

4DVST

TP : Activité de Visualisation de Données

Léo SOHRABI

Rapport de Visualisation de Données

Objectif

Ce projet vise à explorer un jeu de données de football provenant de Kaggle en appliquant les bonnes pratiques de visualisation de données avec Python.

L'objectif est de réaliser plusieurs visualisations lisibles, pertinentes et originales à l'aide de bibliothèques comme Seaborn et Matplotlib, tout en filtrant les données pour ne garder que les matchs des grands tournois internationaux (Coupe du Monde et UEFA Euro).

Jeu de données utilisé :

- Source : Kaggle - International Football Data
 - Fichiers CSV utilisés :
 - **results.csv** : résultats de matchs internationaux
 - **goalscorers.csv** : buteurs par match
-

Justification :

Ce jeu de données est pertinent car il couvre une longue période de l'histoire du football international avec des dimensions variées (temps, lieu, performance). Il permet de produire des analyses temporelles, géographiques, stratégiques ou statistiques avec des graphiques variés.

Choix de la librairie :

J'ai choisi d'utiliser la librairie Seaborn, reposant sur Matplotlib.

Elle permet de créer rapidement des visualisations modernes et claires, avec des palettes de couleurs efficaces et une syntaxe intuitive. Matplotlib a été utilisé en support pour le placement des axes et l'affichage final.

🌐 Visualisations réalisées :

1. Top 10 des meilleurs buteurs (**goalscorers.csv**)

- Type : Barplot horizontal
- Librairie : Seaborn
- Affichage du nombre de buts marqués par les 10 meilleurs buteurs (hors buts contre leur camp)

2. Heatmap des confrontations internationales (**results.csv**)

- Type : Heatmap croisée
- Librairie : Seaborn
- Affiche le nombre de confrontations entre les 15 nations les plus présentes dans les données

3. Victoires et défaites par pays (**results.csv**)

- Type : Barplot horizontal
- Librairie : Seaborn
- Affiche le nombre de victoires et de défaites pour chaque pays dans les matchs sélectionnés. Limitation à un top 15 des équipes les plus présentes pour plus de lisibilité.

⭐ Améliorations et bonnes pratiques :

- **Modularité du code** : Tous les graphiques sont définis via des fonctions réutilisables, ce qui favorise la lisibilité et la réutilisabilité du code.
- **Filtres sur les tournois** : Le dataset a été filtré pour ne garder que les matchs issus des grands tournois comme la Coupe du Monde et l'UEFA Euro, afin de garantir la pertinence des confrontations affichées.
- **Palette et lisibilité** : Utilisation de palettes lisibles (viridis, Blues, flare) pour assurer un bon contraste visuel.
- **Personnalisation des graphiques** : Les axes, titres, grilles et tailles sont soigneusement personnalisés pour garantir une bonne lecture des données.

Plutôt que de créer une version "brute", j'ai directement conçu des visualisations optimisées en partant des principes de design lisible (espacement, étiquettes, palette).

 **Conclusion personnelle :**

Ce projet m'a permis de mettre en pratique mes compétences en visualisation de données et de structurer mon code de manière propre et réutilisable.

J'ai appris à manipuler des données réelles tout en veillant à la lisibilité et la pertinence des résultats.

L'utilisation combinée de Seaborn et Matplotlib m'a permis de créer des visualisations riches tout en conservant une bonne compréhension des données. Ce projet m'a aussi appris à rendre mon code plus lisible et réutilisable, en structurant mes analyses et graphiques sous forme de fonctions paramétrables.

ANNEXE:

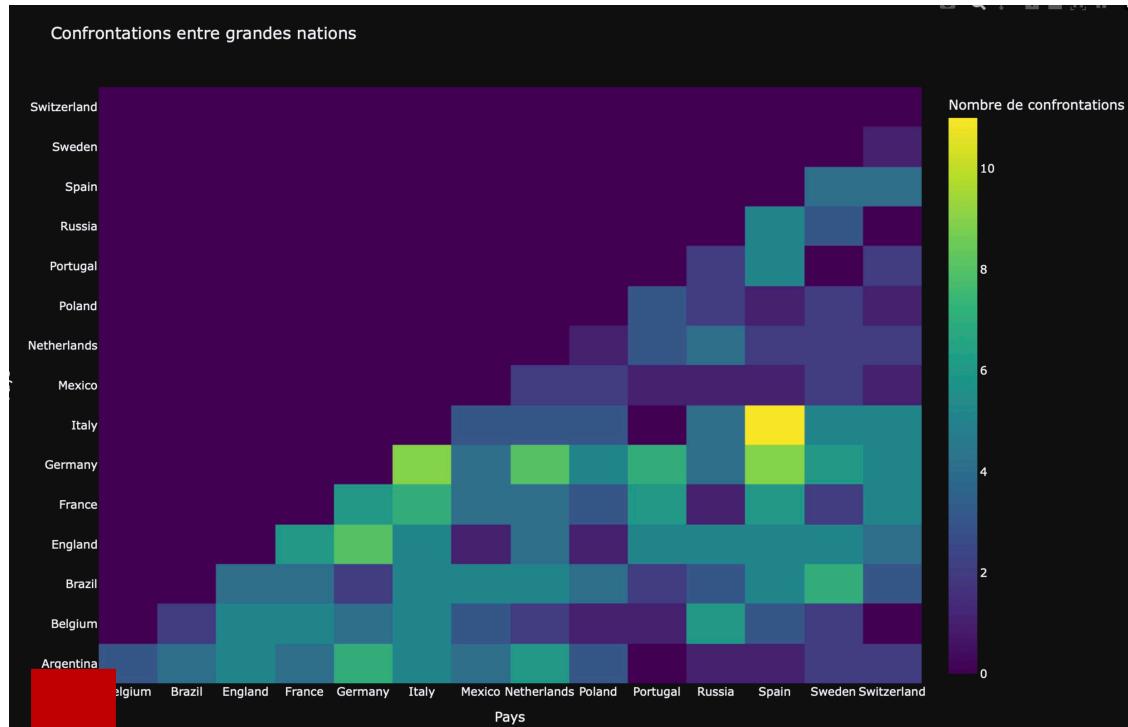


Diagramme : Confrontations entre grandes nations AVANT

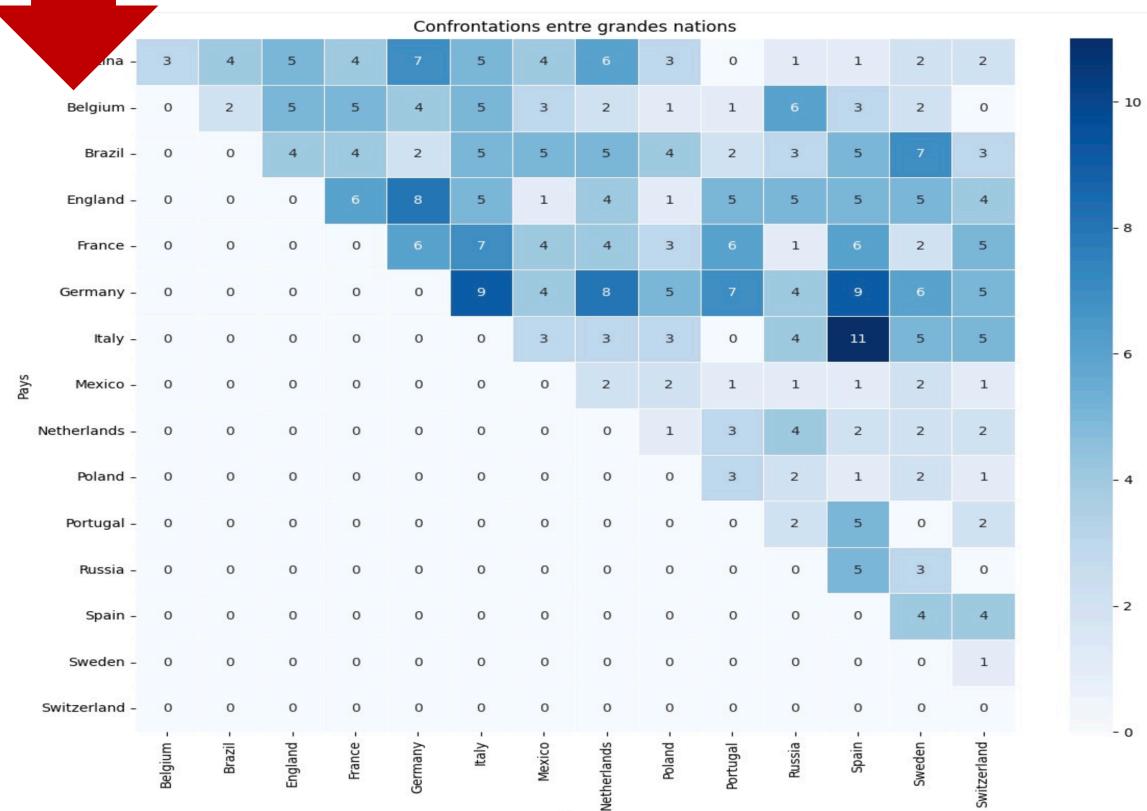


Diagramme : Confrontations entre grandes nations APRÉS

