

METHODES DE CLASSIFICATION

PROJET FINAL



Sujet: Analyse de l'attrition de la clientèle d'un opérateur téléphonique

Description : Cette étude est une mise en récit sur les données de la clientèle d'un opérateur téléphonique pour communiquer tout ce que vous y avez observé après analyse de l'attrition. Plus particulièrement, le but de cette étude est (i) d'identifier et visualiser les facteurs qui contribuent à l'attrition des clients, (ii) de déterminer le profil des clients qui vont potentiellement se désabonner, (iii) et enfin, de proposer des stratégies de fidélisation basées sur les résultats

Objectifs spécifiques :

- Faire une analyse exploratoire complète pour identifier les principaux facteurs d'attrition
- Faire une modélisation statistique pour construire un modèle prédictif des clients qui vont potentiellement se désabonner
- Définir des data-driven stratégies de fidélisation (basées sur les résultats obtenus)

Livrable:

- Notebook (créé avec Google Colab) à rendre sur Google Classroom

Dictionnaire de données:

N°	Variables	Signification	Type
1	user_id	Identifiant unique du client	Qualitative
2	TENURE	la durée dans le réseau	Qualitative
3	TOP_PACK	les pack les plus actifs	Qualitative
4	REGION	la localité de chaque client	Qualitative
5	MONTANT	montant de recharge	Numérique
6	FREQUENCE_RECH	nombre de fois que le client a fait une recharge	Numérique
7	REVENUE	revenu mensuel de chaque client	Numérique
8	ARPU_SEGMENT	revenu sur 90 jours/3	Numérique
9	FREQUENCE	nombre de fois que le client a fait un revenu	Numérique
10	DATA_VOLUME	nombre de connexions	Numérique
11	ON_NET	Appel inter expresso	Numérique
12	ORANGE	Appel vers orange	Numérique
13	TIGO	appel vers Tigo	Numérique
14	ZONE1	appel vers les zone1	Numérique
15	ZONE2	appel vers les zone2	Numérique
16	MRG	Un client qui fait du vas	Numérique
17	REGULARITY	nombre de fois que le client est actif pendant 90 jours	Numérique
18	FREQ_TOP_PACK	nombre de fois que le client a activé les packages top pack	Numérique
19	CHURN	variable à prédire	Numérique

Outil/Langage à utiliser:

- Google colab
- Python

Méthodologie à suivre:

- CRISP-DM (Cross Industry Standard Process for Data Mining)

Tutoriels: https://colab.research.google.com/drive/16dugRdAvo3DJyBHbFYB4TNHm3D_Xuct9?usp=sharing

Annexe : Analyse exploratoire tirée d'un mémoire

La visualisation des données nous permet de voir l'histoire qui se cache derrière les données. Il est basé sur la citation qui dit « une image vaut mieux que mille mots ». Elle est souvent utilisée pour démontrer des performances, communiquer des tendances, comprendre l'impact de nouvelles stratégies, montrer des relations et bien plus encore.

1. Présentation du dictionnaire des variables

Les données fournies par l'entreprise de télécommunication est un fichier contenant des données clients.

Le fichier comprend 19 variables dont 15 variables numériques et 04 variables catégoriques.

N°	Variables	Signification	Type
1	user_id	Identifiant unique du client	Qualitative
2	TENURE	la durée dans le réseau	Qualitative
3	TOP_PACK	les pack les plus actifs	Qualitative
4	REGION	la localité de chaque client	Qualitative
5	MONTANT	montant de recharge	Numérique
6	FREQUENCE_RECH	nombre de fois que le client a fait une recharge	Numérique
7	REVENUE	revenu mensuel de chaque client	Numérique
8	ARPU_SEGMENT	revenu sur 90 jours/3	Numérique
9	FREQUENCE	nombre de fois que le client a fait un revenu	Numérique
10	DATA_VOLUME	nombre de connexions	Numérique
11	ON_NET	Appel inter expresso	Numérique
12	ORANGE	Appel vers orange	Numérique
13	TIGO	appel vers Tigo	Numérique
14	ZONE1	appel vers les zone1	Numérique
15	ZONE2	appel vers les zone2	Numérique
16	MRG	Un client qui fait du vas	Numérique
17	REGULARITY	nombre de fois que le client est actif pendant 90 jours	Numérique
18	FREQ_TOP_PACK	nombre de fois que le client a activé les packages top pack	Numérique
19	CHURN	variable à prédire	Numérique

Tableau 1 : Dictionnaire des variables

1. Analyse univariée

L'analyse univariée consiste à faire une investigation statistique des données pour se familiariser avec elles. On s'intéresse particulièrement à la répartition des valeurs et leurs ordres de grandeur que nous résumons dans le tableau ci-dessous:





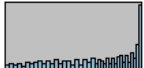


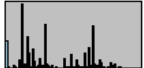





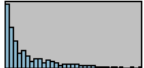


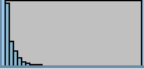






















 MRG		 Discrète	--	--	--	--	--	1	28905
 REGULARITY		 Intervalle	1.000	62.000	41.520	18.488	-0.553	--	28905
 TOP_PACK		 Discrète	--	--	--	--	--	85	26904
 FREQ_TOP_PACK		 Discrète	--	--	--	--	--	118	26904
 Frequence_rech		 Intervalle	1	106	13.100	14.054	1.910	--	28489
 Revenu_		 Intervalle	1	147739	6193.858	7408.770	3.297	--	28905
 Orange_		 Intervalle	0	5841	98.712	210.721	6.568	--	25367
 Tigo_		 Intervalle	0	1594	23.560	61.152	9.343	--	18046
 MontantRecharge_		 Intervalle	50	197400	6154.154	7345.254	3.733	--	28489
 Zone1_		 Intervalle	0	1469	9.011	49.786	17.768	--	3572
 Zone2_		 Intervalle	0	932	7.658	29.477	15.992	--	3095
 Volume de Donnée...		 Intervalle	0	926547	3962.917	14857.329	30.376	--	28905
 CHURN_		 Ensemble	--	--	--	--	--	2	28905

Tableau 2 : Audit complet des données

Valeur ▲	Proportion	%	Comptage
Loyal	<div></div>	95,44	27586
Perdu	<div></div>	4,56	1319

CHURN_

Loyal Perdu

Figure 1: Distribution de la variable churn

Commentaire : à travers ce graphique qui illustre la distribution de la variable churn nous constatons que le jeu de données est déséquilibré avec **95 ;44** qui représente les clients loyaux contre **4 ;56** pour les clients perdus

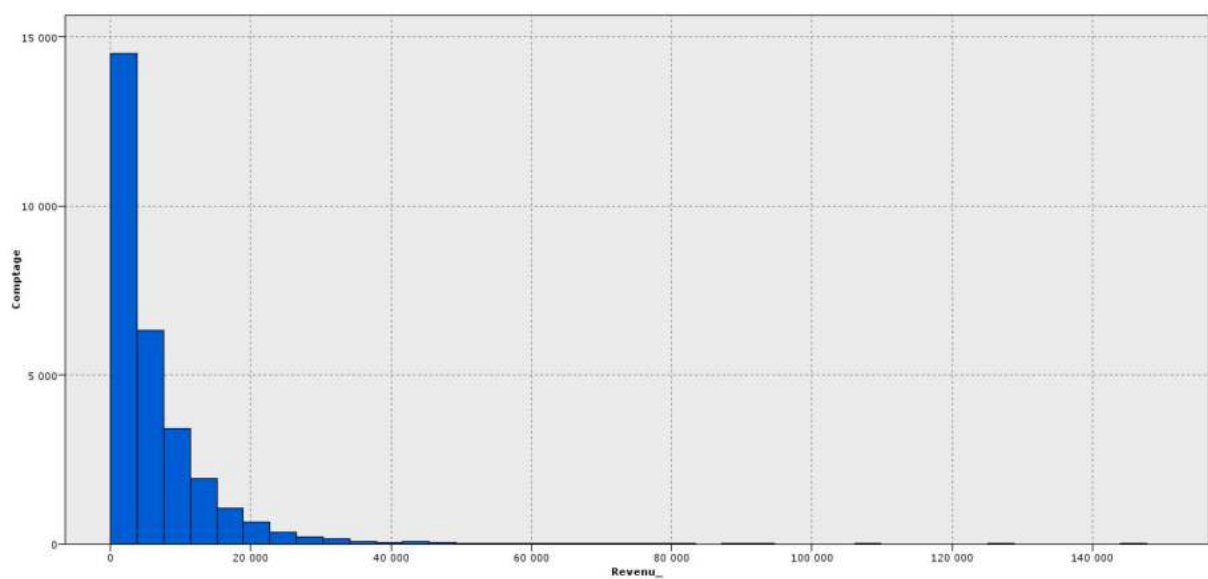


Figure 2: Distribution de la variable revenue

Commentaire : Ce graphique montre que la majorité des clients ont un niveau de revenu faible.

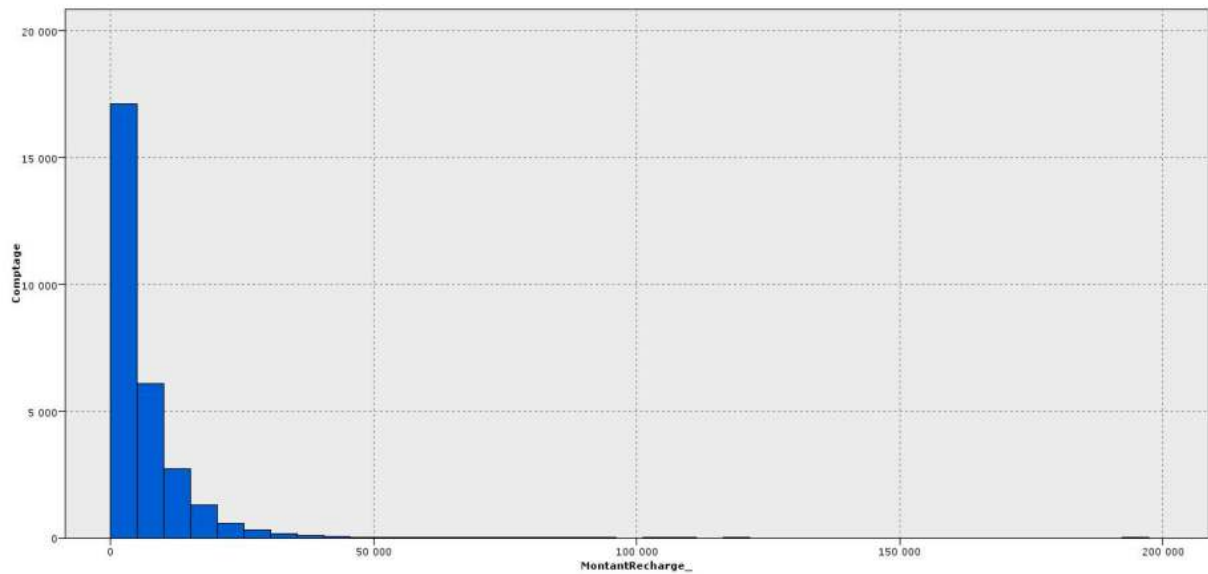


Figure 3: Distribution de la variable montant recharge

Commentaire : Cette représentation nous informe sur le fait que la majorité des clients présente un niveau de Montant_recharge assez bas.

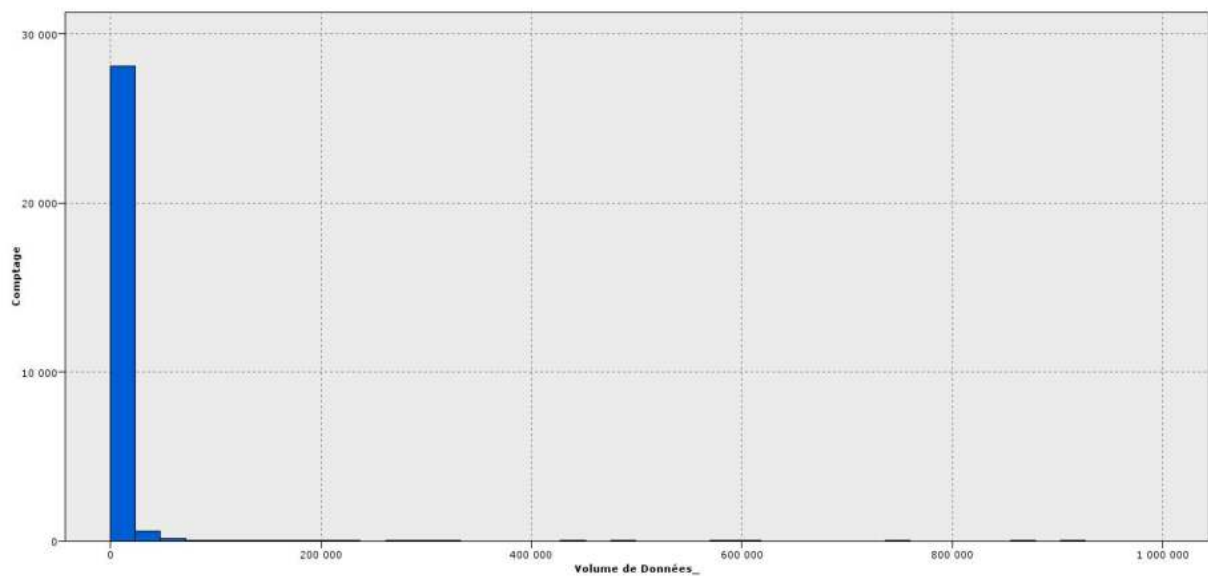


Figure 4: Distribution de la variable volume de données

Commentaire : De même au niveau de cette représentation nous remarquons que la plupart des clients effectuent une consommation très basse des volumes de données.

2. Analyse multivariée

Dans cette section, nous analysons les corrélations entre la variable cible churn et les autres variables prédictives. En d'autres termes, nous aimerions savoir entre autres si:

- Le nombre de fois qu'un client fait une recharge est lié à l'attrition ?
- Le revenu est lié à l'attrition ?
- Le volume de connexion a un impact sur l'attrition ?
- L'ancienneté au réseau a un impact sur l'attrition ?
- Le montant des recharges est-il lié à l'attrition ?
- Les appels effectués vers les opérateurs Orange ou Tigo sont t'ils liés à l'attrition

Présentation des résultats de l'analyse bivariable

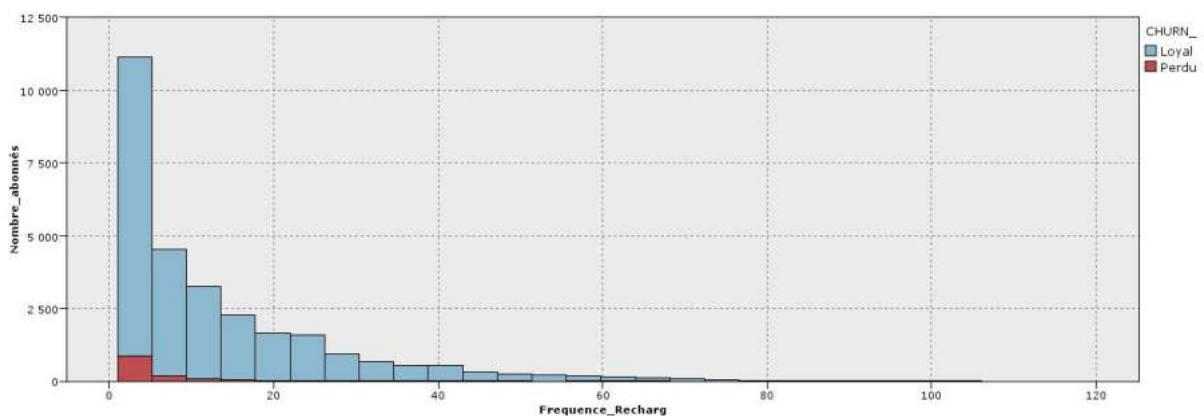


Figure 5: Est-ce que le nombre de fois qu'un client fait une recharge est lié à l'attrition ?

Commentaire : Ce graphique montre que le phénomène du churn est plus en évidence chez les clients à faible fréquence de recharge.

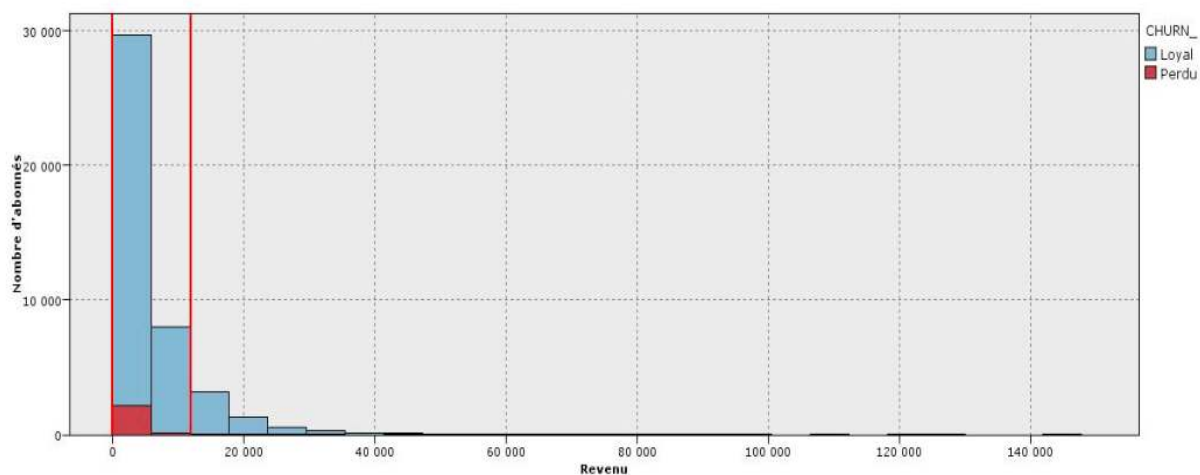


Figure 6: Est-ce que le revenu est lié à l'attrition ?

Commentaire: Ce graphique montre que le phénomène du churn est plus observé chez les abonnés à faible revenu

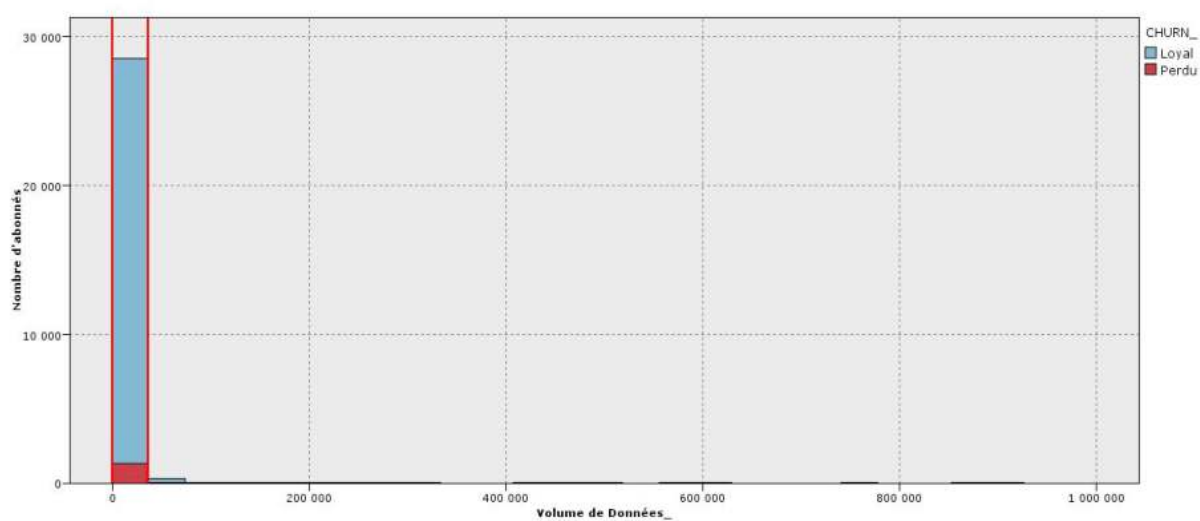


Figure 7: Est-ce que le volume de connexion a un impact sur le churn ?

Commentaire : Ce graphique montre que le phénomène du churn est plus observé chez les abonnés qui utilisent moins de données mobiles.

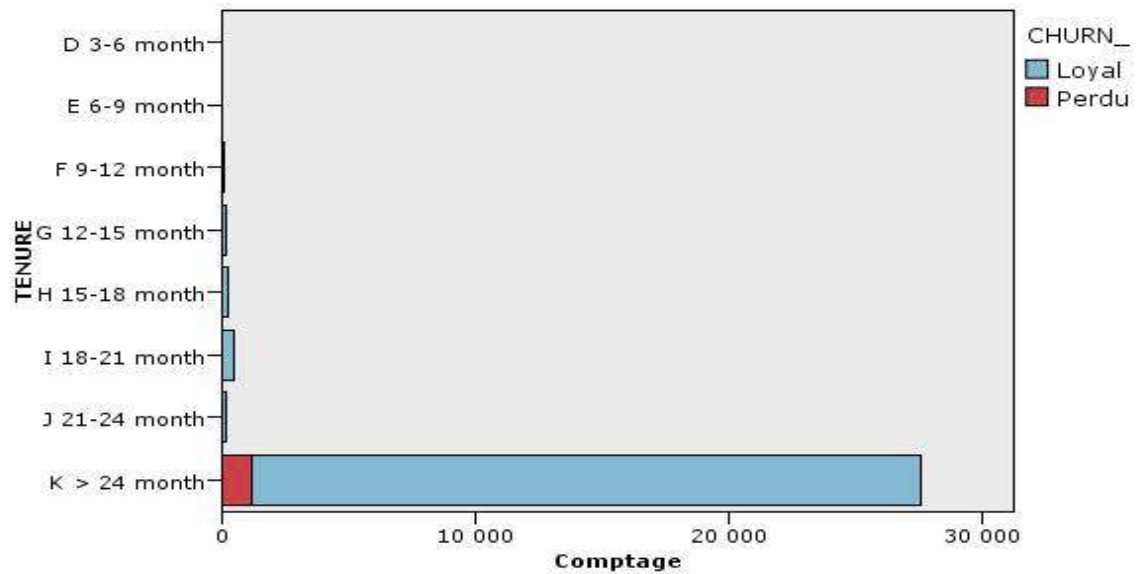


Figure 8: Est-ce que la tenure explique le churn ?

Commentaire : Le phénomène du churn est plus observé chez les abonnés avec une ancienneté de plus de 24 mois dans le réseau.

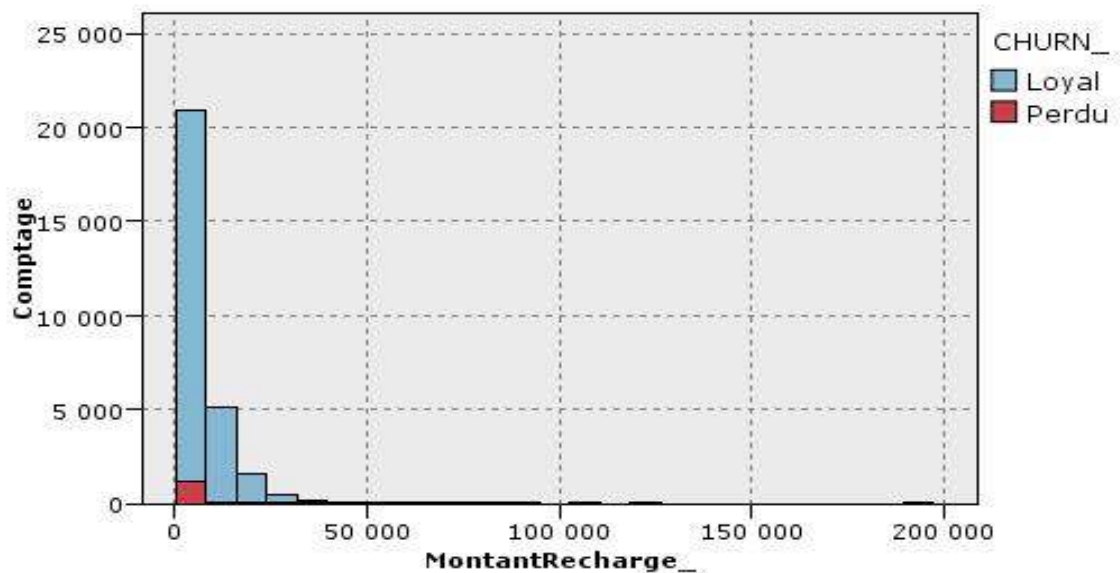


Figure 9: Est-ce que le montant de la recharge explique le churn ?

Commentaire : Ce graphique nous informe que le phénomène du churn est davantage observé chez les clients qui ont en l'occurrence des montants de recharge très bas.

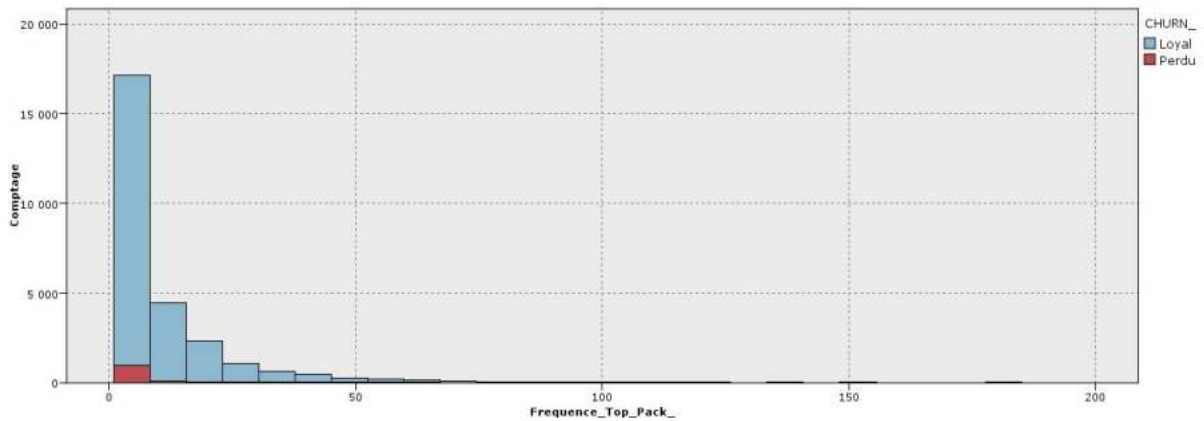


Figure 10: : Est-ce que le nombre de fois où le client a activé les packages Top pack est lié à l'attrition ?

Commentaire : à travers ce graphique nous constatons que le phénomène du churn est plus en exergue chez les clients dont la fréquence d'activation du package Top_Pack est très faible.

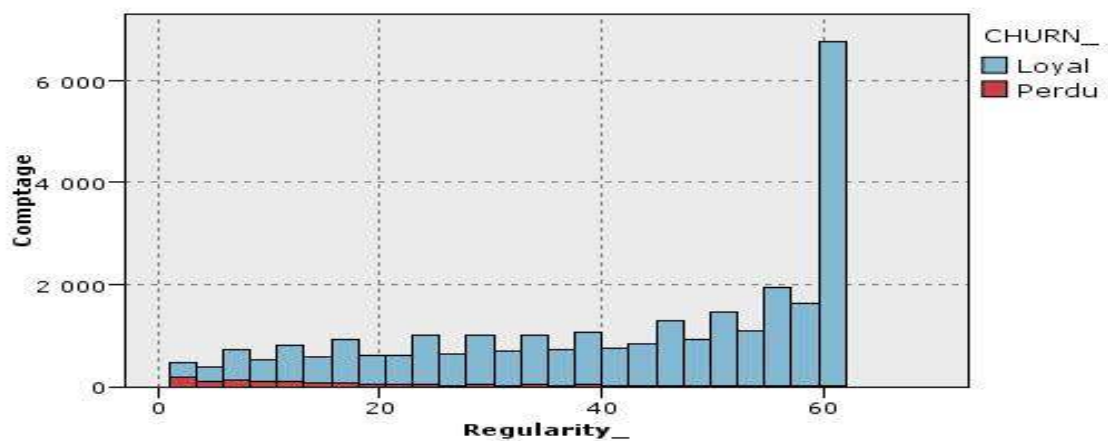


Figure 11: Est-ce que le nombre de fois qu'un client est actif pendant une période de 90 jours est lié à l'attrition ?

Commentaire : Cette représentation nous informe sur le fait que le phénomène du Churn se manifeste au niveau des clients dont le nombre de fois qu'ils ont été actifs pendant 90 jours est faible

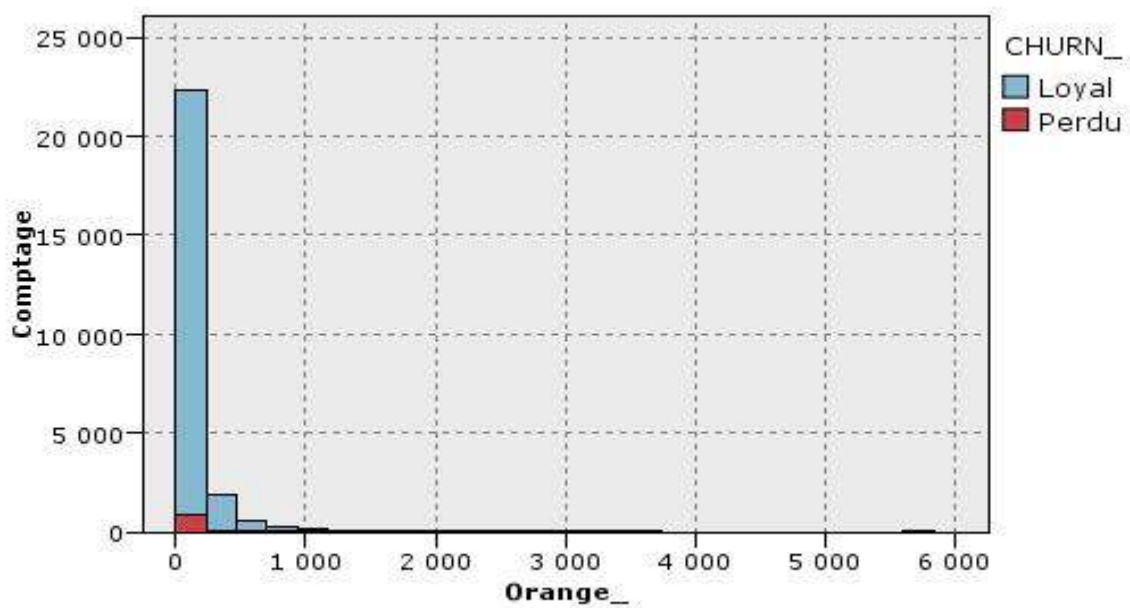


Figure 12: Est-ce que le nombre de fois qu'un client a effectué des appels vers le l'opérateur Orange est lié au churn?

Commentaire : Le phénomène du churn est plus observé chez les abonnés ayant effectué le plus grand nombre d'appels vers le réseaux orange.

Présentation des résultats de l'analyse multivariée

Représentation des abonnés suivant leur revenu et le volume de données mobiles utilisés ?

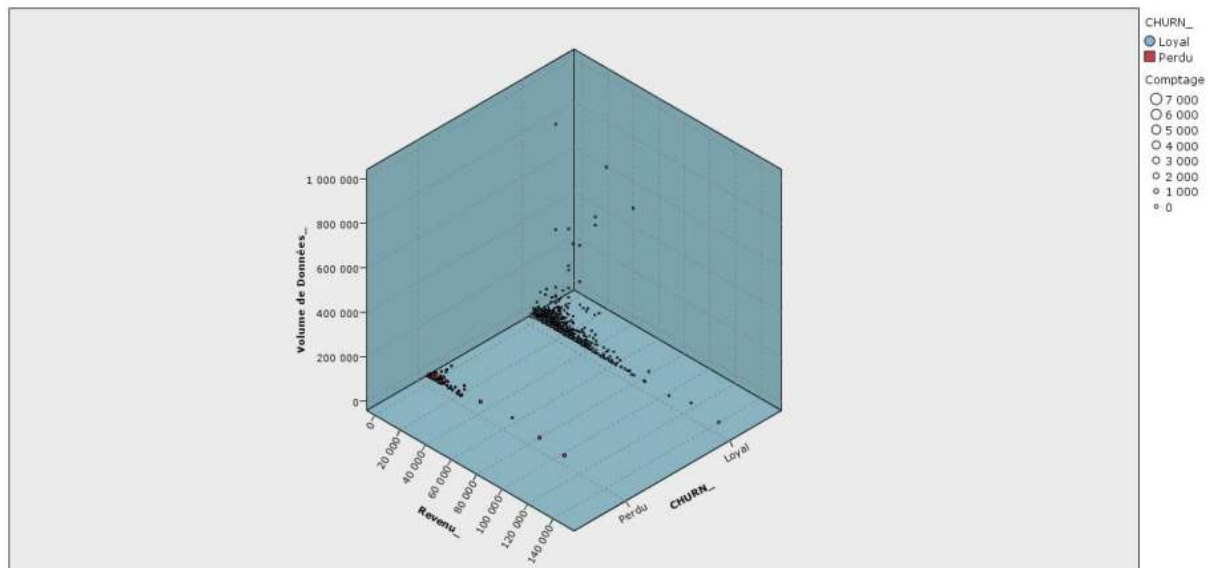


Figure 13: Diagramme de dispersion du churn (3D) suivant le revenu et le volume de données

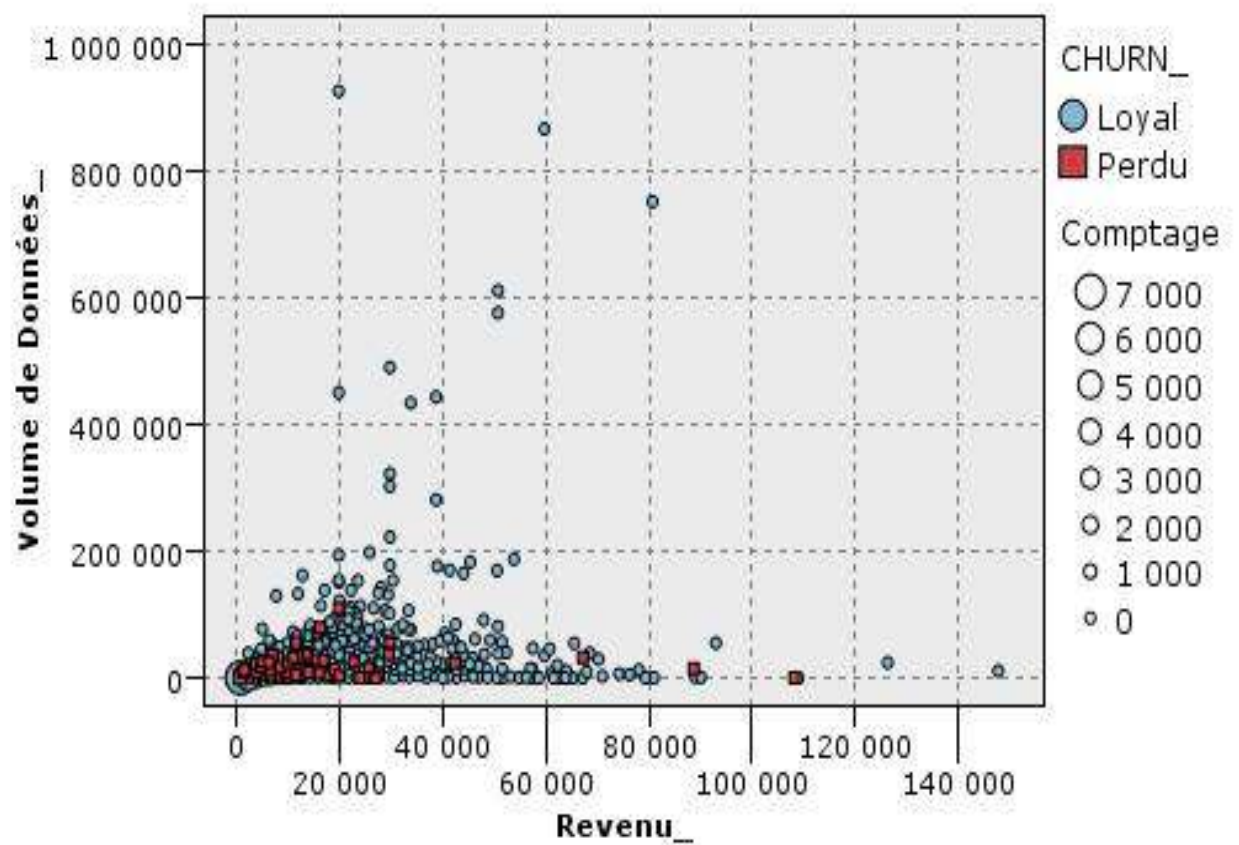


Figure 14: Diagramme de dispersion du churn (2D) suivant le revenu et le volume de données

Commentaire : Ce diagramme confirme les deux visuels précédents à savoir que l'attrition est plus observée chez les abonnés à faible revenu et utilisant moins de données mobiles.

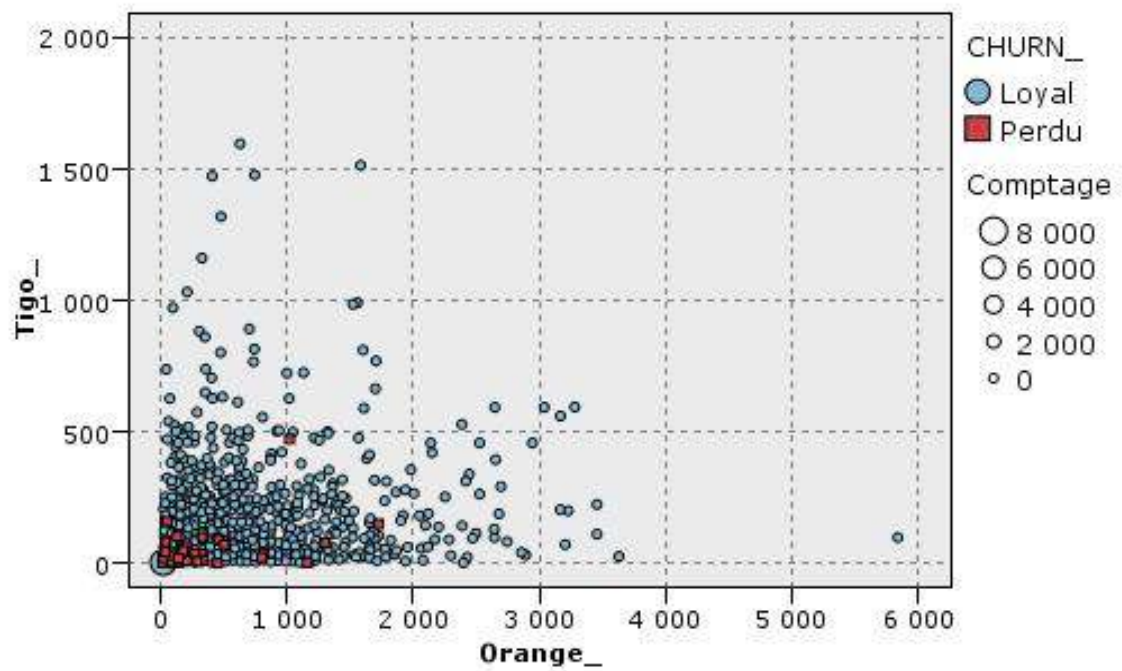


Figure 15: Est-ce que le nombre d'appels vers Orange ou Free explique le churn ?

Commentaire : Ce graphique infirme l'hypothèse comme quoi plus l'abonné passe des appels vers les numéros Tigo ou Orange, plus il a tendance à "churner". Toutefois, on observe que les appels vers Orange suscitent plus le départ que les appels vers Tigo

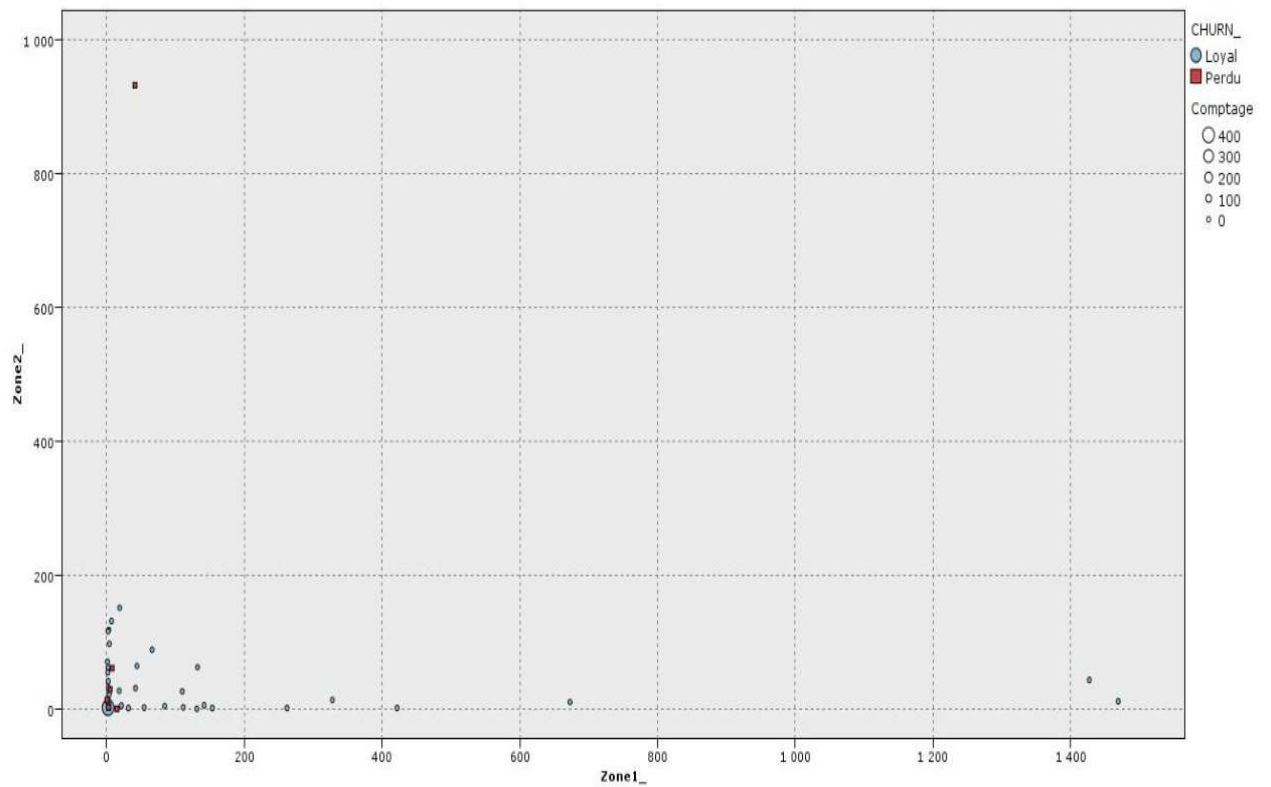


Figure 16: Est-ce que le nombre d'appels à la Zone 1 ou Zone 2 explique l'attrition ?

Commentaire : Ce diagramme illustre que l'attrition est plus liée aux appels à la Zone 2

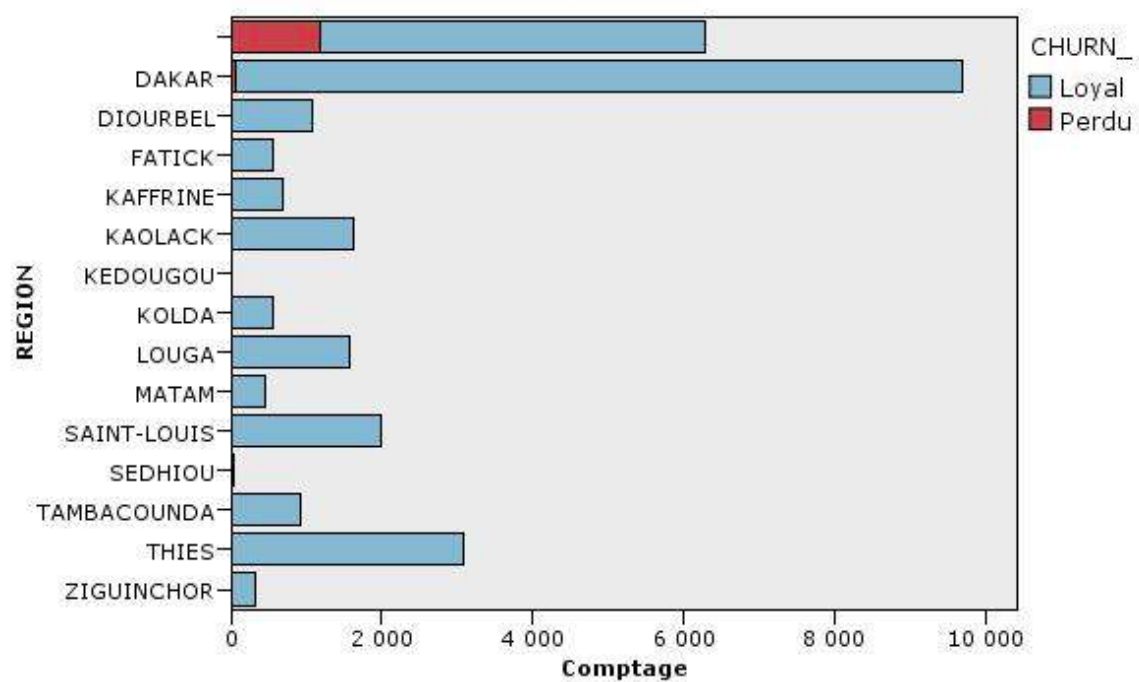


Figure 17: Est-ce que la variable région explique l'attrition ?

Commentaire : Est-ce que l'attrition est liée à la localité ? (Impossible de répondre à la question parce que les données sont anonymisées. En effet, pour les clients qui Perdu, la localité n'est pas renseignée)