

# Soutenance Projet 7: Implémentez un modèle de scoring (18/02/2021)

Miguel José BERNAL MORA

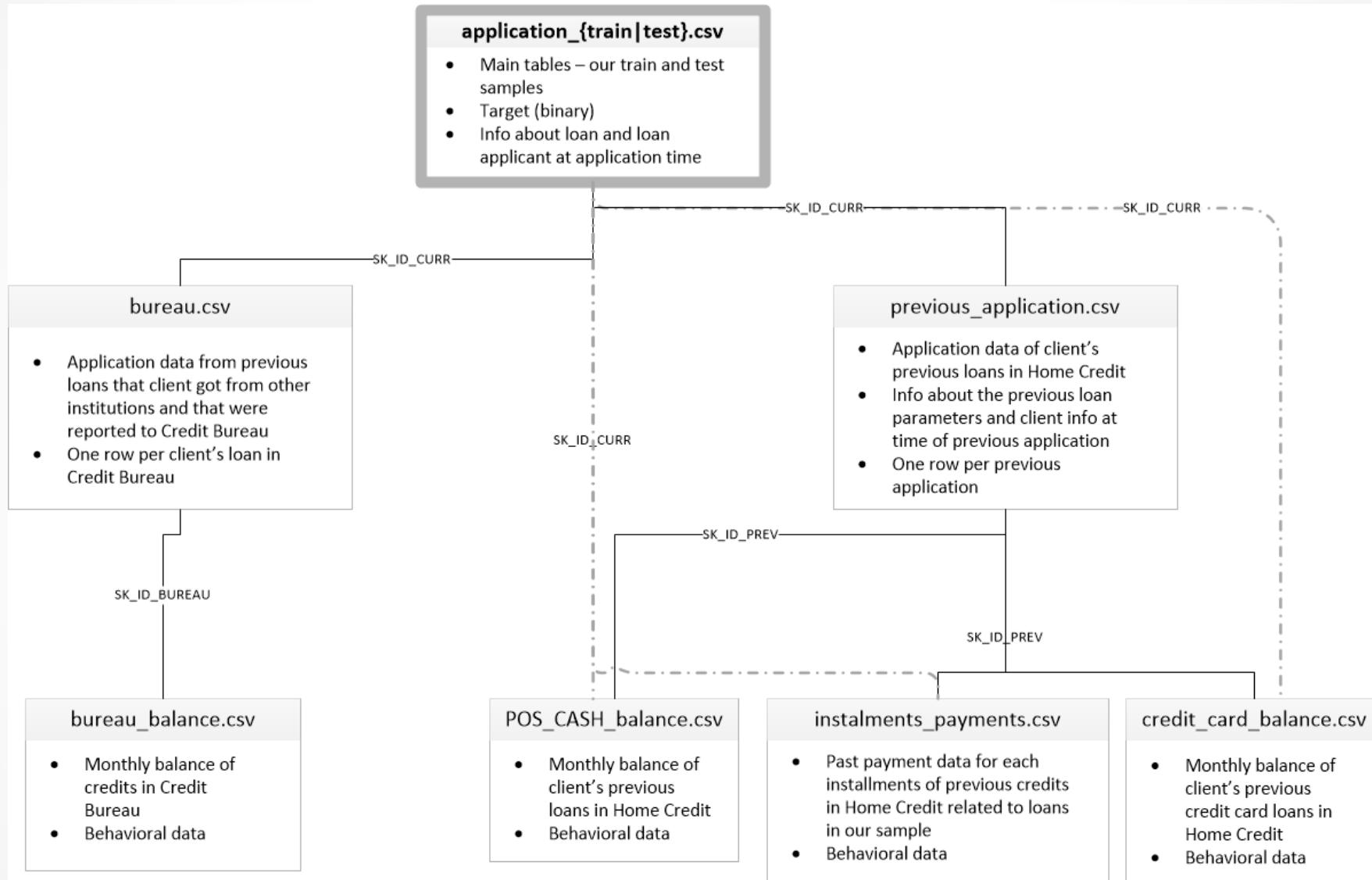
Mentor: Alison Patou

Évaluateur: Olivier Chotin

# Problématique

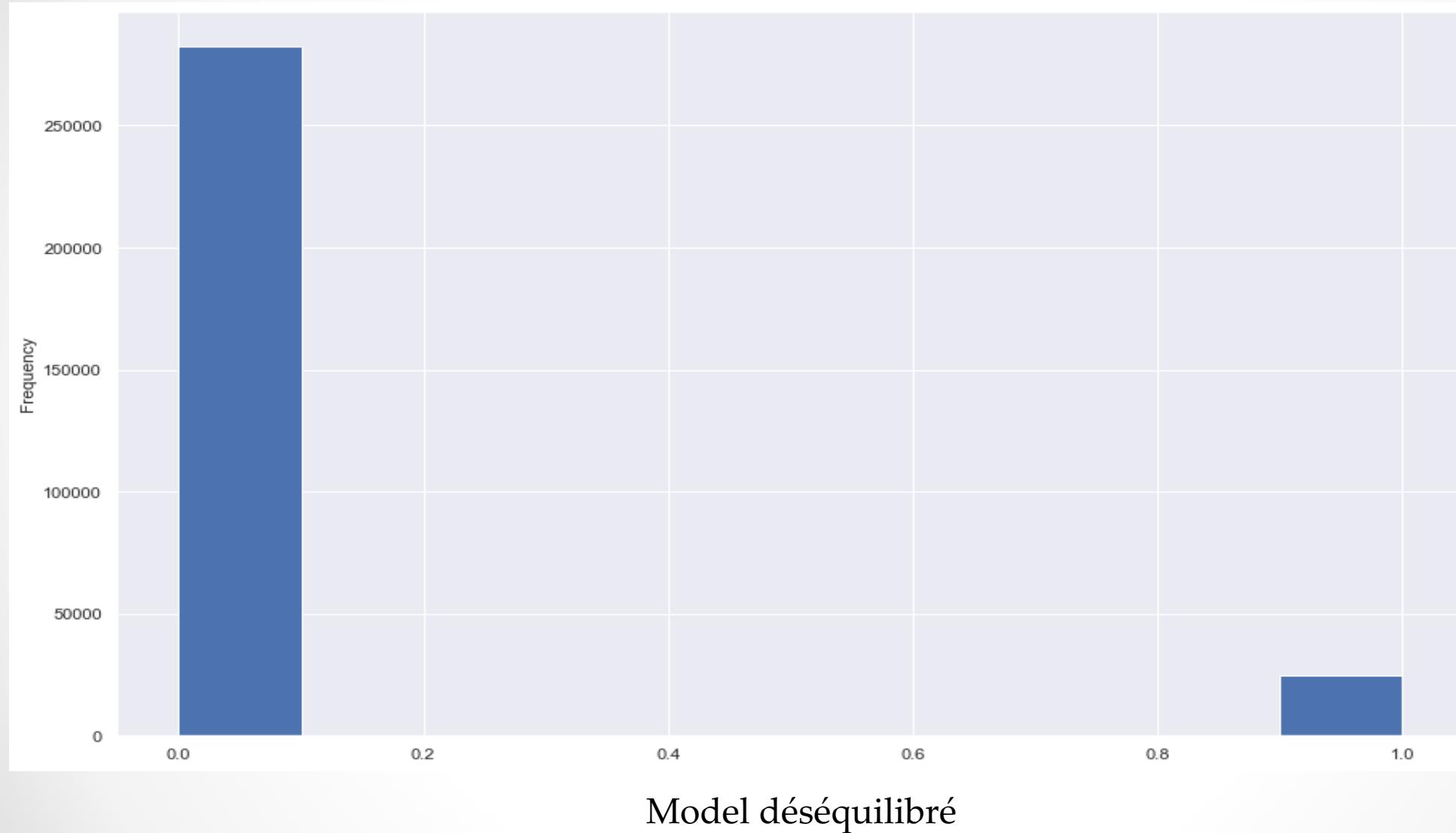
- La société financière "Pret à dépenser", est une société qui est chargée d'accorder des crédits aux personnes qui ont un faible risque selon leur historique de prêt. Pour ce faire, l'entreprise veut développer un modèle de notation de la probabilité de paiement du client pour décider d'accorder ou non le prêt
- 1) Construire un Model Scoring en montrant la probabilité de faillite d'un client (ou réussite)
  - 2) Construire Dashboard Interactive descriptive

# Données

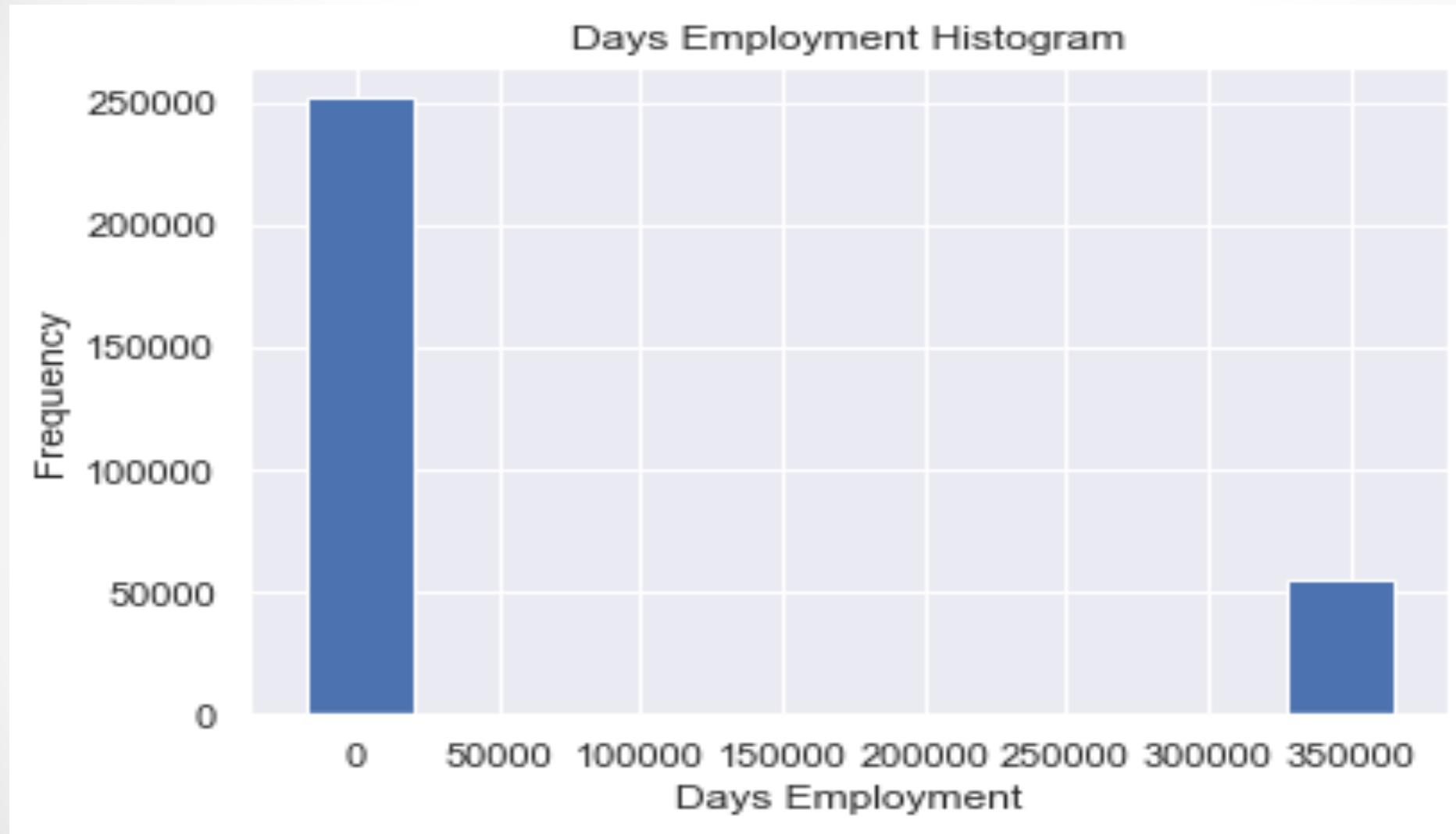


Source: Kaggle Featured Prediction Competition

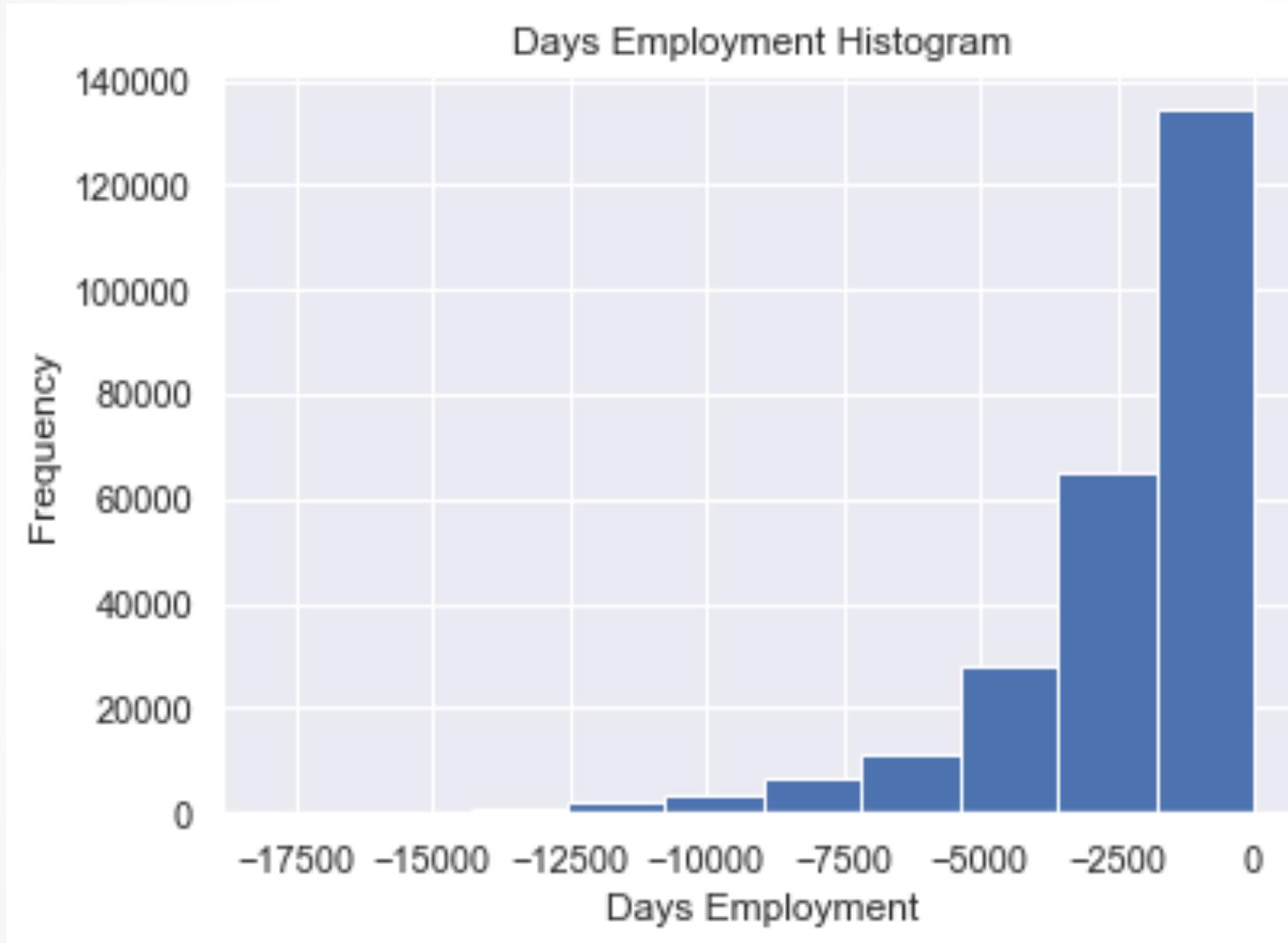
# Analyse exploratoire et « Features Engineering »



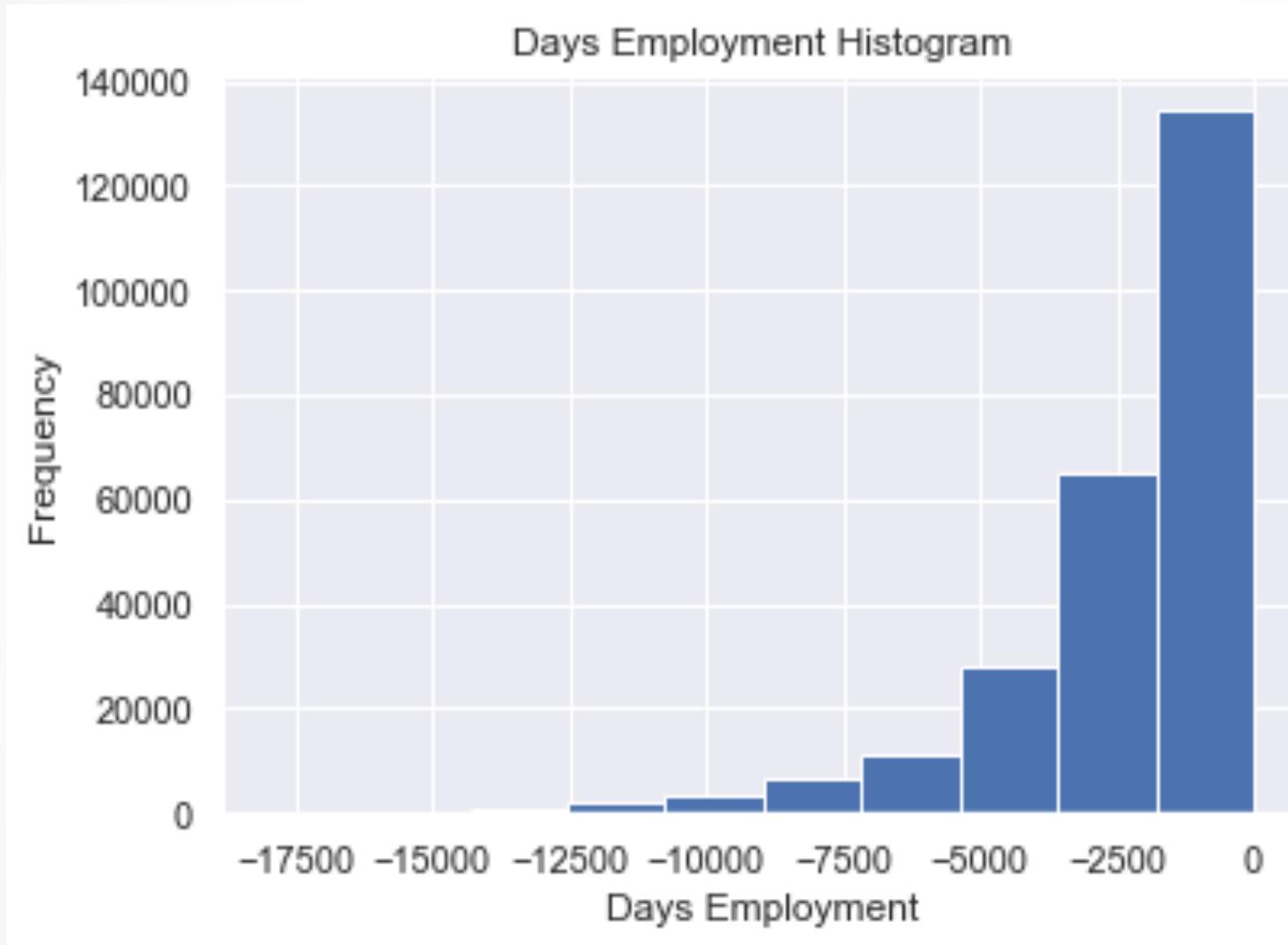
# Analyse exploratoire et « Features Engineering »



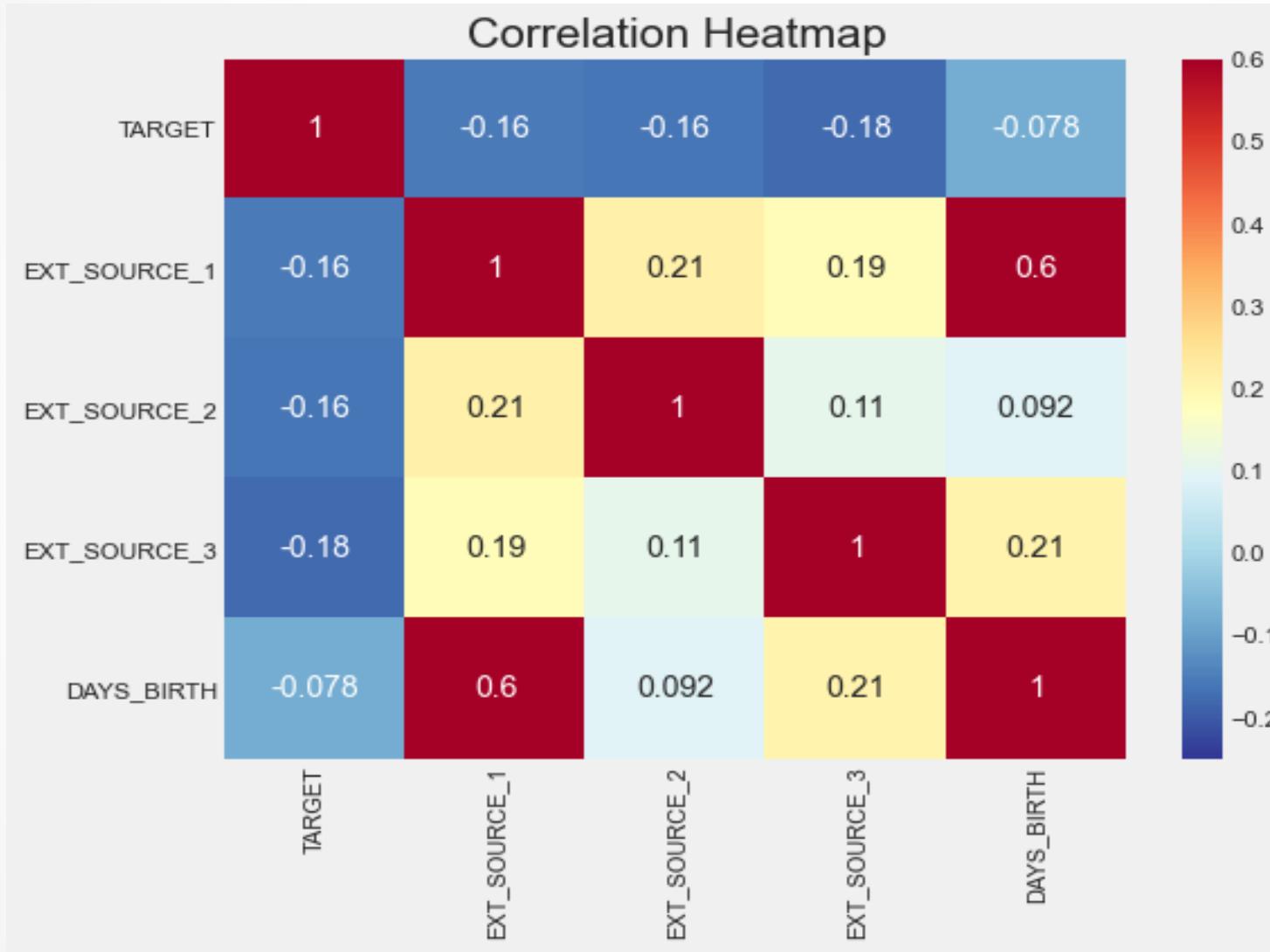
# Analyse exploratoire et « Features Engineering »



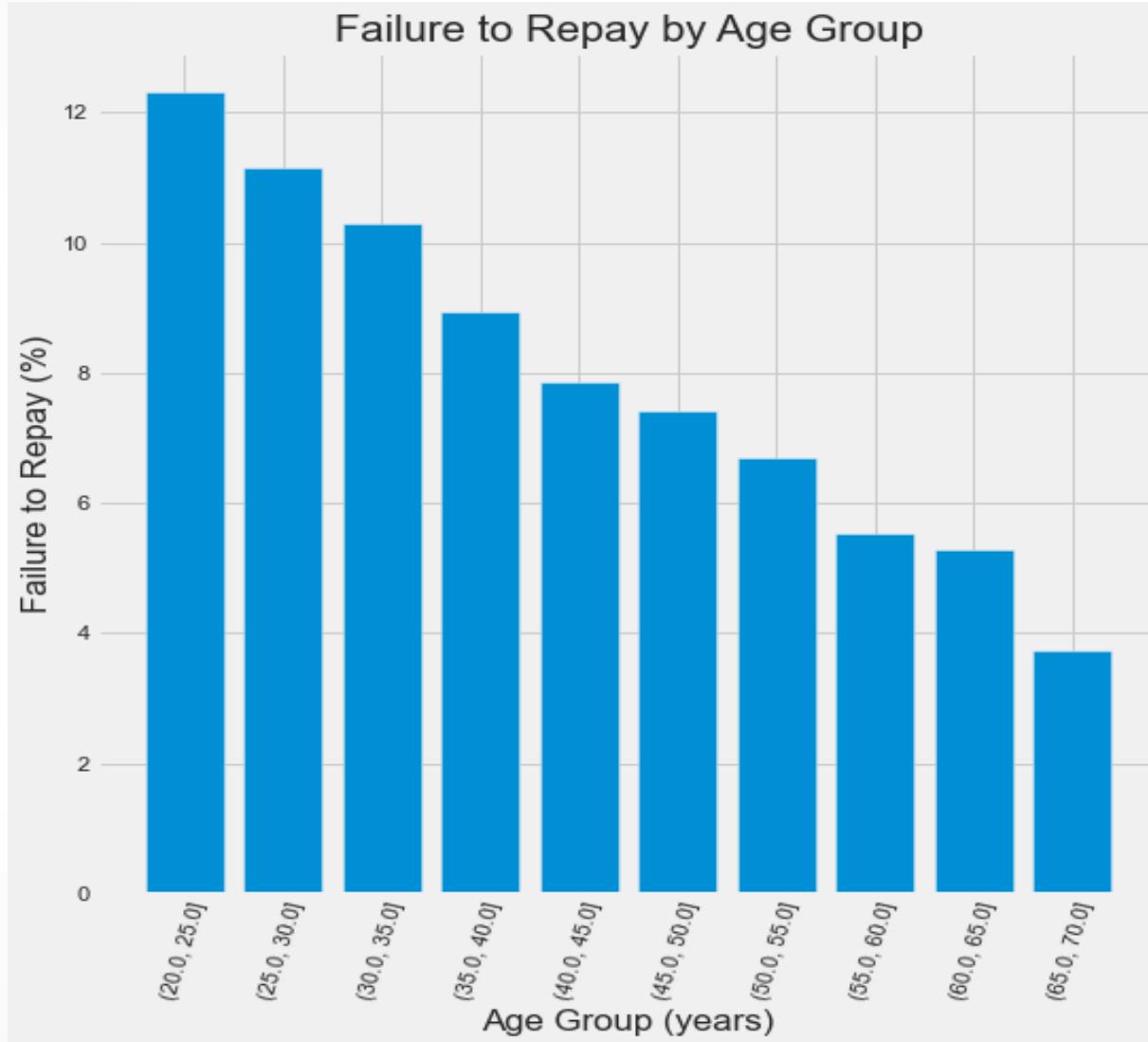
# Analyse exploratoire et « Features Engineering »



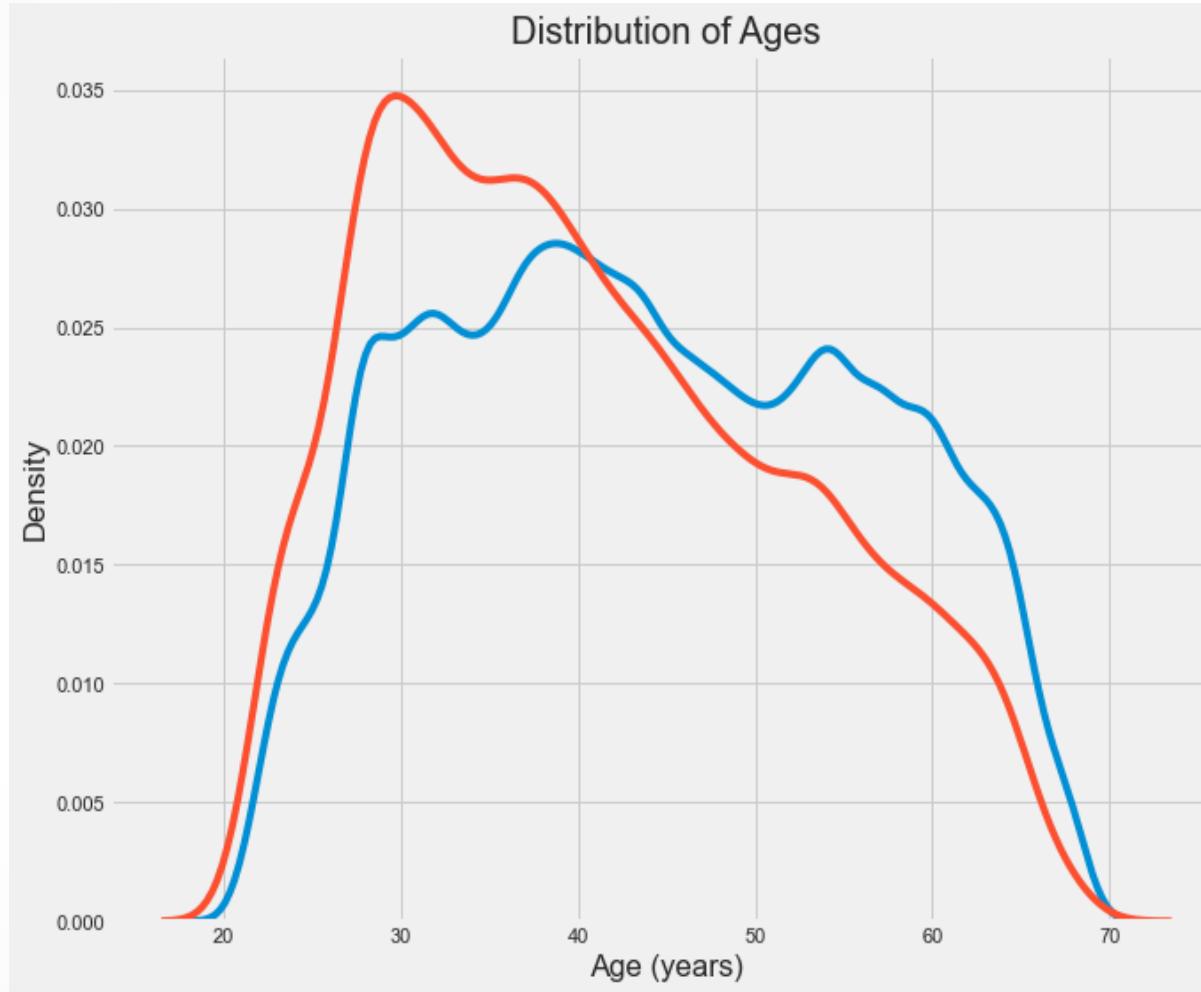
# Analyse exploratoire et « Features Engineering »



# Analyse exploratoire et « Features Engineering »

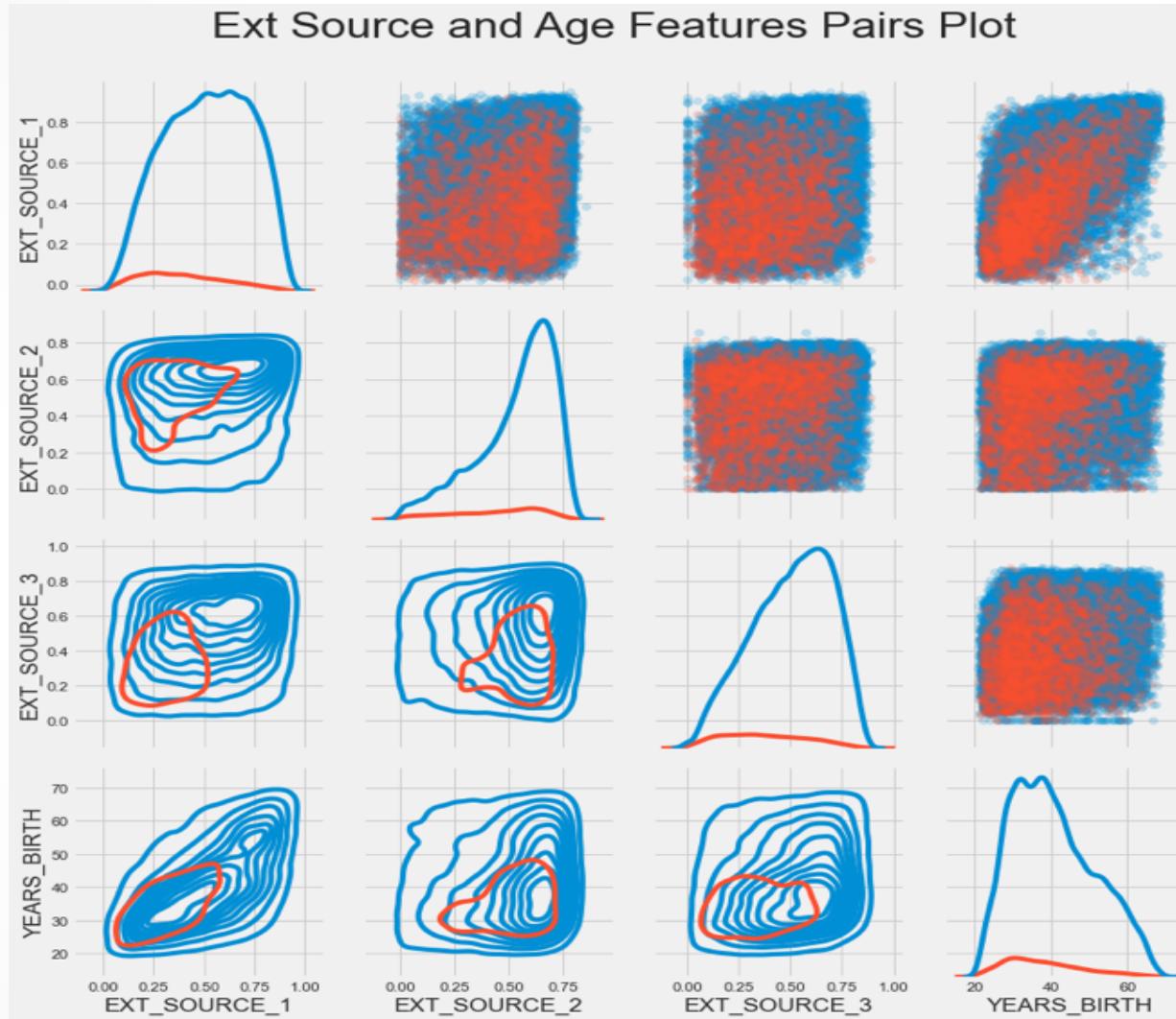


# Analyse exploratoire et « Features Engineering »



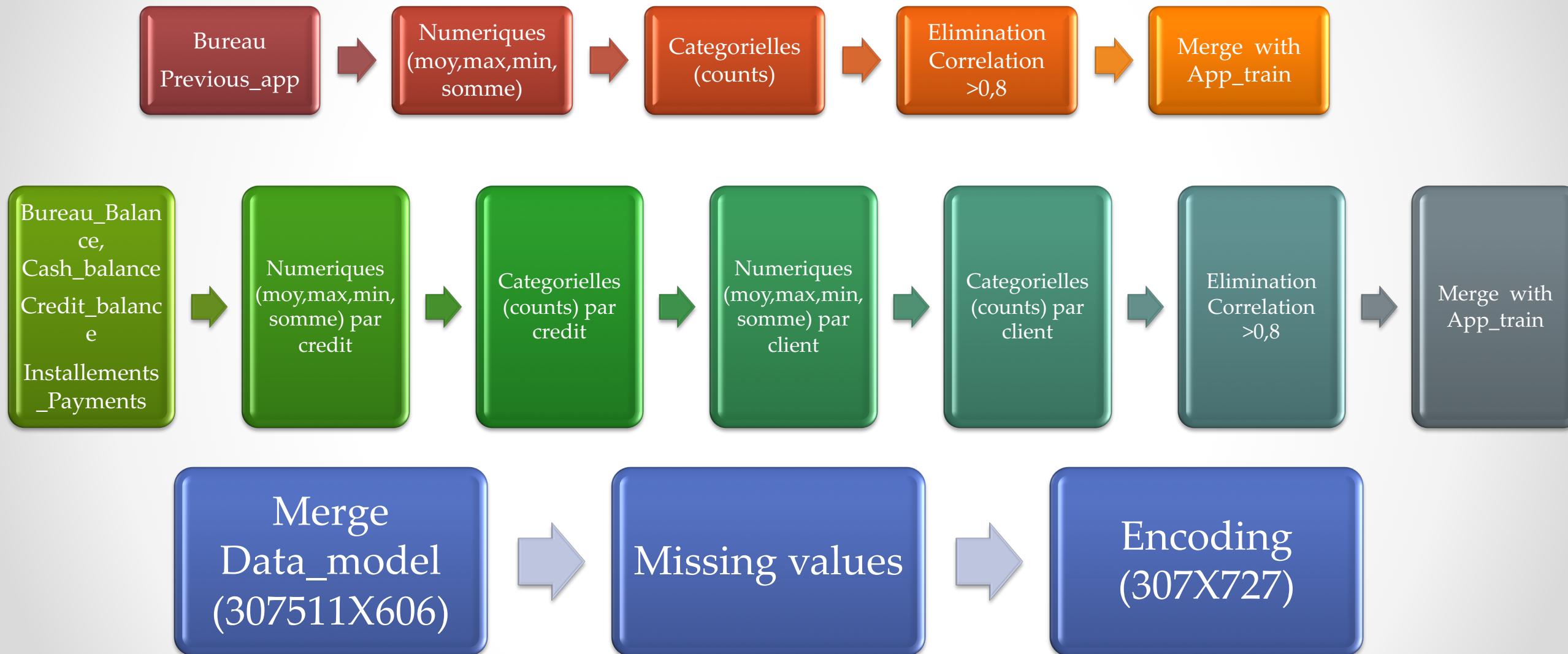
« 1 » Ligne Rouge, « 0 » Ligne Bleue

# Analyse exploratoire et « Features Engineering »

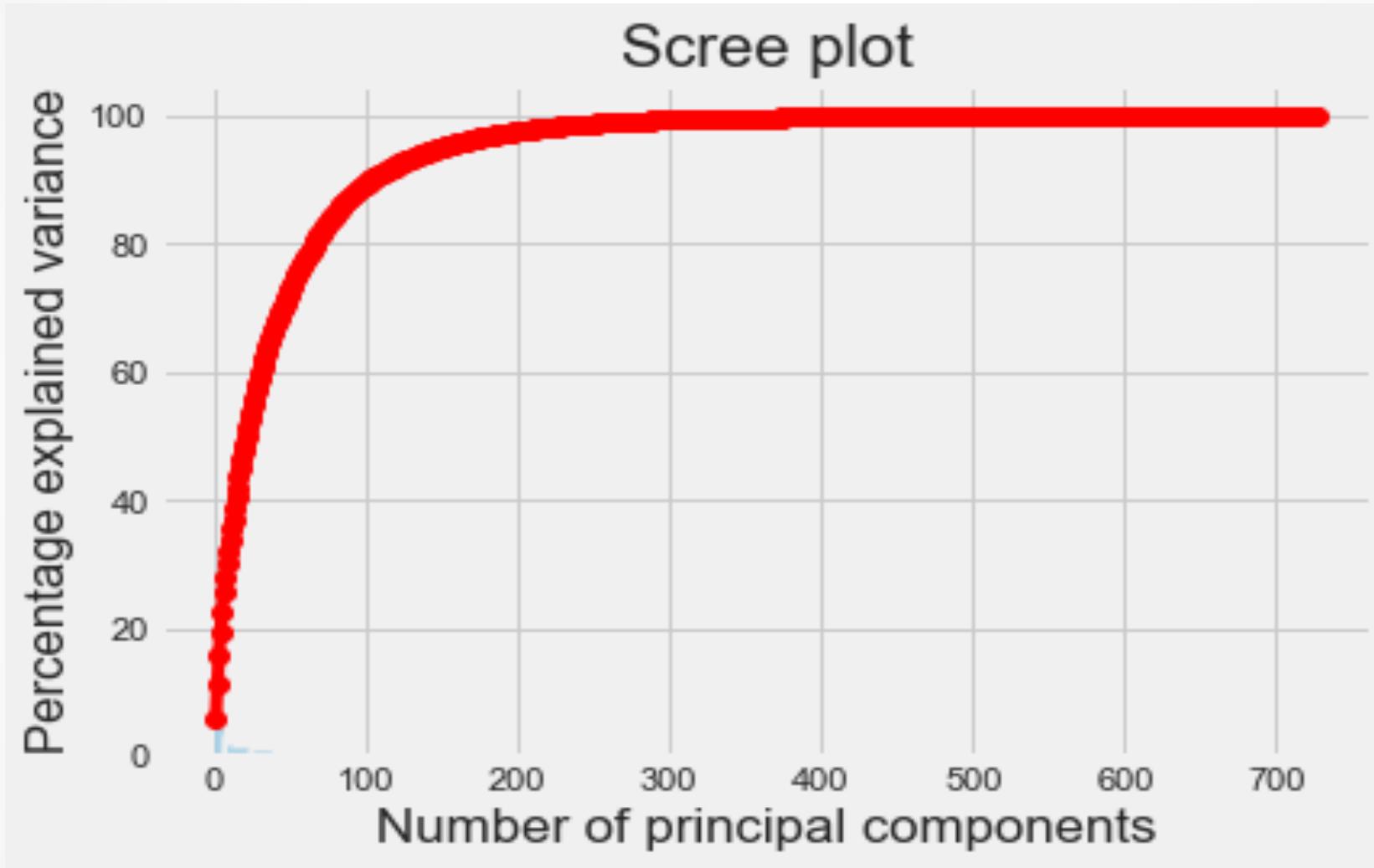


« 1 » Ligne Rouge, « 0 » Ligne Bleue

# Analyse exploratoire et « Features Engineering »



# Modélisation



Data\_model\_Reduced 307511 X100

# Modélisation : Fonction Coût

$$\bullet F_{\beta} = ((1 + \beta^2) * \text{Precision} * \text{Recall}) / (\beta^2 * \text{Precision} + \text{Recall})$$

Les trois valeurs communes du paramètre bêta sont les suivantes :

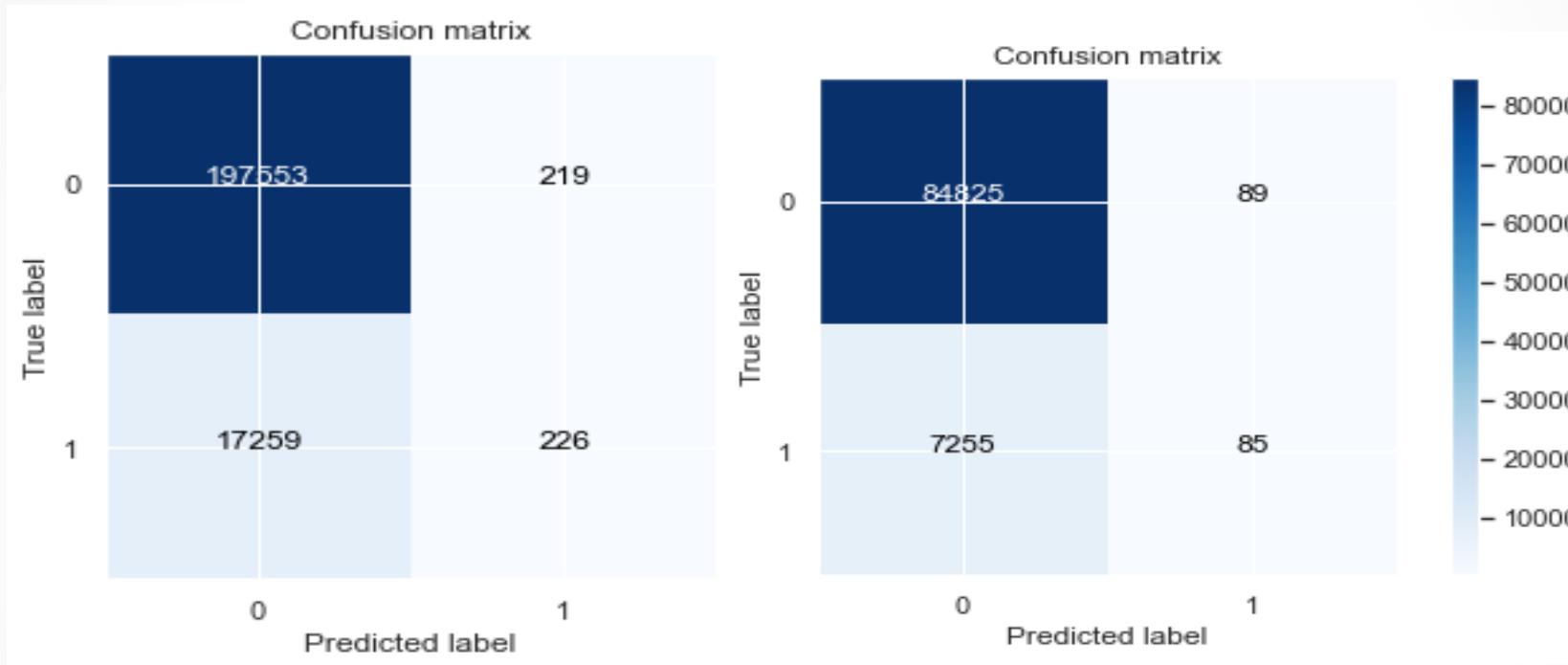
$F_{0,5}$ -Mesure ( $\beta=0,5$ ) : Plus de poids sur la précision, moins de poids sur le rappel.

$F_1$ -Mesure ( $\beta=1,0$ ) : Équilibre le poids de la précision et du rappel.

Mesure  $F_2$  ( $\beta=2,0$ ) : Moins de poids sur la précision, plus de poids sur le rappel.

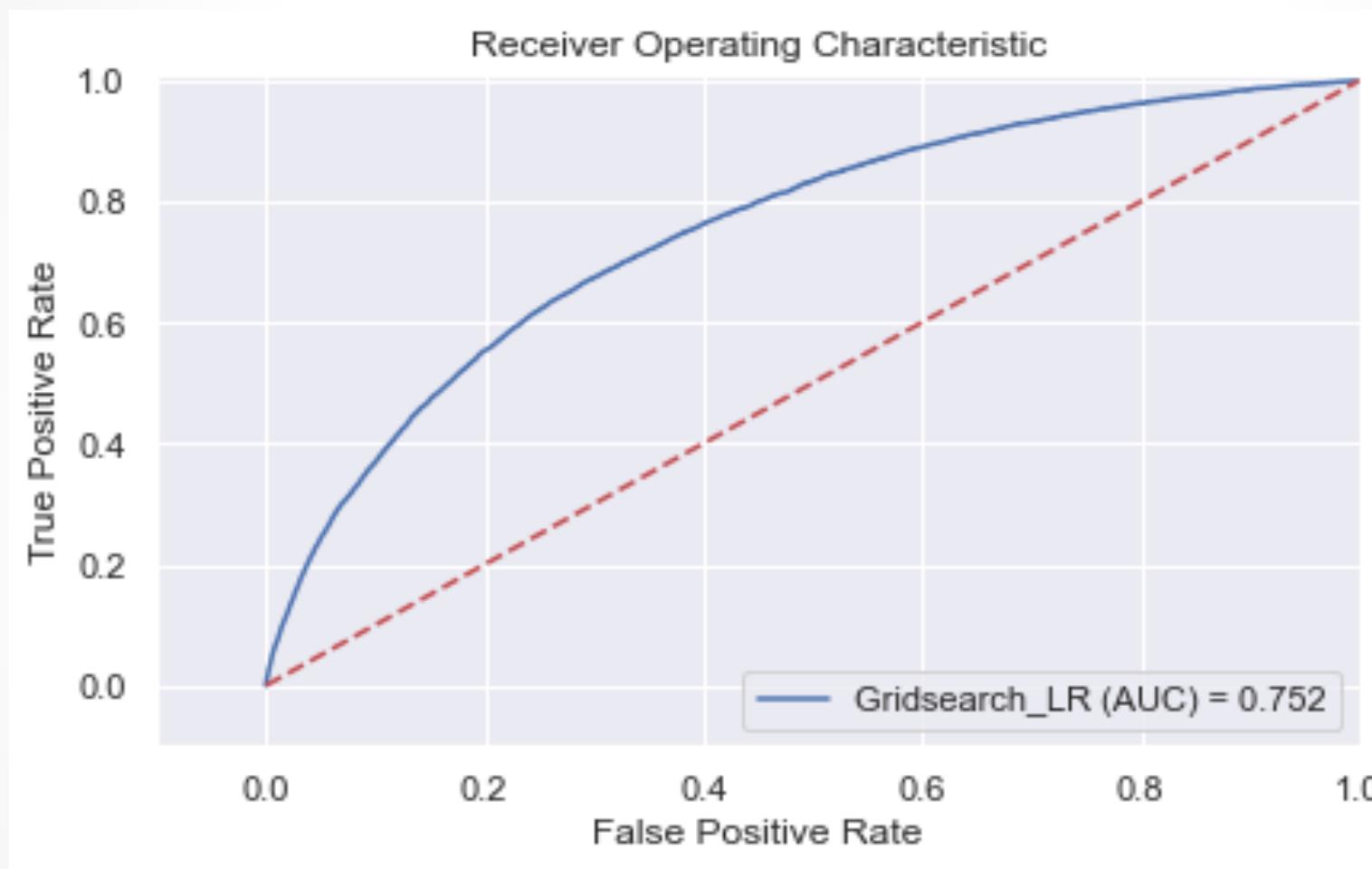
On a utilisé un  $\beta=2,5$

# Modélisation : Regression Logistique

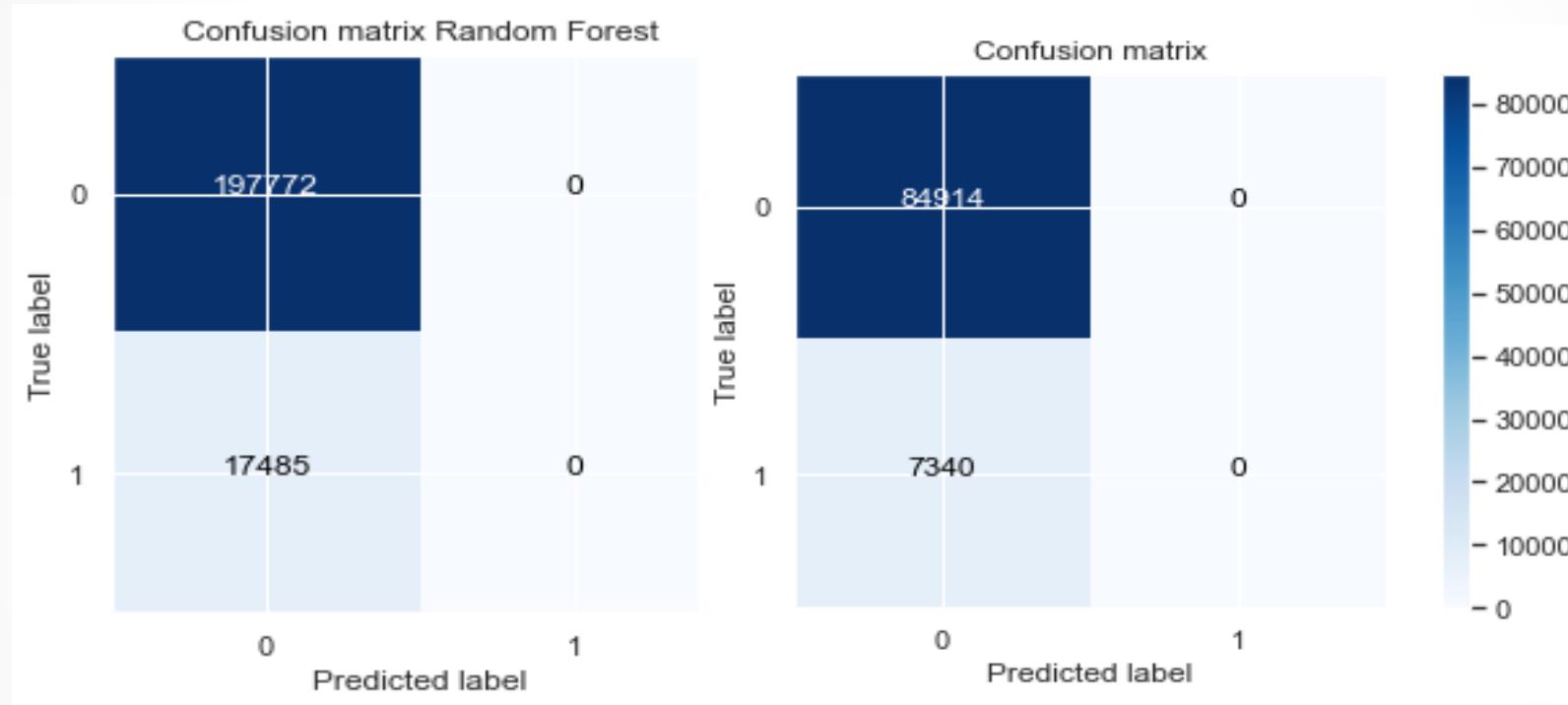


Matrice Confusion X\_train and X\_test LR

# Modélisation : Regression Logistique-ROC

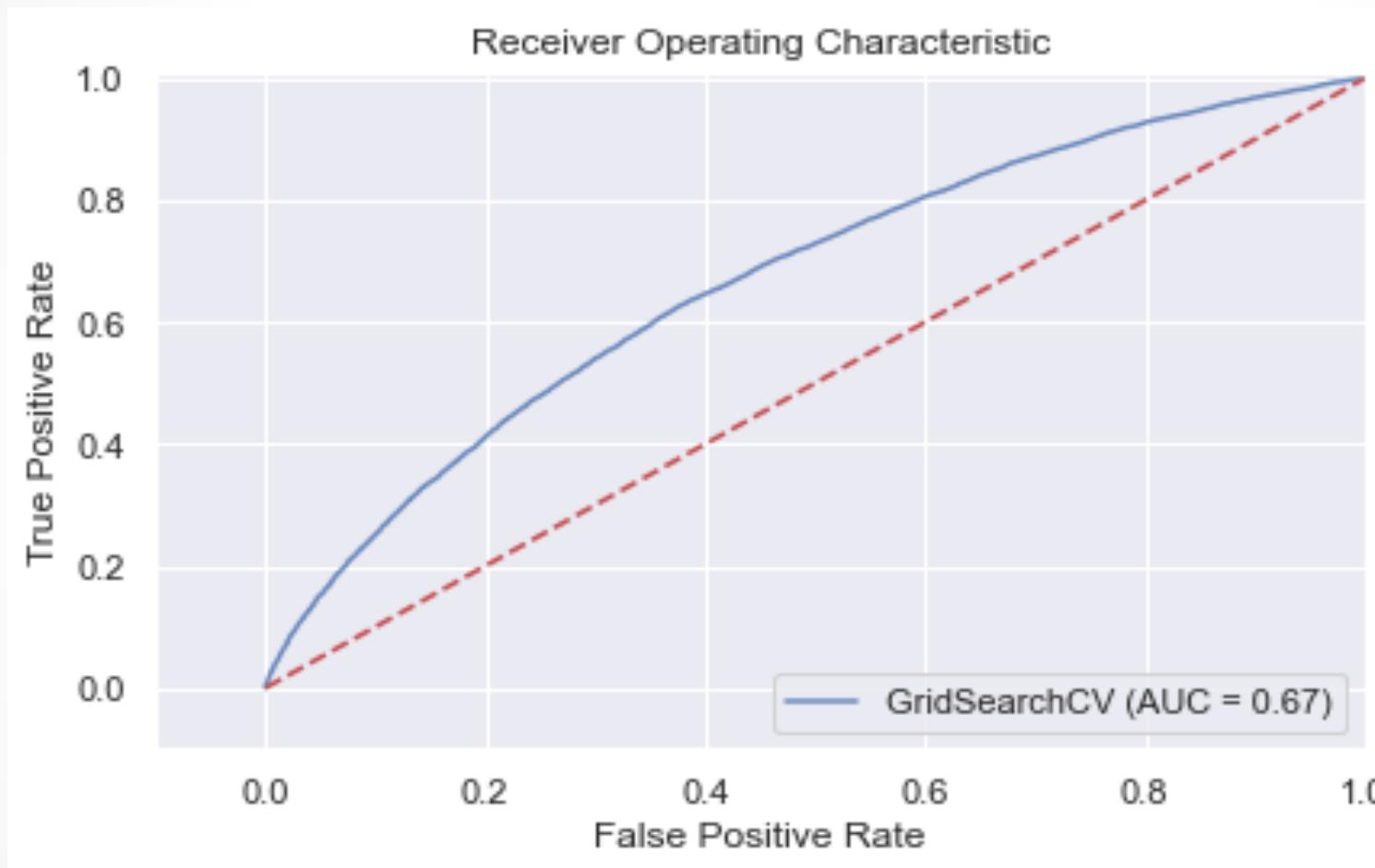


# Modélisation : RandomForest-Matrice Confusion

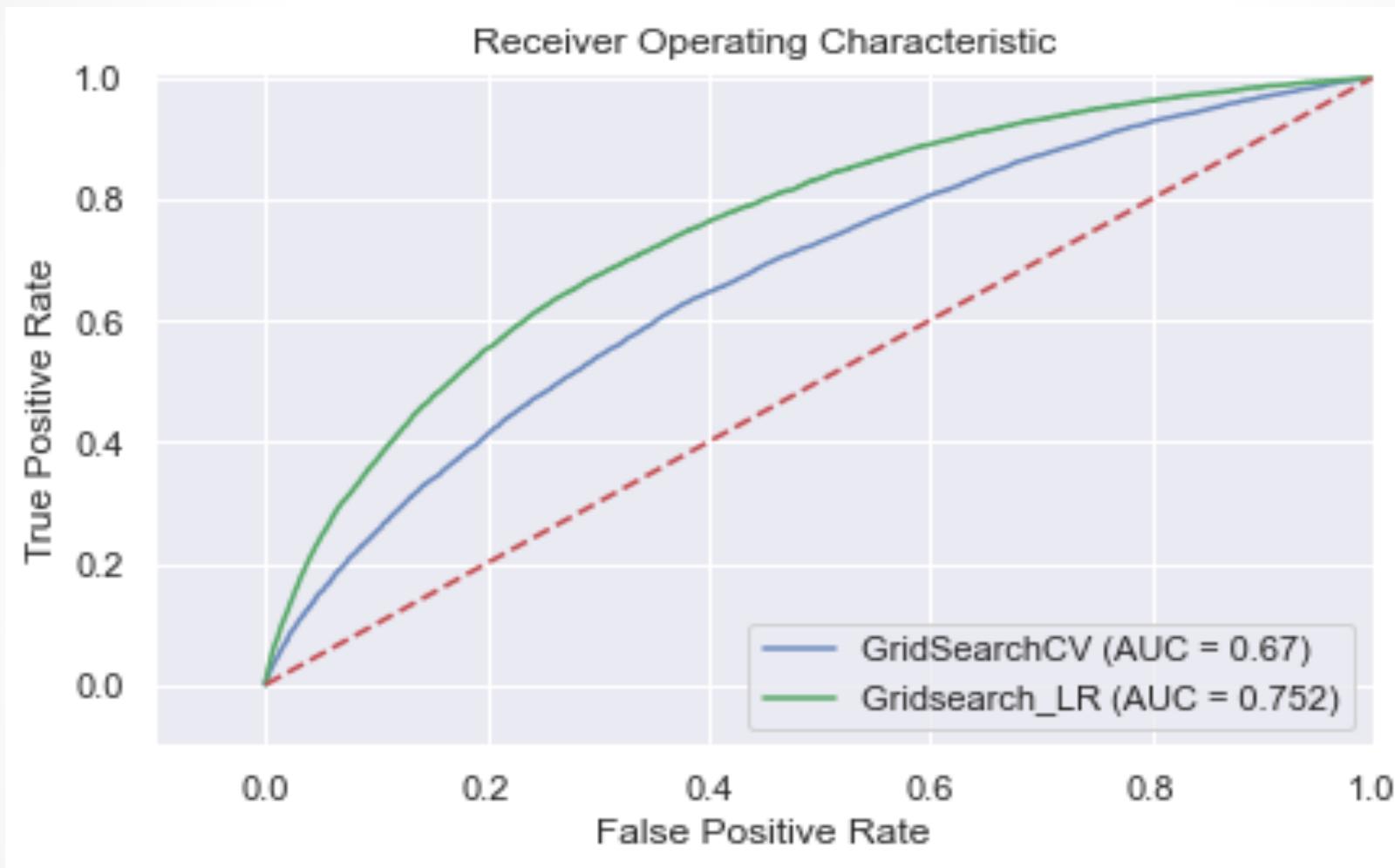


Matrice Confusion X\_train and X\_test RF

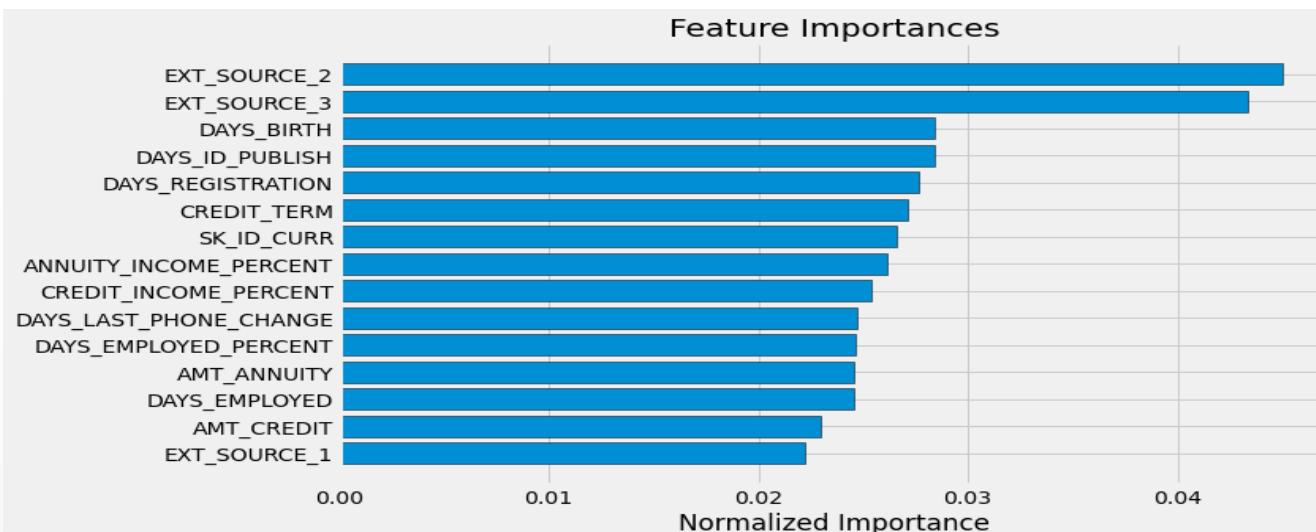
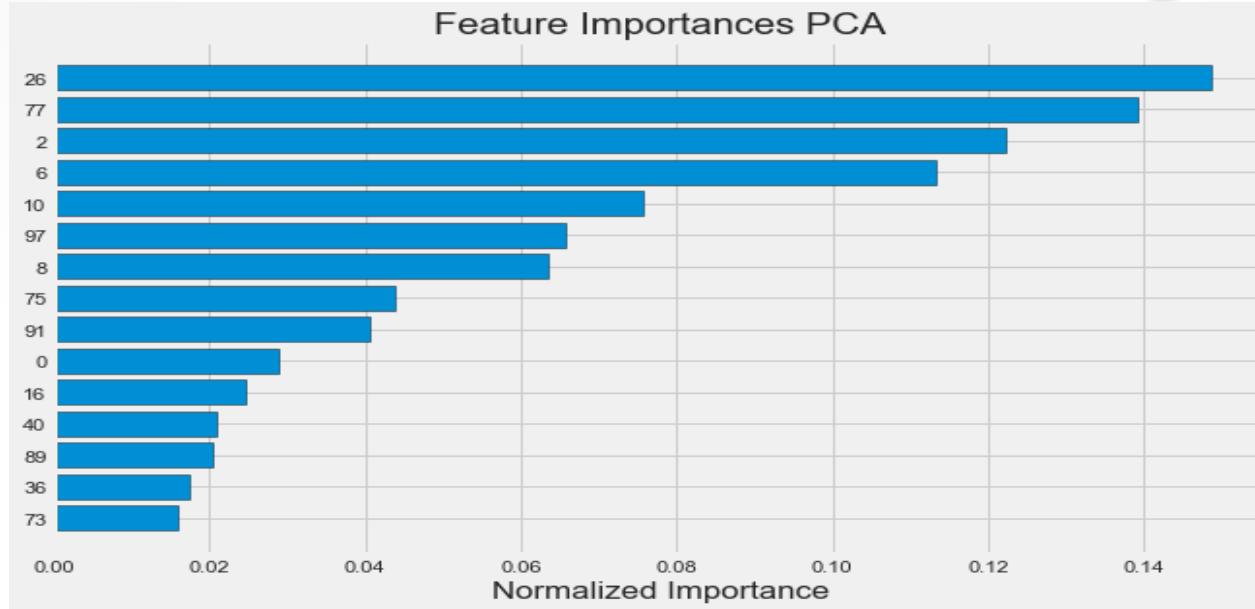
# Modélisation : RandomForest-ROC



# Modélisation : RandomForest vs RegLog-ROC



# Modélisation : RandomForest Featuring Importances



Source: Kaggle Featured Prediction Competition

# Dashboard

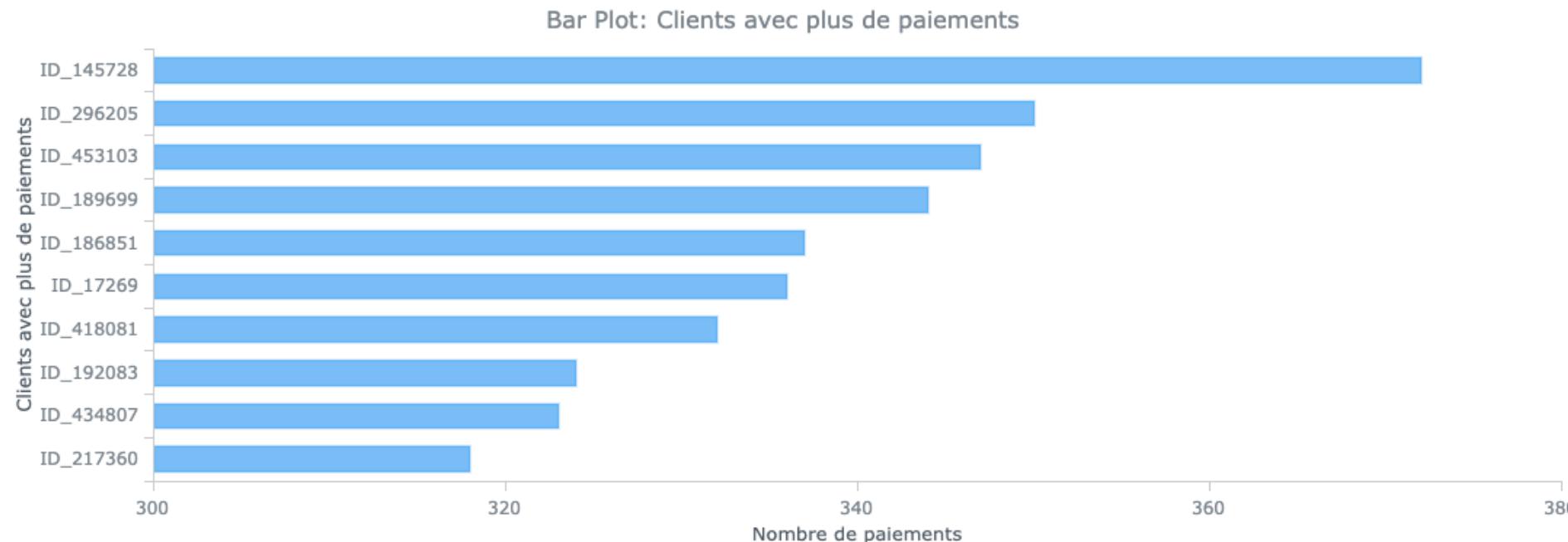
## Dashboard Scoring Paiement Credit

Prediction\_paiement\_reussite\_clients

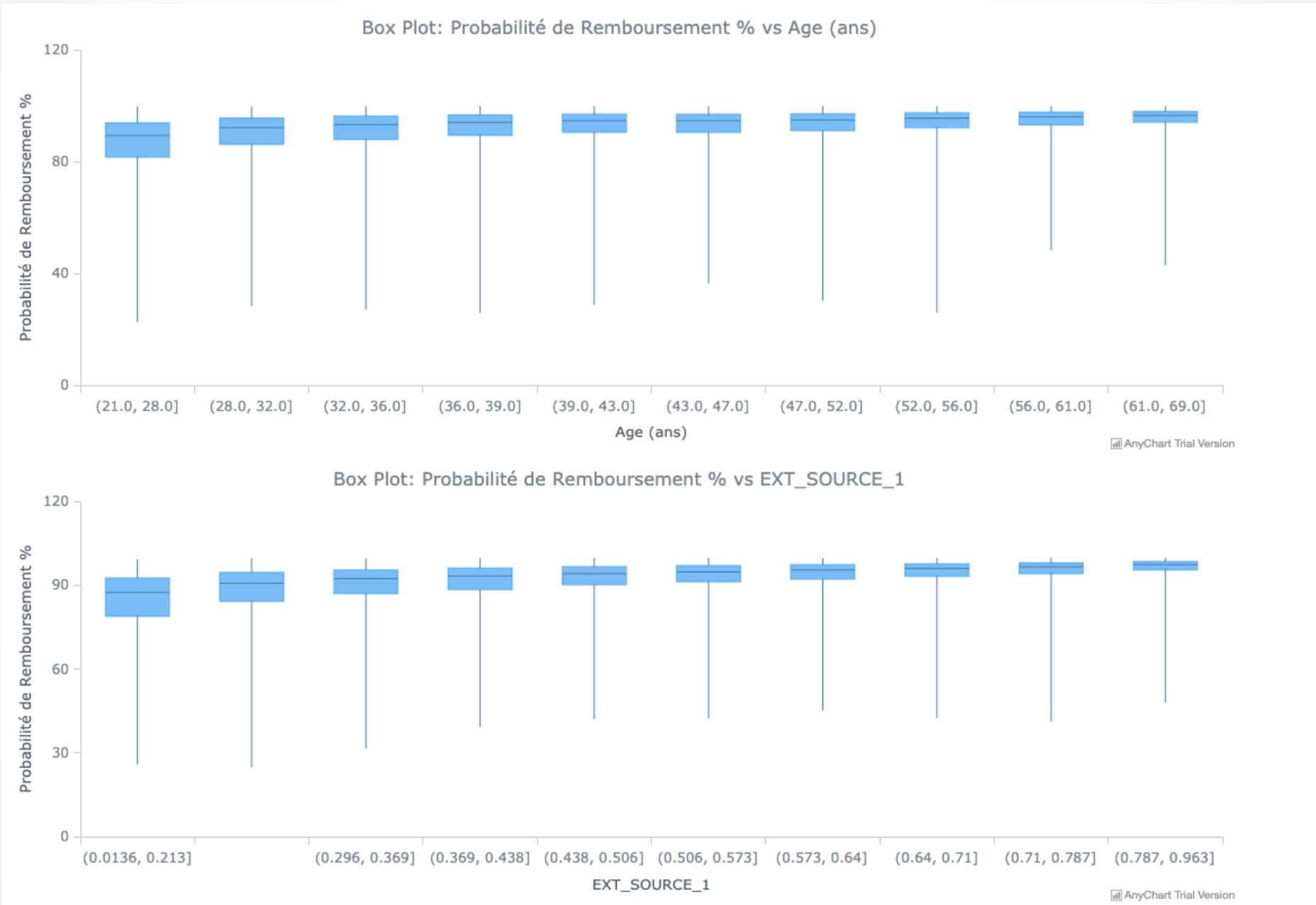
Le client 100002 aura une probabilité de réussite de remboursement de 58.45 %

Pour changer de client saisir le id après predict/ sur la bar de navigation html

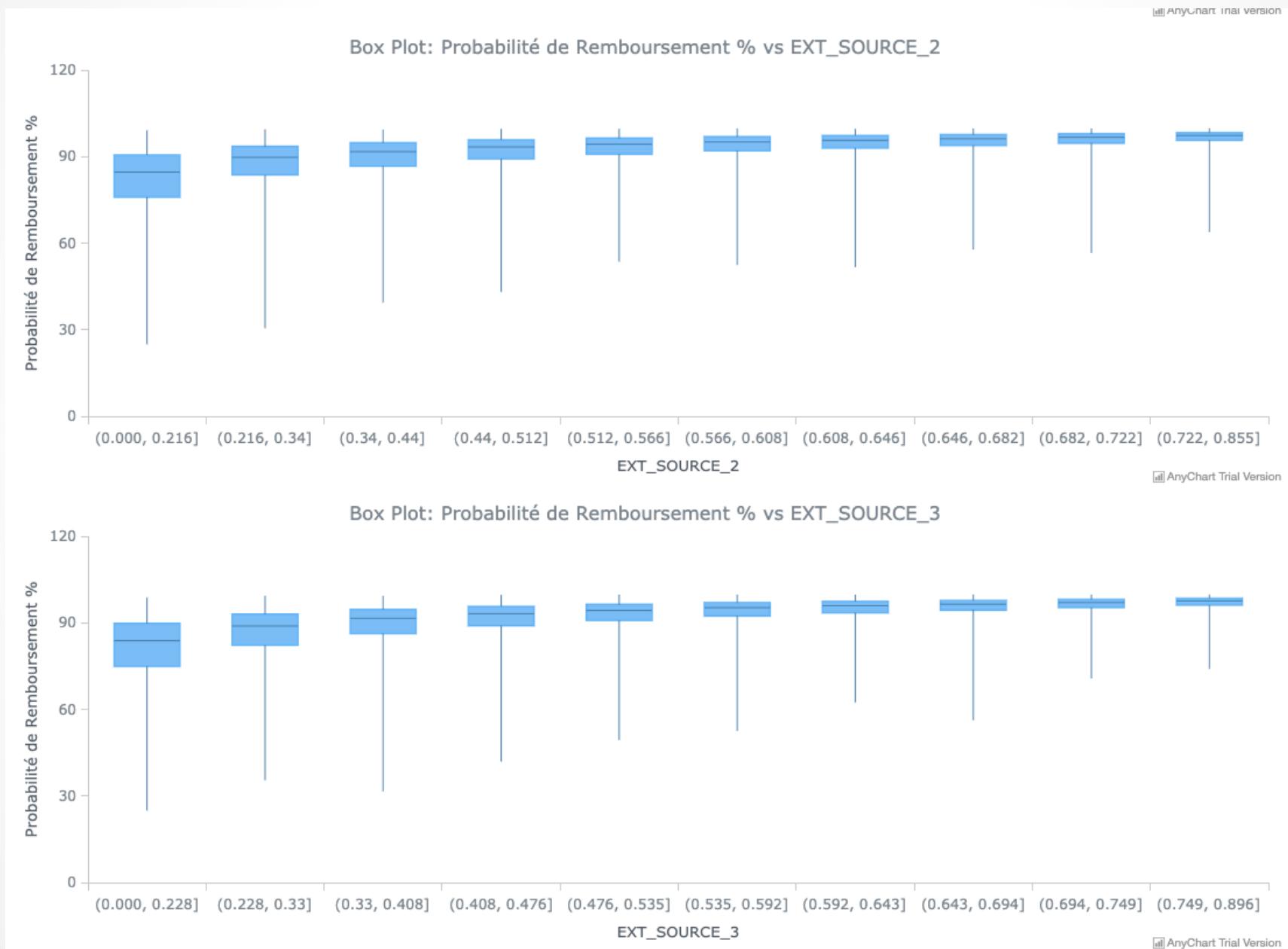
- 1) Bar Plot qui montre les 10 premiers clients avec plus de paiements.
- 2) Box Plot Descriptive pour montrer le changement de probabilité en fonction d'âge des clients
- 3) Box Plot Descriptive pour montrer changement probabilité en fonction d'une source externe normalisée
- 4) Box Plot Descriptive pour montrer changement probabilité en fonction d'une deuxième source externe normalisée
- 5) Box Plot Descriptive pour montrer changement probabilité en fonction d'une troisième source externe normalisée



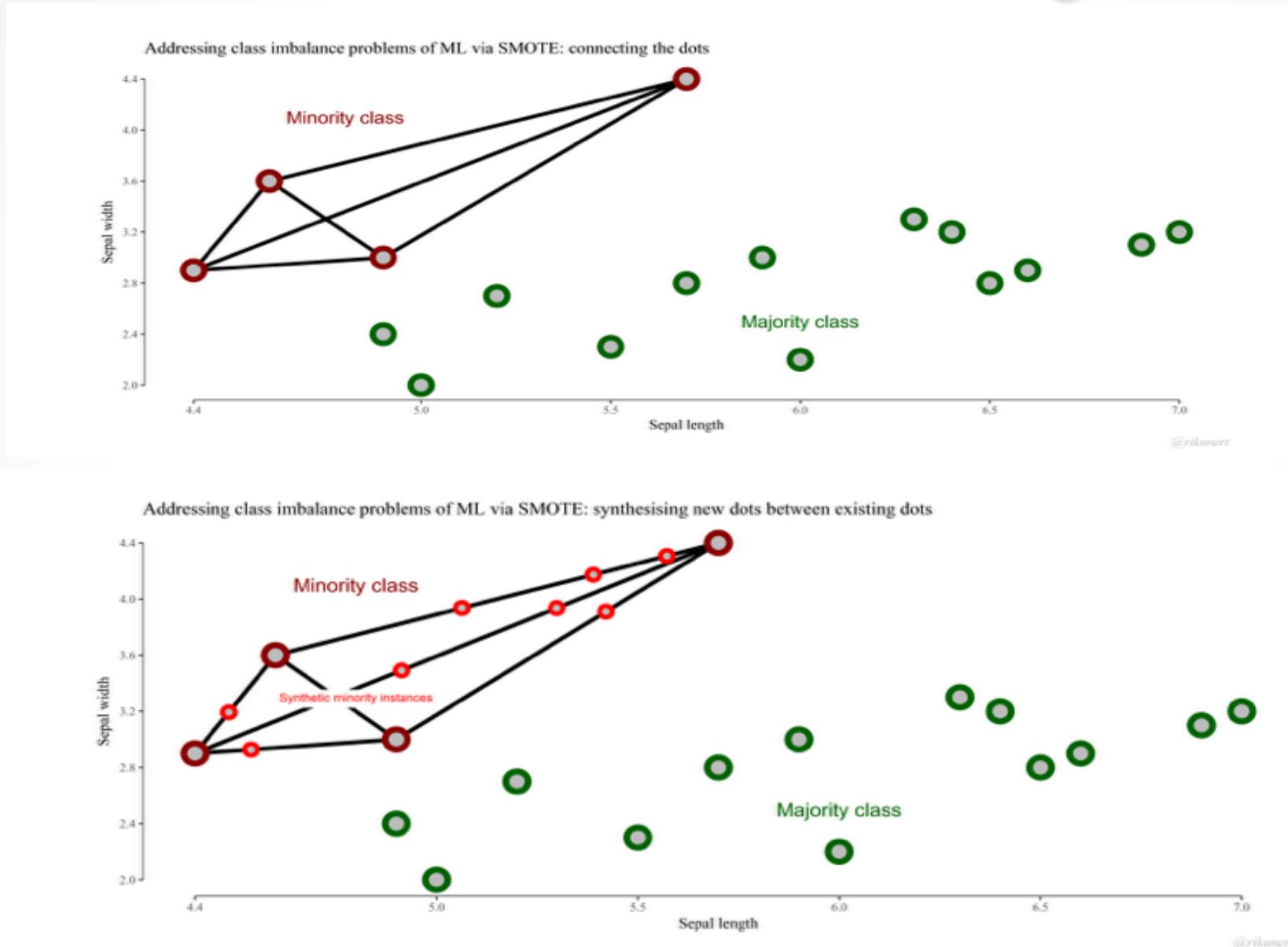
# Dashboard



# Dashboard



# Améliorations et désavantages



Source: [https://rikunert.com/SMOTE\\_explained](https://rikunert.com/SMOTE_explained)

# Questions?



спасибо 谢谢  
**GRACIAS**

**THANK YOU**

ありがとうございました **MERCI**

**DANKE** धन्यवाद

شُكْرًا

**OBRIGADO**