Segmentez des clients d'un site e-commerce

Parcours Al Engineer



Notre mission:

Segmentation des clients pour établir une stratégie de communication et commerciale.

Les étapes d'analyse

Partie 1

Exploration du SQL

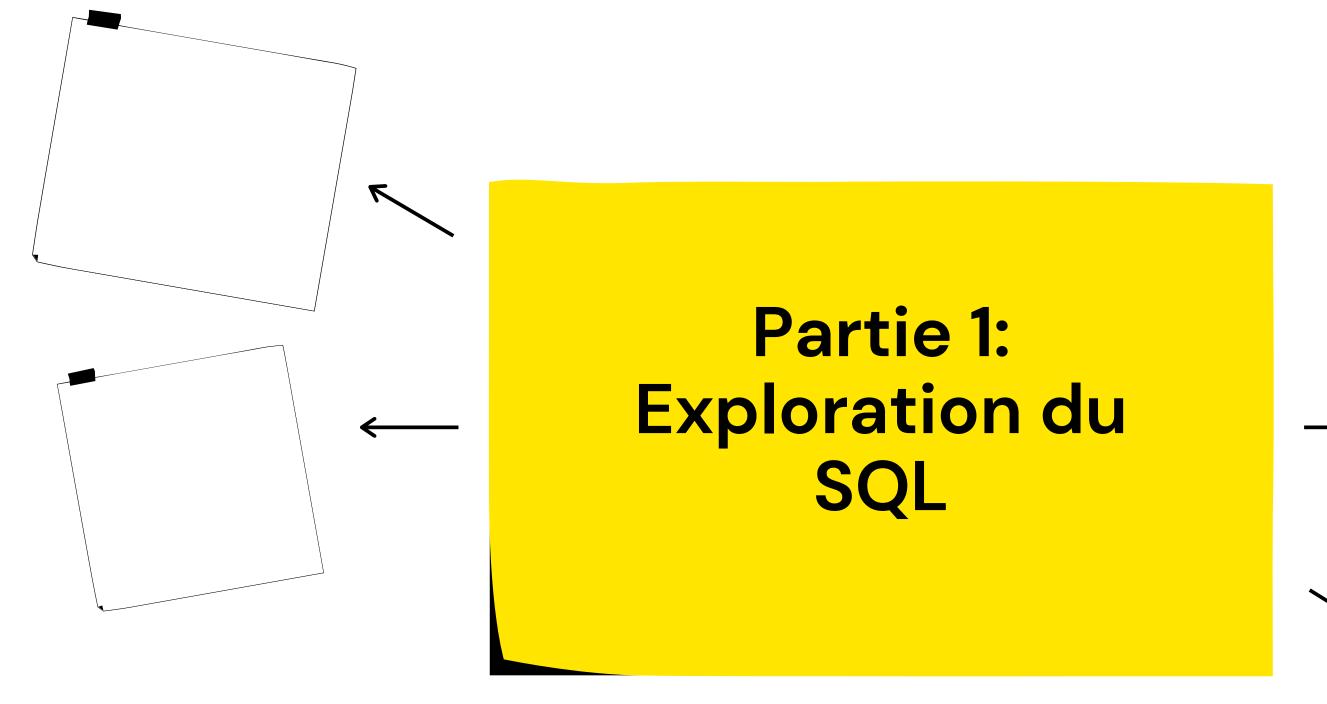
Partie 2

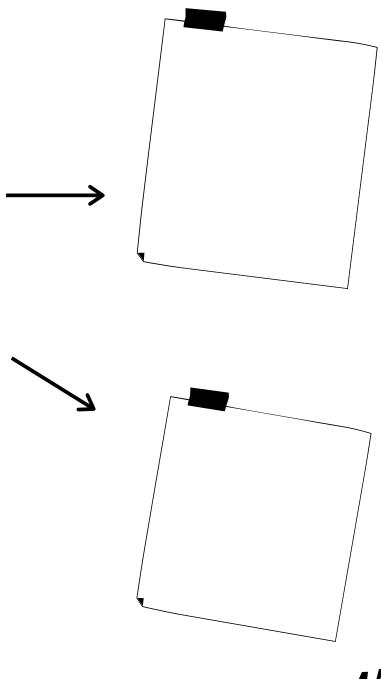
Test de modèles de Clustering Partie 3

Analyse et Stratégie

Partie 4

Contrat de maintenance

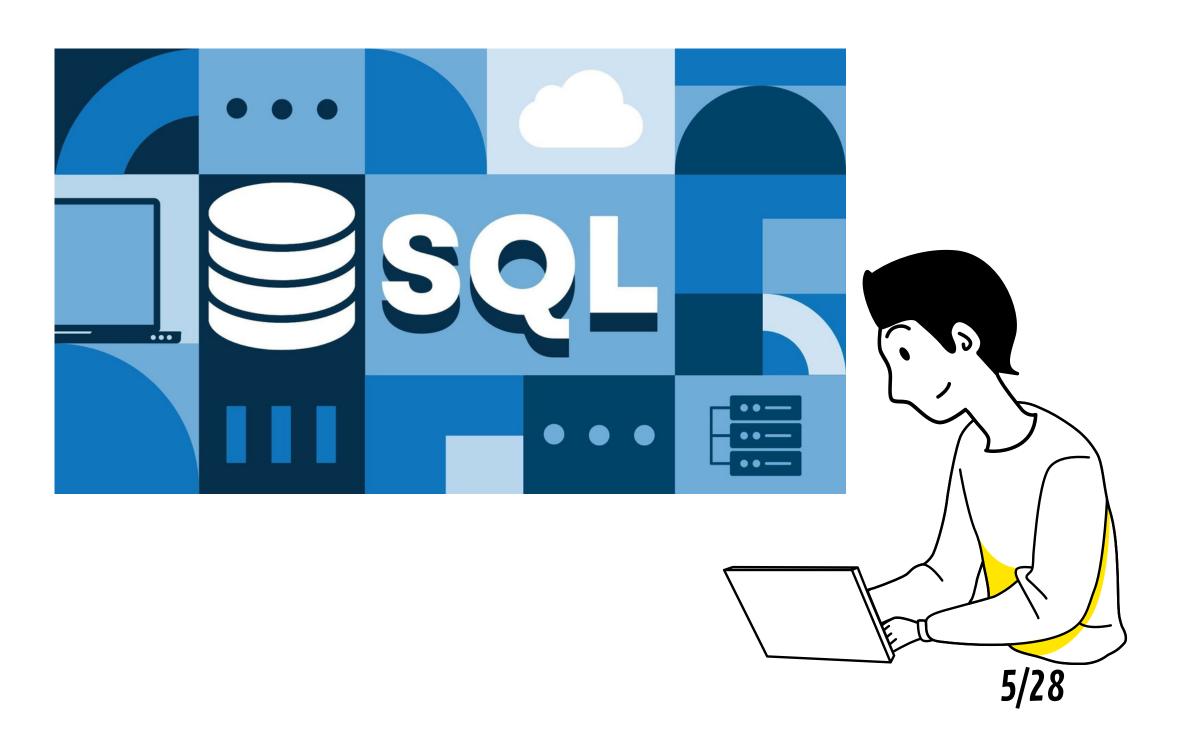




Partie 1: Exploration du SQL

Source SQL

- Exploration
- Réponse aux requêtes



Partie 1: Exploration du SQL

Mise en place sous Python

- Mise en place pour Python
- Création d'un jeu de donnée pour notre exploration.

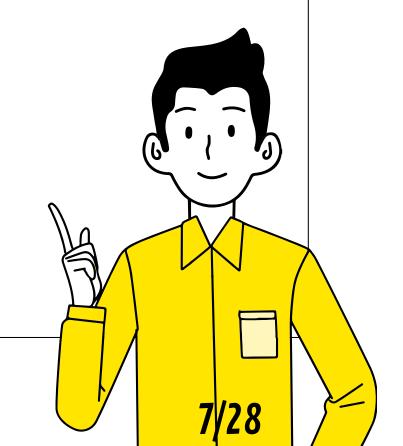
- customer_unique_id
- customer_zip_code_prefix
- customer_city
- customer_state
- nbr_commande
- Date_Premier_Commande
- Date_Derniere_Commande
- Total_achat

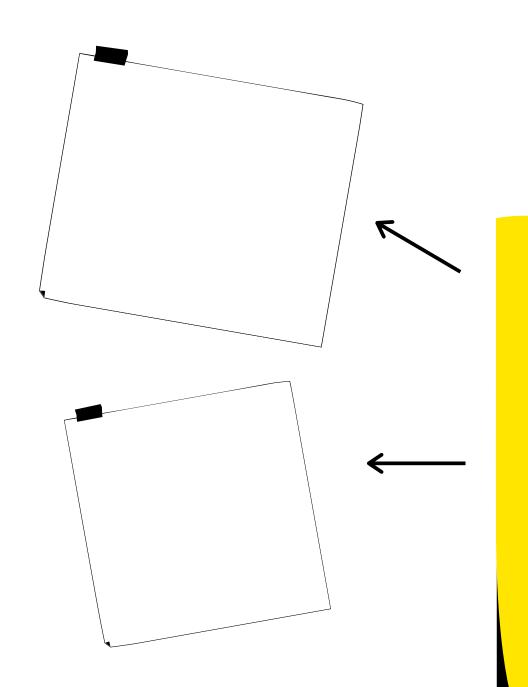


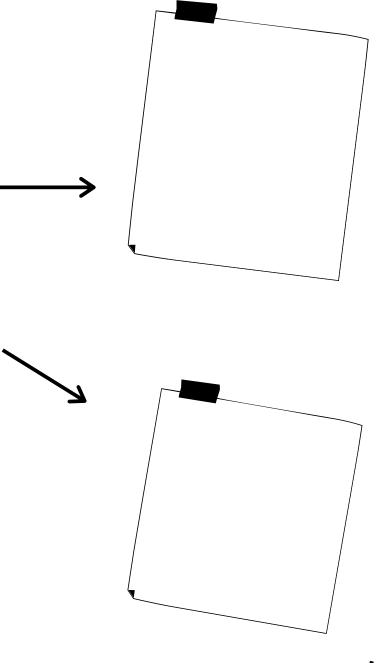
Partie 1: Exploration du SQL

Liste de manipulations

- Normalisation?
- Suppression du montant 13664.08
- Fréquence 17 conservé
- Création de RFM
- Colonne pour le futur





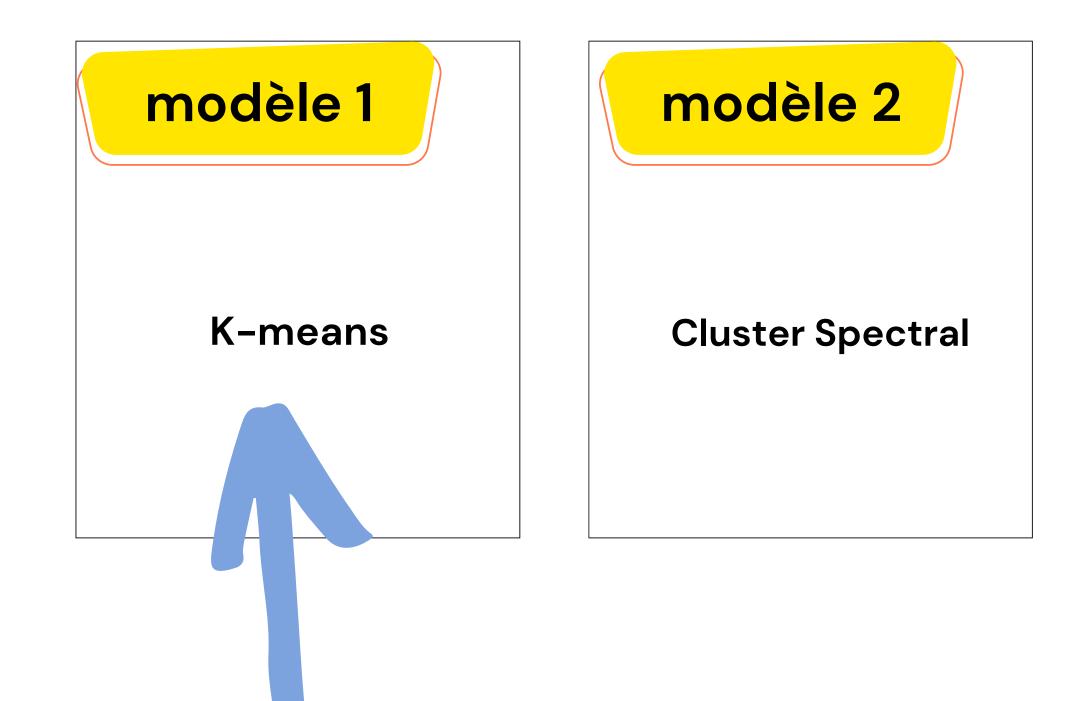


Comparaison de 2 modèles

Nos étapes

- Test global, premier résultat, premier test de performance
- On approfondi avec GridSearch, avec performance.
- Performance moyenne sur plusieurs itérations
- Application d'une échelle de temps
- On observe





modèle 1

K-means

Le modèle K-means est un algorithme de clustering qui partitionne un ensemble de données en K groupes distincts. Chaque observation est assignée au cluster dont le centroïde est le plus proche.

Nos Paramètres:

```
kmeans = KMeans(n_clusters=4, random_state=42)
kmeans.fit(data_rfm_scaled_imputed)
```



modèle 1

Le score moyen de silhouette pour les clusters est : 0.49



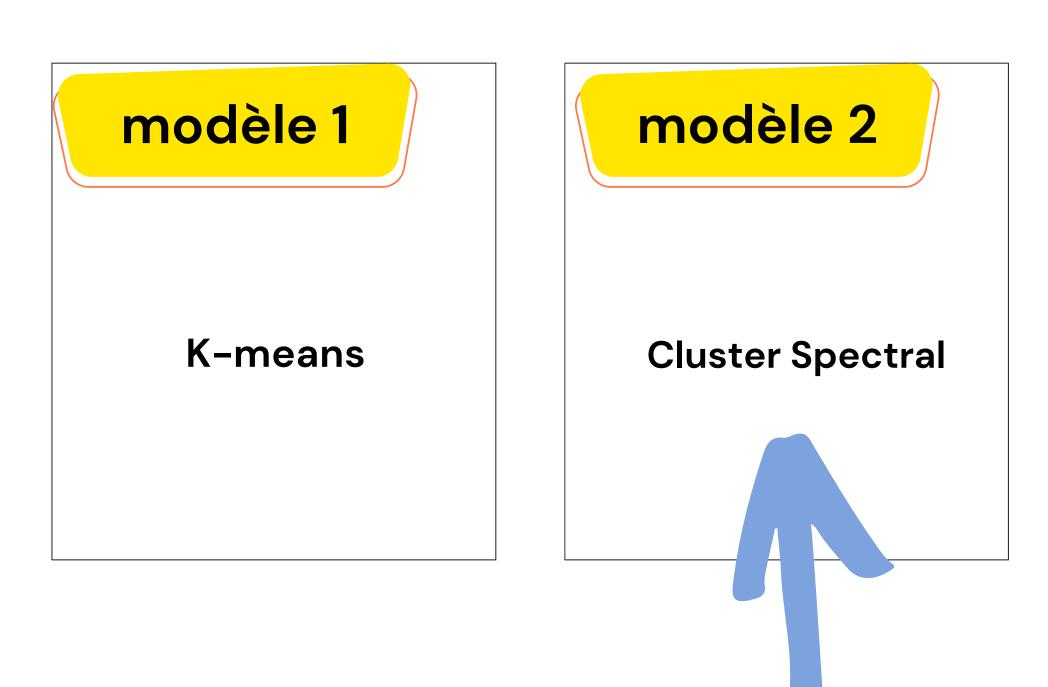
modèle 1

```
Moyenne des scores de Jaccard pour la stabilité des clusters : 0.3026637629545151
```

```
→ Meilleurs paramètres : {'init': 'k-means++', 'max_iter': 300, 'n_clusters': 4, 'n_init': 15}

Meilleur score de silhouette : 0.36858031281810943
```





modèle 2

Cluster Spectral

Le Cluster Spectral est une méthode de clustering basée sur la théorie des graphes et la décomposition spectrale de la matrice de similarité. Cette méthode permet de trouver des clusters de forme complexe en prenant en compte les relations de similarité entre les données.

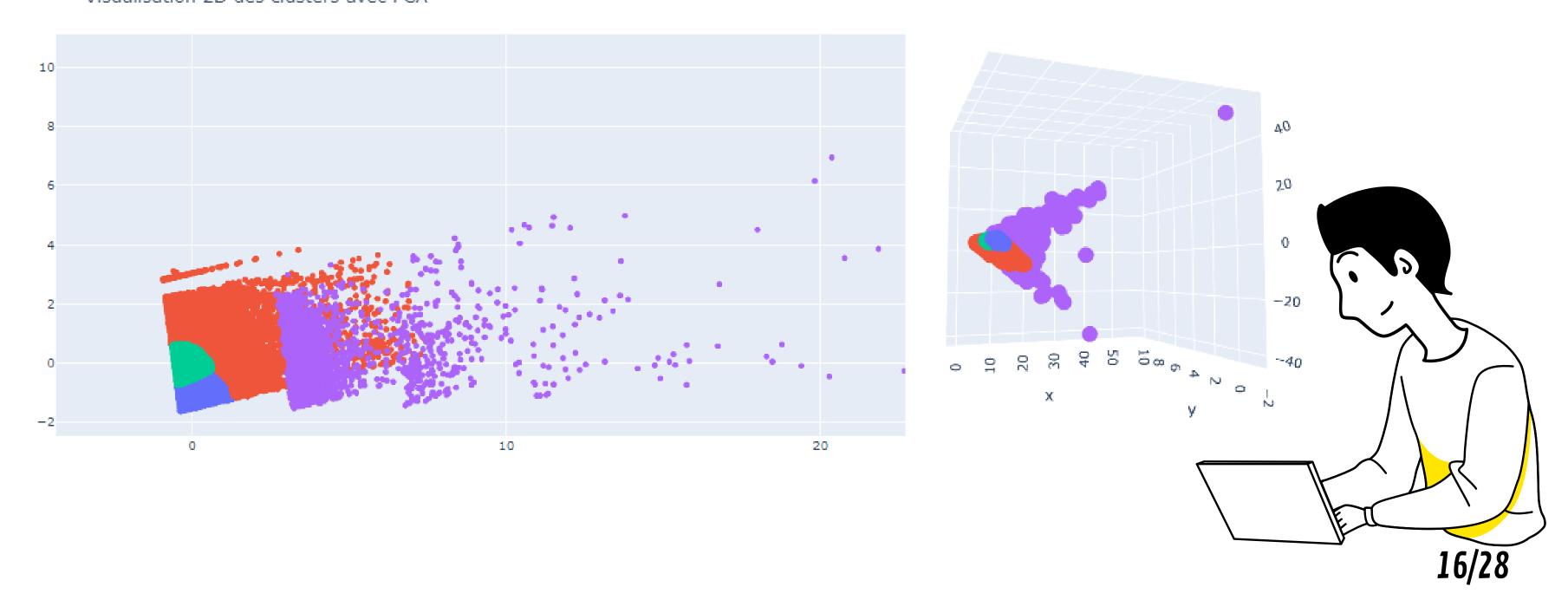
Nos Paramètres:

```
# Choix du nombre de clusters
n clusters = 4
# Création de l'instance de SpectralClustering
spectral clustering = SpectralClustering(n clusters=n clusters, affinity='nearest neighbors', assign labels='kmeans', random state=42)
# Application du clustering aux données
labels_spectral = spectral_clustering.fit_predict(data_rfm_scaled_imputed)
```

modèle 1

Visualisation 2D des clusters avec PCA

Le score moyen de silhouette pour les clusters est : 0.35



modèle 1

```
Moyenne des scores de Jaccard pour la stabilité des clusters : 0.26651244592642925
```

```
Meilleurs paramètres : {'init': 'k-means++', 'max_iter': 600, 'n_clusters': 4, 'n_init': 10}
Meilleur score de silhouette : 0.48791384680547284
```

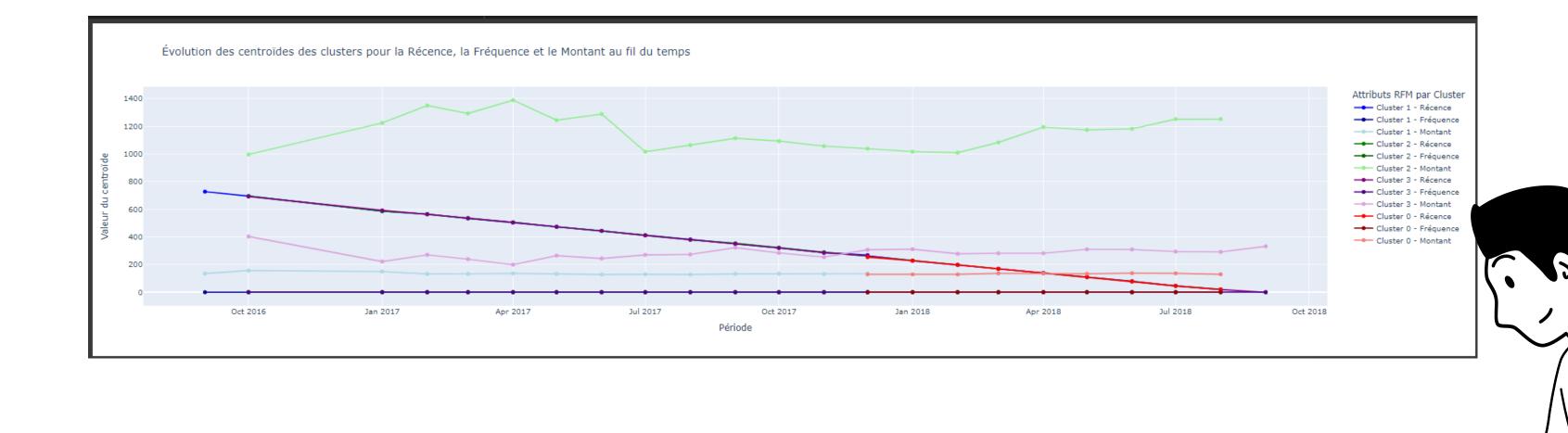


Mise en place des périodes

modèle 1 K-means

Partie 2: Mise en place des périodes

modèle 1



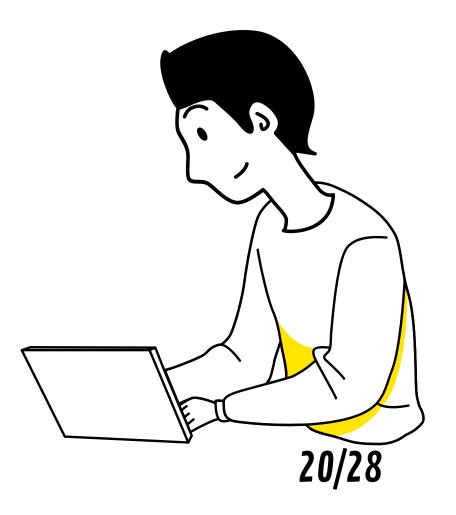
19/28

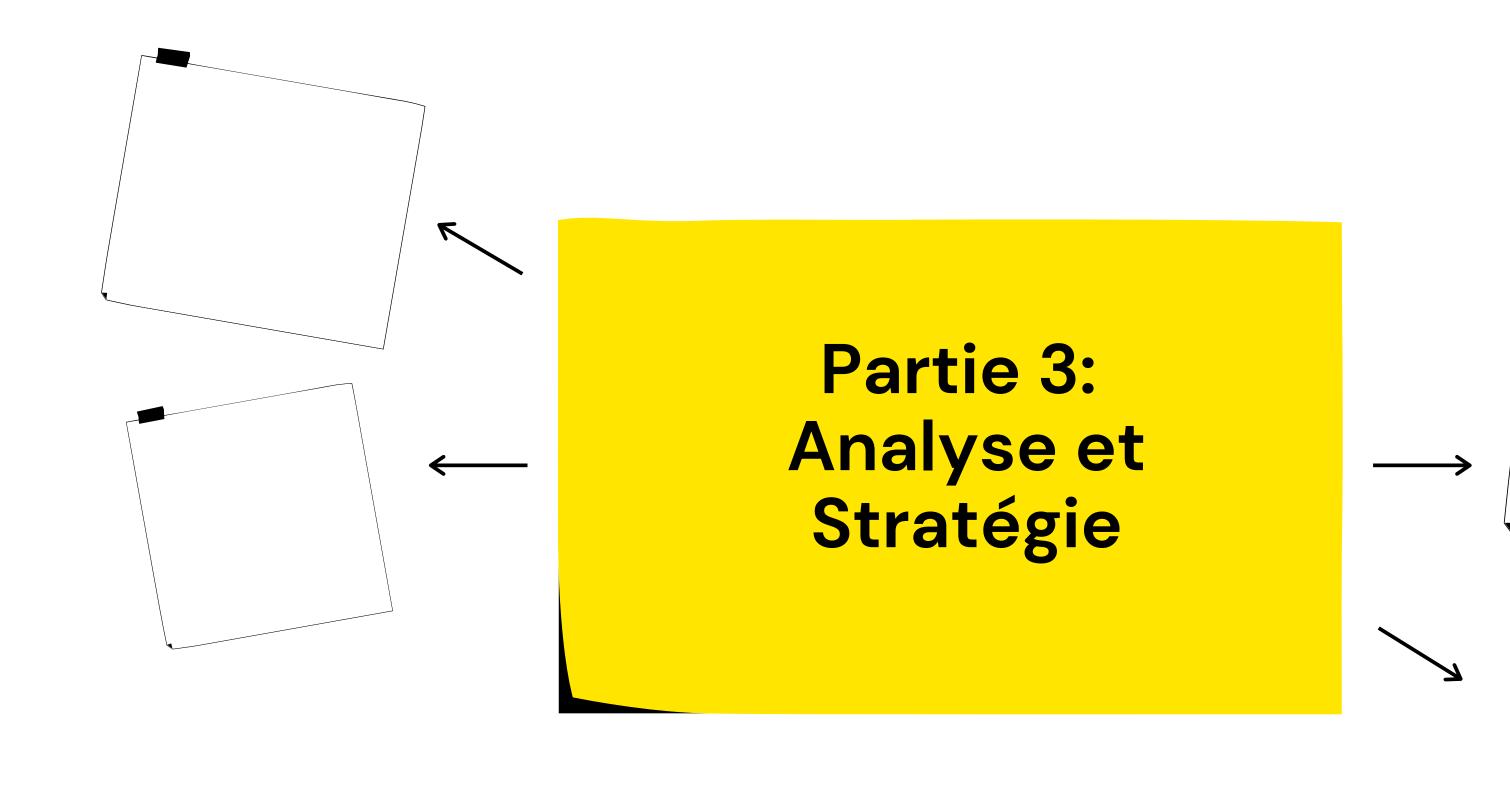
Partie 2: Mise en place des périodes

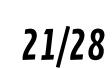
modèle 1

Test de Kolmogorov-Smirnov

```
Test K-S pour l'attribut: Récence
Clusters 0 vs 1: KS Statistic = 0.9997687148253797, P-Value = 0.0
Clusters 0 vs 2: KS Statistic = 0.4422397572964222, P-Value = 0.0
Clusters 0 vs 3: KS Statistic = 0.3907152168829928, P-Value = 0.0
Clusters 1 vs 2: KS Statistic = 0.5575289575289575, P-Value = 0.0
Clusters 1 vs 3: KS Statistic = 0.6090534979423868, P-Value = 0.0
Clusters 2 vs 3: KS Statistic = 0.07104810438143772, P-Value = 3.1692890326533478e-06
Test K-S pour l'attribut: Fréquence
Clusters 0 vs 1: KS Statistic = 0.0, P-Value = 1.0
Clusters 0 vs 2: KS Statistic = 0.011196911196911241, P-Value = 0.9130157603208073
Clusters 0 vs 3: KS Statistic = 1.0, P-Value = 0.0
Clusters 1 vs 2: KS Statistic = 0.011196911196911241, P-Value = 0.917762803915048
Clusters 1 vs 3: KS Statistic = 1.0, P-Value = 0.0
Clusters 2 vs 3: KS Statistic = 0.9888030888030888, P-Value = 0.0
Test K-S pour l'attribut: Montant
Clusters 0 vs 1: KS Statistic = 0.02062029681583455, P-Value = 1.2965740754704273e-08
Clusters 0 vs 2: KS Statistic = 0.997744969547452, P-Value = 0.0
Clusters 0 vs 3: KS Statistic = 0.41979567905993526, P-Value = 0.0
Clusters 1 vs 2: KS Statistic = 0.9963529137890897, P-Value = 0.0
Clusters 1 vs 3: KS Statistic = 0.4315038506601064, P-Value = 0.0
Clusters 2 vs 3: KS Statistic = 0.9281106503328725, P-Value = 0.0
```

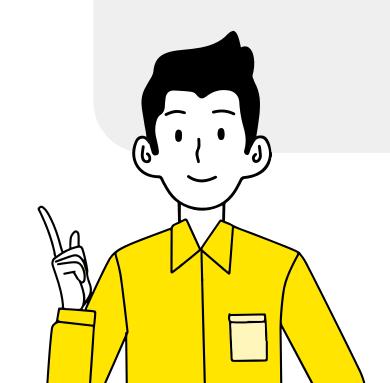






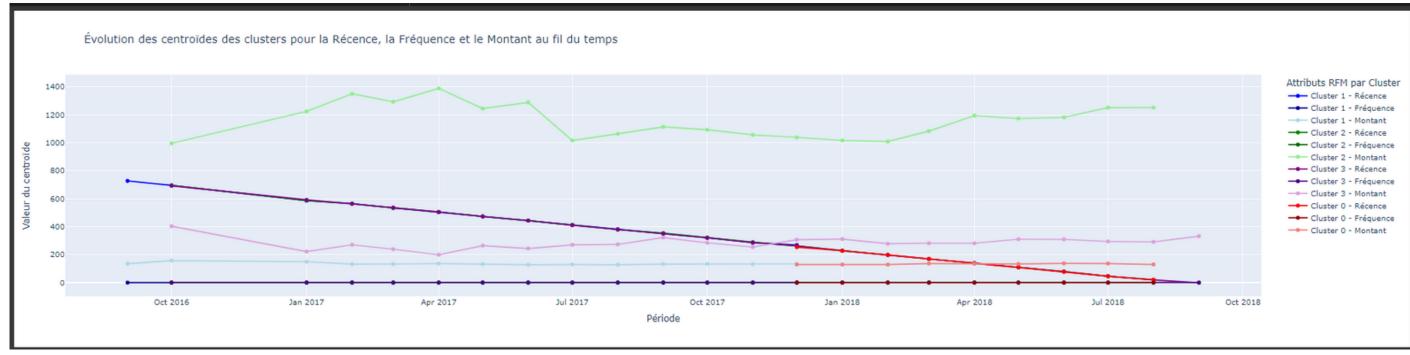
Idée d'utilisation

Confronter nos Clusters avec nos informations clients



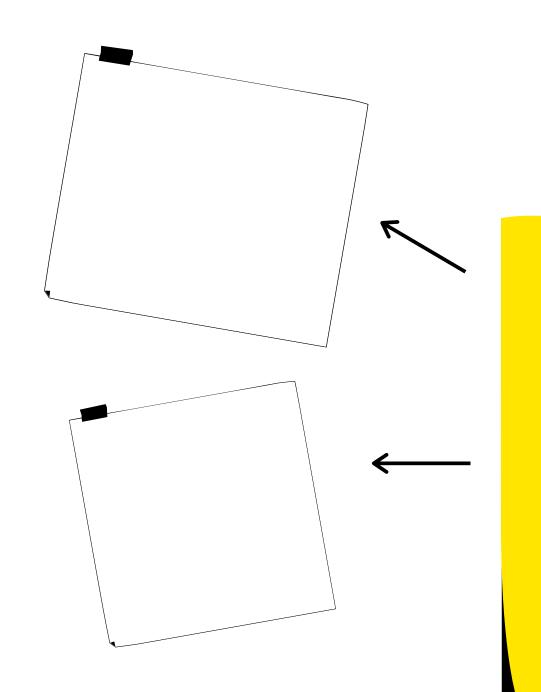
Partie 3: Partie 3: Analyse et Stratégie

modèle 1

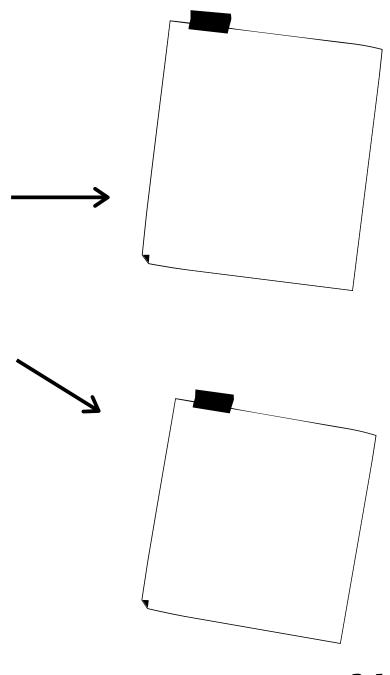


- customer_unique_id
- customer_zip_code_prefix
- customer_city
- customer_state
- nbr_commande
- Date_Premier_Commande
- Date_Derniere_Commande
- Total_achat

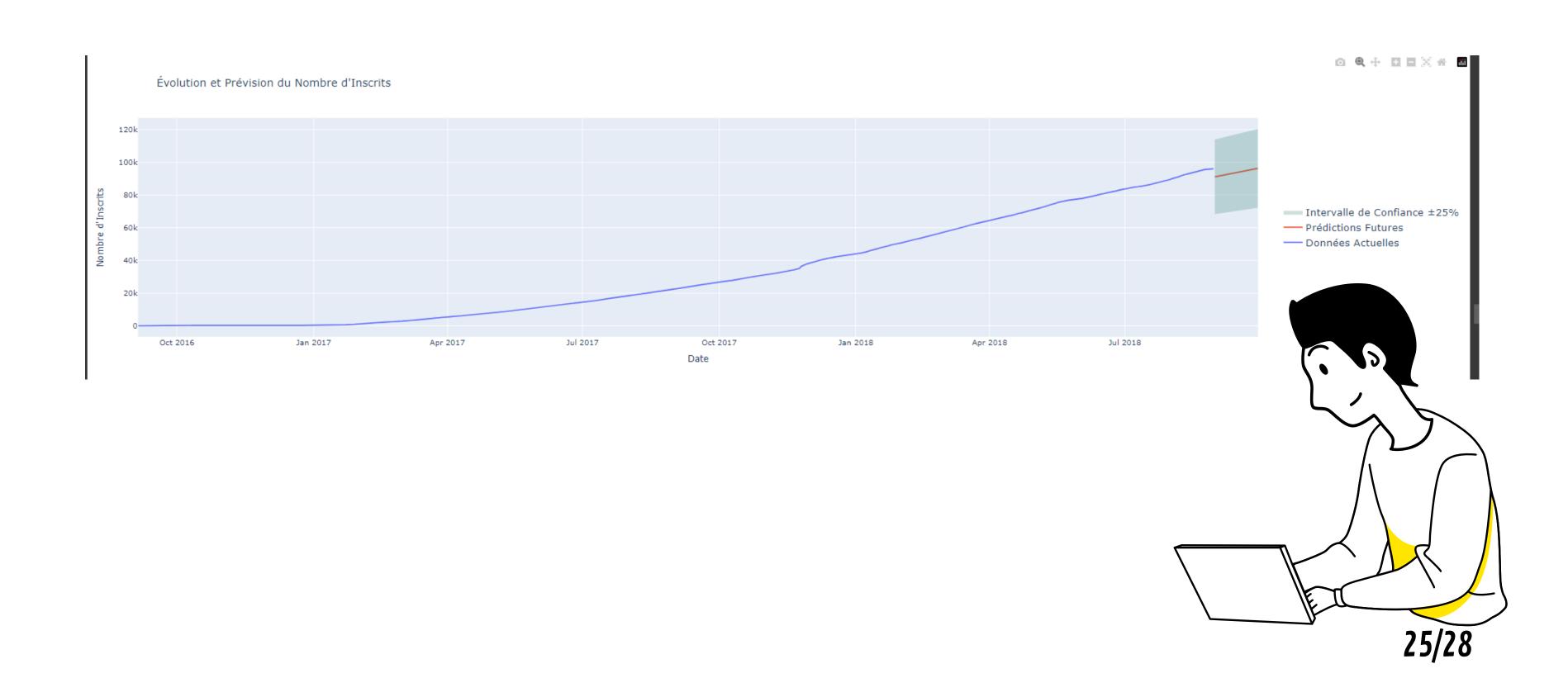




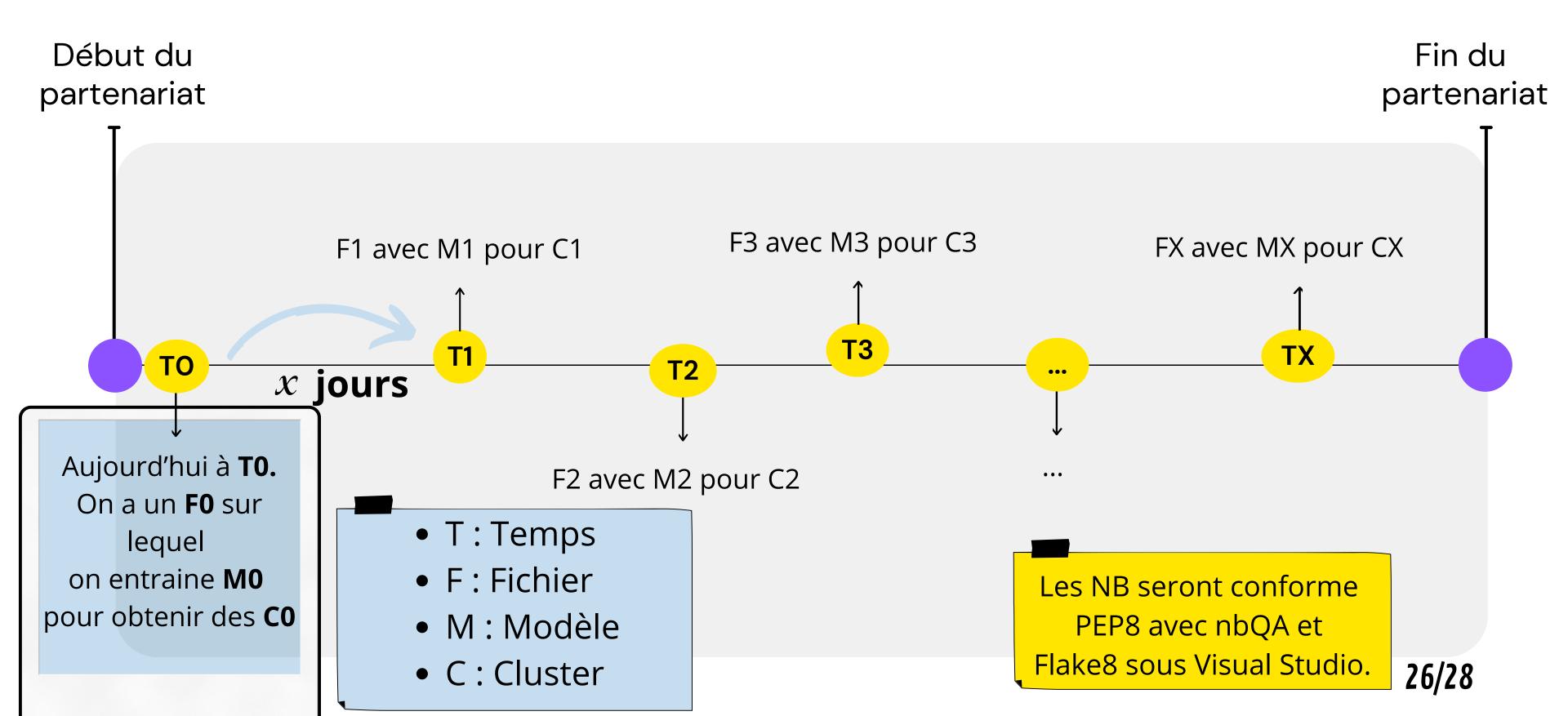
Partie 4: Contrat de maintenance



Partie 4 Contrat de maintenance



Étape 4 : Contrat de maintenance



Partie 4 Contrat de maintenance

- 1. Fréquence
- 2. Adapté les outils de performances
- 3. Mise à jour du modèle
- 4. Besoin d'ajustement si nouvelle donnée
- 5. Coûts en ressources
- 6. Plan de secours



Merci!

