

Jeux de données pour l'Atelier 2

L'équipe des GRAPH Courses

January 2024

Jeux de données de l'Atelier

Lors de l'atelier de cette semaine, vous analyserez un jeu de données de votre choix. Ces jeux de données se trouvent dans le sous-dossier “data”. Choisissez-en un parmi les Maladies Fébriles au Burkina Faso, les Accidents de Moto en Colombie ou les Patients COVID-19 aux États-Unis. Le dernier jeu de données sera utilisé pour une démonstration.

Notez que tous les jeux de données sont en anglais !

Les scripts avec les instructions pour analyser les jeux de données se trouvent dans le sous-dossier “scripts”.

1. Maladies Fébriles au Burkina Faso

Ce jeu de données se concentre sur les maladies fébriles au Burkina Faso, incluant des données cliniques de 914 patients hospitalisés souffrant de maladies causant de la fièvre. Les chercheurs ont cherché à tester la précision d'un algorithme qui utilise les mesures métriques du sang pour prédire la cause de l'infection, en comparant les techniques de culture standard avec un algorithme basé sur l'hémocytométrie (IMS), la procalcitonine (PCT) et la protéine C-réactive (CRP). Le jeu de données fournit des informations sur la démographie des patients, les symptômes et les résultats des tests sanguins, ainsi que la classification finale de la cause de l'infection sous la variable `ims_final_full_classification`.

Pour en savoir plus sur le jeu de données, cliquez [ici](#).

2. Accidents de Moto en Colombie

Ce jeu de données contient les dossiers de police sur les décès résultant de blessures routières liées aux motos dans la ville de Medellin entre 2012 et 2015. Il se concentre exclusivement sur les incidents mortels impliquant des motocyclistes, conducteurs et passagers. Le jeu de données pertinent sur le site lié s'appelle “Dataset 1.xlsx”. Dans la deuxième feuille de ce fichier `xlsx`, vous trouverez un dictionnaire de données. La publication accompagnante est accessible [ici](#).

Accédez aux données connexes [ici](#).

3. Patients COVID-19 aux États-Unis

Ce jeu de données englobe des données anonymisées de 1540 patients hospitalisés pour COVID-19 dans deux hôpitaux de New York : l'Université d'État de New York (SUNY) Downstate Health Sciences University et le Maimonides Medical Center. Les données, trouvées dans le fichier “demographics_both_hospitals.csv”, fournissent une multitude d'informations sur la démographie des patients, les comorbidités connues et les résultats finaux de l'hospitalisation (sortie ou décès).

Pour en savoir plus sur les données, cliquez [ici](#).

4. DÉMO DE L'ENSEIGNANT : Enquête sur les Attitudes Sexuelles en Grande-Bretagne

Ce jeu de données contient des données anonymisées issues de l'Enquête Nationale sur les Attitudes et Styles de Vie Sexuels, troisième vague (Natsal-3), l'une des plus grandes études sur le comportement sexuel en Grande-Bretagne. Réalisée entre 2010 et 2012, Natsal-3 a interrogé plus de 15 000 adultes britanniques âgés de 16 à 74 ans. Les données consistent en les réponses de 3 799 adultes âgés de 16 à 74 ans. Les variables couvrent les caractéristiques démographiques des répondants, leurs attitudes envers le sexe et les relations, ainsi que leur historique reproductif.

Pour en savoir plus sur les données, cliquez **ici**.

Le dictionnaire des variables se trouve **ici**