Evaluation P1 Romain par Nicolas

Pour cette évaluation, j’ai choisi d’accorder des points d’importances particuliers à la présentation du notebook, l’utilisation des données (récupération, nettoyage, exploitation) et pour finir, la méthode et le choix des éléments mis en relations.

# **Présentation du notebook**

La mise en forme du notebook est bien réalisée, aéré et lisible, ce qui me parait très important pour la lecture d’un document de ce type, mêlant code et français.

Un petit bémol concerne les titres, peu mis en valeurs pour bien délimiter chaque partie.

# **Utilisation des données**

## Récupération

Romain a su récupérer les données des fichiers CSV, et les a très bien présenté en nous montrant successivement les premières lignes ou encore les différents types des colonnes.

## Nettoyage

Romain a choisi de nettoyer les données de types string qui pouvaient être convertis en entier, ainsi que les valeurs null.

Pour l’âge, il a insérer une nouvelle colonne alimentée avec les âges médians, il aurait aussi pu directement alimenter les valeurs null dans la même colonne afin d’éviter de recréer une colonne.

## Exploitation

Romain a effectué 3 analyses :

* Influence de l’Age selon le tarif :

Tout d’abord, il a réalisé un graphique avec ces deux paramètres, il en déduit très bien que la plupart des passagers ont payé leur billet moins de 50€, et en particulier les enfants de 10ans et moins. De plus, on se rend compte avec ce graphique que l’âge n’a pas réellement d’influence sur le tarif au-delà de 10 ans.

Ensuite, il a développé une régression linéaire de ces deux facteurs, ce qui lui a permis de voir qu’une légère hausse était présente lorsque l’âge augmentait.

Pour finir, il a effectué des prédictions sur les tarifs en fonction de l’âge grâce à la fonction predict().

* Influence du tarif sur la survie :

Pour débuter, il a effectué un graphique des passagers survivants/décédés en fonction du prix du billet qu’ils avaient payés. Pour moi ce graphique n’est pas parlant, il ne permet pas de tirer de conclusion, ni même d’hypothèses. Cependant, avec la régression linéaire qu’il fait juste après, on se rend très vite compte qu’il y a un très gros impact sur les chances de survie si on paie un billet plus cher.

Petite suggestion : Pourquoi effectuer une régression linéaire sur des prédictions, alors que les données réelles sont à notre disposition ?

* Influence de l’Age selon le tarif en prenant en compte les personnes qui ont survécu (Toutes, puis uniquement les hommes, puis uniquement les femmes)

Tout d’abord, il a réalisé un graphique mettant en lien le nombre d’hommes et de femmes décédés. Il a décidé d’y ajouter aussi le critère d’âge, sans conséquences.

Puis, il a décidé de séparer les hommes des femmes, et de vérifier le taux de survie de chacun en fonction du tarif payé, ceci grâce à des régressions linéaires.

# **Méthode – choix des éléments mis en relation**

Romain est très méthodique, comprend son sujet et a réalisé beaucoup d’essaies et de lecture d’informations dès le début afin de bien s’approprier le sujet.

De plus, il a choisi des exemples convaincants, mis à part peut-être le fait de coupler les 3 éléments [Age, Sexe, Prix] qui aurait pu bénéficier d’une analyse complémentaire.