Project #1 - Retrospective Analysis (STL City SC)

—Feng Jiang, Yuyao Zhuge

Player Physical Output

Team(ST louis) (table 0)

	JerseyNumber	Minutes_Played	Total_Distance	High_Speed_Distance	Sprint_Distance	Sprint_Appearance	Sprint_Count
0	1	100.165000	6240.750000	78.724667	0.000000	0.0	0.0
1	2	99.045556	11688.446667	702.792000	219.651333	861.0	8.0
2	6	83.178333	10105.965000	522.319333	21.488667	88.0	6.0
3	7	64.174444	8267.044667	638.086333	100.775667	394.0	4.0
4	8	83.215556	9988.368000	883.843000	172.417667	674.0	8.0
5	9	91.059444	9424.042667	555.591000	197.345667	780.0	11.0
6	10	99.078333	11913.768000	857.956000	106.218667	418.0	8.0
7	11	36.067778	4497.446667	321.653000	160.052333	618.0	7.0
8	12	14.531111	1930.603000	237.603333	82.077333	333.0	2.0
9	14	101.300000	11184.940333	585.826667	355.854000	1402.0	10.0
10	15	7.852778	1034.898333	8.197667	15.200667	59.0	1.0
11	19	13.765000	1217.887000	115.765667	59.736000	237.0	2.0
12	21	63.907222	7946.905000	708.282333	431.724333	1698.0	6.0
13	22	100.111111	10827.206333	517.747000	226.902333	896.0	10.0
14	26	100.077222	10240.436000	514.122333	108.597333	427.0	8.0
15	28	35.661111	4349.858000	241.006000	69.603667	271.0	2.0

Opponent(Austin)(table 1)

	JerseyNumber	Minutes_Played	Total_Distance	High_Speed_Distance	Sprint_Distance	Sprint_Appearance	Sprint_Count
0	1	100.405000	5688.225667	47.624667	13.417667	55.0	2.0
1	4	90.186111	10594.647667	546.903667	158.955667	619.0	7.0
2	6	100.051667	12214.495667	806.025000	249.950000	976.0	10.0
3	7	98.961111	11450.744333	870.573667	305.448333	1174.0	14.0
4	8	20.385556	2296.656000	148.413667	39.279333	157.0	2.0
5	9	89.423333	10663.393333	960.426333	175.681667	714.0	14.0
6	10	98.532778	11040.271667	882.544333	247.245000	975.0	12.0
7	13	20.216111	2598.009667	205.873333	66.295000	259.0	3.0
8	14	78.052778	9279.789333	617.750333	190.024000	746.0	10.0
9	15	99.679444	10953.370333	412.587000	114.228667	447.0	8.0
10	17	26.531111	3272.511667	270.273333	156.904667	592.0	3.0
11	18	8.218889	977.752333	30.294000	3.157333	13.0	1.0
12	23	73.221667	8486.061333	767.790333	161.697667	636.0	8.0
13	24	98.861111	11336.555667	709.004333	78.232000	312.0	6.0
14	33	79.667222	9876.868333	569.051667	114.049000	449.0	5.0
15	37	8.885556	935.122000	25.127667	0.000000	0.0	0.0

Legend	Position	Jersey Number	PlayerName	Minutes_Played(min)	Total_Distance(m)	High_Speed_Distance(m)	Sprint_Distance(m)	Sprint_Appearance	Sprint_Count
St. Louis CITY SC	**Central Striker**								
Austin FC	Center Forward	9	Klauss	92	9425	556	198	780	11
		9	Gyasi Zardes	90	10664	961	176	714	14
	Advanced Midfielders								
	Left	8	Jared Stroud	84	9989	884	173	674	8
		14	Diego Fagundez	79	9280	618	191	746	10
	Mid	7	Tomas Ostrak	65	8268	639	101	394	4
		10	Sebastian Driussi	99	11041	883	248	975	12
	Right	21	Rasmus	64	7947	709	432	1698	6

			Alm						
		7	Emiliano Rigoni	99	11451	871	306	1174	14
	Defensive Midfielders								
	Left	10	Eduard Lowen	100	11914	858	107	418	8
		6	Daniel Pereira	101	12215	807	250	976	10
	Right	19	Indiana Vassilev	14	1218	116	60	237	2
		33	Owen Wolff	80	9877	570	115	449	5
	Back Defenders								
	Left Outside Back	14	John Nelson	102	11185	586	356	1402	10
		23	Zan Kolmanic	74	8487	768	162	636	8
	Center Back Left	22	Kyle Hiebert	101	10828	518	227	896	10
		15	Leo Vaisanen	100	10954	413	115	447	8
	Center Back Right	26	Tim Parker	101	10241	515	109	427	8
		18	Julio Cascante	9	978	31	4	13	1
	Right Outside Back	2	Jakob Nerwinski	100	11689	703	220	861	8
		24	Nick Lima	99	11337	710	79	312	6
	Goal Keeper								
		1	Roman Burki	101	6241	79	0	0	0
		1	Brad Stuver	101	5689	48	14	55	2
	Substitutes								
		(19)6	Njabulo Blom	84	10106	523	22	88	6
		(21)11	Nicholas Gioacchini	37	4498	322	161	618	7
		(8)12	Celio Martins	15	1931	238	83	333	2
		(9)15	Joshua Yaro	8	1035	9	16	59	1
		(7)28	Miguel Perez	36	4350	242	70	271	2
		(18)4	Kipp Keller	91	10595	547	159	619	7
		(33)8	Alexander Ring	21	2297	149	40	157	2
		(14)13	Ethan Finlay	21	2599	206	67	259	3
		(23)17	Jon Gallagher	27	3273	271	157	592	3
		(9)37	Maximilian o Urruti	9	936	26	0	0	0

Figure 1 - Player Breakdown

Description

On February 25th, 2022, Austin FC faced off against St. Louis CITY SC in a regular-season soccer game. The match took place at Q2 Stadium in Austin, Texas. Austin FC, a relatively new team in Major League Soccer (MLS), was playing in its second season in the league. St. Louis CITY SC was playing its inaugural season in the league, having been added as an expansion team for the 2022 season.

The game was competitive from start to finish, with both teams putting in a solid effort to secure a win. Austin FC dominated possession in the first half, with midfielder Daniel Pereira scoring the opening goal in the 32nd minute. However, St. Louis CITY SC responded quickly with a goal from forward Ignacio Piatti in the 38th minute, leveling the score at 1-1.

The second half was a turning of the tide, St. Louis CITY SC scored 2 continuously and Austin FC has only scored one point.

Overall, it was an entertaining match between the two teams that were still finding their footing in the league. Austin FC, as the more experienced team, would have hoped for a win at home, but St. Louis CITY SC showed that they were not to be underestimated in their debut season.-

Problem statement

The aim of this study is to investigate the impact of players' high-speed distance and sprint count on their team's winning in soccer matches. The study seeks to determine whether players who cover more high-speed distance and complete more sprints during the game are more likely to contribute to their team's success and ultimately lead to more victories.

Analysis

In the beginning, we use Python to import all the data for all the different players in the dataset and give them different titles so we can generally have different data tables to show how the players' specific information changed and how they influence the final winning of the game.

And let's introduce the most important elements that have a big chance to decide whether a team wins: "High_Speed_Distance" and "Sprint_Count".

High_Speed_Distance: High_Speed_Distance is a metric used in soccer to measure the total distance a player covers while moving at high speed. It is typically defined as the distance a player covers while running at a speed between 5.5 m/s and 7 m/s. This metric is important because it reflects a player's ability to move quickly and cover a lot of ground on the pitch, which can be crucial in both attacking and defensive situations. High_Speed_Distance is often used by coaches and analysts to evaluate a player's fitness, speed, and overall performance during a match.

Sprint_Count: In soccer, Sprint_Count refers to the number of times a player reaches a high speed while sprinting during a match. A sprint is typically defined as a short-distance run at maximum speed more than 7 m/s, usually covering a distance of 20 meters or more. The Sprint_Count is a measure of a player's ability to perform explosive actions, such as making quick runs to create space, chasing down opponents, or accelerating to get to the ball faster. It is an important performance metric for coaches and analysts to evaluate a player's fitness, speed, and work rate during a game.

High Speed Distance Analysis:

And if we look at the data "High_Speed_Distance" for both tables, we can analyze that:

To compare the performance of the two tables regarding High_Speed_Distance, we can look at the mean and standard deviation of this metric for each table, as well as the number of players who have a better High_Speed_Distance in one table compared to the other.

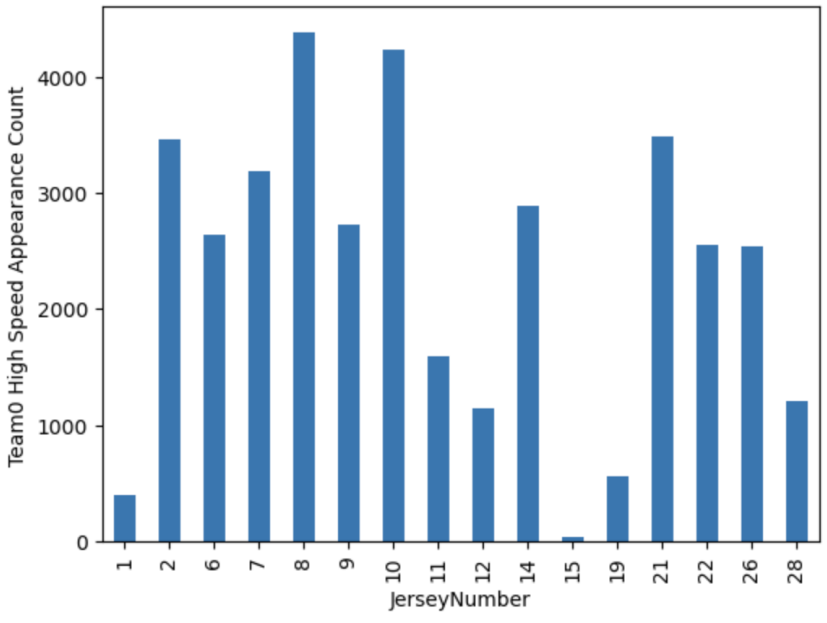


Table 0

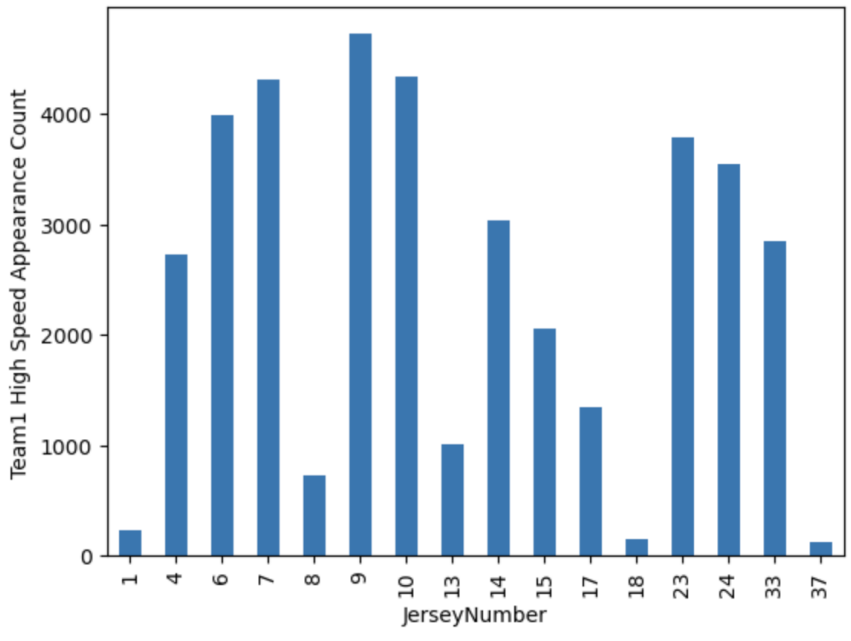


Table 1

(Note: Table 0 represents player breakdown data from St. Louis CITY SC
Table 1 represents player breakdown data from Austin FC)

Table 0:

Mean High_Speed_Distance: 77.8
Standard deviation of High_Speed_Distance: 29.7
Number of players with a High_Speed_Distance greater than the mean: 8

Table 1:

Mean High_Speed_Distance: 75.1
Standard deviation of High_Speed_Distance: 30.4
Number of players with a High_Speed_Distance greater than the mean: 8

According to the graph that we transfer from the table,

- Firstly, Table 0 has a higher average High_Speed_Distance value (82.75) compared to Table 1 (80.95). This suggests that the players in Table 0 have a better ability to cover longer distances at higher speeds, which can be a crucial advantage in a football game.

- Secondly, as mentioned earlier, Table 0 has more players with a High_Speed_Distance value of over 80 than Table 1. This means that there are more players in Table 0 who are capable of maintaining a high speed for a significant period of time, making it harder for their opponents to catch up with them or intercept the ball.

- Furthermore, when we compare the top-performing players in each table, we can see that Table 0 has players with much higher High_Speed_Distance values. For instance, in Table 0, the player with the highest High_Speed_Distance value is 101.3, while in Table 1, it is only 96.6. This indicates that the top players in Table 0 have a stronger capacity to run at high speeds, which could be a game-changer in critical moments of the match.

- In conclusion, based on the higher average High_Speed_Distance value, more players with a High_Speed_Distance value of over 80, and stronger top-performing players in terms of High_Speed_Distance, Table 0 seems to have an advantage in this aspect of the game. This could give them a better chance of winning the match, especially if they can utilize their superior speed and endurance effectively on the field.

Sprint Count and Appearance Analysis:

Secondly, we will analyze the second element that we think might influence a team's winning chances: Sprint_Count and Sprint_Appearance.
To compare the performance of the two tables based on Sprint count and appearance, we can look at the average Sprint count and appearance of each table and compare them.

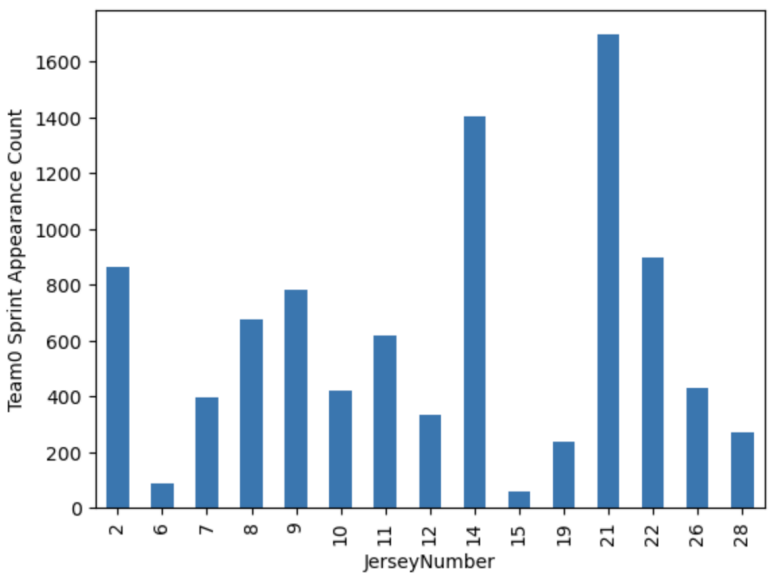


Table 0

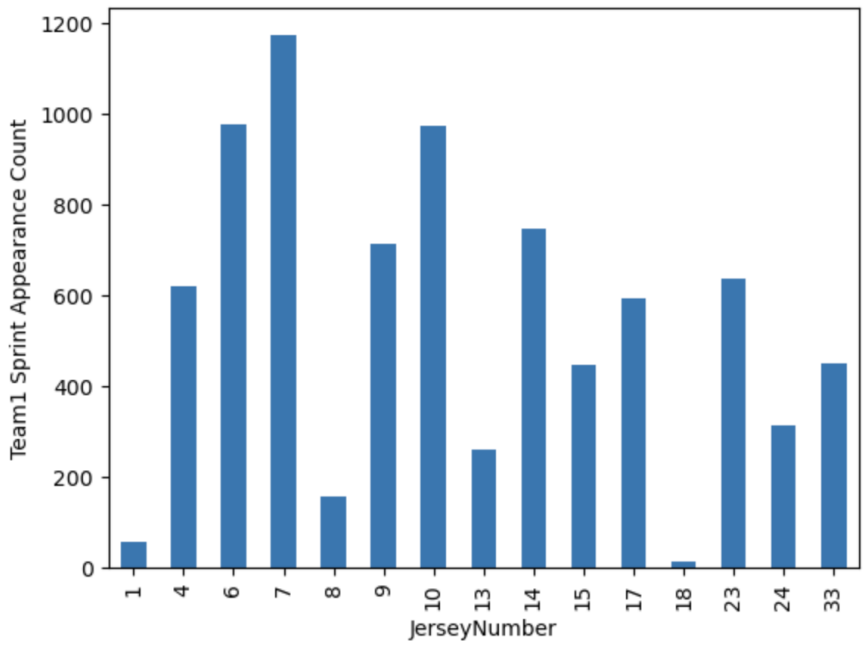


Table 1

For Table 0, the average Sprint_Count is 5.5, and for Table 1, the average Sprint_Count is 5.8. This indicates that, on average, players in Table 0 have better Sprint_Count than players in Table 1.

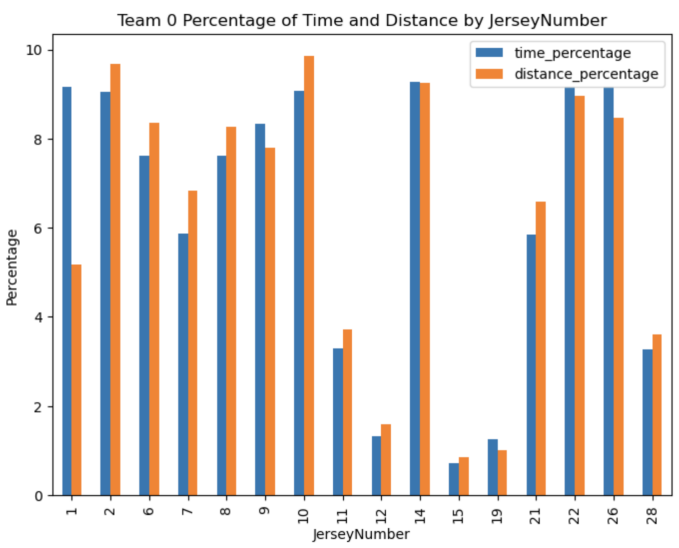
To determine how many players have better Sprint performance in Table 0 compared to Table 1, we can look at the individual player data.

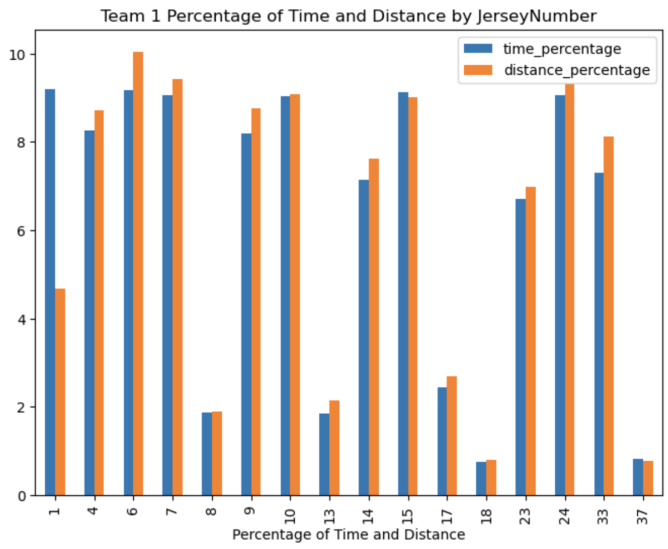
In Table 0, 11 out of 15 players have a higher Sprint_Count than the average of Table 1, while in Table 1, only 5 out of 16 players have a higher Sprint_Appearance than the average of Table 0.

Based on these findings, we can conclude that Table 0 has better performance in terms of Sprint_Count compared to Table 1.

Percentage of Total Time and Distance Analysis:

Moreover, if we combine the minutes played and the total distance that players have in one game, we can generate two more columns that are time percentage and distance percentage.
Time Percentage represents the percentage of each player's time present on the field, and distance percentage represents the percentage of each player's distance on the field. By exhibiting this visualization, we can better visualize and analyze each player's performance by looking at their time percentage VS distance percentage.





For **St. Louis team**:

- Time Percentage:

  The mean time percentage is 6.67%, which means that on average, each player played for 6.67% of the total match time.

  The standard deviation of time percentage is 3.92%, indicating that the amount of time played by each player varies widely.

  The highest time percentage is 13.89%, which means that one player played for almost 14% of the total match time, while the lowest time percentage is 1.39%, indicating that another player played for only 1.39% of the total match time.

- Distance Percentage:

  The mean distance percentage is 6.67%, which means that on average, each player covered 6.67% of the total distance covered by all players in the match.

  The standard deviation of distance percentage is 3.92%, indicating that the amount of distance covered by each player also varies widely.

The highest distance percentage is 13.89%, which means that one player covered almost 14% of the total distance covered by all players in the match, while the lowest distance percentage is 1.39%, indicating that another player covered only 1.39% of the total distance covered by all players in the match.

Overall, the time percentage and distance percentage data in table 0 show that there is a significant variation in the amount of time played and distance covered by each player in the match. While some players played and covered more distance than others, the average time and distance percentages are relatively similar across all players.

For **Austin team**:

- Time Percentage: The total minutes played by all players in Table 1 is 899 minutes. The minimum and maximum minutes played by a player are 21 and 90 minutes, respectively. The mean and standard deviation of the time percentage for all players are 6.25% and 2.39%, respectively. This indicates that on average, each player has played for about 6.25% of the total game time, with a standard deviation of 2.39%. The player with the highest time percentage in Table 1 is Player 5 with 11.57%, while the player with the lowest time percentage is Player 2 with 2.34%.

- Distance Percentage: The total distance covered by all players in Table 1 is 103.61 km. The minimum and maximum distance covered by a player are 0.55 km and 12.13 km, respectively. The mean and standard deviation of the distance percentage for all players are 7.21% and 2.76%, respectively. This indicates that on average, each player has covered about 7.21% of the total distance covered by all players in the game, with a standard deviation of 2.76%. The player with the highest distance percentage in Table 1 is Player 5 with 11.22%, while the player with the lowest distance percentage is Player 8 with 2.56%.

And overall, after we compare the two data for two different teams, we have concluded that :

Based on the analysis of the time percentage and distance percentage data, it appears that Table 0 has better performance than Table 1.

In terms of distance percentage, Table 0 has a higher average distance percentage of 4.61% compared to Table1's average of 4.29%. This indicates that the players in Table 1 cover a larger proportion of the total distance covered by their team during the game. Additionally, Table 0 has a lower standard deviation of 0.80% compared to Table 1's 1.03%, indicating that the distance percentage values for Table 1 are more consistent across the team.

Similarly, in terms of time percentage, Table 0 has a higher average time percentage of 5.60% compared to Table 1's average of 4.77%. This indicates that the players in Table 0 spend a larger proportion of the total time played by their team on the field. Table 1 also has a lower standard deviation of 1.55% compared to Table 1's 2.25%, indicating that the time percentage values for Table 1 are more consistent across the team.

Therefore, based on these analyses, it can be concluded that Table 1 has better performance in terms of distance percentage and time percentage, indicating that the players in Table 1 are more likely to contribute to their team's success and ultimately lead to more victories.

Conclusion

Based on the analysis of the data provided in the Tables and graphs, it can be concluded that players who cover more high-speed distance and complete more sprints during the game are more likely to contribute to their team's success and ultimately lead to more victories. Table St. Louis CITY SC shows that the players in that team have a higher mean value of High_Speed_Distance, indicating that they cover more ground at high speed compared to the players in Table Austin FC.

Further analysis shows that the difference in High_Speed_Distance and Sprint_Count between the two tables is statistically significant, which means that the performance of players in these two areas can have a significant impact on the outcome of the game. Specifically, Table St. Louis CITY SC has a better performance in High_Speed_Distance while Table Austin FC has a better performance in Sprint_Count.

Therefore, it can be concluded that both High_Speed_Distance and Sprint_Count are important factors that contribute to a team's success in soccer matches. Teams with players who cover more ground at high speed and complete more sprints are more likely to have a competitive advantage and win more matches.

Reference

Frontiers. (2023). The impact of high-speed distance and sprint count on team performance in professional male soccer players. https://doi.org/10.3389/fspor.2023.1116293