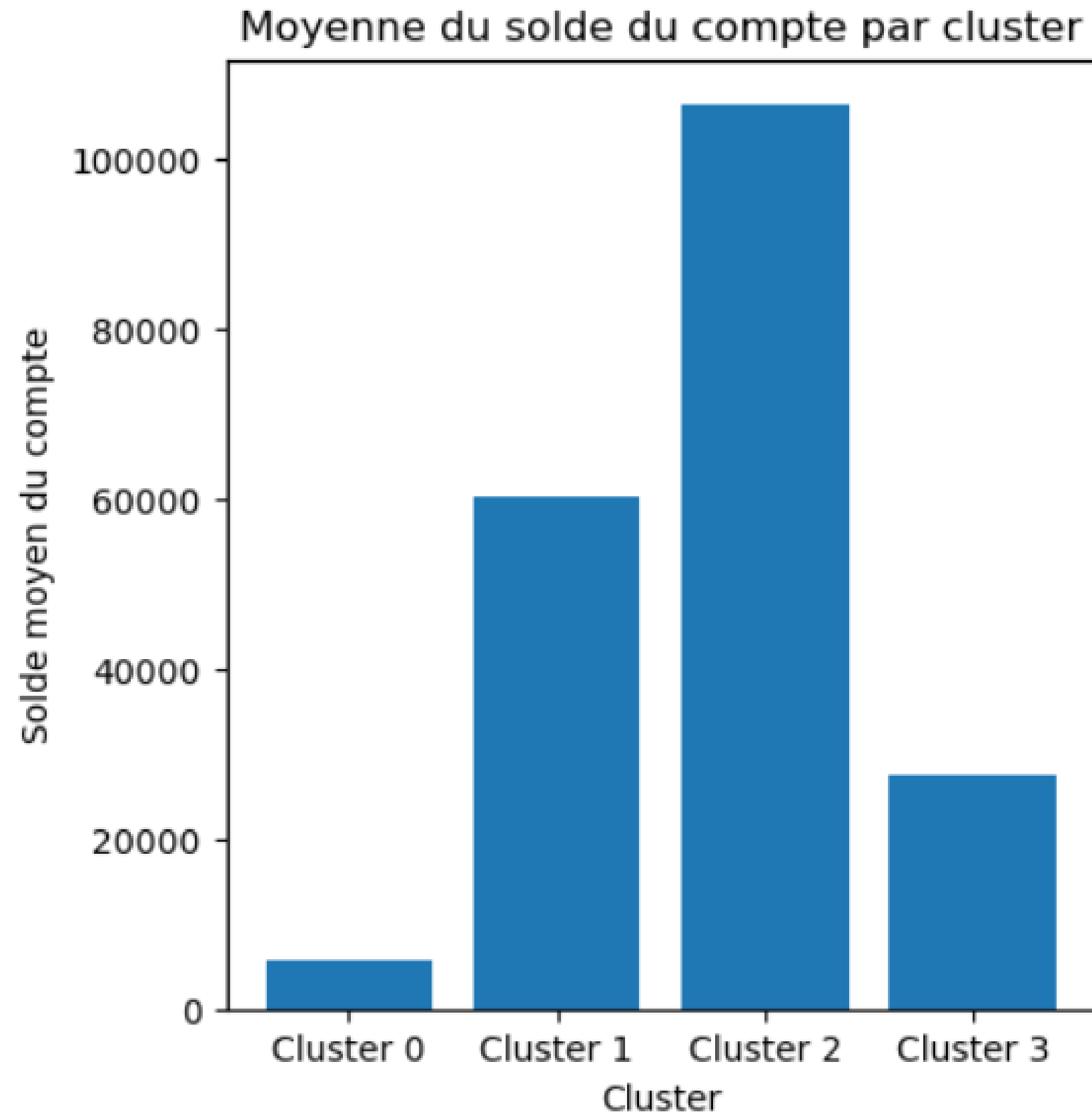
II- Processus de recommandation

A-Etude statistique :

cluster	CustAccountBalance		TransactionAmount (INR)	
	mean	std	mean	std
0	5772.295981	4140.820187	2202.715447	2535.518232
1	60359.003468	11186.478970	994.503707	1119.311375
2	106474.739111	14759.715228	1106.577922	1205.862502
3	27469.574859	7639.515711	847.853940	1032.556958

-
- Représentent les moyennes et écarts-types des caractéristiques de chaque cluster, en fonction de (CustAccountBalance et TransactionAmount).

visualisation des statistiques



La recommandation collaborative :

- Implémentation d'une fonction qui se base sur l'étude statistique déjà fait



Cluster 0 : reduction des frais des transaction

proposer des credit à court terme avec des montants faibles

Cluster 1 : Des services de conseil financier personnalisés pour chaque client en fonction de son objectifs

Cluster 2 : proposer un compte d'épargne rémunéré

investissements à long terme

Cluster 3 : investir dans des produits à faible risque tels que des obligations d'État ou des fonds de trésor

cartes de crédit à faible taux d'intérêt

le but général :

- Fédiliser les clients
- Assurer la liquidité financière

Conclusion

- la segmentation des clients est une tâche importante pour toute entreprise souhaitant comprendre les caractéristiques de ses clients et concevoir des recommandations pointu
- L'utilisation de Python a permis de gagner du temps et d'automatiser certaines parties de l'analyse en utilisant les multiples bibliothèques et technique optimisé



**MERCI POUR
VOTRE
ATTENTION**

**ENCADRER PAR :
OMAR SEFRIOUI**

