

bo_2024

Gama Junior Vertu

2024-11-11

1. Introduction

Contexte

Le Ballon d'Or est une récompense emblématique dans le monde du football, où les meilleurs joueurs de l'année sont sélectionnés par des journalistes du monde entier. Mais le système de points utilisé pour ces votes pourrait-il influencer le classement final ? L'idée ici est de tester différentes distributions de points pour voir si elles auraient changé les résultats pour les classements de 2024 et 2019. **C'est un exercice qui permet de mieux comprendre l'impact du système de notation, sans prétention d'analyse approfondie ou de recherche.**

Objectifs de l'analyse

Cette analyse cherche à observer l'effet de systèmes de points alternatifs sur les classements du Ballon d'Or pour deux années récentes, juste pour voir si le Top 10 aurait changé. Plus précisément :

- Regarder si le **Top 10 du Ballon d'Or 2024** aurait été différent avec la distribution des points de 2023.
- Regarder si le **Top 10 du Ballon d'Or 2019** aurait été différent en appliquant le système de points de 2017.

2. Méthodologie

Description des systèmes de points

En 2024, les journalistes attribuent des points aux 10 meilleurs joueurs selon une échelle : 15, 12, 10, 8, 7, 5, 4, 3, 2, 1. En 2023, la notation se limite aux 5 premiers avec une répartition des points en 6, 4, 3, 2, et 1. En 2017, seuls les 3 premiers recevaient des points (5, 3, et 1). La variation entre ces systèmes permet d'observer les effets potentiels sur les classements finaux.

Sources de données

Les données sont issues de sources en ligne officielles :

- Ballon d'Or 2024
- Ballon d'Or 2019

Ces données incluent les scores par pays, permettant de simuler divers systèmes de points.

Processus d'analyse

Pour chaque année, les données de classement sont récupérées et organisées pour correspondre aux différentes distributions de points. Ensuite, les scores sont calculés pour chaque joueur en appliquant les systèmes de points alternatifs, permettant d'établir un classement simulé et de le comparer avec le classement réel.

Chargement des données

les packages

```
library(rvest)
library(janitor)
library(dplyr)
library(tidyr)
library(formattable)
```

les données Importation des données

```
urls <- c("https://en.wikipedia.org/wiki/2024_Ballon_d%27Or" , "https://fr.wikipedia.org/wiki/Ballon_d%27Or")
bo_2024 <- read_html(urls[1])
bo_2019 <- read_html(urls[2])
```

1. Classement 2024 avec le système actuel :

Les journalistes attribuent des points à 10 joueurs, avec les points suivants : 15, 12, 10, 8, 7, 5, 4, 3, 2 et 1.

```
#récupération de la table sur les détails des votes par pays (2024)
votes_pays <- bo_2024 %>%
  html_elements("table") %>%
  html_table() %>%
  .[[4]] %>% # Erreur de retour : La 2em colonne devrait s'appeler "rodri"
  rename(rodri = colnames(.)[2]) %>% #Renomme la 2em colonne en "rodri"
  clean_names()

votes_pays %>%
  select(-country) %>%
  summarise(across(everything(), sum)) %>%
  pivot_longer(cols= everything() , names_to = "joueur" , values_to = "score") %>%
  top_n(10 , score) %>%
  arrange(desc(score)) %>%
  mutate(place = C(1:10)) %>%
  select(place , everything())
```

```
## # A tibble: 10 x 3
##   place joueur      score
##   <fct> <chr>      <int>
## 1 1      rodri      1170
## 2 2      vinicius_junior 1129
## 3 3      jude_bellingham 917
## 4 4      dani_carvajal 550
```

```
## 5 5      erling_haaland      432
## 6 6      kylian_mbappe      420
## 7 7      lautaro_martinez    402
## 8 8      lamine_yamal        383
## 9 9      toni_kroos          291
## 10 10    harry_kane          201
```

2. Classement 2024 avec le système alternatif de 2023 :

Et si comme en 2023 , les journalistes choisissaient que 5 joueurs et leur attribuaient les points suivants : 6, 4, 3, 2 et 1.

```
#resultat dans le 6 4 3 2 1
votes_pays %>%
  select(-country) %>%
  mutate(across(everything() , ~ case_when(
    . == 15 ~ 6 , #remplacement des points
    . == 12 ~ 4 ,
    . == 10 ~ 3 ,
    . == 8 ~ 2 ,
    . == 7 ~ 1 ,
    TRUE ~ 0 #les autres valeurs sont remplaces par zero
  ))) %>%
  summarise(across(everything(), sum)) %>%
  pivot_longer(cols= everything() , names_to = "joueur" , values_to = "score") %>%
  top_n(10 , score) %>%
  arrange(desc(score))%>%
  mutate(place = C(1:10)) %>%
  select(place , everything())
```

```
## # A tibble: 10 x 3
##   place joueur      score
##   <fct> <chr>      <dbl>
## 1 1      rodri        420
## 2 2      vinicius_junior 389
## 3 3      jude_bellingham 263
## 4 4      dani_carvajal   132
## 5 5      kylian_mbappe    71
## 6 6      lautaro_martinez 67
## 7 7      erling_haaland   64
## 8 8      lamine_yamal    61
## 9 9      toni_kroos      47
## 10 10    harry_kane      15
```

Classement 2019

Jusqu'en 2017, les journalistes étaient invités à choisir leurs 3 joueurs favoris, avec une attribution de points répartis de la manière suivante : 5 pour le premier, 3 pour le deuxième et 1 pour le troisième. Cependant, à partir de 2018, ce système a évolué : les journalistes devaient désormais sélectionner leurs 5 joueurs, avec une attribution de points suivant le classement 6 pour le premier, 4 pour le deuxième, 3 pour le troisième, 2 pour le quatrième et 1 pour le cinquième. faisons la meme analyse pour voir si le top 10 aurait changer si on applique la repartition de 2017

```

#recuperation de la table sur les details des votes par pays (2024)
pays_vote <- bo_2019 %>%
  html_elements("table") %>%
  html_table() %>%
  .[[5]] %>%      # recupre la 5 emme table
  #tout les colonnes de la table on le meme noms "pays votant"
  `colnames<-`(.[, ]) %>% #renome les colonnes avec le nom de la premiere ligne
  slice(-1 , -177) %>%
  clean_names %>%
  #les colonnes etant de type char , on les transforme en numeric
  mutate(across(everything(), as.numeric))

```

resultat dans le 6 4 3 2 1

```

## Warning: There were 2 warnings in 'mutate()'.
## The first warning was:
## i In argument: 'across(everything(), as.numeric)'.
## Caused by warning:
## ! NAs introduced by coercion
## i Run 'dplyr::last_dplyr_warnings()' to see the 1 remaining warning.

```

```

## Resultat dans le 6 4 3 2 1
pays_vote %>%
  select(-confederation , -pays_votant) %>%
  summarise(across(everything(), sum , na.rm= TRUE)) %>%
  pivot_longer(cols= everything() , names_to = "joueur" , values_to = "score") %>%
  top_n(10 , score) %>%
  arrange(desc(score))%>%
  mutate(place = c(1:10)) %>%
  select(place , everything())

```

```

## Warning: There was 1 warning in 'summarise()'.
## i In argument: 'across(everything(), sum, na.rm = TRUE)'.
## Caused by warning:
## ! The '...' argument of 'across()' is deprecated as of dplyr 1.1.0.
## Supply arguments directly to '.fns' through an anonymous function instead.
##
## # Previously
##   across(a:b, mean, na.rm = TRUE)
##
## # Now
##   across(a:b, \(x) mean(x, na.rm = TRUE))

```

```

## # A tibble: 10 x 3
##   place joueur score
##   <int> <chr>  <dbl>
## 1     1   mes    1371
## 2     2  vv_d    1355
## 3     3  cr7     951
## 4     4  mne     688
## 5     5  sal     354

```

```
## 6      6 mbp      178
## 7      7 bck      134
## 8      8 lew       88
## 9      9 slv       79
## 10     10 mhz       66
```

Classement 2019 avec le système de points de 2017

```
# resultat dans le 5 3 1
pays_vote %>%
  select(-confederation , -pays_votant) %>%
  mutate(across(everything() , ~case_when(
    . == 6 ~ 5 ,
    . == 4 ~ 3 ,
    . == 3 ~ 1 ,
    TRUE ~ 0
  ))) %>%
  summarise(across(everything(), sum , na.rm= TRUE)) %>%
  pivot_longer(cols= everything() , names_to = "joueur" , values_to = "score") %>%
  top_n(10 , score) %>%
  arrange(desc(score)) %>%
  mutate(place = c(1:12)) %>%
  select(place , everything())
```

```
## # A tibble: 12 x 3
##   place joueur score
##   <int> <chr>   <dbl>
## 1      1 vv_d     486
## 2      2 mes     474
## 3      3 cr7     249
## 4      4 mne     182
## 5      5 sal      58
## 6      6 mbp      28
## 7      7 bck      27
## 8      8 lew      15
## 9      9 mhz      11
## 10     10 slv       8
## 11     11 d_jg       8
## 12     12 haz       8
```

4. Discussion

Comparaison des classements

Les résultats indiquent qu'en passant d'un système basé sur 10 joueurs à un système limité à 5 ou 3 joueurs, les positions finales des joueurs sont modifiées. Par exemple, des joueurs ayant des scores modérés dans un système basé sur 10 joueurs voient souvent leurs positions décliner dans les classements issus des systèmes plus restreints, où la concentration des points est plus élevée.

Implications de la distribution des points

Dans les systèmes de points qui limitent le nombre de joueurs sélectionnés, les meilleurs joueurs sont davantage favorisés en tête des classements, en raison d’une répartition plus focalisée des points. Cela introduit des écarts notables dans les positions des joueurs qui, dans un système à 10 joueurs, se trouveraient dans des positions intermédiaires ou inférieures.

Limites de l’analyse

Les conclusions de cette analyse sont limitées aux données disponibles et ne cherchent pas à prédire les effets de changements de règles au-delà de ce contexte spécifique. De plus, cette simulation ne prend en compte que certains systèmes de points, et d’autres facteurs pouvant influencer les votes n’ont pas été inclus.

5. Conclusion

Résumé des principaux résultats

Cette analyse montre que la distribution des points influence la composition du Top 10 du Ballon d’Or, en particulier pour les joueurs dans les positions intermédiaires ou inférieures. En effet, les systèmes qui concentrent les points sur un nombre réduit de joueurs renforcent la position des leaders dans le classement, ce qui peut complexifier l’identification d’un “meilleur joueur” pour des saisons sans domination nette d’un joueur.