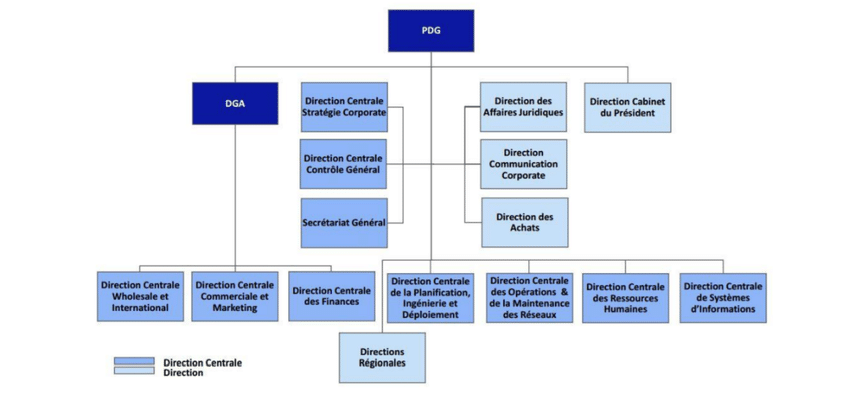
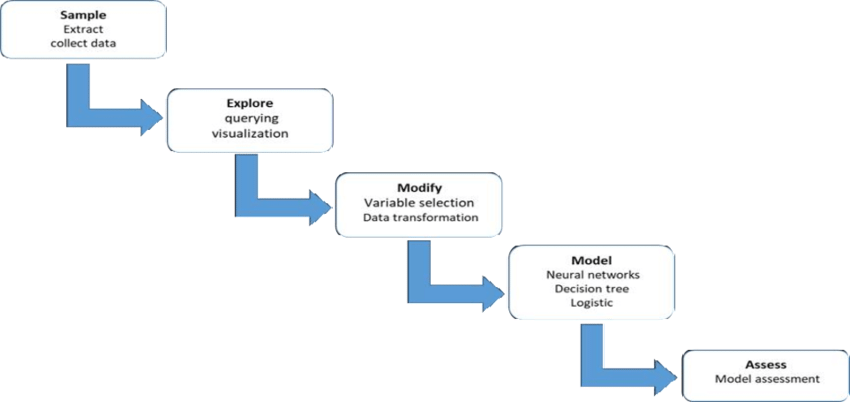
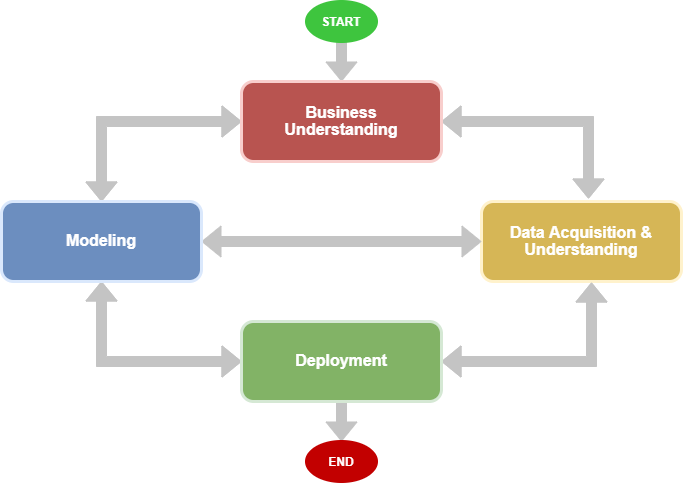
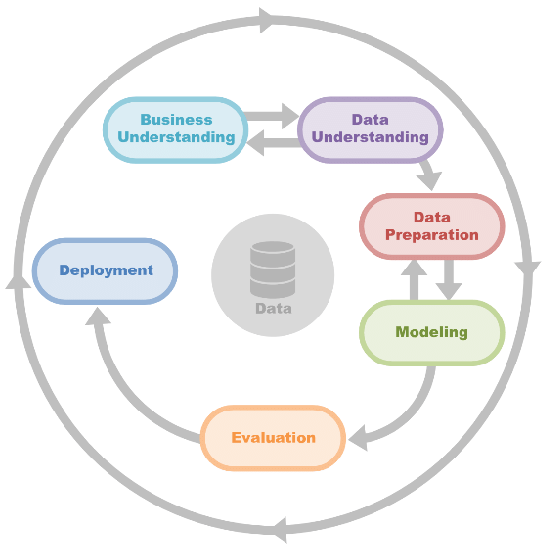
**Ministère de l’Enseignement Supérieur et de la Recherche Scientifique**

Université de la Mannouba

**Institut Supérieur des Arts Multimédias**

****  
  
  
  
  
Rapport De projet de fin d’études   
Présenté en vue de l’obtention du diplôme de   
  
**Licence en Big Data et Analyse de données**Sujet  
 **Prédiction du churn des clients Tunisie Télécom** Élaboré par : **Fatma Ezzahra Benabdallah**  
 Organisme d’accueil: **Tunisie Télécom  
  
  
  
  
  
  
  
 Encadrant Universitaire : Mr Bechikh Chedi   
Encadrant Professionnel : Mr Jemai Hatem   
  
  
 Année Universitaire 2023 / 2024  
  
  
  
Dédicace  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
Remerciements  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
Résumé** L'objectif principal de ce rapport de stage est de développer un modèle de prédiction de churn (taux de désabonnement) qui identifiera les clients les plus proches d'annuler leurs abonnements. Nous utiliserons des techniques d'apprentissage automatique pour prédire l'attrition des clients et intégrerons le modèle dans une application Web.  
 Cette application aidera les agences de télécommunications tunisiennes, dans notre cas la Tunisie Télécom, à faire des prévisions facilement et contiendra un lien vers un rapport BI qui permet de visualiser les données historiques des clients. En fait, pour mener à bien ce projet, nous avons suivi la méthodologie CRISP-DM.   
  
  
Mots clés : client churn, apprentissage automatique, prédiction churn, CRISP-DM, Python, BI

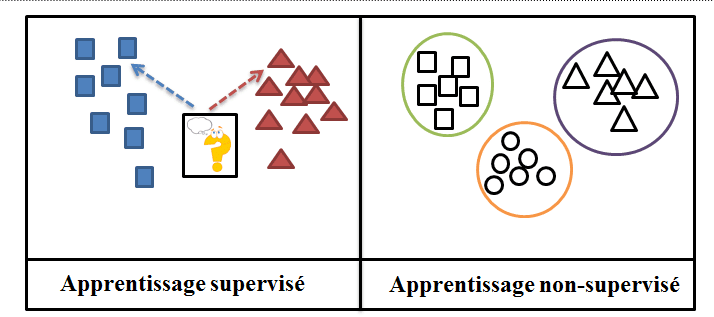
Abstract  
  
  
 The main objective of this internship report is to develop a churn (churn rate) prediction model that will identify the customers closest to cancel their subscriptions. We will use machine learning techniques to predict customer churn and embed the model in a web application.  
 This application will help Tunisian telecommunications agencies, in our case Tunisia Telecom, to make forecasts easily and will contain a link to a BI report which allows to visualize the historical data of the customers. In fact, to carry out this project, we followed the CRISP-DM methodology.   
  
Keywords : churn costumers, machine learning, churn prediction, CRISP-DM, Python, BI  
  
  
  
Table des matières  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
Liste des figures  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
Liste des Tableaux  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
Liste des abréviations  
  
  
  
  
  
  
  
  
  
  
  
  
  
Introduction générale  
 De nos jours, le secteur des télécommunications s’améliore de jour en jour, qui connaît une concurrence intense et une évolution technologique rapide. Cette situation a un impact considérable sur le taux de résiliation des clients, connu sur le nom de churn, ce qui en fait une préoccupation majeure pour ce secteur. Le churn se produit lorsque les clients décident de résilier leur abonnement et de passer à un autre fournisseur de services téléphoniques.  
 Pour les entreprises de télécommunications, il est devenu essentiel de mettre en place une gestion efface des relations clients afin d’accroître leurs revenus. En fait, la perte de clients ou d’abonnés reste un défi majeur pour l’industrie des télécommunications, car les clients n’hésitent pas à se désabonner ou changer d’opérateur s’ils ne trouvent pas sa satisfaction.  
 De nombreuses études ont démontré l’efficacité du Machine Learning dans la prévision de cette situation. L’intelligence artificielle joue un rôle crucial dans la prédiction du churn des clients.   
 En exploitant les techniques d’apprentissage automatique, il est possible de traiter les données et de développer des modèles de prédiction du taux de désabonnement. Ces modèles aident les opérateurs de télécommunications à identifier les clients les plus susceptibles de résilier leur abonnement, ce qui leur permet de prendre des mesures préventives pour les retenir.   
 Actuellement, l’opérateur Tunisie Télécom ne dispose d’aucun système de prédiction du churn. C’est pourquoi elle a décidé de faire une première étape en cherchant une solution au problème de l’attrition des clients. Alors, dans le but de résoudre ce problème, notre projet de fin d’étude intitulé « Prédiction du churn des clients Tunisie Télécom » a été initié. Ce projet s’inscrit dans le cadre d’une Licence en Big Data et Analyse de données à l’ISAMM. Il consiste à proposer un modèle de prédiction de churn pour identifier les clients les plus proches à arrêter leurs lignes téléphoniques avec Tunisie Télécom on utilise les techniques de Machine Learning et de Data Science. Par la suite, nous évaluerons les performances de ce modèle en analysant les résultats de la prédiction, en suivant la méthodologie de travail CRISP-DM. Le présent rapport comporte cinq chapitres :   
  
 • Le premier chapitre de notre rapport, intitulé "Étude de projet et compréhension du métier ", est consacré à la présentation du cadre de notre projet, l’organigramme d’accueil, la problématique que nous traitons, la solution que nous proposons, ainsi que les méthodologies utilisées en Data Science et les 2 outils adoptés pour la réalisation du projet. Ensuite, il se focalise sur les techniques de base de la Data Science  
 • Le deuxième chapitre, intitulé " Compréhension des données", se concentré sur l’exploration et de l’analyse des données de notre data base.   
 • Le troisième chapitre, intitulé "Prétraitement des données", présente en détail toutes les étapes nécessaires pour prétraiter et nettoyer les données avant de les modéliser.   
 • Le quatrième chapitre, intitulé "La modélisation", nous explorons en détail les différents modèles et les mesures de performance appropriées à appliquer dans notre projet.  
 • Le cinquième chapitre, intitulé "Évaluation, optimisation de la performance des modèles et déploiement du modèle choisi", se concentre sur les résultats obtenus dans chaque modèle, ainsi que la comparaison en utilisant les mesures nécessaires.  
 • Le dernier chapitre, intitulé "Déploiement", se focalise sur la création d’une interface web et la construction d’un tableau de bord afin de rendre notre projet facilement accessible aux utilisateurs.   
  
 Finalement, ce rapport est clôturé par une conclusion générale et quelques perspectives.  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
 Chapitre 1: Étude de projet et compréhension du métier  
  
  
  
  
**1.1 Introduction**  
 Dans ce chapitre, nous exposons l'objectif de notre projet à travers le contexte général des télécommunications en Tunisie, en mettant l'accent sur Tunisie Télécom. Nous débuterons par présenter l'organisme d'accueil, puis la problématique spécifique et la solution proposée. Ensuite, nous décrirons brièvement l'architecture du projet, la méthodologie adaptée et l'environnement de travail. Enfin, nous aborderons les concepts essentiels du projet.  
  
**1.2 Présentation de l’organisme d’accueil   
 1.2.1 Présentation du Tunisie Télécom** Tunisie Télécom est une entreprise de télécommunications tunisienne qui fournit des services de téléphonie fixe et mobile, d’internet et de transmission de données. Actuellement, c’est l’un des plus grands opérateurs des télécommunications de la région. Tunisie Télécom a été fondée en 1995 et est ouverte au Grand public qu’aux entreprises et opérateurs tiers.  
  
   
 **1.2.2 Historique**   
 L’Agence nationale des télécommunications a été créée par la promulgation de la loi n° 36 du 17 avril 1995. Le bureau a ensuite changé son statut juridique pour devenir une société anonyme en vertu du décret n° 30 du 5 avril 2004 dénommé "Tunisie Télécom". En juillet 2006, le capital de Tunisie Télécom a ouvert 35% au profit du consortium émirati « TeCom-DIG ». L’entreprise vise à accroître la rentabilité de Tunisie Télécom et à en faire l’un des principaux opérateurs internationaux.  
 **1.2.3 Organisation :** Tunisie Télécom est composée de 24 directions régionales, 80 Actuels et points de vente, Plus de 13 000 emplacements privés. Elle emploie plus de 8 000 agents. Cet opérateur historique dispose aussi six centres de support clients de téléphonie fixe et Mobile et données. La figure 1.2 présente l’organisation fonctionnelle de Tunisie Télécom.  
  
  
**1.3 Présentation du projet  
 1.3.1 Cadre général du projet**   
  Ce projet se présente dans le cadre d’un projet fin d’étude à l’École Supérieure d’Économie Numérique pour le but d’obtenir le diplôme d’une Licence en Big Data et Analyse de données. Le stage est effectué au sein de la société Tunisie Télécom.  
  
 **1.3.2 Étude de l’existant :** La phase d'étude de l'existant revêt une grande importance dans le projet. Elle permet d'obtenir une vision claire et précise des concepts de l'environnement de travail afin d'identifier les besoins et de déterminer les problématiques à résoudre. L'objectif primordial de Tunisie Telecom est de satisfaire ses clients, ce qui nécessite une approche basée sur la compréhension de leurs besoins et de leurs comportements d'achat. Dans le domaine des télécommunications, les clients recherchent généralement les produits ou services qui leur procurent le maximum de satisfaction.  
   
 Pour assurer la satisfaction de ses clients, Tunisie Telecom utilise différentes méthodes motivées par diverses raisons, notamment la détection des causes d'insatisfaction et des dysfonctionnements potentiels, ainsi que l'évaluation des opinions des clients.  
 Actuellement, Tunisie Telecom met en place une démarche qualité visant à mesurer la satisfaction de ses clients et à fournir un diagnostic précis de leurs satisfactions et 7 insatisfactions. Ces évaluations conduisent à des actions concrètes pour améliorer le niveau de satisfaction.  
 Ainsi, la mesure de la satisfaction client devient un outil essentiel pour guider l'entreprise dans sa démarche de satisfaction client, en permettant de déterminer dans quelle mesure les clients sont satisfaits des services et d'évaluer le niveau de réponse de l'entreprise aux attentes des clients.  
  
 **1.3.3 Problématique** Tunisie Télécom, en tant que premier opérateur de télécommunication en Tunisie, propose différentes offres et une variété de services pour répondre aux besoins de sa clientèle. Malgré les services proposés, certains clients ne sont pas satisfaits et ont même décidé de résilier leur abonnement et de passer à un opérateur concurrent, ce qui représente un problème critique car le taux de désabonnement entraîne une baisse des revenus et une diminution de la part de marché. Les raisons de résiliation incluent les coûts élevés des appels, des problèmes de couverture réseau et un service client insatisfaisant. Ces facteurs peuvent influencer leur décision de choisir un autre opérateur.  
 **Comment pouvons-nous donc trouver les clients qui peuvent résilier leurs contrats et quittent l’opérateur Tunisie Télécom ?  
  
 1.3.4 Solution**   
 Tunisie Télécom s'engage à développer des moyens permettant de prédire le taux de désabonnement de ses clients. L'entreprise cherche à identifier les différents segments de sa clientèle, car le coût d'acquisition d'un nouveau client est généralement plus élevé que celui de fidélisation d'un client existant.   
 La solution envisagée consiste à créer un modèle de prédiction du taux de désabonnement en utilisant divers algorithmes d'apprentissage supervisé. L'objectif est de sélectionner le meilleur modèle en fonction de son taux de précision, en analysant les données historiques des clients et en visualisant les résultats de la prédiction ainsi que les performances du modèle. Les modèles seront évalués en termes de précision dans la prédiction des clients résiliant leur abonnement et de ceux qui décident de rester.   
 Un data scientist recherche toujours la satisfaction de l'utilisateur. Dans ce contexte, notre meilleur modèle de prédiction sera intégré dans une page web simple et facile à utiliser.  
  
**1.4 Méthodologie de gestion de Projet adoptée** Avant de commencer la réalisation de notre projet, il est crucial d'examiner et d'évaluer la méthode de gestion de projet à adopter afin de garantir une coordination efficace entre les parties prenantes et les différentes tâches. Alors, nous allons étudier trois méthodologies qui sont : SEMMA, TDSP et CRISP-DM. Notre objectif est d’analyser chaque méthode afin de choisir celle qui conviendra le mieux à notre projet.   
  
 **1.4.1 Méthodologie SEMMA**  
 *SEMMA* est une méthode utilisée en Data Mining pour résoudre des problèmes d’analyse de données en cinq étapes : échantillonnage, exploration, modification, modélisation et évaluation. Cette méthodologie a été développée par SAS pour aider les analystes à utiliser les données pour résoudre des problèmes commerciaux et analytiques.  
 *1. Échantillonner les données (Sample) :* Cette étape consiste à choisir les données pertinentes qui sera utilisé pour l’analyse et extraire un échantillon représentatif.   
 *2. Explorer et visualiser les données (Explore) :* Exploration et visualisation des données pour comprendre leur structure et permet de détecter les anomalies et identifier les relations entre les variables.   
 *3. Modifier et nettoyer les données (Modify) :* Nettoyage et préparation des données pour l’analyse.   
 *4. Modéliser les données (Model) :* Construction des modèles de Machine Learning pour prédire ou expliquer les phénomènes étudiés.   
 *5. Analyser les résultats (Assess) :* L’évaluation des résultats obtenus à l’aide des modèles et les interpréter en les comparant à des critères de performance prédéfinis.  
  
   
 **1.4.2 Méthodologie TDSP**   
 Le TDSP est une méthodologie de gestion de projet agile et itérative développée spécifiquement pour la science des données. Il a été créé par Microsoft, pour faciliter la collaboration entre les membres d’une équipe. La méthode TDSP se divise en cinq étapes :   
 *1. Business Understanding :* L’objectif de cette première phase est de bien comprendre les enjeux métiers, d’identifier les objectifs du projet ainsi que les critères de réussite et les parties prenantes impliquées.   
 *2. Data Acquisition and Understanding :* Collecte, nettoyage et préparation des données nécessaires pour répondre aux objectifs métier.   
 *3. Modeling :* Au cours de cette phase, des modèles de données sont développés et testés pour répondre aux objectifs métiers du projet et sélectionner le modèle le plus performants.   
 *4. Deployment* : Une fois que les modèles ont été créés, ils sont prêts à être déployés dans un environnement opérationnel pour l’utilisation.   
 *5. Acceptance :* Cette étape a pour but de mesurer les résultats du déploiement des modèles pour vérifier la réussite du projet et la réalisation des objectifs métier.  
   
 **1.4.3 Méthodologie CRISP-DM**   
 CRISP-DM a été publié en 1999 pour normaliser les processus d’exploration de données dans tous les secteurs. C’est devenu la méthode la plus populaire pour les projets d’exploration de données, d’analyse et de science des données.   
 Le cycle de vie de CRISP-DM est divisé en 6 étapes :   
 *1. Business Understanding :* La première étape est d’identifier le problème que l’organisation est en train d’essayer de résoudre en se basant sur données et établir une architecture bien définie pour la mise en œuvre du projet.   
 *2. Data UnderStanding :* Cette étape comprend la collecte des informations initiales, la compréhension et la description du type de données à analyser et établir des liens entre les données et leur signification d’un point de vue métier.   
 *3. Data Preparation :* Cette phase a pour but de préparer les données à analyser. Il consiste notamment à nettoyer les données, à les transformer afin qu’ils soient compatibles avec les algorithmes qui seront utilisés.   
 *4. Modeling* *:* La modélisation est basée sur le choix de paramétrage ainsi que le modèle qui sera utilisé après l’essai de plusieurs modèles. Cette étape comprend 4 fonctions : sélection de 9 Chapitre 1. Étude du projet la technique de modélisation, conception du prototype, construction du modèle, évaluation du modèle.   
 *5. Evaluation :* Avant de passer au déploiement définitif du modèle, il faut contrôler et vérifier les modèles ou les connaissances obtenues pour s’assurer qu’ils atteignent les objectifs énoncés dès le début du processus, permet aussi de prendre la décision de déploiement du modèle ou de l’améliorer.   
 *6. Deployment :* C’est la phase finale du processus. Elle s’agit de déployer les analyses pour une utilisation effective. Son objectif est de mettre en forme les connaissances obtenues par la modélisation et les intégrer dans le processus de prise de décision.  
  
  
  
   
  
  
  
  
  
  
  
  
  
 **1.4.8 Comparaison des méthodologies**   
 Les trois méthodologies de gestion de projet de science des données SEMMA, TDSP et CRISP-DM ont le même objectif, mais il y a quelques différences entre eux.   
 Le tableau 1.1 présente la comparaison entre ces trois méthodologies.

|  |  |  |  |
| --- | --- | --- | --- |
| Critère | SEMMA | TDSP | CRISP-DM |
| Domaine d’application | Data Mining et analyse prédictive | Projet de science des données en équipe | Data Mining et exploration de données |
| Phases | 5 phases | 5 phases | 6 phases |
| Approche itérative | Non | Oui | Oui |
| Flexibilité | Faible | Moyenne | Élevée |
| Priorités | Accent sur l’exploitation et la modification | Accent sur l’acquisition et la préparation | Accent sur la compréhension et la préparation |

Il est crucial de choisir la méthodologie de gestion de projet adéquate pour garantir le succès d'un projet de science des données. Après une comparaison des trois méthodologies disponibles, nous avons opté pour **CRISP-DM** en raison de sa souplesse et de son approche itérative, qui permet une adaptation aux changements tout au long du projet.  
 **1.5 Architecture de Projet** Notre système de prédiction de churn repose sur une architecture composée de huit étapes essentielles, assurant ainsi des prévisions précises et efficaces :   
1- Collecte des données des clients.   
2- Compréhension des données clients.   
3- Prétraitement des données clients.   
4- Sélection des fonctionnalités pertinentes pour le développement des modèles de prédiction.   
5- Construction des modèles.   
 6- Évaluation et calcul de la précision des modèles.  
 7- Prise de décision pour la rétention des clients.   
 8- Création du tableau de bord.  
 **1.6 Environnement de travail**   
 Dans cette section, nous déterminons les outils logiciels et matériels qui ont été utilisés pour la réalisation et la mise en place de notre projet.

**1.7 Conclusion**  
 Ce premier chapitre a inclus une présentation de l’organisme d’accueil, ensuite la problématique et la solution proposée. Puis la méthodologie de gestion de projet que nous allons adopter tout au long du projet. Et enfin, nous avons décrit l’environnement de travail.

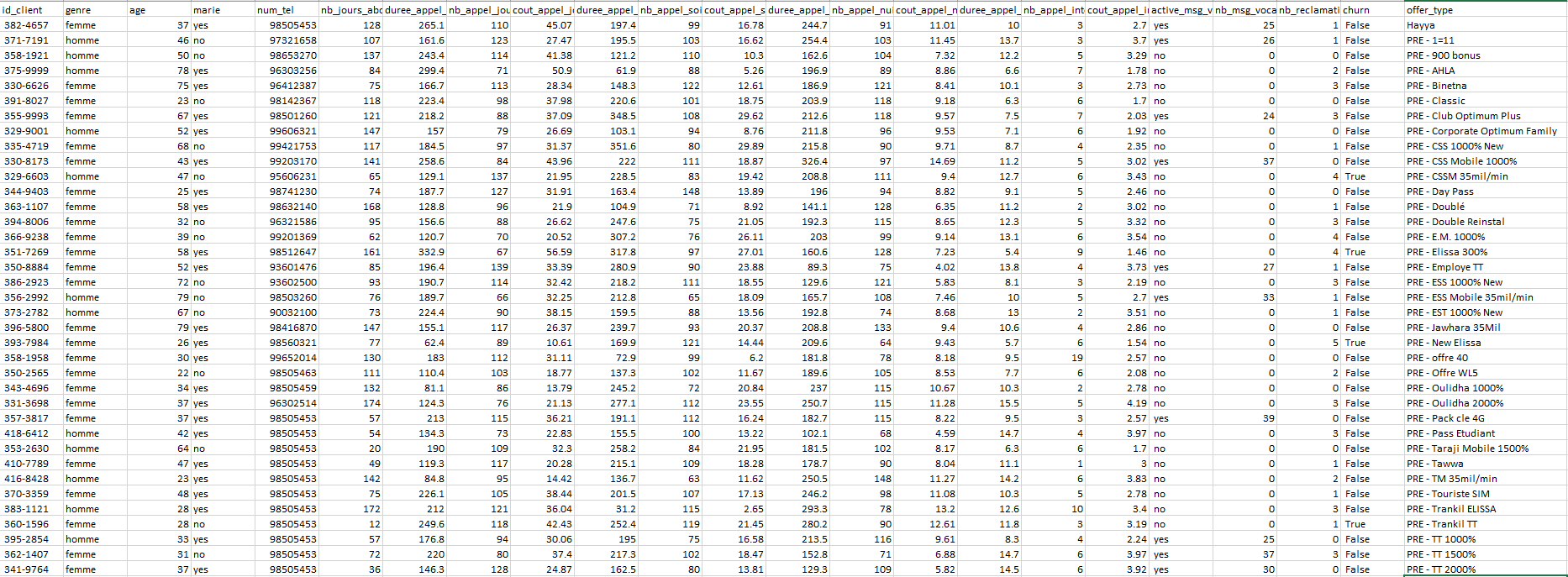
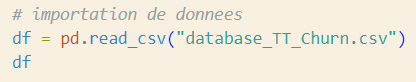
Chapitre 2 : Compréhension du problème métier et des données

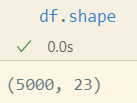
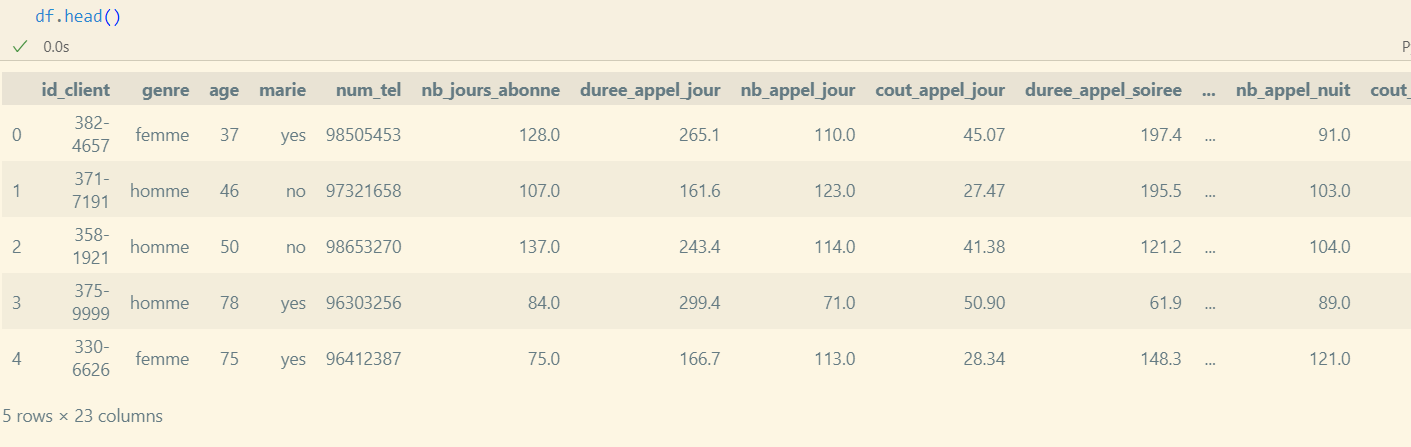
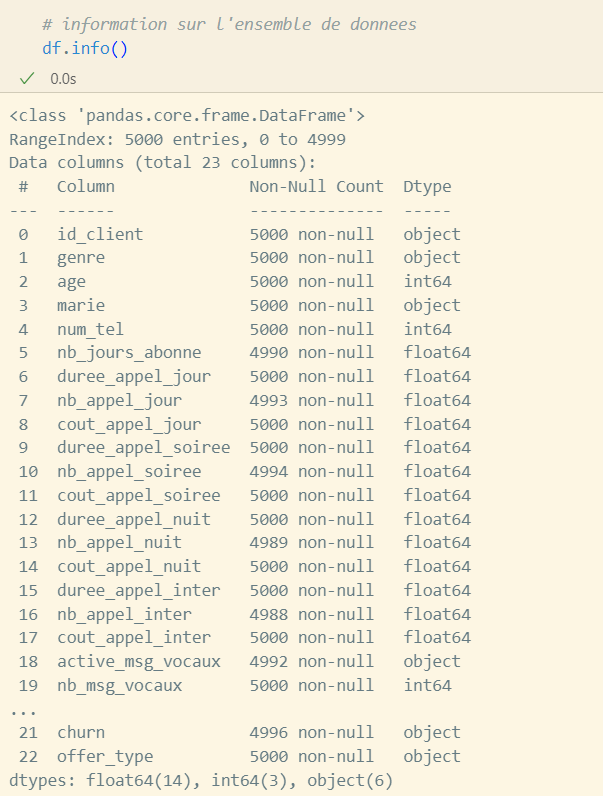
**2.1 Introduction** Au cours de ce chapitre, nous allons présenter les concepts fondamentaux de notre projet, puis analyser en détails les données collectées. Nous mettrons en œuvre des techniques de visualisation pour explorer et interpréter les données, dans le but d’enrichir leur compréhension.  
  
**2.2 Notions théoriques liées au projet**  
 Dans cette section, nous allons explorer les différents concepts liés à notre projet. Les principaux concepts qui nous intéressent sont le churn, apprentissage automatique et les défis de la modélisation. Nous allons commencer par la définition du churn.   
 **2.2.1 Churn**   
 Selon B. Bathelot Churn est un terme anglais qui désigne la perte de clientèle ou d’abonnées. Le taux de churn, ou taux d’attrition, est un indicateur important pour mesurer la performance des services d’une entreprise. Il représente le pourcentage de clients qui cessent d’utiliser les produits ou services de l’entreprise sur une période donnée. Cet indicateur permet de déterminer si le service proposé par l’entreprise est en phase avec les attentes des clients ou non.  
  
 Taux de churn = Nombre de clients perdus / Nombre de clients total   
 Equation 1 : Formule de taux de churn  
  
 Les clients churn de télécommunication peuvent être divisés en deux catégories principales : involontaire et volontaire.   
• Client volontaire : Décrit comme la fin du service par l’abonnée.   
• Client involontaire : Les abonnées que la société de télécommunications décide de supprimer pour diverses raisons, notamment la fraude et le non-paiement de leurs factures.   
 **2.2.2 Apprentissage automatique**   
 L’apprentissage automatique, également connu sous le nom de Machine Learning en anglais, est un sous-domaine de l’intelligence artificielle qui se concentre sur la création de modèles permettant  
Compréhension du problème métier et des données à une machine d’apprendre à partir de données et de s’améliorer automatiquement en les exploitant.  
  
 Les techniques d’apprentissage automatique peuvent être regroupés en deux catégories :   
• Supervisé   
• Non-supervisé  
 La figure présentée ci-dessous représente la différence entre les 2 catégories  
****

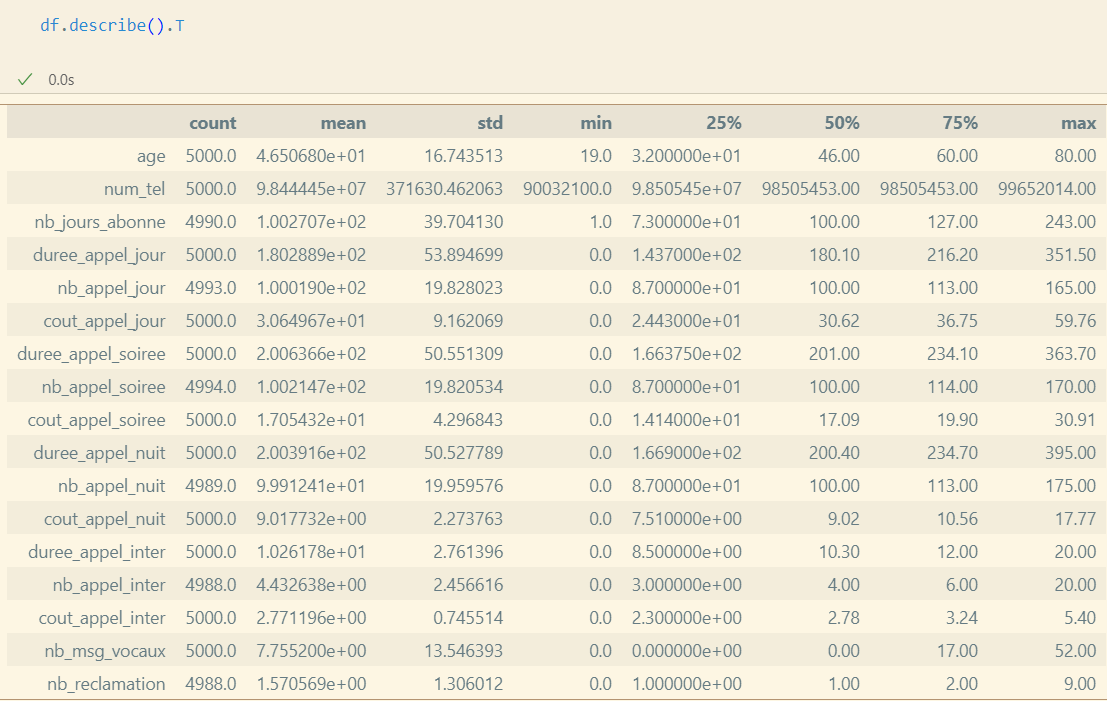
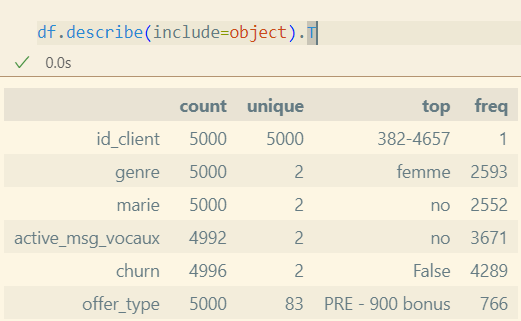
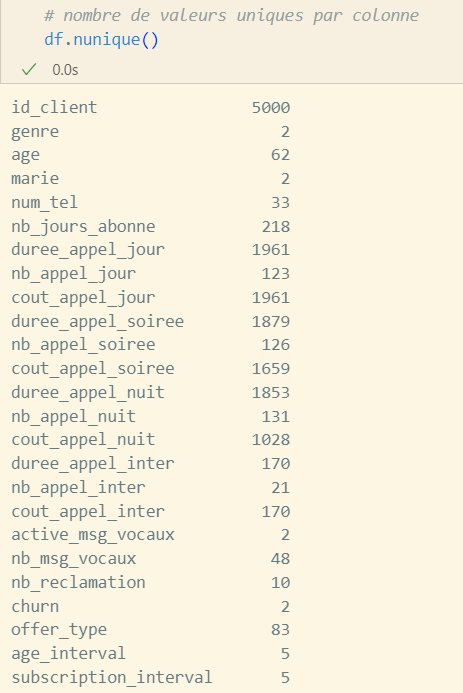
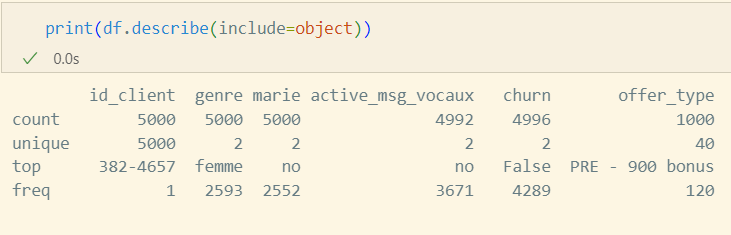
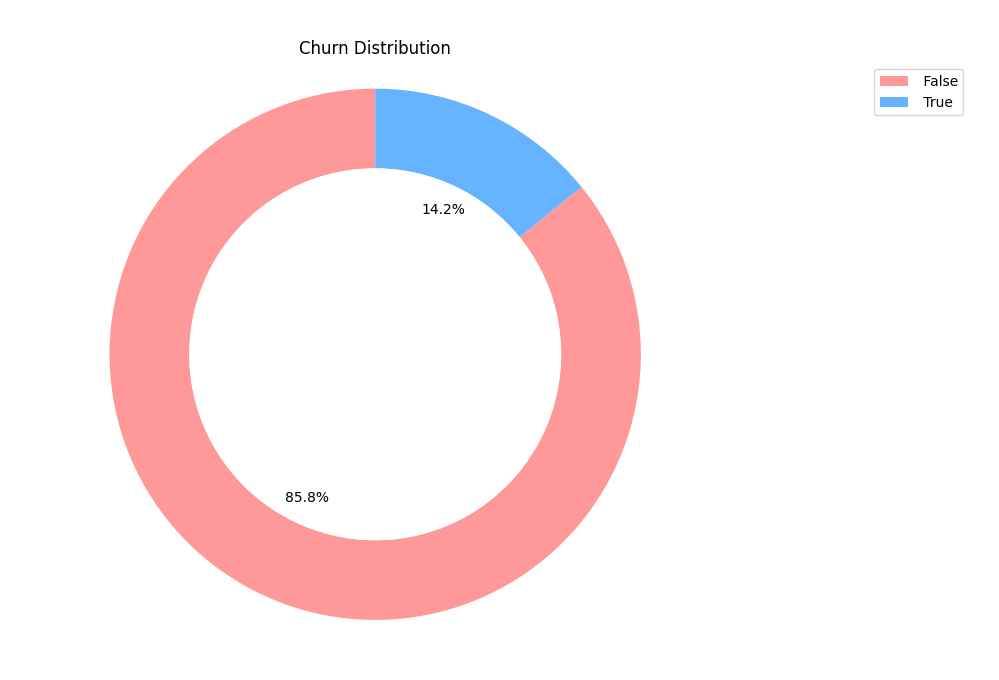
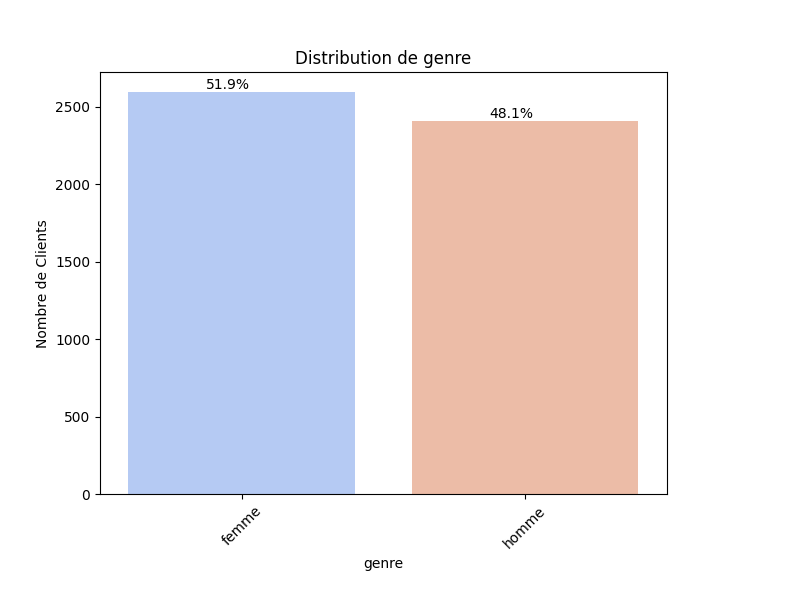
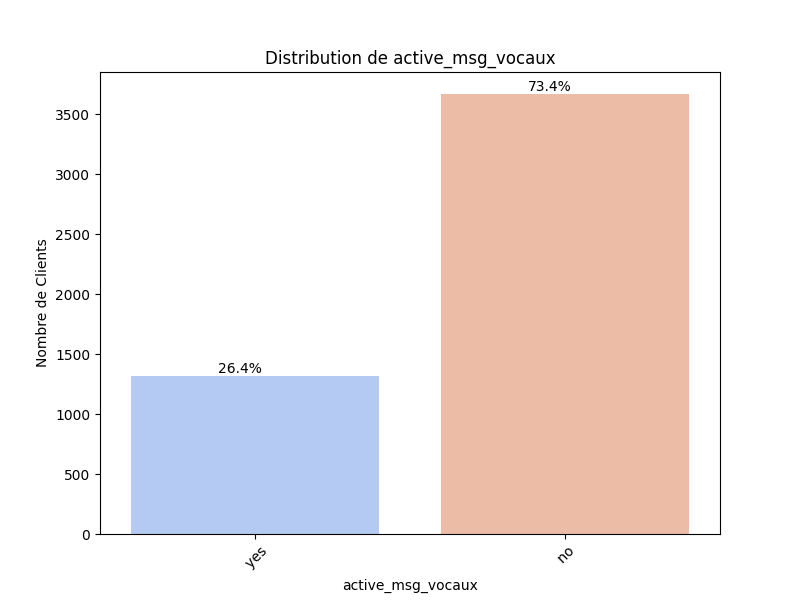
**2.3 Choix de la base de données** Notre dataframe disponible pour cette recherche comprend des renseignements sur l'historique des opérations d'appels nationaux et internationaux, les messages vocaux et les réclamations des clients de Tunisie Télécom, ainsi que leurs données démographiques et leur numéro téléphonique.

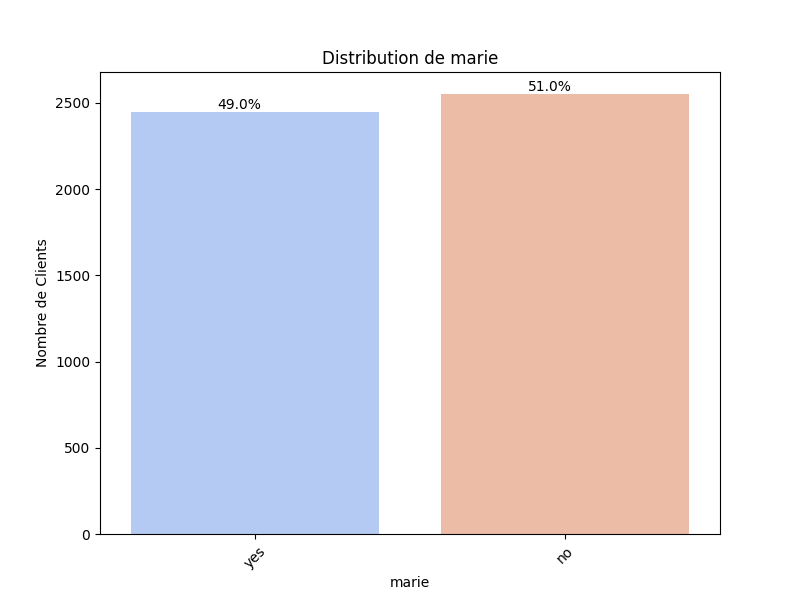
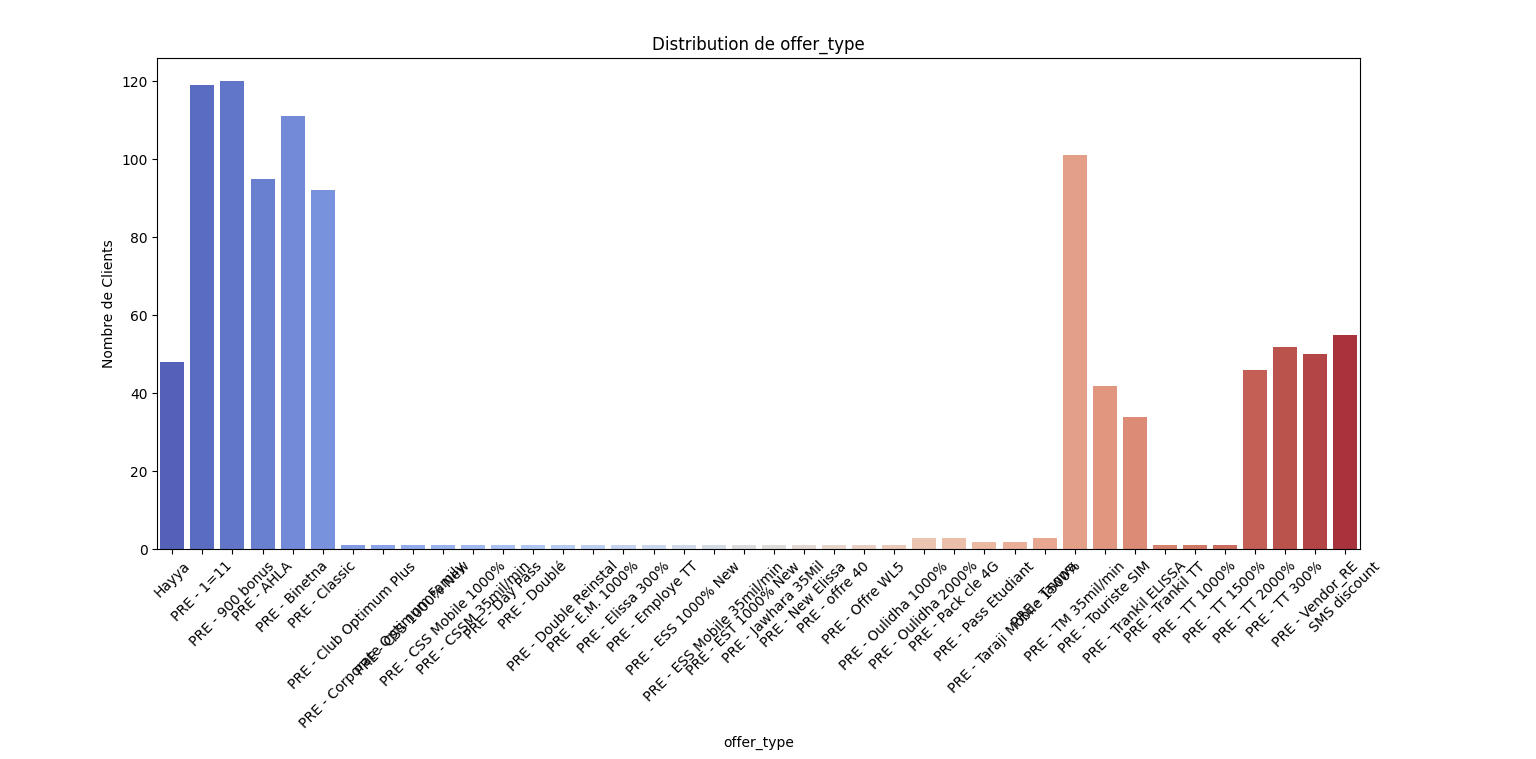
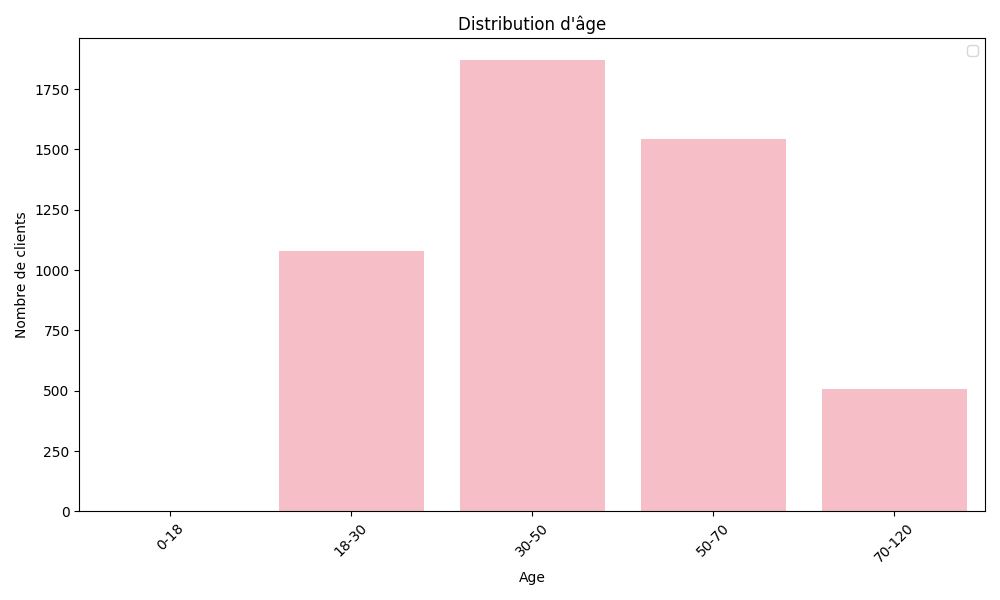
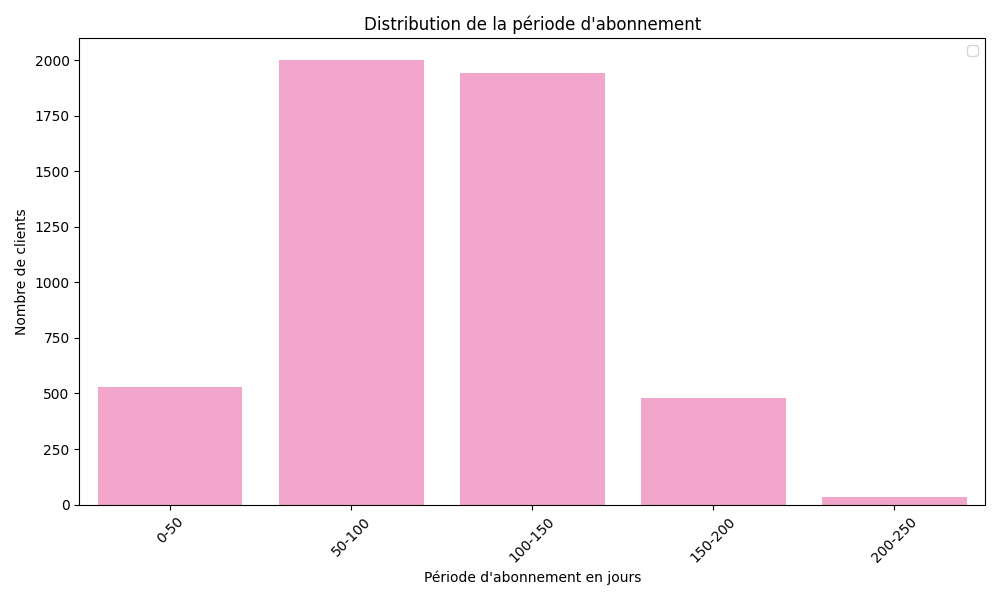
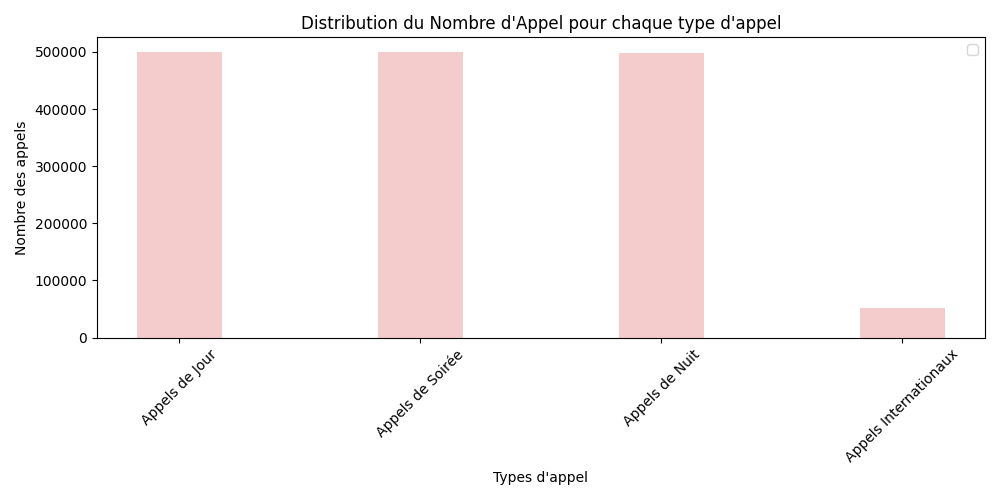
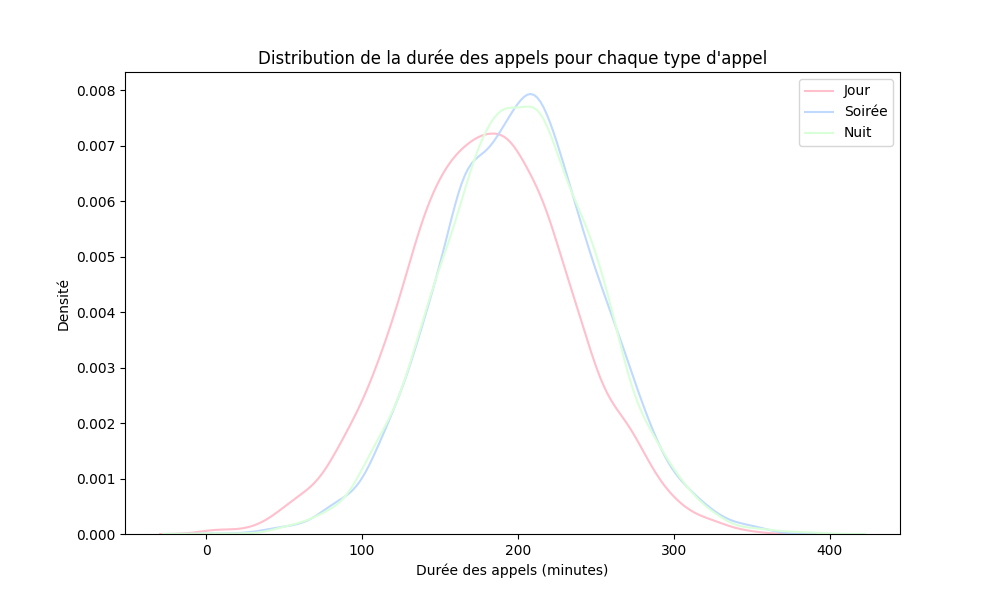
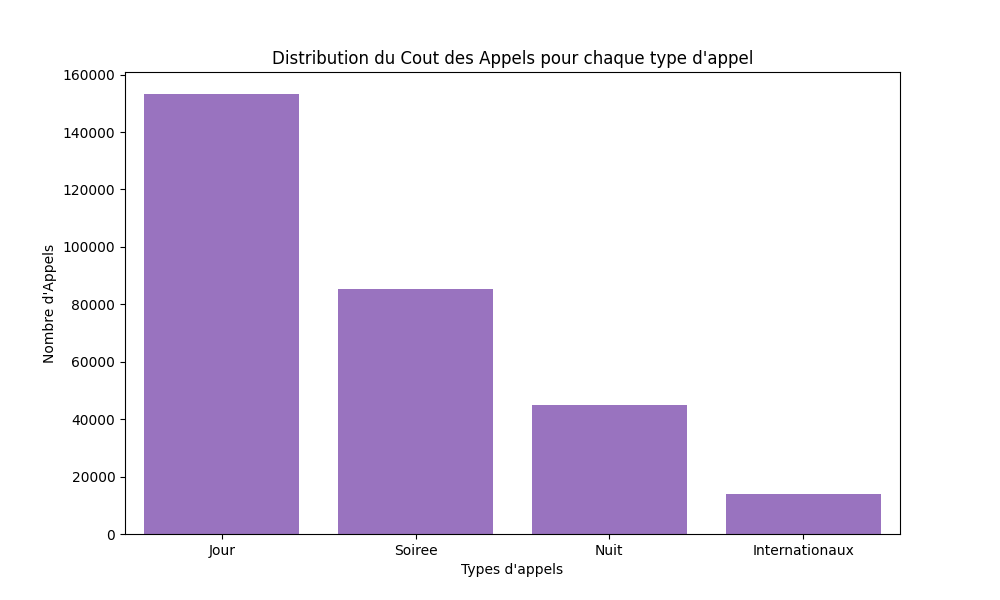
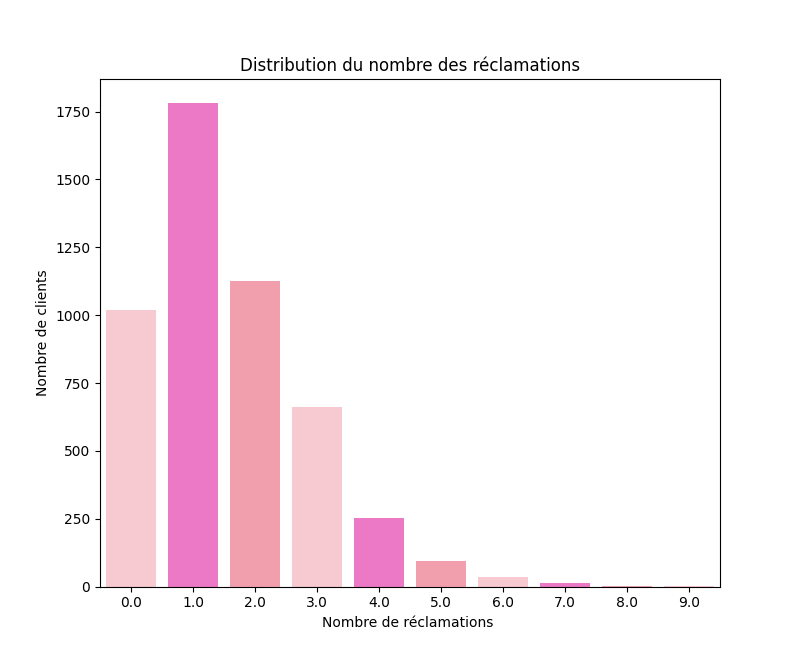
Le tableau 2.1 donne une vue d’ensemble des données. Il identifie les variables, leurs types et une simple description de leur signification.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Description** | **Type** |
| Id\_client | Identifiant du client | Chaine de caractère |
| genre | La catégorisation sociale des individus en fonction de leur sexe | Numérique |
| age | L’âge de chaque client | Numérique |
| marié | L'état civil de chaque client | Numérique |
| num\_tel | Le numéro téléphonique pour chaque client | Numérique |
| nb\_jours\_abonne | Nombre de jours que le client a été abonné à Tunisie Télécom | Numérique |
| durée\_appel\_jour (minutes) | La durée totale des appels effectués pendant la journée (en minutes) | Numérique |
| nb\_appel\_jour | Le nombre total d’appels effectués pendant la journée | Numérique |
| cout\_appel\_jour | Le coût total des appels effectués pendant la journée. | Numérique |
| durée\_appel\_soirée (minutes) | La durée totale des appels effectués pendant la soirée (en minutes) | Numérique |
| nb\_appel\_soirée | Nombre de jours que le client a été abonné à Tunisie Télécom | Numérique |
| cout\_appel\_soirée | Le coût total des appels effectués pendant la soirée | Numérique |
| durée\_appel\_nuit (minutes) | La durée totale des appels effectués pendant la nuit (en minutes) | Numérique |
| nb\_appel\_nuit | Le nombre total d’appels effectués pendant la nuit | Numérique |
| cout\_appel\_nuit | Le coût total des appels effectués pendant la nuit | Numérique |
| durée\_appel\_inter (minutes) | La durée totale des appels internationaux (en minutes) | Numérique |
| nb\_appel\_inter | Le nombre total d’appels internationaux | Numérique |
| cout\_appel\_inter | Le coût total des appels internationaux | Numérique |
| active\_msg\_vocaux | Indique si le client a activé la messagerie vocale (‘Yes’, ‘No’) | Chaine de caractère |
| nb\_msg\_vocaux | Le nombre de messages vocaux reçus par le client | Numérique |
| nb\_reclamation | Le nombre des réclamations déposées par le client | Numérique |
| churn | False si le client est fidèle, True si le client churn | Chaine de caractère |
| offer\_type | Le type d’offre tarifaire choisis par le client | Chaine de caractère |

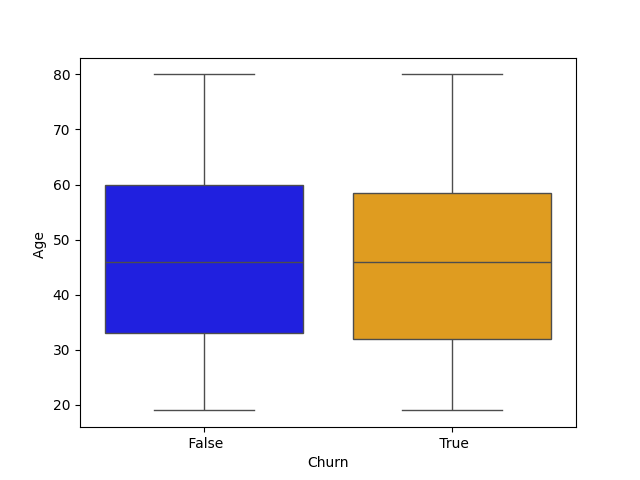
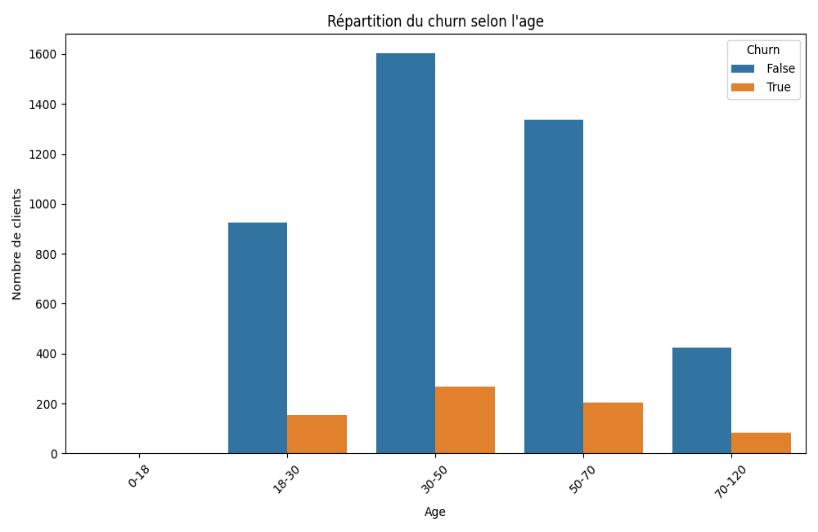
Notre base de données contient 23 variables dont 5 variables "id\_client", "marie" "offer\_type”, “active\_msg\_vocaux" et "churn" sont des variables de type "Chaîne de caractères".  
  
  
 Afin de réaliser notre étude, nous disposons d’un échantillon de 5000 clients qui ont été répartis de la manière suivante :  
 • 4289 : Clients fidèles (non-churn)   
 • 707 : Clients ayant résilié leur abonnement (churn)  
  
 Nous allons utiliser la fonction "read\_csv" de bibliothèque Pandas pour importer les données d’un fichier CSV et les stocker dans un DataFrame.  
 

**2.4 Exploration des données** L’exploration de données est une étape cruciale dans l’analyse de données qui permet de mieux comprendre les données en les examinant et en les interprétant dans leur contexte. Cette étape comprend généralement des tâches telles que l’analyse de statistiques descriptives et la visualisation de données.  
  
  
**• data.shape :** Permet de retourner le nombre de lignes et de colonnes du DataFrame   
  
**• data.head() :** Permet d’afficher les premières lignes d’un DataFrame. Par défaut, renvoie les 5 premières lignes.  
  
  
**• data.info() :** Permet d’afficher un résumé du DataFrame (les noms et types de colonne, le nombre de valeurs non-nulles et l’utilisation de la mémoire)  
  
**• data.describe() :** Permet d’afficher des statistiques descriptives d’un DataFrame (le nombre de valeurs, la moyenne, l’écart type, les valeurs minimum et maximum, et les quartiles pour chaque colonne)

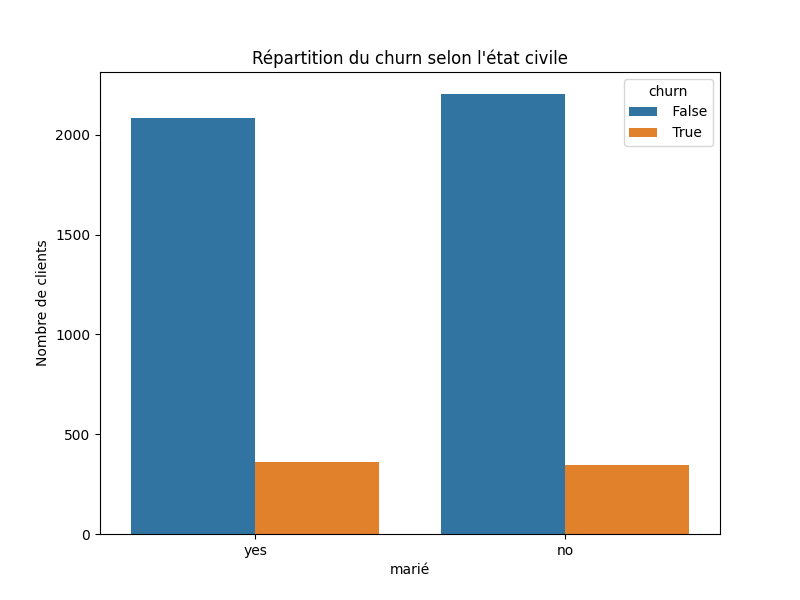
  
  
   
 Nous allons afficher des statistiques pour les colonnes qui contiennent des données de type "Object".  
  
  
**• data.nunique() :** renvoie le nombre de valeurs uniques pour chaque colonne.  
  
 **2.5 Analyse des données   
 2.5.1 Analyse univarié**   
 L’analyse univariée est une méthode d’analyse statistique qui permet d’explorer une seule variable à la fois. Elle permet de résumer et de visualiser les caractéristiques de la variable étudiée. Cette analyse nous permet d’améliorer notre compréhension des données.   
La répartition des clients en fonction de leur statut de churn est représentée dans la figure   
  
  
  
☆ Variables catégorielles :  
  
 • Churn : La répartition des clients en fonction de leur statut de churn est représentée dans la figure suivante   
   
 Après l’analyse sur le taux de désabonnements de la clientèle, nous remarquons que la majorité des clients restent actifs, il représente une proportion de 85.8%, alors que 14.2% des clients ont résilié leurs services, ce qui peut être considéré comme relativement élevé. Cela signifie qu’il pourrait avoir un impact négatif important sur les modèles finaux.  
  
  
  
  
  
 • Genre :  
   
 D’après le graphique, nous pouvons observer que notre jeu de données a une répartition entre les femmes et les hommes presque équilibrée, avec une proportion de 51,9% de femmes et de 48,1% d'hommes.  
  
 • Active Messages Vocaux : La figure 2.14 présente la distribution des clients en fonction de leur activation des messages vocaux.  
 

D’après le graphique, nous pouvons observer que la majorité des clients ne sont pas intéressés par l’activation de la fonctionnalité de messages vocaux, avec un pourcentage de 73.50% qui n’a pas activé cette fonctionnalité. Seulement 26.50% des clients ont activé les messages vocaux. Cette analyse indique que la fonctionnalité de messages vocaux ne semble pas être très utilisée par la majorité des clients.  
  
 • Marié :   
   
  
 D’après le graphique, nous pouvons observer que les clients sont répartis, selon leur état civil, d’une manière presque équivalente avec une proportion de 49% sont mariés et de 51% sont non mariés.  
  
 • Offre Tarifaire :   
   
  
☆ Variables numériques :  
 • Age :  
   
  
  
  
 Le graphique montre que la majorité des clients de Tunisie Télécom ont entre 30 et 50 ans, suivis par ceux âgés de 50 à 70 ans. Les tranches d'âge de 18 à 30 ans et de 70 à 120 ans sont moins représentées. Cette diversité suggère que Tunisie Télécom cible un large éventail d'âges, ce qui pourrait être bénéfique pour l'entreprise.  
  
 • Nombre de jours d’abonnement :  
   
 D’après le graphique, nous pouvons observer que la majorité des clients sont abonnés entre 50 et 100 jours, indiquant l'importance de cette période pour la fidélisation. Les abonnements de 100 à 150 jours sont également significatifs, montrant une bonne rétention. Cependant, un petit groupe de clients préfère des abonnements plus courts, peut-être des essais ou des changements fréquents d'opérateur. Enfin, quelques clients sont fidèles depuis plus de 200 jours, soulignant la satisfaction à long terme avec les services de Tunisie Télécom.  
  
 • Nombre des Appels pour chaque type d’appel:  
   
  
   
 • Durée des Appels pour chaque type d’appel:  
   
  
  
  
 • Cout des Appels pour chaque type d’appel:  
   
 • Nombre de réclamations :  
   
  
 **2.5.2 Analyse bivariée**

L’analyse bivariée est une méthode statistique qui permet d’examiner la relation entre deux variables en les comparant ou en les associant, ce qui peut aider à comprendre comment elles sont liées et comment elles interagissent.

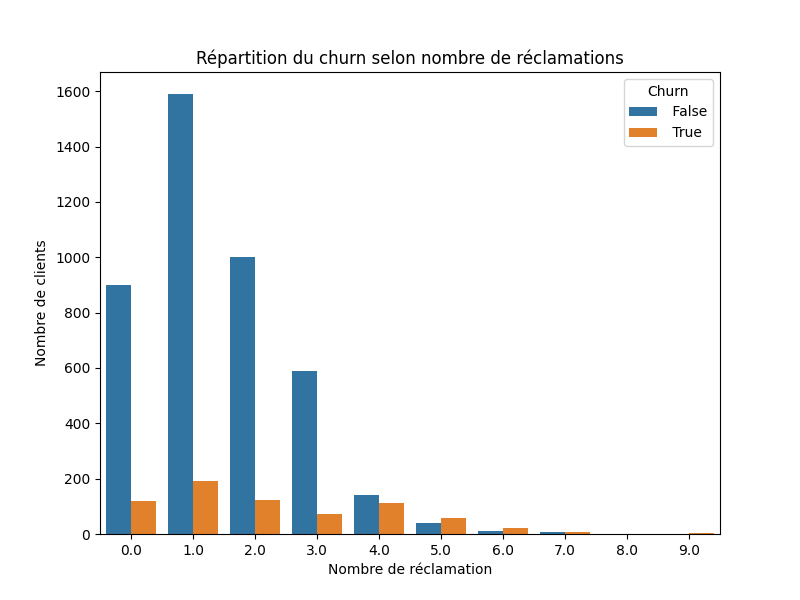
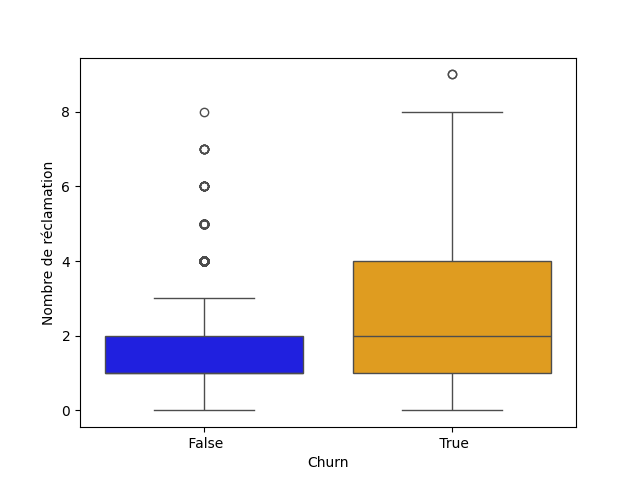
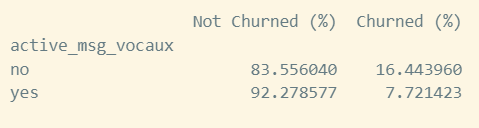
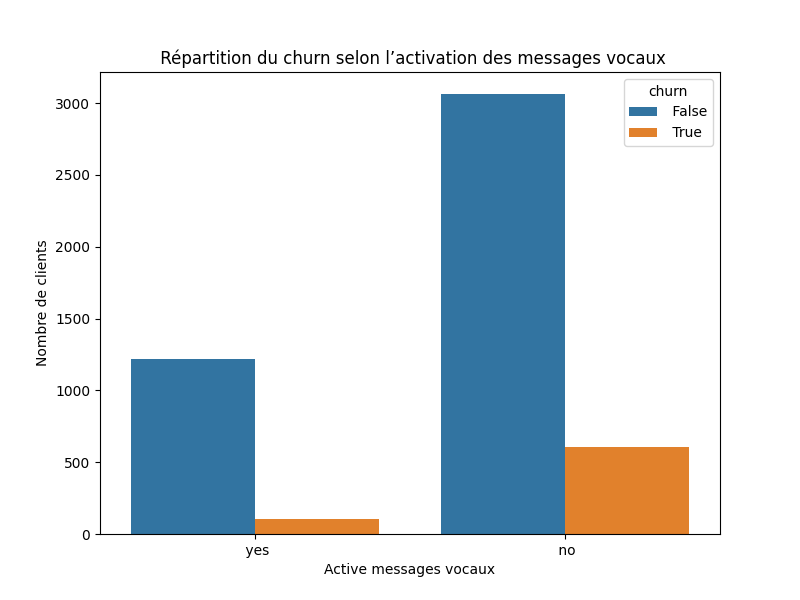
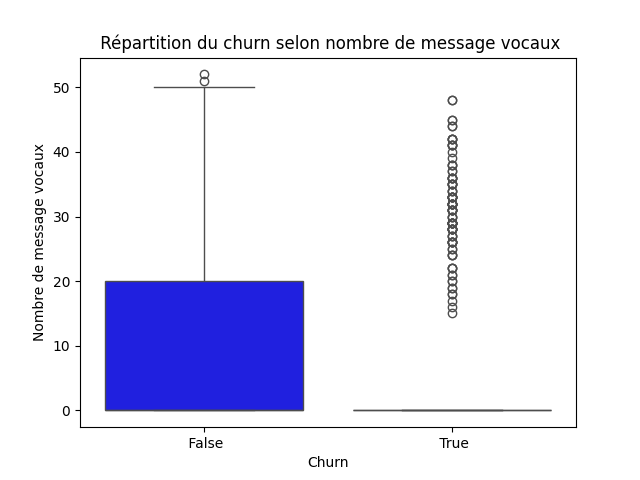
 • La figure illustre la distribution du churn selon l’age.  
   
 

Cette analyse révèle une corrélation entre l'âge des clients et leur taux de churn. Les clients âgés de 18 à 30 ans présentent le taux de churn le plus élevé, d'environ 14%. Ce taux diminue progressivement avec l'âge des clients. Les jeunes résilient leur abonnement plus fréquemment que les clients plus âgés. Une analyse approfondie est recommandée pour comprendre les raisons de ces différences d'attrition entre les groupes d'âge.

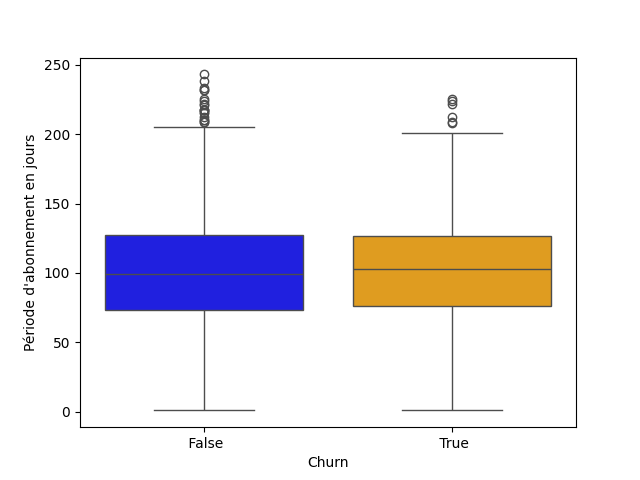
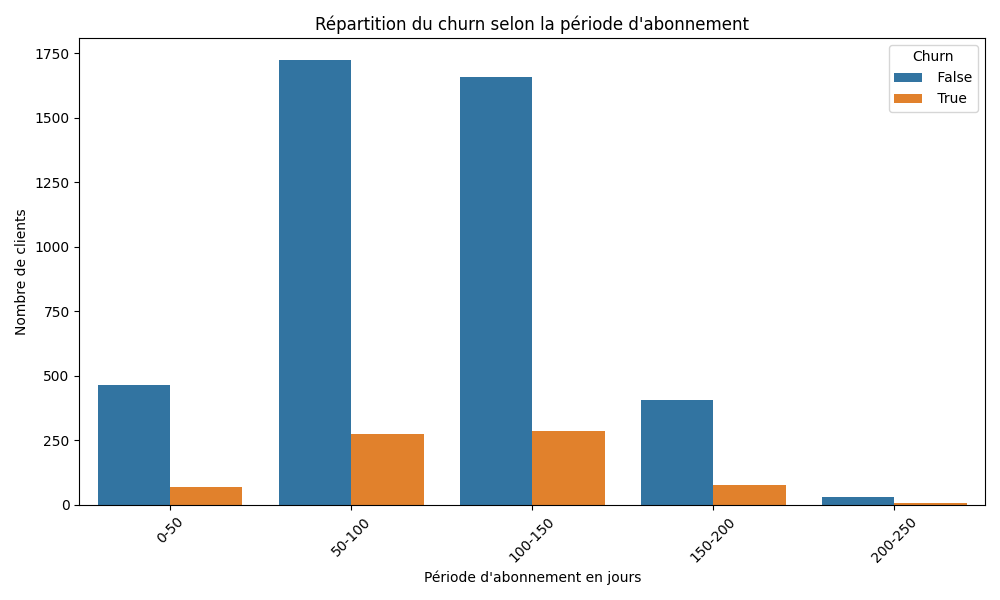
• La figure illustre la distribution du churn selon l’état civile.  
 

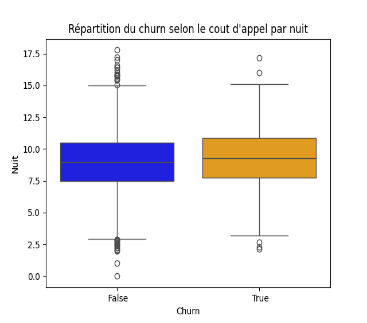
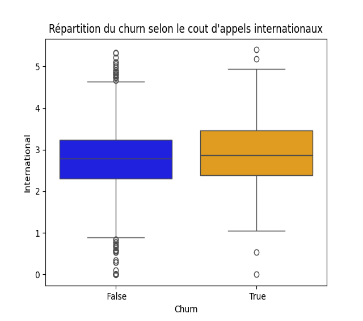
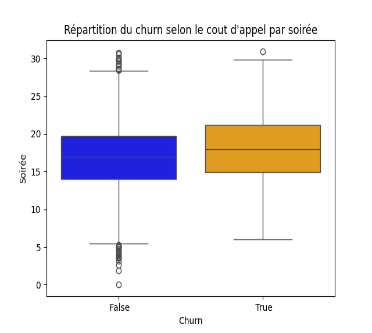
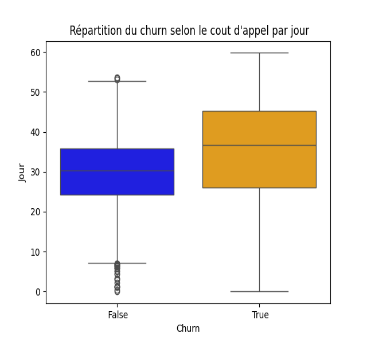
L'analyse de Tunisie Telecom révèle une corrélation entre l'état civil des clients et leur taux de churn. Les non-mariés ont un taux de churn plus élevé (environ 11%) par rapport aux mariés (environ 7%). Cette différence suggère que l'état civil influence la résiliation d'abonnement. Une exploration approfondie des raisons de cette disparité est nécessaire pour élaborer des stratégies de rétention ciblées.

• La figure illustre la distribution du churn selon le nombre de réclamations.

  
 La figure indique clairement que le taux de désabonnement augmente fortement à partir de 5 réclamations.  
  
 • La figure illustre la distribution du churn selon l’activation des messages vocaux.  
  
 D’après la figure ci-dessus, on peut conclure que les clients qui n’utilisent pas la messagerie vocale ont un risque élevé de résiliation de leur abonnement.   
  
 • La figure illustre la distribution du churn selon le nombre des messages vocaux.  
 

En analysant les données des messages vocaux, nous pouvons conclure que les clients qui reçoivent plus de 15 messages vocaux présentent un risque plus élevé de résilier leur abonnement.

• La figure illustre la distribution du churn selon la période d’abonnement en jours.  


 Cette analyse montre que la majorité des clients ayant résilié leur abonnement l’ont fait au cours des 3 premiers mois. La période de 50 à 150 jours après l’abonnement se révèle être une période critique pour la rétention des clients, car le taux de churn y est particulièrement élevé.  
  
• La figure illustre la distribution du churn selon le coût d’appel des différents types d’appels.  


Les coûts totaux qui correspondent aux clients churneurs sont plus élevées par rapport aux autres qui restent. Cela pourrait montrer que les clients qui quittent l’entreprise ne sont pas satisfaits du montant qu’ils paient pour leur plan.