

Exploitation des paradonnées

Objectifs :

Mon alternance à L'Institut national de la statistique et des études économiques ([Insee](#)) sur l'exploitation des paradonnées à débuter en décembre 2023. Les paradonnées 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). L'exploitation de celles-ci, s'est faite essentiellement sur R, mais j'ai également pu faire de la programmation en python afin de convertir au préalable les paradonnées en un format csv.

Au terme de la mission, j'ai pu atteindre les objectifs qui consistaient à analyser et à proposer une réorganisation de la base de paradonnées afin de la rendre plus facilement exploitable. La mission a également abouti à la publication des programmes réalisés qui pourront être réutilisés afin d'exploiter les paradonnées issues d'autres enquêtes.

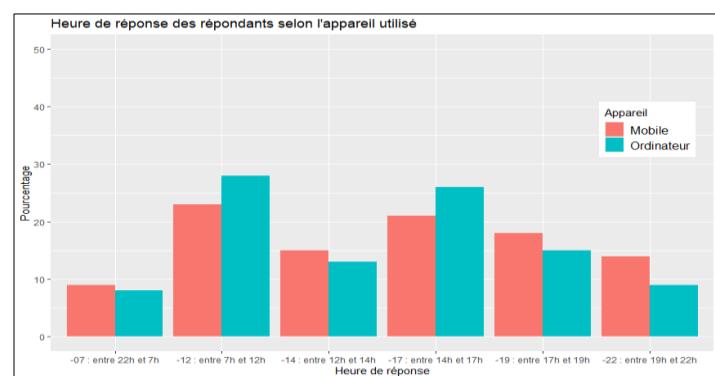
Démarches Méthodologique :

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 j'ai déterminé les heures où les répondants sont le plus susceptible de répondre 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.

Pour les heures de réponses des répondants, j'ai réalisé le graphique ci-contre qui représente la proportion de répondants ayant répondu sur un créneau horaire précis. L'analyse s'est faite selon l'appareil utilisé afin de comparer les heures de réponses entre les répondants sur ordinateur et ceux sur un appareil mobile.



Compétences développées :

Ce travail sur les paradonnées m'a permis de grandement m'améliorer en programmation sur le langage R tout en apprenant de nouvelles méthodes afin de mieux structurer mon programme. Cette mission m'a également permis de développer mes compétences en analyse statistique et en traitement de données.