

TP 1 Manipulation des DataFrames

MI - IA / GL

Ilyas Bambrik

Table des matières



I - Exercice : Lecture d'un DataFrame	3
II - Exercice : Sélection	4

Exercice : Lecture d'un DataFrame



Le DataFrame utilisé dans ce TP est similaire au DataFrame du cours mais propose plus d'informations. Le représente *Times Higher Education World University Rankings 2024* qui comprend 1 904 universités dans 108 pays et régions. *THE (Times Higher Education)* fournit depuis 2004 des données fiables sur les performances des universités aux étudiants, leurs familles, aux dirigeants universitaires, aux gouvernements et à l'industrie.

Question

- Lisez le DataFrame à partir du fichier *timeswr.csv*. Combien de colonnes contient ce DataFrame ?
- Quelles sont les cinq meilleurs universités selon ce classement ?
- Quelle est le nom de la colonne qui désigne le pays de l'université ?
- Affichez les dix premières lignes du DataFrame.
- Affichez les dix dernières lignes du DataFrame.
- Changez l'index du DataFrame tel que la colonne *name* devient l'index.

Exercice : Sélection



La colonne *unaccredited* indique si l'université n'est pas reconnue pour des raisons réglementaires.

Question 1

- Quel est le type de cette colonne ?
- Trouvez les noms des universités non accréditées dans le classement s'ils existent.

Indice :

Utilisez l'attribut *dtype* de la colonne (Serie)

Question 2

- Combien d'universités Algériennes existent dans le classement ?
- Quelle est l'université Algérienne la mieux classée selon les données ?

Ce classement comporte plusieurs colonnes de score comme *scores_overall* et *scores_research*.

Question 3

- Sélectionnez les colonnes du DataFrame allons de la colonne *scores_overall* jusqu'à *scores_international_outlook_rank*.
- Sélectionnez les lignes du DataFrame avec une valeur *scores_research* ou *scores_teaching* supérieure à 80
- Sélectionnez les universités classées parmi les cinq meilleurs dans *selon scores_teaching_rank* ou *scores_research_rank* ou *scores_citations_rank*.