Nashville SC Data Project

Prateek Senapati
prateek1210s@gmail.com

All the work I have done assumes that I am at the end of the MLS 2019 season.

- **MLS team I am going to take control of:**
  I looked at the regular season results from 2014-2019, found out the teams who could not make it to the play-offs in most of these seasons, and looked at their performance before 2014. **Houston Dynamo** stood out as they had won the MLS Cup twice in the recent past. I decided to choose them in order to help them qualify for the play-offs in the 2020 season.

- **Position I need to strengthen:**
  I checked out Houston Dynamo's 2019 season review. I realised that they are lacking a **ball-moving central midfielder**. What they need is a central midfielder who would create meaningful possession, set up good passes to the attacking players, and would act as a channel between the defense and the attack.

- **Player I would try to acquire in a trade (max of 3 targets):**

#	Player	Team	Guaranteed Compensation ($)
1	Mark-Anthony Kaye	Los Angeles Football Club	177,810
2	Jackson Yueill	San Jose Earthquakes	190,000
3	Marco Delgado	Toronto FC	253,000

- **The process I used to make these selections:**
  First of all, I downloaded three kinds of data I needed for my analysis:

- 2019 season salary details of all the midfielders from all the MLS teams (except Houston Dynamo)
- 2019 season xGoals stats (per 96 minutes) of all the players from all the MLS teams
- 2019 season xPass stats (per 96 minutes; only for attacking and middle passing thirds) of all the players from all the MLS teams

I selected only the attacking and middle passing thirds as I wanted to focus on how the players performed specifically in these areas.

I selected relevant metrics from each type of data:

Data	**Metrics**
Salaries	Base Salary, Guaranteed Compensation
xGoals	Minutes Played, Key Passes (KeyP) p96, Assists (A) p96, xA p96, A-xA p96, xG+xA p96
xPass	Minutes Played, Passes p96, Pass%, xPass%, Score p96, Per100, Vertical, Touch%

After this, I decided to filter the players on certain minimum values for some of these metrics. For the xGoals data, I found out the average minutes played and average KeyP p96 for all the players. For the xPass data, I found out the average minutes played and average Passes p96 for all the players. Based on the results, I decided the following minimum values:

- Minimum minutes played = 1500
- Minimum KeyP p96 = 0.9
- Minimum Passes p96 = 30
- As I wanted my targets to perform as good as or better than their expected values, I also set the following filters:
    - A-xA p96 >= 0
    - Score p96 >= 0
    - Per100 >= 0
    - Vertical > 0 (I wanted my targets to play forward passes on average)

I filtered the xGoals and xPass data based on the above criteria, and also removed Houston Dynamo players from these data.

I performed an inner join on the three datasets to find players common to each of these datasets, and finally got five players who satisfied all the criteria I had set so far:

	Player	Team	Minutes Played	Passes per 96	Pass %	xPass %	Touch %	Key Passes per 96	A (Assists) per 96	xA per 96	xG+xA per 96	Base Salary ($)	Guaranteed Compensation ($)
0	Sebastian Lletget	LAG	2710	50.9756	87.7	83.77	11.67	1.20443	0.177122	0.135785	0.264	300000	312667
1	Jackson Yueill	SJE	2953	57.8016	87.06	85.14	12.91	1.52794	0.130037	0.109482	0.189536	145000	190000
2	Marco Delgado	TOR	3194	55.6644	84.23	82.45	12.72	1.29242	0.120225	0.11532	0.206021	253000	253000
3	Cristhian Paredes	POR	1900	37.2884	84.69	83.04	8.35	0.96	0.101053	0.0847478	0.280805	252000	283500
4	Mark-Anthony Kaye	LAFC	2787	59.7287	83.56	83.49	12.29	1.75673	0.241119	0.239676	0.344067	166248	177810

All of these players have really good (and, almost similar) Pass% and xPass% values.

To find a maximum of three players who would best fit our requirements, I decided two categories of metrics which would help me make a decision:

- *Primary*: Passes p96, Touch%, Minutes Played
- *Secondary*: KeyP p96

Based on the *primary* selection metrics, I eliminated Paredes as he has significantly less playing time, considerably less Passes p96, and a low Touch% compared to the other four candidates. Yueill and Kaye stood out on Passes p96 and Touch%, respectively.

I used the *secondary* selection metrics to decide between Lletget and Delgado. Delgado has a better KeyP p96 value than Lletget.

Finally, I decided to go with Kaye, Yueill and Delgado - they are young, they have impressive xPass stats, really good xGoals stats, and they have very low Guaranteed Compensation values. Lletget came close, but he is a little bit expensive when compared to my three target players, especially when they have similar stats.

I didn't put a lot of emphasis on Assists, xA, and xG+xA values as I believe these stats should be primarily focussed on when looking for forwards and attacking midfielders. I am looking for targets who would deliver good pre-assist passes and passes that would find attacking players in the final third.

NOTE: ***Please have a look at the detailed interactive visualisations supporting my findings in the corresponding Jupyter Notebook. They will be really helpful and give you more insights into the selection process I used.***

- **Information that was not available (that could have changed my perspective):**
  Having pass maps for the targets would have helped me see how and where they distribute their passes. Heat maps would have helped as well - showing me details about the targets' off the ball movements and positioning. I would also find packing data and pre-assist data of my targets to be very vital in studying their contribution. I believe information for the aforementioned metrics can prove to be really essential for analysing a midfielder's performance.