

Analyse exploratoire de données

d'un site web de livraison de produits alimentaires

Testez vos compétences en matière de manipulation de données sur un ensemble de données réelles.

Cet ensemble de données anonymes contient plus de 3 millions de commandes et 200 000 utilisateurs d'un site web de livraison de produits alimentaires.

Ensemble de données

Chaque utilisateur a entre 4 et 100 commandes, avec la séquence des produits achetés dans chaque commande.

La semaine et l'heure de la commande sont également indiquées, ainsi qu'une mesure relative du temps écoulé entre les commandes.

Les produits se trouvent dans des allées qui sont elles-mêmes divisées en catégories.

Une banane verte se trouve dans l'allée des fruits frais, qui fait elle-même partie de la catégorie des fruits et légumes.



Objectif du projet

Vous êtes libre de choisir les analyses et les graphiques que vous souhaitez présenter.

Vous êtes libre d'utiliser:

- Bokeh : <https://bokeh.org/>
- Plotly : <https://plotly.com/>
- Pandas + D3.js : <https://d3js.org/>

Lors de l'examen oral de la semaine prochaine, vous devrez expliquer en détail vos choix.

Votre travail doit être aussi organisé et professionnel que possible.

Pour commencer

Voici des exemples de ce que vous pourriez faire.

- Quand les gens commandent-ils ?
- Combien d'articles achètent-ils ?
- Quels sont les articles les plus vendus ?
- Combien de fois les clients commandent-ils à nouveau les mêmes articles ?
- Quel article les clients mettent-ils en premier dans le panier ?
- Quel est le rapport entre les commandes contenant du bio et celles n'en contenant pas ?
- Combien de produits uniques sont proposés dans chaque rayon/allée ?
- Existe-t-il un lien entre la date de la dernière commande et la probabilité de commander à nouveau ?

