****

**Réalisé par : IKRAM Jennane**

**Prédiction de Désabonnement Télécom**

Introduction

La prévision du taux d'attrition dans le secteur des télécommunications consiste à identifier les clients susceptibles de quitter un service de télécommunications ou de s'en désabonner. Il s'agit d'une prévision importante pour les entreprises de télécommunications, car l'acquisition de nouveaux clients coûte souvent plus cher que la fidélisation des clients existants. En prévoyant le désabonnement des clients, les entreprises de télécommunications peuvent prendre des mesures proactives pour conserver leurs clients et minimiser la perte de revenus.

La prédiction du désabonnement des clients dans l'industrie des télécommunications implique l'utilisation de techniques d'apprentissage automatique pour analyser les données des clients et identifier des modèles ou des indicateurs associés au désabonnement. Ces modèles peuvent inclure des facteurs tels que les données démographiques des clients, les habitudes d'utilisation, les informations de facturation, les interactions avec le service client, etc.



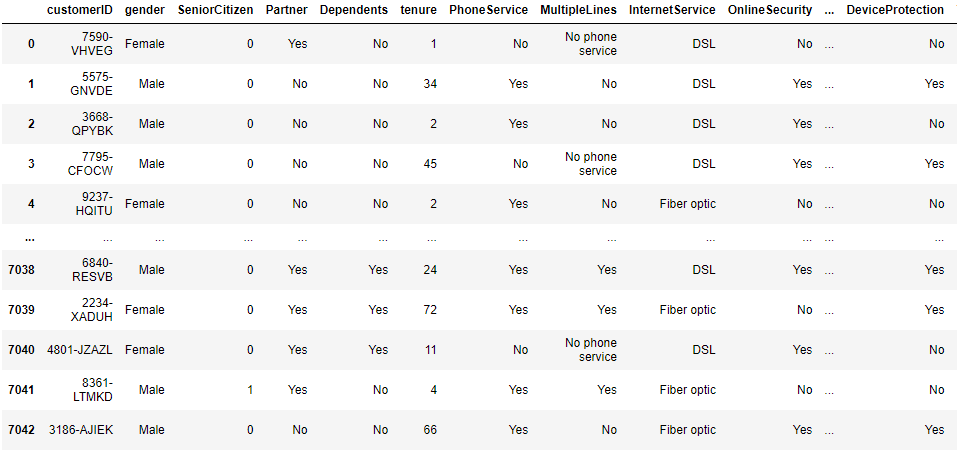
Pour prédire le taux de désabonnement dans les télécommunications à l'aide de l'exploration de données, on peut suivre une série d'étapes. Bien que la mise en œuvre spécifique puisse varier, voici un aperçu général du processus :

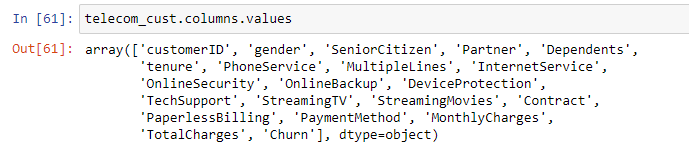
1. Compréhension de l'activité :

Comprendre clairement le problème et les objectifs de l'entreprise en ce qui concerne la prédiction du taux de désabonnement dans le secteur des télécommunications. Identifier les facteurs clés qui contribuent au désabonnement des clients et l'impact potentiel sur l'entreprise.

2. Collecte des données :

Il est difficile pour les chercheurs d'acquérir l'ensemble des données réelles des industries des télécommunications. Cela s'explique par le fait que les données privées des clients peuvent être utilisées à mauvais escient. Les chercheurs doivent donc prendre des précautions pour protéger la vie privée des individus tout en collectant des données pertinentes pour leurs études. Dans le cas de cette étude, l'ensemble de données a été acquis sur Kaggle.





3. Nettoyage et prétraitement des données :

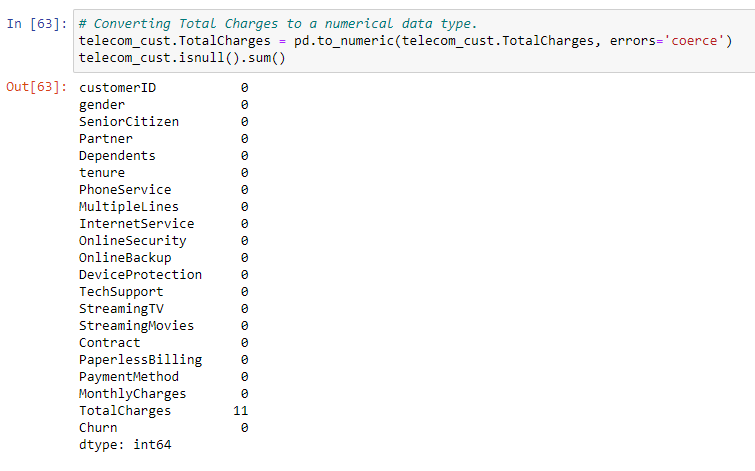
Effectuer une analyse exploratoire des données (AED) pour comprendre la structure et la qualité des données. Traiter les valeurs manquantes, les valeurs aberrantes et les incohérences dans l'ensemble des données. Prétraitez les données en transformant les variables, en normalisant ou en standardisant les caractéristiques et en codant les variables catégorielles.

**Valeurs manquantes**

Dans notre analyse des clients ayant changé de fournisseur de services de télécommunications, nous avons identifié 11 valeurs manquantes dans la colonne "Total des frais" de notre ensemble de données. Il est important de traiter ces valeurs manquantes avant de poursuivre l'analyse.

Les valeurs manquantes peuvent avoir un impact significatif sur la précision et la fiabilité de notre analyse. Il est donc essentiel de les traiter de manière appropriée. Une approche courante consiste à supprimer de l'ensemble de données les lignes comportant des valeurs manquantes. En supprimant ces lignes, nous nous assurons que notre analyse est basée sur des données complètes et fiables.

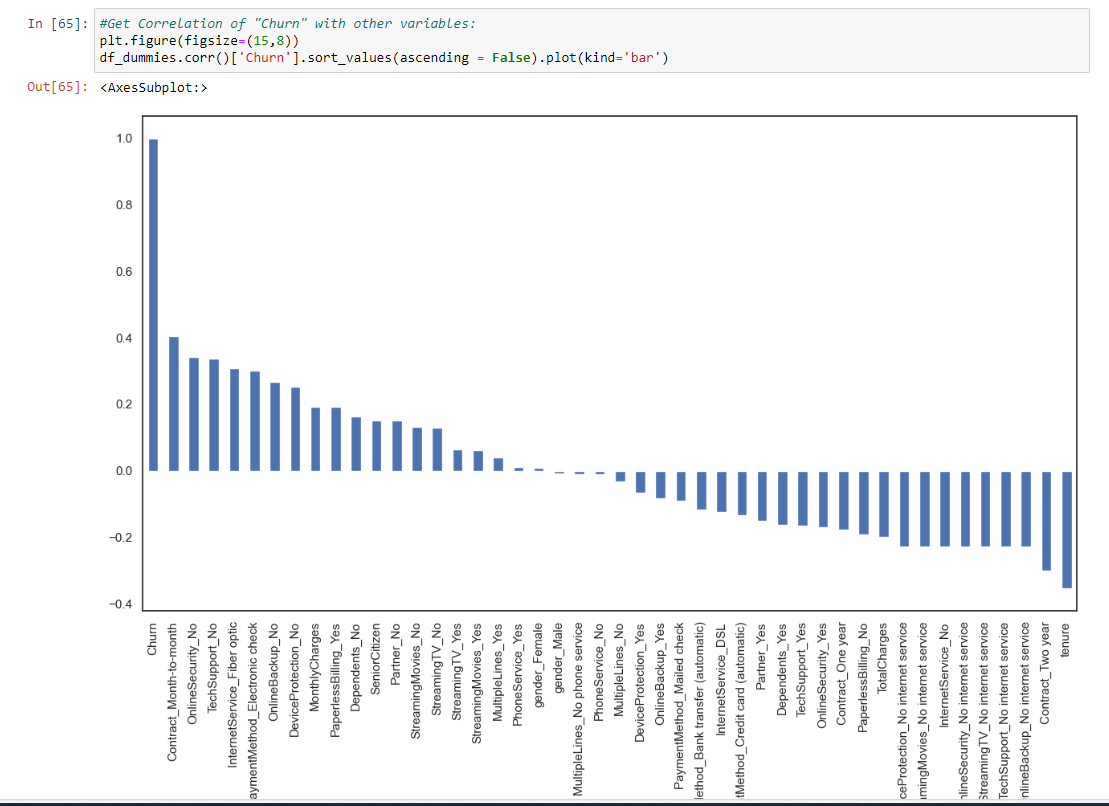
La suppression des lignes contenant des valeurs manquantes peut être effectuée à l'aide de différents langages et outils de programmation. Par exemple, en Python, nous pouvons utiliser la bibliothèque pandas pour filtrer les lignes contenant des valeurs manquantes.



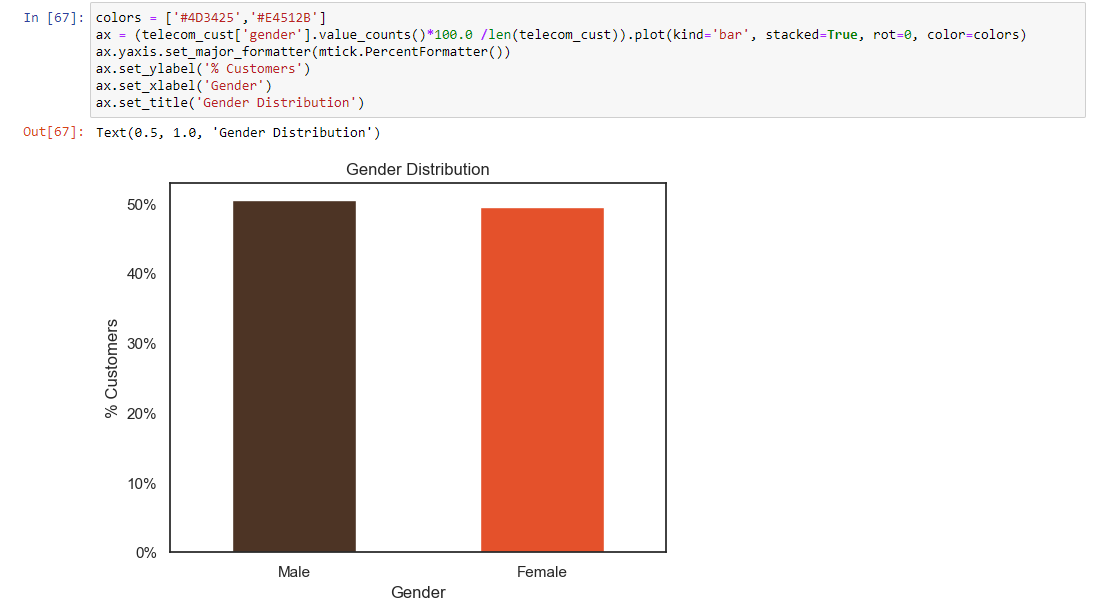
**Correlation**

Les contrats au mois, le manque de sécurité en ligne et l'absence d'assistance technique semblent avoir une relation positive avec la perte de clientèle. En revanche, une durée d'engagement plus longue et des contrats de deux ans semblent être associés négativement au désabonnement.

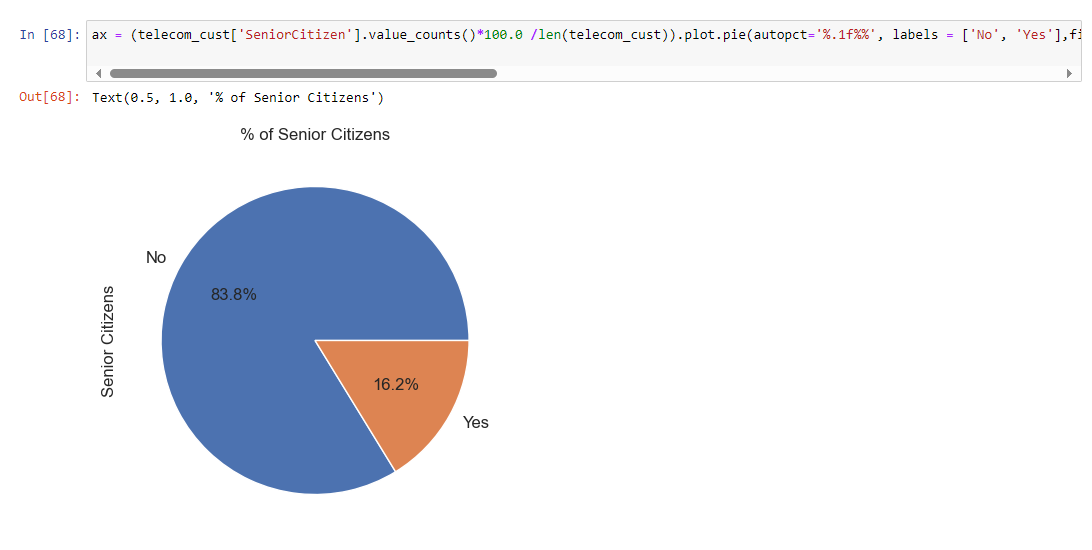
Il est intéressant de noter que les services tels que la sécurité en ligne, la télévision en continu, la sauvegarde en ligne et l'assistance technique qui ne nécessitent pas de connexion à l'internet sont négativement liés au taux de désabonnement.



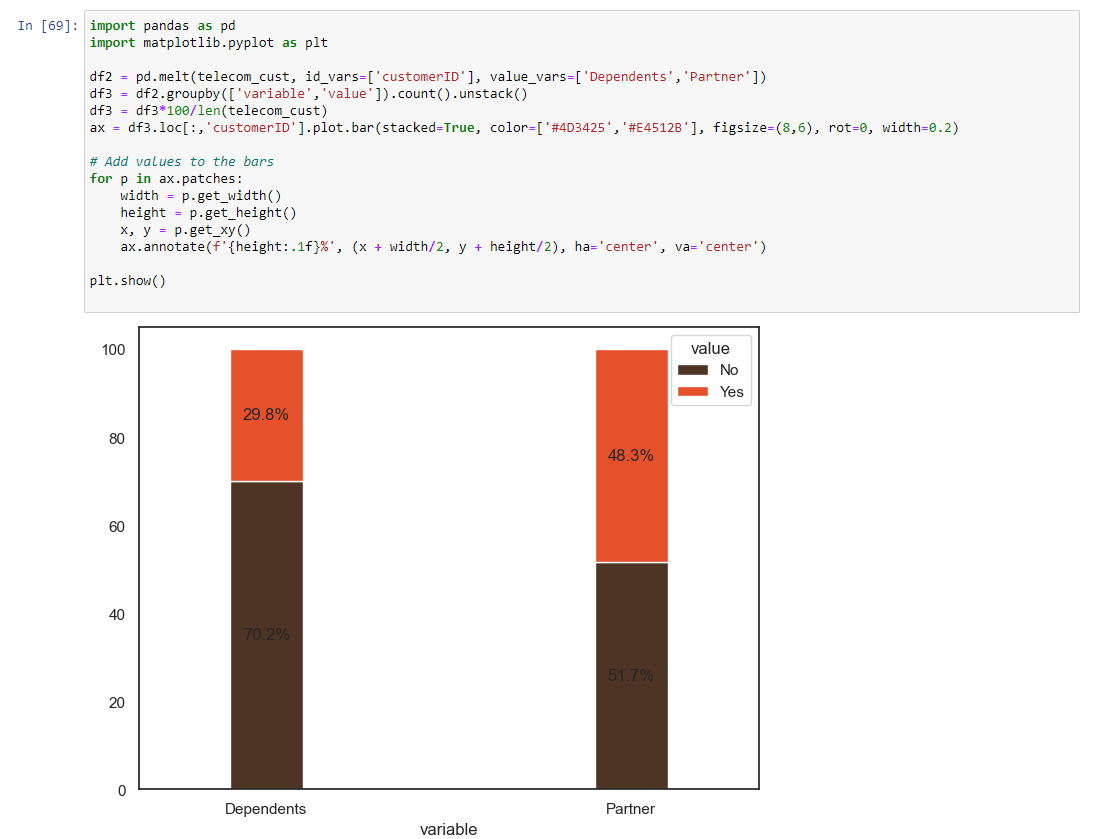
**Répartition par sexe** - Environ la moitié des clients de notre ensemble de données sont des hommes et l'autre moitié des femmes.



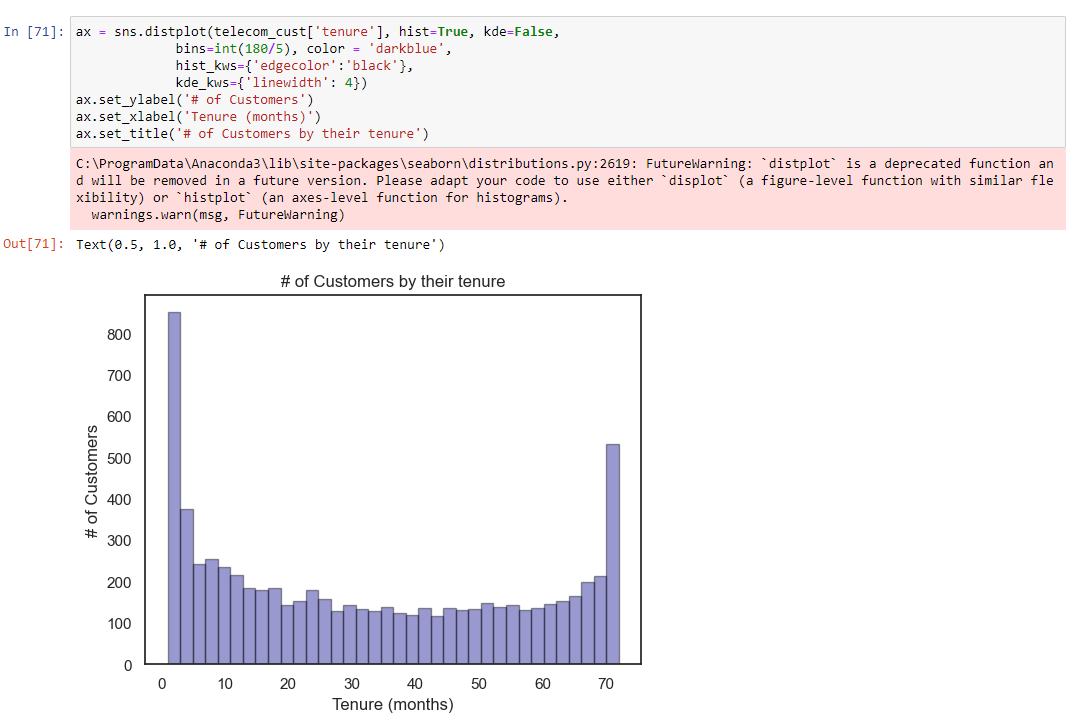
**Personnes âgées** - Seuls 16 % des clients sont des personnes âgées. La plupart de des clients sont donc des personnes plus jeunes.



**Statut de partenaire et de personne à charge** - Environ 50 % des clients ont un partenaire, tandis que seulement 30 % du total des clients ont des personnes à charge.

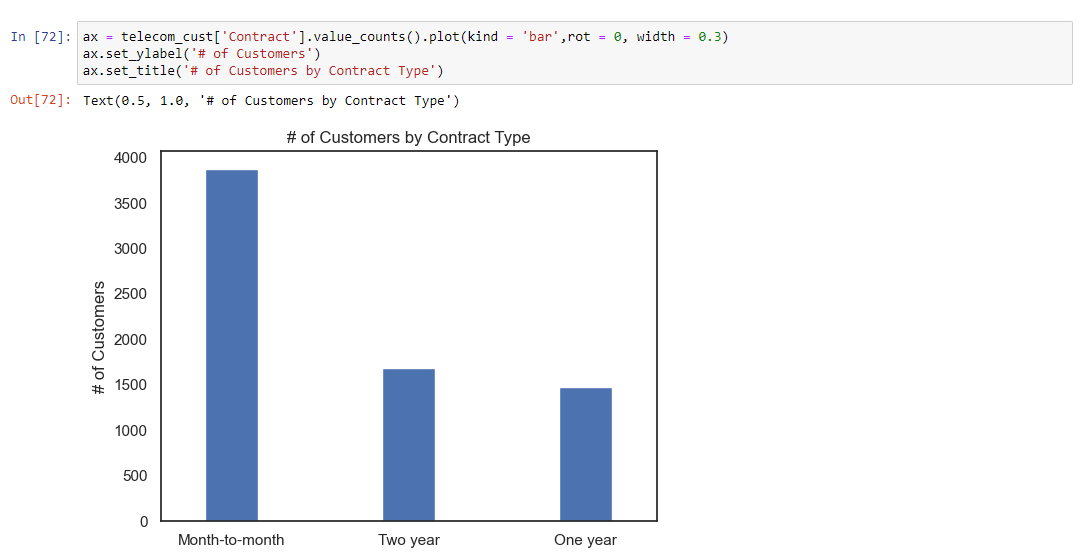


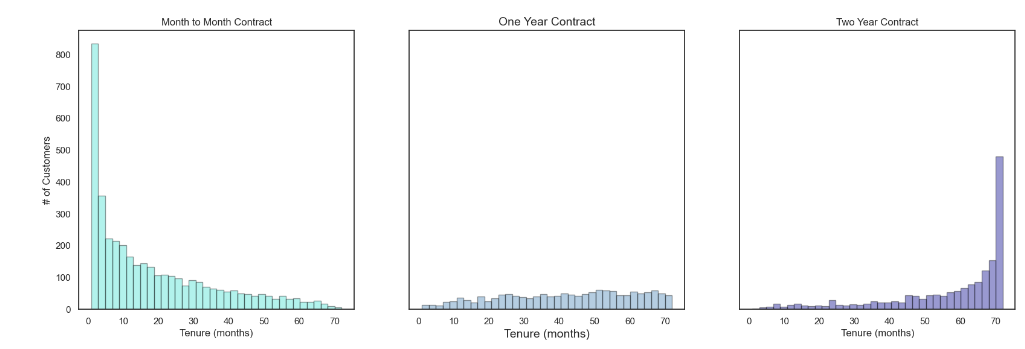
**Durée d'utilisation** : L'histogramme ci-dessous montre qu'un grand nombre de clients n'ont rejoint l'entreprise de télécommunications que depuis un mois, alors qu'un grand nombre d'entre eux sont là depuis environ 72 mois. Cela pourrait s'expliquer par le fait que les clients n'ont pas tous le même contrat. Ainsi, en fonction du contrat qu'ils ont souscrit, il pourrait être plus ou moins facile pour les clients de rester ou de quitter l'entreprise de télécommunications.



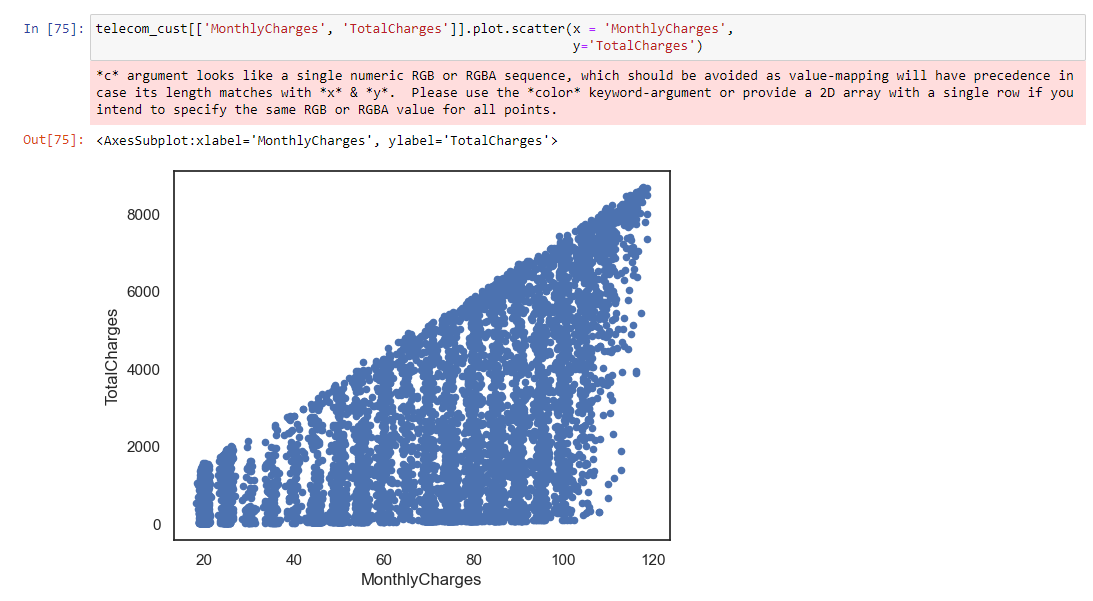
**Contrats** : Pour comprendre le graphique ci-dessus, examinons d'abord le nombre de clients selon les différents contrats. Ce graphique montre que la plupart des clients ont un contrat mensuel. Il y a un nombre égal de clients dans les contrats d'un an et de deux ans.

Il est intéressant de noter que la plupart des contrats mensuels ont une durée de 1 à 2 mois, alors que les contrats de 2 ans ont une durée d'environ 70 mois. Cela montre que les clients qui souscrivent un contrat plus long sont plus fidèles à l'entreprise et ont tendance à rester avec elle plus longtemps.

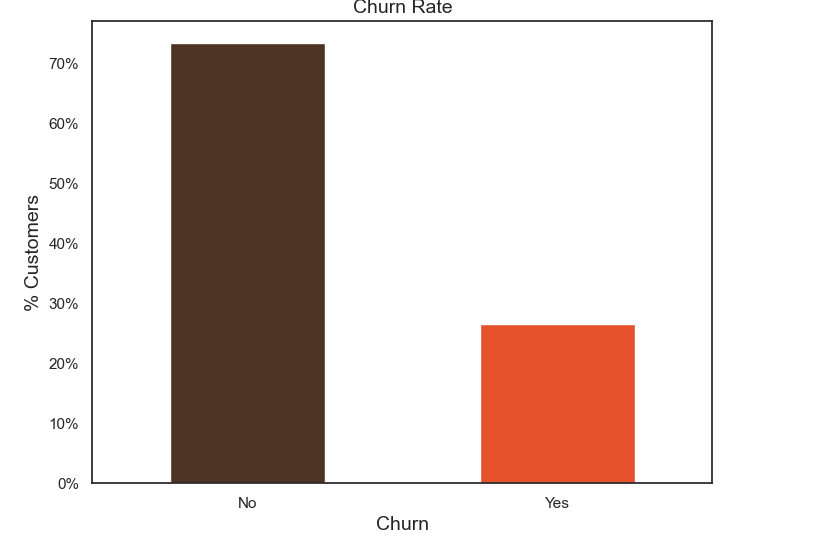




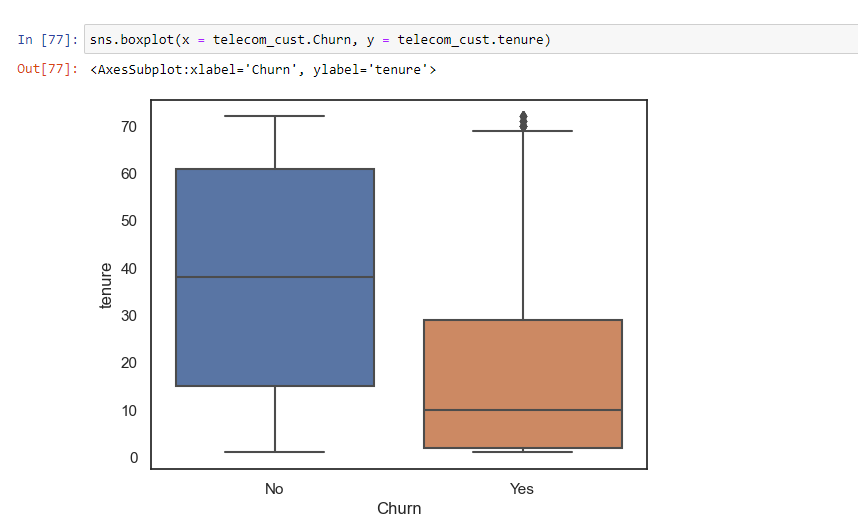
**Total charges** increases as the monthly bill for a customer increases.



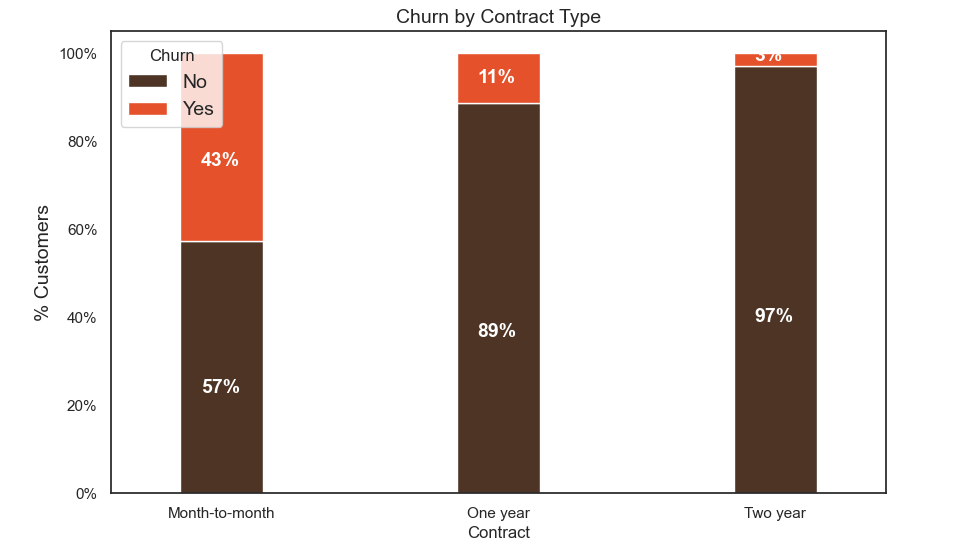
**Churn situation** : Dans les données, 74 % des clients ne se désabonnent pas. Il est clair que les données sont faussées, car nous nous attendrions à ce qu'une grande majorité des clients ne se désabonnent pas. Il est important de garder cela à l'esprit pour notre modélisation, car l'asymétrie pourrait entraîner un grand nombre de faux négatifs. Nous verrons dans la section consacrée à la modélisation comment éviter l'asymétrie des données.



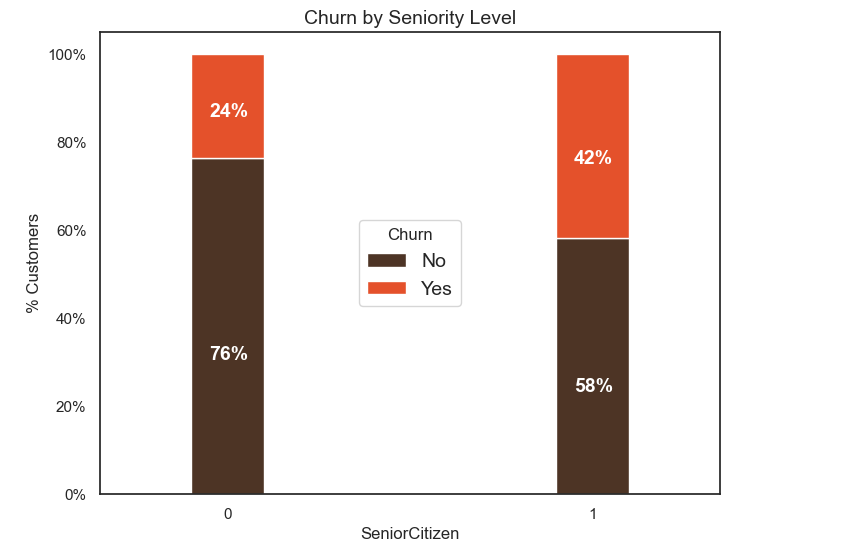
**Taux de désabonnement par rapport à l'ancienneté** : Comme le montre le graphique ci-dessous, les clients qui ne changent pas de fournisseur ont tendance à rester plus longtemps dans l'entreprise de télécommunications.



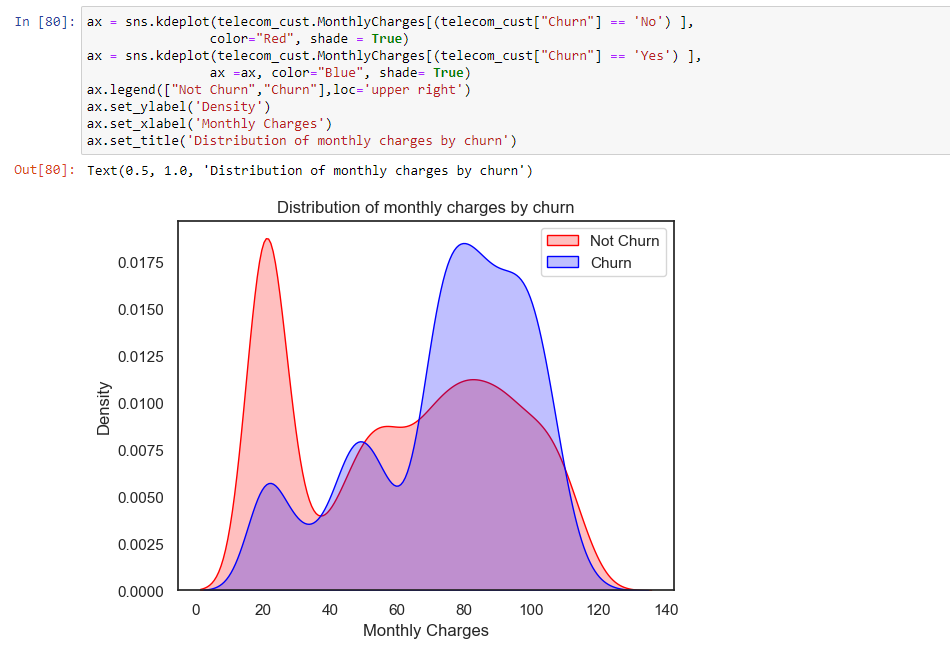
**Taux de désabonnement par type de contrat** : Comme dans le graphique de corrélation, les clients qui ont un contrat mensuel ont un taux de désabonnement très élevé.



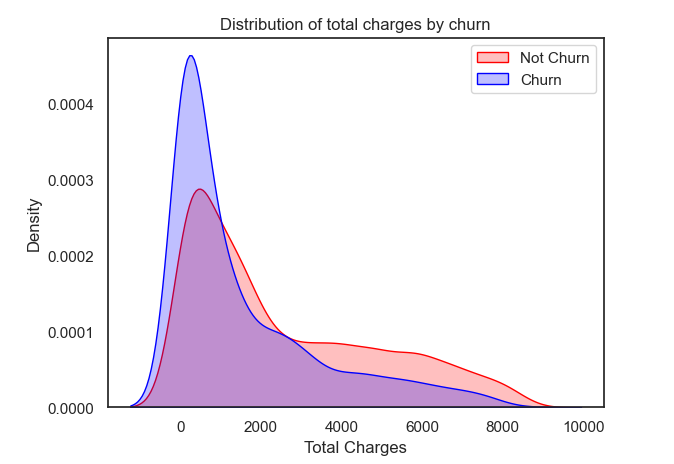
**Taux de désabonnement selon l'ancienneté** : Les seniors ont un taux de désabonnement presque deux fois plus élevé que la population plus jeune.



**Désabonnement en fonction des frais mensuels** : Le pourcentage de désabonnement est plus élevé lorsque les frais mensuels sont élevés.



**Taux de désabonnement en fonction des frais totaux** : Il semble que le taux de désabonnement soit plus élevé lorsque les frais totaux sont moins élevés.



4. Ingénierie des caractéristiques :

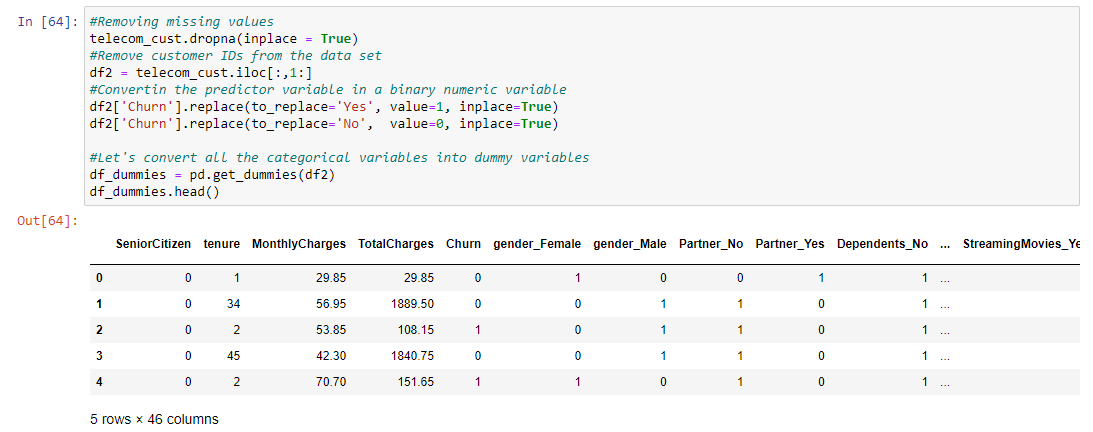
Identifier et créer de nouvelles caractéristiques susceptibles de prédire le taux de désabonnement. Il peut s'agir de dériver de nouvelles variables à partir de celles qui existent déjà ou de combiner plusieurs variables pour en extraire des informations significatives. L'ingénierie des caractéristiques joue un rôle crucial dans l'amélioration des performances du modèle de prédiction du désabonnement.

Suppression des valeurs manquantes : nous commençons par supprimer toutes les valeurs manquantes de l'ensemble de données à l'aide de la fonction `dropna()`. Cette fonction supprime les lignes contenant des valeurs manquantes de l'ensemble de données. Le paramètre `inplace=True` garantit que les modifications sont apportées directement à l'ensemble de données original.

Suppression des identifiants des clients : L'étape suivante consiste à supprimer les identifiants des clients de l'ensemble de données. Le code crée un nouveau DataFrame appelé `df2` en sélectionnant toutes les colonnes sauf la première (`telecom\_cust.iloc[ :,1 :]`). Cette étape permet d'exclure les identifiants des clients de l'analyse.

Conversion de la variable prédictive : Le code convertit ensuite la variable prédictive "Churn" en une variable numérique binaire. Il remplace les valeurs "Yes" par 1 et "No" par 0 à l'aide de la fonction `replace()`. Cette étape est nécessaire car de nombreux algorithmes d'apprentissage automatique nécessitent des entrées numériques.

Conversion des variables catégorielles en variables nominales : La dernière étape consiste à convertir toutes les variables catégorielles de l'ensemble de données en variables muettes. Pour ce faire, on utilise la fonction `get\_dummies()` de la bibliothèque pandas. Les variables muettes sont des variables binaires qui représentent la présence ou l'absence d'une catégorie particulière. Cette étape est importante car les algorithmes d'apprentissage automatique ne peuvent généralement pas traiter directement les variables catégorielles.



5. Développement du modèle :

Sélectionner un algorithme d'apprentissage automatique approprié ou un ensemble d'algorithmes pour la prédiction du taux d'attrition. Les algorithmes couramment utilisés comprennent la régression logistique, les arbres de décision, les forêts aléatoires, les machines à vecteurs de support (SVM) et les méthodes de gradient boosting. Entraînez le modèle à l'aide de l'ensemble de données préparé et réglez les hyperparamètres pour optimiser ses performances.

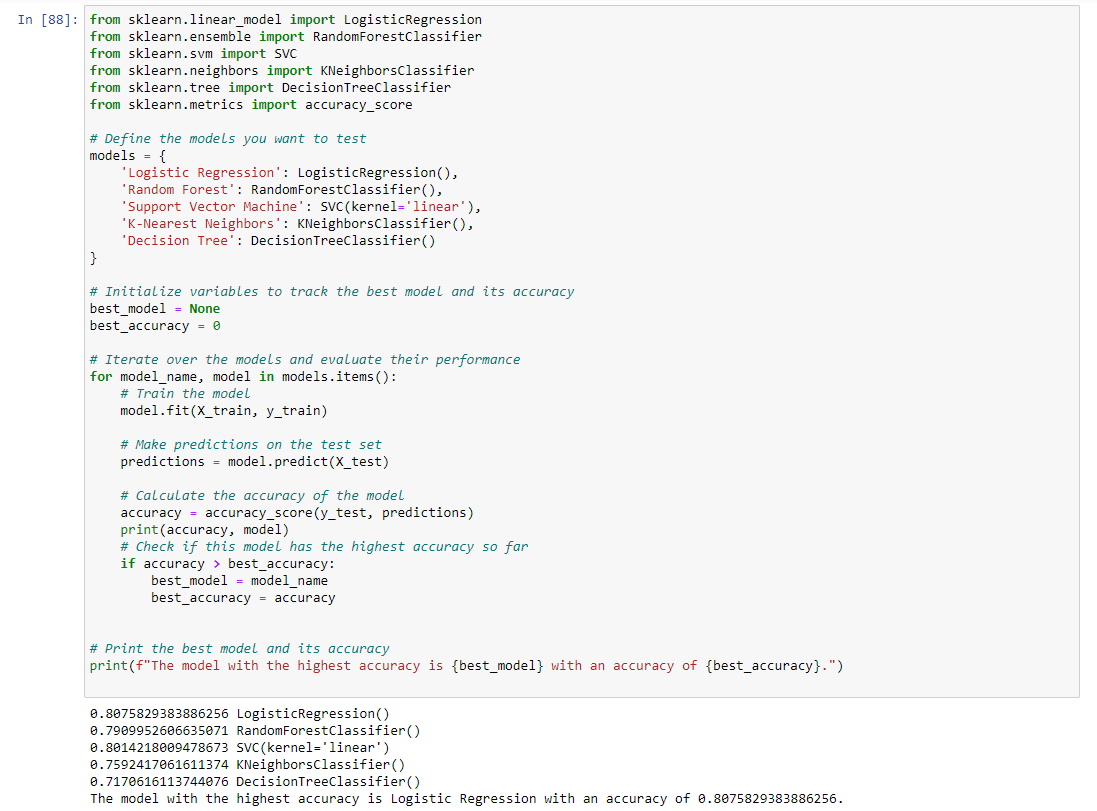
6. Évaluation du modèle :

Évaluer les performances du modèle de prédiction du désabonnement à l'aide de mesures d'évaluation appropriées telles que l'exactitude, la précision, le rappel, le score F1 et l'aire sous la courbe caractéristique d'exploitation du récepteur (AUC-ROC). Utilisez des techniques telles que la validation croisée ou la validation d'attente pour évaluer la capacité de généralisation du modèle.



7. Interprétation et perspectives :

Interpréter les résultats du modèle de prédiction du désabonnement pour mieux comprendre les facteurs qui contribuent au désabonnement des clients. Identifier les caractéristiques les plus importantes et leur impact sur la probabilité de désabonnement. Ces informations peuvent aider à concevoir des stratégies de fidélisation efficaces et à prendre des décisions commerciales fondées sur des données.



8. Déploiement et suivi :

Déployer le modèle de prévision du taux de désabonnement dans l'environnement de production, en l'intégrant aux systèmes ou processus existants. Contrôler en permanence les performances du modèle et le mettre à jour lorsque de nouvelles données sont disponibles. Évaluer régulièrement l'efficacité du modèle et l'affiner si nécessaire.