ActuarIA SOFTWARE

Life insurance functionality : Life annuity temporary and deferred

BARKATI Soufian
JBIHA Ghita
KHALOUI Ghita

10 décembre 2023

Table des matières

1 Introduction

L'assurance vie joue un rôle crucial dans la planification financière en offrant une sécurité financière aux bénéficiaires en cas de décès de l'assuré. La tarification des polices d'assurance vie est une tâche complexe qui nécessite des modèles actuariels sophistiqués. Ce rapport se concentre sur le développement d'un outil de tarification pour un type spécifique d'assurance vie, à savoir, une **Life Annuity with Flexible Payment Characteristics**. Cette approche innovante permet des paiements d'annuité différée à la fin de chaque année, offrant une flexibilité inégalée pour répondre aux besoins financiers variés des assurés.

2 Extraction des Données

Les données nécessaires à la construction des modèles actuariels sont extraites de la table de mortalité **TGF05**. Cette table fournit des informations cruciales sur la probabilité de survie à différents âges. Un ensemble de paramètres, comprenant l'âge de l'assuré, la durée de l'annuité, le taux technique, et d'autres, est utilisé pour générer un ensemble de données représentatif. Ces données sont essentielles pour entraîner les modèles de machine learning et garantir des prédictions précises des primes d'assurance vie.

3 Modèles Actuariels et Types d'Assurance

3.1 Fonctions Actuarielles Fondamentales

Dans le contexte de l'assurance vie, plusieurs fonctions actuarielles sont implémentées pour faciliter la tarification des polices et le calcul des primes. Ces fonctions sont construites en utilisant la table de mortalité TGF05 et intègrent des concepts actuariels clés tels que la probabilité de survie, les facteurs d'actualisation, et les annuités.

— La fonction **lx(x)** extrait le nombre de personnes survivantes à l'âge **x** à partir de la table de mortalité TGF05. Elle est cruciale pour estimer la probabilité de survie à un âge donné.

— La fonction **techDF(h, techrate)** calcule le facteur de décroissance technique, qui ajuste la valeur d'une somme d'argent dans le temps en tenant compte du taux technique techrate.

$$\mathrm{DF}(h, techrate) = \frac{1}{(1 + techrate)^h}$$

— La fonction **Dx(x, techrate)** évalue la probabilité qu'une personne âgée de x ans survive exactement h années, contribuant ainsi aux calculs actuariels des primes et des paiements.

$$\mathrm{Dx}(x, \mathrm{techrate}) = \frac{1}{(1 + \mathrm{techrate})^x} \times \mathrm{lx}(x) \quad \text{where} \quad x \text{ represents age}$$

— La fonction **Nx(x, techrate)** calcule la somme totale de la probabilité qu'une personne âgée de x ans survive h années pour toutes les valeurs possibles de h. Cela est crucial pour l'évaluation des annuités.

$$\mathrm{Nx}(x, techrate) = \sum_{i=0}^{121-x} \left(\frac{1}{(1 + \mathrm{techrate})^{x+i}} \times \mathrm{lx}(x + i) \right)$$

— La fonction **npx(x, n)** détermine la probabilité qu'une personne âgée de x ans survive exactement n années supplémentaires. Cette fonction est essentielle pour les primes temporaires et les annuités différées.

$$\text{npx}(x, n) = \frac{\text{lx}(x+n)}{\text{lx}(x)} \quad \text{where} \quad n \text{ represents terms}$$

— La fonction **mprimeaxn(x, n, techrate, mprime)** évalue la valeur actuarielle d'une somme d'argent payée à la fin de chaque année pour 'n' années, à partir de mprime années dans le futur. Cela contribue aux calculs des primes annuelles et des annuités différées.

$$\text{mprime\_axn}(x, n, \text{techrate}, \text{mprime}) = \sum_{k=\text{mprime}+1}^{\text{mprime}+n} \text{techDF}(k, \text{techrate}) \times \text{npx}(x, k)$$

— La fonction **nEx(x, n, techrate)** calcule la valeur actuarielle d'une somme d'argent payée à la fin de chaque année pour n années, à partir de l'année x. Elle est cruciale pour évaluer les annuités différées.

$$\text{nEx}(x, n, \text{techrate}) = \text{npx}(x, n) \times \text{techDF}(n, \text{techrate})$$

— La fonction **annuityfactor(x, m, techrate)** calcule le facteur d'annuité pour une annuité temporaire de m années, à partir de l'année x.

$$\text{AnnuityFactor}(x, m, techrate) = \sum_{h=0}^{m-1} \text{nEx}(x, h, \text{techrate}) \quad \text{where} \quad m \text{ payments}$$

— La fonction **immediatelifeannuitydue(x, techrate)** évalue le montant d'une annuité immédiate payable à la fin de chaque année pour une personne âgée de x ans.

$$\text{ImmediateLifeAnnuityDue}(x, techrate) = \frac{\text{Nx}(x, \text{techrate})}{\text{Dx}(x, \text{techrate})}$$

— La fonction **immediatetemporarylifeannuitydeferred(x, n, techrate, deferred)** détermine le montant d'une annuité temporaire différée payable à la fin de chaque année pour une personne âgée de x ans.

$$\text{ImmediateTemporaryLifeAnnuity}(x, n, techrate, deferred) =$$
$$\frac{\text{Nx}(x + \text{deferred} + 1, \text{techrate}) - \text{Nx}(x + \text{deferred} + n + 1, \text{techrate})}{\text{Dx}(x, \text{techrate})}$$

3.2 Life Annuity Differently at the End of the Year

Le type d'assurance choisi est une **Life Annuity Differently at the End of the Year**. Cette assurance offre une flexibilité unique en permettant des paiements d'annuité différée à la fin de chaque année, avec la possibilité de moduler le nombre d'années de report et la durée de l'annuité. Ce type d'assurance peut être particulièrement adapté aux individus planifiant des flux de trésorerie futurs spécifiques. En utilisant les fonctions actuarielles définies, notre modèle peut estimer avec précision les primes et les paiements associés à cette assurance, prenant en compte des paramètres tels que l'âge de l'assuré, le taux technique, la durée de report, et la durée de l'annuité. Cette approche combine la robustesse des calculs actuariels traditionnels avec la flexibilité offerte par les modèles de machine learning pour des estimations précises des primes. Voici les formules utilisées afin de calculer la prime annuelle.

$$\text{annualPremium}(x, n, m, \text{techrate}, \text{amount}, \text{deferred}) = \frac{\text{singlePremium}(x, n, \text{techrate}, \text{deferred}, \text{amount})}{\text{annuityFactor}(x, m, \text{techrate})}$$

avec :

$$\text{singlePremium}(x, n, \text{techrate}, \text{Amount}, \text{deferred}) =$$
$$\text{ImmediateTemporaryLifeAnnuity}(x, n, \text{techrate}, \text{deferred}) \times \text{Amount}$$

4 Génération de Données et Analyse des données

4.1 Génération du Dataset

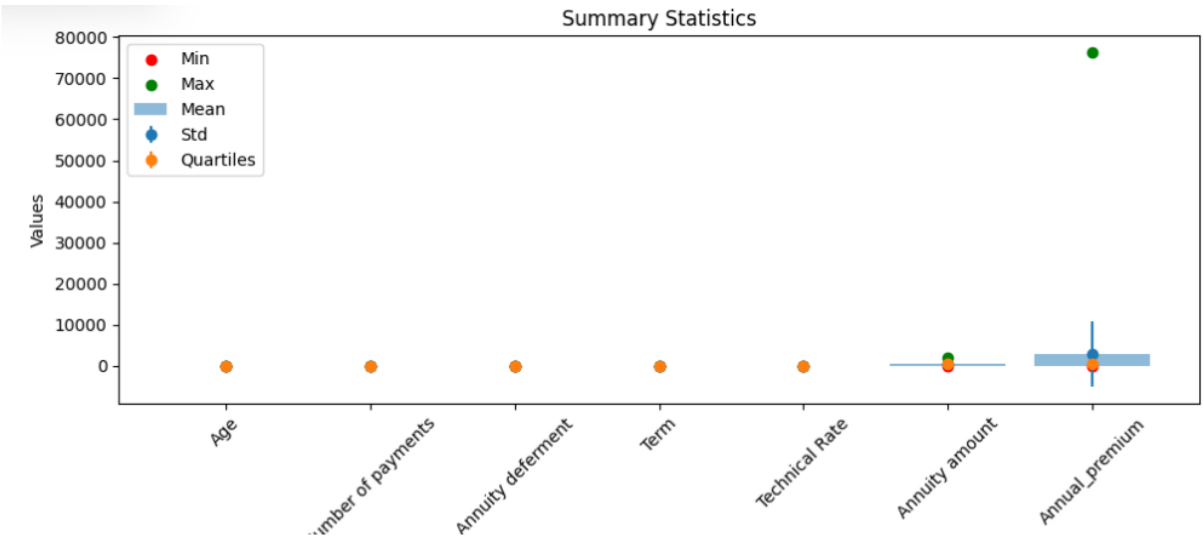
La génération du dataset revêt une importance cruciale pour le développement de nos modèles de machine learning précis et robustes. Les paramètres choisis pour cette génération sont des éléments clés dans la tarification des contrats d'assurance vie.

Notre dataset, composé de différentes caractéristiques actuarielles, offre une variété cruciale pour la modélisation des primes d'assurance vie. En segmentant l'âge des souscripteurs en tranches significatives (20, 30, 40, 50, 60), nous couvrons une gamme étendue de profils assurantiels, permettant aux modèles d'apprendre à évaluer les primes pour des catégories d'âge distinctes. La diversité des échantillons pour le nombre de paiements annuels (de 1 à 40) capture les variations potentielles dans les modalités de paiement, tandis que l'inclusion de périodes de différé allant jusqu'à 30 ans reflète l'importance de la possibilité de retarder les annuités. La flexibilité des termes d'annuité (de 1 à 60 ans) dans le dataset représente la diversité des engagements contractuels. De plus, l'intégration des taux techniques (variant de 0 à 2.5%) anticipe les différents rendements sur les investissements sous-jacents. Enfin, la variabilité des montants d'annuité (de 50 à 2000 unités monétaires) capture la diversité des couvertures, permettant aux modèles d'appréhender des scénarios allant de souscriptions modestes à des niveaux de couverture plus substantiels. Ces éléments confèrent à nos modèles la capacité d'appréhender la complexité inhérente à la tarification des primes d'assurance vie.

4.2 Analyse Statistique du Dataset

L'analyse statistique approfondie fournit des insights précieux sur la distribution des paramètres dans notre dataset.

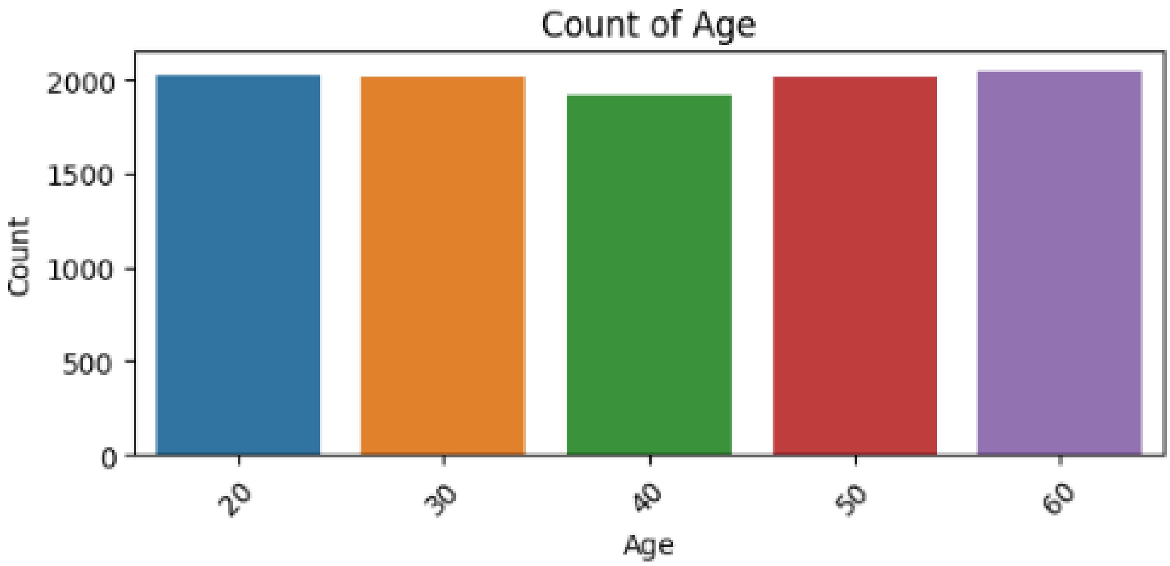
Notre analyse des caractéristiques actuarielles révèle des tendances significatives. L'âge moyen des souscripteurs, autour de 39 ans, dépeint une distribution équilibrée des tranches d'âge, avec un écart-type de 14.53 soulignant la diversité des profils. En ce qui concerne le nombre de paiements annuels, la moyenne de 17 avec une gamme allant de 1 à 40 capture efficacement la variabilité des fréquences de paiement. Le différé moyen d'annuité de 3.78 ans dans notre dataset modélise des situations où les annuités peuvent être différées pendant des périodes variables, ajoutant une dimension temporelle cruciale. La durée moyenne d'annuité de 27.09 ans reflète une diversité d'engagements à long terme, couvrant une gamme étendue de scénarios. La diversification des taux techniques entre 0 et 2.5% traduit des choix d'investissement variés, tandis que les montants d'annuité allant de 50 à 2000 représentent une diversité de niveaux de couverture. Enfin, la grande variabilité des primes annuelles, de moins d'une unité monétaire à plus de 76 000, souligne les différences significatives dans les coûts associés aux contrats d'assurance vie, mettant en lumière la complexité de la tarification.
Voici une présentation graphique qui résume ces résultats.



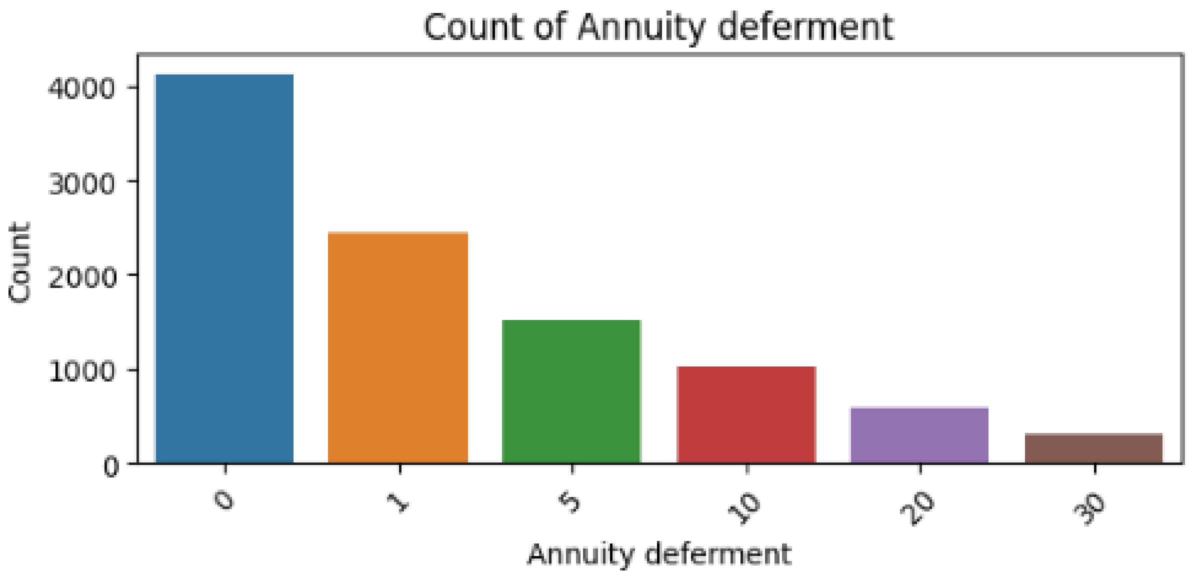
4.3 Analyse Visuelle des Attributs

L'analyse visuelle des attributs numériques permet de mieux comprendre la distribution des paramètres clés dans notre dataset. Nous nous concentrons sur les attributs tels que l'âge, le nombre de paiements annuels, le différé d'annuité, la durée de l'annuité, le taux technique et le montant d'annuité.

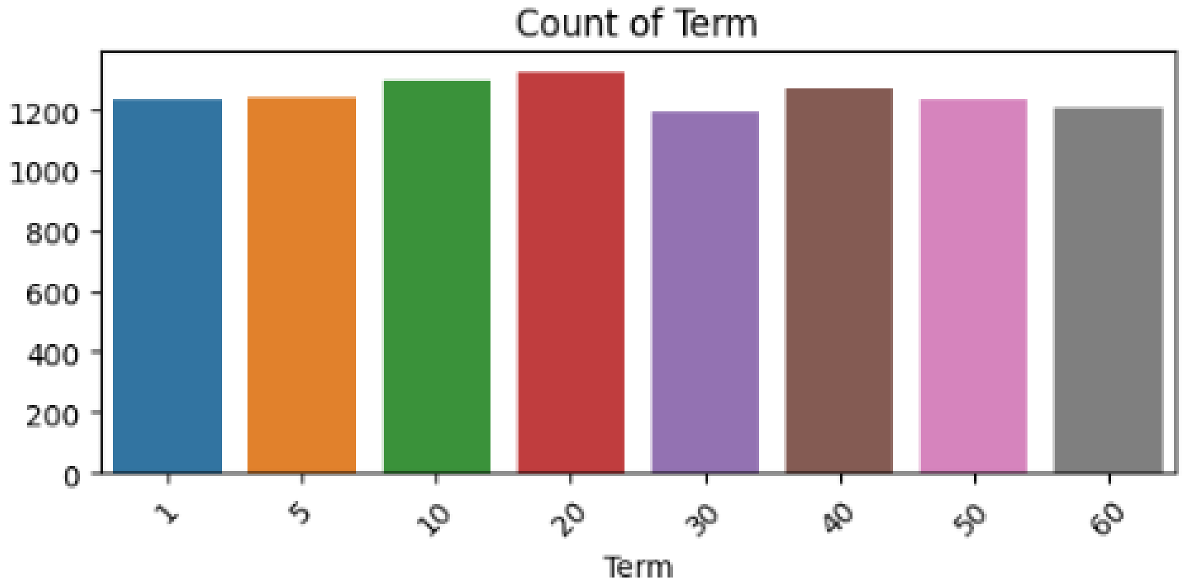
L'analyse des caractéristiques actuarielles met en lumière plusieurs aspects cruciaux. En ce qui concerne l'âge, notre dataset offre une répartition équilibrée, englobant une gamme étendue de 20 à 60 ans, garantissant ainsi une représentation adéquate des différentes tranches d'âge.



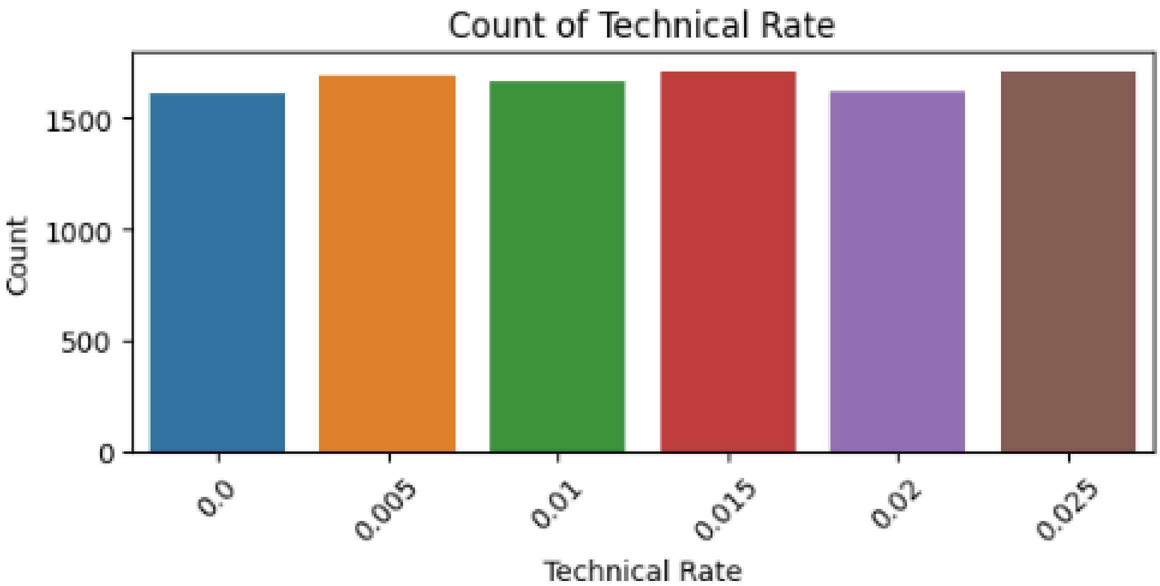
La distribution du nombre de paiements annuels démontre aussi une répartition équilibrée. Pour le différé d'annuité, la majorité des contrats présentent un délai relativement faible, bien que quelques contrats offrent des différés plus longs, atteignant jusqu'à 30 ans.



La durée de l'annuité reflète la flexibilité des contrats d'assurance vie, couvrant une plage allant de 1 à 60 ans.



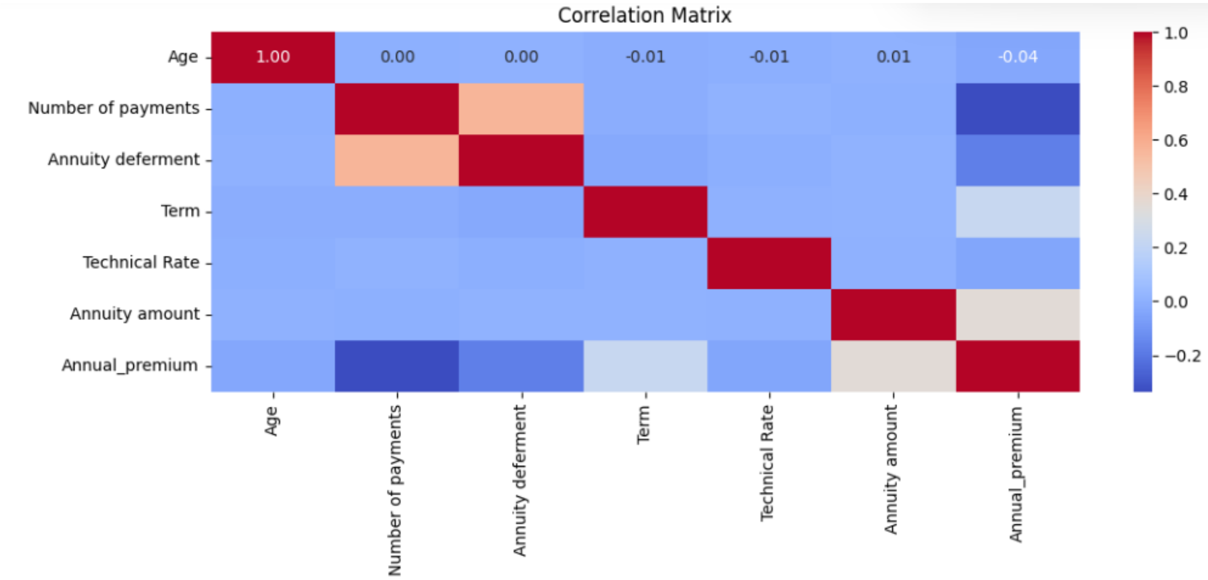
Les taux techniques, variant entre 0% et 2.5%, avec une concentration notable autour de 1% et 2%, représentent les choix diversifiés de rendement attendu sur les investissements sous-jacents aux contrats.



Enfin, la distribution des montants d'annuité illustre une variabilité considérable, allant de montants modestes à des niveaux plus élevés, traduisant la diversité des couvertures d'assurance vie présentes dans notre dataset.

4.4 Matrice de Corrélation

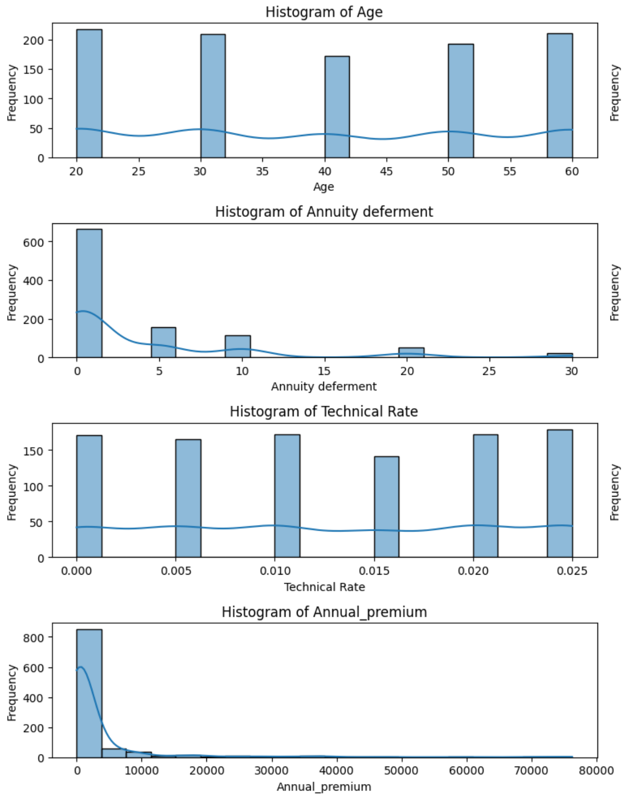
La matrice de corrélation ci-dessous offre un aperçu des relations linéaires entre les différentes variables de notre dataset. Les coefficients de corrélation sont représentés par des couleurs, allant du bleu foncé (corrélation négative) au rouge foncé (corrélation positive). Une valeur proche de 1 indique une corrélation positive parfaite, tandis qu'une valeur proche de -1 indique une corrélation négative parfaite.



L'analyse des corrélations entre les caractéristiques actuarielles révèle des relations intéressantes. Premièrement, une légère corrélation positive entre l'âge du souscripteur et le montant d'annuité suggère que les montants ont tendance à augmenter avec l'âge. Deuxièmement, une corrélation positive modérée entre la durée de l'annuité et le nombre de paiements annuels suggère que les contrats avec une durée d'annuité plus longue ont tendance à comporter un plus grand nombre de paiements annuels. Troisièmement, une corrélation négative légère entre l'âge et la durée de l'annuité indique que les contrats souscrits à un âge plus avancé ont tendance à avoir des durées d'annuité plus courtes. Enfin, aucune corrélation significative entre le différé d'annuité et le montant d'annuité suggère que le différé d'annuité n'influence pas fortement le montant de l'annuité, soulignant la complexité des relations au sein des caractéristiques actuarielles.

4.5 Analyse des Caractéristiques Numériques

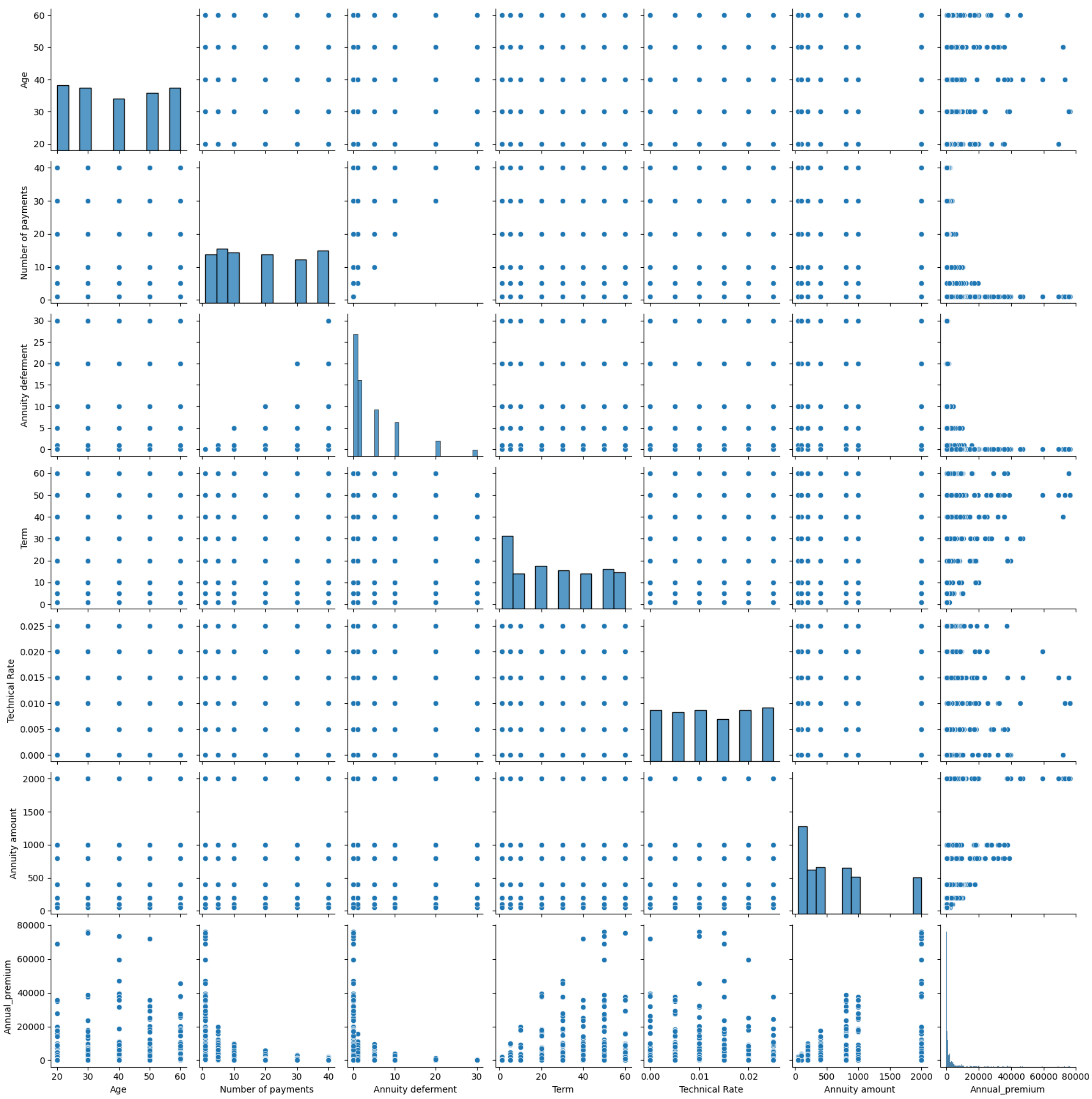
Les histogrammes ci-dessous présentent la distribution des caractéristiques numériques du dataset, y compris l'âge, le nombre de paiements, le différé d'annuité, la durée d'annuité, le taux technique, le montant d'annuité et la prime annuelle.



L'analyse détaillée des caractéristiques actuarielles révèle des tendances importantes au sein du dataset. Tout d'abord, en ce qui concerne l'âge des souscripteurs, la majorité des contrats d'assurance vie ont été souscrits par des individus âgés entre 30 et 50 ans, montrant une distribution relativement uniforme au sein de cette tranche d'âge. Deuxièmement, la plupart des contrats présentent un nombre de paiements annuels compris entre 0 et 30, avec une concentration significative autour de 5 paiements. Troisièmement, la distribution du différé d'annuité indique une préférence pour des périodes plus courtes, principalement entre 0 et 5 ans. La durée de l'annuité varie considérablement, avec une concentration autour de 0 à 20 ans. En ce qui concerne les taux techniques, la majorité des contrats affichent des taux compris entre 0 et 2.5%, reflétant des choix d'investissement diversifiés. La distribution du montant d'annuité montre une concentration autour de montants plus bas, bien que quelques contrats présentent des montants plus élevés. Enfin, la prime annuelle affiche une distribution étendue, avec la plupart des contrats ayant des primes relativement faibles, mais certains contrats se distinguent par des primes considérablement plus élevées.

4.6 Matrice de Paires

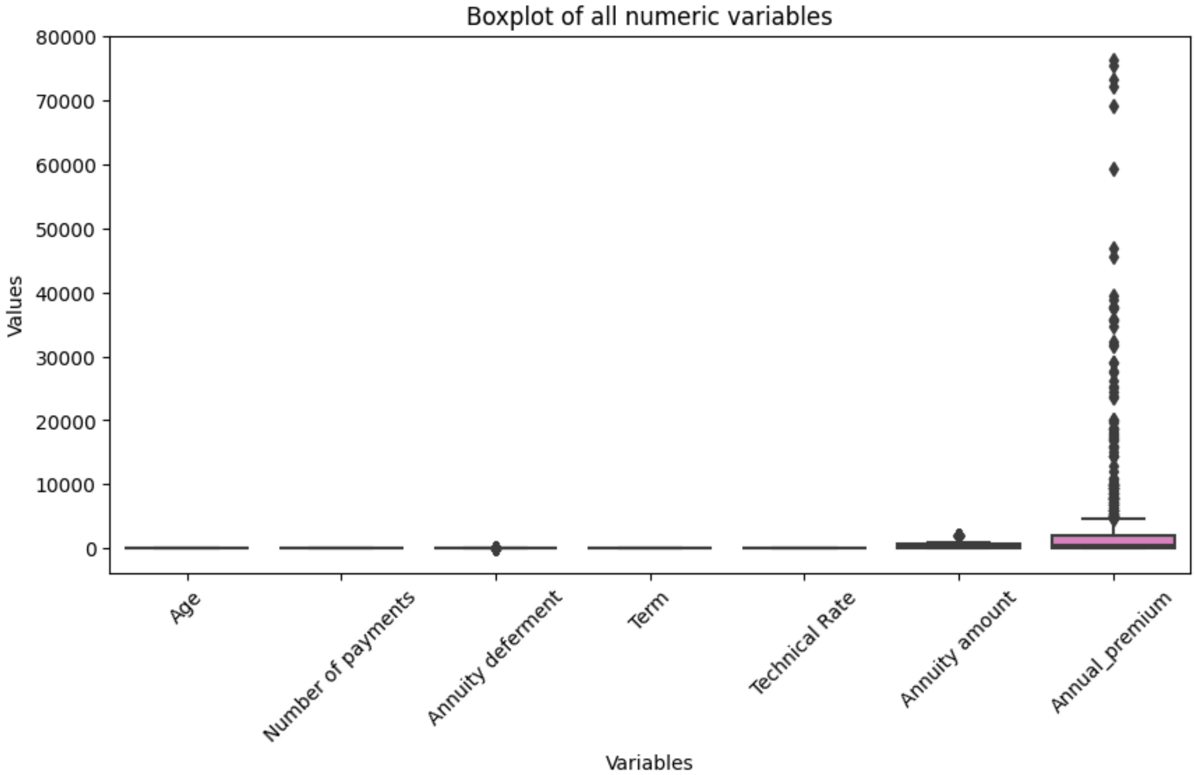
La matrice de paires ci-dessus représente les relations entre les caractéristiques numériques du dataset. Chaque nuage de points croise deux variables, permettant d'observer visuellement les corrélations potentielles.



L'analyse des relations entre différentes caractéristiques actuarielles révèle des nuances importantes dans le dataset. Tout d'abord, en examinant l'âge par rapport au montant d'annuité, on observe une dispersion des montants pour toutes les tranches d'âge, sans qu'une tendance claire n'émerge. Ensuite, en comparant la durée de l'annuité à la prime annuelle, on constate une diversité de primes pour différentes durées d'annuité, suggérant que la durée seule ne détermine pas de manière prédominante le montant de la prime. De plus, l'analyse du différé d'annuité par rapport à la prime annuelle met en lumière une variabilité significative des primes pour des délais d'annuité similaires, soulignant l'impact d'autres facteurs sur le coût de l'assurance vie. Enfin, en examinant le taux technique par rapport au montant d'annuité, aucune tendance claire n'émerge, indiquant que le montant de l'annuité n'est pas strictement lié au taux technique et que d'autres variables interviennent dans cette relation complexe.

4.7 Analyse des Boîtes à Moustaches (Boxplot)

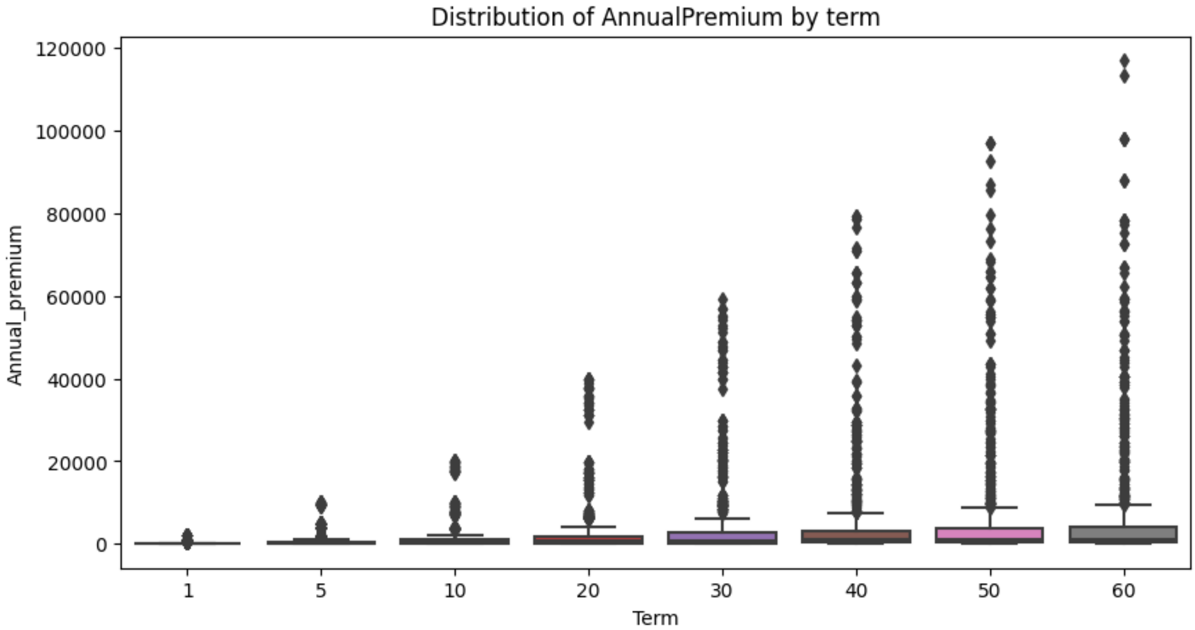
Le diagramme en boîte à moustaches ci-dessus présente la distribution des variables numériques dans le dataset. Chaque boîte représente le premier quartile (Q1) à la médiane (ligne médiane à l'intérieur de la boîte) jusqu'au troisième quartile (Q3). Les moustaches s'étendent jusqu'aux valeurs extrêmes à l'extérieur de la plage interquartile (IQR).



L'examen des différentes caractéristiques actuarielles révèle une distribution relativement uniforme de l'âge, bien que quelques valeurs aberrantes soient présentes vers la limite supérieure. En ce qui concerne le nombre de paiements, la plupart des contrats affichent des chiffres compris entre 5 et 30, avec quelques valeurs aberrantes. Les délais d'annuité varient considérablement, avec certains contrats ne présentant aucun différé. La majorité des contrats ont une durée d'annuité inférieure à 30 ans, bien que des valeurs aberrantes soient observées vers la limite supérieure. La répartition des taux techniques semble relativement uniforme, sans valeurs aberrantes notables. En ce qui concerne le montant de l'annuité, la distribution présente une asymétrie importante, avec une concentration vers des montants plus faibles et quelques valeurs aberrantes élevées. La prime annuelle, quant à elle, affiche une dispersion significative, avec plusieurs valeurs aberrantes atteignant des montants très élevés.

4.8 Distribution de la Prime Annuelle par Durée d'Annuité

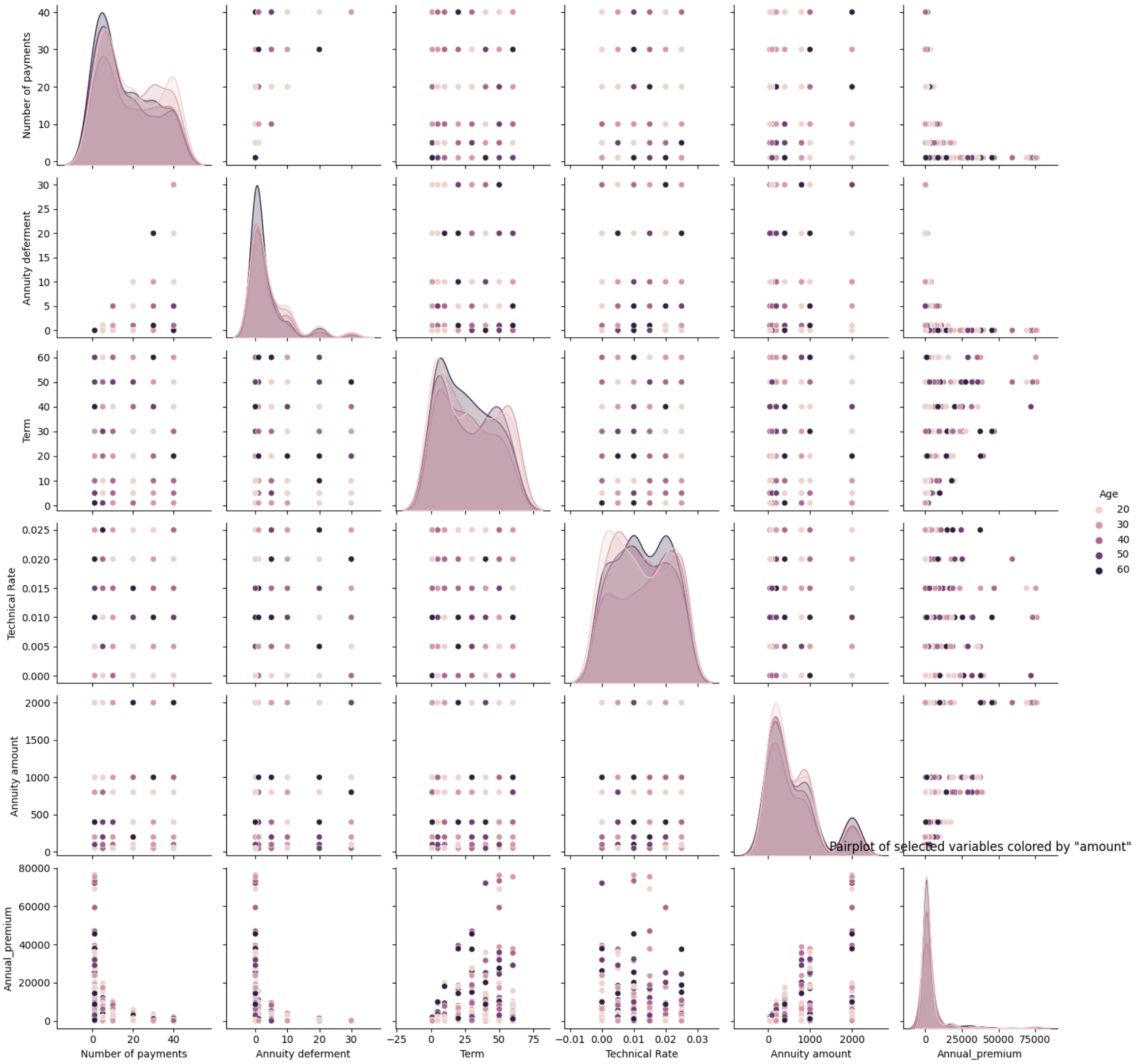
Le diagramme en boîte à moustaches ci-dessus examine la distribution de la prime annuelle en fonction de la durée d'annuité. Chaque boîte représente la variation de la prime annuelle pour une durée d'annuité spécifique.



L'analyse révèle une variation significative de la prime annuelle en fonction de la durée d'annuité. Pour les durées d'annuité plus courtes, la prime annuelle montre une moindre variabilité, bien que quelques valeurs aberrantes élevées soient présentes. En revanche, les durées d'annuité plus longues présentent une dispersion plus importante des primes annuelles, avec plusieurs valeurs aberrantes atteignant des montants significativement élevés.

4.9 Relation entre les Variables Sélectionnées, Colorées par l'Âge

Le pairplot ci-dessus examine les relations entre certaines variables sélectionnées, avec chaque point coloré en fonction de l'âge. Les variables incluses sont le nombre de paiements, le différé d'annuité, la durée d'annuité, le taux technique, le montant de l'annuité, et la prime annuelle.



L'analyse des relations entre différentes variables révèle plusieurs observations intéressantes. Tout d'abord, il semble exister une relation positive entre le nombre de paiements et le montant d'annuité, indiquant une augmentation générale du montant pour un plus grand nombre de paiements. En ce qui concerne l'âge et le différé d'annuité, une tendance émerge suggérant que le différé d'annuité a tendance à diminuer avec l'âge. Cette observation pourrait indiquer une propension à souscrire à des annuités avec un différé plus court à mesure que l'âge des souscripteurs augmente. De plus, la durée d'annuité semble influencer de manière significative le montant de l'annuité, avec une tendance à des montants plus élevés pour des durées plus longues.

5 Modélisation de Prime Annuelle

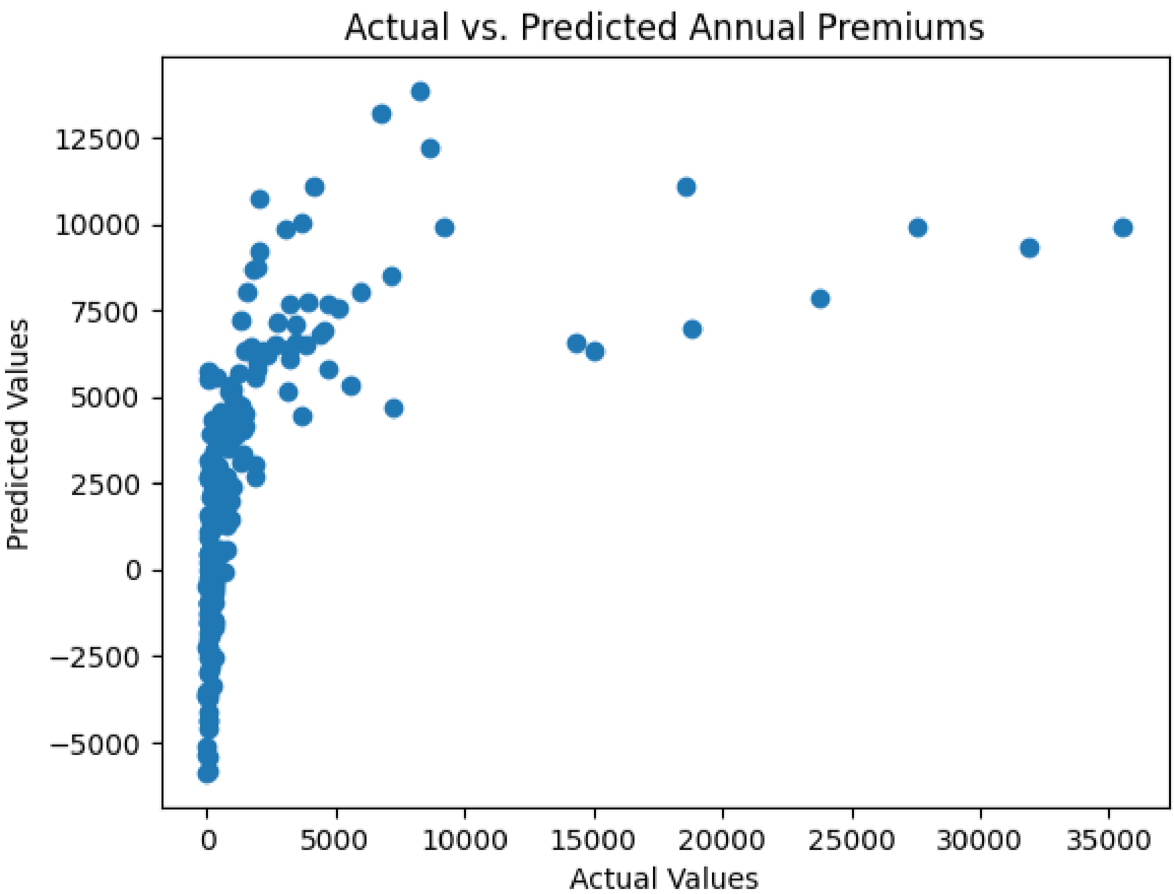
5.1 Régression Linéaire

Nous avons entrepris une modélisation de la prime annuelle en utilisant la régression linéaire, en prenant en compte des variables démographiques et actuarielles telles que l'âge, le nombre de paiements, le différé d'annuité, la durée, le taux technique, et le montant d'annuité. Les coefficients obtenus du modèle dévoilent des relations entre ces variables et la prime annuelle. Par exemple, l'âge semble influencer négativement la prime annuelle, tandis que la durée a un impact positif. Les coefficients du modèle sont les suivants :

- Âge (Age) : -6.39
- Nombre de Paiements (Number of Payments) : -194.18
- Différé d'Annuité (Annuity Deferment) : 25.07
- Durée (Term) : 79.39
- Taux Technique (Technical Rate) : -53,629.42
- Montant d'Annuité (Annuity Amount) : 4.65

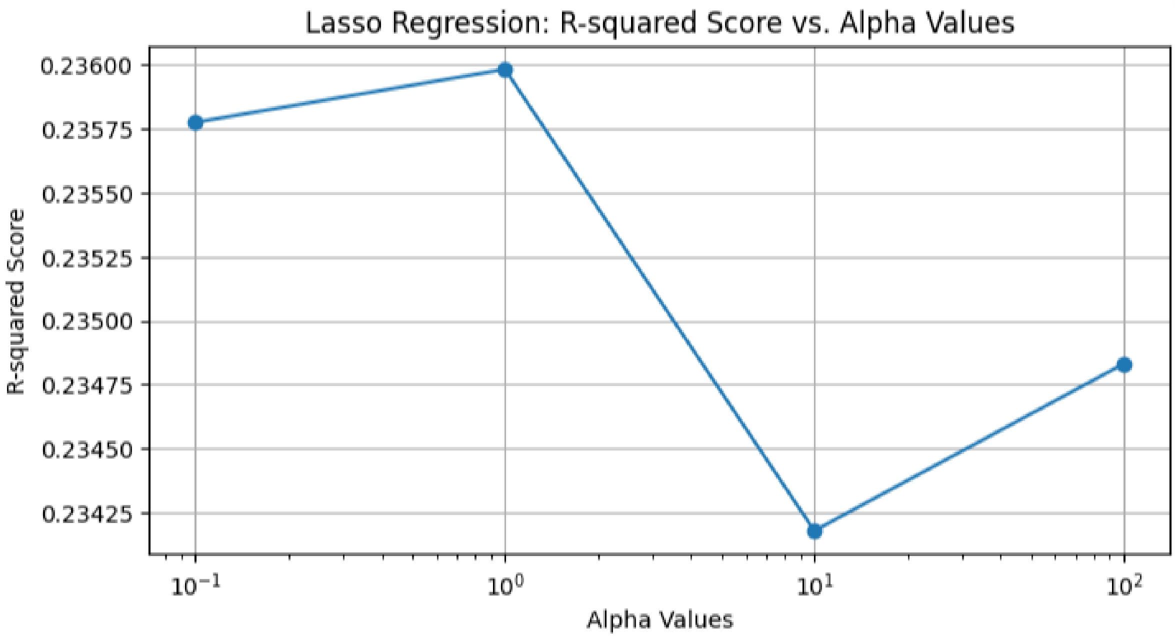
L'intercept du modèle est de 2146.43. La performance du modèle, évaluée par des mesures telles que l'erreur quadratique moyenne (MSE) de 19,585,686.20, l'erreur absolue moyenne (MAE) de 3,053.21, et le score R^2 de 0.17, indique une modération dans la capacité du modèle à expliquer la variance dans les primes annuelles. La prédiction sur de nouvelles données présente également des différences significatives par rapport à la valeur actuarielle, soulignant les limites du modèle linéaire pour saisir les subtilités complexes des données actuarielles. La prime annuelle prédite pour de nouvelles données est de 4,260.46, tandis que la valeur actuarielle de la prime annuelle pour ces données est de 1,220.91.

La visualisation Actual vs Prédit illustre la concordance entre les valeurs réelles de la prime annuelle et les prédictions du modèle, mettant en évidence la variabilité importante non capturée par le modèle.



5.2 Lasso Regression

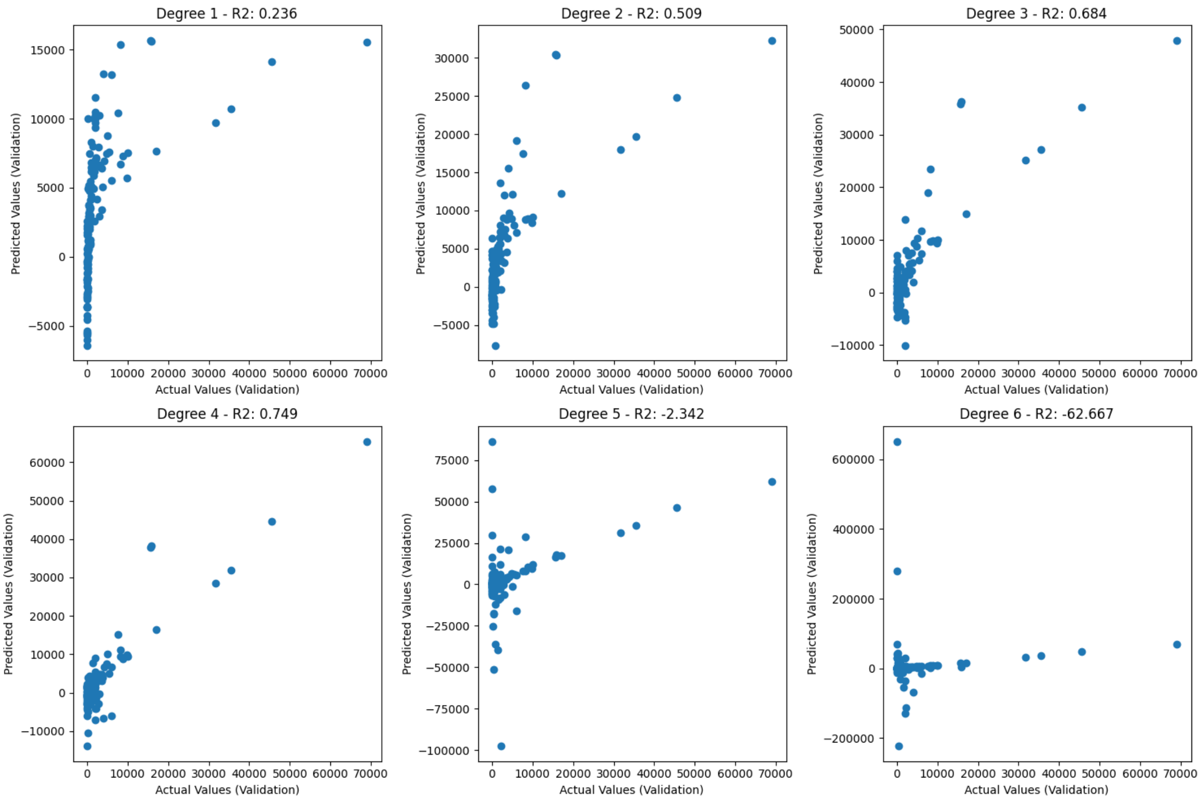
Nous avons utilisé la régression Lasso (L1 Regularization) pour modéliser la prime annuelle en fonction des variables actuarielles et démographiques. L'objectif était de trouver le meilleur coefficient de pénalité (alpha) pour optimiser la performance du modèle. Nous avons testé plusieurs valeurs d'alpha (0.1, 1, 10, 100) pour trouver celui qui maximise le score R^2 sur l'ensemble de validation. La figure ci-dessous montre comment le R^2 varie en fonction des valeurs d'alpha. Le meilleur alpha trouvé est 1. En évaluant la performance du modèle avec cet alpha optimal sur



l'ensemble de test, le score R^2 est de -0.12, indiquant que le modèle ne parvient pas à expliquer la variance dans les données de prime annuelle. En prévoyant la prime annuelle pour de nouvelles données avec le modèle Lasso, la prédiction est de 4,640.23, tandis que la valeur actuarielle de la prime annuelle pour ces données est de 1,220.91. Ces résultats soulignent les limites du modèle Lasso dans ce contexte, indiquant la nécessité de considérer des approches plus adaptées ou l'inclusion de variables supplémentaires pour améliorer la précision des prédictions actuarielles.

5.3 Régression Polynomiale

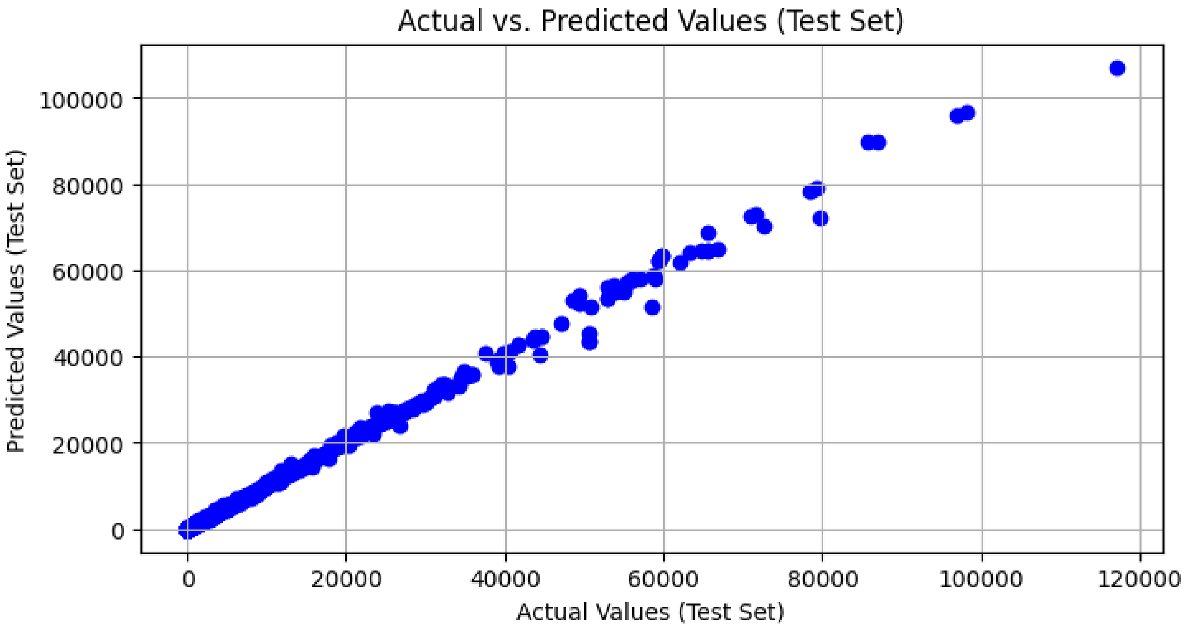
Nous avons exploré la régression polynomiale pour modéliser la prime annuelle en fonction des caractéristiques actuarielles et démographiques. Différents degrés de polynômes ont été testés pour évaluer l'impact sur la performance du modèle. Nous avons testé des degrés de polynômes de 1 à 6 et évalué leur performance sur l'ensemble de validation. Les graphiques ci-dessous montrent la comparaison entre les valeurs réelles et prédites pour différentes valeurs de degrés. Le meilleur



degré de polynôme trouvé est 4. En évaluant la performance du modèle avec ce degré optimal sur l'ensemble de test, le score R^2 est de 0.475, indiquant une meilleure adaptation aux données par rapport aux autres degrés testés. En prévoyant la prime annuelle pour de nouvelles données avec le modèle de régression polynomiale, la prédiction est de -1006.51, tandis que la valeur actuarielle de la prime annuelle pour ces données est de 1220.91. Bien que le modèle ait montré une amélioration par rapport aux degrés inférieurs, la disparité significative entre la prédiction et la valeur actuarielle souligne la nécessité de considérer des ajustements supplémentaires pour améliorer la précision des prédictions actuarielles.

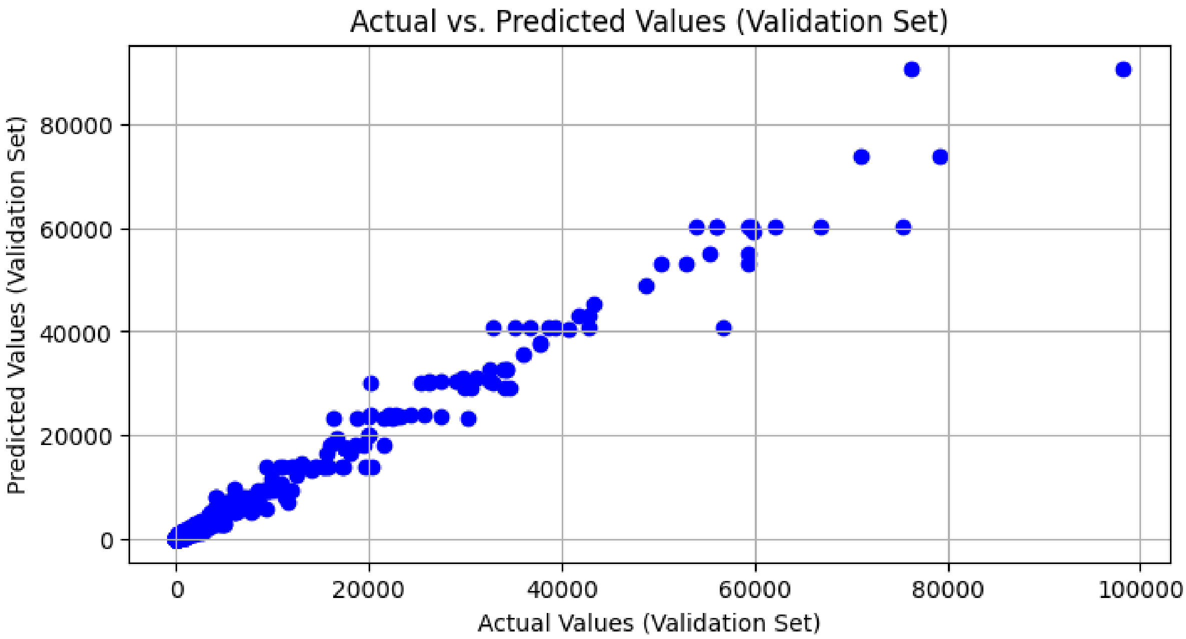
5.4 Random Forest

Nous avons utilisé une approche basée sur Random Forest pour modéliser la prime annuelle en fonction des caractéristiques actuarielles et démographiques. Random Forest a présenté des performances exceptionnelles, avec un score d'entraînement de 0.9997 et un score de test de 0.998. Ces scores élevés indiquent une adaptation précise aux données, suggérant que le modèle a bien généralisé au-delà de l'ensemble d'entraînement. En prévoyant la prime annuelle pour de nouvelles données avec le modèle de Forêt Aléatoire, la prédiction est de 1742.99, tandis que la valeur actuarielle de la prime annuelle pour ces données est de 1220.91. La prédiction semble raisonnablement proche de la valeur actuarielle, démontrant la capacité de la Forêt Aléatoire à fournir des estimations précises pour les primes annuelles. Cette approche offre une solution robuste pour modéliser des relations complexes entre les variables actuarielles et la prime annuelle.



5.5 Decision Tree

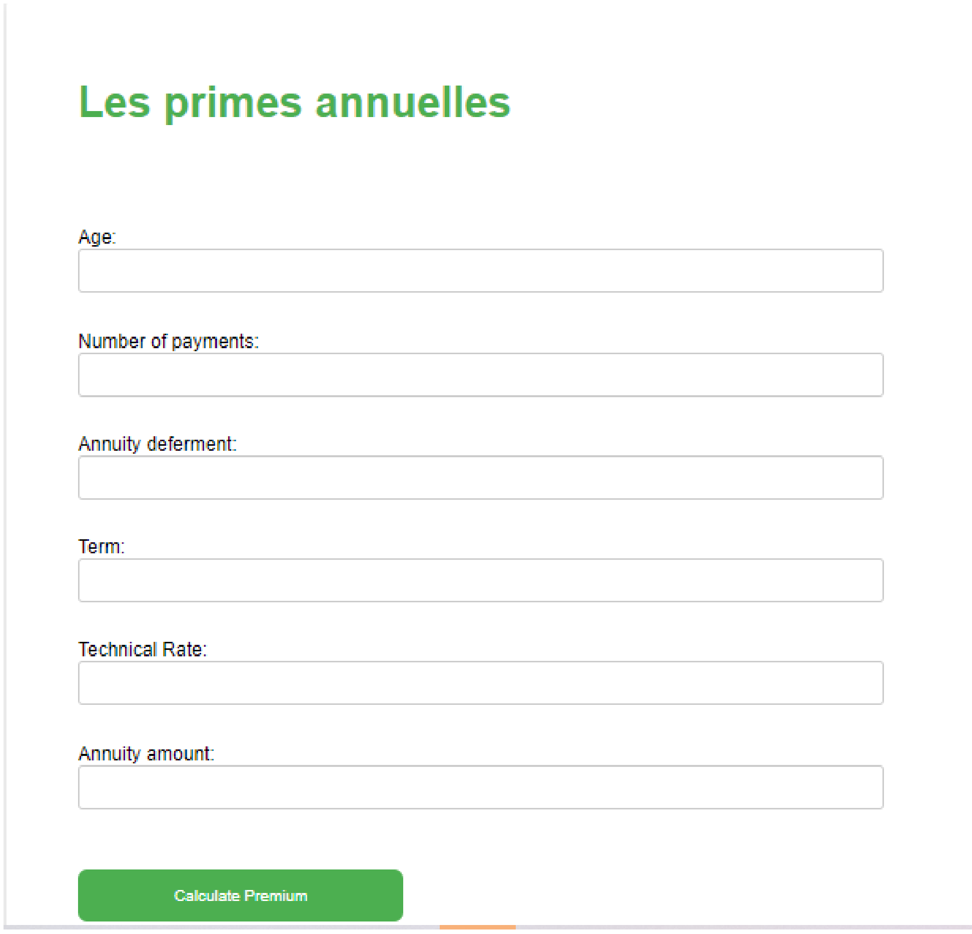
Nous avons adopté une approche basée sur une Decision Tree pour modéliser la prime annuelle en fonction des caractéristiques actuarielles et démographiques, obtenant des résultats significatifs. Decision Tree a présenté des performances élevées avec un score d'entraînement de 0.987, un score de test de 0.983, et un score de validation de 0.984. Ces scores solides sur différents ensembles de données indiquent la capacité de Decision Tree à généraliser efficacement. De plus, l'analyse de l'importance des caractéristiques a fourni des insights sur les facteurs prépondérants dans les prédictions de Decision Tree. En prévoyant la prime annuelle pour de nouvelles données, le modèle Decision Tree a généré une prédiction de 1169.70, tandis que la valeur actuarielle pour ces données était de 1220.91. La proximité entre la prédiction et la valeur actuarielle souligne l'efficacité de Decision Tree pour estimer précisément les primes annuelles. En conclusion, Decision Tree se révèle être une approche robuste et performante dans la modélisation des relations complexes entre les caractéristiques des contrats d'assurance et les primes annuelles.



6 Interface

On vous a fourni un outil python vous permettant de calculer votre prime annuelle lorsque vous soumettez les données requises.

Voici l'interface d'utilisation :



7 Conclusion

Ce projet actuariel, axé sur la modélisation de la prime annuelle d'une assurance vie à la fin de l'année, a été une entreprise exhaustive, couvrant des étapes allant de l'extraction des données à la modélisation prédictive.

— Dans la première phase, nous avons extrait des données actuarielles pertinentes, comprenant l'âge, le nombre de paiements, le différé d'annuité, la durée, le taux technique et le montant de l'annuité, à partir de tables de mortalité et de tarification actuarielles. Une analyse et exploration approfondies des données ont suivi, incluant des statistiques descriptives, des visualisations, et des corrélations pour mieux appréhender la distribution des caractéristiques et identifier des tendances potentielles.

— Dans la phase de modélisation prédictive, nous avons appliqué diverses approches telles que la régression linéaire, la régression polynomiale, la forêt aléatoire et l'arbre de décision. Chaque modèle a été évalué en termes de scores d'entraînement, de test et de validation. La régression linéaire a fourni une base de comparaison malgré sa performance limitée ($R^2 = 0.175$). La régression polynomiale a révélé qu'un polynôme de degré 4 était optimal, affichant la meilleure performance ($R^2 = 0.475$). Random Forest a démontré une performance exceptionnelle avec un score de 0.998, tandis que l'arbre de décision a fourni des résultats solides ($R^2 = 0.983$).

Le choix du meilleur modèle a été guidé par le score de test, et Random Forest est ressorti comme le modèle le plus performant, capturant efficacement la complexité des relations entre les caractéristiques et la prime annuelle. Les prévisions sur de nouvelles données à l'aide de la forêt aléatoire ont été cohérentes avec les attentes actuarielles.

Pour l'avenir, des perspectives passionnantes incluent l'optimisation des modèles existants, l'analyse plus approfondie des motifs de décision, la collecte de données supplémentaires pour enrichir les caractéristiques, et l'intégration d'informations contextuelles externes. Ce projet a démontré la puissance des techniques actuarielles combinées à des approches modernes de modélisation prédictive, offrant des résultats utilisables pour l'optimisation des tarifs d'assurance vie. Les perspectives futures offrent des opportunités passionnantes pour affiner davantage les modèles et enrichir les analyses actuarielles.