



Data Science Bootcamp

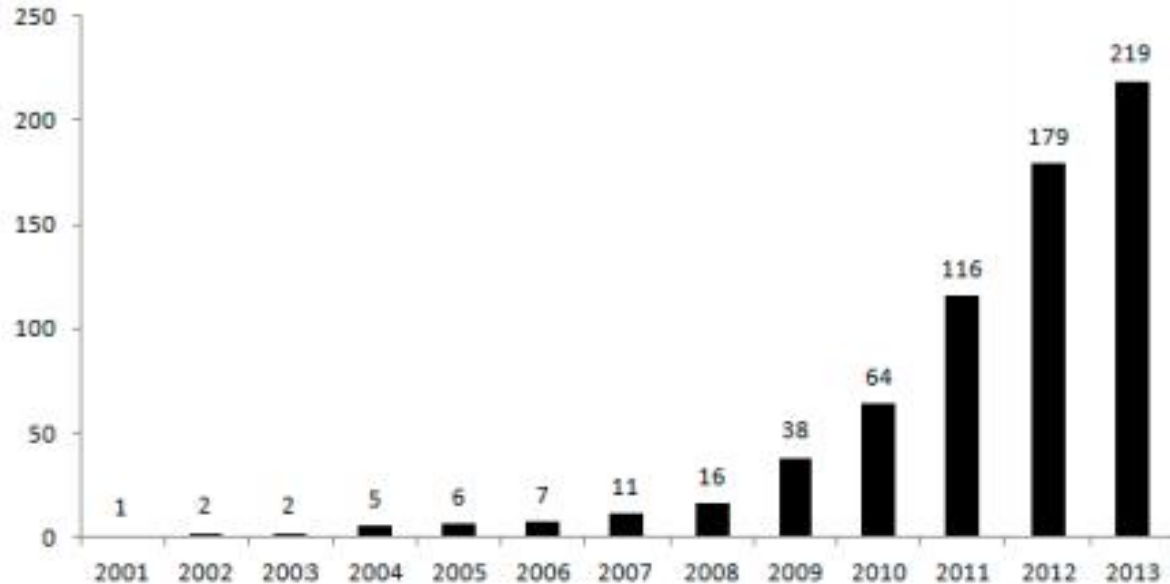
Regis Obiang



***Prédire une transaction financière
frauduleuse via mobile money***

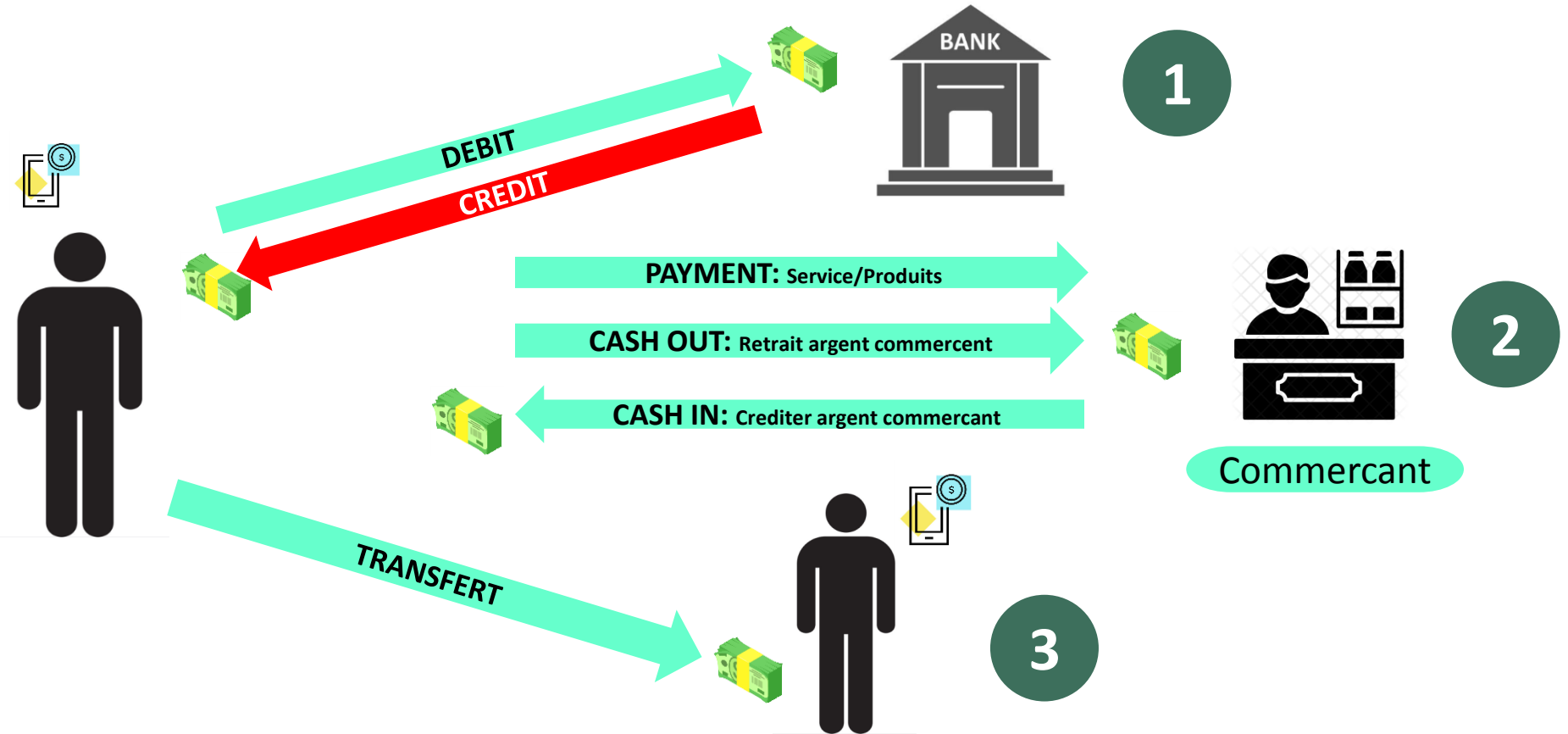
Contexte

Live Deployments of Mobile Money, 2001-2013

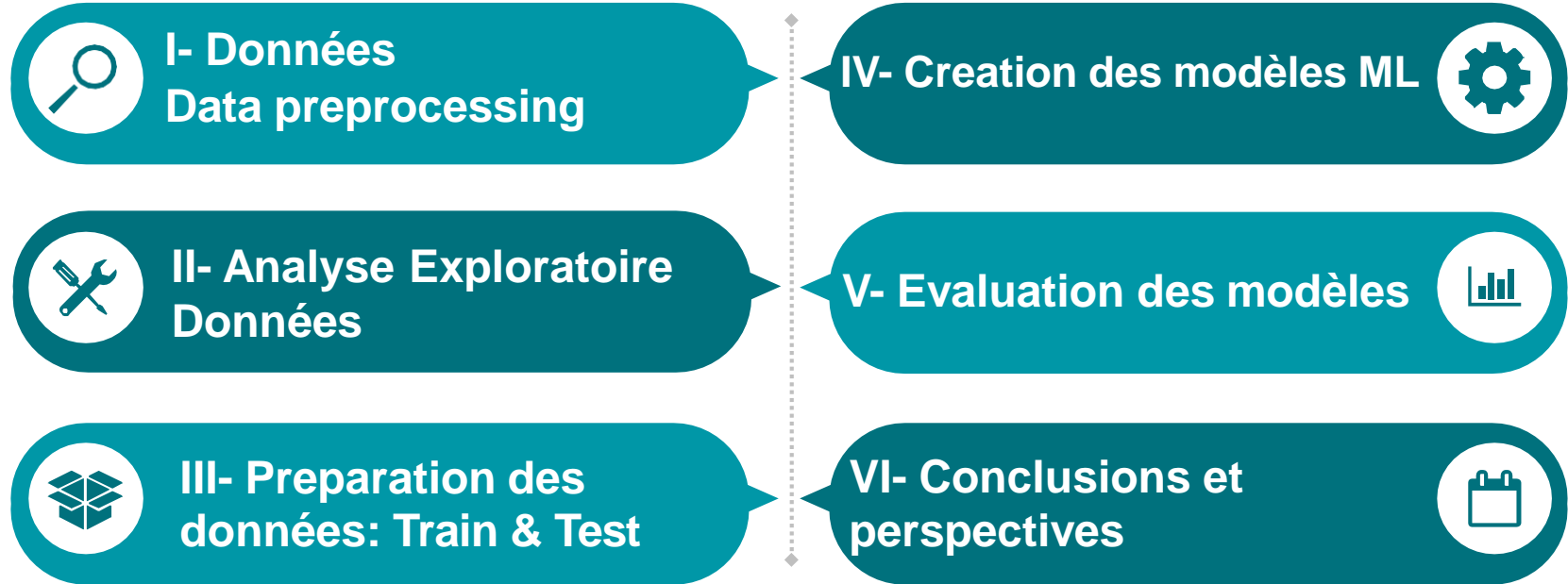


Source: Groupe Speciale Mobile Association (2014).

Schéma des flux d'un service mobile money



Phases



Données & Data preprocessing

- Source: <https://www.kaggle.com/ntnu-testimon/paysim1>
 - Données issues d'un simulateur de transactions des services de paiements mobiles
- ~6,4 Millions entrées
- 11 caractéristiques (variables)
- Pas de données manquantes



Variables clés

Variable à prédire

- ☐ Transaction frauduleuse

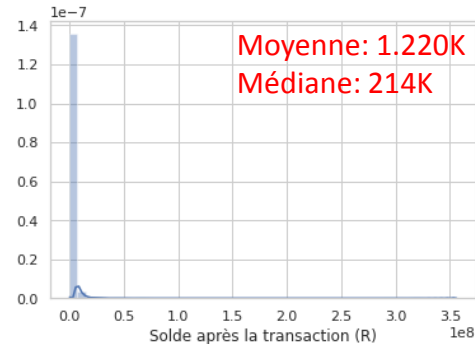
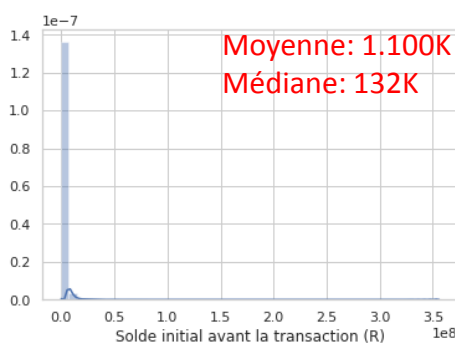
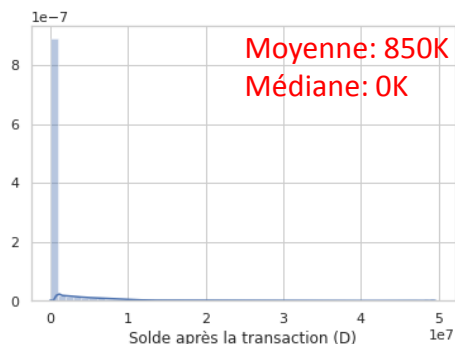
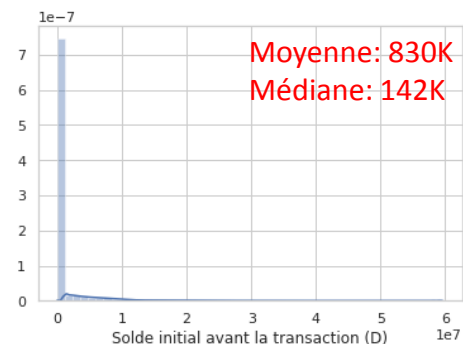
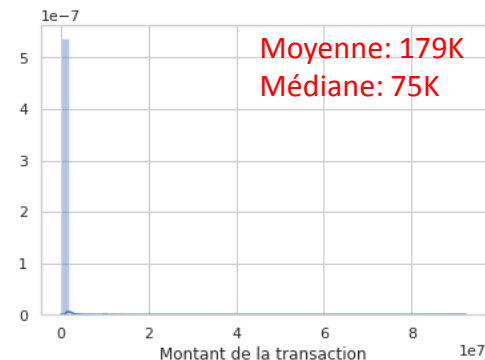
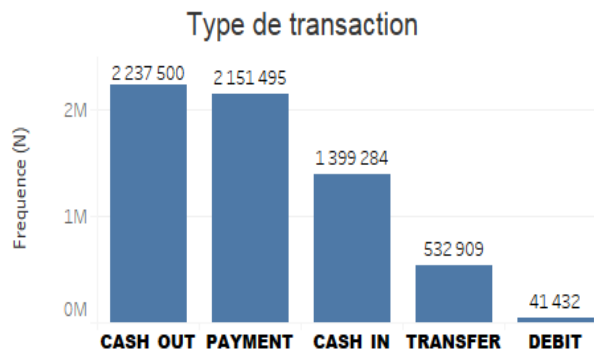
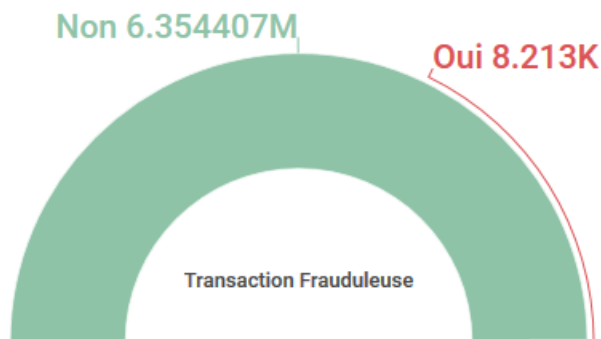
Covariables

- ☐ Type de transaction
- ☐ Montant de la transaction

Covariables

- ☐ Solde initial avant la transaction (D)
- ☐ Solde après la transaction (D)
- ☐ Solde initial avant la transaction (R)
- ☐ Solde après la transaction (R)

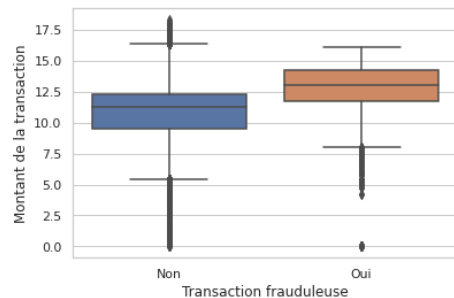
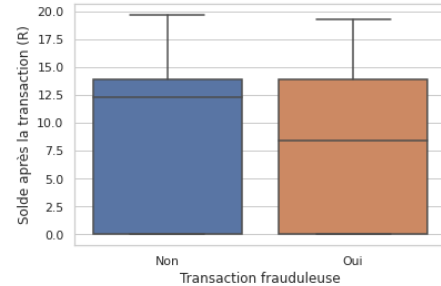
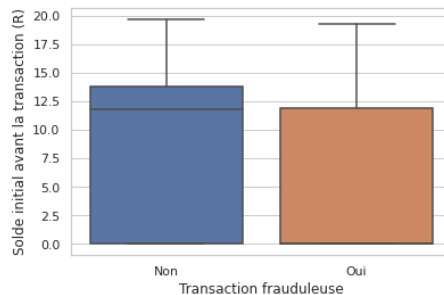
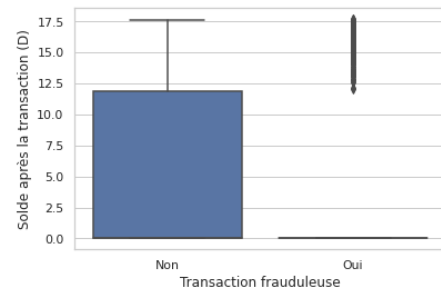
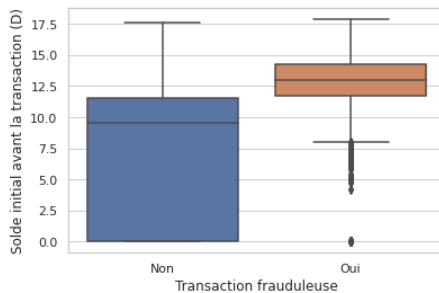
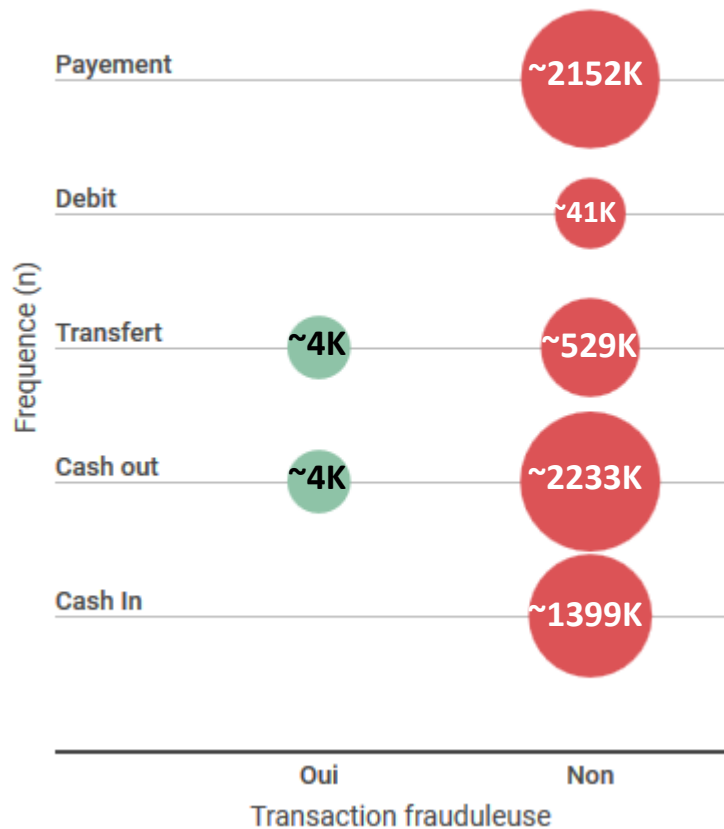
Analyse Exploratoire Données (1)



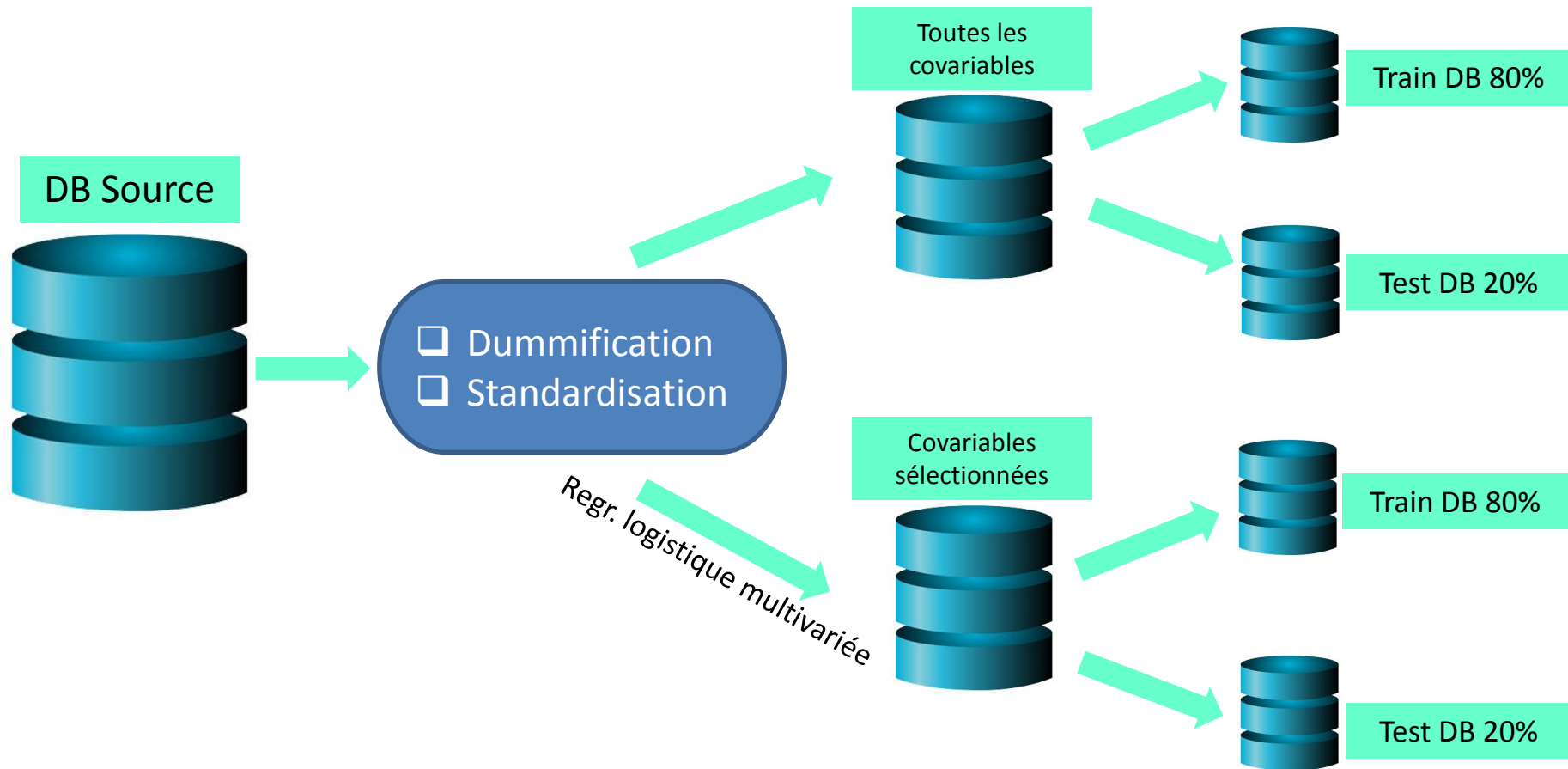
☐ Groupe déséquilibrée (Trans. Frauduleuse)

☐ Données asymétriques centrées vers la droite

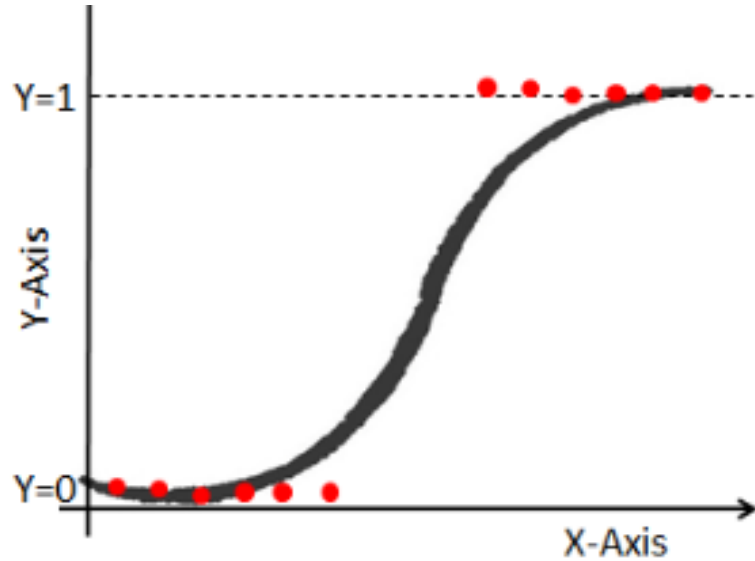
Analyse Exploratoire Données (2)



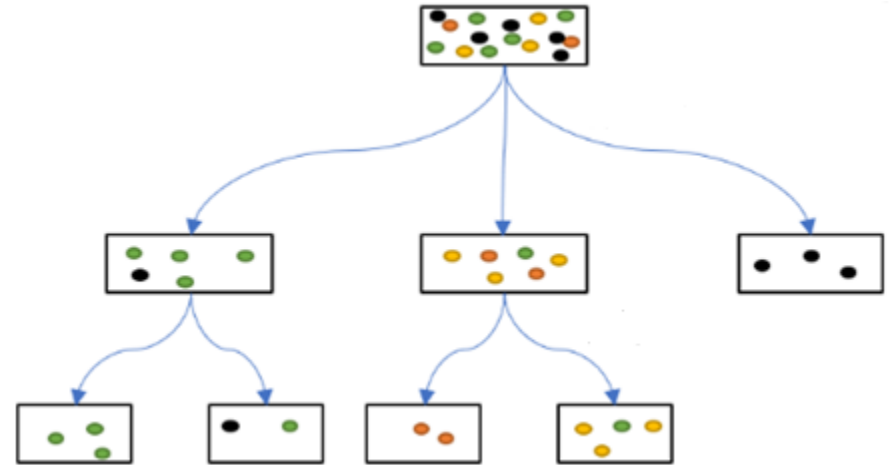
Préparation des données: Train et Test



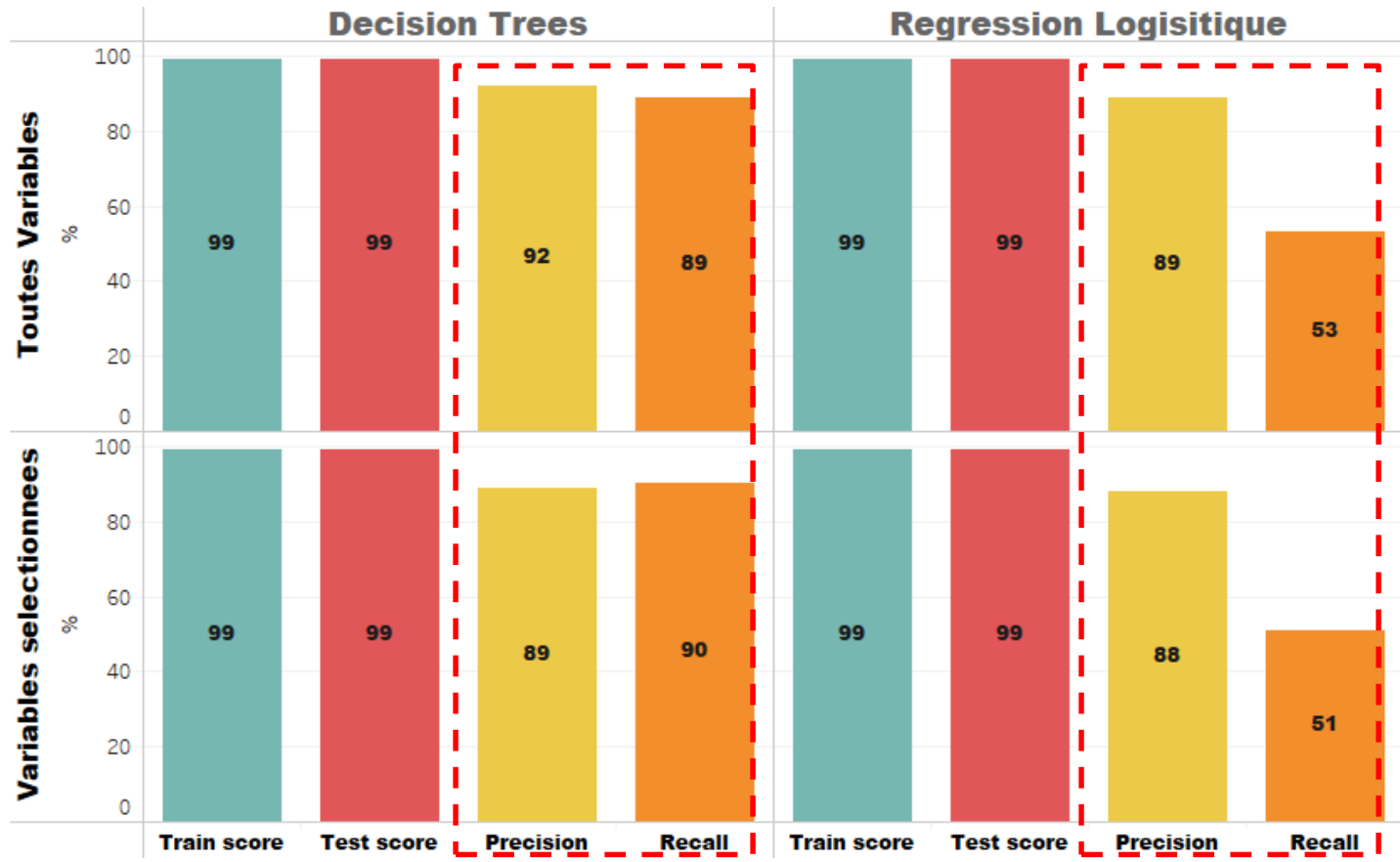
Regression Logistique



Decision trees



Évaluation des modèles



$$\text{Précision} = \frac{VP}{VP + FP}$$

$$\text{Rappel} = \frac{VP}{VP + FN}$$

Conclusions et perspectives

- ❑ Les Scores des data (Train et Test) semblent excessivement élevés: **Overfitting**
- ❑ Modele ML avec Decision Trees semble donner des meilleurs résultats comparativement à la régression logistique
- ❑ Envisager plus tard, un modèle de ML plus robuste comme XgBoost



Data Science Bootcamp

Des questions ?

