

Introduction

In this project, we will be analyzing data from the 2021/2022 season of the UEFA Champions League which is the biggest professional soccer tournament in Europe. The tournament comprises of the top teams from each regional league in Europe from the previous year, so it represents the highest level of the sport across Europe. The data comes from Kaggle which is linked [here](#). It contains multiple datasets, each containing important player information, team information, and statistics about a different aspect of the game throughout the tournament. Specifically, the eight datasets are: *Attacking*, *Defending*, *Goalkeeping*, *Goals*, *Attempts*, *Disciplinary*, *Distribution*, *Key Stats*.

The purpose of this project is to play the role of a data analyst for Real Madrid, who is one of the teams in the Champions League. Although Real Madrid won the Champions League in the 2021/2022 season, our task is to create a report for the players, coaching staff, and scouting team that discusses interesting and notable statistics from the past season of play. We want to see where our club and players stand with respect to the other clubs in terms of both the good and the bad. We may also be interested in which players and clubs are our biggest competition in certain areas of the sport, as well as who should be considered underdogs, dark horses, or maybe even overrated. Although there are a number of interesting questions we could ask about the data, we will focus on three general areas to investigate in this project. These are:

1. Goalkeeper analysis for the scouting team
2. Goals and tackle analysis for defenders
3. Passing, assists, and shots analysis for attackers

We go into depth about each of these areas and specify relevant questions of interest in the following sections. Below each question, we provide a table along with a brief answer to the question. Given our audience, we need to keep each answer straightforward and to the point. The players and staff simply look to our analysis as guidance when making their own decisions about the game.

We will utilize SAS and several skills learned in BIOS 669 to complete this task, including PROC SQL, PROC REPORT, derived variables, joins, and macros. All code for generating the figures and tables below can be found in the ZIP file containing this report.

1. Goalkeeper analysis for the scouting team

Question of interest for scouting team:

- *Our current goalkeeper (Thibaut Courtois) is out with an ACL injury, and we are in search of a temporary replacement for next year. To do this, we will analyze goalkeepers in the Champions League last year based on the three key metrics present in the goalkeeping data set. In particular, we want to know which goalkeepers had the most saves per match, conceded the least goals, and had the most clean sheets (games where they concede no goals). Ideally, we want to target the best performing goalkeeper in terms of all of these metrics, but the scouting team does prioritize the metrics by the order of them mentioned previously.*

2021/2022 UEFA Champions League: Top Performing Goalkeepers

Statistics				
Player	Club	Saves per game	Goals Conceded per game	Average Clean Sheets
Athanasiadis	Sheriff	5.80	1.60	0.20
Mignolet	Club Brugge	4.67	3.33	0.00
Ersin Destanoğlu	Beşiktaş	4.00	3.40	0.00
Vlachodimos	Benfica	3.80	1.70	0.50
Donnarumma	Paris	3.60	1.20	0.40
Casteels	Wolfsburg	3.60	1.60	0.20
Musso	Atalanta	3.50	2.17	0.17
Rulli	Villarreal	3.42	1.33	0.25
Diogo Costa	Porto	3.00	1.83	0.33
De Gea	Man. United	2.71	1.29	0.14

Table 1.

Job proj1_madhurir run on 26APR24 at 20:35

Based on Table 1, we see that the top three goalkeepers in the Champions League that have played at least five games, based on statistics alone, are Athanasiadis of Club Sheriff, Mignolet of Club Brugge, and Ersin Destanoglu of Besiktas. It is interesting to note that these top three keepers (and most of the keepers in this top ten list) are actually not the most famous or well known, and they don't play for the biggest teams either. This is great news because it means that we can likely buy a good goalkeeper for a fairly cheap price from these clubs. Although these goalkeepers play for not well known clubs, their statistics in the Champions League are evidence that they are capable of performing at the highest level, which is what we need from a replacement for Courtois.

2. Goals and tackle analysis for defenders

Questions of interest for defenders:

- *What is the most common type and area of goal scored by each team in the tournament? In other words, what kinds of goals should we be wary of?*

In Table 2 below, we see the most common type of goal scored by each team in the Champions League 2021/2022 season, both by types of goals and the areas the goals were scored from. We also see the specific break down of goals per team by type and area. Not surprisingly, most teams' goals were scored from inside the penalty area, which makes sense since it is closer to the goal. Most teams seem to prefer scoring with their right foot, although about 8 teams prefer scoring with their left foot more and two teams, Sheriff and Besiktas, have scored the most by headers. These details are good for the defenders to know when preparing for each opponent next year, since they can try to predict which players and shots are the most dangerous.

**2021/2022 UEFA Champions League:
Most Common Types of Goals Scored by Each Team**

Club	Type				Most Common Type	Area			Most Common Area
	Right Foot	Left Foot	Headers	Other		Inside Area	Outside Area	Penalties	
Ajax	5	12	3	1	Left Foot	20	1	1	Inside Area
Atalanta	4	7	1	0	Left Foot	10	2	0	Inside Area
Atlético	3	4	2	0	Left Foot	9	0	1	Inside Area
Barcelona	2	0	0	0	Right Foot	2	0	0	Inside Area
Bayern	13	12	3	1	Right Foot	26	4	3	Inside Area
Benfica	8	2	3	0	Right Foot	13	0	1	Inside Area
Beşiktaş	0	1	2	0	Headers	3	0	1	Inside Area
Chelsea	13	4	4	0	Right Foot	21	0	2	Inside Area
Club Brugge	4	1	0	0	Right Foot	5	0	0	Inside Area
Dortmund	5	3	2	0	Right Foot	10	0	2	Inside Area
Dynamo Kyiv	1	0	0	0	Right Foot	1	0	0	Inside Area
Inter	6	2	1	0	Right Foot	8	1	0	Inside Area
Juventus	2	6	3	0	Left Foot	11	0	2	Inside Area
LOSC	5	3	0	0	Right Foot	7	1	2	Inside Area
Leipzig	9	2	4	0	Right Foot	14	1	2	Inside Area
Liverpool	11	13	6	0	Left Foot	25	5	1	Inside Area
Malmö	1	0	0	0	Right Foot	1	0	0	Inside Area
Man. City	14	9	5	0	Right Foot	26	2	2	Inside Area
Man. United	9	2	1	0	Right Foot	9	3	0	Inside Area
Milan	3	1	1	0	Right Foot	5	0	0	Inside Area
Paris	7	7	1	0	Right Foot	13	2	2	Inside Area
Porto	1	0	1	0	Right Foot	2	0	1	Inside Area
Real Madrid	17	7	4	0	Right Foot	22	6	3	Inside Area
Salzburg	4	6	0	0	Left Foot	9	1	3	Inside Area
Sevilla	4	0	1	0	Right Foot	5	0	2	Inside Area
Shakhtar Do	2	0	0	0	Right Foot	2	0	0	Inside Area
Sheriff	1	2	3	1	Headers	5	2	0	Inside Area
Sporting CP	5	6	2	1	Left Foot	11	3	2	Inside Area
Villarreal	8	9	3	0	Left Foot	20	0	2	Inside Area
Wolfsburg	2	2	1	0	Right Foot	5	0	0	Inside Area
Young Boys	4	2	1	0	Right Foot	5	2	0	Inside Area
Zenit	5	2	2	0	Right Foot	7	2	1	Inside Area

Table 2.
Job proj2a_madhurir run on 28APR24 at 01:45

- *Who is the best at tackling on our team? In other words, who wins the highest percentage of tackles and who wins the most tackles on average per match? What about tackle attempts?*

2021/2022 UEFA Champions League: Top Tacklers on Real Madrid

<i>Player</i>	<i>Position</i>	<i>Tackling Statistics</i>		
		<i>Percentage of Tackles Won overall</i>	<i>Average Tackles Won per match</i>	<i>Average Tackles Attempted per match</i>
Mendy	Defender	83.33%	0.50	0.60
Alaba	Defender	66.67%	1.00	1.50
Valverde	Midfielder	53.85%	0.64	1.18
Camavinga	Midfielder	52.94%	0.90	1.70
Carvajal	Defender	50.00%	0.55	1.09
Kroos	Midfielder	45.45%	0.42	0.92
Casemiro	Midfielder	41.94%	1.18	2.82
Nacho	Defender	40.00%	0.44	1.11
Modrić	Midfielder	33.33%	0.23	0.69
Éder Militão	Defender	31.58%	0.50	1.58

Table 3.

Job proj2b_madhurir run on 28APR24 at 01:45

We see from Table 3 above an ordered list of the best tacklers on our team (Real Madrid). Specifically, we calculated the percentage of tackles won overall, average tackles won per match, and average tackles attempted per match for each player on the team and ranked the players in order of percentage of tackles won overall. We highlight in green the players with over 50% of tackles won overall. Based on this, Ferland Mendy is the best tackler on the team in this Champions League season. It is also great to see that even though this was Eduardo Camavinga's first season on Real Madrid at 19 years old, and even though he is a midfielder, he was a great tackler in this tournament. This is probably why he has recently been playing in defensive positions with David Alaba out with an ACL injury as well He is a very promising and flexible young player! On the other hand, Eder Militao and Nacho are two defenders with a bit worse tackle winrates, so perhaps this is an area that they can focus on improving.

3. Passing, assists, and shots analysis for attackers

Questions of interest for attackers (midfielders and forwards):

- *Who, on our team, has the best ratio of shots on target to shots off target? In other words, who is more likely to shoot on target when they take a shot?*

2021/2022 UEFA Champions League: Top Shot Takers on Real Madrid

Raw Data				Shot Statistics	
Player	Position	Shots On Target	Total Shot Attempts	Ratio of Shots On Target vs. Off Target	Percentage of Shots on Target
Kroos	Midfielder	7	12	7.00	58.33%
Benzema	Forward	23	45	1.77	51.11%
Modrić	Midfielder	5	14	1.67	35.71%
Rodrygo	Forward	10	23	1.43	43.48%
Éder Militão	Defender	4	9	0.80	44.44%
Vinícius Júnior	Forward	8	27	0.80	29.63%
Casemiro	Midfielder	2	10	0.33	20.00%
Valverde	Midfielder	0	9	0.00	0.00%

Table 4.

Job proj3a_madhurir run on 28APR24 at 01:45

Table 4 shows the top shot takers for Real Madrid and highlights the two players with over 50% of shots on target on average. Karim Benzema is one of the best players in the world and won the Ballon d’or (best player in the world of the year) right after this Champions League season, so the quality of his shots is not surprising at all. Toni Kroos is a highly experienced and talented midfielder, so we are also not surprised to see him top the list with seven shots on target to every one shot off target in the Champions League. Based on these statistics, we see that Vini Jr. and Valverde enjoy taking a of shots but most are not on target, so this can definitely be an area of improvement for them.

- *Who had the most assists per game and most passes completed per game in the league overall? What about from our team? And are these lists the same? In other words, are players who assist the most goals also the best passers in general during a game? This will help inform our game strategy.*

Finally, we examine Table 5 below which contains a list of the top 30 best passers in this Champions League season along with their respective number of assists, for all players with at least five matches played. This list is sorted by overall pass accuracy (total passes completed / total passes attempted) followed by average number of passes completed per game. Pass accuracy is an indicator of the quality of their passes, and average passes completed per game is an indicator of the actual number of good passes they make (quantity).

Before this analysis, we hypothesized that the best passers may also be more likely to assist goals, due to both the quantity and quality of their passes. However, we clearly see from this table that the players with highest pass accuracy (highlighted in grey) all have only one assist in this Champions League season. They have both quantity and quality of passes but very low assist numbers. When we think about this some more, it actually makes sense. During a soccer game, defenders often try to slow down the game and "regroup" by keeping the ball in their half of the field and passing slowly between each other and their goalkeeper, before making an attacking play. They are rarely ever contested by opposing players when doing this, which explains why their pass accuracy is so high. Additionally, it is rare for defenders to assist goals by nature of their role and position on the field, which explains the low number of assists. We also see our players bolded in this list. Ferland Mendy, who was our top tackler, is fairly high on this list (top 10) based on pass accuracy and has two assists. Luka Modric is also on this list with high pass accuracy and four assists as well. This is not surprising at all because like Toni Kroos, Modric is a very experienced player often nicknamed "the magician", so his passes as a midfielder are more likely to be contested by opposing defenders but are still very high quality!

2021/2022 UEFA Champions League: Best Passers in the Tournament

<i>Pass Statistics</i>					
<i>Player</i>	<i>Club</i>	<i>Position</i>	<i>Assists</i>	<i>Pass Accuracy</i>	<i>Average Passes Completed per game</i>
Martínez	Ajax	Defender	1	94.1%	79.13
Christensen	Chelsea	Defender	1	93.4%	64.63
Thiago Silva	Chelsea	Defender	1	93.1%	63.44
Marquinhos	Paris	Defender	1	92.5%	59.38
Aurier	Villarreal	Defender	1	91.8%	11.60
Bernardo Silva	Man. City	Midfielder	3	91.6%	38.82
Jorginho	Chelsea	Midfielder	2	91.5%	59.63
Mendy	Real Madrid	Defender	2	91.1%	37.60
Neuer	Bayern	Goalkeeper	1	90.9%	35.33
Gündoğan	Man. City	Midfielder	2	90.8%	36.70
Van Dijk	Liverpool	Defender	1	90.7%	67.67
Vidal	Inter	Midfielder	2	90.4%	31.71
Hudson-Odoi	Chelsea	Midfielder	2	90.2%	26.80
Pavard	Bayern	Defender	1	90.2%	61.30
Modrić	Real Madrid	Midfielder	4	89.8%	55.62
Kanté	Chelsea	Midfielder	1	89.8%	40.83
Zappacosta	Atalanta	Defender	1	89.7%	27.00
Danilo	Juventus	Defender	1	89.6%	66.20
Sterling	Man. City	Forward	2	88.4%	15.42
João Cancelo	Man. City	Defender	3	88.0%	65.11
Azpilicueta	Chelsea	Defender	1	88.0%	48.22
Zinchenko	Man. City	Defender	1	88.0%	39.25
Keïta	Liverpool	Midfielder	2	87.9%	31.30
Camavinga	Real Madrid	Midfielder	1	87.7%	15.80
Gravenberch	Ajax	Midfielder	1	87.4%	42.75
Carvajal	Real Madrid	Defender	1	87.3%	45.73
Capoue	Villarreal	Midfielder	4	87.2%	36.50
Bernardeschi	Juventus	Midfielder	2	87.2%	22.20
Grealish	Man. City	Midfielder	1	87.2%	30.57
Dybala	Juventus	Forward	1	87.0%	31.20

Table 5.

Job proj3b_madhurir run on 28APR24 at 06:05

Conclusion

In this project, we conducted data analysis of UEFA Champions League data from the 2021/2022 season as an analyst for the club Real Madrid. We utilized tools in SAS that we learned from BIOS 669, including PROC SQL, PROC REPORT, joins, derived variables, etc. We created several reports to address questions of interest from a variety of people on the team. Specifically, our analysis benefitted/was for the scouting team, the defenders, and the attackers on the team. We hope that the key takeaways we summarized for each question and table will be useful for the team's improvement and decision-making for next season.