

## **A . Pour le projet Assurance en SAS**

La compagnie d'assurance Zenor, spécialisée dans l'assurance des véhicules à deux roues, nous a soumis un projet pour le compte de leur entreprise. Afin d'évaluer la situation de la société et d'atteindre ses objectifs, nous avons analysé trois bases de données distinctes : clients, contrat1 et contrat2. Ces bases nous ont fourni des informations détaillées sur nos assurés ainsi que sur les termes de leurs contrats.

Zenor a défini trois objectifs qu'elle vise à réaliser grâce à notre analyse. Elle ambitionne d'accroître sa part de marché de 10 %, de réduire de 3 points son taux de résiliation et d'augmenter de 5 points la proportion de clients de moins de 30 ans.

Cette analyse a été menée à l'aide du langage SAS, largement utilisé par de nombreuses compagnies d'assurance pour effectuer des analyses statistiques. Ce langage permet la création et la gestion efficace de bases de données pour mener des études approfondies. SAS facilite également la génération et la diffusion de rapports synthétiques, ce qui s'avère précieux dans l'analyse statistique.

Dans un premier temps, pour garantir une analyse pertinente, nous avons procédé à une segmentation du profil client afin de simplifier la manipulation et la compréhension de nos données. Il a été essentiel de sélectionner les variables les plus pertinentes pour notre analyse, en nous basant sur leur importance dans la réalisation de nos objectifs.

Ensuite, nous avons réalisé une analyse descriptive de notre base de données traitée pour obtenir les statistiques nécessaires à notre analyse. Nous avons utilisé différentes commandes, telles que "proc freq", "proc means" et "proc summary", pour afficher des tableaux de fréquence et des tableaux descriptifs. De plus, nous avons enrichi notre analyse en illustrant nos résultats à l'aide de graphiques et de cartes produits sur Excel.

Enfin, nous avons élaboré un tableau de bord sur Excel, dans lequel nous développerons des recommandations sur la manière d'aider l'entreprise à atteindre ses objectifs.

## **B. Projet d'Analyse de Données Assurance sous Python**

### Introduction

En tant que groupe de data scientists/Data Analyst juniors travaillant pour une insurtech, notre mission est d'exploiter de nouvelles bases de données d'assurance en utilisant exclusivement

Python. Ce rapport présente notre analyse des données, comprenant des traitements, des analyses descriptives, des modélisations, et des interprétations pertinentes.

## 1. Traitement des Données

Nous avons débuté par comprendre la structure des données, en vérifiant la complétude et la conformité des informations. Les étapes de traitement incluent la gestion des valeurs manquantes, l'identification et la correction des anomalies, ainsi que l'exclusion des données aberrantes. Toutes les manipulations ont été documentées pour assurer la transparence du processus.

## 2. Jointures et Contrôles

Des jointures ont été réalisées entre les différentes bases de données pour consolider les informations. Des contrôles de cohérence ont été effectués pour garantir l'intégrité des données.

## 3. Analyses Descriptives

Des analyses univariées et multivariées ont été menées pour comprendre la distribution des variables et les relations entre elles. Des techniques d'analyse statistique des données (ASD) ont été utilisées pour explorer en profondeur les caractéristiques des données.

## 4. Data Visualization

Nous avons créé des visualisations graphiques avancées pour interpréter et comprendre les données. L'utilisation de bibliothèques telles que Matplotlib, Seaborn et Plotly nous a permis de représenter graphiquement les tendances, les distributions et les relations entre les variables.

## 5. Modélisation

Nous avons utilisé des techniques de modélisation supervisée et non supervisée pour prédire les tendances et les comportements. Des modèles de régression et de classification ont été mis en œuvre pour analyser les relations entre les variables et pour réaliser des prévisions.

## 6. Analyse des Résultats

Les résultats obtenus ont été interprétés et expliqués en détail, en mettant en lumière les relations significatives entre les variables et en identifiant les facteurs clés influençant les résultats.

## Conclusion

Ce rapport détaille notre approche méthodique pour exploiter les nouvelles bases de données d'assurance, en utilisant Python comme outil principal. Les analyses et modélisations réalisées ont permis de fournir des insights précieux pour guider les décisions stratégiques de l'entreprise.

**Deux liens pour mes applications sur Rshiny**

1. Cette application Shiny constitue une plateforme conviviale dédiée à l'exploration, à l'analyse et à la modélisation des séries temporelles. Elle met à disposition une gamme d'outils pour une compréhension approfondie des données, facilitant ainsi la prise de décisions basées sur des prévisions précises. Les utilisateurs ont le choix entre plusieurs modèles, tels que la régression linéaire, ARIMA, Holt-Winters et AUTO ARIMA, pour modéliser leurs séries temporelles. Chaque modèle est configurable grâce à des paramètres et des options spécifiques, permettant un ajustement précis. Une fonctionnalité intégrée permet aux utilisateurs de visualiser graphiquement les prévisions générées par le modèle sélectionné. De plus, l'application propose une option de transformation log pour les données, ouvrant la voie à la modélisation de séries transformées. Cet outil s'avère particulièrement bénéfique pour les analystes de données et toute personne travaillant avec des données temporelles, offrant une meilleure compréhension des séries chronologiques et la capacité à anticiper des scénarios futurs grâce à des prévisions fiables. Ci-joint le lien

<https://mamadoudiop.shinyapps.io/SerieTemporelle/>

2. Cette application utilise des techniques de Machine Learning pour évaluer avec précision la probabilité qu'un individu puisse faire défaut sur un prêt. En se basant sur des variables telles que les revenus, le nombre de virements, le sexe, l'âge, l'épargne retraite, le nombre de conseils, le niveau d'étude, le livret A et le type de produits, elle crée des modèles prédictifs et offre une vision détaillée des données à travers des statistiques et des graphiques interactifs. Elle propose plusieurs algorithmes de Machine Learning et permet aux utilisateurs d'entraîner, d'évaluer ces modèles et de faire des prédictions sur de nouveaux emprunteurs. En automatisant ce processus, elle aide les institutions financières à prendre des décisions éclairées, économisant temps et ressources tout en améliorant la précision des évaluations de crédit. Cette application représente un outil précieux pour prédire le risque de défaut, contribuant ainsi à une gestion plus rigoureuse des prêts et à une réduction des risques financiers dans le secteur financier. Ci-joint le lien

<https://mamadoudiop.shinyapps.io/MonAPP/>