Exploitation des paradonnées

Objectifs:

Mon travail en entreprise sur les paradonnées à débuter en décembre 2023. Les paradonnées, ce sont l'ensemble des enregistrements horodatés de chaque action du répondant (saisie dans un champ texte, clic sur le bouton de retour arrière, clic sur le bouton suivant). Cette mission, je l'ai essentiellement réalisée sur R, mais j'ai aussi pu faire de la programmation python pour la conversion des données en un format csv. Au terme de cette mission, un rapport texte est attendu. Le programme que j'aurai réalisé sur R, une fois corrigé et documenter, sera partagé avec une fiche résumant les objectifs et la démarche de mon travail.

<u>Démarches Méthodologique:</u>

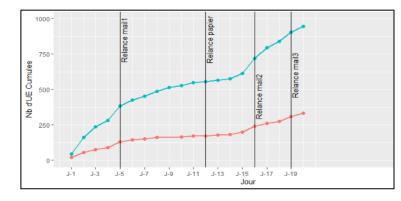
La première étape concerne la récupération et la mise en forme des données. La difficulté était de manipuler les données temporelles. De plus, les données n'étaient pas aussi complètes qu'attendues, j'ai donc dû travailler spécialement à l'enrichissement de données via diverses sources. Pour la seconde étape, qui concerne l'exploitation, j'ai réalisé des analyses sur les temps passés par questions, et ai déterminé les heures où le plus de personnes réponde au questionnaire. J'ai également vérifié l'efficacité des relances par mail et par papier et identifier les répondant ayant effectué un changement d'appareil.

Résultats:

J'ai ainsi pu déterminer que les relances par mail étaient plus efficaces que les relances papier, que les réponses se font majoritairement entre 7h et 12, puis entre 14h et 17h. Pour les changements d'appareil, il se faisait le plus souvent d'un appareil mobile vers un ordinateur, et se produisait régulièrement sur une même page.

Pour l'efficacité des relances, j'ai par exemple réalisé le graphique ci-contre, qui représente le nombre de questionnaires cumulé récoltés selon l'appareil. Au 9ème jour de collecte, nous avions donc déjà récolté environ 500 questionnaires sur ordinateur, en bleu, et environ 130 sur appareil mobile, en rouge. J'ai pu faire figurer sur les axes verticaux, les relances qui ont eut lieu et leurs types, et c'est en regardant l'augmentation du nombre de questionnaires récoltés dans les jours qui suivent que j'ai pu déterminer que les relances par mails étaient plus efficaces que les relances papiers.

Nombre de questionnaires cumulé récoltés



Compétences développées :

Ce travail sur les paradonnées m'a permis de grandement m'améliorer en programmation R et d'apprendre de nouvelles méthodes afin de mieux structurer mon programme. J'ai également gagné en autonomie et en force de proposition en prenant moi-même l'initiative de proposer des idées pour l'exploitation des paradonnées.



