

Analyse statistique descriptive du churn client bancaire

Approche descriptive fondée sur une logique métier bancaire

Malek Hattab

2026-01-02

Introduction

Le churn client correspond à la perte d'un client lorsque l'ensemble de ses comptes bancaires sont fermés.

Dans le secteur bancaire, la compréhension des mécanismes de churn constitue un enjeu stratégique majeur, notamment pour orienter les actions de rétention et d'amélioration de la relation client.

Objectif du travail

- Construire un jeu de données client fiable
- Réaliser une analyse statistique **descriptive et comparative** du churn
- Approche **strictement non prédictive**

Données utilisées

- **CUSTOMERS** : informations clients
- **ACCOUNTS** : comptes bancaires
- **BALANCES** : soldes financiers

Les données proviennent de différentes tables d'un même système d'information bancaire et présentent des relations de type client-comptes, nécessitant un traitement spécifique.

Fusion brute (diagnostic)

- ▶ Afficher le code

```
[1] 373156      22
```

Cette fusion est volontairement brute afin de mettre en évidence les duplications, les valeurs manquantes et les incohérences structurelles avant toute phase de nettoyage ou d'agrégation.

Nettoyage : types & valeurs manquantes

- Afficher le code

Le traitement des valeurs manquantes repose sur une logique métier bancaire, visant à conserver l'information tout en assurant la cohérence analytique du jeu de données.

Variables analytiques & valeurs aberrantes

- ▶ Afficher le code

Les variables âge et ancienneté sont calculées à partir de la date courante, conformément à l'approche adoptée dans le notebook de référence.

Agrégation au niveau client

- ▶ Afficher le code

Le solde moyen est utilisé comme indicateur synthétique du comportement financier du client.

Définition du churn

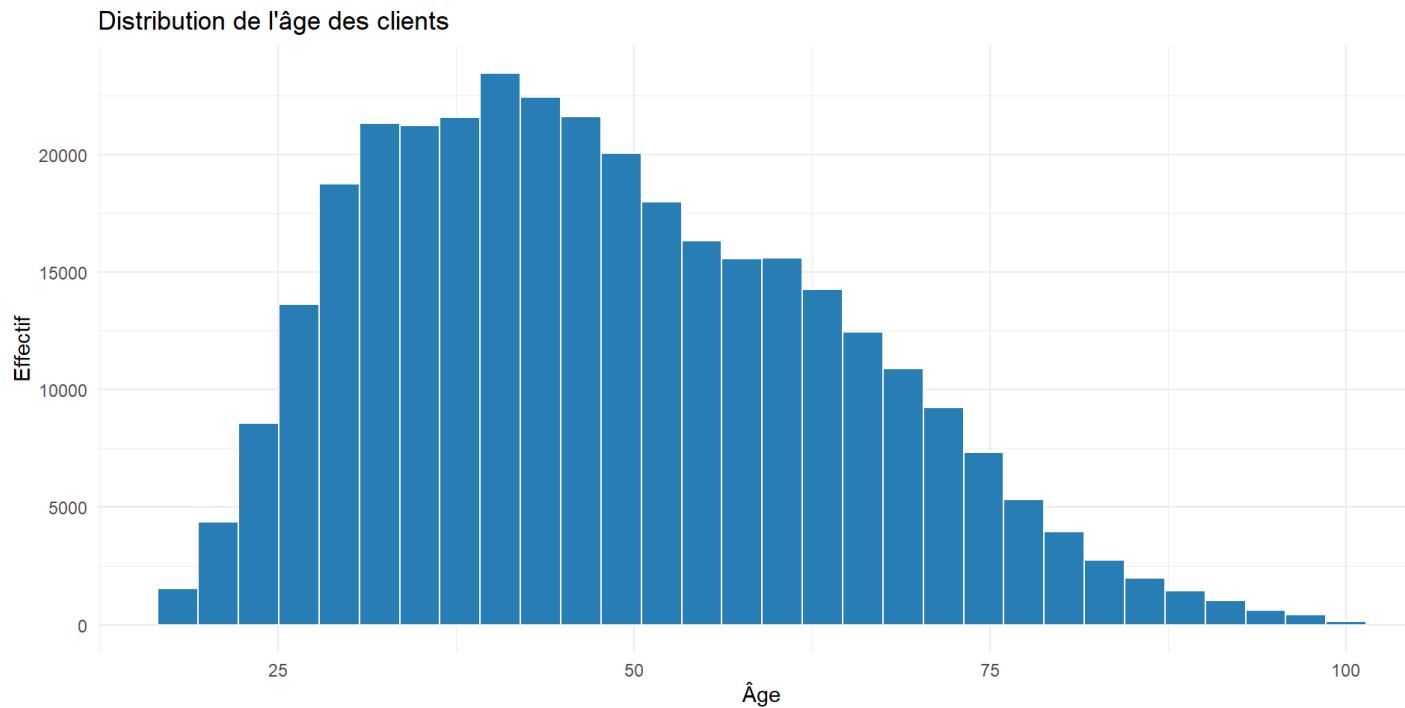
- Afficher le code

0	1
248959	114354

Un client est considéré comme churné lorsqu'il ne détient plus aucun compte actif.

Histogramme : âge (global)

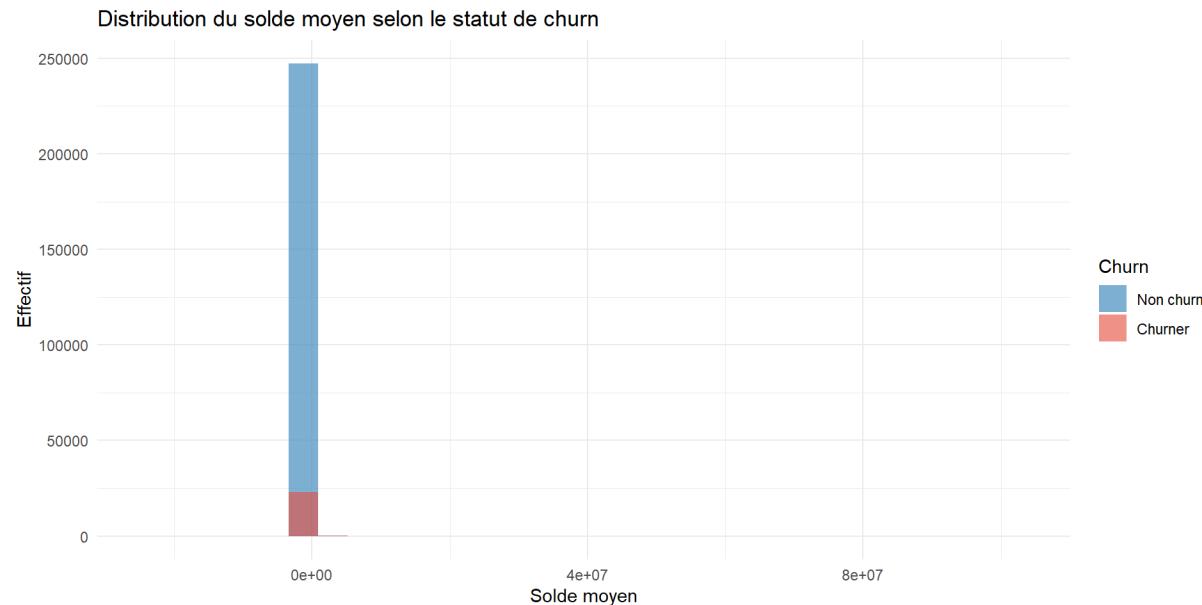
► Afficher le code



La distribution de l'âge met en évidence une clientèle majoritairement concentrée entre 30 et 60 ans, caractéristique d'une population bancaire active.

Histogramme clé : solde moyen selon churn

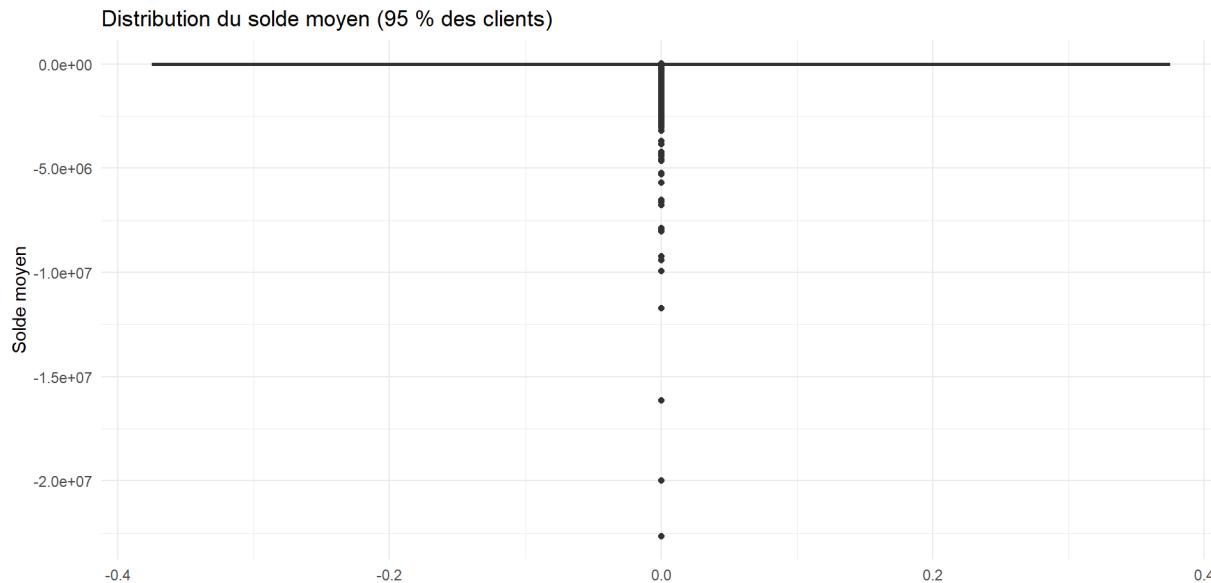
► Afficher le code



Les deux distributions présentent une forte concentration autour de soldes proches de zéro, suggérant des profils financiers globalement comparables entre churners et non-churners.

Boxplot : solde moyen (95 %)

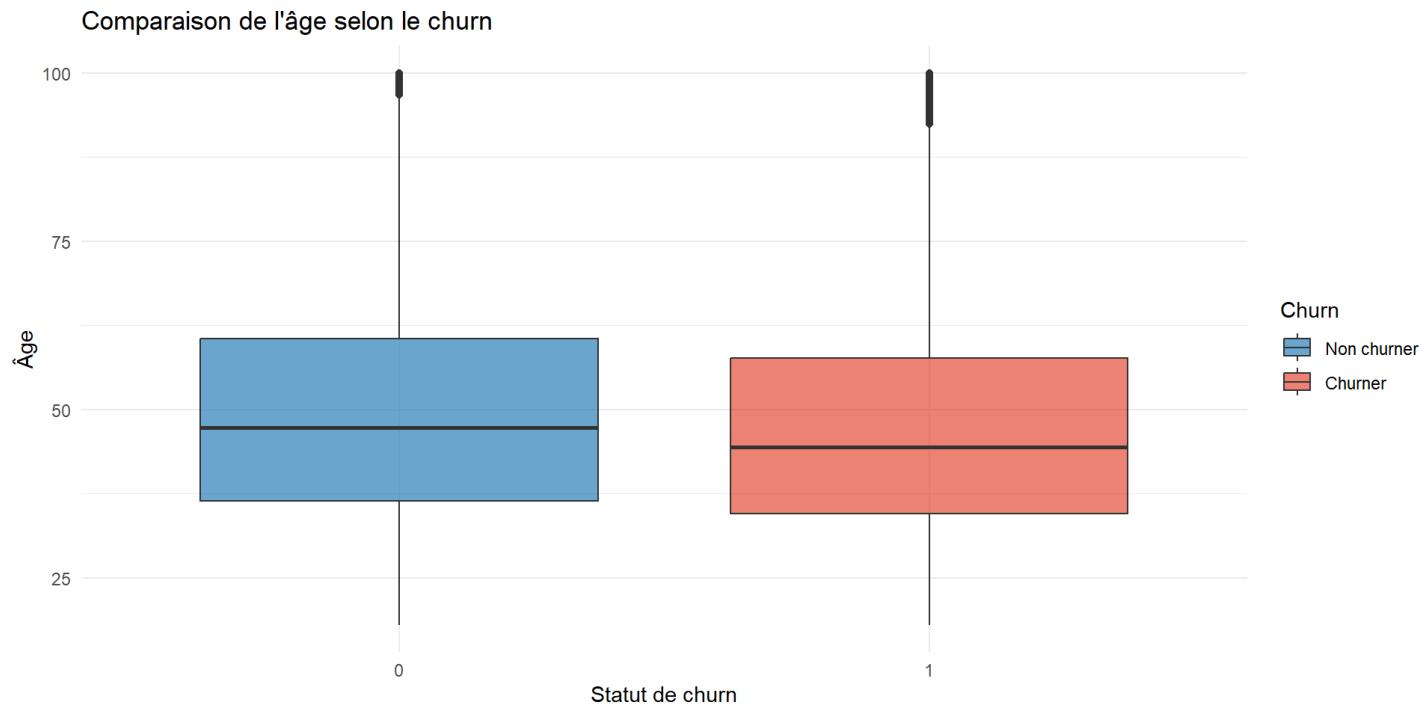
► Afficher le code



La limitation de l'analyse aux 95 % des clients permet de neutraliser l'effet des valeurs extrêmes et met en évidence une dispersion des soldes moyens globalement comparable entre les groupes.

Comparaison âge selon churn

► Afficher le code



Les churners apparaissent légèrement plus jeunes en médiane, bien que les distributions se recouvrent largement, ce qui traduit une différence modérée.

Validation statistique

- Normalité non vérifiée
- Utilisation de tests **non paramétriques (Wilcoxon)**
- Objectif : **confirmer les différences observées, sans prédition**

Résultat des comparaisons

- ▶ Afficher le code

```
# A tibble: 2 × 5
  churn age_mean seniority_mean balance_mean n_clients
  <dbl>    <dbl>        <dbl>      <dbl>      <int>
1     0      48.9       10.9      8135.    248959
2     1      47.0       11.0      9615.    114354
```

Conclusion

- Des différences existent entre churners et non-churners
- L'âge, l'ancienneté et le solde moyen présentent des différences modérées entre les groupes.

Perspectives

- Intégration de variables comportementales
- Analyse temporelle de l'évolution des soldes
- Extension possible vers une analyse descriptive longitudinale ou segmentée.