

Manipulation des données

Exercice 1 Manipulation des textes

1. Importer la base de données relatives aux joueurs de la Coupe du Monde 2014 >> `Players_WC2014.xlsx`
2. Déterminer le nombre de joueurs dans chaque équipe et créer un dictionnaire { équipe : Nombre de joueurs }
3. Déterminer quels sont les 3 joueurs qui ont couvert le plus de distance. Y a t il un biais de sélection ?
4. Parmi les joueurs qui sont dans le premier décile des joueurs plus rapides, qui a passé le plus clair de son temps à courrir sans la balle ?

Exercice 2 Les villes

1. Importer la base des villes villes.xls
2. Les noms de variables et les observations contiennent des espaces inutiles (exemple : 'MAJ') : commencer par nettoyer l'ensemble des chaines de caractères (à la fois dans les noms de colonnes et dans les observations)
3. Trouver le nombre de codes INSEE différents (attention aux doublons)
4. Comment calculer rapidement la moyenne, le nombre et le maximum pour chaque variable numérique ? (une ligne de code)
5. Compter le nombre de villes dans chaque Region et en faire un dictionnaire où la clé est la région et la valeur le nombre de villes
6. Représenter les communes en utilisant
 - matplotlib

- une librairie de cartographie (ex : folium)

Exercice 3 Disponibilité des vélib

1. Importer les données sous la forme d'un dataframe

- `velib_t1.txt` - avec les données des stations à un instant t
 - `velib_t2.txt` - avec les données des stations à un instant $t+1$
- !

2. Représenter la localisation des stations vélib dans Paris

- représenter les stations en fonction du nombre de places avec un gradient

3. Comparer pour une station donnée l'évolution de la disponibilité (en fusionnant les deux bases t et $t+1$)

- représenter les stations qui ont connu une évolution significative (plus de 5 changements) avec un gradient de couleurs