r-eurocup-soccer

June 12, 2023

1 Challenge 2: EuroCup Soccer

[1]:
```
library(tidyverse)

url <- "https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/
 ↪02_Filtering_%26_Sorting/Euro12/Euro_2012_stats_TEAM.csv"
euro_stats <- read_csv(url)

print(euro_stats)
```

```
-- Attaching core tidyverse packages
------------------------------------------------------------------ tidyverse 2.0.0
--
v dplyr     1.1.2     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.4.2     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.1
-- Conflicts ------------------------------------------------------------
------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package
(<http://conflicted.r-lib.org/>) to force all conflicts to
become errors
Rows: 16 Columns: 35
-- Column specification --------------------------------
--------------------------------------------------------------------
Delimiter: ","
chr  (5): Team, Shooting Accuracy, % Goals-to-shots, Passing Accuracy,
Saves…
dbl (30): Goals, Shots on target, Shots off target, Total shots (inc.
Blocke…

i Use `spec()` to retrieve the full column specification for this
data.
i Specify the column types or set `show_col_types = FALSE` to quiet
```

this message.

```
# A tibble: 16 x 35
   Team           Goals `Shots on target` `Shots off target` `Shooting Accuracy`
   <chr>          <dbl>
<dbl>              <dbl>
<chr>
 1 Croatia            4                13                 12 51.9%
 2 Czech Republic     4                13                 18 41.9%
 3 Denmark            4                10                 10 50.0%
 4 England            5                11                 18 50.0%
 5 France             3                22                 24 37.9%
 6 Germany           10                32                 32 47.8%
 7 Greece             5                 8                 18 30.7%
 8 Italy              6                34                 45 43.0%
 9 Netherlands        2                12                 36 25.0%
10 Poland             2                15                 23 39.4%
11 Portugal           6                22                 42 34.3%
12 Republic of I~     1                 7                 12 36.8%
13 Russia             5                 9                 31 22.5%
14 Spain             12                42                 33 55.9%
15 Sweden             5                17                 19 47.2%
16 Ukraine            2                 7                 26 21.2%
# i 30 more variables: `% Goals-to-shots` <chr>,
#   `Total shots (inc. Blocked)` <dbl>, `Hit Woodwork` <dbl>,
#   `Penalty goals` <dbl>, `Penalties not scored` <dbl>, `Headed goals`
<dbl>,
#   Passes <dbl>, `Passes completed` <dbl>, `Passing Accuracy` <chr>,
#   Touches <dbl>, Crosses <dbl>, Dribbles <dbl>, `Corners Taken`
<dbl>,
#   Tackles <dbl>, Clearances <dbl>, Interceptions <dbl>,
#   `Clearances off line` <dbl>, `Clean Sheets` <dbl>, Blocks <dbl>,
…
```

1.1 How many teams participated in the Euro2012?

[2]:
```
num_distinct_teams <- n_distinct(euro_stats$Team)

print(num_distinct_teams)
```

```
[1] 16
```

1.2 What is the number of columns in the dataset?

[3]:
```
num_columns <- ncol(euro_stats)

print(num_columns)
```

```
[1] 35
```

1.3 View only the columns Team, Yellow Cards and Red Cards and assign them to a dataframe called discipline.

[4]:
```
discipline <- euro_stats %>%
  select(Team, `Yellow Cards`, `Red Cards`)

print(discipline)
```

```
# A tibble: 16 x 3
   Team                `Yellow Cards` `Red Cards`
   <chr>                        <dbl>
<dbl>
 1 Croatia                          9           0
 2 Czech Republic                   7           0
 3 Denmark                          4           0
 4 England                          5           0
 5 France                           6           0
 6 Germany                          4           0
 7 Greece                           9           1
 8 Italy                           16           0
 9 Netherlands                      5           0
10 Poland                           7           1
11 Portugal                        12           0
12 Republic of Ireland              6           1
13 Russia                           6           0
14 Spain                           11           0
15 Sweden                           7           0
16 Ukraine                          5           0
```

1.4 Sort the teams by Red Cards, then to Yellow Cards.

[5]:
```
sorted_discipline <- discipline %>%
  arrange(`Red Cards`, `Yellow Cards`)

print(sorted_discipline)
```

```
# A tibble: 16 x 3
   Team                `Yellow Cards` `Red Cards`
   <chr>                        <dbl>
<dbl>
 1 Denmark                          4           0
```

2	Germany	4	0
3	England	5	0
4	Netherlands	5	0
5	Ukraine	5	0
6	France	6	0
7	Russia	6	0
8	Czech Republic	7	0
9	Sweden	7	0
10	Croatia	9	0
11	Spain	11	0
12	Portugal	12	0
13	Italy	16	0
14	Republic of Ireland	6	1
15	Poland	7	1
16	Greece	9	1

1.5 Calculate the mean Yellow Cards given per Team.

[6]:
```
average_yellow_cards <- mean(discipline$`Yellow Cards`)

print(average_yellow_cards)
```

```
[1] 7.4375
```

1.6 Filter teams that scored more than 6 goals.

[7]:
```
filtered_teams <- euro_stats %>%
  filter(Goals > 6) %>%
  pull(Team)

print(filtered_teams)
```

```
[1] "Germany" "Spain"
```

1.7 Select the teams that start with the letter G.

[8]:
```
selected_teams <- euro_stats %>%
  filter(str_detect(Team, "^G")) %>%
  pull(Team)

print(selected_teams)
```

```
[1] "Germany" "Greece"
```

1.8 Select the first 7 columns.

[9]:
```
first_7_columns <- select(euro_stats, 1:7)

print(first_7_columns)
```

```
# A tibble: 16 x 7
   Team           Goals `Shots on target` `Shots off target` `Shooting Accuracy`
   <chr>          <dbl>
<dbl>              <dbl>
<chr>
 1 Croatia            4                13                 12 51.9%
 2 Czech Republic     4                13                 18 41.9%
 3 Denmark            4                10                 10 50.0%
 4 England            5                11                 18 50.0%
 5 France             3                22                 24 37.9%
 6 Germany           10                32                 32 47.8%
 7 Greece             5                 8                 18 30.7%
 8 Italy              6                34                 45 43.0%
 9 Netherlands        2                12                 36 25.0%
10 Poland             2                15                 23 39.4%
11 Portugal           6                22                 42 34.3%
12 Republic of I~     1                 7                 12 36.8%
13 Russia             5                 9                 31 22.5%
14 Spain             12                42                 33 55.9%
15 Sweden             5                17                 19 47.2%
16 Ukraine            2                 7                 26 21.2%
# i 2 more variables: `% Goals-to-shots` <chr>,
#   `Total shots (inc. Blocked)` <dbl>
```

1.9 Select all columns except the last 3.

[10]:
```
selected_columns <- euro_stats %>%
  select(-all_of((num_columns-2):num_columns))

print(selected_columns)
```

```
# A tibble: 16 x 32
   Team           Goals `Shots on target` `Shots off target` `Shooting Accuracy`
   <chr>          <dbl>
<dbl>              <dbl>
<chr>
 1 Croatia            4                13                 12 51.9%
 2 Czech Republic     4                13                 18 41.9%
 3 Denmark            4                10                 10 50.0%
 4 England            5                11                 18 50.0%
 5 France             3                22                 24 37.9%
 6 Germany           10                32                 32 47.8%
 7 Greece             5                 8                 18 30.7%
```

```
 8 Italy              6                34                45 43.0%
 9 Netherlands        2                12                36 25.0%
10 Poland             2                15                23 39.4%
11 Portugal           6                22                42 34.3%
12 Republic of I~     1                 7                12 36.8%
13 Russia             5                 9                31 22.5%
14 Spain             12                42                33 55.9%
15 Sweden             5                17                19 47.2%
16 Ukraine            2                 7                26 21.2%
# i 27 more variables: `% Goals-to-shots` <chr>,
#   `Total shots (inc. Blocked)` <dbl>, `Hit Woodwork` <dbl>,
#   `Penalty goals` <dbl>, `Penalties not scored` <dbl>, `Headed goals`
<dbl>,
#   Passes <dbl>, `Passes completed` <dbl>, `Passing Accuracy` <chr>,
#   Touches <dbl>, Crosses <dbl>, Dribbles <dbl>, `Corners Taken`
<dbl>,
#   Tackles <dbl>, Clearances <dbl>, Interceptions <dbl>,
#   `Clearances off line` <dbl>, `Clean Sheets` <dbl>, Blocks <dbl>,
…
```

1.10 Present only the Shooting Accuracy from England, Italy and Russia.

[11]:
```r
accuracy_subset <- euro_stats %>%
  filter(Team %in% c("England", "Italy", "Russia")) %>%
  select(Team, `Shooting Accuracy`)

print(accuracy_subset)
```

```
# A tibble: 3 x 2
  Team    `Shooting Accuracy`
  <chr>   <chr>
1 England 50.0%
2 Italy   43.0%
3 Russia  22.5%
```