

M1 G3 – Informatique pour les géosciences
TP4 – Introduction au module Pandas

Le fichier rock_properties_TP2.xlsx contient des données pétrophysiques (densité et vitesses d'ondes P et S) mesurées sur des échantillons de roches prélevés à divers endroits d'Amérique.

1. Chargement des données

1.1. Vérification/ Changement de répertoire

A l'aide du module os, listez les fichiers présents dans le répertoire courant. Si le fichier de données n'est pas dans le répertoire courant, changez de répertoire.

1.2. Import des données à partir du fichier Excel

En utilisant la fonction `pd.read_excel()`, chargez le fichier de données.

Utilisez les méthodes `df.head()`, `df.tail()`, `df.shape`, `df.info()` et `df.describe()` pour vérifier que les données ont été chargées correctement.

2. Analyse des données

2.1. Types de roches

Utilisez la fonction `values_counts()` pour afficher le nombre d'échantillons pour chaque type de roche.

2.2 Analyse des densités

Dans cette partie, on va s'intéresser uniquement aux valeurs de densité pour les différents types de roches.

- Créez un nouveau dataframe ne contenant que les colonnes 'Rock type/ VISUAL' et 'Density/(gm/cc)'.
- En utilisant les méthodes `df.groupby()` et `.mean()` calculez la moyenne des densités pour chaque type de roche.
- En utilisant les méthodes `df.groupby()` et `.std()` calculez la déviation standard des densités pour chaque type de roche.
- En utilisant la méthode `df.pivot_table()`, calculez les minimum, maximum, moyenne et déviation standard des densités pour chaque type de roche.
- En utilisant la méthode `df.plot()`, représentez l'histogramme des densités.
- En utilisant les méthodes `df.groupby()` et `df.plot()`, représentez l'histogramme des densités pour chaque type de roche.

2.3 Vitesses en fonction des densités

En utilisant la méthode `df.plot()`, représentez les vitesses en fonction des densités sous forme de nuage de points (scatter plot). A quoi correspondent les vitesses les plus élevées ? les vitesses les plus faibles ?

Filtrez le dataframe de manière à ne représenter que les vitesses d'ondes P.