

Projet Python

Étude comparative entre le championnat d'Europe de Football 2016 et de la coupe du monde de Rugby 2023 (Tous deux en France)



Mise en contexte :

Monsieur Etienne Guerin est un célèbre journaliste/chroniqueur d'une chaîne de radio sportive très populaire. Il souhaiterait obtenir plusieurs informations comparatives entre l'Euro de Football 2016 et la coupe du monde de Rugby 2023, malheureusement il ne sait pas coder et manque de temps. Il a donc décidé de vous contacter, vous qui savez coder, afin d'extraire les différentes informations qu'il recherche pour sa prochaine émission. Pour accomplir votre mission, il vous a transféré des bases de données qu'un ami lui a données.

Bases de données :

Vous avez à votre disposition trois bases de données format txt, deux bases portant sur le championnat d'Europe de football 2016 et une sur la coupe du monde de rugby 2023 :

- Une base de données portant sur les matchs de la phase de poule de l'Euro de football 2016. Chaque ligne correspond à un match avec pour données : 1^{ère} équipe, 2^{ème} équipe, le stade où a eu lieu le match, la ville où se trouve le stade, une estimation du nombre minimal de supporters dans le stade, une estimation du nombre maximal de supporters dans le stade et la date du match (aa/bb/cccc). Les éléments de données sont séparés les uns des autres par un espace (**non par une tabulation**). Les matchs sont classés par groupe puis ordre chronologique. Le passage d'un groupe à l'autre se fait par le saut d'une ligne vierge.

Nom du fichier : groupes euro 2016

Exemple :

France	Roumanie	Stade_de_France	Saint-Denis	75 000	80 000	10/06/2016
--------	----------	-----------------	-------------	--------	--------	------------

- Une base de données portant sur les matchs de la phase à élimination directe de l'Euro de football 2016. Chaque ligne correspond à un match avec pour données : 1^{ère} équipe, 2^{ème} équipe, le stade où a eu lieu le match, la ville où se trouve le stade, une estimation du nombre minimal de supporters dans le stade, une estimation du nombre maximal de supporters dans le stade et la date du match (aa/bb/cccc). Les éléments de données sont séparés les uns des autres par un espace (**non par une tabulation**). Les matchs sont classés des huitièmes de finale à la finale puis par ordre chronologique. Le passage d'une phase à élimination à l'autre se fait par le saut d'une ligne vierge.

Nom du fichier : phase à élimination directe euro 2016

- Une base de données sur les matchs de la phase de poule de la coupe du monde de rugby 2023. Chaque ligne correspond à un match avec pour données : 1^{ère} équipe, 2^{ème} équipe, le stade où a eu lieu le match, la ville où se trouve le stade, une estimation de nombre minimal de supporters dans le stade et une estimation du nombre maximal de supporters dans le stade. Les éléments de données sont séparés les uns des autres par un espace (**non par une tabulation**). Les matchs sont classés par groupe puis ordre chronologique. Le passage d'un groupe à l'autre se fait par le saut d'une ligne vierge.

Nom du fichier : groupes coupe du monde rugby 2023

Exemple :

France	Nouvelle_Zelande	Stade_de_France	Saint-Denis	75 000	80 000
--------	------------------	-----------------	-------------	--------	--------

Astuces :

- Pensez à lire toutes les consignes avant de commencer votre travail
- Pour transformer une chaîne de caractères en liste vous pouvez utiliser la commande suivante :
`L = my_string.split()`

Partie 1 : Mise en forme des bases de données

Les bases de données que Mr Guerin vous a transmises sont partiellement structurées. Dans un premier temps vous allez devoir effectuer un travail de mise en forme des données.

- 1) Vous pouvez constater qu'il existe une colonne indiquant la date du match (dernière colonne) pour les deux fichiers portant sur l'Euro. Cependant cette colonne n'existe pas dans le fichier sur la coupe du monde de Rugby. Mr Guerin ne s'intéresse pas à la date du match, supprimer cette colonne pour les deux fichiers.
- 2) Durant ces deux événements sportifs les équipes de France ont joué à domicile, les Français se sont déplacés en masse et ont rempli les stades pour encourager leurs équipes. Pour chacun des trois fichiers, modifier la colonne de l'estimation du nombre minimum de spectateurs dans le stade par la valeur de l'estimation du nombre maximum de spectateurs dans le stade pour chaque match des français. (La valeur de l'estimation maximale de supporteurs restera inchangée).
- 3) Il y a une erreur dans le nom de certains stades. « Stade_de_Nice » alors que son nom est « Allianz_Riviera ». De même pour la ville de Bordeaux, le nom affiché est « Stade_de_Bordeaux » alors que son nom est « Matmut_Atlantique ». Modifiez les noms des stades pour chacun des trois fichiers.

Partie 2 : Analyse global

- 1) Mr Guerin aimerait connaître la ville qui a le plus profité de ces événements sportifs et celle qui en a le moins. Indiquer le nom de la ville dont le stade a accueilli le plus de matchs et celle qui en a accueilli le moins, pour chacun des fichiers et pour tous les fichiers cumulés.
- 2) Il souhaiterait aussi connaître les noms des nations européennes majeures en Football et Rugby. On peut considérer qu'une nation est considérée comme une nation majeure en Football et Rugby à partir du moment où elle participe à la coupe du monde de Rugby et l'Euro de Football. Afficher les noms de ces nations majeures. Attention, au rugby la République d'Irlande et l'Irlande du Nord jouent sous même bannière contrairement au football, penser à en tenir compte dans votre réponse.
- 3) Il aurait aussi besoin de certaines informations statistiques (pour cette question vous ne tiendriez pas compte du fichier « phase à élimination directe euro 2016 ») :
 - Le nombre moyen de spectateurs à chaque match (borne minimum + borne maximum / 2)
 - Le nombre moyen de spectateurs présents lors de la phase de poules de l'Euro 2016
 - Le nombre de spectateurs moyen présents lors de la coupe du monde de Rugby 2023
 - Le nombre de spectateurs moyen phase de poules de l'Euro 2016 et coupe du monde de Rugby 2023 réunies.

Partie 3 : Étude économique sur le prix des places

À partir de maintenant, vous n'avez plus à utiliser le fichier « phase à élimination direct euro 2016 »

- 1) Les données économiques sur le prix des places sont absentes des fichiers, ça va être à vous de les rajouter. On estime que chaque match de poule de l'Euro 2016 a coûté aux alentours de 70€, sauf pour les matchs de la France et de l'Allemagne. Les matchs de la France étaient en moyenne à 90€ et les matchs du favori l'Allemagne étaient à 100€. Pour le Rugby le prix des places est en moyenne de 50€, sauf pour la France et l'Irlande première nation mondiale. Le prix moyen pour assister aux matchs des français et des irlandais était le même, soit 80€.
- 2) Enfin vous allez devoir faire une étude comparative des prix/revenus :
 - Calculer le revenu de chaque match. (prix * nombre moyen de spectateurs présents lors du match, pour chaque match).
 - Calculer la somme des revenus de la billetterie pour chacun des deux événements, puis pour les deux confondus.
 - Calculer le revenu moyen d'un match pour chacune des deux compétitions (somme des revenus/ nombre de matchs).
 - Calculer la variance des prix des matchs pour chaque tournoi.

Rappel de la formule de la Variance :

$$V(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

(Avec « x_i » le prix du i -ème match, « \bar{x} » le prix moyen d'un match et « n » le nombre total de match).

Conclusion :

Faites en sorte que quand Mr Guerin va lancer votre code, un message qui résume toutes les informations que vous avez obtenues s'affiche dans la console.

Mr Guerin vous remercie du travail bénévole que vous effectuez pour lui. 🙏