

Projet data: départs suspects de clients dans une banque

Contexte:

Nous allons étudier un ensemble de données issu d'un fichier d'une banque, répertoriant 10000 clients choisis au hasard, cet ensemble de données nous fournit :

- le numéro de client
- le nom
- les crédits,
- l'âge, le sexe
- le nombre de produits (nombre de produits détenus)
- s'il a une carte de crédit
- s'il est actif sur son compte, l'estimation de son salaire
- s'il a quitté ou pas la banque.

Le problème avec cette banque est que de nombreux clients ont quitté cette banque et ils aimeraient savoir pourquoi.

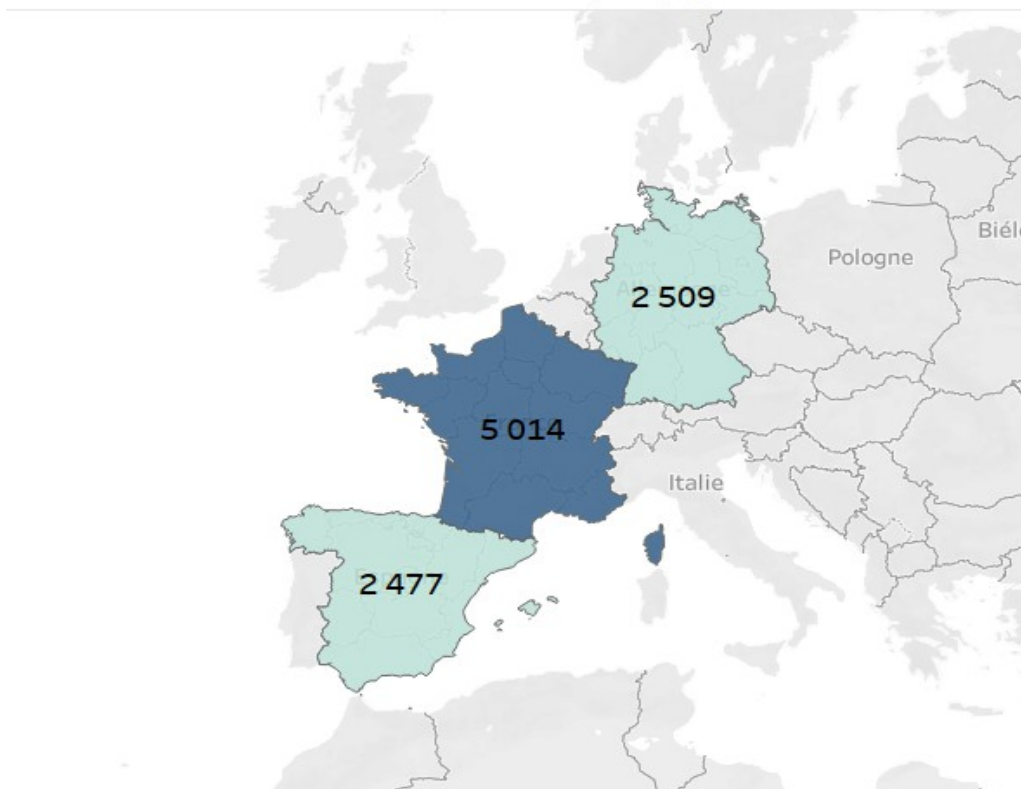
Déroulement du projet

1ere Partie: Un peu de DataMining

1)Question

Combien avons nous de clients dans chaque pays ?

Nombre de clients par pays

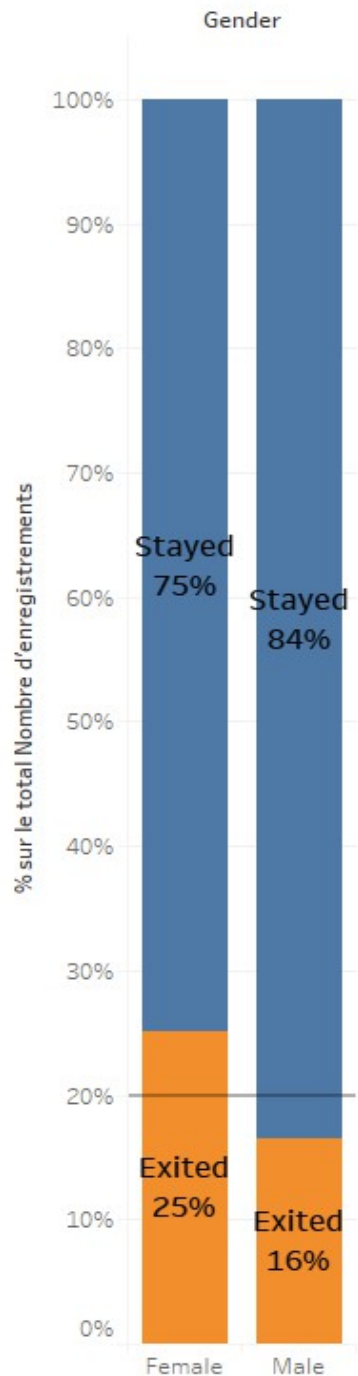


Nous pouvons voir que **5014 personnes** vivent en **France**, **2509** en **Allemagne** et **2477** en **Espagne**. Un **total de 10000 personnes** sur **3 pays**.

2)Question

Existe t-il une **correlation** entre le **nombre de départ** et le **sexe**?

Pourcentage de personnes parti ou resté par genre

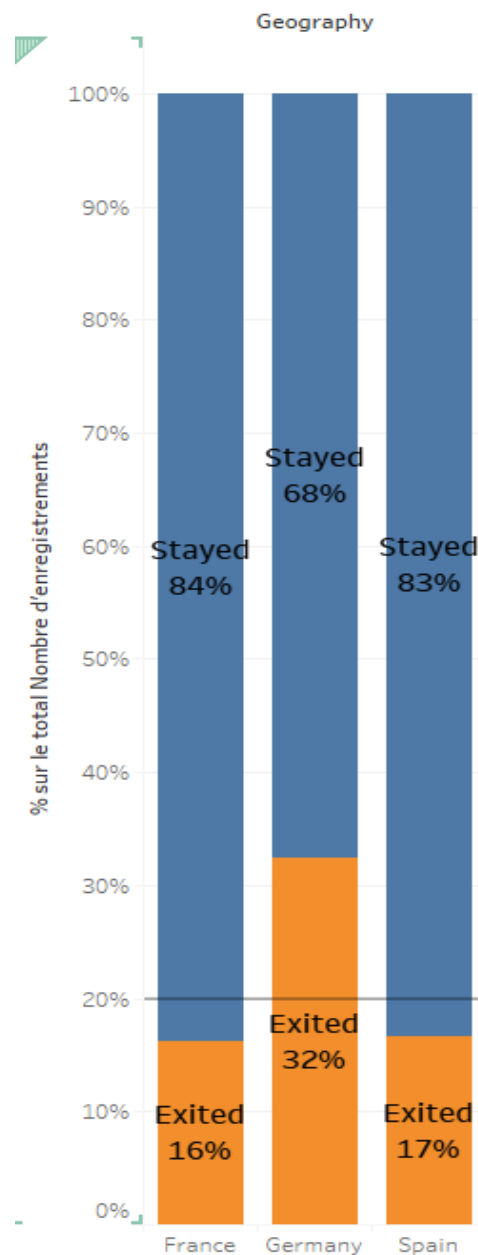


Mon **premier calcul** a été de **fixer mon taux de départ globale de la banque**. Ce taux était de **20%** de clients partit et 80% des clients sont restés. Cette étape permettra de **comparer le taux de départ global** à celui des hommes ou des femmes. On voit donc que chez **les femmes**, le **taux de départ est anormalement supérieur** à la moyenne des départ globale de la banque.

3)Question

Existe t-il une corrélation entre le nombre de départ et le pays du client?

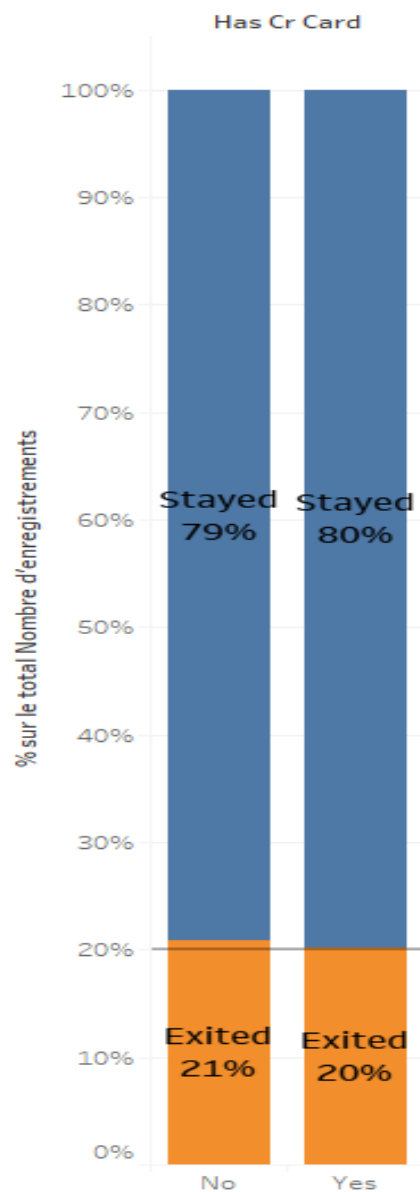
Nombre de départ/rester par pays



Ce diagramme nous montre, toujours sur une moyenne à 20% des départs de la banque, qu'il y a **plus de départs en Allemagne que dans les autres pays**.

4)Question

Existe t-il une corrélation entre le nombre de départ et le fait d'avoir une carte de crédit?

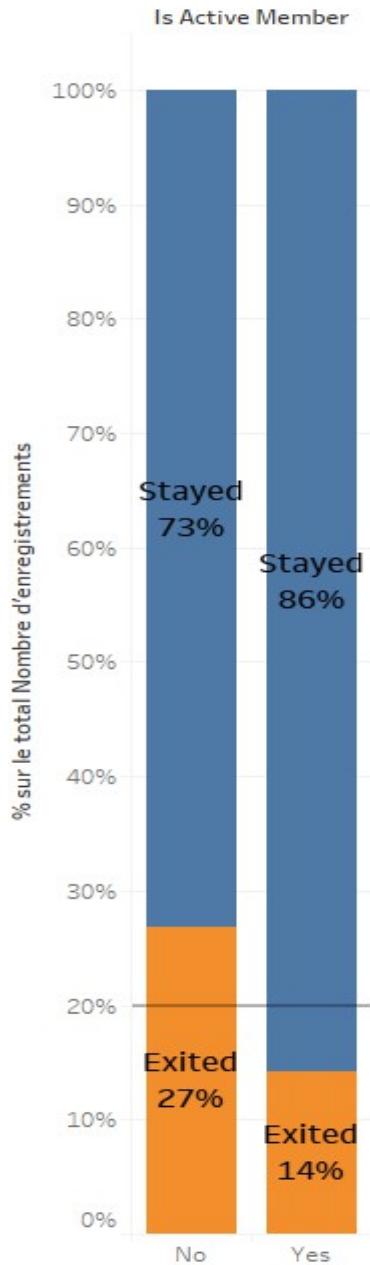


Nous réalisons que **le fait d'avoir une carte de crédit n'affecte pas le départ**, car finalement proche de la moyenne globale des départs dans cet ensemble de données de 20%.

5)Question

Existe t-il une corrélation entre le nombre de départ et le une carte de crédit?

Nombre de personnes restés /partit en étant actif/ non

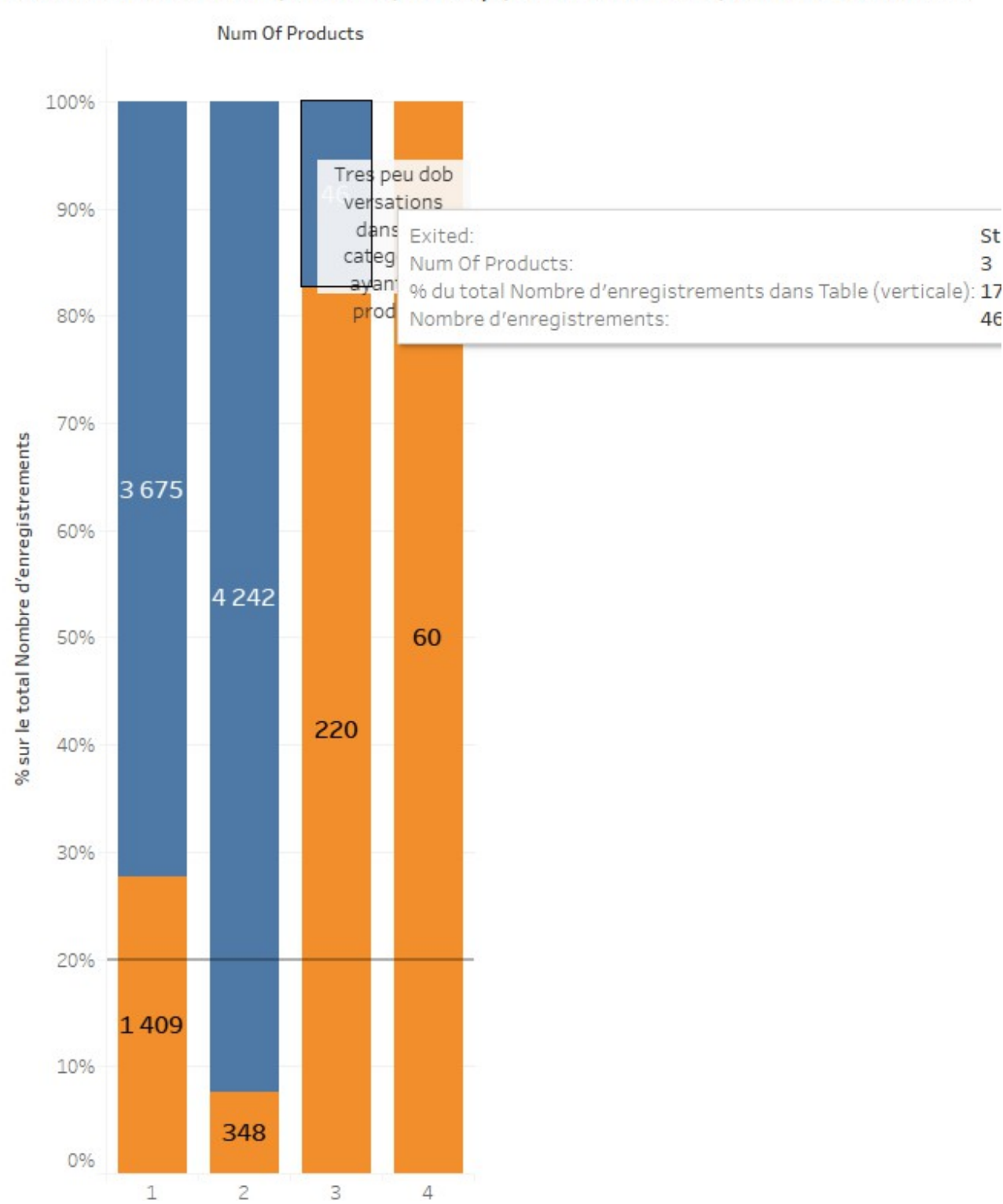


Ce diagramme nous révèle qu' **un usager non-actif** sur son compte a **plus de chance de partir** qu' **un usager actif**.

6)Question

Existe t-il une corrélation entre le nombre de départ et le nombre de produit possédé dans cette banque?

Nombre de clients qui sont partis / par nombre de produits détenus

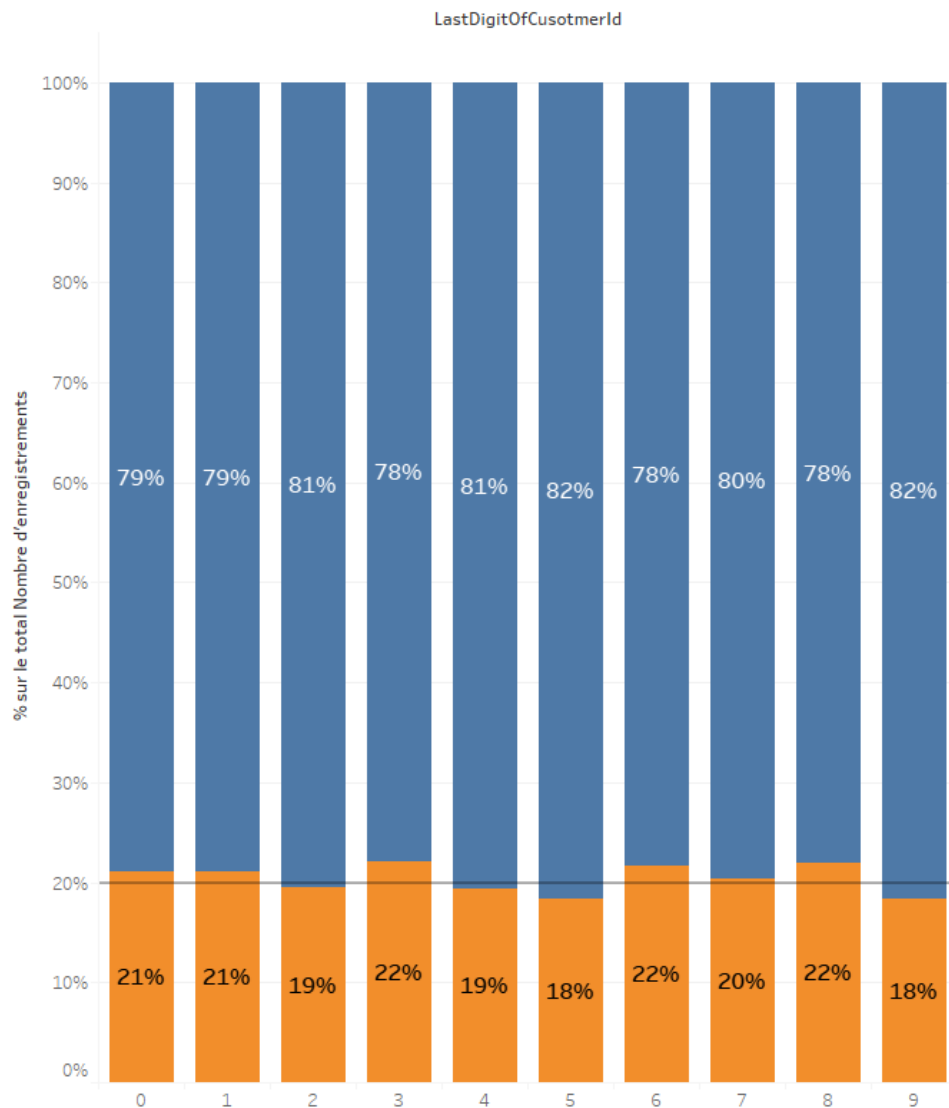


Ce diagramme nous montre que les clients qui détiennent un produit sont plus susceptibles de partir que les clients qui en détiennent 2. Pour les valeurs des produits détenus 3 et 4, nous pouvons voir des anomalies, cela peut s'expliquer par pur hasard, c'est dire que sur l'échantillon des données, la banque a pris 60 personnes parmi tant d'autres, détenant 4 produits et ayant tous quitté la banque. Même cas pour les clients avec 3 produits. En conséquence, nous pouvons tirer cette conclusion: **les clients avec 1 produit sont plus susceptibles de partir que les clients avec 1 produit.**

Validation de nos données

Pour valider nos données, prenons une variable n'ayant aucune corrélation avec le nombre de départ de la banque

Validation



Le fait d'avoir cette distribution uniforme dans ce dernier diagramme, nous assure que tout les modèles démontrés précédemment sont valides.