

# Questionnaire d'analyse exploratoire (EDA)

## Dataset Customer Churn

### **I-Analyse de forme (structure et qualité des données)**

1. Combien de lignes (clients) et de colonnes contient la dataset ?
2. Chaque observation correspond-elle bien à un client unique ?
3. Quels sont les types de variables présentes (numériques, catégorielles, binaires) ?
4. Quelle est la variable cible du problème ?
5. La variable *Churn* est-elle bien définie et cohérente ?
6. Existe-t-il des valeurs manquantes dans la dataset ?
7. Quelles variables contiennent des valeurs manquantes ?
8. Comment expliquer la présence de ces valeurs manquantes ?
9. Quelles stratégies peuvent être utilisées pour traiter ces valeurs ?
10. Observe-t-on des valeurs aberrantes dans les variables numériques ?
11. Certaines variables sont-elles redondantes ou peu informatives ?
12. Faut-il supprimer ou transformer certaines colonnes ?

### **II-Analyse de fond – distributions des variables**

13. Quelle est la distribution de la variable cible *Churn* ?
14. Observe-t-on un déséquilibre entre les clients qui partent et ceux qui restent ?
15. Quelle est la distribution de l'ancienneté (*tenure*) des clients ?
16. Les clients récents sont-ils plus nombreux que les clients anciens ?
17. Comment se répartissent les charges mensuelles (*MonthlyCharges*) ?
18. Existe-t-il des groupes de clients avec des charges très élevées ?
19. Quelle est la distribution du montant total payé (*TotalCharges*) ?
20. Comment *TotalCharges* évolue-t-il en fonction de l'ancienneté ?

### **III-Analyse du churn selon les variables clés**

21. Le taux de churn varie-t-il selon l'ancienneté du client ?
22. Les clients avec une faible ancienneté churnent-ils davantage ?
23. Les clients ayant des charges mensuelles élevées quittent-ils plus souvent le service ?

- 24. Le type de contrat influence-t-il le taux de churn ?
- 25. Les contrats mensuels sont-ils plus à risque que les contrats longue durée ?
- 26. La méthode de paiement a-t-elle un impact sur le churn ?
- 27. La facturation électronique influence-t-elle le comportement des clients ?

## **IV-Analyse des variables catégorielles et démographiques**

- 28. Le sexe du client a-t-il une influence sur le churn ?
- 29. Les clients seniors quittent-ils plus souvent le service ?
- 30. Les clients ayant un partenaire churnent-ils moins ?
- 31. Les clients avec des personnes à charge sont-ils plus fidèles ?
- 32. L'utilisation de plusieurs services réduit-elle le risque de churn ?
- 33. Le support technique a-t-il un impact sur la fidélité du client ?

## **V-Analyse des corrélations et relations entre variables**

- 34. Existe-t-il une corrélation entre *tenure* et *Churn* ?
- 35. Quelle est la relation entre *MonthlyCharges* et *Churn* ?
- 36. Les variables financières sont-elles corrélées entre elles ?
- 37. Peut-on identifier des variables fortement liées au churn ?
- 38. Les corrélations observées confirment-elles les hypothèses métier ?

## **VI-Synthèse exploratoire et lien avec le MLOps**

- 39. Quelles sont les variables les plus importantes pour expliquer le churn ?
- 40. Peut-on identifier un profil type de client à risque ?
- 41. Quelles transformations de données sont nécessaires avant la modélisation ?
- 42. Quelles variables doivent être encodées ?
- 43. Quels enseignements principaux peut-on tirer de l'analyse exploratoire ?
- 44. Comment ces résultats peuvent-ils aider à améliorer la performance du modèle ?
- 45. Pourquoi cette analyse exploratoire est-elle essentielle dans un pipeline MLOps ?