A/ vous choisissez un angle d'étude

c'est le cas de figure classique pour un datascientist :

on doit expliquer des phénomènes, mais on ne sait pas ce que l'on cherche exactement

et le client final n'est pas toujours apte à formuler son besoin correctement

il s'agit donc d'une mise en situation réaliste

EXEMPLE pour les "arbres remarquables" :

=> chercher des corrélations (entre taille/longévité et type d'arbre, leur répartition par quartier en fonction de la densité et la pollution...)

vous êtes libre : le seul impératif est de prouver des choses à votre interlocuteur (moi en l'occurrence) et d'être dans la véracité : évidemment pas d'erreur de calcul, mais surtout : pas en essayant d’éviter les erreurs d'interprétation (vous pouvez très bien formuler des choses non 'certaines' à condition d'y apporter cette mention)

vous pouvez utiliser un logiciel de statistique (R, matlab…)

B/ou aussi envisager de développer de “mini-algorithmes” pour vous aider à charger, traiter et exporter vos données

le langage de programmation est totalement libre

ATTENTION :

* votre étude ne peut pas se résumer à une simple analyse descriptive des données qui se contenterait de présenter différents charts (ex : répartition des arbres par taille),
* il faut chercher à prouver quelque chose à travers les données (ex: corrélation entre la taille des arbres et la pollution de l’air ?) => les charts sont ainsi des moyens intermédiaires étayant votre démarche intellectuelle sans être une fin en soi

f/ vous projetez vos analyses sous forme de dataviz

les dataviz de bureautique (excel, google sheet, open office....) sont acceptées

le niveau au-dessus également (par exemple data-studio <-> google sheet, ou elastic <-> kibana...)

vous pouvez vous appuyer sur un logiciel de statistique (R, matlab…)

ainsi que les outils de dataviz avancés (type metabase/Seektable/PowerBi…) si vous vous sentez

vous êtes libres :

\* le seul impératif est de dénouer par la dataviz une explication complexe (ex : illustrer une corrélation contre-intuitive)

\* ne seront pas acceptées des dataviz triviales - comme représenter basiquement une évolution mois/mois d'une simple variable

\* le choix des charts a son importance (map, nuage, bubble, diagramme...), mais attention à ne pas en faire trop (avec des diagrammes inutilement complexes ou à pousser l'esthétisme trop loin) : il faut privilégier **l'efficacité, l'efficience, la simplicité et le pragmatisme**