

# 리그별 이적현황 크롤링

<https://github.com/Trigger21>

포지션 별 분류

# 리그별 이적현황 크롤링 이유

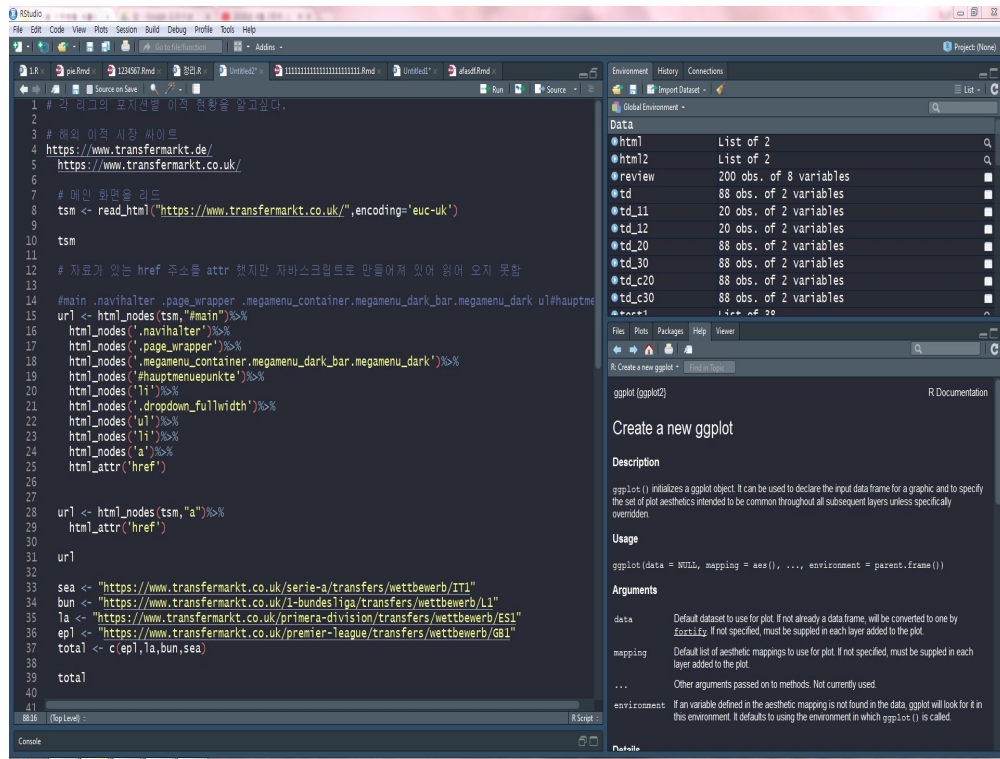
- 리그별 이적현황에서 알수있는 것들
  - 어느 포지션의 선호도가 높았는가?
  - 어느 포지션이 이적 금액이 높았는가?
  - 어느 리그가 이적시장이 활발한가?
  - 어느 리그가 어떤 스타일의 축구를 하는가?
- 크롤링 후 얻을수 있는 예측 정보들
  - 리그 팀별 ppr(페어 플레이 룰)을 알수있음
  - 리그별 선호도와 연도별 선호도를 비교
  - 이적시장 선수 포지션별 선수 금액을 알수있음
  - 이적시장 평균 나이를 분석 각 리그가 원하는 연령 확인

# 크롤링 모델 선정

크롤링을 어떤 모델로 할것인지  
결정

수업시간에 배운 R을 사용한 크롤링  
결정

- 사용할 패키지
  - library(rvest)
  - library(dplyr)
  - library(stringr)
  - library(wordcloud)
  - library(wordcloud2)
  - library(ggplot2)



The screenshot shows the RStudio interface. The left pane contains R code for web scraping using the rvest package. The code reads HTML from a URL, extracts specific elements using CSS selectors, and processes the data. The right pane shows the 'Environment' tab with a list of installed packages and their characteristics.

```
1 # 각 리그의 포지션별 이적 현황을 알고싶다.
2
3 # 해외 이적 시장 사이트
4 https://www.transfermarkt.de/
5 https://www.transfermarkt.co.uk/
6
7 # 메인 화면을 로드
8 tsm <- read_html("https://www.transfermarkt.co.uk/", encoding='euc-uk')
9
10 tsm
11
12 # 자료과 있는 href 주소를 attr 했지만 자바스크립트로 만들어져 있어 읽어 오지 못함
13
14 #main .navihalter .page_wrapper .megamenu_container.megamenu_dark_bar.megamenu_dark_u1#hauptne
15 url <- html_nodes(tsm, "#main") %>%
16   html_nodes(".navihalter") %>%
17   html_nodes(".page_wrapper") %>%
18   html_nodes(".megamenu_container.megamenu_dark_bar.megamenu_dark") %>%
19   html_nodes("#hauptmenuepunkte") %>%
20   html_nodes("li") %>%
21   html_nodes(".dropdown_fullwidth") %>%
22   html_nodes("ul") %>%
23   html_nodes("a") %>%
24   html_attr("href")
25
26
27 url <- html_nodes(tsm, "a") %>%
28   html_attr("href")
29
30 url
31
32 sea <- "https://www.transfermarkt.co.uk/serie-a/transfers/wettbewerb/IT1"
33 bun <- "https://www.transfermarkt.co.uk/1-bundesliga/transfers/wettbewerb/L1"
34 la <- "https://www.transfermarkt.co.uk/primera-division/transfers/wettbewerb/ES1"
35 ep1 <- "https://www.transfermarkt.co.uk/premier-league/transfers/wettbewerb/GB1"
36 total <- c(ep1, la, bun, sea)
37
38
39
40
41
```

The 'Environment' tab on the right shows the following data:

Package	Description
html	List of 2
html2	List of 2
review	200 obs. of 8 variables
td	88 obs. of 2 variables
td_11	20 obs. of 2 variables
td_12	20 obs. of 2 variables
td_20	88 obs. of 2 variables
td_30	88 obs. of 2 variables
td_c20	88 obs. of 2 variables
td_c30	88 obs. of 2 variables

The 'Packages' tab shows the following packages:

Package	Version
ggplot2	3.3.5

The 'Description' tab shows the following text:

Create a new ggplot

Description

ggplot() initializes a ggplot object. It can be used to declare the input data frame for a graphic and to specify the set of plot aesthetics intended to be common throughout all subsequent layers unless specifically overridden.

Usage

```
ggplot(data = NULL, mapping = aes(), ..., environment = parent.frame())
```

Arguments

data: Default dataset to use for plot. If not already a data frame, will be converted to one by `fortify`. If not specified, must be supplied in each layer added to the plot.

mapping: Default list of aesthetic mappings to use for plot. If not specified, must be supplied in each layer added to the plot.

...: Other arguments passed on to methods. Not currently used.

environment: If an variable defined in the aesthetic mapping is not found in the data, ggplot() will look for it in this environment. If defaults to using the environment in which ggplot() is called.

# 크롤링 대상

- 이적시장 현황을 알수있는 홈페이지 선별
- 대상 홈페이지에서 리그별 이적현황을 확인할수 있는지 확인
- 이적시장 기간 확인
- 크롤링 대상 홈페이지 선정
- <https://www.transfermarkt.co.uk/>

The screenshot displays the Premier League Transfermarkt website. The top navigation bar includes links for NEWS, TRANSFERS & RUMOURS, MARKET VALUES, COMPETITIONS, FORUMS, MY TM, and LIVE. The main content area is divided into several sections:

- PREMIER LEAGUE**: Overview of the league, including league level, size, players, foreign players, and market value. It also lists record champions, defending champions, and most valuable players.
- TRANSFERS 17/18**: A section for viewing transfers for the 17/18 season. It includes filters for select season, time of transfer, and transfer type (loans, club internal transfers).
- CHELSEA FC**: A table showing player transfers for Chelsea FC, including player name, age, position, market value, and transfer fee.
- TRANSFER RESULTS 17/18**: A section for viewing transfer results for the 17/18 season, including departures, arrivals, and total balance.

In	Age	Nat.	Position	Market value	From	Transfer fee
Alvaro Morata	24	Spain	Centre-Forward	€50.0m	Real Madrid	€50.0m
Tiémoué Bakayoko	22	Ivory Coast	Defensive Midfield	€24.4m	Monaco	€36.0m
Danny Drinkwater	27	England	Central Midfield	€5.0m	Leicester	€24.2m
Antonio Rüdiger	24	Germany	Centre-Back	€22.5m	AS Roma	€15.5m
David Dappena	25	France	Right-Back	€7.0m	Trondheim	€22.5m
Emerson	23	Brazil	Left-Back	€12.0m	AS Roma	€16.0m
Olivier Giroud	32	France	Centre-Forward	€10.0m	Arsenal	€15.5m
Ross Barkley	24	England	Attacking Midfield	€22.5m	Leiston	€15.2m

# 크롤링 이후 전처리

필요없는 정보, 광고등을 삭제하고  
알고자 하는 정보만을 추리고  
잘못입력되어 있거나 오류가 있던부분  
을 고쳐 다시 입력하는 작업 수행

원하는 이적시장 정보와 선수 이름,  
나이, 클럽명, 포지션, 이적금액등을  
전처리하여 정리

[[1]] 정제작업후 예

[1] ""	"Arrivals"
"Age"	
[4] "Nat."	"Position"
"Pos"	
[7] "Market value"	"Moving from"
"Transfer fee"	
[10] "Alvaro MorataA. Morata"	"24"
"Centre-Forward"	
[13] "CF"	"£36.00m"
"Real Madrid"	
[16] "£55.80m"	"Tiemoue
BakayokoT. Bakayoko"	"22"
[19] "Defensive Midfield"	"DM"
"£14.40m"	

## 크롤링 결과 단어

어떤 단어가 많이 나왔고 어떤  
포지션이 많이 언급되었으며  
어떤 클럽이 많은 영입을 했는지  
보기 쉽게 워드 클라우드  
형식으로 출력하였다.

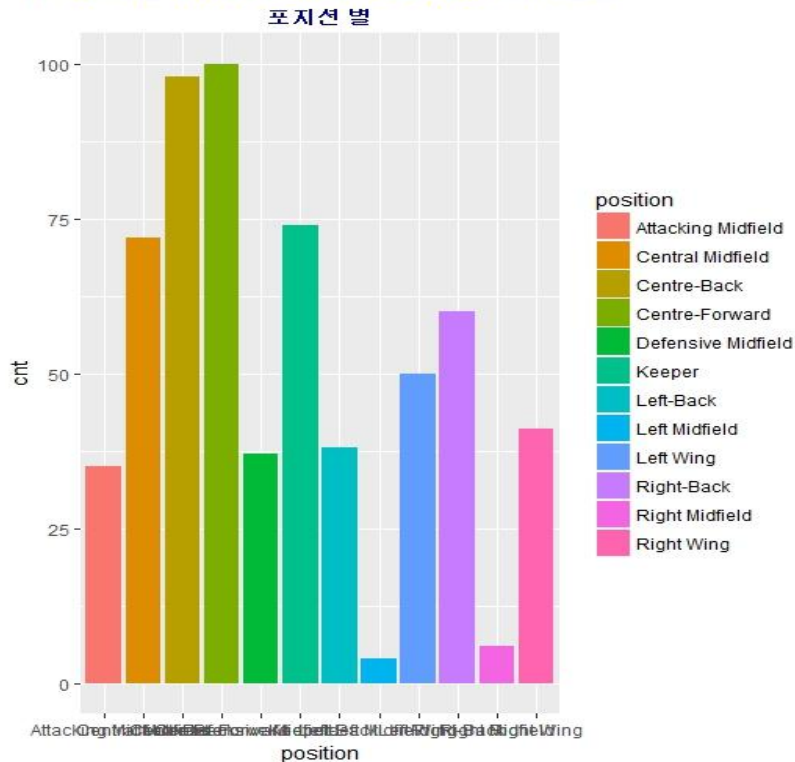
이번 이적시장에서 이적이 많았던 나이와 포지션 금액이 눈에 띄고 있다.



# EPL 이적 현황

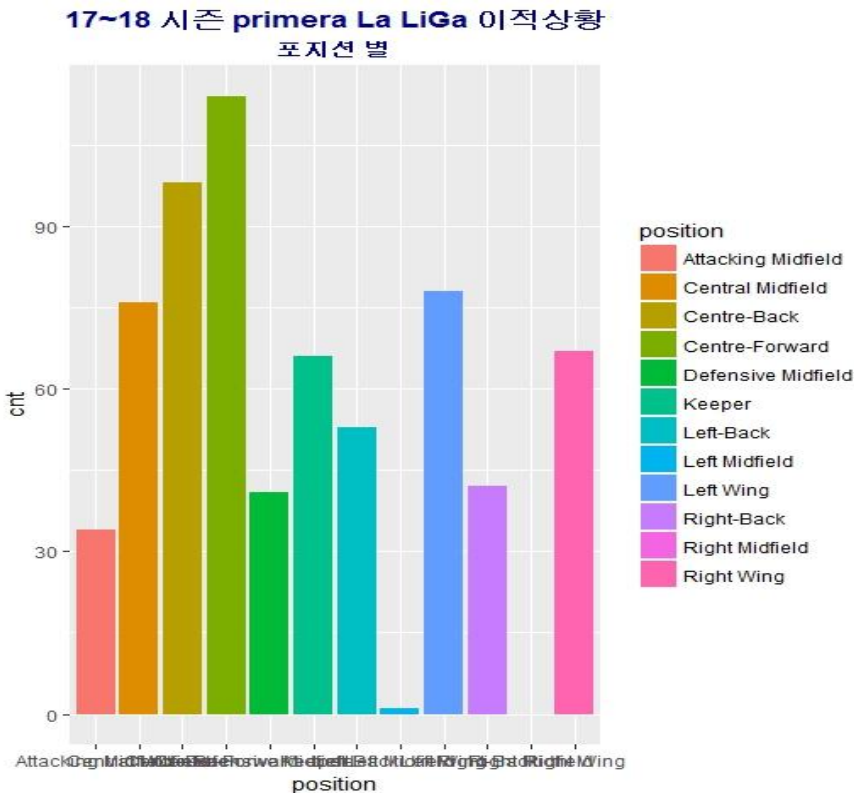
프리미어 출범이후 많은 감독과 선수들이 모여 잉글랜드 전통의 축구스타일을 벗어났다는 평가를 받고 있지만 아직 많은 구단들이 전통적인 잉글랜드식 롭볼을 구사하고 있고 이적시장 포지션현황에서도 센터 포워드, 센터 미드필더, 센터백의 이적 비율이 높은것으로 보아 아직 롭볼 플레이 스타일이 중요하다 여겨짐.

17~18 시즌 ENG PREMIER LEAGUE 이적상황



# La Liga 현황

메이저 리그중 가장 기술적인 부분이 부각되는 스페인 리그이다  
전체적인 부분에서 골고른 포지션이 영입되었다 스페인 리그의 짧은 패스 플레이 스타일이 그래프에서도 반영 되어 있지만  
메이저 리그 중에서 가장 영입수가 적었다 유스 시스템을 중시하는 리그의 특성과 이미 고착화 되어있는 팀간의 밸런스가 크게 작용한것 같다.



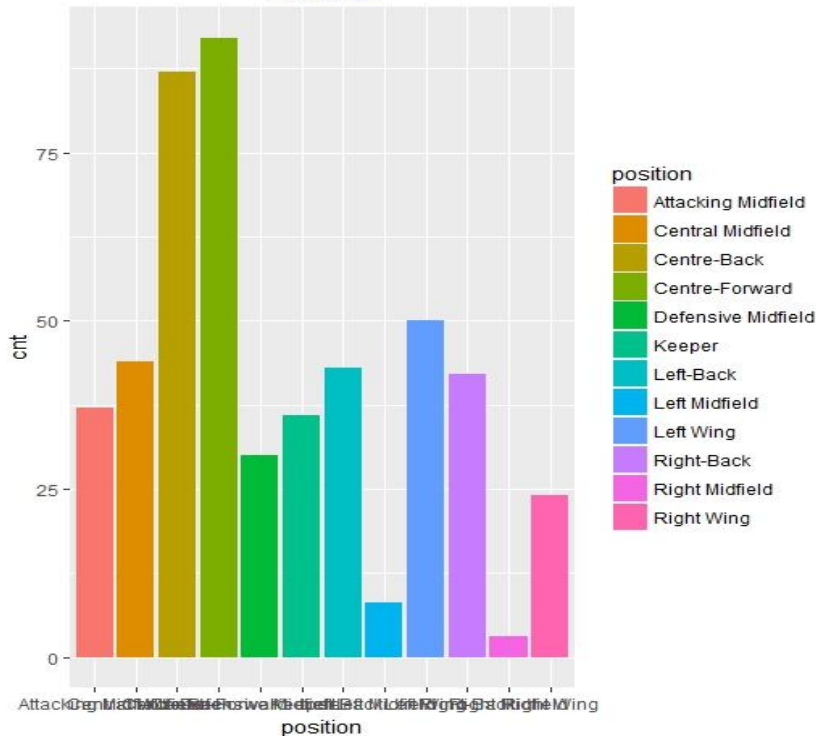


# BundesLiga 현황

새로이 떠오르고 있는 메이저 리그인 독일 리그이다. Epl과 플레이 스타일이 비슷하며 독일 특유의 힘과 압박을 중시하는 리그이다. 이적 현황 그래프에서도 고른 포지션을 원하였지만 센터 포워드와 센터백의 비중이 가장 높은 것을 볼수 있듯이 힘을 중시하는 리그의 성격이 들어났다.

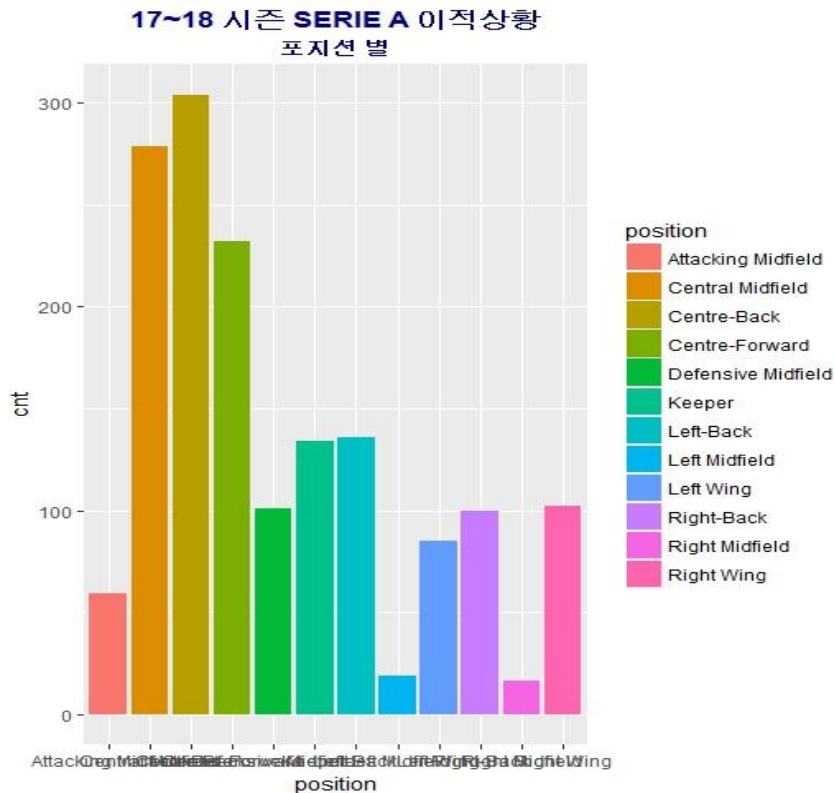
17~18 시즌 1 division.BUNDESLIGA 이적상황

포지션 별



# Serie A 현황

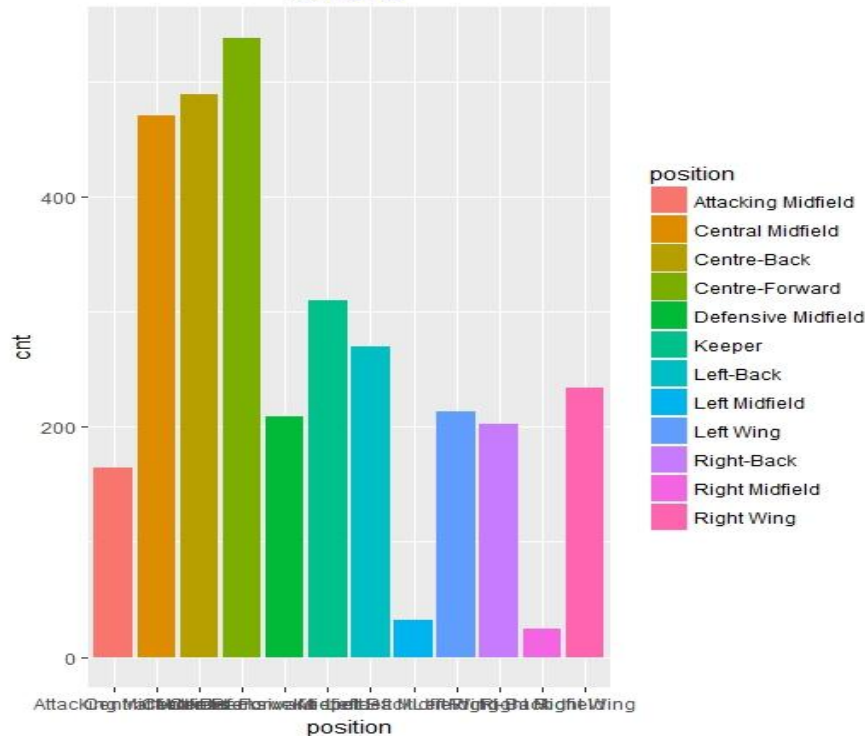
지금은 4대 메이저 리그에서 밀려난 세리아지만 인지도 면에서 프랑스 리그앙 보다 높아 추가를 하였다. 이탈리아 특유의 수비 축구와 전술 감독의 리그 라는 인식이 있지만 메이저 리그에 다시 오르겠다는 공격적인 투자와 중국 구단주 자본의 유입으로 인해 많은 포지션이 추가되었고 4대 리그중 가장 많은 이적수가 나왔다.



# 유럽 시장 현황

유럽 시장에서 가장 원하는 포지션은 척추라인(센터백, 센터미드필더, 센터포워드)으로 나왔다 많은 포지션이 이적하고 중요하다고 여겨지고 있지만 축구의 가장 기본인 척추라인과 몇년째 이어오고 있는 대형 센터포워드와 센터백의 부재가 많은 포지션 이동을 부추긴것 같다.

17~18 시즌 4대 메이저 리그 이적상황  
포지션 별



# 결론

이번 크롤링은 내가 선수를 가지고 마케팅 영업을 하는 에이전시라는 가정하에 진행해보았다.

내가 어느 포지션의 선수를 지금 영입해야하고 어느 리그에 소개, 영업을 해야하는지 알수 있도록 결과가 나온 크롤링이었다.

나온 결과와 같이 몇년간의 이적 시장을 크롤링 하면 미래의 이적시장 방향성을 구축할수 있을것 같다.