tut_06

R Markdown

Which countries have more than half of their squad playing their club games within their home country?
Store the abbreviated country names in q1_vec. The countries are Eng,Esp,Ger,Ksa,Rus

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##      filter, lag
```

```
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```
library(readxl)
```

```
country<-read_excel("C:/Users/Noorus Suhaina/Documents/DSA2101/data/country_fifa2018.xlsx")
player_info<-read_excel("C:/Users/Noorus Suhaina/Documents/DSA2101/data/player_info_fifa2018.xlsx")
result<-player_info%>%
  left_join(country,by=c("Country"="ShortName"))
#extract out their home countries in paranthesis
library(stringr)
```

```
## Warning: package 'stringr' was built under R version 4.0.5
```

```
home <- str_extract(player_info$Club,"(?<=\\()[^)]+")
#str_extract(player_info$Club,"(?<=\\()(\\w)+")
q1_vec<-player_info%>%
  mutate(home_country=home)%>%
  group_by(home_country)%>%
  summarize(count=sum(home_country==Country))%>%
  filter(count>12)%>%pull(home_country)
```

Question 2: Did any players play a match on their birthday? If yes, store the player names in q2_vec. If
no, set q2_vec to be NULL. left-join p-stats with player-info(jersey and country columns)

```
player_stats<-read_excel("C:/Users/Noorus Suhaina/Documents/DSA2101/data/player_stats_fifa2018.xlsx")
match_sched<-read_excel("C:/Users/Noorus Suhaina/Documents/DSA2101/data/match_schedule.xlsx")

player_info$DOB<-as.Date(player_info$DOB,format="%d/%m/%y")
player_stats$match_id<-as.character(player_stats$match_id)

vec<-left_join(player_stats,player_info,by=c("player_name"="Name"))%>%
  mutate(Day_month=strftime(DOB,format="%d/%m"))%>%
  left_join(match_sched,by=c("match_id"="match_id"))
vec$date_time<-as.Date(vec$date_time,format="%d %b %Y -%H:%M")
q2_vec<-vec%>%
  mutate(match_date=strftime(date_time,format="%d/%m"))%>%
  filter(Day_month==match_date)%>%
  select(player_name)%>%
  unique()%>%
  pull()
```

Question 3: Which country squads have defenders who play together at club level? Store the abbreviated country names in q3_vec.

```
tmp2<-player_info%>%filter(Pos=="DF")%>%
  group_by(Country)
q3_vec<-tmp2%>%summarize(distinct=n_distinct(Club),count=n())%>%
  filter(distinct<count)%>%
  pull()
```

Question 4

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

```
library(dplyr)
player_stats$Total[is.na(player_stats$Total)]="0/0"
tmp3<-player_stats%>%
  filter(Activity=="Tackles gaining/not gaining the ball")%>%
  separate(Total,sep="/",into=c("success","unsucc"))
tmp3$success<-as.numeric(tmp3$success)
tmp3$unsucc<-as.numeric(tmp3$unsucc)
tmp4<-tmp3%>%
  rowwise()%>%
  mutate(total_tackles=sum(c_across(success:unsucc)))%>%
  separate(time,into=c("minutes","seconds"))
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 1790 rows [1, 2, 3,
## 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
tmp4$minutes<-as.integer(tmp4$minutes)
tmp4$seconds<-as.integer(tmp4$seconds)
q4_vec<-tmp4%>%
  rowwise()%>%
```

```
  mutate(seconds=seconds/60,total_time=sum(c_across(minutes:seconds)))%>%
  filter(total_time>30)%>%
  mutate(average=total_tackles/total_time*90)%>%
  ungroup%>%
  group_by(player_name)%>%
  summarise(mean=mean(average))%>%
  arrange(desc(mean))%>%
  select(player_name)
q4_vec<-q4_vec[1:10,]%>%
  pull()
q4_vec
```

```
##  [1] "LEE Seungwoo"      "Yohan BEN ALOUANE" "Jacek GORALSKI"
##  [4] "KOO Jacheol"       "Daniel ARZANI"     "YAHIA ALSHEHRI"
##  [7] "AHMED FATHI"       "Ever BANEGA"       "Ghaylen CHAALELI"
## [10] "Cristhian STUANI"
```