**Diapositive 2 : Segmentation – Diviser pour mieux comprendre**

**Exemple : *Segmentation des clients par comportement de paiement dans un projet de recouvrement***

* **Méthodologie** :
  + **Clustering** : Vous pouvez utiliser des algorithmes de segmentation comme le *K-Means* ou le *DBSCAN* pour regrouper les clients en fonction de variables clés comme le taux de prélèvement réussi, le nombre de prélèvements échoués, la durée moyenne en recouvrement, ou encore le solde moyen du compte.
  + Le choix des variables utilisées pour la segmentation est crucial. Par exemple, en analysant l'historique des comportements de paiement, on pourrait regrouper les clients en fonction de leur régularité dans les paiements (ex : jamais en retard, parfois en retard, systématiquement en retard).
* **Déroulement technique** :
  + On utilise des variables comportementales comme input pour l’algorithme de clustering, après les avoir normalisées. Si vous optez pour *K-Means*, il faudra tester plusieurs valeurs de k (le nombre de clusters) et évaluer la qualité de la segmentation à l'aide d'un indice comme le *Silhouette Score*.
  + Une autre option serait d’utiliser des algorithmes de clustering hiérarchique si le nombre de segments à déterminer est flou, ce qui permet de visualiser les distances entre groupes grâce à un dendrogramme.
* **Application dans le recouvrement** :
  + La segmentation permet de cibler des stratégies différentes selon les groupes : par exemple, les clients à "faible risque" peuvent être soumis à des relances automatiques et moins coûteuses, tandis que les clients à "haut risque" pourraient être contactés par des conseillers spécialisés plus rapidement. Cette segmentation permettrait d'optimiser les ressources et les interventions.

**Discussion avec le chef :**

* Vous pourriez discuter des différentes méthodes de clustering envisageables (K-Means vs DBSCAN vs clustering hiérarchique), en fonction de la distribution des données et du volume d’échantillons disponibles.
* Le choix des features pour la segmentation est également un sujet clé à approfondir avec lui, en veillant à ce qu'elles reflètent le comportement à analyser de manière pertinente.

**Diapositive 3 : Rule Mining – Découvrir des patterns cachés**

**Exemple : *Découverte de relations cachées entre échec de prélèvement et risque de recouvrement***

* **Méthodologie** :
  + Le *rule mining*, et plus spécifiquement l'algorithme **Apriori**, est utilisé pour identifier des règles d’association à partir des transactions historiques.
  + L'algorithme cherche des ensembles d'items (ou événements) qui coexistent fréquemment dans les données. Par exemple, vous pouvez identifier que les clients qui manquent deux prélèvements consécutifs ont une probabilité élevée d'entrer en recouvrement dans les 30 jours.
* **Déroulement technique** :
  + Les étapes clés de cette méthode incluent la définition de seuils comme le **support**, la **confiance** et le **lift** pour chaque règle. Le support reflète la fréquence avec laquelle une combinaison d'événements se produit (par exemple, X = "deux prélèvements échoués"), et la confiance représente la probabilité que Y (passage en recouvrement) suive X.
  + On applique l’algorithme Apriori sur un dataset de prélèvements en échec, avec des caractéristiques comme le nombre de prélèvements échoués, la fréquence de ces échecs, et le délai entre les échecs.
* **Application dans le recouvrement** :
  + Une règle du type "Si deux prélèvements échouent consécutivement, alors 80 % des clients passent en recouvrement dans les 30 jours" pourrait permettre de définir des stratégies préventives : par exemple, contacter ces clients dès le deuxième prélèvement échoué pour éviter leur passage en recouvrement.

**Discussion avec le chef :**

* Vous pourriez parler de l'ajustement des seuils de support et de confiance pour filtrer les règles les plus pertinentes et éviter les fausses corrélations.
* Discuter de l’implémentation d’algorithmes alternatifs comme *FP-Growth* si le volume de données est très large ou si les règles générées par Apriori sont trop nombreuses et peu exploitables.
* Vous pourriez aussi envisager de combiner le rule mining avec une analyse séquentielle si les événements doivent être analysés dans l’ordre.

**Diapositive 4 : Scoring – Anticiper le futur pour mieux agir**

**Exemple : *Prédiction du risque de recouvrement à partir d'un score calculé sur les contrats***

* **Méthodologie** :
  + Utilisation de modèles de machine learning supervisés pour prédire la probabilité qu'un contrat passe en recouvrement dans un certain délai.
  + Les algorithmes classiques comme la **régression logistique**, les **arbres de décision**, ou des méthodes plus avancées comme les **forêts aléatoires** ou le **gradient boosting** (*XGBoost*) peuvent être utilisés ici.
  + Les variables (features) peuvent inclure : l’historique des paiements (taux de réussite/échec des prélèvements), la durée de la relation avec la banque, le type de contrat, le solde du compte, les incidents passés, etc.
* **Déroulement technique** :
  + Le modèle apprend sur des données historiques où l’issue (passage en recouvrement ou non) est connue.
  + Chaque contrat se voit attribuer un **score** entre 0 et 1 (ou entre 0 et 100, selon le modèle), correspondant à la probabilité de passage en recouvrement. Plus le score est élevé, plus le contrat est à risque.
  + Le modèle est évalué sur des métriques telles que l'**AUC-ROC** (courbe ROC), qui montre la capacité du modèle à discriminer les contrats à risque de ceux qui ne le sont pas.
* **Application dans le recouvrement** :
  + Ce score permet de prioriser les contrats pour les relances. Par exemple, tous les contrats ayant un score supérieur à 0,7 pourraient être soumis à des actions préventives immédiates, tandis que ceux en dessous de 0,3 pourraient être surveillés avec moins d’attention.

**Discussion avec le chef :**

* Avec votre chef, vous pouvez discuter de la pertinence des différentes approches de modélisation. Par exemple, la régression logistique est souvent un bon point de départ pour un modèle interprétable, mais des modèles plus complexes comme **XGBoost** peuvent offrir une meilleure performance au prix d'une interprétabilité plus faible.
* Il serait également intéressant de discuter du choix des **features** (variables) à inclure dans le modèle et de la gestion des données déséquilibrées (si le nombre de contrats qui passent en recouvrement est faible par rapport au nombre total de contrats).
* Enfin, vous pourriez aborder la question de la calibration des scores et l'ajustement du seuil de décision pour maximiser les bénéfices tout en minimisant les faux positifs (contrats à faible risque classés comme à haut risque).

**Conclusion pour vos échanges avec le chef :**

Chacune de ces méthodes a ses spécificités techniques, et la discussion avec votre chef pourrait porter sur les choix méthodologiques (algorithmes à utiliser, paramétrage, etc.), les défis liés à la qualité et la disponibilité des données (manque de données, données déséquilibrées), ainsi que l'intégration pratique des résultats (comment ces méthodes s'intègrent dans les processus décisionnels de la banque).