

Mini Projet – Traitement de Données avec Python - Pandas

Le but de cet exercice est d'extraire un sous-ensemble de données à partir d'une base de données existante.

Supposons que vous ayez besoin de réaliser une application qui prend en entrée, des noms de lieux du Burkina Faso, avec des informations telles que la latitude/longitude.

Pour ce faire, nous décidons d'extraire ces informations à partir du server de référencement géoname (<http://www.geonames.org/>). Vous allez procéder de la façon suivante :

1. Exporter la base de données qui recense les informations sur le Burkina Faso (<https://download.geonames.org/export/dump/>).
 - Pour ce faire, référez-vous au Readme, décrit à la fin de la page pour identifier le code iso correspondant à celui du Burkina Faso
2. Télécharger le fichier zip correspondant
3. Appliquer les opérations de prétraitement et filtres nécessaires à ce fichier, pour ne garder que les colonnes correspondantes :
 - Identifiants, Noms de lieux, latitudes, longitudes
 - Renommez les avec les noms suivants : `'ID'`, `'location_name'`, `'lat'`, `'long'`
 - Sauvegarder ces données dans un fichier CSV, nommez-le *burkina_location.csv*
4. Opérations sur le fichier CSV *burkina_location.csv*.
 - Extraire les données contenant le nom *'gounghin'*, enregistrez-le sous le fichier *gounghin.csv*
 - Extraire la sous-partie de la base de données (fichier *burkina_location.csv*), dont les premières lettres des noms de lieux sont compris entre *'A'* et *'P'* (ordre alphabétique).
 - Identifiez respectivement, la latitude, la longitude minimale et les noms de lieux correspondants dans le fichier *burkina_location.csv*.
 - Quels sont les lieux dont les coordonnées sont comprises entre ($lat \geq 11$ et $lon \leq 0.5$)
5. Sorties Excel

À partir des extractions de l'étape 4 :

- Créer un fichier Excel et nommer le : *mini_projet*
- Créer une feuille dans ce fichier, du nom *gounghin* et enregistrer les données contenant le nom '*gounghin*' obtenues dans 4.1
- Créer une second feuille dans ce même fichier, du nom *A_to_P* et enregistrer les données de 4.2

Chacun devrait rendre :

- 1- le fichier du code source (Jupyter ou Google Colab).
- 2- Un repertoire github (lien d'accès) associé au nom (*mini_projet_data_processing*), dans lequel il mettra les données utilisées, les codes , et les données issues des différentes manipulations.
- 3 – Le délai pour rendre l'exercice : 07.11.2021