

DataScientest

# Projet - Campagne de Marketing d'une banque

Rendu 1

Farida CHETTIH  
15/01/2026

## Table des matières

1. Introduction au projet .....	2
1.1 Contexte professionnel et métier .....	2
1.2. Point de vue technique : .....	2
1.3 Point de vue économique : .....	2
2. Objectifs du projet .....	2
2.1 Objectifs principaux .....	2
2.2 Niveau d'expertise et interactions métier.....	2
3. Compréhension et manipulation des données .....	2
3.1 Cadre et données utilisées .....	2
3.2 Volumétrie .....	3
4. Pertinence des variables .....	3
4.1 Variables pertinentes .....	3
4.2 Particularités.....	3
5. Pre-processing et Feature Engineering.....	3
5.1 Nettoyage des données .....	3
5.2 Valeurs extrêmes : .....	4
6. Visualisations et Statistiques .....	5
6.1 Distributions .....	5
6.2 Relations entre variables : .....	5
6.3 Statistiques descriptives .....	6
⇒ Ce qui a été réalisé : .....	6
⇒ Ce qui reste à faire .....	6

# 1. Introduction au projet

## 1.1 Contexte professionnel et métier

Analyser les données d'une campagne marketing bancaire afin d'identifier des leviers d'amélioration et comprendre les facteurs associés à la souscription (variable cible)

## 1.2. Point de vue technique :

Le projet demande des compétences techniques d'analyse de données, nettoyage, pré-traitement et visualisation à l'aide de Python et des bibliothèques Pandas, Matplotlib et Seaborn.

## 1.3 Point de vue économique :

Dans un contexte de forte concurrence bancaire, l'optimisation des campagnes marketing permet de mieux cibler les clients, de réduire les coûts de prospection et d'améliorer le taux de souscription.

# 2. Objectifs du projet

## 2.1 Objectifs principaux

L'objectif de ce projet est d'analyser et préparer un nouveau jeu de données basé sur les anciennes campagnes marketing bancaires afin d'identifier des pistes d'amélioration pour les futures actions commerciales.

Afin de mieux répondre à l'objectifs du projet nous devons :

- Comprendre la structure et le contenu du jeu de données
- Identifier les variables pertinentes pour l'analyse des campagnes marketing
- Nettoyer et préparer les données afin de les rendre exploitables
- Mettre en évidence des relations entre variables à l'aide de visualisations
- Poser les bases d'analyses futures visant à améliorer l'efficacité des campagnes

## 2.2 Niveau d'expertise et interactions métier

Le projet est mené dans un cadre de montée en compétence en data analysis.

Aucune interaction directe avec des experts métiers n'a été réalisée. Toutefois, la problématique est largement documentée dans le secteur bancaire et marketing.

# 3. Compréhension et manipulation des données

## 3.1 Cadre et données utilisées

Le jeu de données que nous utilisons provient d'une campagnes marketing téléphonique menées par une banque Le fichier 'bank.csv' (Campagne marketing d'une banque)

=> Les données sont disponibles librement via des plateformes publiques (Kaggle).

## 3.2 Volumétrie

Le jeu de données contient :

- 11 162 lignes (clients contactés),
- 17 colonnes (variables explicatives et variable cible).

Il comprend :

- 7 variables quantitatives (âge, durée des appels, balance, nombre de contacts...),
- 10 variables qualitatives (profession, type de contact, crédits...).

Le jeu de données est composé majoritairement de variables qualitatives, ce qui nécessite une analyse approfondie des modalités.

## 4. Pertinence des variables

### 4.1 Variables pertinentes

Les variables jugées les plus pertinentes sont :

- Variables de campagne : duration, campaign, pdays, previous,
- Variables clients: age, job, education, marital,
- Variables financières : balance, housing, loan,
- Variable cible : deposit.

### 4.2 Particularités

Plusieurs particularités ont été identifiées :

- Présence fréquente de la modalité unknown
- Valeurs spécifiques comme pdays = -1
- Distributions asymétriques et valeurs extrêmes

Ces éléments constituent des limites analytiques à prendre en compte.

## 5. Pre-processing et Feature Engineering

### 5.1 Nettoyage des données

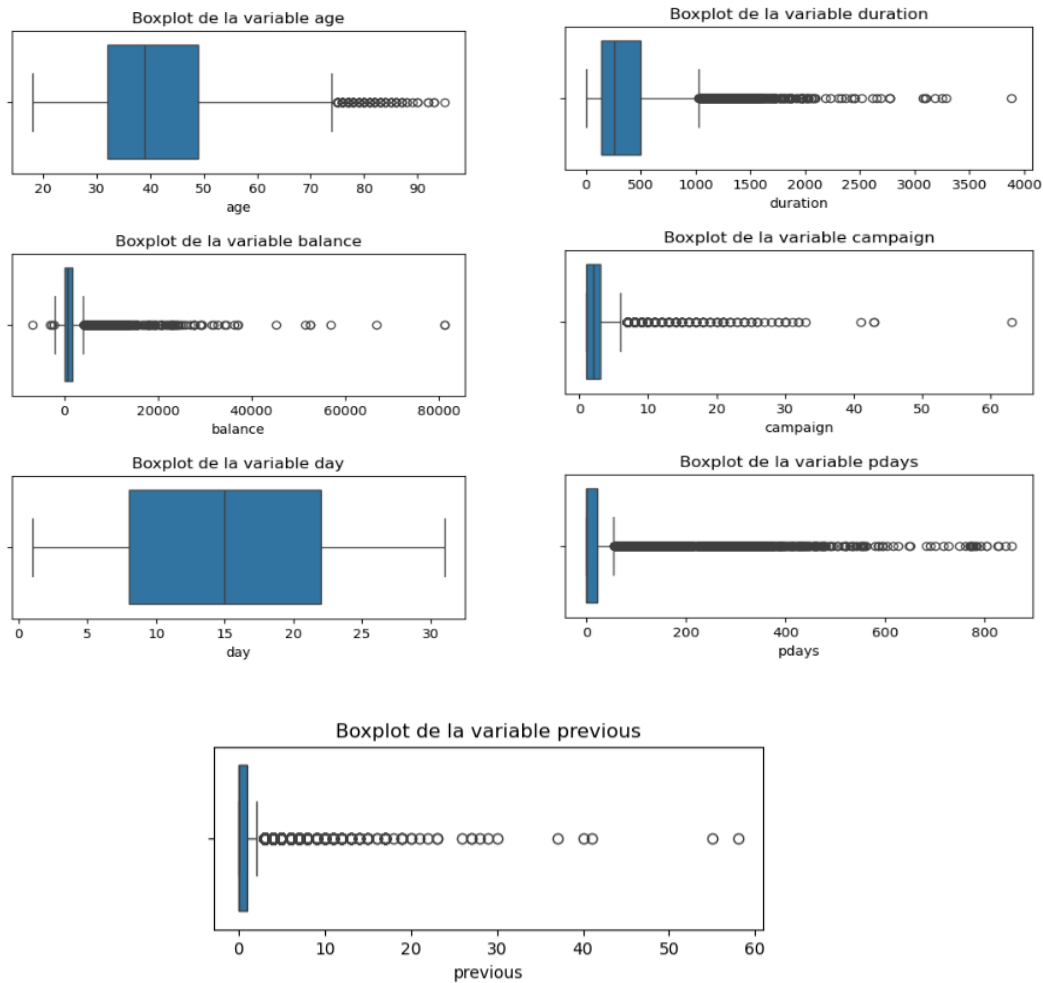
Les étapes suivantes ont été réalisées :

- Vérification des valeurs manquantes (absence de NaN significatifs)
- Détection des doublons
- Analyse spécifique des modalités unknown
- Séparation des variables quantitatifs et qualitatives

Les modalités "unknown" ont été conservées et renommées afin de préserver l'information liée à l'absence de renseignement. Cette stratégie permet d'éviter toute perte d'information car celle-ci pouvant avoir un impact sur le comportement des clients lors des campagnes marketing

## 5.2 Valeurs extrêmes :

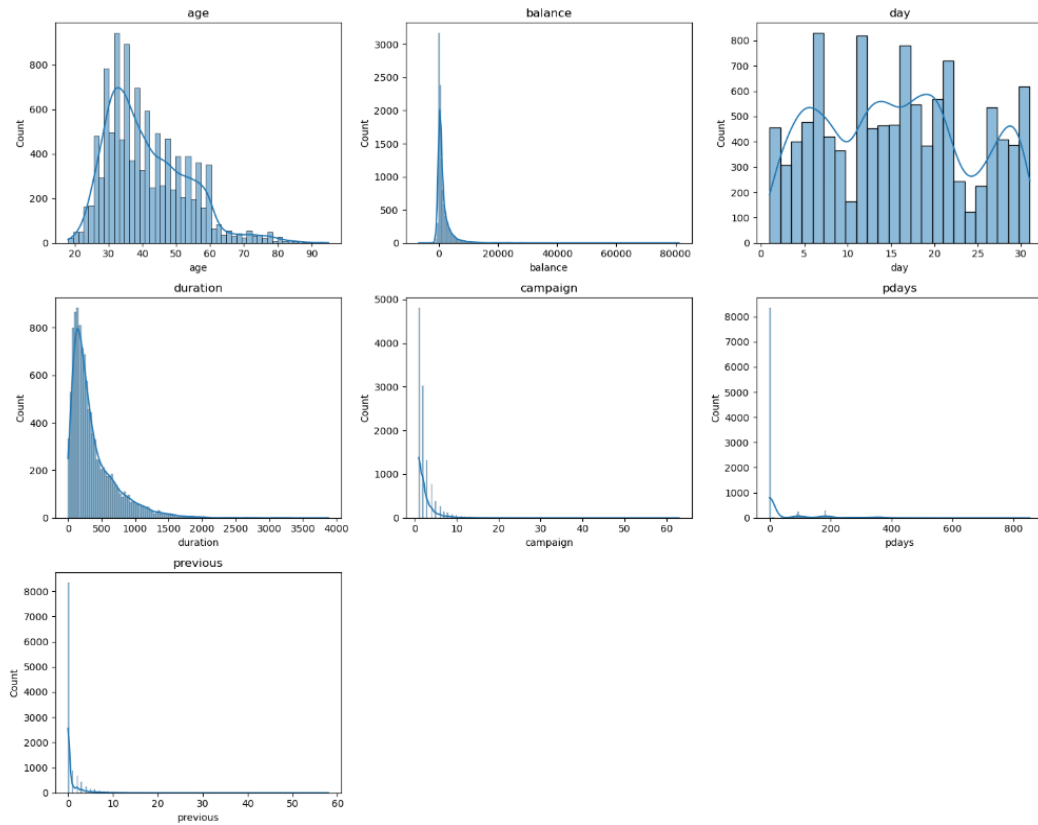
Les boxplots ont mis en évidence la présence de valeurs extrêmes sur certaines variables numériques, notamment âge, balance et duration. Ces valeurs extrêmes observées correspondent à des profils clients atypiques mais réalistes. Elles ont donc été conservées pour l'analyse exploratoire.



## 6. Visualisations et Statistiques

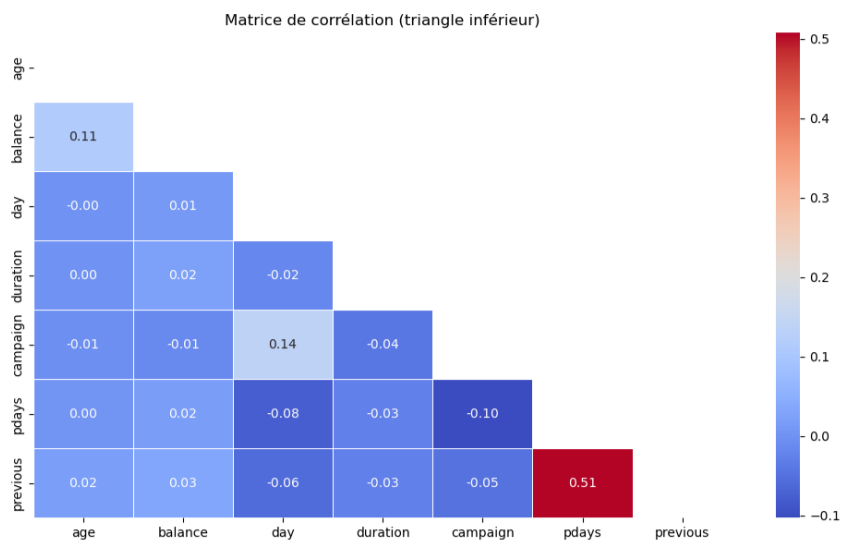
### 6.1 Distributions

Les histogrammes montrent des distributions souvent asymétriques, notamment pour la balance et la durée des appels.



### 6.2 Relations entre variables :

L'analyse des corrélations met en évidence des relations entre certaines variables numériques et une influence notable des variables de campagne sur la souscription.



## 6.3 Statistiques descriptives

Les statistiques descriptives (moyenne, médiane, quartiles, écart-type) confirment les tendances observées graphiquement et permettent de quantifier la dispersion des données.

	age	balance	day	duration	campaign	pdays	previous
<b>count</b>	11162.000000	11162.000000	11162.000000	11162.000000	11162.000000	11162.000000	11162.000000
<b>mean</b>	41.231948	1528.538524	15.658036	371.993818	2.508421	51.330407	0.832557
<b>std</b>	11.913369	3225.413326	8.420740	347.128386	2.722077	108.758282	2.292007
<b>min</b>	18.000000	-6847.000000	1.000000	2.000000	1.000000	-1.000000	0.000000
<b>25%</b>	32.000000	122.000000	8.000000	138.000000	1.000000	-1.000000	0.000000
<b>50%</b>	39.000000	550.000000	15.000000	255.000000	2.000000	-1.000000	0.000000
<b>75%</b>	49.000000	1708.000000	22.000000	496.000000	3.000000	20.750000	1.000000
<b>max</b>	95.000000	81204.000000	31.000000	3881.000000	63.000000	854.000000	58.000000

### ⇒ Ce qui a été réalisé :

Un travail de visualisation et de pre-processing qui a permis :

- Une compréhension approfondie du jeu de données
- L'identification et le nettoyage des variables clés
- La préparation d'un jeu de données exploitable pour la modélisation

### ⇒ Ce qui reste à faire

Les prochaines étapes consisteront à développer des modèles prédictifs et à formuler des recommandations opérationnelles pour améliorer l'efficacité des campagnes marketing via l'outil BI (Business Intelligence).