



# K-MEANS VS DBSCAN

Comparaison des Algorithmes K-Means et DBSCAN  
pour la Segmentation Client : Analyse et  
Implémentation

Réalisé par :

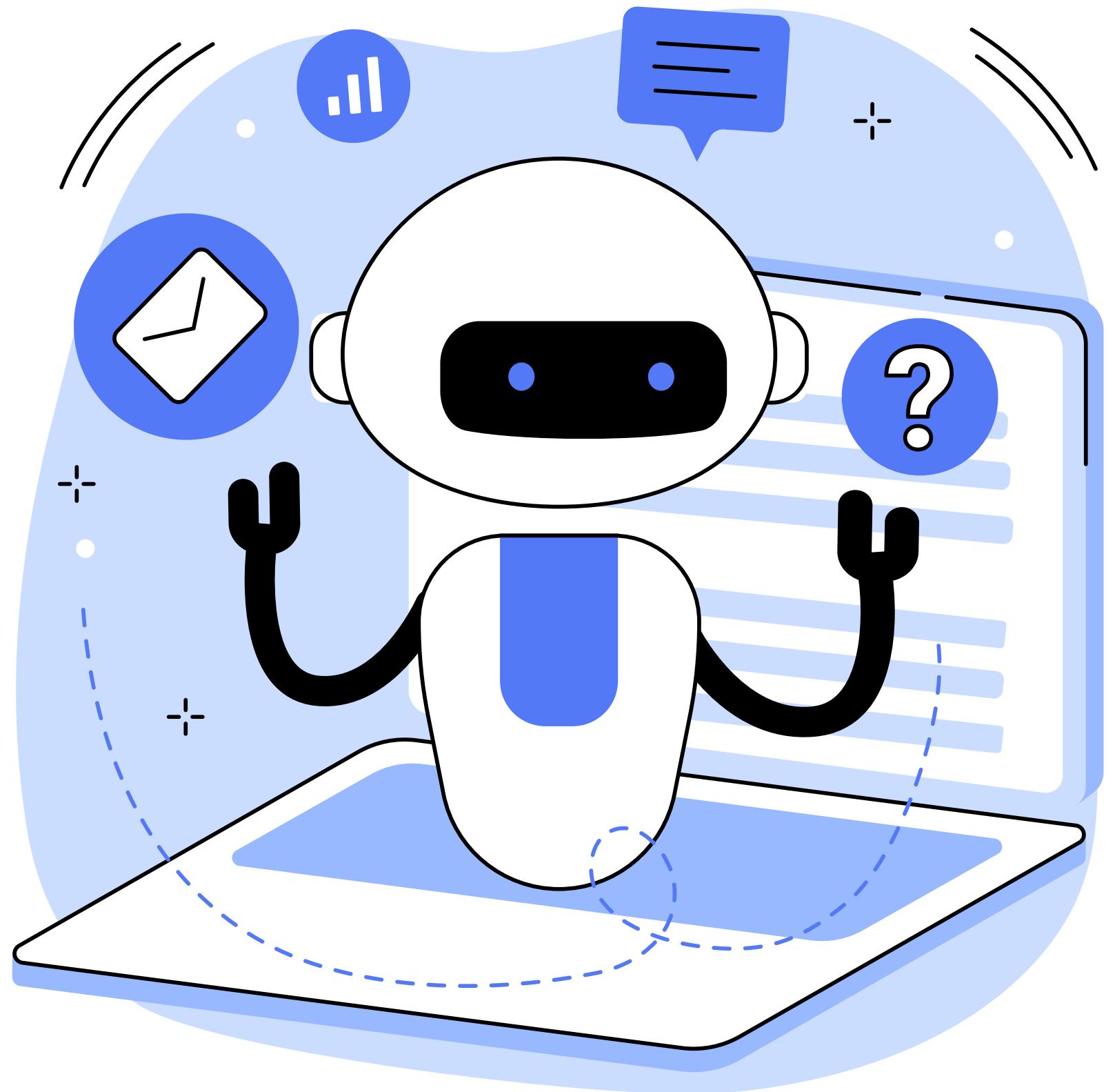
M. DOUKKANI Mohamed  
M. ZAGHOU Mohamed

Encadré par :

Mme. MOUNIR Ilham

2024/2025

# INTRODUCTION



# PLAN

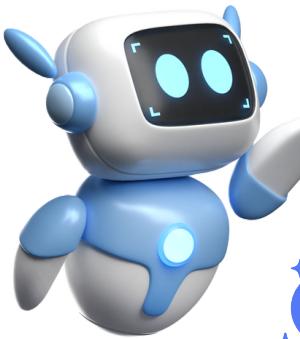


- 1 Présentation du sujet et du dataset**
- 2 Présentation des algorithmes de clustering**
- 3 Outils et bibliothèques utilisés**
- 4 Processus de Réalisation de la segmentation client**
- 5 Démonstration dans Jupyter Notebook**
- 6 Conclusion**

# PRÉSENTATION DU SUJET ET DU DATASET



# PRÉSENTATION DU SUJET ET DU DATASET



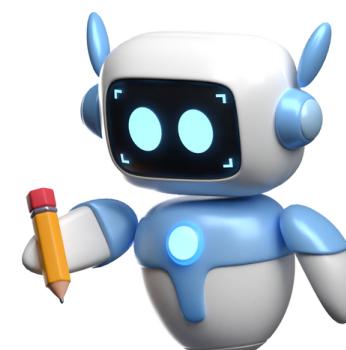
## 💡 Introduction.

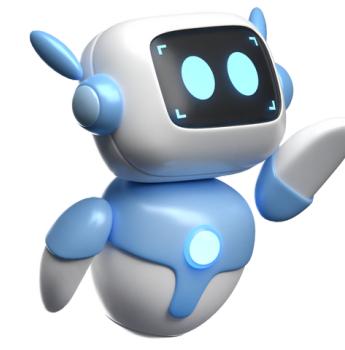
- La **segmentation client** est une démarche essentielle pour comprendre les différents comportements des clients. Elle permet de regrouper ces derniers en fonction de leurs caractéristiques communes, offrant ainsi une meilleure compréhension de leurs besoins. Ces informations sont cruciales pour adapter les stratégies marketing, personnaliser les offres et améliorer la satisfaction client.



## 💡 Objectif.

- Dans ce projet, nous analyserons un **dataset** contenant des informations sur **des clients de centres commerciaux**. L'objectif est de les **regrouper en différents clusters** pour mieux comprendre leurs habitudes et comportements.





# DESCRIPTION DU DATASET



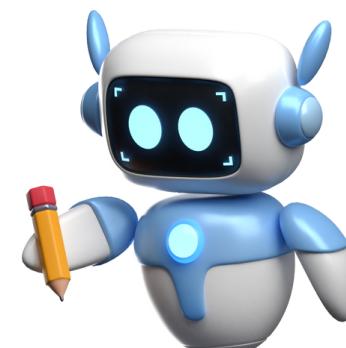
## Structure.

- Nombre total de lignes : **200**
- Nombre total de colonnes : **5**

## Colonnes principales.

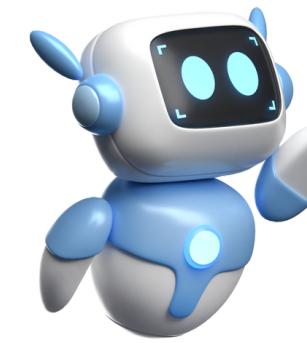
- CustomerID** : Identifiant unique de chaque client.
- Genre** : Sexe du client (Masculin/Féminin).
- Age** : Âge du client.
- Annual Income (k\$)** : Revenu annuel en milliers de dollars.
- Spending Score (1-100)** : Score indiquant les habitudes de dépenses du client (1 = Faible, 100 = Élevé).

CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
1	Male	19	15	39
2	Female	21	15	81
3	Female	20	16	6
4	Male	23	17	77



# PRÉSENTATION DES ALGORITHMES DE CLUSTERING



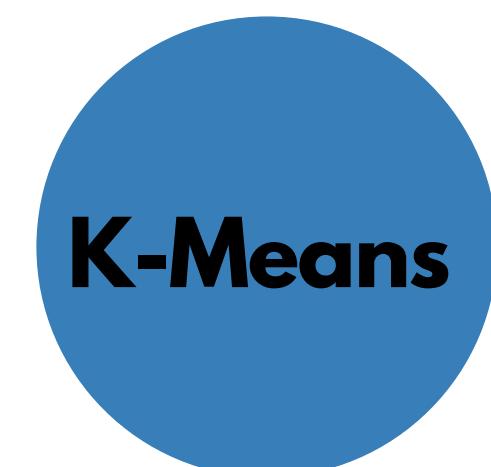
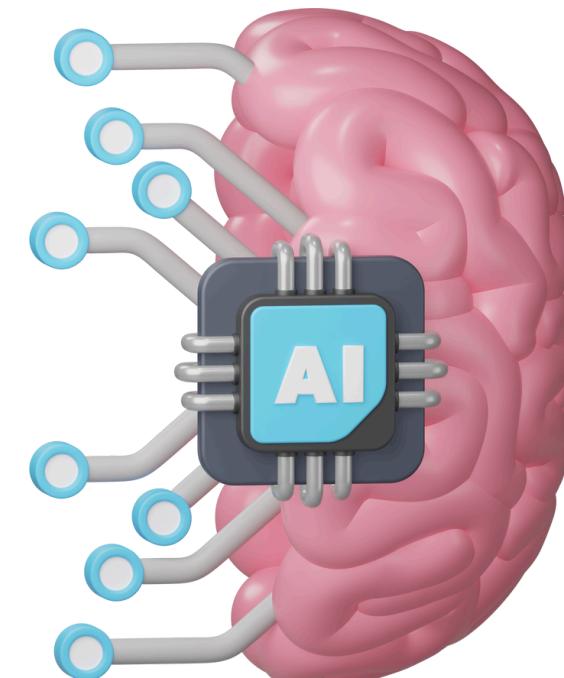


# INTRODUCTION AU CLUSTERING ET SES ALGORITHMES

## 💡 Clustering.

- Le clustering, ou regroupement, est le processus de division d'un ensemble de données en groupes homogènes où les éléments au sein d'un groupe sont similaires entre eux. Il repose sur plusieurs algorithmes.

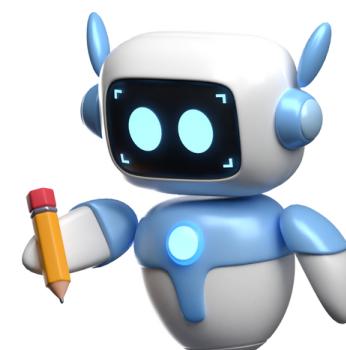
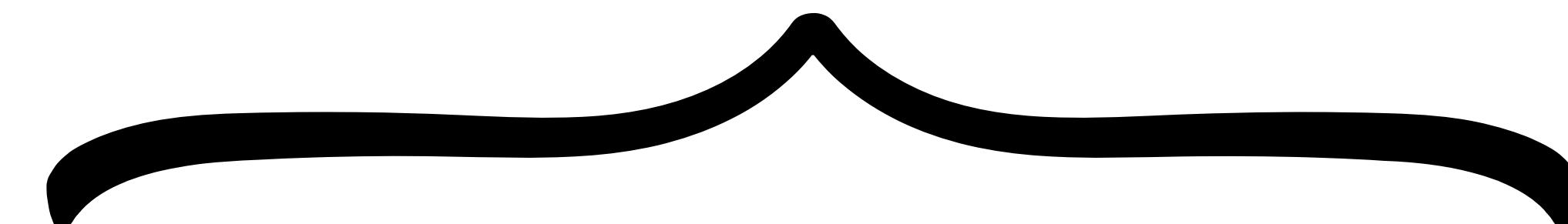
## 💡 Principaux Algorithmes de Clustering.



**K-Means**



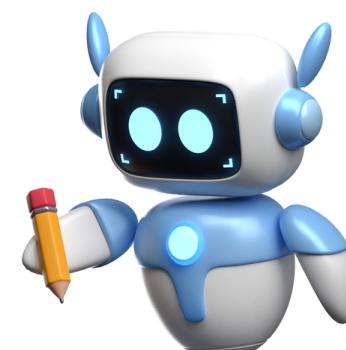
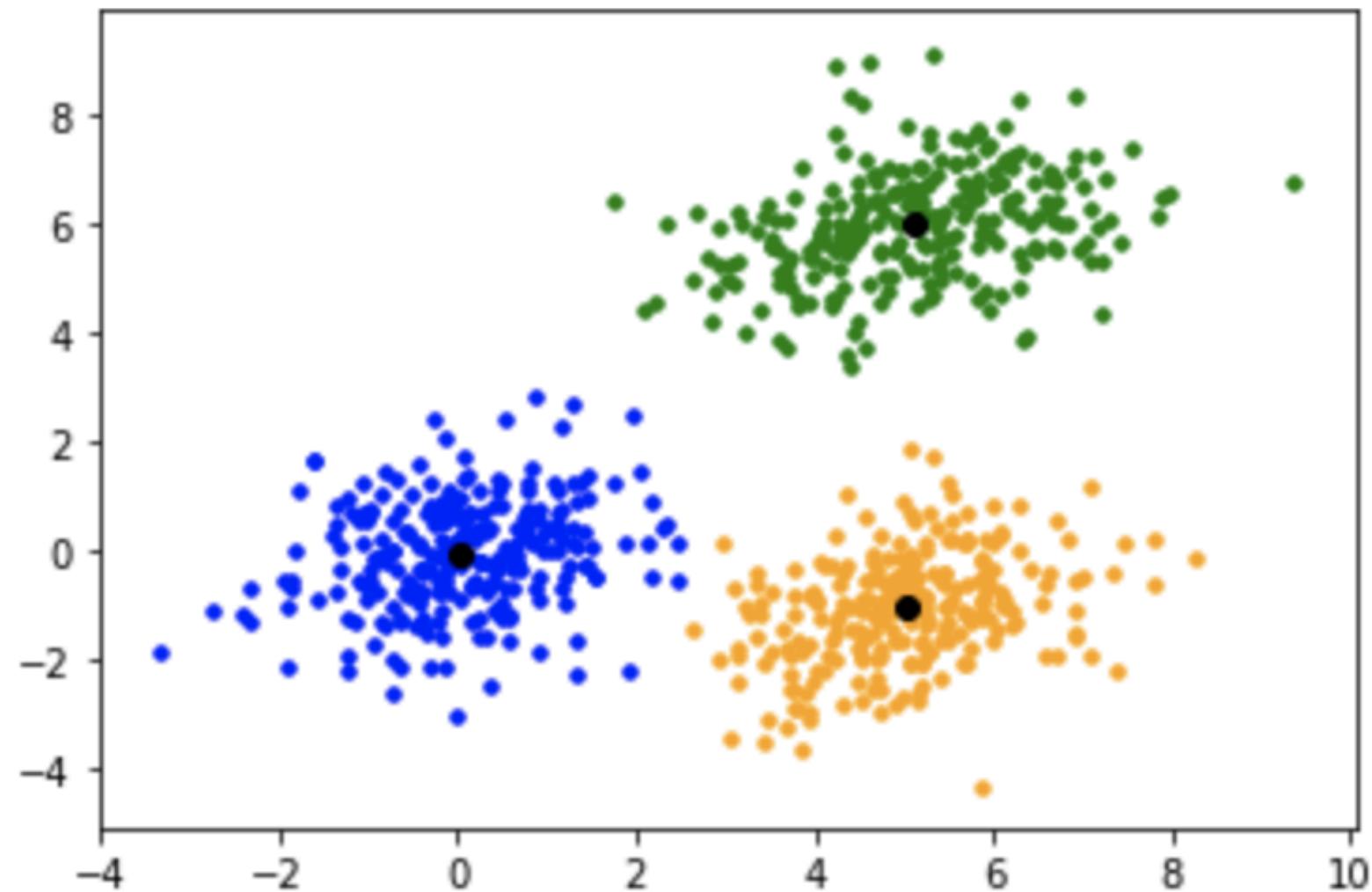
**DBSCAN**

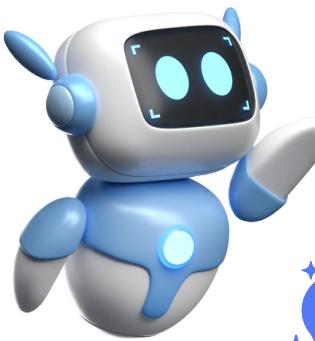


# L'ALGORITHME K-MEANS



- L'algorithme K-Means est déjà bien connu et a été étudié en cours. Dans ce projet, il sert de point de comparaison pour évaluer les résultats obtenus avec l'algorithme DBSCAN. K-Means regroupe les données en fonction de leur proximité par rapport à des centroides et est particulièrement efficace pour identifier des clusters bien séparés et de forme sphérique.

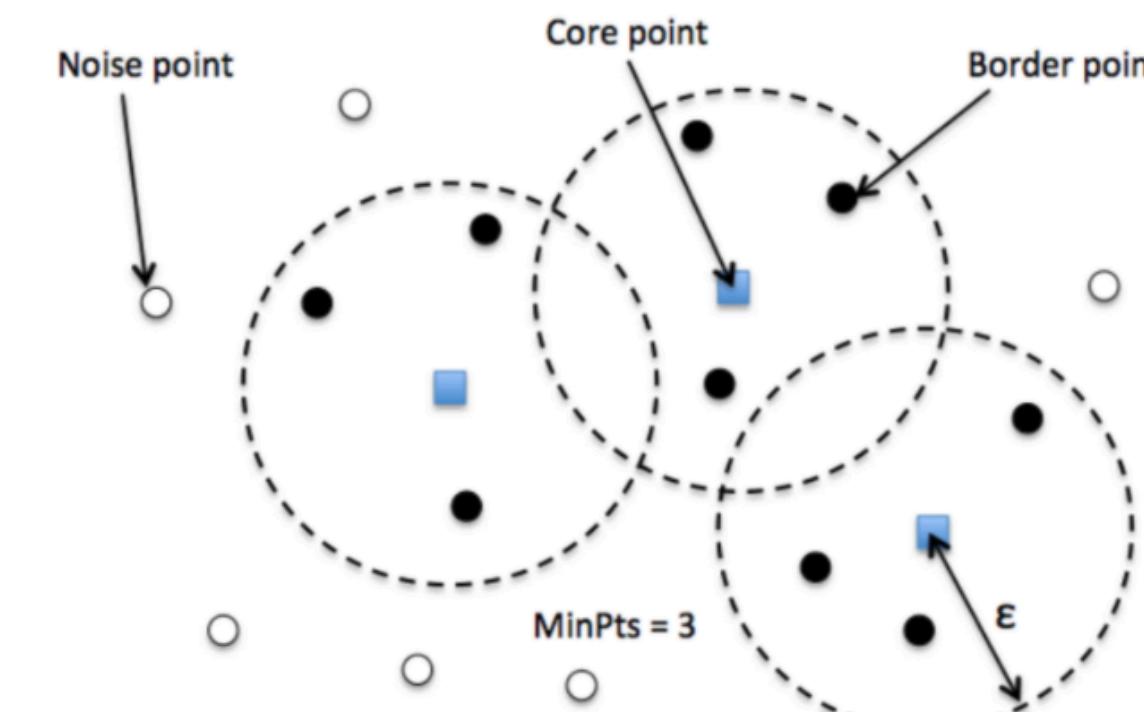
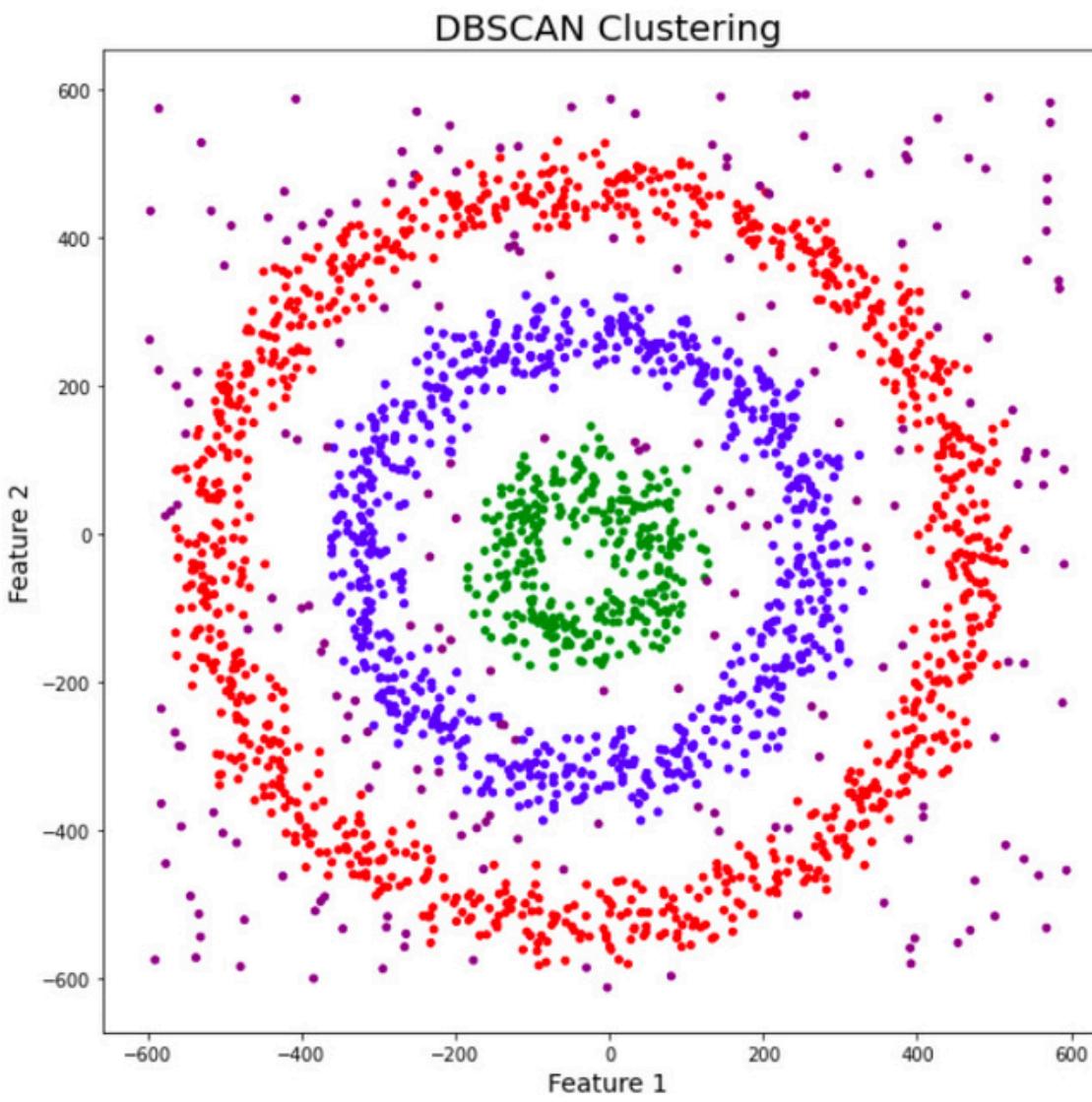




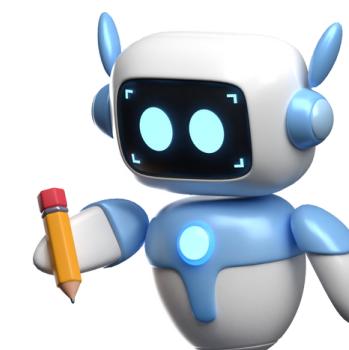
# DÉFINITION DE L'ALGORITHME DBSCAN

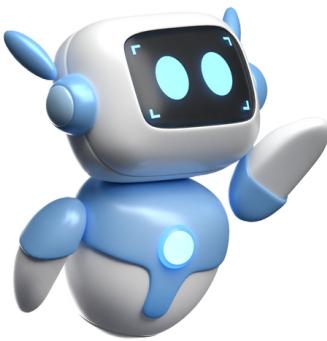
## DBSCAN.

- Le **DBSCAN** est un algorithme de regroupement basé sur la densité. Contrairement à K-Means, il identifie les clusters en fonction des zones de haute densité et est capable de détecter des formes complexes tout en isolant les points aberrants (bruit).

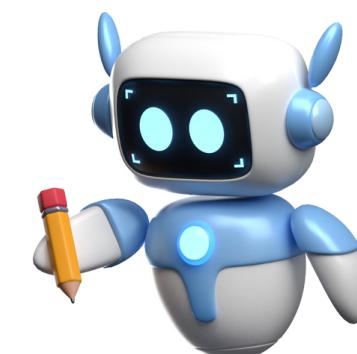
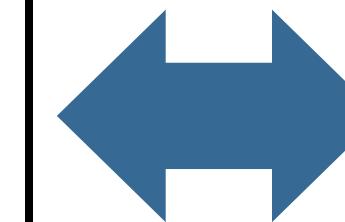
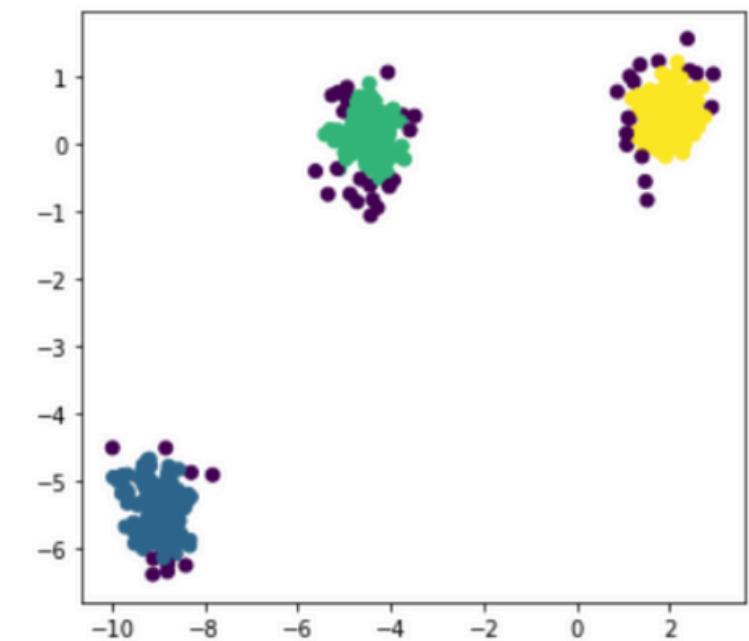
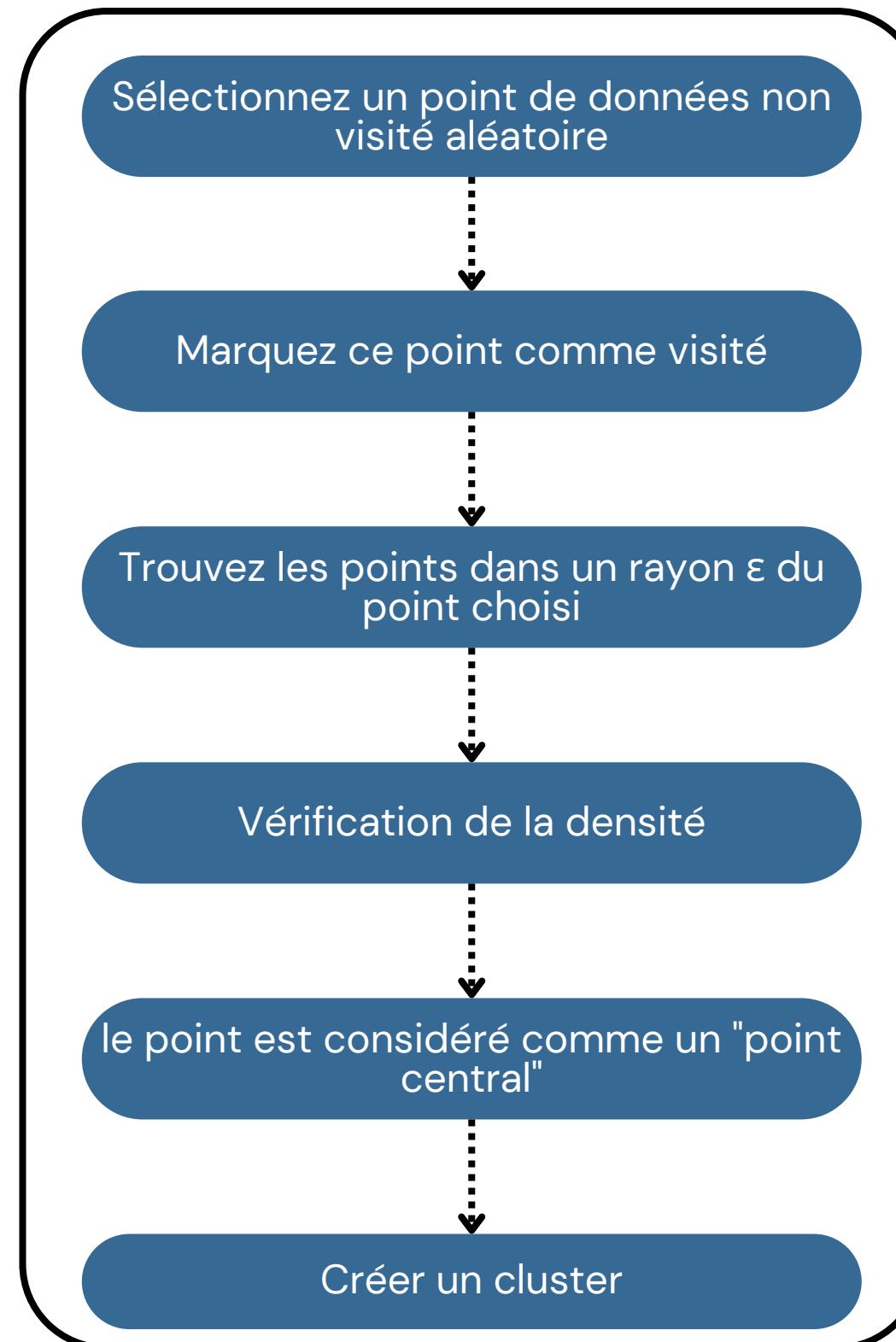
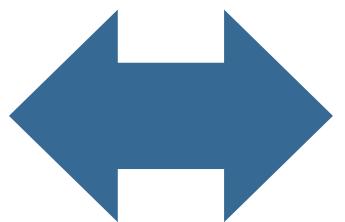
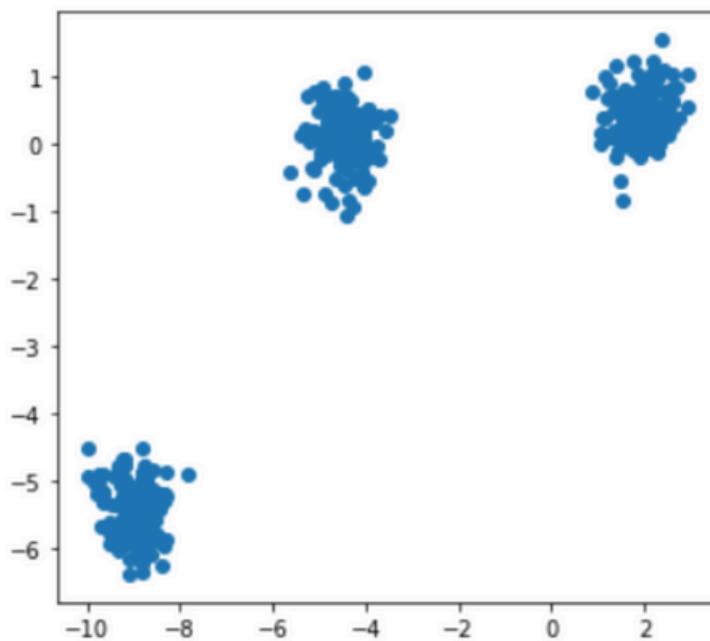


- Points centraux** : points ayant au moins  $\text{MinPts}$  voisins dans un rayon  $\epsilon$ .
- Points de bordure** : voisins d'un point central, mais ayant moins de  $\text{MinPts}$ .
- Points bruit** : points isolés, ne faisant partie d'aucun cluster.

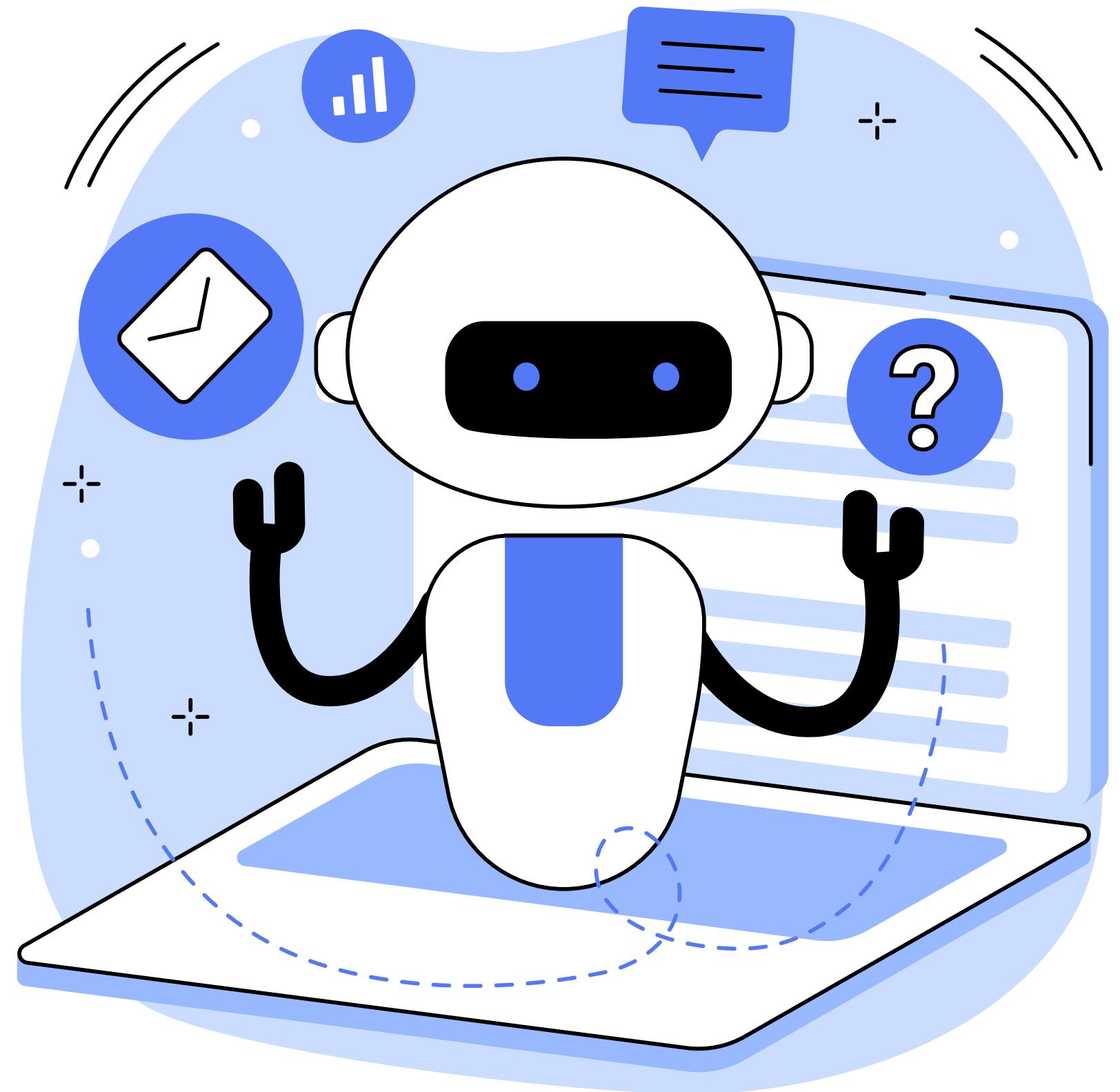


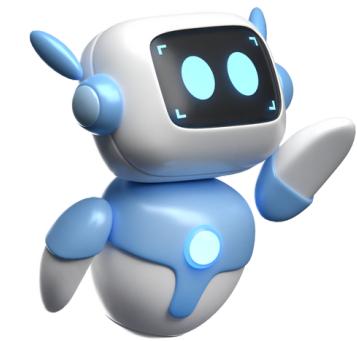


# FONCTIONNEMENT DE L'ALGORITHME DBSCAN



# OUTILS ET BIBLIOTHÈQUES UTILISÉS





# OUTILS ET LANGAGES UTILISÉS



ANACONDA®

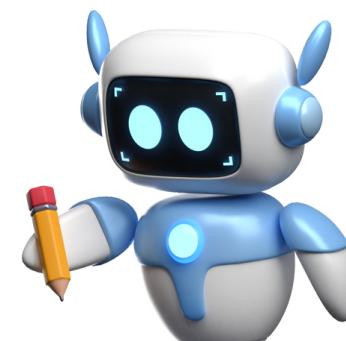


Anaconda

Jupyter

Github

Python

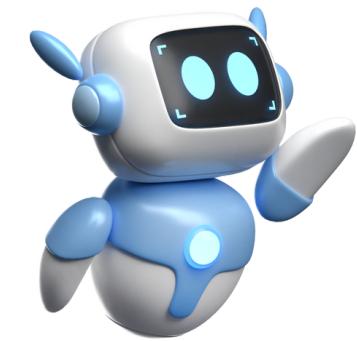


# BIBLIOTHÈQUES UTILISÉS

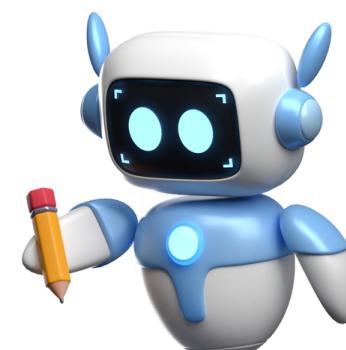


# PROCESSUS DE RÉALISATION DE LA SEGMENTATION CLIENT





# PROCESSUS DE RÉALISATION DE LA SEGMENTATION CLIENT



Importation des  
Bibliothèques  
Nécessaires

Chargement et  
Prétraitement du  
Dataset

Analyse  
Exploratoire des  
Données (EDA)

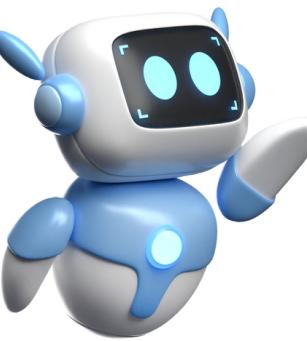
Création des  
Clusters avec K-  
Means

Application du  
Clustering avec  
DBSCAN

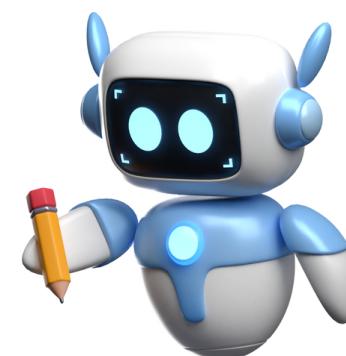
Comparaison  
entre K-Means et  
DBSCAN

# DÉMONSTRATION DANS JUPYTER NOTEBOOK





# DÉMONSTRATION EN DIRECT



# CONCLUSION





**MERCI POUR  
VOTRE ATTENTION!**