

Analyse Médical et Reporting

Contexte

Dans le cadre de l'activité de notre entreprise d'assurance-vie, il est essentiel de disposer de données médicales fiables et exploitables pour mieux comprendre les risques, ajuster nos offres et optimiser nos processus décisionnels. Ce projet a pour but d'extraire des informations pertinentes du dataset médical afin d'identifier les tendances en matière de risques et d'améliorer la gestion des sinistres.

Objectifs

L'objectif principal de ce projet est de fournir des analyses détaillées des données médicales, qui permettront de :

1. **Analyser les risques médicaux** : Identifier les pathologies les plus fréquentes et comprendre leur impact sur les coûts.
2. **Optimiser les politiques de tarification** : Utiliser les données pour ajuster les primes et mieux gérer le risque des sinistres.
3. **Suivi des indicateurs de performance** : Mettre en place des indicateurs permettant de suivre l'évolution des sinistres et des coûts médicaux.
4. **Fournir des rapports réguliers** : Créer des tableaux de bord et des rapports automatisés pour la direction et les équipes métier.

Livrables Attendus

1. **Rapport Analytique Détaillé**:
 - Une analyse descriptive des données, incluant les informations démographiques (sexe, âge, etc.) et les types de pathologies.
 - Analyse des coûts médicaux : Coûts moyens par pathologie, comparaison entre les sexes (male/female).
 - Analyse des facteurs de risque, avec des insights sur l'impact de l'âge, du sexe et des antécédents médicaux sur la probabilité de sinistres.
2. **Tableaux de Bord Dynamiques**
 - Tableau de bord interactif permettant de filtrer les résultats par genre (Male/Female), par âge, par type de pathologie, et par région.
 - Indicateurs clés (KPIs) comme le coût moyen par sinistre, ratio sinistres/primes, et prévalence des pathologies.
 - Comparaison des risques entre différents groupes démographiques.

3. Modèles Prédictifs

- Modèles de machine learning pour prédire la probabilité de sinistre en fonction des caractéristiques démographiques et médicales.
- Estimations des coûts futurs basées sur les tendances actuelles.

4. Automatisation des Rapports

- Rapports automatisés (quotidiens, hebdomadaires ou mensuels) générés à partir des tableaux de bord et des analyses.
- Exportations en formats Excel et PDF, avec des graphiques et des résumés clairs pour la prise de décision.

Spécifications Fonctionnelles

1. Préparation des Données

- Nettoyage des données (suppression des doublons, gestion des valeurs manquantes, standardisation des valeurs).
- Transformation des données qualitatives, comme la variable **Gender**, où **Male** sera codé en 0 et **Female** en 1 ou en une autre codification définie par l'équipe.
- Vérification de la cohérence des données, en particulier pour les âges, les coûts et les types de pathologies.

2. Analyse Exploratoire des Données (EDA)

- Statistiques descriptives pour chaque variable : Moyenne, médiane, écart-type, répartition par genre, âge et pathologie.
- Corrélation entre les variables démographiques (genre, âge) et les coûts médicaux.
- Visualisation des données sous forme de graphiques : histogrammes, bar charts, pie charts, et heatmaps.

3. Modélisation Prédictive

- Utilisation d'algorithmes de machine learning pour prédire les risques de sinistres : régression logistique, arbres de décision, ou random forests.
- Création d'un modèle de régression pour estimer le coût moyen des sinistres en fonction des pathologies et des facteurs de risque démographiques.

4. Visualisation et Reporting

- Création de tableaux de bord sous Power BI ou Tableau, connectés aux données pour visualiser en temps réel les indicateurs clés.

- Export des rapports sous format Excel et PDF, avec les graphiques et tableaux résumant les informations.

Contraintes

1. Confidentialité et Sécurité des Données

- Le dataset contient des informations médicales sensibles. Toutes les analyses et reportings doivent respecter les normes de confidentialité des données (GDPR et autres réglementations en vigueur).

2. Qualité des Données

- Les données doivent être nettoyées, validées, et vérifiées avant toute analyse. Les valeurs manquantes doivent être traitées de manière appropriée (par imputation ou suppression).

Délais et Ressources

1. Délais

- Phase de préparation des données : 1 semaine.
- Analyse exploratoire des données : 1 semaine.
- Création des tableaux de bord et rapports : 2 semaines.
- Modélisation prédictive : 1 semaine.

2. Ressources Nécessaires

- Accès aux données médicales et aux outils de visualisation de données (Power BI/Tableau).
- Outils pour l'analyse de données : Python ou R avec les bibliothèques de machine learning (pandas, scikit-learn, etc.).
- Collaboration avec l'équipe métier pour comprendre les besoins spécifiques de reporting.