

Rapport final sur l'étude de polluant

LI Arnold/EL ASSRI Aziz

Introduction

Dans la vie de tous les jours, nous sommes tous exposés plus ou moins à de la pollution que ce soit chez soi ou même au travail mais cependant chaque personne est exposé différemment au même polluant, c'est pourquoi en utilisant un questionnaire auquel plusieurs participants ont répondu, on va essayer de comprendre pourquoi.

Dans cette étude on va étudier les PM1.0, les PM2.5, les PM10, le BC et le NO2. Les PM sont des particules fines et ultrafines dans l'air qui peuvent à un taux élevé avoir un risque sanitaire sur la population, le black carbone (BC) est aussi un polluant de l'air qui est essentiellement émis par les pots d'échappement et par la combustion domestique et est également dangereux pour la santé puis le dioxyde d'azote (NO2) est un polluant majeur de l'atmosphère terrestre produit par les moteurs à combustion interne et les centrales thermiques.

Objectif global

On s'est fixé plusieurs objectifs:

- pouvoir comprendre des données à partir d'un questionnaire et de données fournies
- faire le lien entre les données de mesure et de questionnaire
- pouvoir les analyser
- savoir interpréter les résultats obtenus

Principal partie et résultats

Etude polluant en fonction des catégories

On a d'abord fait une étude générale c'est-à-dire que l'on a fait des requêtes visant tous les participants du questionnaire en fonction de leurs catégories comme on peut le voir sur le pdf 'rapport1.pdf' on a calculé la moyenne des polluants pour chaque catégorie. On remarque que les non sportifs et les personnes 35 à 50 ans sont les plus exposés aux particules fines.

Etude polluant en fonction des participants

Ensuite sur le rapport2, on a calculé les moyennes des polluants en fonction des participants et on avait remarqué que c'est le participant 9999932 qui était le plus exposé mais qu'il avait aussi des valeurs aberrantes et qu'il était aussi le seul non sportif du coup on s'est aussi intéressé au participant 9999944 et 9999946 car ils avaient une forte exposition avec partiellement aucune valeur aberrante.

Etude sur la fiabilité

On s'est posé la question sur la fiabilité des données que l'on avait qui est plus explicite sur le rapport3, on a listé sur le pdf les anomalies des activités et événements auxquelles cela ne correspondait pas

Etude polluant en fonction des activités

Sur ce même pdf on a calculé les moyennes des polluants en fonction des activités et on a remarqué que dans les restaurants, il y avait une forte exposition aux polluants

Etude sur la corrélation

On a essayé de chercher s'il existait des corrélations entre les différentes données et en effet entre les différents polluants ce qui est évident puisqu'ils ne se différencient pas par leur taille et on a trouvé que les autres corrélations étaient soit trop faibles ou très peu intéressantes

Etude polluant en fonction des événements

D'après le pdf sur plot_polluant on en a déduit que les événements allumages de cheminée et marcher étaient ceux qui produisaient le plus de polluant. Ce qui paraît logique puisque marcher on est forcément dehors, on est plus exposé au polluant de l'air et allumage de cheminée qui crée de la fumée, ce qui est un signe que le bois brûle mal et donc qui génère des particules cancérogènes.

Outliers

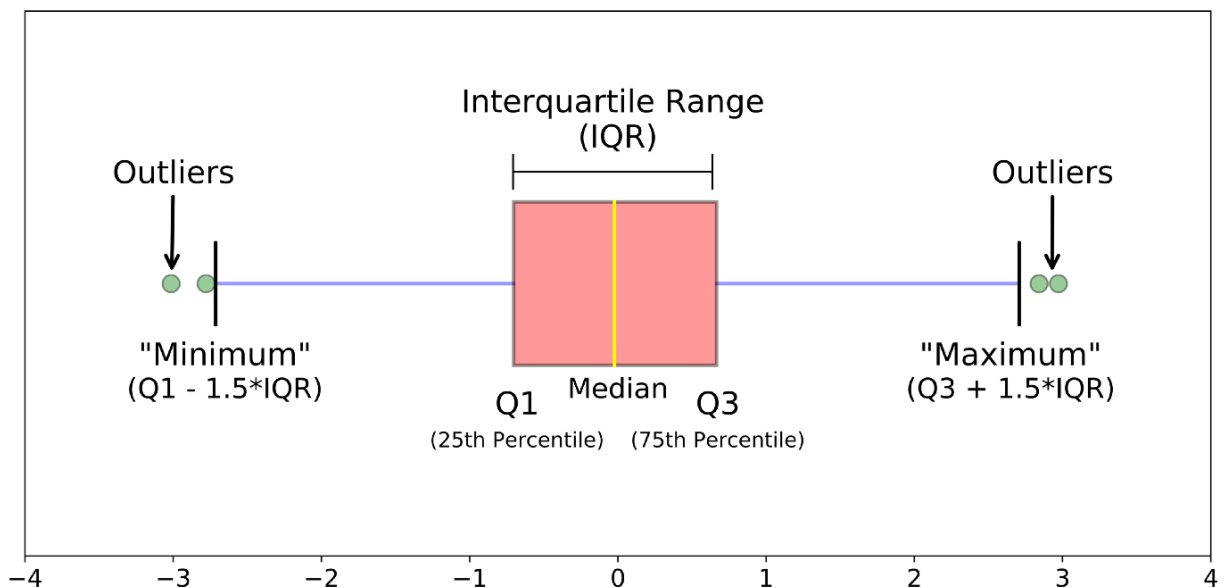


Figure 1: A caption

Une valeur aberrante est une valeur extrême de la distribution d'une variable, elle définit les valeurs extrêmes comme les valeurs supérieures ou inférieures à 1.5 (généralement) fois l'écart interquartile pour pallier à cela nous avons utilisé les outliers. Toutes nos boxplots sont le dossier Plots/Outliers. Pour expliquer ce que l'on

a fait, tout simplement on a remplacé les valeurs aberrantes par le median ou bien par le Q3 si ils se trouvent au dessus de max et si ils se trouvaient au dessous du min on les remplace par le Q1 25%

Outils utilisé

Les outils qu'on a utilisé sont:

- R pour le langage de programmation car le langage R est très pratique pour faire des statistiques et analyser des données
- SQLite pour les traitements des requêtes on a utilisé et les dataframes pour le stockage des tables
- Shiny pour faire une mini app de notre projet et avoir de l'interaction avec l'utilisateur
- R markdown pour la rédaction de rapport

Problème rencontrée

Au début, nous avons remarqué qu'il y avait des anomalies sur les données fournies car il est difficile de se fier à des données incomplète ou même fausse.

Les valeurs NA sont des valeurs que un participant n'a pas répondu et difficile à traiter en effet on ne savait pas comment les analyser

Perspective

Ce qui pourrait être améliorer ou fait si on avait plus de temps:

- on pourrait faire une application qui pourrait avoir la fonction qui calcul la moyenne de pollution d'un endroit à telle moment pour que l'utilisateur sache si il va dans une zone trop pollué ou pas

Conclusion

Ce projet nous a beaucoup appris et a été très enrichissant que ce soit au niveau des connaissances et du travail d'équipe. Il nous a permis aussi de découvrir les enjeux lié à la pollution notre planète, les risques possibles sur les différents polluants. Grâce à ce projet, on aimerait que nos futurs projets seront encore mieux