Data Analysis

We inserted a large volume of actual data (EA Sports, 2021) into our database schema. To illustrate that our database is capable of provide business intelligence, we raised a few practical questions.

1. How does the total salaries of the BIG 6 clubs of English Premiere League change from season 14/15 to 20/21?

a. Code Implementation

-- The total salaries of the BIG 6 clubs of English Premiere League

-- from season 14/15 to 20/21

drop table if exists wage\_epl6;

create table wage\_epl6 as

select club\_name, player.season, sum(wage\_eur)

from player join club using(club\_id, season)

where club\_name in ("Arsenal", "Chelsea", "Tottenham Hotspur", "Manchester United", "Manchester City", "Liverpool")

group by club\_name, player.season;

select \*

from wage\_epl6;

b. Data Visualization

c. Key Business Intelligence

From season 14/15 to 20/21, in general, Manchester City had the highest total wage; Tottenham Hotspur had the lowest total wage before Arsenal took its position in season 18/19.

From season 16/17 to 19/20, the total wage of Arsenal dropped significantly from about 3.1 million to 2 million.

From season 14/15 to 20/21, Manchester City, Liverpool, and Tottenham Hotspur generally had an upward trend in total wage.

1. The rank of clubs in each of the "Big Five" European football leagues in terms of the total value of players in season 20/21.

a. Code Implementation

-- The rank of clubs in each of the "Big Five" European football

-- leagues in terms of the total value of players in season 20/21

drop table if exists club\_wage\_20\_21;

create table club\_wage\_20\_21 as

select league\_id, league\_name, club\_id, club\_name, sum(wage\_eur) as total\_wage\_eur

from player join club using(club\_id, season) join league using (league\_id, season)

where season = "20/21"

and tier = 1

and league.country\_region\_name in ("England", "Germany", "France", "Italy", "Spain")

group by club\_id;

select league\_name, club\_name, total\_wage\_eur, rank()

over (partition by league\_id order by total\_wage\_eur desc) as "rank"

from club\_wage\_20\_21;

b. Data Visualization

c. Key Business Intelligence

In season 20/21, compared with clubs of other leagues, clubs of English Premiere League and German 1. Bundesliga had relatively close total wages. In other words, their distributions of wages were more uniform. In English Premiere League, the highest total wage (Manchester City) was about only 6 times of the lowest one (Sheffield United). While in German 1. Bundesliga, the highest total wage (FC Bayern München) was also about 6 times of the lowest one (DSC Arminia Bielefeld).

In French League 1, Paris Saint-Germain had a total wage that was approximately two times of the second highest one (AS Monaco) and 10 times of the lowest one (Dijon FCO).

In Italian Serie A, Inter had a total wage that was about 13 times of the lowest one (Crotone).

In Spanish La Liga, Real Madrid and FC Barcelona had very close total wages (nearly 5 million euros), which were also the highest among the clubs of the Big Five leagues. They were over 20 times higher than the lowest (Elche CF).

1. A Comparison of the Big 3 Clubs of La Liga in Terms of Overall Ratings of Different Positions from Season 18/19 to 20/21.
2. Code Implementation

-- The average ratings of forwards, midfielders, defenders

-- (including goalkeepers) of the BIG 3 clubs of La Liga in   
-- season 18-19, 19-20, and 20-21

-- Assuming each player plays in his best position

-- (where he has the highest rating)

-- If a player has the same rating in two or more positions

-- that belong to different classes,

-- e.g. LW -> forward and CM -> midfielder,

-- then, he will be counted in both classes.

drop table if exists player\_best\_rating;

create table player\_best\_rating as

select player\_id, season, max(rating) as best\_rating

from player join club using(club\_id, season) join player\_positional\_rating using(player\_id, season) join general\_player using(player\_id, season)

where season in ("18/19", "19/20", "20/21")

and club\_name in ("Real Madrid", "FC Barcelona", "Atlético Madrid")

group by player\_id, season;

drop table if exists player\_best\_position\_class;

create table player\_best\_position\_class as

select distinct player\_id, season, player\_name, club\_id, club\_name, position\_class, rating as player\_best\_rating

from player join club using(club\_id, season) join player\_positional\_rating using (player\_id, season)

join positions using (position\_name) join player\_best\_rating using (player\_id, season)

where season in ("18/19", "19/20", "20/21")

and club\_name in ("Real Madrid", "FC Barcelona", "Atlético Madrid")

and rating = best\_rating;

insert into player\_best\_position\_class

select player\_id, season, player\_name, club\_id, club\_name, "defender" as position\_class, overall as player\_best\_rating

from player join club using(club\_id, season) join goalkeeper using(player\_id, season)

where season in ("18/19", "19/20", "20/21")

and club\_name in ("Real Madrid", "FC Barcelona", "Atlético Madrid");

select \* from player\_best\_position\_class;

select club\_name, season, position\_class,

avg(player\_best\_rating) as "average rating"

from player\_best\_position\_class

group by club\_name, season, position\_class with rollup;

1. Data Visualization
2. Key Business Intelligence

During the three seasons, Read Madrid and FC Barcelona had close overall ratings, while Atlético Madrid had a lower overall rating.

In season 19/20 and 20/21, Atlético Madrid’s ratings of defenders and midfielders were not far behind those of FC Barcelona and Real Madrid. However, the front line of Atlético Madrid had a great disparity with the others.

The rating of defenders, midfielders as well as overall rating of Atlético Madrid showed an upward trend during the three seasons.

References

SoFIFA.com. (2020). EA Sports FIFA series database. Retrieved April 10, 2021, from https://www.kaggle.com/stefanoleone992/fifa-21-complete-player-dataset?select=players\_21.csv