

Project Presentation

Kush Lalwani

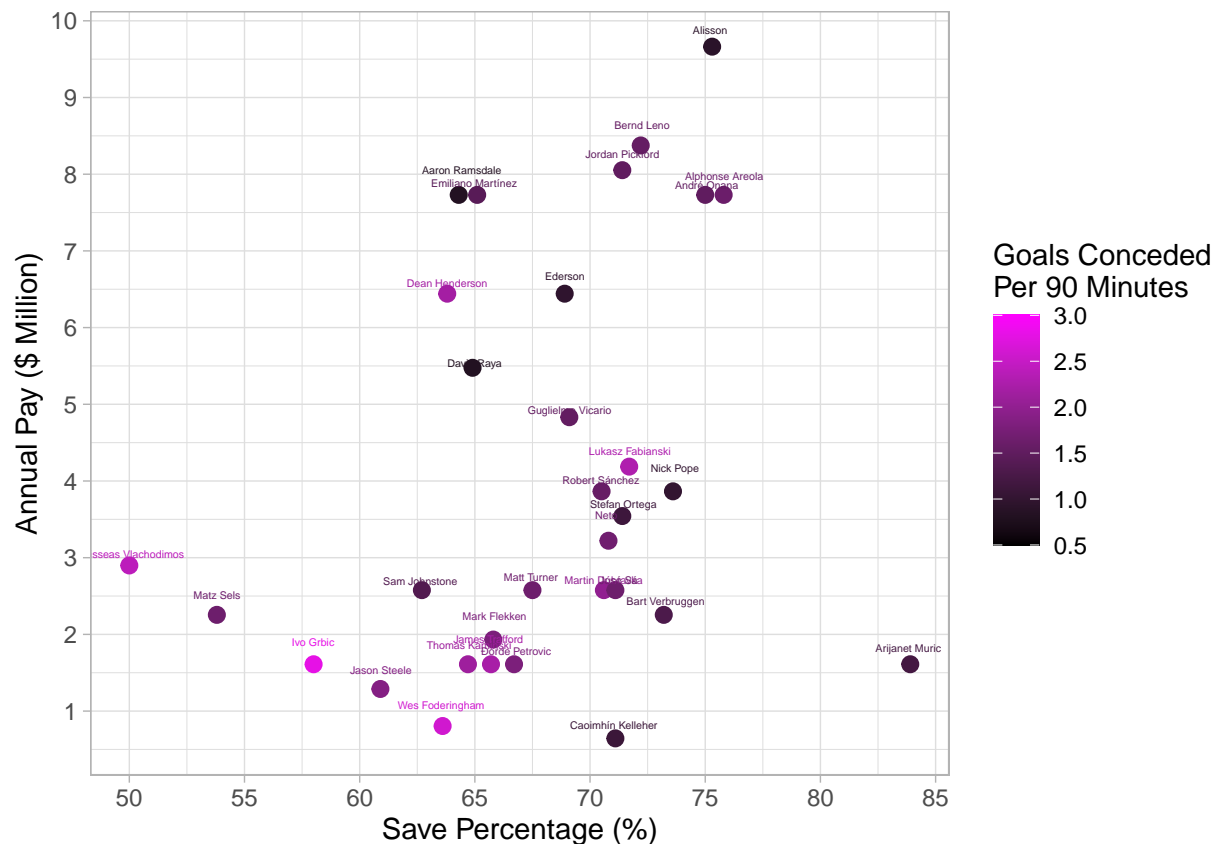
2024-04-23

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.4.4      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.0
## v purrr      1.0.2
## -- 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
##
## Attaching package: 'rvest'
##
##
## The following object is masked from 'package:readr':
##
##     guess_encoding
```

Research Question

I will be analyzing the relationship between a player's wages and their in game performance. The question is: Does a player's wages affect their performance in game? This question is important because it is important to understand if players are playing better or worse depending on their salary. This could help the teams determine if they should give their best players more money to keep their performances at a high level. Or it could also be used to decide the future of struggling players; whether they should get new contracts or sold.

Plot



Through this plot we can see the relationship between a goalkeeper's save percentage and their annual pay in millions of dollars. We also have the dots colored by the amount of goals each goalkeeper concedes per 90 minutes. We can see a slight positive linear trend in this data meaning that typically the higher paid players will have a higher save percentage. We also see that there are more pink colored points in the bottom left quarter of the graph. This means that players that are paid less and have a lower save percentage might concede more goals. This means shows that there is a relationship between the goalkeepers performance(save percentage/goals conceded) and their pay.

Challenge

One challenge that I faced was that the table of the goalkeeper's statistics would not read using the normal `read_html()` function. This was because the website used javascript to generate the table, meaning the function couldn't read it directly from the HTML code. To solve this issue, I used the `read_html_live()` function to read my table. This function uses google chrome in the background to open the website and read it directly from the chrome tab. This partially solved my problem, the one issue I still have is that sometimes the chrome tab cannot read the table within the 10 seconds the function allows.