TF DA DJPPR: 3-session Workshop

Arif P. Sulistiono / @arifpras

2025-04-28

```
# Open directory: ctrl + shift + H
getwd()
```

```
## [1] "/Users/arifpras/Library/CloudStorage/OneDrive-Kemenkeu/01_Kemenkeu/TF_DataAnalytics"
```

```
# data/2023-24/gws/merged_gw.csv --> raw

tic()

# Method 1: Directly read the CSV file from URL
players00 <- read.csv("https://raw.githubusercontent.com/vaastav/Fantasy-Premier-League/refs/heads/maste

toc()
```

```
## 0.698 sec elapsed
```

```
# View the first few rows
head(players00)
```

```
##                     name position          team  xP assists bonus bps
## 1            Femi Seriki      DEF Sheffield Utd 0.5       0     0   0
## 2       Jack Hinshelwood      MID      Brighton 1.5       0     0   0
## 3           Jadon Sancho      MID       Man Utd 3.0       0     0   4
## 4 Rhys Norrington-Davies      DEF Sheffield Utd 0.1       0     0   0
## 5          Vitaly Janelt      MID     Brentford 2.1       0     0   6
## 6             Ionuț Radu       GK   Bournemouth 2.4       0     0   0
##   clean_sheets creativity element expected_assists expected_goal_involvements
## 1            0        0.0     653            0.00                       0.00
## 2            0        0.0     621            0.00                       0.00
## 3            0       11.3     397            0.05                       0.05
## 4            0        0.0     487            0.00                       0.00
## 5            0       11.5     105            0.01                       0.03
## 6            0        0.0     607            0.00                       0.00
##   expected_goals expected_goals_conceded fixture goals_conceded goals_scored
## 1           0.00                    0.00       7              0            0
## 2           0.00                    0.00       4              0            0
## 3           0.00                    1.08      10              0            0
## 4           0.00                    0.00       7              0            0
## 5           0.02                    1.26       8              2            0
## 6           0.00                    0.00       3              0            0
##   ict_index influence         kickoff_time minutes opponent_team own_goals
## 1       0.0       0.0 2023-08-12T14:00:00Z       0             8         0
## 2       0.0       0.0 2023-08-12T14:00:00Z       0            12         0
## 3       2.3       3.8 2023-08-14T19:00:00Z      22            20         0
```

```
## 4       0.0       0.0 2023-08-12T14:00:00Z       0             8         0
## 5       4.3      14.6 2023-08-13T13:00:00Z      90            18         0
## 6       0.0       0.0 2023-08-12T14:00:00Z       0            19         0
##   penalties_missed penalties_saved red_cards round saves selected starts
## 1                0               0         0     1     0        0      0
## 2                0               0         0     1     0      822      0
## 3                0               0         0     1     0    83993      0
## 4                0               0         0     1     0     6456      0
## 5                0               0         0     1     0     6508      1
## 6                0               0         0     1     0     8745      0
##   team_a_score team_h_score threat total_points transfers_balance transfers_in
## 1            1            0      0            0                 0            0
## 2            1            4      0            0                 0            0
## 3            0            1      8            1                 0            0
## 4            1            0      0            0                 0            0
## 5            2            2     17            2                 0            0
## 6            1            1      0            0                 0            0
##   transfers_out value was_home yellow_cards GW
## 1             0    40     True            0  1
## 2             0    45     True            0  1
## 3             0    70     True            0  1
## 4             0    40     True            0  1
## 5             0    55     True            0  1
## 6             0    45     True            0  1
```

```
tic()
# Method 2: Download first, then read
download.file(url = "https://raw.githubusercontent.com/vaastav/Fantasy-Premier-League/refs/heads/master/
              destfile = "merged_gw.csv")

# Then read the downloaded file
players00 <- read.csv("merged_gw.csv")
toc()
```

```
## 3.933 sec elapsed
```

```
# Method 3: Using readr from tidyverse
# if (!require("readr")) install.packages("readr")
# library(readr)

tic()
players00 <- read_csv("https://raw.githubusercontent.com/vaastav/Fantasy-Premier-League/refs/heads/maste
toc()
```

```
## 0.458 sec elapsed
```

```
# spec(players00)
```

```
# str(players00)
glimpse(players00)
```

```
## Rows: 29,725
## Columns: 41
## $ name                      <chr> "Femi Seriki", "Jack Hinshelwood", "Jadon S~
## $ position                  <chr> "DEF", "MID", "MID", "DEF", "MID", "GK", "M~
## $ team                      <chr> "Sheffield Utd", "Brighton", "Man Utd", "Sh~
## $ xP                        <dbl> 0.5, 1.5, 3.0, 0.1, 2.1, 2.4, 0.5, 4.1, 0.0~
```

```
## $ assists                   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ bonus                     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ bps                       <dbl> 0, 0, 4, 0, 6, 0, 3, 0, 0, 0, 0, 0, 1, 15, ~
## $ clean_sheets              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ creativity                <dbl> 0.0, 0.0, 11.3, 0.0, 11.5, 0.0, 0.0, 0.0, 0~
## $ element                   <dbl> 653, 621, 397, 487, 105, 607, 655, 354, 417~
## $ expected_assists          <dbl> 0.00, 0.00, 0.05, 0.00, 0.01, 0.00, 0.01, 0~
## $ expected_goal_involvements <dbl> 0.00, 0.00, 0.05, 0.00, 0.03, 0.00, 0.01, 0~
## $ expected_goals            <dbl> 0.00, 0.00, 0.00, 0.00, 0.02, 0.00, 0.00, 0~
## $ expected_goals_conceded   <dbl> 0.00, 0.00, 1.08, 0.00, 1.26, 0.00, 0.00, 0~
## $ fixture                   <dbl> 7, 4, 10, 7, 8, 3, 7, 1, 6, 10, 1, 4, 9, 9,~
## $ goals_conceded            <dbl> 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0~
## $ goals_scored              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ ict_index                 <dbl> 0.0, 0.0, 2.3, 0.0, 4.3, 0.0, 0.0, 0.0, 0.0~
## $ influence                 <dbl> 0.0, 0.0, 3.8, 0.0, 14.6, 0.0, 0.0, 0.0, 0.~
## $ kickoff_time              <dttm> 2023-08-12 14:00:00, 2023-08-12 14:00:00, ~
## $ minutes                   <dbl> 0, 0, 22, 0, 90, 0, 3, 0, 0, 0, 0, 0, 24, 7~
## $ opponent_team             <dbl> 8, 12, 20, 8, 18, 19, 8, 6, 2, 14, 13, 5, 7~
## $ own_goals                 <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ penalties_missed          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ penalties_saved           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ red_cards                 <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ round                     <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ saves                     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ selected                  <dbl> 0, 822, 83993, 6456, 6508, 8745, 0, 488362,~
## $ starts                    <dbl> 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0~
## $ team_a_score              <dbl> 1, 1, 0, 1, 2, 1, 1, 3, 1, 0, 3, 1, 1, 1, 2~
## $ team_h_score              <dbl> 0, 4, 1, 0, 2, 1, 0, 0, 5, 1, 0, 4, 1, 1, 2~
## $ threat                    <dbl> 0, 0, 8, 0, 17, 0, 0, 0, 0, 0, 0, 0, 0, 3, ~
## $ total_points              <dbl> 0, 0, 1, 0, 2, 0, 1, 0, 0, 0, 0, 0, 1, 2, 0~
## $ transfers_balance         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ transfers_in              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ transfers_out             <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ value                     <dbl> 40, 45, 70, 40, 55, 45, 45, 75, 45, 45, 40,~
## $ was_home                  <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, F~
## $ yellow_cards              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ GW                        <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
```

```
colSums(is.na(players00))
```

```
##                         name                     position
##                            0                            0
##                         team                          xP
##                            0                            0
##                      assists                        bonus
##                            0                            0
##                          bps                clean_sheets
##                            0                            0
##                   creativity                      element
##                            0                            0
##           expected_assists expected_goal_involvements
##                            0                            0
##             expected_goals    expected_goals_conceded
##                            0                            0
##                      fixture             goals_conceded
```

```
##                         0                         0
##               goals_scored                 ict_index
##                         0                         0
##                  influence              kickoff_time
##                         0                         0
##                    minutes             opponent_team
##                         0                         0
##                  own_goals          penalties_missed
##                         0                         0
##            penalties_saved                red_cards
##                         0                         0
##                      round                     saves
##                         0                         0
##                   selected                   starts
##                         0                         0
##              team_a_score              team_h_score
##                         0                         0
##                     threat              total_points
##                         0                         0
##          transfers_balance              transfers_in
##                         0                         0
##             transfers_out                     value
##                         0                         0
##                  was_home              yellow_cards
##                         0                         0
##                        GW
##                         0
```

```
# https://www.premierleague.com/news/3537201

teams00 <- read_csv("https://raw.githubusercontent.com/vaastav/Fantasy-Premier-League/refs/heads/master/
```

```
glimpse(teams00)
```

```
## Rows: 20
## Columns: 21
## $ code                  <dbl> 3, 7, 91, 94, 36, 90, 8, 31, 11, 54, 14, 102, 43~
## $ draw                  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ form                  <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ id                    <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 1~
## $ loss                  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ name                  <chr> "Arsenal", "Aston Villa", "Bournemouth", "Brentf~
## $ played                <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ points                <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ position              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ short_name            <chr> "ARS", "AVL", "BOU", "BRE", "BHA", "BUR", "CHE",~
## $ strength              <dbl> 5, 4, 3, 3, 3, 2, 3, 3, 3, 3, 4, 2, 5, 3, 3, 3, ~
## $ team_division         <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ unavailable           <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE,~
## $ win                   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ strength_overall_home <dbl> 1350, 1160, 1100, 1100, 1100, 1045, 1125, 1070, ~
## $ strength_overall_away <dbl> 1365, 1285, 1100, 1100, 1210, 1050, 1190, 1100, ~
## $ strength_attack_home  <dbl> 1370, 1140, 1055, 1110, 1070, 1050, 1080, 1080, ~
## $ strength_attack_away  <dbl> 1370, 1220, 1130, 1055, 1180, 1050, 1150, 1120, ~
## $ strength_defence_home <dbl> 1330, 1180, 1145, 1090, 1135, 1040, 1170, 1060, ~
```

```
## $ strength_defence_away <dbl> 1360, 1350, 1075, 1150, 1240, 1050, 1230, 1085, ~
## $ pulse_id              <dbl> 1, 2, 127, 130, 131, 43, 4, 6, 7, 34, 10, 163, 1~
```

Data Cleaning

```
# select()

teams01 <- teams00 %>%
  select(id, name, short_name)
```

```
# right_join(., by = c("" = ""))
# rename (new_name = old_name)
# relocate

players01 <- players00 %>%
  left_join(teams01, by = c("opponent_team" = "id")) %>%
  rename(
    "name" = name.x,
    "opponent_long" = name.y,
    "opponent_short" = short_name
  ) %>%
  relocate(opponent_long, opponent_short, .after = opponent_team)
```

```
# mutate
# unselected
# janitor::clean_names

players02 <- players01 %>%
  mutate("value_dec" = value/10) %>%
  relocate(value_dec, .after = value) %>%
  select(-value) %>%
  janitor::clean_names()
```

```
# arrange
# recode old = new

players03 <- players02 %>%
  arrange(team, position, name, gw) %>%
  mutate(position_long = recode (
    position,
    GK = "01gk ",
    DEF = "02def ",
    MID = "03mid ",
    FWD = "04fwd "
  )) %>%
  relocate(position_long, .after = position)

unique(players03$position_long)
```

```
## [1] "02def " "04fwd " "01gk "  "03mid "
```

```
# if_else
# str_squish
```

```r
players04 <- players03 %>%
  mutate(was_home_dv = if_else(was_home, 1, 0)) %>%
  mutate(across(where(is.character), str_squish)) %>%
  arrange(team, desc(position_long), name, gw)

unique(players04$position_long)
```

```
## [1] "04fwd" "03mid" "02def" "01gk"
```

```r
# function

squish_all_chars <- function(df) {
  message("Squishing all character columns...")
  df %>% mutate(across(where(is.character), str_squish))
}
```

```r
players05 <- squish_all_chars(players03)
```

```
## Squishing all character columns...
```

```r
unique(players05$position_long)
```

```
## [1] "02def" "04fwd" "01gk"  "03mid"
```

```r
(players06 <- players04 %>%
  filter(was_home_dv == 1))
```

```
## # A tibble: 14,867 x 45
##    name      position position_long team    x_p assists bonus   bps clean_sheets
##    <chr>     <chr>    <chr>         <chr> <dbl>   <dbl> <dbl> <dbl>        <dbl>
##  1 Eddie Nk~ FWD      04fwd         Arse~   2.4       0     2    26            1
##  2 Eddie Nk~ FWD      04fwd         Arse~   6.5       0     0    24            0
##  3 Eddie Nk~ FWD      04fwd         Arse~   5         0     0     8            0
##  4 Eddie Nk~ FWD      04fwd         Arse~   3         0     0     0            0
##  5 Eddie Nk~ FWD      04fwd         Arse~   2.2       0     0    -1            1
##  6 Eddie Nk~ FWD      04fwd         Arse~   7.3       0     3    77            1
##  7 Eddie Nk~ FWD      04fwd         Arse~   5.8       0     0     2            0
##  8 Eddie Nk~ FWD      04fwd         Arse~   1.7       0     0    -2            0
##  9 Eddie Nk~ FWD      04fwd         Arse~   1.9       1     0    15            0
## 10 Eddie Nk~ FWD      04fwd         Arse~   1.7       0     0     1            0
## # i 14,857 more rows
## # i 36 more variables: creativity <dbl>, element <dbl>, expected_assists <dbl>,
## #   expected_goal_involvements <dbl>, expected_goals <dbl>,
## #   expected_goals_conceded <dbl>, fixture <dbl>, goals_conceded <dbl>,
## #   goals_scored <dbl>, ict_index <dbl>, influence <dbl>, kickoff_time <dttm>,
## #   minutes <dbl>, opponent_team <dbl>, opponent_long <chr>,
## #   opponent_short <chr>, own_goals <dbl>, penalties_missed <dbl>, ...
```

```r
(players06 <- players04 %>%
  filter(team == "MUN" | opponent_short == "NFO"))
```

```
## # A tibble: 1,464 x 45
##    name      position position_long team    x_p assists bonus   bps clean_sheets
##    <chr>     <chr>    <chr>         <chr> <dbl>   <dbl> <dbl> <dbl>        <dbl>
##  1 Eddie Nk~ FWD      04fwd         Arse~   2.4       0     2    26            1
##  2 Eddie Nk~ FWD      04fwd         Arse~   3         0     0     2            0
##  3 Folarin ~ FWD      04fwd         Arse~   1.5       0     0     0            0
```

```
##  4 Folarin ~ FWD      04fwd         Arse~   0         0     0     0            0
##  5 Gabriel ~ FWD      04fwd         Arse~   0         0     0     0            0
##  6 Gabriel ~ FWD      04fwd         Arse~   8         1     2    31            1
##  7 Albert S~ MID      03mid         Arse~   2.1       0     0     0            0
##  8 Bradley ~ MID      03mid         Arse~   0.5       0     0     0            0
##  9 Bukayo S~ MID      03mid         Arse~   3.4       0     3    37            0
## 10 Bukayo S~ MID      03mid         Arse~   7         0     3    35            0
## # i 1,454 more rows
## # i 36 more variables: creativity <dbl>, element <dbl>, expected_assists <dbl>,
## #   expected_goal_involvements <dbl>, expected_goals <dbl>,
## #   expected_goals_conceded <dbl>, fixture <dbl>, goals_conceded <dbl>,
## #   goals_scored <dbl>, ict_index <dbl>, influence <dbl>, kickoff_time <dttm>,
## #   minutes <dbl>, opponent_team <dbl>, opponent_long <chr>,
## #   opponent_short <chr>, own_goals <dbl>, penalties_missed <dbl>, ...
```

```
(players06 <- players04 %>%
  filter(was_home_dv == 1 & team == "Man Utd"))
```

```
## # A tibble: 790 x 45
##    name      position position_long team    x_p assists bonus   bps clean_sheets
##    <chr>     <chr>    <chr>         <chr> <dbl>   <dbl> <dbl> <dbl>        <dbl>
##  1 Anthony ~ FWD      04fwd         Man ~   0.7       0     0     0            0
##  2 Anthony ~ FWD      04fwd         Man ~   1.2       0     0     3            0
##  3 Anthony ~ FWD      04fwd         Man ~   1.8       0     0     6            0
##  4 Anthony ~ FWD      04fwd         Man ~   1.2       0     0     5            0
##  5 Anthony ~ FWD      04fwd         Man ~   1.2       0     0     1            0
##  6 Anthony ~ FWD      04fwd         Man ~   0.5       0     0     3            0
##  7 Anthony ~ FWD      04fwd         Man ~   2         0     0     3            0
##  8 Anthony ~ FWD      04fwd         Man ~   4         0     0     0            0
##  9 Anthony ~ FWD      04fwd         Man ~   3.5       0     0     1            0
## 10 Anthony ~ FWD      04fwd         Man ~   0.2       0     0     0            0
## # i 780 more rows
## # i 36 more variables: creativity <dbl>, element <dbl>, expected_assists <dbl>,
## #   expected_goal_involvements <dbl>, expected_goals <dbl>,
## #   expected_goals_conceded <dbl>, fixture <dbl>, goals_conceded <dbl>,
## #   goals_scored <dbl>, ict_index <dbl>, influence <dbl>, kickoff_time <dttm>,
## #   minutes <dbl>, opponent_team <dbl>, opponent_long <chr>,
## #   opponent_short <chr>, own_goals <dbl>, penalties_missed <dbl>, ...
```

```
(players06 <- players04 %>%
  filter(was_home_dv == 1, team == "Man Utd"))
```

```
## # A tibble: 790 x 45
##    name      position position_long team    x_p assists bonus   bps clean_sheets
##    <chr>     <chr>    <chr>         <chr> <dbl>   <dbl> <dbl> <dbl>        <dbl>
##  1 Anthony ~ FWD      04fwd         Man ~   0.7       0     0     0            0
##  2 Anthony ~ FWD      04fwd         Man ~   1.2       0     0     3            0
##  3 Anthony ~ FWD      04fwd         Man ~   1.8       0     0     6            0
##  4 Anthony ~ FWD      04fwd         Man ~   1.2       0     0     5            0
##  5 Anthony ~ FWD      04fwd         Man ~   1.2       0     0     1            0
##  6 Anthony ~ FWD      04fwd         Man ~   0.5       0     0     3            0
##  7 Anthony ~ FWD      04fwd         Man ~   2         0     0     3            0
##  8 Anthony ~ FWD      04fwd         Man ~   4         0     0     0            0
##  9 Anthony ~ FWD      04fwd         Man ~   3.5       0     0     1            0
## 10 Anthony ~ FWD      04fwd         Man ~   0.2       0     0     0            0
```

```
## # i 780 more rows
## # i 36 more variables: creativity <dbl>, element <dbl>, expected_assists <dbl>,
## #   expected_goal_involvements <dbl>, expected_goals <dbl>,
## #   expected_goals_conceded <dbl>, fixture <dbl>, goals_conceded <dbl>,
## #   goals_scored <dbl>, ict_index <dbl>, influence <dbl>, kickoff_time <dttm>,
## #   minutes <dbl>, opponent_team <dbl>, opponent_long <chr>,
## #   opponent_short <chr>, own_goals <dbl>, penalties_missed <dbl>, ...
```

```
(players06 <- players04 %>%
  filter(was_home_dv == 1, team %in% c("Arsenal", "Liverpool", "Man City")))
```

```
## # A tibble: 1,949 x 45
##    name      position position_long team    x_p assists bonus   bps clean_sheets
##    <chr>     <chr>    <chr>         <chr> <dbl>   <dbl> <dbl> <dbl>        <dbl>
##  1 Eddie Nk~ FWD      04fwd         Arse~   2.4       0     2    26            1
##  2 Eddie Nk~ FWD      04fwd         Arse~   6.5       0     0    24            0
##  3 Eddie Nk~ FWD      04fwd         Arse~   5         0     0     8            0
##  4 Eddie Nk~ FWD      04fwd         Arse~   3         0     0     0            0
##  5 Eddie Nk~ FWD      04fwd         Arse~   2.2       0     0    -1            1
##  6 Eddie Nk~ FWD      04fwd         Arse~   7.3       0     3    77            1
##  7 Eddie Nk~ FWD      04fwd         Arse~   5.8       0     0     2            0
##  8 Eddie Nk~ FWD      04fwd         Arse~   1.7       0     0    -2            0
##  9 Eddie Nk~ FWD      04fwd         Arse~   1.9       1     0    15            0
## 10 Eddie Nk~ FWD      04fwd         Arse~   1.7       0     0     1            0
## # i 1,939 more rows
## # i 36 more variables: creativity <dbl>, element <dbl>, expected_assists <dbl>,
## #   expected_goal_involvements <dbl>, expected_goals <dbl>,
## #   expected_goals_conceded <dbl>, fixture <dbl>, goals_conceded <dbl>,
## #   goals_scored <dbl>, ict_index <dbl>, influence <dbl>, kickoff_time <dttm>,
## #   minutes <dbl>, opponent_team <dbl>, opponent_long <chr>,
## #   opponent_short <chr>, own_goals <dbl>, penalties_missed <dbl>, ...
```

```
(players06 <- players04 %>%
  filter(was_home_dv == 1, gw >= 19))
```

```
## # A tibble: 8,386 x 45
##    name      position position_long team    x_p assists bonus   bps clean_sheets
##    <chr>     <chr>    <chr>         <chr> <dbl>   <dbl> <dbl> <dbl>        <dbl>
##  1 Eddie Nk~ FWD      04fwd         Arse~   1.7       0     0     1            0
##  2 Eddie Nk~ FWD      04fwd         Arse~   2.8       1     0    16            0
##  3 Eddie Nk~ FWD      04fwd         Arse~   1.7       0     0     0            0
##  4 Eddie Nk~ FWD      04fwd         Arse~   0         0     0     8            0
##  5 Eddie Nk~ FWD      04fwd         Arse~   1.5       0     0     0            0
##  6 Eddie Nk~ FWD      04fwd         Arse~   1.8       0     0    -1            0
##  7 Eddie Nk~ FWD      04fwd         Arse~   1.3       0     0     0            0
##  8 Eddie Nk~ FWD      04fwd         Arse~   3         0     0     0            0
##  9 Eddie Nk~ FWD      04fwd         Arse~   1.2       0     0     0            0
## 10 Eddie Nk~ FWD      04fwd         Arse~   1         0     0     0            0
## # i 8,376 more rows
## # i 36 more variables: creativity <dbl>, element <dbl>, expected_assists <dbl>,
## #   expected_goal_involvements <dbl>, expected_goals <dbl>,
## #   expected_goals_conceded <dbl>, fixture <dbl>, goals_conceded <dbl>,
## #   goals_scored <dbl>, ict_index <dbl>, influence <dbl>, kickoff_time <dttm>,
## #   minutes <dbl>, opponent_team <dbl>, opponent_long <chr>,
## #   opponent_short <chr>, own_goals <dbl>, penalties_missed <dbl>, ...
```

```r
(players06 <- players04 %>%
  filter(gw >= 10 & gw <= 20))
```

```
## # A tibble: 8,152 x 45
##    name      position position_long team    x_p assists bonus   bps clean_sheets
##    <chr>     <chr>    <chr>         <chr> <dbl>   <dbl> <dbl> <dbl>        <dbl>
##  1 Eddie Nk~ FWD      04fwd         Arse~   7.3       0     3    77            1
##  2 Eddie Nk~ FWD      04fwd         Arse~   6.3       0     0     1            0
##  3 Eddie Nk~ FWD      04fwd         Arse~   5.8       0     0     2            0
##  4 Eddie Nk~ FWD      04fwd         Arse~   1.8       0     0     3            0
##  5 Eddie Nk~ FWD      04fwd         Arse~   1.7       0     0    -2            0
##  6 Eddie Nk~ FWD      04fwd         Arse~   1.8       0     0     0            0
##  7 Eddie Nk~ FWD      04fwd         Arse~   1.3       0     0     2            0
##  8 Eddie Nk~ FWD      04fwd         Arse~   1.9       1     0    15            0
##  9 Eddie Nk~ FWD      04fwd         Arse~   1.2       0     0    -2            0
## 10 Eddie Nk~ FWD      04fwd         Arse~   1.7       0     0     1            0
## # i 8,142 more rows
## # i 36 more variables: creativity <dbl>, element <dbl>, expected_assists <dbl>,
## #   expected_goal_involvements <dbl>, expected_goals <dbl>,
## #   expected_goals_conceded <dbl>, fixture <dbl>, goals_conceded <dbl>,
## #   goals_scored <dbl>, ict_index <dbl>, influence <dbl>, kickoff_time <dttm>,
## #   minutes <dbl>, opponent_team <dbl>, opponent_long <chr>,
## #   opponent_short <chr>, own_goals <dbl>, penalties_missed <dbl>, ...
```

```r
(players06 <- players04 %>%
    filter(between(gw, 10, 20)))
```

```
## # A tibble: 8,152 x 45
##    name      position position_long team    x_p assists bonus   bps clean_sheets
##    <chr>     <chr>    <chr>         <chr> <dbl>   <dbl> <dbl> <dbl>        <dbl>
##  1 Eddie Nk~ FWD      04fwd         Arse~   7.3       0     3    77            1
##  2 Eddie Nk~ FWD      04fwd         Arse~   6.3       0     0     1            0
##  3 Eddie Nk~ FWD      04fwd         Arse~   5.8       0     0     2            0
##  4 Eddie Nk~ FWD      04fwd         Arse~   1.8       0     0     3            0
##  5 Eddie Nk~ FWD      04fwd         Arse~   1.7       0     0    -2            0
##  6 Eddie Nk~ FWD      04fwd         Arse~   1.8       0     0     0            0
##  7 Eddie Nk~ FWD      04fwd         Arse~   1.3       0     0     2            0
##  8 Eddie Nk~ FWD      04fwd         Arse~   1.9       1     0    15            0
##  9 Eddie Nk~ FWD      04fwd         Arse~   1.2       0     0    -2            0
## 10 Eddie Nk~ FWD      04fwd         Arse~   1.7       0     0     1            0
## # i 8,142 more rows
## # i 36 more variables: creativity <dbl>, element <dbl>, expected_assists <dbl>,
## #   expected_goal_involvements <dbl>, expected_goals <dbl>,
## #   expected_goals_conceded <dbl>, fixture <dbl>, goals_conceded <dbl>,
## #   goals_scored <dbl>, ict_index <dbl>, influence <dbl>, kickoff_time <dttm>,
## #   minutes <dbl>, opponent_team <dbl>, opponent_long <chr>,
## #   opponent_short <chr>, own_goals <dbl>, penalties_missed <dbl>, ...
```

```r
# specific column
(players06 <- players04 %>%
  filter(is.na(position_long)))
```

```
## # A tibble: 0 x 45
## # i 45 variables: name <chr>, position <chr>, position_long <chr>, team <chr>,
## #   x_p <dbl>, assists <dbl>, bonus <dbl>, bps <dbl>, clean_sheets <dbl>,
```

```
## #   creativity <dbl>, element <dbl>, expected_assists <dbl>,
## #   expected_goal_involvements <dbl>, expected_goals <dbl>,
## #   expected_goals_conceded <dbl>, fixture <dbl>, goals_conceded <dbl>,
## #   goals_scored <dbl>, ict_index <dbl>, influence <dbl>, kickoff_time <dttm>,
## #   minutes <dbl>, opponent_team <dbl>, opponent_long <chr>, ...
```

```
# all columns
(players06 <- players04 %>%
  filter(if_any(everything(), is.na)))
```

```
## # A tibble: 0 x 45
## # i 45 variables: name <chr>, position <chr>, position_long <chr>, team <chr>,
## #   x_p <dbl>, assists <dbl>, bonus <dbl>, bps <dbl>, clean_sheets <dbl>,
## #   creativity <dbl>, element <dbl>, expected_assists <dbl>,
## #   expected_goal_involvements <dbl>, expected_goals <dbl>,
## #   expected_goals_conceded <dbl>, fixture <dbl>, goals_conceded <dbl>,
## #   goals_scored <dbl>, ict_index <dbl>, influence <dbl>, kickoff_time <dttm>,
## #   minutes <dbl>, opponent_team <dbl>, opponent_long <chr>, ...
```

```
(
  players07 <- players04 %>%
    group_by(team) %>%
    mutate(points_avgteam = mean(total_points)) %>%
    ungroup()
)
```

```
## # A tibble: 29,725 x 46
##    name      position position_long team    x_p assists bonus   bps clean_sheets
##    <chr>     <chr>    <chr>         <chr> <dbl>   <dbl> <dbl> <dbl>        <dbl>
##  1 Eddie Nk~ FWD      04fwd         Arse~   2.4       0     2    26            1
##  2 Eddie Nk~ FWD      04fwd         Arse~   7         1     0     6            1
##  3 Eddie Nk~ FWD      04fwd         Arse~   6.5       0     0    24            0
##  4 Eddie Nk~ FWD      04fwd         Arse~   5         0     0     8            0
##  5 Eddie Nk~ FWD      04fwd         Arse~   4         0     0     3            1
##  6 Eddie Nk~ FWD      04fwd         Arse~   3         0     0     0            0
##  7 Eddie Nk~ FWD      04fwd         Arse~   3.2       1     0    16            1
##  8 Eddie Nk~ FWD      04fwd         Arse~   2.2       0     0    -1            1
##  9 Eddie Nk~ FWD      04fwd         Arse~   2.8       0     0    -3            0
## 10 Eddie Nk~ FWD      04fwd         Arse~   7.3       0     3    77            1
## # i 29,715 more rows
## # i 37 more variables: creativity <dbl>, element <dbl>, expected_assists <dbl>,
## #   expected_goal_involvements <dbl>, expected_goals <dbl>,
## #   expected_goals_conceded <dbl>, fixture <dbl>, goals_conceded <dbl>,
## #   goals_scored <dbl>, ict_index <dbl>, influence <dbl>, kickoff_time <dttm>,
## #   minutes <dbl>, opponent_team <dbl>, opponent_long <chr>,
## #   opponent_short <chr>, own_goals <dbl>, penalties_missed <dbl>, ...
```

```
(
  players07 <- players04 %>%
    group_by(team) %>%
    summarise(points_avgteam = mean(total_points)) %>%
    ungroup() %>%
    arrange(desc(points_avgteam))
)
```

```
## # A tibble: 20 x 2
```

```
##    team           points_avgteam
##    <chr>                   <dbl>
##  1 Man City                1.80
##  2 Arsenal                 1.64
##  3 Liverpool               1.36
##  4 West Ham                1.24
##  5 Everton                 1.22
##  6 Crystal Palace          1.20
##  7 Fulham                  1.17
##  8 Aston Villa             1.17
##  9 Newcastle               1.12
## 10 Spurs                   1.12
## 11 Bournemouth             1.06
## 12 Brentford               1.01
## 13 Man Utd                 0.993
## 14 Wolves                  0.913
## 15 Brighton                0.910
## 16 Chelsea                 0.865
## 17 Burnley                 0.754
## 18 Luton                   0.743
## 19 Nott'm Forest           0.685
## 20 Sheffield Utd           0.659
```

```
(
  players07 <- players04 %>%
    group_by(team, name, position) %>%
    summarise(
      points_sumplayer = sum(total_points),
      minutes_played = sum(minutes),
      .groups = "drop") %>%
    ungroup() %>%
    arrange(desc(points_sumplayer))
)
```

```
## # A tibble: 893 x 5
##    team        name            position points_sumplayer minutes_played
##    <chr>       <chr>           <chr>                <dbl>          <dbl>
##  1 Chelsea     Cole Palmer     MID                    243           2607
##  2 Man City    Phil Foden      MID                    230           2860
##  3 Aston Villa Ollie Watkins   FWD                    228           3222
##  4 Arsenal     Bukayo Saka     MID                    226           2922
##  5 Man City    Erling Haaland  FWD                    217           2553
##  6 Spurs       Son Heung-min   MID                    213           2934
##  7 Liverpool   Mohamed Salah   MID                    211           2531
##  8 Arsenal     Martin Ødegaard MID                    186           3098
##  9 Newcastle   Anthony Gordon  MID                    183           2896
## 10 Arsenal     Benjamin White  DEF                    182           2987
## # i 883 more rows
```

```
(
  players08 <- players07 %>%
    relocate(position, .before = team) %>%
    group_by(position, team) %>%
    arrange(desc(points_sumplayer)) %>%
    slice_head(n=2) %>%
```

```
    ungroup()
)
```

```
## # A tibble: 160 x 5
##    position team        name                points_sumplayer minutes_played
##    <chr>    <chr>       <chr>                          <dbl>          <dbl>
##  1 DEF      Arsenal     Benjamin White                   182           2987
##  2 DEF      Arsenal     William Saliba                   164           3420
##  3 DEF      Aston Villa Ezri Konsa Ngoyo                  91           3069
##  4 DEF      Aston Villa Lucas Digne                       85           2405
##  5 DEF      Bournemouth Marcos Senesi                    102           2239
##  6 DEF      Bournemouth Illia Zabarnyi                    84           3330
##  7 DEF      Brentford   Nathan Collins                    84           2649
##  8 DEF      Brentford   Ethan Pinnock                     77           2521
##  9 DEF      Brighton    Lewis Dunk                        89           2869
## 10 DEF      Brighton    Jan Paul van Hecke                61           2368
## # i 150 more rows
```

```
(
  players08 <- players07 %>%
    relocate(position, .before = team) %>%
    group_by(position, team) %>%
    slice_max(order_by = points_sumplayer, n = 2) %>%
    ungroup()
)
```

```
## # A tibble: 170 x 5
##    position team        name              points_sumplayer minutes_played
##    <chr>    <chr>       <chr>                        <dbl>          <dbl>
##  1 DEF      Arsenal     Benjamin White                 182           2987
##  2 DEF      Arsenal     William Saliba                 164           3420
##  3 DEF      Aston Villa Ezri Konsa Ngoyo                91           3069
##  4 DEF      Aston Villa Lucas Digne                     85           2405
##  5 DEF      Aston Villa Pau Torres                      85           2462
##  6 DEF      Bournemouth Marcos Senesi                  102           2239
##  7 DEF      Bournemouth Illia Zabarnyi                  84           3330
##  8 DEF      Brentford   Nathan Collins                  84           2649
##  9 DEF      Brentford   Ethan Pinnock                   77           2521
## 10 DEF      Brighton    Lewis Dunk                      89           2869
## # i 160 more rows
```

Challenge: Find the Top 20 Players per Position (based on their average points per game, not total points)

```
(
  players09 <- players04 %>%
    filter(minutes != 0) %>%
    group_by(team, position_long, name) %>%
    summarise(
      count_played = n(),
      minutes_played = sum(minutes),
      points_sumplayer = sum(total_points),
      points_avgplayer = mean(total_points),
      .groups = "drop"
    ) %>%
    relocate(position_long, name, team) %>%
```

```
    group_by(position_long) %>%
    slice_max(order_by = points_avgplayer, n = 20) %>%
    ungroup()
)
```

```
## # A tibble: 80 x 7
##    position_long name         team  count_played minutes_played points_sumplayer
##    <chr>         <chr>        <chr>        <int>          <dbl>             <dbl>
##  1 01gk          Mark Travers Bour~            4            360                19
##  2 01gk          Arijanet Mu~ Burn~           10            900                47
##  3 01gk          David Raya ~ Arse~           32           2880               135
##  4 01gk          Jordan Pick~ Ever~           38           3420               153
##  5 01gk          Alisson Ram~ Live~           28           2520               107
##  6 01gk          Nick Pope    Newc~           15           1345                57
##  7 01gk          Alphonse Ar~ West~           31           2700               116
##  8 01gk          Robert Sánc~ Chel~           16           1433                59
##  9 01gk          Daniel Bent~ Wolv~            5            382                18
## 10 01gk          Bernd Leno   Fulh~           38           3420               133
## # i 70 more rows
## # i 1 more variable: points_avgplayer <dbl>
```

```
(
  players10 <- players04 %>%
    filter(minutes != 0) %>%
    group_by(team, position, name) %>%
    summarise(
      count_played = n(),
      minutes_played = sum(minutes),
      points_sumplayer = sum(total_points),
      points_avgplayer = mean(total_points),
      value_avg = mean(value_dec),
      ict_avg = mean(ict_index),
      .groups = "drop") %>%
    ungroup() %>%
    relocate(position, name, team) %>%
    group_by(position) %>%
    slice_max(order_by = points_sumplayer, n = 10) %>%
    # mutate(position_rank = row_number(desc(points_sumplayer))) %>%
    mutate(position_rank = min_rank(desc(points_sumplayer))) %>%
    ungroup()
)
```

```
## # A tibble: 40 x 10
##    position name              team  count_played minutes_played points_sumplayer
##    <chr>    <chr>             <chr>        <int>          <dbl>             <dbl>
##  1 DEF      Benjamin White    Arse~           37           2987               182
##  2 DEF      William Saliba    Arse~           38           3420               164
##  3 DEF      Gabriel dos Sant~ Arse~           36           3042               149
##  4 DEF      Pedro Porro       Spurs           35           3090               136
##  5 DEF      Jarrad Branthwai~ Ever~           35           3116               124
##  6 DEF      Joško Gvardiol    Man ~           28           2327               123
##  7 DEF      Kyle Walker       Man ~           32           2766               123
##  8 DEF      Fabian Schär      Newc~           36           3055               123
##  9 DEF      Trent Alexander-~ Live~           28           2153               122
## 10 DEF      Joachim Andersen  Crys~           38           3416               121
```

```
## # i 30 more rows
## # i 4 more variables: points_avgplayer <dbl>, value_avg <dbl>, ict_avg <dbl>,
## #   position_rank <int>
```

```
(
  players11 <- players10 %>%
    pivot_longer(
      cols = count_played:position_rank,
      names_to = "variable01",
      values_to = "value01"
    )
)
```

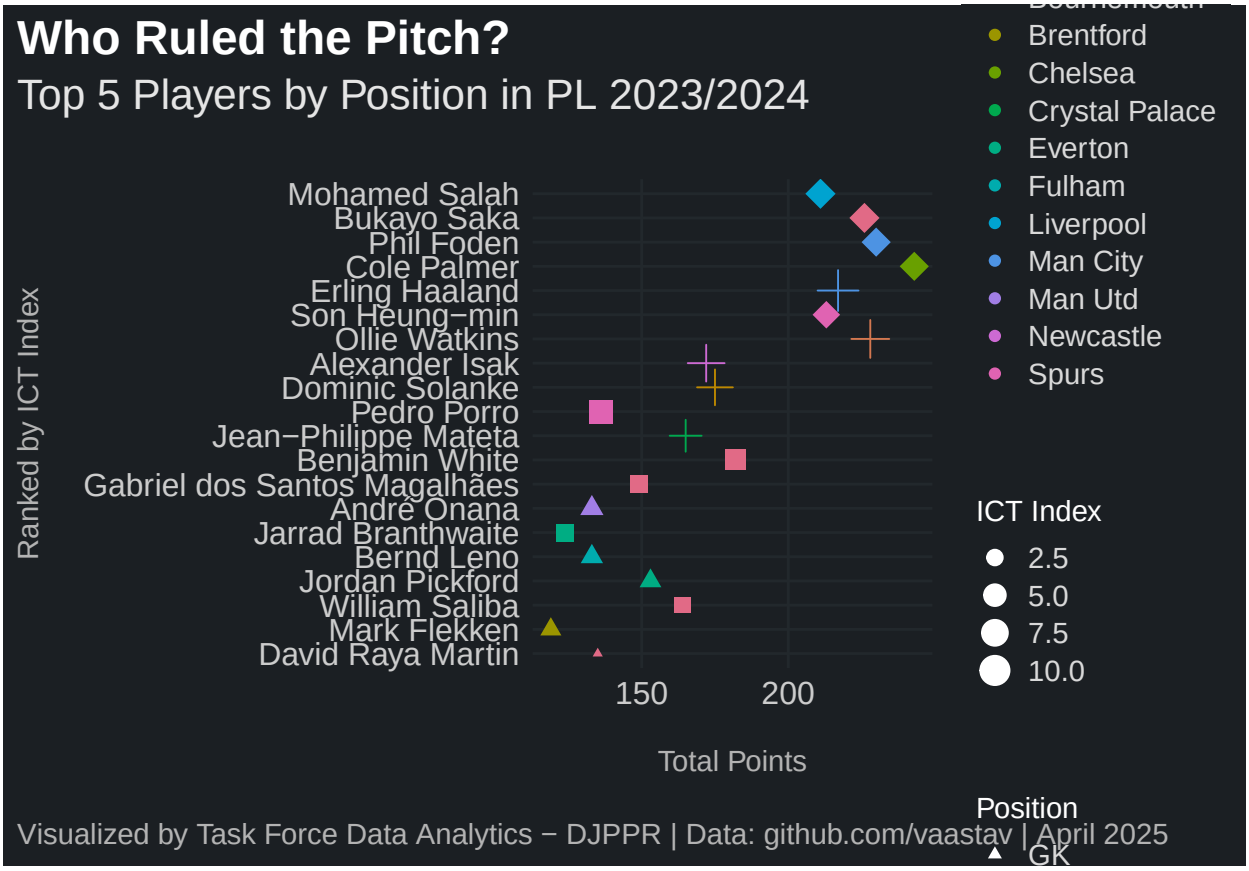
```
## # A tibble: 280 x 5
##    position name           team    variable01       value01
##    <chr>    <chr>          <chr>   <chr>              <dbl>
##  1 DEF      Benjamin White Arsenal count_played       37
##  2 DEF      Benjamin White Arsenal minutes_played   2987
##  3 DEF      Benjamin White Arsenal points_sumplayer  182
##  4 DEF      Benjamin White Arsenal points_avgplayer    4.92
##  5 DEF      Benjamin White Arsenal value_avg           5.69
##  6 DEF      Benjamin White Arsenal ict_avg             4.30
##  7 DEF      Benjamin White Arsenal position_rank       1
##  8 DEF      William Saliba Arsenal count_played       38
##  9 DEF      William Saliba Arsenal minutes_played   3420
## 10 DEF      William Saliba Arsenal points_sumplayer  164
## # i 270 more rows
```

```
writexl::write_xlsx(x = list(table_wide = players10, table_long = players11),
                    path = "fpldataset.xlsx")

# write_csv(players10, "fpldataset_tablewide.csv")
```

Data Viz

```
# Load nord palette (if needed)
# pacman::p_load(nord, scico, colorspace)

plot01 <- players10 %>%
  mutate(position = fct_relevel(position, "GK", "DEF", "MID", "FWD")) %>%
  group_by(position) %>%
  slice_max(order_by = points_sumplayer, n = 5) %>%
  ungroup() %>%
  arrange(desc(points_sumplayer)) %>%
  ggplot() +
  geom_point(
    aes(
      x = reorder(name, ict_avg),
      y = points_sumplayer,
      color = factor(team),
      size = ict_avg,
      shape = position
    )
  ) +
```

```
  scale_shape_manual(values = c("GK" = 17, "DEF" = 15, "MID" = 18, "FWD" = 3)) +
  colorspace::scale_color_discrete_qualitative(palette = "Dark 3") +
  scale_size_continuous(range = c(1, 5)) +
  coord_flip() +
  labs(
    title = "Who Ruled the Pitch?",
    subtitle = "Top 5 Players by Position in PL 2023/2024\n",
    caption = "\nVisualized by Task Force Data Analytics - DJPPR | Data: github.com/vaastav | April 2025
    x = "Ranked by ICT Index\n",
    y = "\nTotal Points"
  ) +
  guides(
    size = guide_legend("ICT Index", override.aes = list(color = "white")),
    color = guide_legend("Team"),
    shape = guide_legend("Position", override.aes = list(color = "white"))

  ) +
  theme_minimal(base_family = "sans") +
  theme(
    plot.background = element_rect(fill = "#1b1f23", color = NA),
    panel.background = element_rect(fill = "#1b1f23", color = NA),
    panel.grid.major.y = element_line(color = "#22282C"),
    panel.grid.major.x = element_line(color = "#22282C"),
    panel.grid.minor = element_blank(),

    plot.title = element_text(hjust = 0, size = 18, face = "bold", color = "white"),
    plot.subtitle = element_text(hjust = 0, size = 16, color = "gray90"),
    plot.caption = element_text(hjust = 0, size = 11, color = "gray70"),
    plot.title.position = "plot",
    plot.caption.position = "plot",

    axis.text.x = element_text(size = 12, color = "gray80"),
    axis.text.y = element_text(size = 12, color = "gray80"),
    axis.title.x = element_text(size = 11, color = "gray70"),
    axis.title.y = element_text(size = 11, color = "gray70"),
    axis.ticks = element_blank(),

    legend.background = element_rect(fill = "#1b1f23", color = NA),
    legend.title = element_text(size = 11, color = "white"),
    legend.text = element_text(size = 11, color = "gray85"),
    legend.key = element_rect(fill = "#1b1f23"),
    legend.key.size = unit(0.5, "cm"),
    legend.spacing = unit(1, "cm"),
    legend.position = "right",

    strip.text.x = element_text(size = 12, color = "white"),
    panel.ontop = FALSE,
    panel.spacing = unit(1, "lines")
  )

plot01
```




```r
ggsave(
  filename = "20250428_plot01_darkmode.png",
  plot = plot01,
  width = 3840,
  height = 2160,
  units = "px",
  dpi = 300
)
```

```r
# ggsave(
#   filename = "20250428_rcourse_plot01.pdf",
#   plot = plot01,
#   device = cairo_pdf,
#   width = 297,
#   height = 210,
#   units = "mm"
# )
```

— END —