



K-MEANS VS DBSCAN

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

Réalisé par :

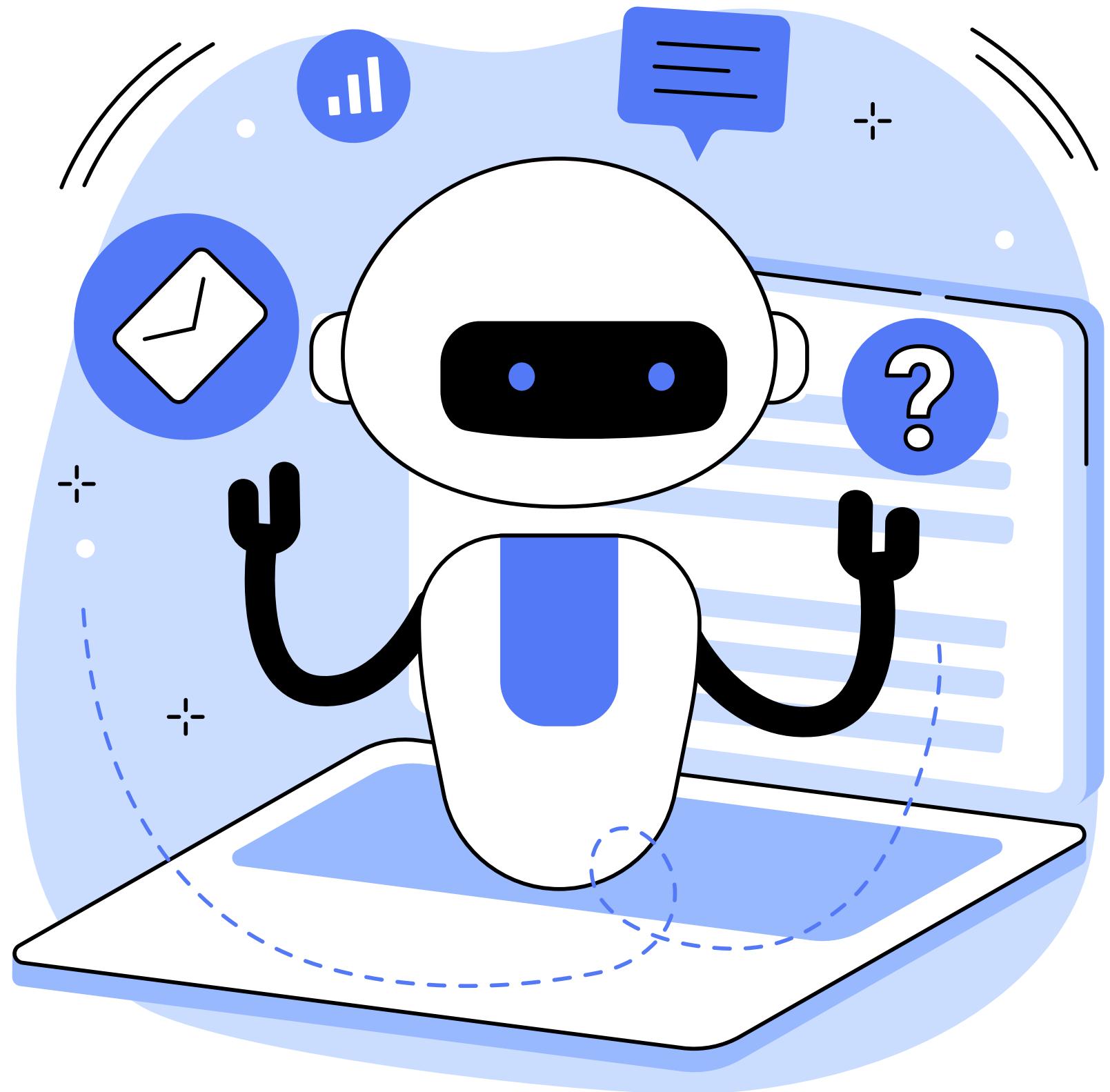
M. DOUKKANI Mohamed
M. ZAGHOU Mohamed

Encadré par :

Mme. MOUNIR Ilham

2024/2025

INTRODUCTION



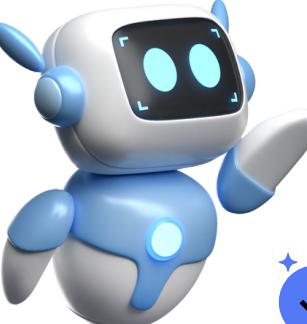
PLAN



- 1 Concepts Clés en Data Mining et Clustering**
- 2 K-MEANS Clustering**
- 3 DBSCAN Clustering**
- 4 Implémentation Pratique du Projet**
- 5 Conclusion**

CONCEPTS CLÉS EN DATA MINING ET CLUSTERING





INTRODUCTION AU DATA MINING



Data Mining.

- Le Data Mining, ou fouille de données, consiste à analyser de grandes quantités de données pour en extraire des informations utiles et des modèles cachés. Cela permet de mieux comprendre les données et de prendre des décisions éclairées. Il se divise en deux types de méthodes principales :



Méthodes Prédictives et Descriptives.

Méthodes Prédictives

“

Elles servent à prédire une valeur ou un résultat futur à partir de données existantes (exemple : prévision des ventes).

”

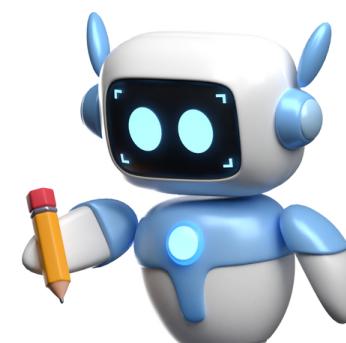


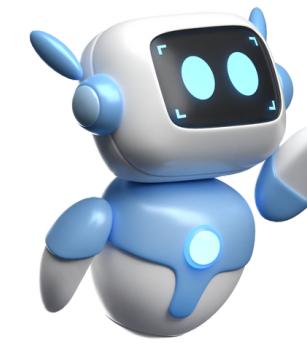
Méthodes Descriptives

“

Elles explorent les données pour identifier des relations ou des modèles sans chercher à prédire (exemple : regroupement de clients en fonction de leur comportement).

”





INTRODUCTION AU CLUSTERING ET SES ALGORITHMES

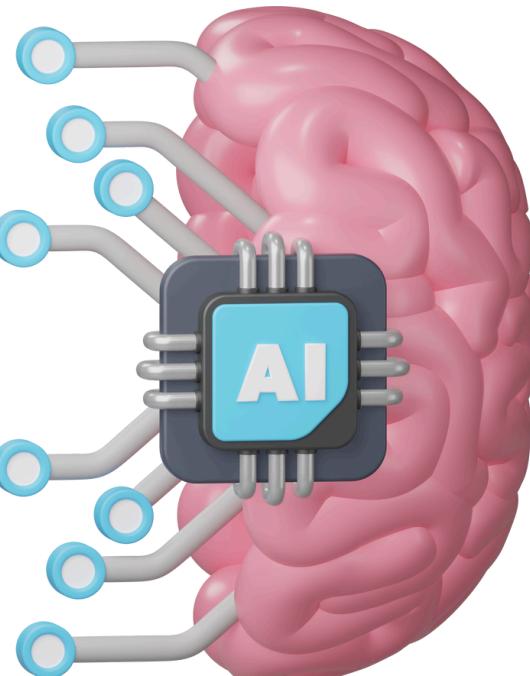
💡 Clustering.

- 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.

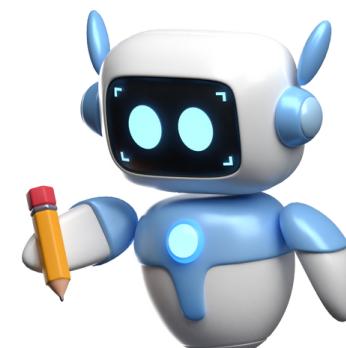
K-Means

DBSCAN



“ Regroupe les données en un nombre fixé de clusters. Simple et rapide, mais nécessite de définir le nombre de groupes à l'avance. ”

“ Crée des clusters basés sur la densité des points et gère bien les valeurs aberrantes ou bruits. ”



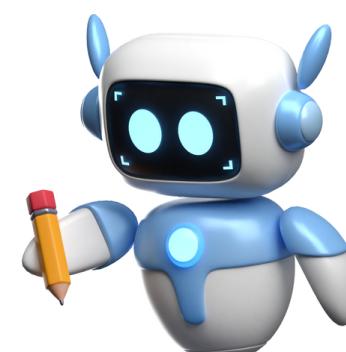
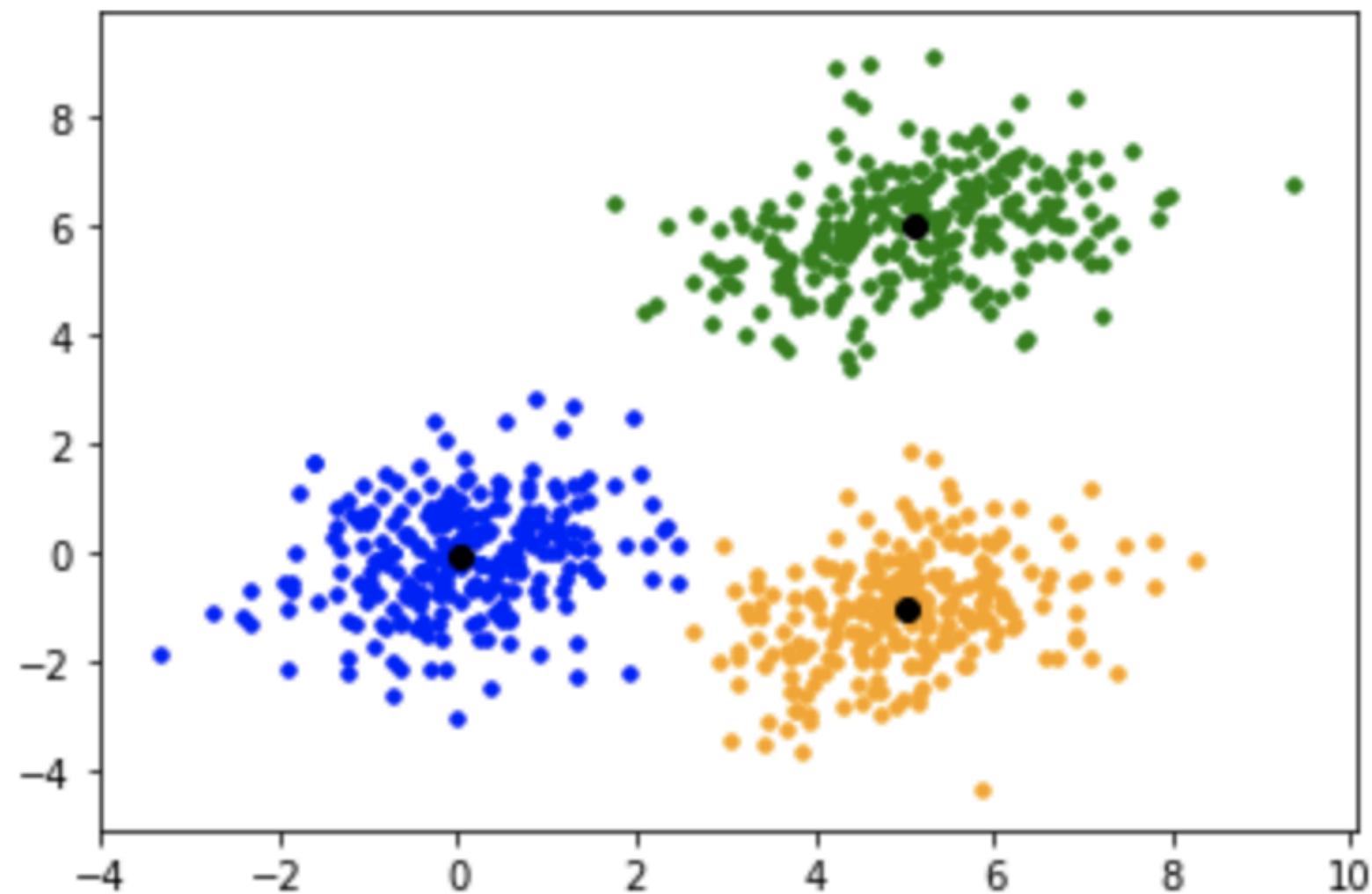
K-MEANS

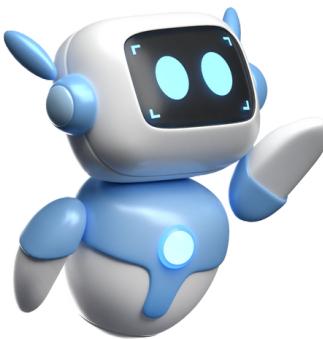


DÉFINITION DE L'ALGORITHME K-MEANS

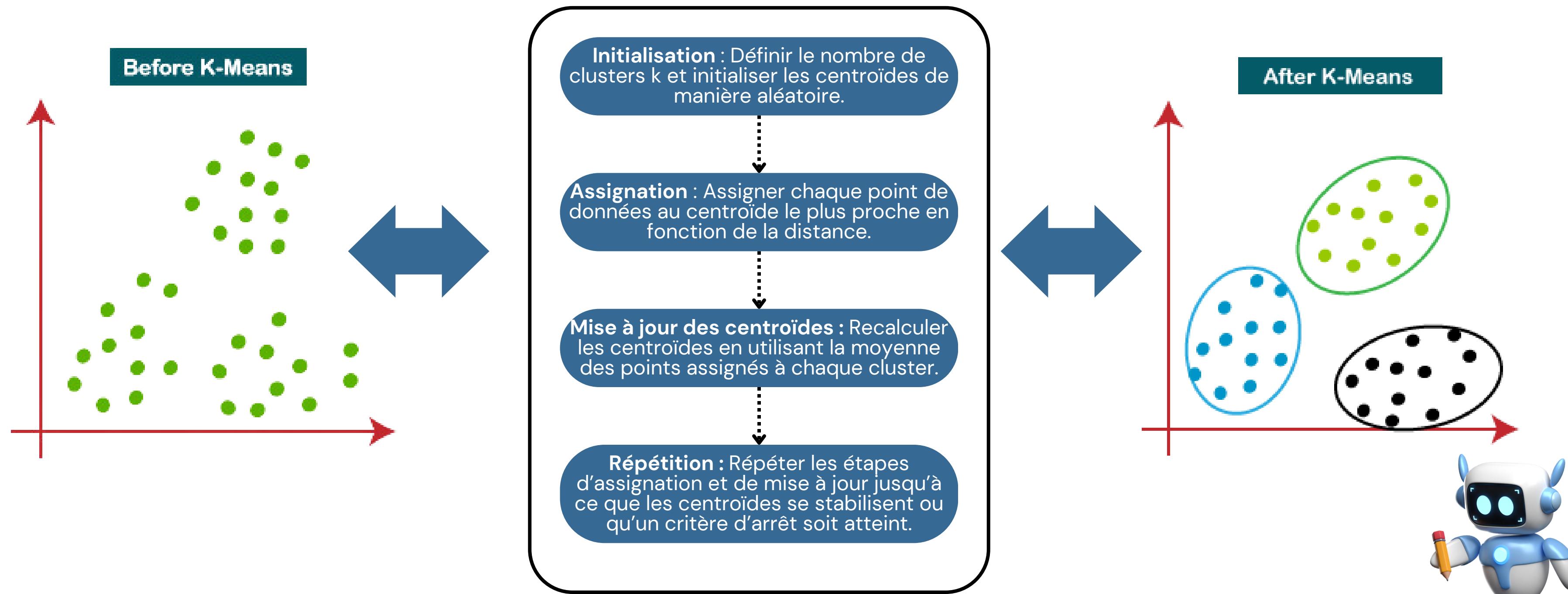


- Le K-Means Clustering est un algorithme d'apprentissage non supervisé utilisé pour regrouper des données en plusieurs clusters basés sur leurs similarités. Il vise à minimiser la distance entre chaque point de données et le centre de son cluster, appelé centroïde.

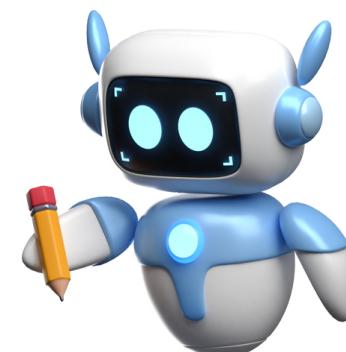
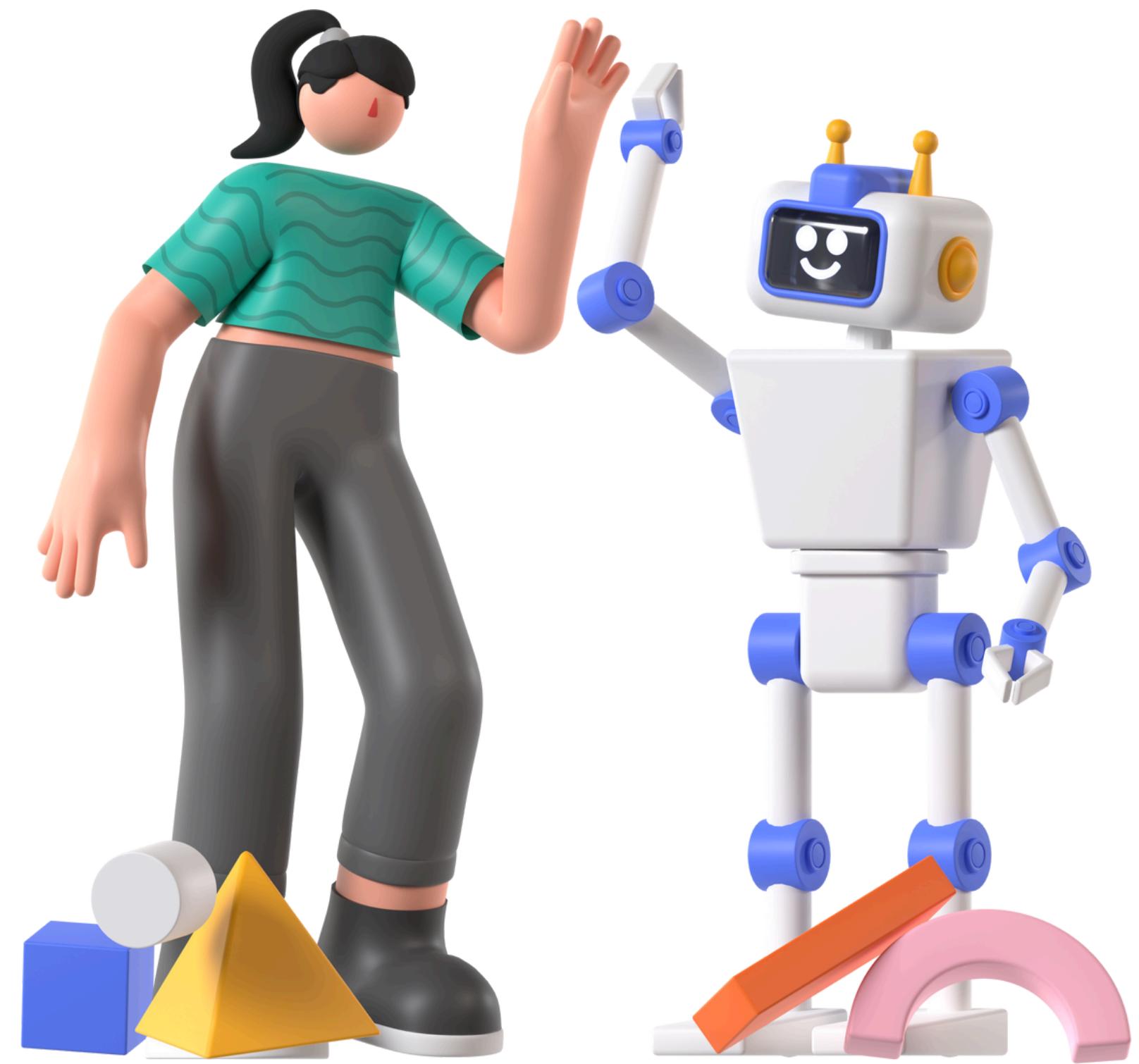




FONCTIONNEMENT DE L'ALGORITHME K-MEANS

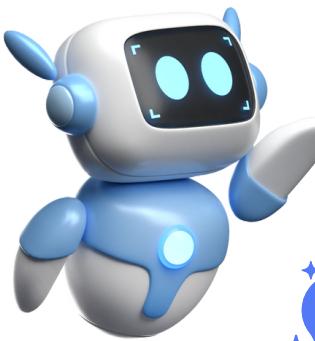


EXEMPLE D'APPLICATION



DBSCAN

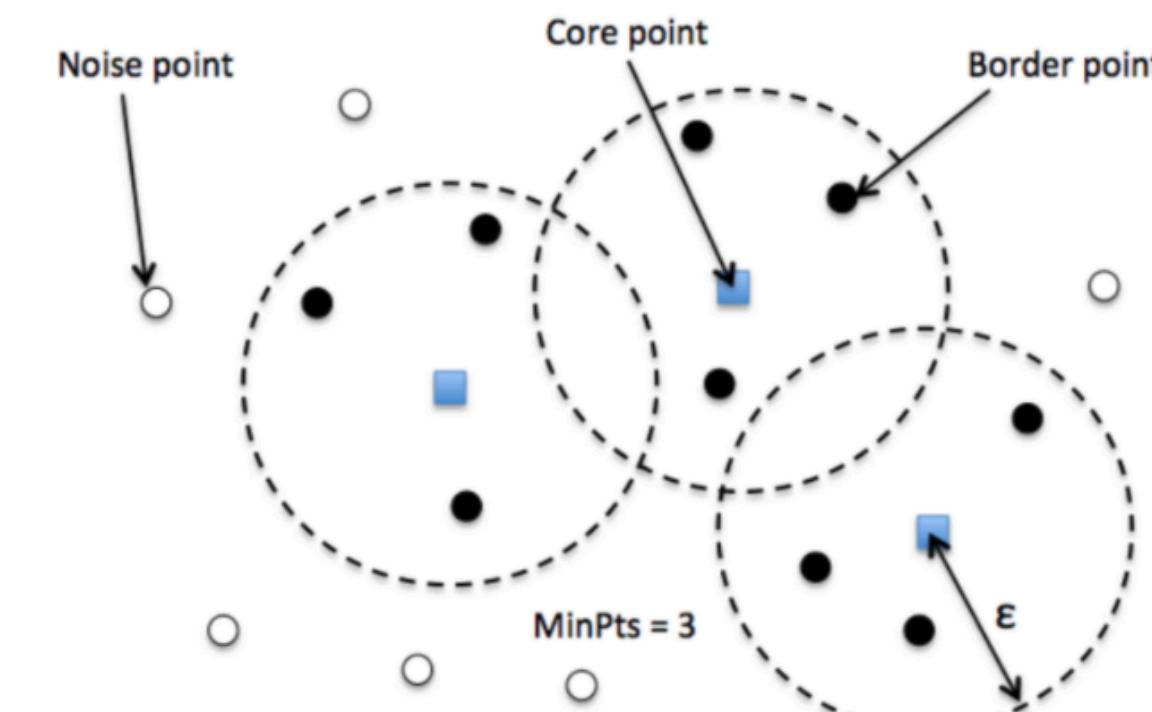
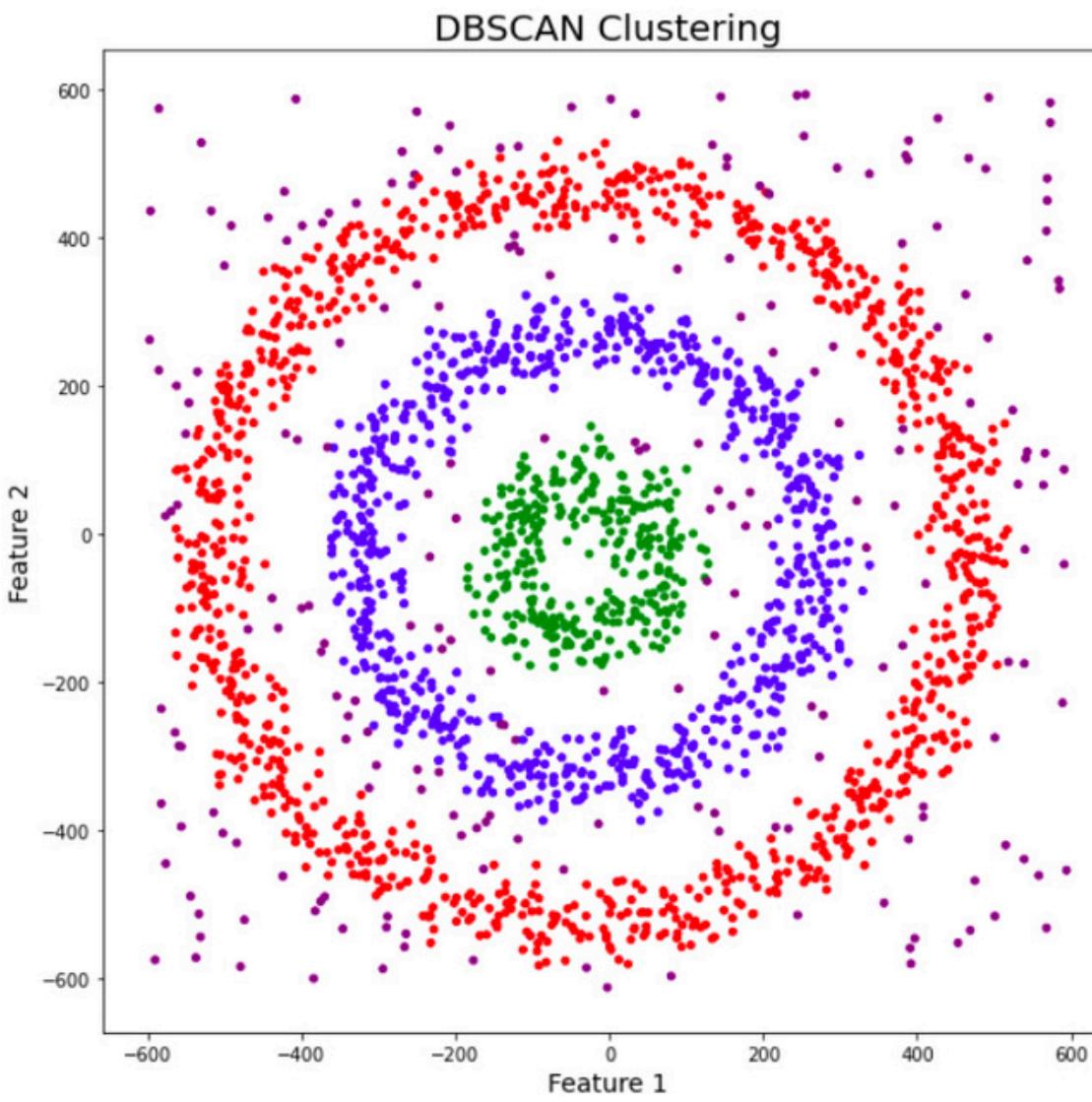




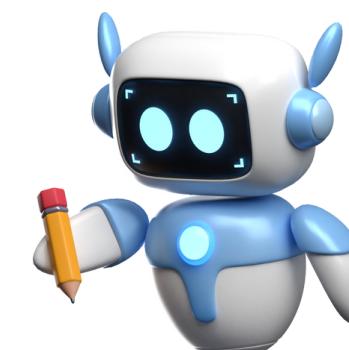
DÉFINITION DE L'ALGORITHME DBSCAN

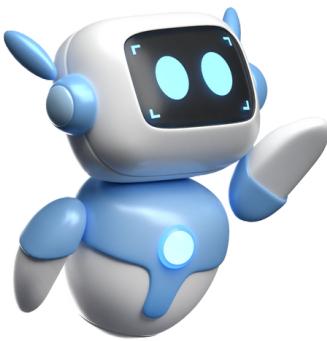
DBSCAN.

- 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).

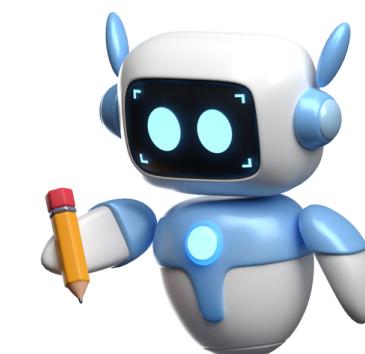
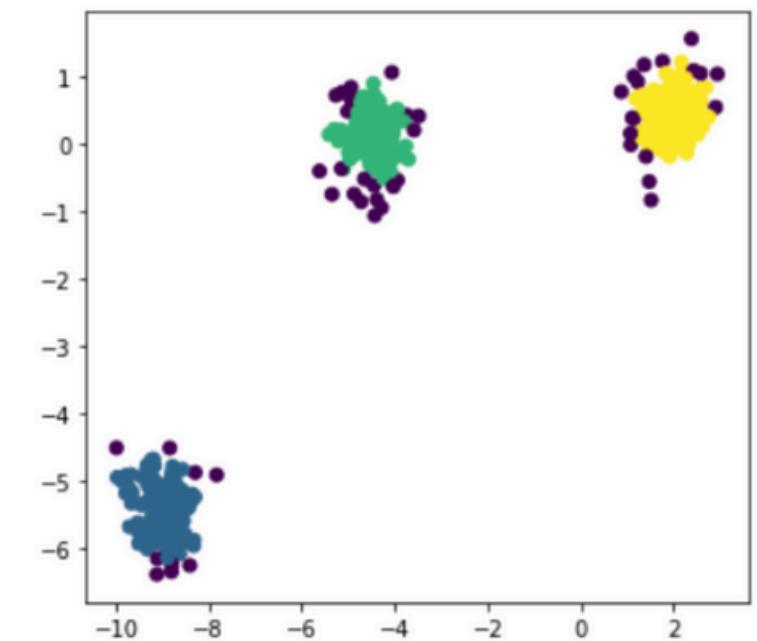
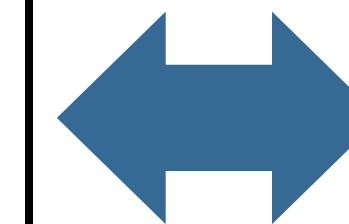
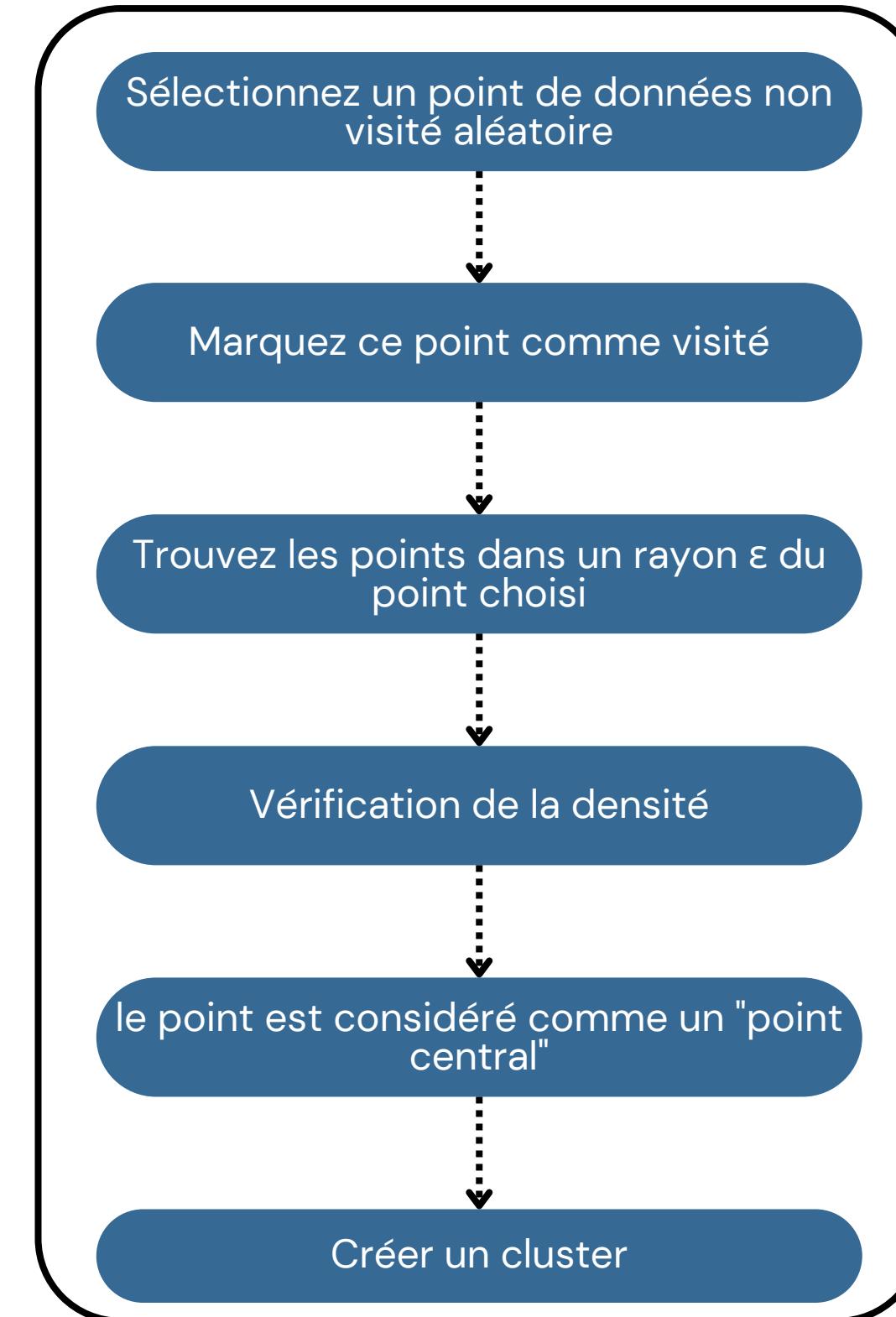
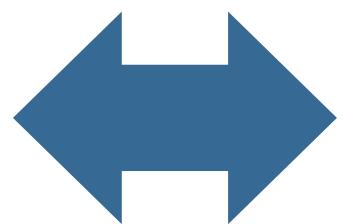
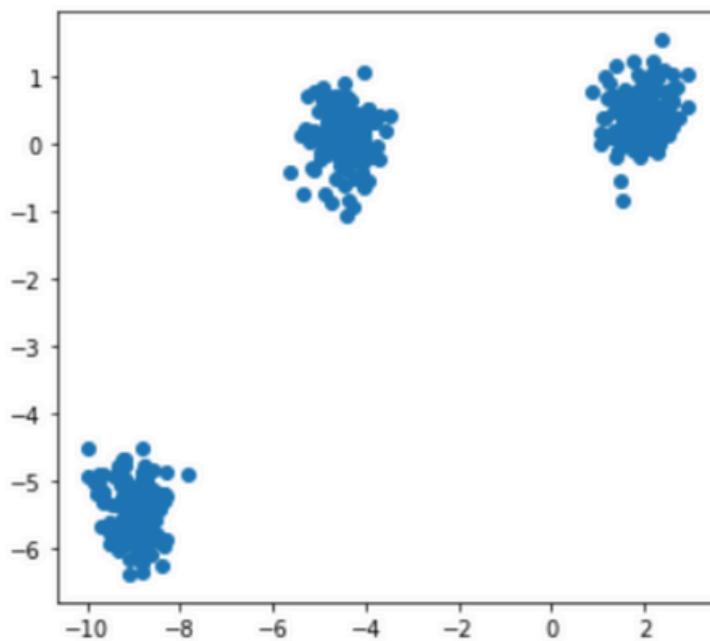


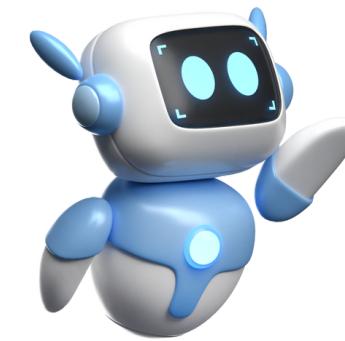
- Points centraux** : points ayant au moins MinPts voisins dans un rayon ϵ .
- Points de bordure** : voisins d'un point central, mais ayant moins de MinPts .
- Points bruit** : points isolés, ne faisant partie d'aucun cluster.



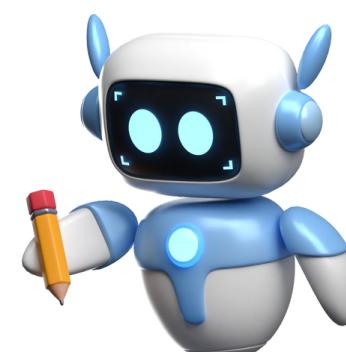


FONCTIONNEMENT DE L'ALGORITHME DBSCAN

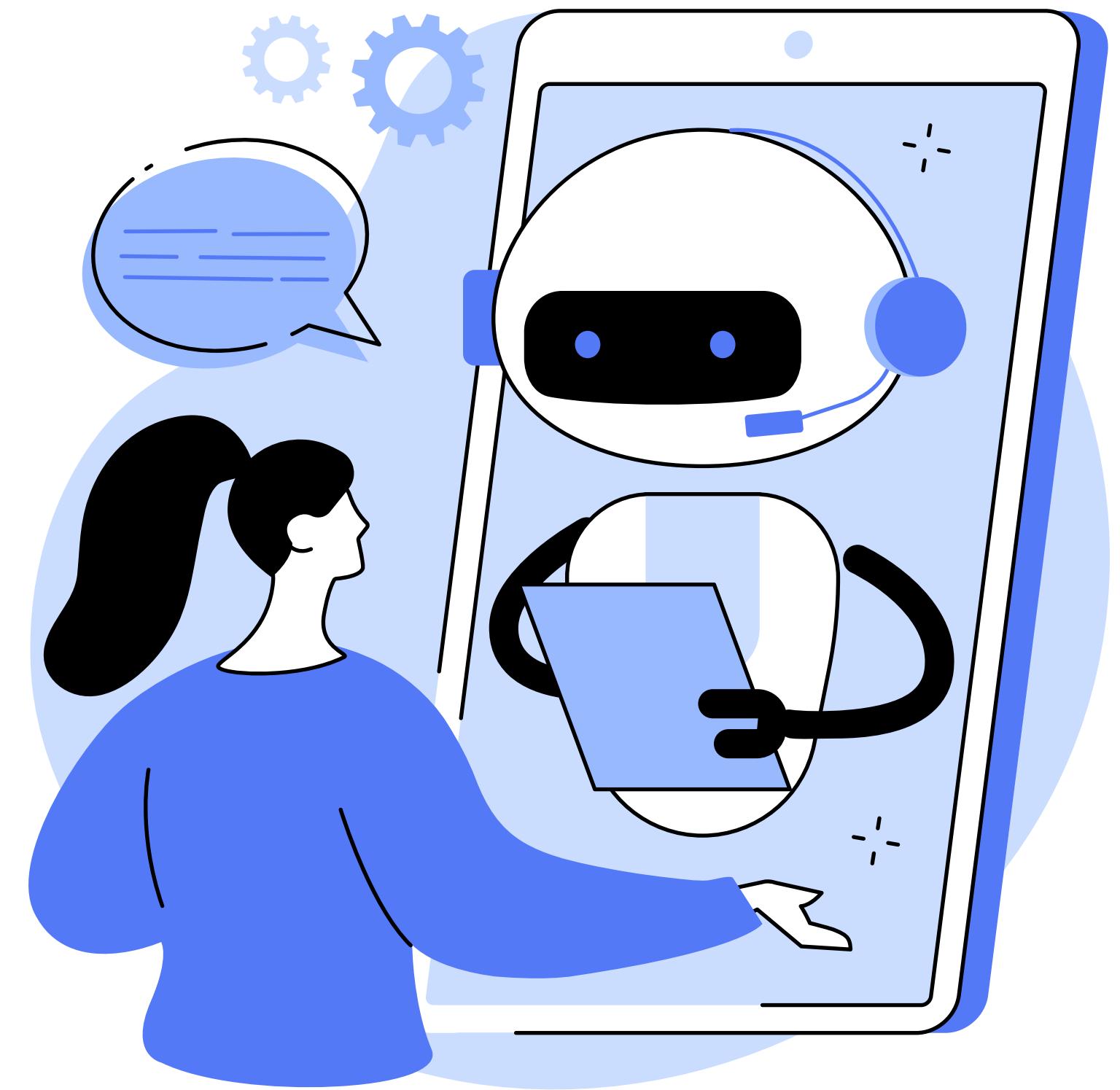


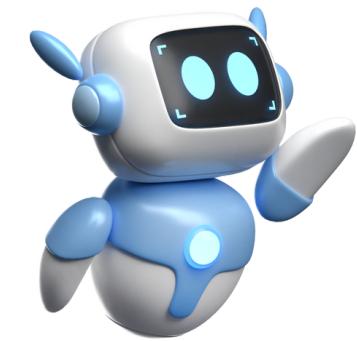


EXEMPLE D'APPLICATION



IMPLEMENTATION





OUTILS ET LANGAGES UTILISÉS



ANACONDA®

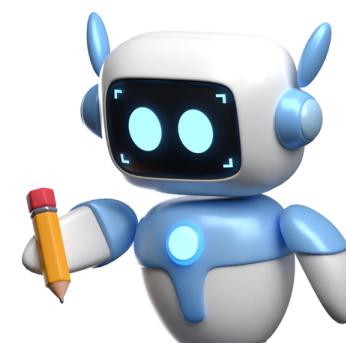


Anaconda

Jupyter

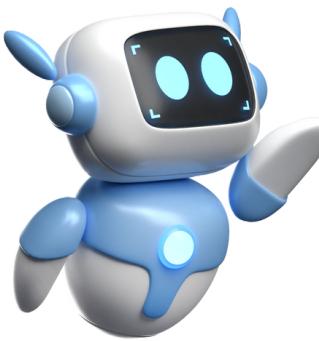
Github

Python



BIBLIOTHÈQUES UTILISÉS





STRUCTURE ET COLONNES DU DATASET

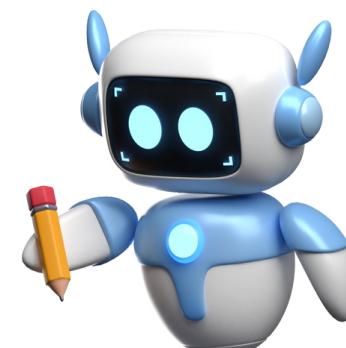
Structure.

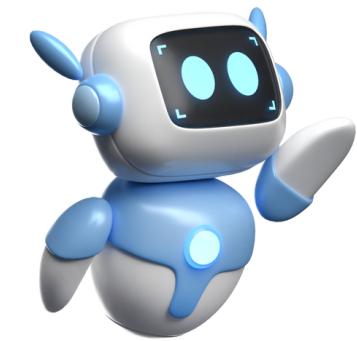
- Nombre total de lignes : **200**
- Nombre total de colonnes : **5**

Colonnes principales.

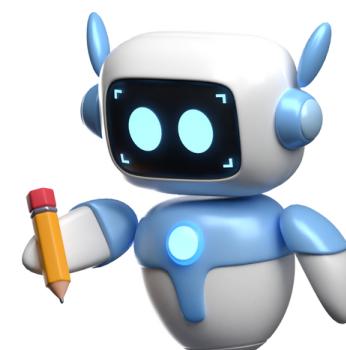
- CustomerID** : Identifiant unique de chaque client.
- Genre** : Sexe du client (Masculin/Féminin).
- Age** : Âge du client.
- Annual Income (k\$)** : Revenu annuel en milliers de dollars.
- Spending Score (1-100)** : Score indiquant les habitudes de dépenses du client (1 = Faible, 100 = Élevé).

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





PROCESSUS DE RÉALISATION DE LA SEGMENTATION CLIENT



Importation des
Bibliothèques
Nécessaires

Chargement et
Prétraitement du
Dataset

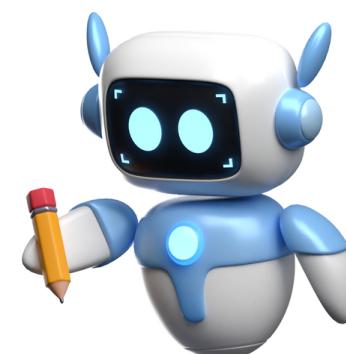
Analyse
Exploratoire des
Données (EDA)

Création des
Clusters avec K-
Means

Application du
Clustering avec
DBSCAN

Comparaison
entre K-Means et
DBSCAN

DÉMONSTRATION EN DIRECT



CONCLUSION





**MERCI POUR
VOTRE ATTENTION!**