

# PREDICTION DU DESABONNEMENT DES CLIENTS BANCAIRES



## GROUPE I

Professeur:  
**FLORIAN SAWADOGO**

◆ **OUEDRAOGO WENDTOIN ISSAKA**  
◆ **SOME FIARMA LANDRY**

Le désabonnement des clients représente un défi majeur pour les banques, impactant directement leur rentabilité. L'objectif de ce projet est de développer un modèle prédictif pour identifier les clients à risque de désabonnement.

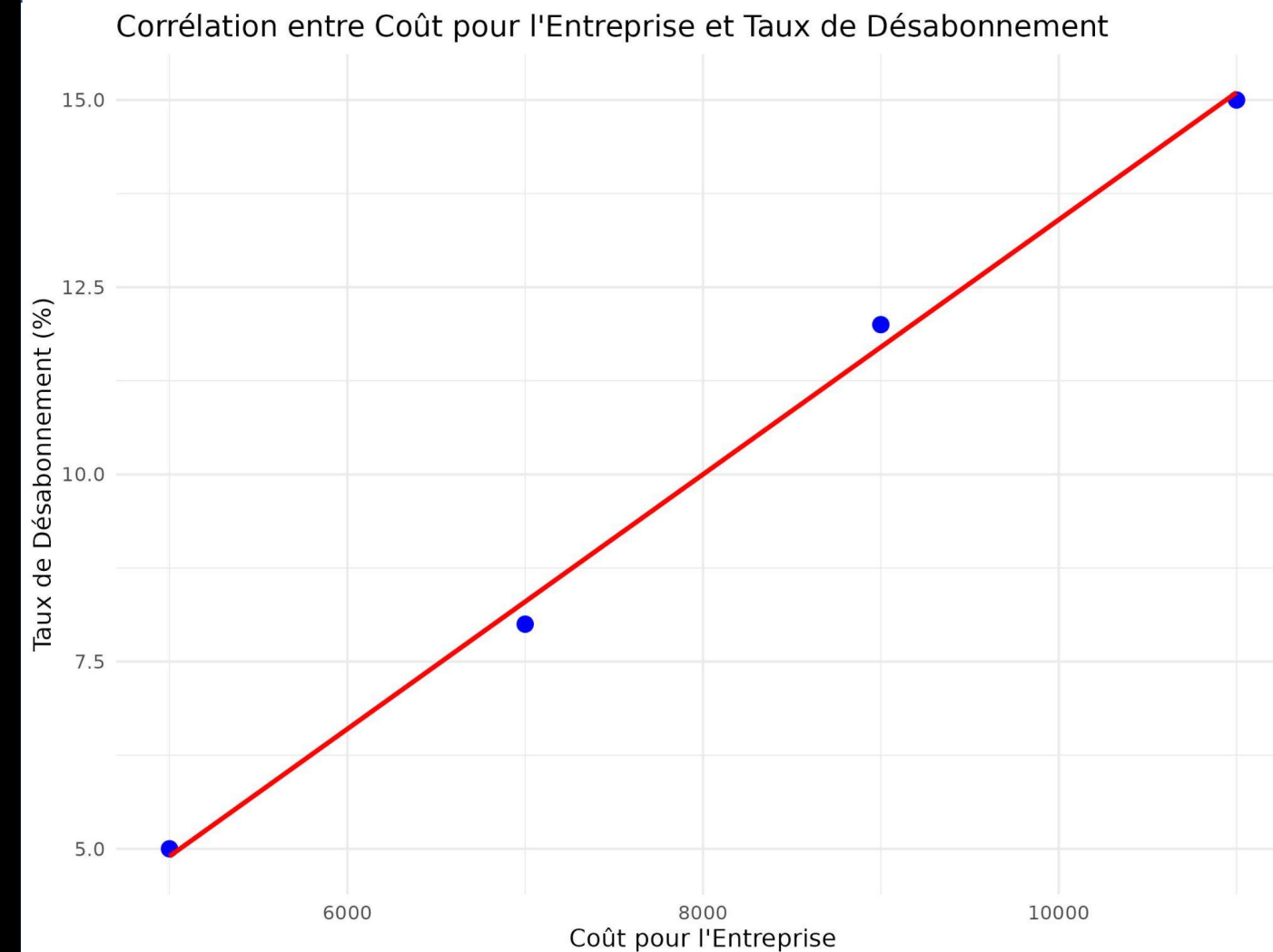
# PROBLEMATIQUE

---

Pourquoi la prédiction du désabonnement est-elle cruciale ?  
Des études ont montré qu'acquérir un nouveau client peut coûter jusqu'à cinq fois plus cher que de fidéliser un client existant (Bain & Company).  
Ce projet vise à fournir un outil pour anticiper les départs et agir en conséquence.

---

.....





# BASE DE DONNEES

❑ Source des données : [Kaagle](#)

❑ Contenu :

La base de données sur laquelle on va travailler, contient 5 000 enregistrements clients et 18 Variables, pas de données manquantes.

- ❖ Variables numeriques(14) : [RowNumber](#), [CustomerId](#), [CreditScore](#), [Age](#), [Tenure](#), [Balance](#), [NumOfProducts](#), [HasCrCard](#), [IsActiveMember](#), [EstimatedSalary](#), [Exited](#), [Complain](#), [Satisfaction.Score](#), [Point.Earned](#)
- ❖ Variables catégorielles(4) : [Surname](#), [Geography](#), [Gender](#), [Card.Type](#)

Description: df [5,000 × 14]

RowNum...	CustomerId	CreditScore	Age	Tenure	Balance	NumOfProducts
<int>	<int>	<int>	<int>	<int>	<dbl>	<int>
6603	15580872	761	38	1	120530.13	2
4061	15593250	640	29	4	0.00	2
2013	15780124	841	74	9	108131.53	1
161	15692132	717	22	6	101060.25	1
8058	15620836	816	34	2	108410.87	2
8504	15794101	559	48	2	0.00	2
4062	15605333	529	31	6	0.00	1
7038	15789611	568	46	8	150836.92	1
1244	15593331	693	25	6	146580.69	1
8667	15785920	687	35	1	125141.24	2

1-10 of 5,000 rows | 1-7 of 14 columns

Previous 1 2 3 4 5 6 ... 100 Next

Description: df [6 × 18]

Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
<dbl>	<int>	<int>	<int>	<dbl>	<int>
120530.1	2	1	0	109394.62	0
0.0	2	1	0	44904.26	0
108131.5	1	0	1	60830.38	0
101060.2	1	0	1	84699.56	0
108410.9	2	1	0	102908.91	0
0.0	2	0	1	137961.41	0

6 rows | 10-15 of 18 columns

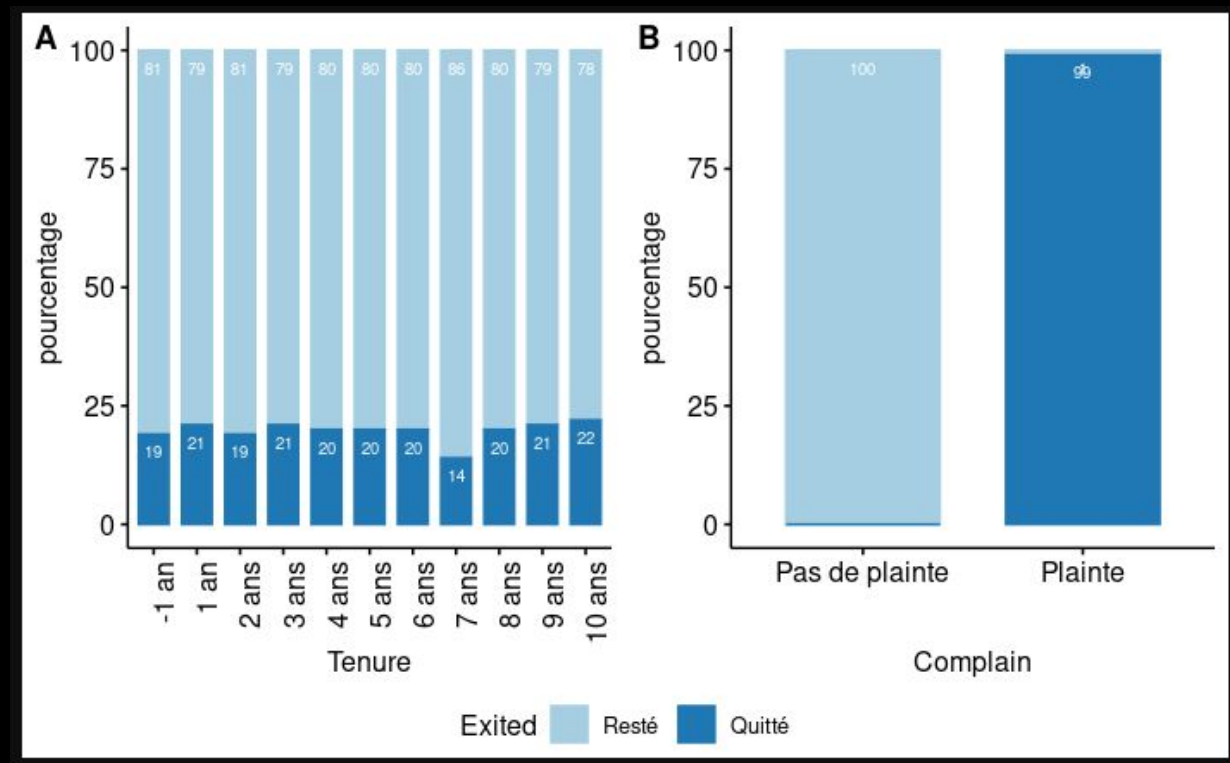
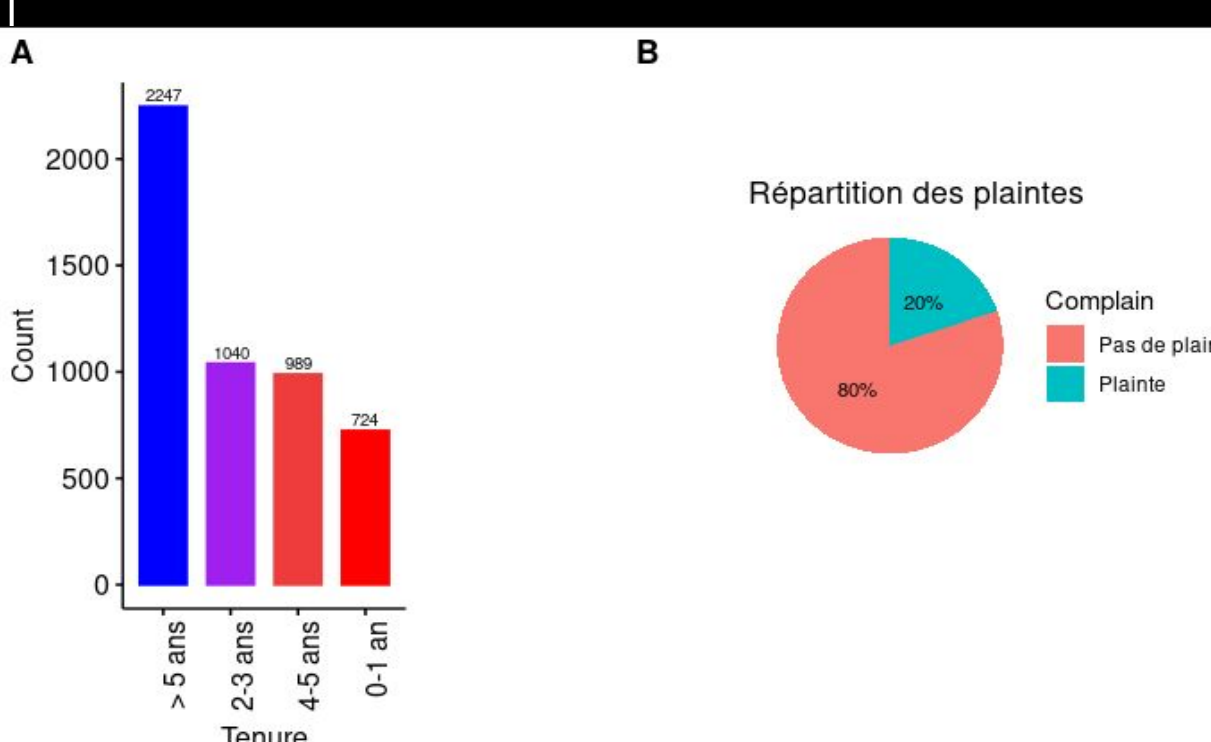
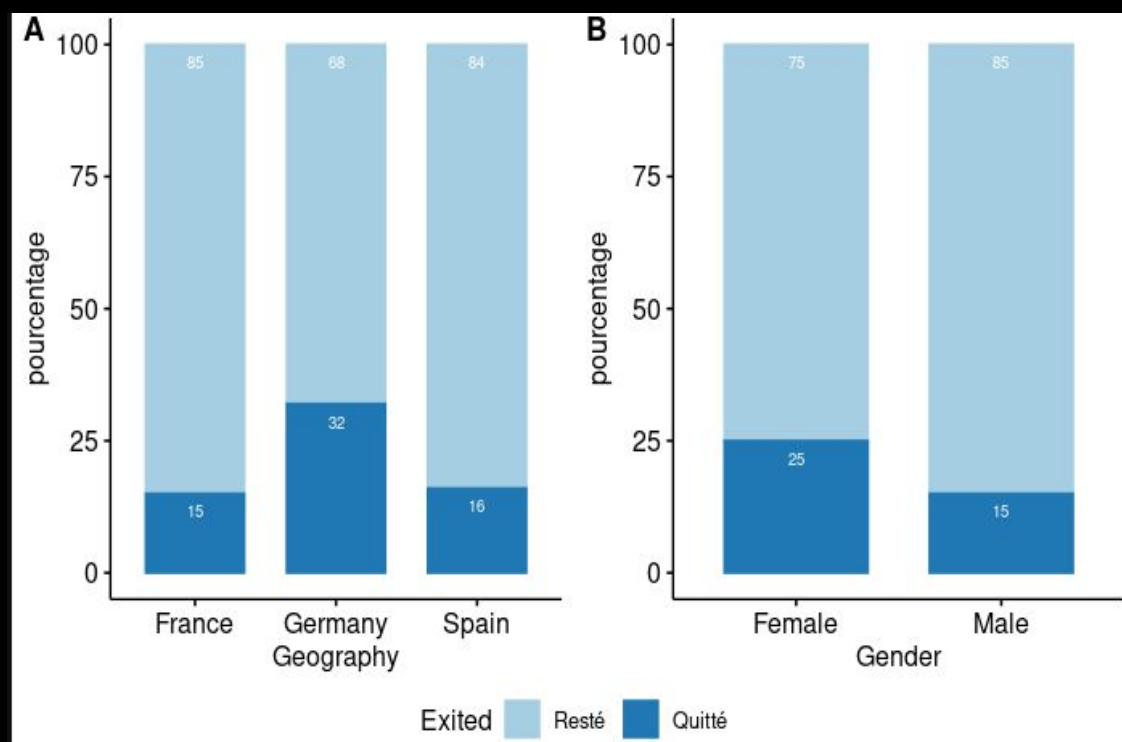
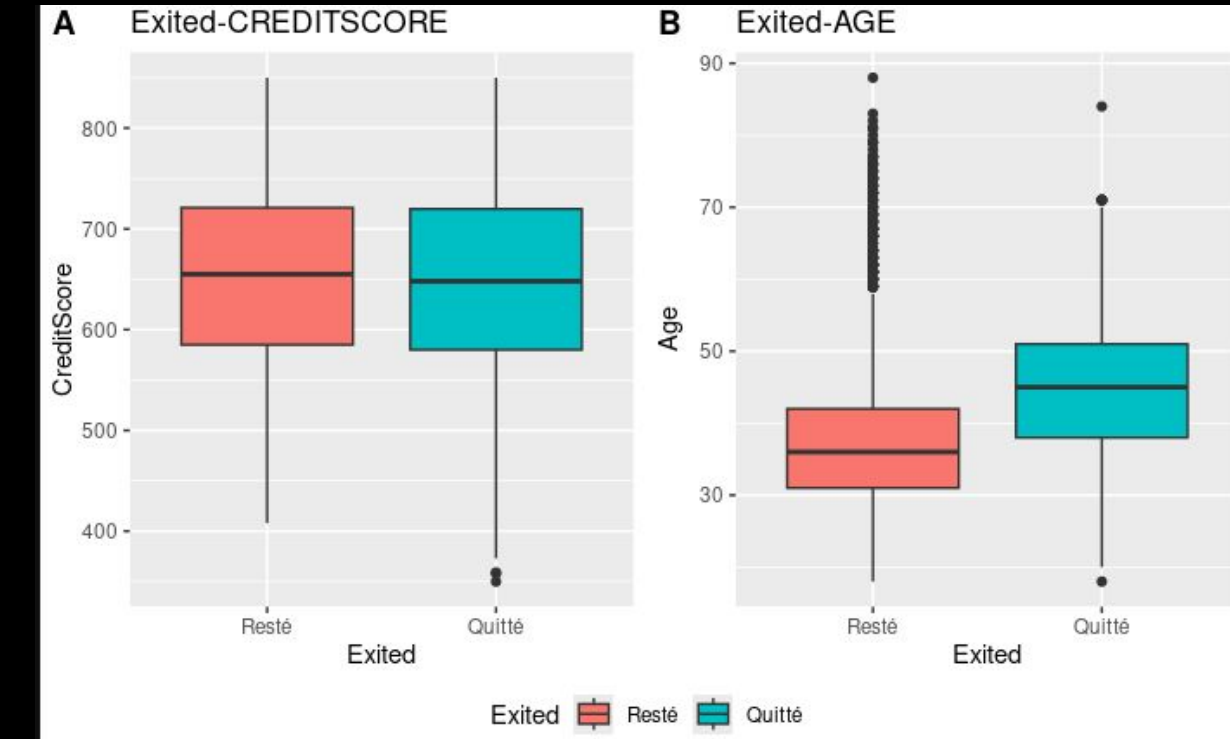
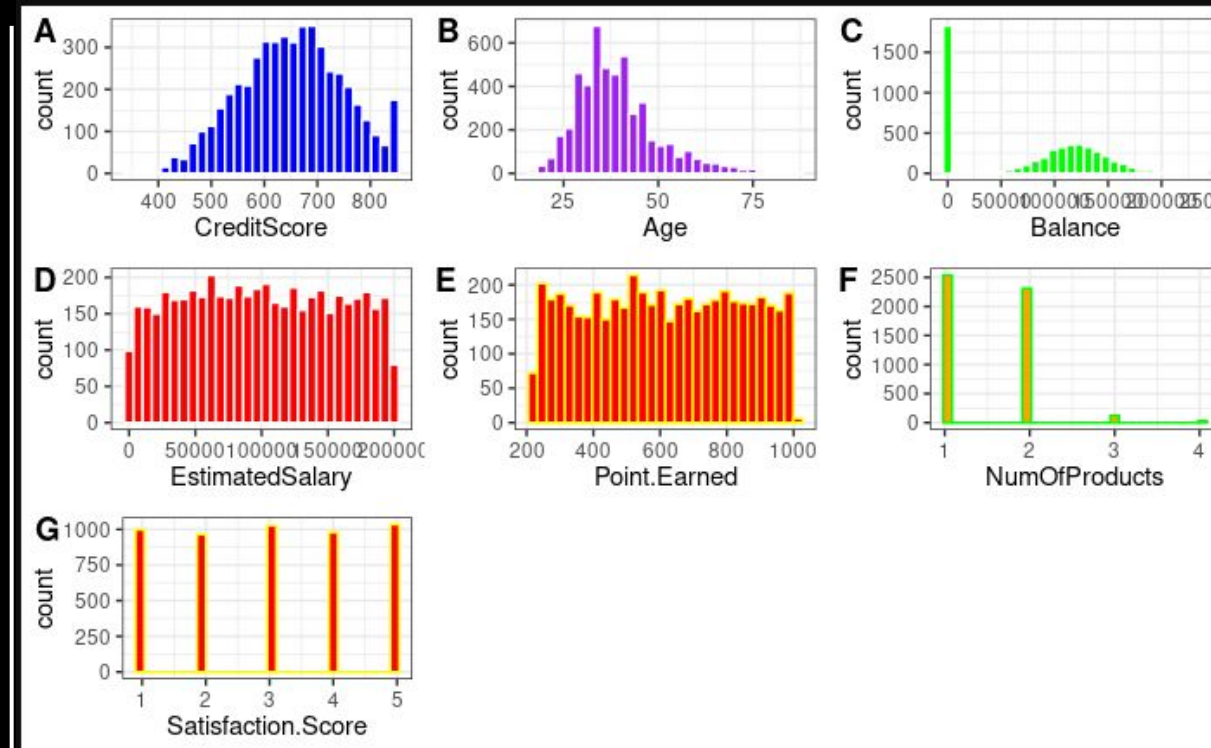
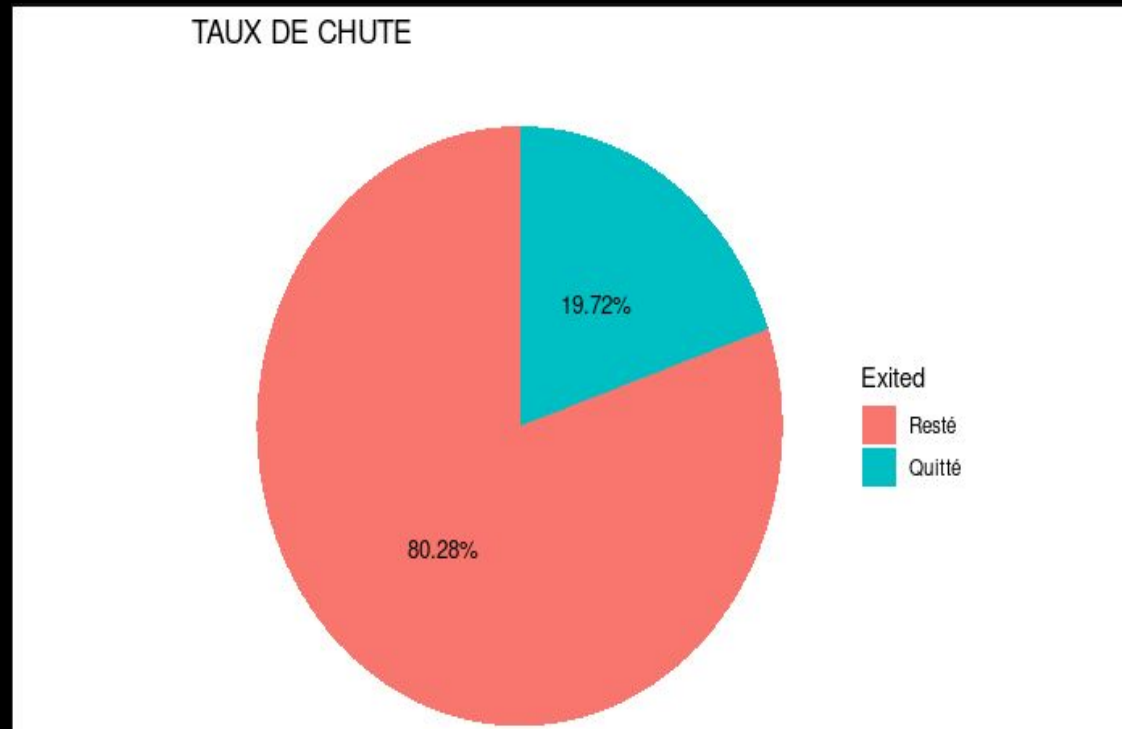


Description: df [6 × 18]

EstimatedSalary	Exited	Complain	Satisfaction.Score	Card.Type	Point.Earned
<dbl>	<int>	<int>	<int>	<chr>	<int>
109394.62	0	0	2	GOLD	815
44904.26	0	0	3	GOLD	758
60830.38	0	0	5	SILVER	698
84699.56	0	0	2	GOLD	666
102908.91	0	0	4	PLATINUM	555
137961.41	0	0	3	GOLD	640

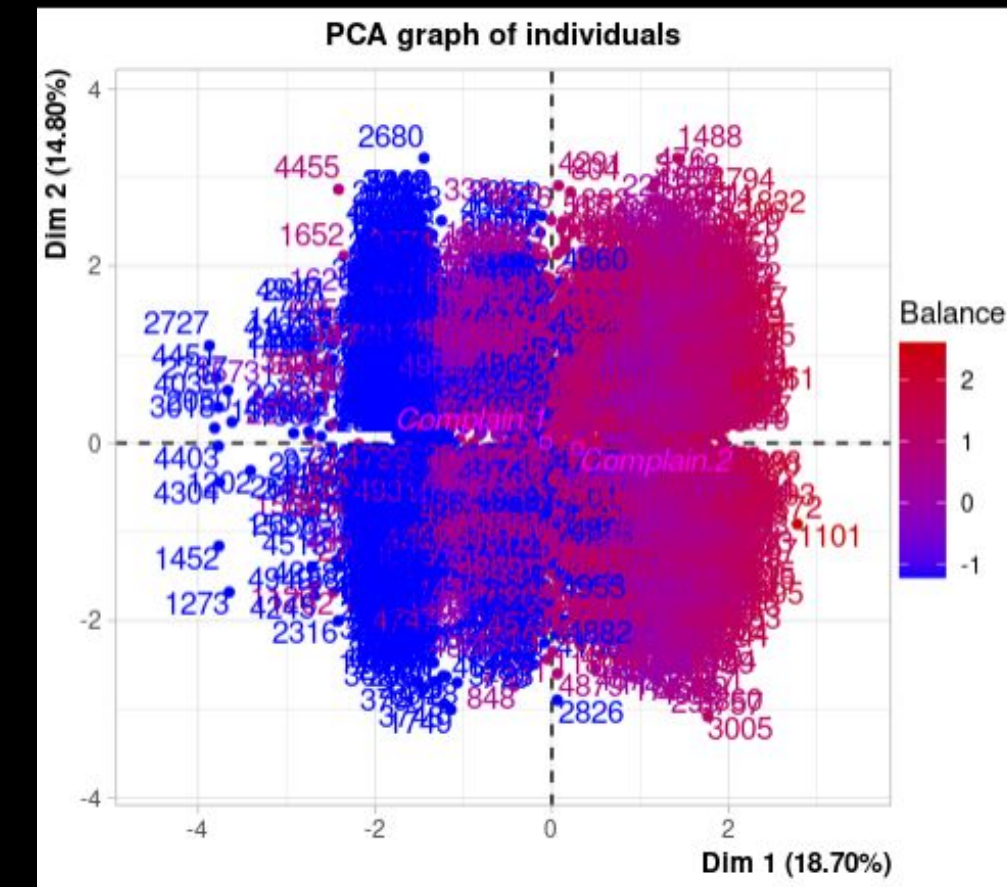
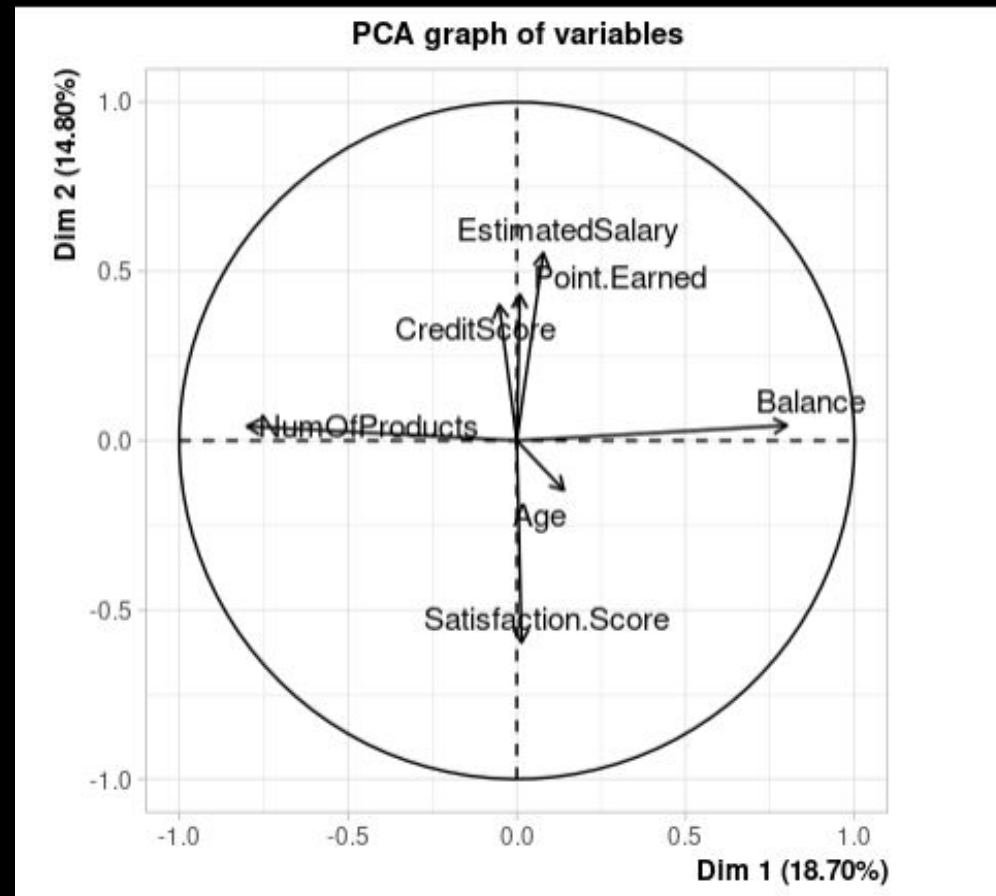
6 rows | 14-19 of 18 columns

# ANALYSE DESCRIPTIVE





# ANALYSE FACTORIELLE : ACP



- ❖ Le graphe des variables montre que la dimension 1 est principalement influencée par le solde bancaire (Balance), tandis que la dimension 2 est plus liée à la satisfaction et à l'âge des clients.
- ❖ Les autres variables ont une influence plus faible sur ces dimensions.
- ❖ Ces deux dimensions expliquent ensemble 33,5% de la variance totale.

- ❖ Le graphe montre la répartition des clients sur les deux premières dimensions de l'ACP, expliquant ensemble 33,59% de la variance totale.
- ❖ La dimension 1 semble principalement liée au solde bancaire (Balance), avec les clients à solde élevé regroupés à droite et ceux à solde faible à gauche.
- ❖ La variable Complain (Oui/Non) n'influence pas fortement la séparation des individus sur ces dimensions

# MODÉLISATION ET RÉSULTATS

- ❑ Nous avons développé un modèle de régression logistique pour prédire si un client quittera la banque ou non. Après avoir ajusté le modèle sur l'ensemble de données d'entraînement(70%) et évalué ses performances sur un ensemble de test(30%), voici les résultats clés.
- ❑ **Prédictions** : Le modèle a prédit les probabilités de départ des clients avec une grande précision
- ❑ **Exactitude** : Le modèle a atteint une exactitude (accuracy) de 99.8%
- ❑ **Matrice de confusion(clients =1499; 0:1204; 1:295)** :
  - ❑ 1202 vrais négatifs (clients restés correctement identifiés).
  - ❑ 295 vrais positifs (clients quittés correctement identifiés).
  - ❑ 2 faux positifs (clients prédits comme quittant alors qu'ils sont restés).
  - ❑ 0 faux négatif (client prédit comme resté alors qu'il a quitté).



```
1745      2277      4240      2942      1993      4774
0.0001241055 0.0004292380 0.0002271780 0.9921250496 0.0001242309 0.9999664393
[1] "Logistic Regression Accuracy 0.99866577718479"
[1] "Confusion Matrix for Logistic Regression"

      FALSE TRUE
0    1202    2
1      0   295
```



# CONCLUSION

- ❖ **Résultats Clés** : Les modèles d'apprentissage automatique permettent de prédire efficacement le désabonnement des clients. Les caractéristiques les plus influentes identifiées sont la Plainte, l'Age, la Géographie, selon que le client soit un membre actif ou pas de la banque.
- ❖ **Impact** : Ces prédictions offrent à la banque la possibilité de cibler ses efforts de rétention, réduisant ainsi les coûts d'acquisition de nouveaux clients.

Merci pour votre attention!

- ❖ APPORTS
- ❖ QUESTIONS?

Lien GitHub Script : [https://github.com/wendtoinissaka/UVBF\\_L3\\_ANALYSE\\_DE\\_DONNEES/tree/main/Data\\_mining\\_R](https://github.com/wendtoinissaka/UVBF_L3_ANALYSE_DE_DONNEES/tree/main/Data_mining_R)