

IFT3700 A2023

Travail final

Directives

- Le travail peut-être rédigé en français, anglais ou espagnol.
- 30% de la note totale du cours
- **Un seul fichier, pas de .zip**

Vous devrez remettre le rapport en format Ipython Notebook

- [Scikit-learn](#)
- [matplotlib.pyplot — Matplotlib 3.5.3 documentation](#)

Description

Vous allez devoir collecter et nettoyer des données numériques et catégoriques au sujet des pays du monde. Les données proviennent de 40 tableaux disponibles sur Wikipédia. En plus de traiter les valeurs numériques données, toutes les caractéristiques (colonne) .Une fois les données nettoyées, une analyse en plusieurs points sera effectuée. La qualité de l'analyse et de la présentation sont d'égale importance.

Il faut se rappeler que les données observationnelles seules ne permettent pas d'établir des liens causaux. Il est impératif que votre analyse ne contienne **pas de jugement de valeur inapproprié**.

Partie 1: Collecte

Collecter les 40 colonnes telles que spécifiées dans la section donnée. Nettoyer les données en conservant les 40 colonnes mais en éliminant ou regroupant les pays problématiques. Vous allez faire face à une grande quantité de problèmes que vous devrez régler au meilleur de vos capacités. Vous devez sommairement justifier toutes vos décisions. Il faut remettre un fichier Python Notebook et un fichier text CVS. Le fichier CVS contient le tableau final nettoyer et avec des blancs pour les données manquantes. Vous pouvez discuter entre vous sur le forum.

Les données

La liste suivante contient une référence précise à la page, au tableau, à la colonne pour chaque caractéristique à étudier. Notez qu'il existe des outils qui permettent de faciliter le téléchargement de ce genre de tables.

Lorsque les valeurs sont données sous forme d'intervalle, vous devez choisir le point médian entre la limite inférieure et la limite supérieure de l'intervalle. Dans les très rares cas où deux valeurs sont données, c'est-à-dire lorsque différentes sources différentes valeurs, vous devez aussi choisir le point médian.

1. [List of countries by GDP \(nominal\) per capita](#), GDP (in US\$) per capita by country, United Nations, estimate.
2. [List of countries by Internet connection speeds](#), Fixed broadband, Average download speed (Mbit/s) (Ookla).
3. [List of countries by alcohol consumption per capita](#), Recorded per capita consumption of pure alcohol (litres) per adult 15 years of age and over per year, 2016.
4. [List of countries by intentional homicide rate](#), Intentional homicide victims per 100,000 inhabitants. From UNODC, rate.
5. [List of countries by military expenditures](#), List by the Stockholm International Peace Research Institute, % of GDP.
6. [List of countries by Human Development Index](#), 2019 data.
7. [Democracy Index](#), By country, 2020.
8. [List of countries by tertiary education attainment](#), Countries by level of tertiary education, at least a 2-year tertiary degree or its equivalent %.
9. [Importance of religion by country](#), Countries/Districts, Yes important.
10. [Christianity by country](#), UN members and dependent territories, % Christian.
11. [Islam by country](#), Table, Muslim percentage of total population.
12. [Buddhism by country](#), Buddhist population by country, % population.
13. [Jewish population by country](#), Table, Core population, pct.
14. [List of countries by infant and under-five mortality rates](#), Under-five mortality (deaths/1,000 live births) – 2019 estimates
15. [Age of criminal responsibility](#), By country, age reduced.
16. [List of countries by minimum wage](#), List, Annual, US\$.
17. [List of countries by external debt](#), List of countries with respect to external debt, % GDP
18. [List of countries by income equality](#), UN, World Bank and CIA list – income ratios and Gini indices, WB Gini %.

19. [List of countries by total health expenditure per capita](#), Total health expenditure per capita in 2018 PPP international U.S. dollars, 2018.
20. [List of countries by suicide rate](#), Suicide rates by sex and country, All.
21. [List of sovereign states and dependencies by total fertility rate](#), 2019 List by the World Bank, 2019.
22. [Tobacco consumption by country](#), Prevalence of current tobacco use among persons aged 15 years and older, 2000.
23. [List of countries by obesity rate](#), list, 2016 %.
24. [List of countries by number of Internet users](#), Table, %.
25. [List of countries by median age](#), Median age per CIA World Factbook 2018 and 2020 (ranked) estimates, 2020 median combined.
26. [List of countries by economic freedom](#), 2019 Economic Freedom of the World Index, score.
27. [List of countries by oil production](#), per capita 2017. **Valeur absente = 0, valeur manquante [-] = comme les autres.**
28. [List of countries by population growth rate](#), Table, UN 2015-20.
29. [List of countries by life expectancy](#), List by the World Health Organization (2019), Life expectancy at birth, All.
30. [List of countries by meat consumption](#), Meat consumption by country, KG per capita **2002**.
31. [List of countries by incarceration rate](#), Incarceration populations and rates. From World Prison Brief, rate per 100000.
32. [List of countries by literacy rate](#), List of UN member states by age group and gender disparity, Elderly (65+), rate.
33. [List of countries by age at first marriage](#), Woman. Les différents tableaux doivent être combinés.
34. [List of countries by spending on education](#), % of GDP
35. [List of countries by homeless population](#), per 100000.
36. [List of countries by milk consumption per capita](#), 2013.
37. [List of countries by number of scientific and technical journal articles](#), per capita.
38. [Books published per country per year](#), titles (total).
39. [List of countries by food energy intake](#), kilocalories.
40. [List of countries by average yearly temperature](#), celsius.