ETC5521 Assignment 1

Your title

FILL FILL

2020-08-26

Contents

This assignment is for ETC5521 Assignment 1 by Team FILL comprising of FILL and FILL.

```
## Parsed with column specification:
## cols(
##   year = col_double(),
##   team = col_character(),
##   score = col_double(),
##   round = col_character(),
##   yearly_game_id = col_double(),
##   team_num = col_double(),
##   win_status = col_character()
## )
```

```
## Parsed with column specification:
## cols(
##   squad_no = col_double(),
##   country = col_character(),
##   pos = col_character(),
##   player = col_character(),
##   dob = col_datetime(format = ""),
##   age = col_double(),
##   caps = col_double(),
##   goals = col_double(),
##   club = col_character()
## )
```

```
## Parsed with column specification:
## cols(
##   country = col_character(),
##   team = col_character()
## )
```

1 Introduction and motivation:

Women's soccer has gained a lot of media attention in the last few years, and deservingly so. Countries like the USA, a dominant force in women's soccer, have produced some top quality players like Megan Rapinoe, Carli Lloyd, Alex Morgan and Julie Ertz. Just north of the USA, Christine Sinclair from Canada became the first player to win the Lou Marsh Award as Canadian Athlete of the Year^1, which is unsurprising considering her outstanding haul of 182 career international goals. Ada Hegerberg from Norway became the first woman to win the prestigious Ballon d'Or award, an yearly honor awarded to best player in the world, which was first introduced for the women's game in 2018^1.

The advancement of women's soccer has been parallel with the understanding and acceptance of data and analytics in the game. The technological advancement in capturing and analysing data has exploded since the late 1990s, when it began. Clubs have used this technology to identify and scout talent from all over the world.

Every country has a different approach towards coaching and developing talent. Success at an international level could be down to the ability of world class players playing with a certain level of chemistry, which could be challenging as it is common for them to be playing for different rival clubs at the domestic level. In the men's game, we saw Italy, Spain and Germany winning world cups in 2006, 2010 and 2014 respectively. Many partly attributed the success of these teams towards the fact that majority of these players playing in the same domestic league (or country), or even in the same club, especially those clubs that produce the most successful international players.

In this paper, we will approach the women's game with the same lens. We will look for any specific domestic clubs that contribute towards more successful world cup players, than others. We believe that these findings could prove meaningful to coaches, scouts and even young players in the game looking to choose clubs to play for with future international success in mind.

2 Data description:

This dataset is about Women's World Cup, which originally comes from website data.world https://data.world/sportsvizsunday/womens-world-cup-data. It consists of final score and win/loss status data from 1991-2019. Additionally, the 2019 world cup rosters for each team are included. After being cleaned by a data analyst, we downloaded the data from his github https://github.com/rfordatascience/tidytuesday/tree/master/data/2019/2019-07-09. There are three .csv files in this dataset to manage and store the data, and they are named "codes", "squads" and "wwc_outcomes". "wwc_outcomes.csv" has seven variables: "year" represents the year of tournament, "team" is the abbreviated team name, "score" means the score by team, "round" means the round of the tourament, "yearly_game_id" is a grouping variable for matches in each world cup, "team_num" is 1 or 2 representing teams in each match, and "win_status" represents a team win or loss. In "squads.csv", "squad_no" is the player jersey number from 1 to 23, "country" shows the country of the team, "pos" is the position of player, "player" shows the name of the player, "dob" is date of birth, "age" is the players' ages, "caps" displays the number of international games played by the player, "goal" is the number of international goals scored by the player, and "club" shows which domestic club the player belongs to.

3 Analysis and findings

[FILL] Should include at least one plot or numerical summary for each of your questions, that helps the reader arrive at an answer. You should also write paragraphs describing the methods, summaries and findings.

3.0.1 Section 1 - Clubs that provide the most number of world cup players:

3.0.1.1 Section 1a - Clubs that contribute the highest number of goal scorers - individual players assessment and overall number of goals scored.

Table 1: Top five clubs that providing most capped players

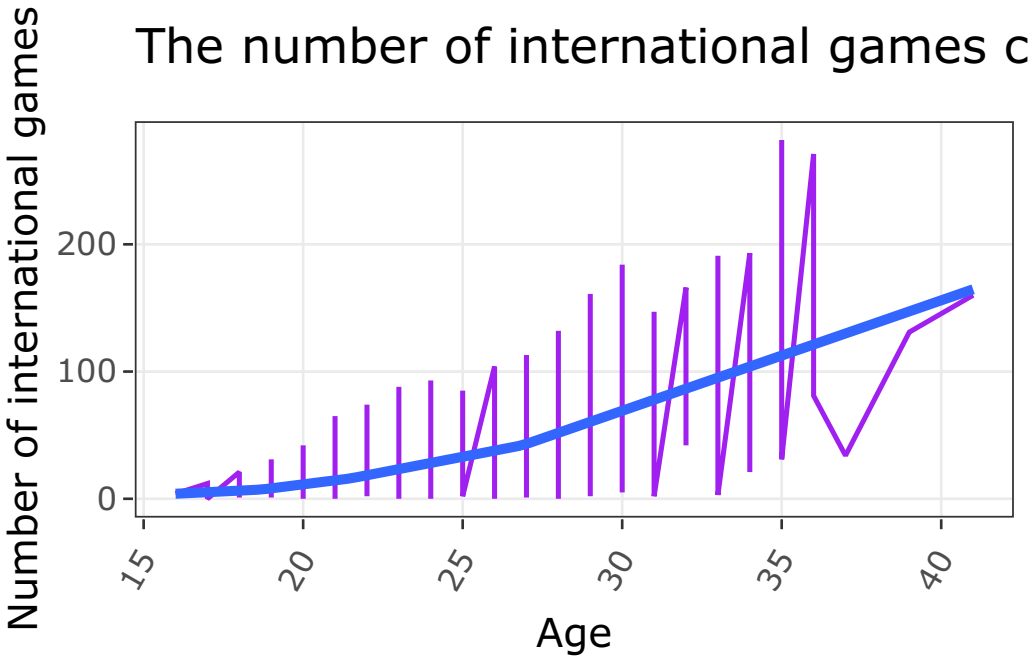
squad_no	country	pos	player	dob	age	caps	goals	club
12	Canada	FW	Christine Sinclair	1983-06-12	35	282	181	Portland Thorns
10	US	FW	Carli Lloyd	1982-07-16	36	271	107	Sky Blue FC
17	Sweden	MF	Caroline Seger	1985-03-19	34	193	27	Rosengård
6	Scotland	MF	Joanne Love	1985-12-06	33	191	13	Glasgow City
15	France	MF	Élise Bussaglia	1985-09-24	33	186	29	Dijon

3.0.2 Section 2 - Clubs that provide the most capped players:

In order to find the club providing player with the highest caps value, the "squads.csv" file is sorted in descending order, and the results top five clubs are displayed in the Table 1.

In this table, we could see that the club with most capped player is Portland Thorns. Caps represents the number of times a player has participated in international competitions. And we can notice that the top five capped players are all over 30 years old! Is there any relationship between caps and the age of players? To answer this question, I made a plot of caps changing by age. In Figure **??** the purple curve represents the trend of caps with age, while the blue curve represents the trend of conditional mean value of caps with age.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Conditional mean refers to the average level of caps predicted under the fixed age, so the characteristics of caps changing with age can be obtained more intuitively by observing the blue line. It can be seen that the blue line shows an upward trend with the increase of age, that is to say, the older a player is, the more times he participates in international competitions. It's not hard to understand that the older he is, the longer his career may be, the more games he will play. But we can't help but wonder whether there are players who are both young and highly capped? What club is he from?

3.0.2.1 Section 2a - Which are the clubs that provide players that are young and also highly capped - check for players under 25 years of age. First of all, we set the age under 25 as "young", and the caps value higher than the mean value belongs to "highly capped", and then it comes to analysis.

Table 2: Summary of clubs with players under 25 years old

squad_no	country	pos	player	dob	age	c
Min. : 1	Length:246	Length:246	Length:246	Min. :1993-06-10 00:00:00	Min. :16.00	M
1st Qu.: 9	Class :character	Class :character	Class :character	1st Qu.:1994-09-11 06:00:00	1st Qu.:21.00	1
Median :15	Mode :character	Mode :character	Mode :character	Median :1996-03-19 12:00:00	Median :23.00	M
Mean :14	NA	NA	NA	Mean :1996-07-01 09:45:21	Mean :22.45	M
3rd Qu.:19	NA	NA	NA	3rd Qu.:1997-12-18 18:00:00	3rd Qu.:24.00	3

Table 3: Top five clubs providing both young and highly capped players

squad_no	country	pos	player	dob	age	caps	goals	club
7	China PR	MF	Wang Shuang	1995-01-23	24	93	25	Paris Saint-Germain
3	Canada	DF	Kadeisha Buchanan	1995-11-05	23	88	3	Lyon
10	Australia	MF	Emily van Egmond	1993-07-12	25	85	18	Orlando Pride
14	Australia	DF	Alanna Kennedy	1995-01-21	24	77	7	Orlando Pride
10	Canada	DF	Ashley Lawrence	1995-06-11	23	76	5	Paris Saint-Germain

Make a summary to get the mean value of caps and the results are in Table 2. The mean value of caps is 22.45, which means the player with caps higher than 22.5 can be called “highly capped”.

Then select the players whose caps are higher than 22.15, and presenting the top five players with their clubs in Table 3. It can be seen that clubs Paris Saint-Germain,Lyon,Orlando Pride and Paris Saint-Germain all provide both young and highly capped players. Specially, there are two qualified players from the same club Orlando Pride, which means that it may be a club with more young players and more emphasis on training new players.

3.0.3 Section 3 - Clubs that provide the most:

a. GKs
b. DFs
c. MFs
d. FWDs

4 References

1. Best female soccer players of 2019 - https://sportmob.com/en/article/152786-best-female-soccer-players-of-2019
2. Smoothed conditional means - https://ggplot2.tidyverse.org/reference/geom_smooth.html