

Projet : WORLD HAPPINESS REPORT

Pour ce projet, vous devez télécharger et utiliser le fichier csv suivant : worldHappinessReport.csv (source : WHR).

Remarque : la signification des différentes variables peut être trouvée:

https://s3.amazonaws.com/happiness-report/2019/WHR19_Ch2A_Appendix1.pdf.

Pour ce projet, tous vos graphiques, calculs, tests, etc. devront être introduits, justifiés et interprétés (surtout interpréter vos résultats). Une rédaction professionnelle est attendue.

A. REPORTING

1. Créez un reporting avec R

B. STATISTIQUE UNIVARIE

Regardons en détail la variable GDP

2. Que signifie t'elle ?
3. Faire un résumé statistique de cette variable et calculer les coefficients de skewness et kurtosis interprétez
4. Existe t'il des données aberrantes dans cette variable ? Expliquez
5. Cette variable suit elle la loi normale ? Expliquez
6. Donnez l'intervalle de confiance à 90% de la moyenne de cette variable
7. Affichez les 10 premiers pays avec les GDP les plus élevées
8. Calculez la moyenne des GDP par region du monde. Dessinez un graphique à barres. Interpréter.
9. Étudiez la variable Region : fréquences, proportions, graphiques.

C STATISTIQUE GENERALE ET INFERENTIELLE

Dans cette partie, il est conseillé de choisir 2 variables pour chaque questions

10. Faites un test d'indépendance entre 2 variables catégorielles (visualisation et test)
11. Faites un test d'indépendance entre 2 variables quantitatives (visualisation et test)
12. Faites un test d'indépendance entre 1 variable quantitative et 1 variable catégorielle (visualisation et test)

D MODELE DE REGRESSION

13. Étudiez comment happiness dépend de la région et des autres variables.
15. Effectuez une régression simple pour prédire happiness en fonction du GDP. Donnez l'équation de la droite
16. Effectuez une régression multiple entre happiness et certains variables quantitatives que vous choisirez.
17. Standardisez ou mettez aux mêmes échelles vos données si besoin

E PCA

18. Effectuez une analyse en composantes principales avec les principales variables caractéristiques des pays.

F CLASSIFICATION : DECISION TREE

18. Remplacer la colonne happiness par 1 si happiness>5 sinon par 0 (happiness <=0).

19. Peut on créer un modèle de classification (tres simple avec une decision tree) pour predire cette nouvelle variable ?.

20. Que pensez vous de cet exercice en générale ?

BONUS

G CLUSTERING

1. Effectuez un algorithme de clustering k-means avec un nombre de cluster de 5.

2. Déterminez le nombre optimal de clusters et Dessinez un graphique correspondant à ce clustering.

3. Même question avec un algorithme de clustering hiérarchique. .

4. Représenter les clusters précédents sur un plan défini par les deux premiers composants.

H DETECTION D'ANOMALIE

5. Détectez les 5% des pays qui pourraient être considérés comme une anomalie, interprétez.

6. Documentez vous sur folium et affichez une carte du monde avec la variable happiness

I AUTOMATISATION

7. Automatisez l'ensemble de votre travail (même partiellement).