Vigneshwaran Kannan

Bharath Mohan

Ashish Nangia

Mohamed Salihdeen Rafiqdeen

Surendar Rajavijayan Sundar

October 17, 2016

Data Visualization – Group 10

Contents

[**DATASET:** 2](#_Toc464510927)

[**INSIGHT 1** 3](#_Toc464510928)

[**INSIGHT 2** 4](#_Toc464510929)

[**INSIGHT 3.a** 5](#_Toc464510930)

[**INSIGHT 3.b** 6](#_Toc464510931)

[**INSIGHT 3.c** 7](#_Toc464510932)

[**INSIGHT 4** 8](#_Toc464510933)

[**INSIGHT 5** 9](#_Toc464510934)

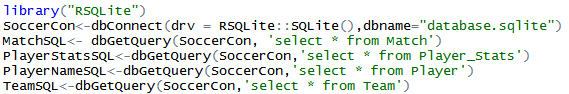
[**ACTUAL DECISIONS MADE BY MANCHESTER UNITED** 10](#_Toc464510935)

[**APPENDIX A** 10](#_Toc464510936)

# **DATASET:**

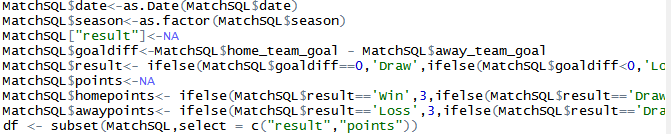
***Link****:* [**https://www.kaggle.com/hugomathien/soccer**](https://www.kaggle.com/hugomathien/soccer)No table of figures entries found.

* We acquired European soccer database from Kaggle, which was in “\*.sqlite” format.
* We explored the dataset in R using “RSQLite” library.
* Our dataset had 7 tables, out of which we identified Match as our primary data source and Playerstats as our secondary data source.
* We further merged the data’s from player to playerstats and team to match dataset.



* ***Match:*** It contained match stats for all the primary league matches played throughout Europe between 2008/09 season and 2015/16 season.
* ***PlayerStats:*** It contained player stats for the players who were part of European teams during the 2008/09 – 2015/16 period.
* Next, we constructed Points Table for each team during each season across all leagues from the match dataset using R functions and excel. This later became our primary dataset which is linked to player stats using team id.

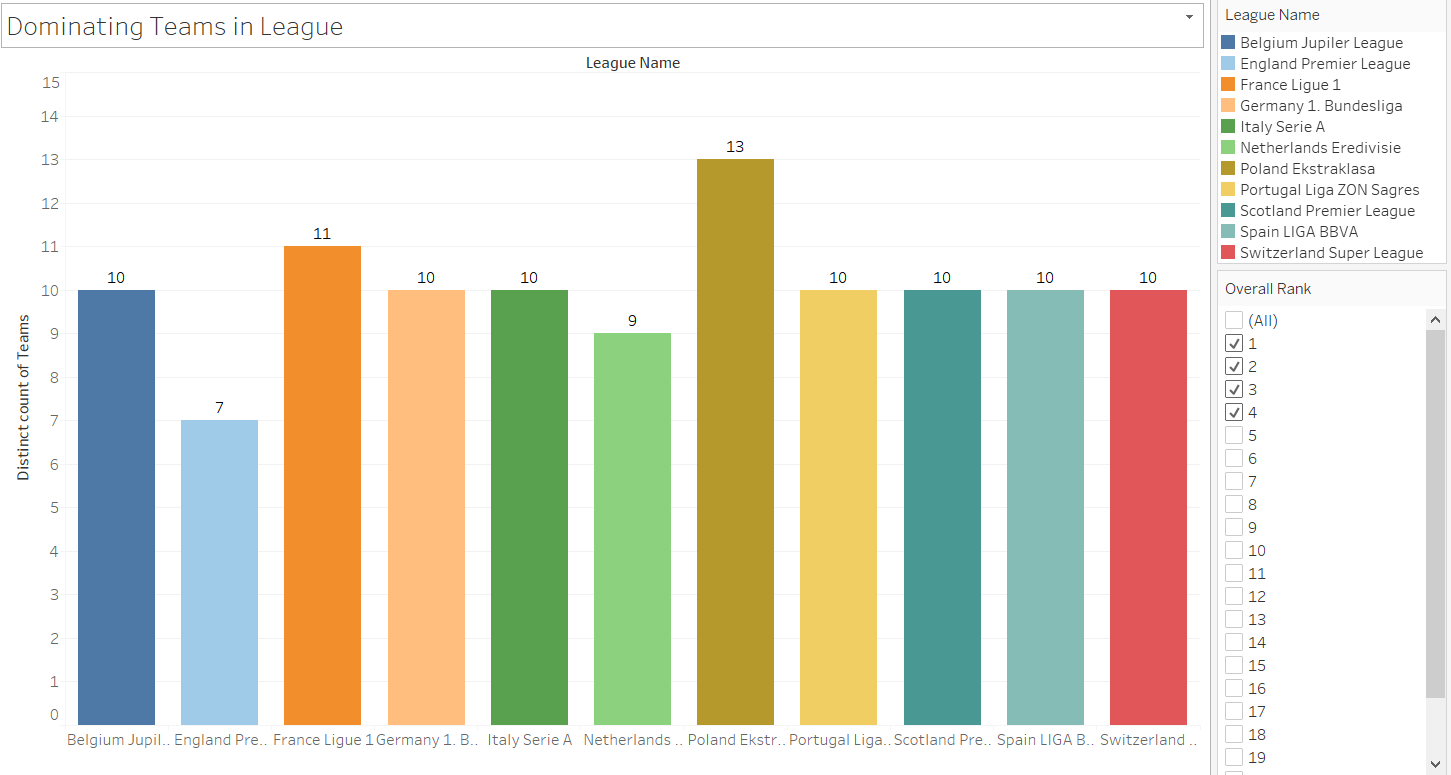


******

* Player stats had multiple stats for a player calculated at various intervals during the course of the season. To reduce ambiguity we kept only the end of season stats for a player and removed all intermediary stats calculated during the season.

# **INSIGHT 1**

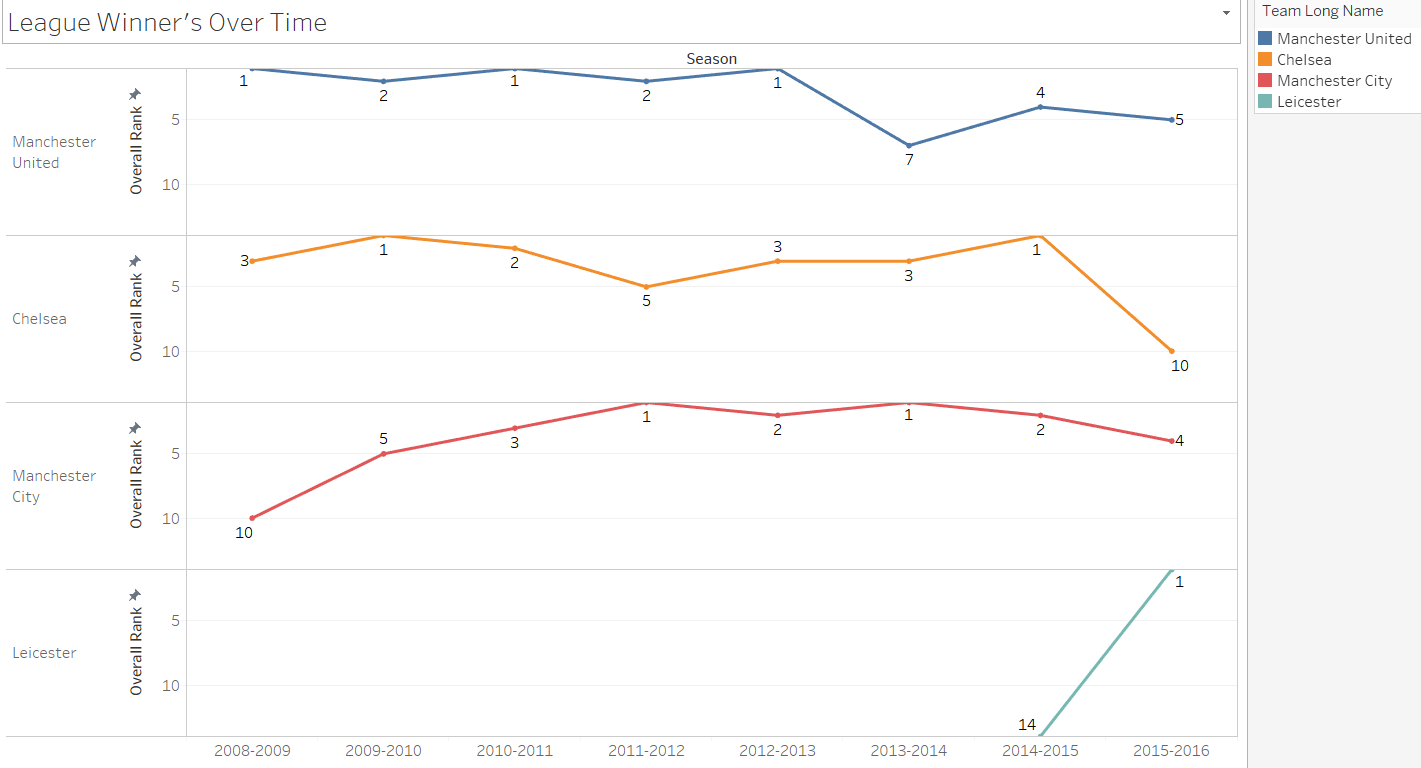
* To understand our dataset better, we decided to visualize how many teams in the top division leagues of 11 Europe countries have consistently played well.
* A good parameter to decide the same was to see how many teams have finished top 4 in these different leagues over the 8 seasons from 2008-2016.
* Below is the screenshot that displays the unique number of different teams that finished top 4 in these leagues.



* We found that the English Premier league had the least amount of unique teams that had finished in the top 4.
* This meant that the EPL had the most consistent and dominating teams which managed to come in the top 4 season after season.
* This does not necessarily mean that the teams have finished in the same positions every season. There has been a shuffle in how the teams have finished in the top 4.
* It does however display that the same teams have been dominating throughout seasons.

# **INSIGHT 2**

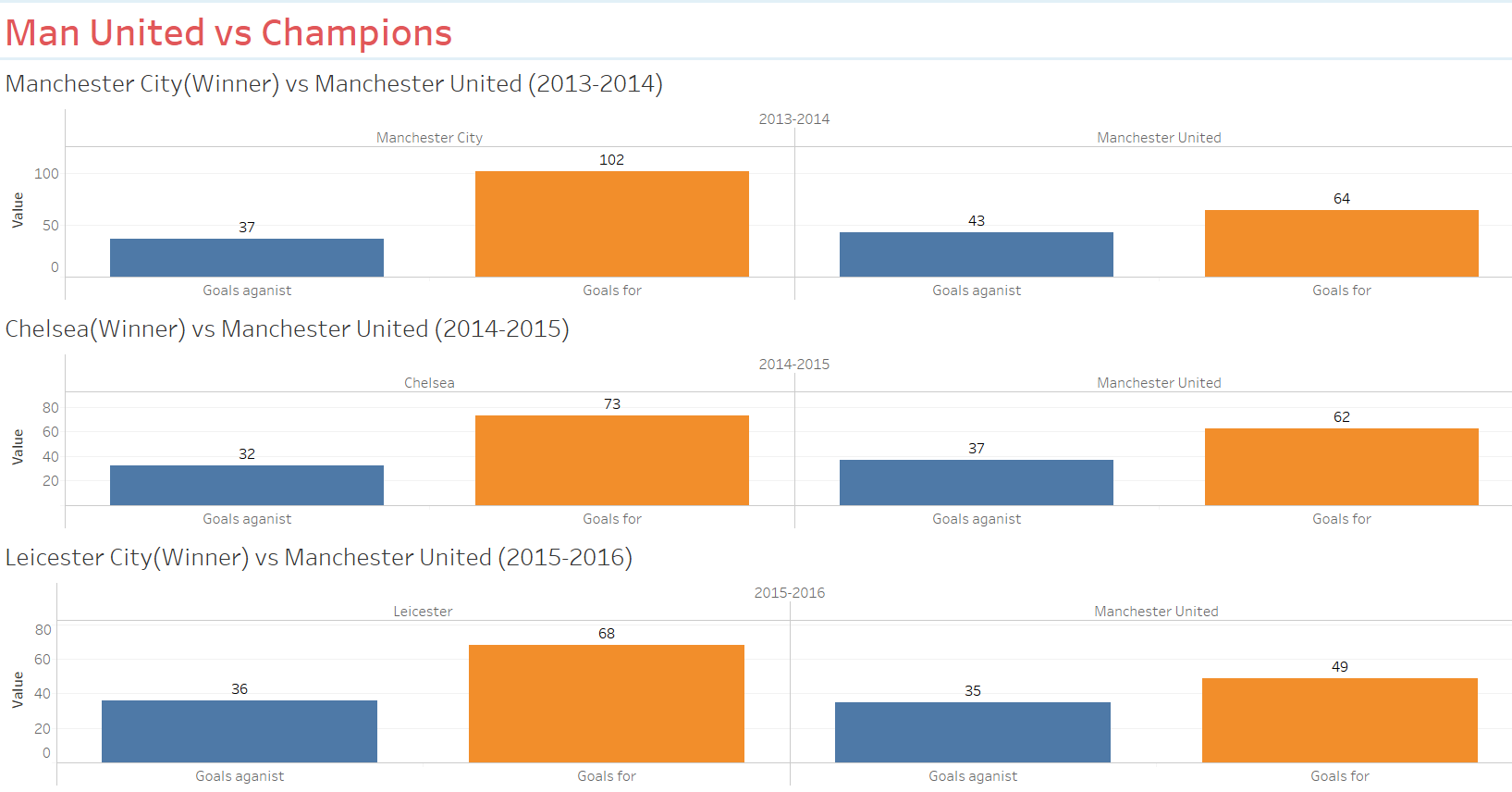
* Based on insight 1, we decided to focus on EPL since it had the most dominating teams within the league.
* Next we decided to visualize the teams that have won the league at least once and their positions during the other seasons.
* Below is the screenshot of a line chart that displays the 4 teams that have won the league in the past 8 seasons and their performance in the other seasons.



* Leicester City is an outlier in this situation, since it came into the top division only in 2014 and won the league the next season.
* It is quite evident that Manchester United had been very consistent from 2008-2013, but then had a considerable drop and have not been able to finish in the top 3 since and got into the top 4 just once.
* To figure out why there was this drop in performance for Manchester United, we have decided to further dive into the reasons for this drop and give suggestions of how they can improve.

# **INSIGHT 3.a**

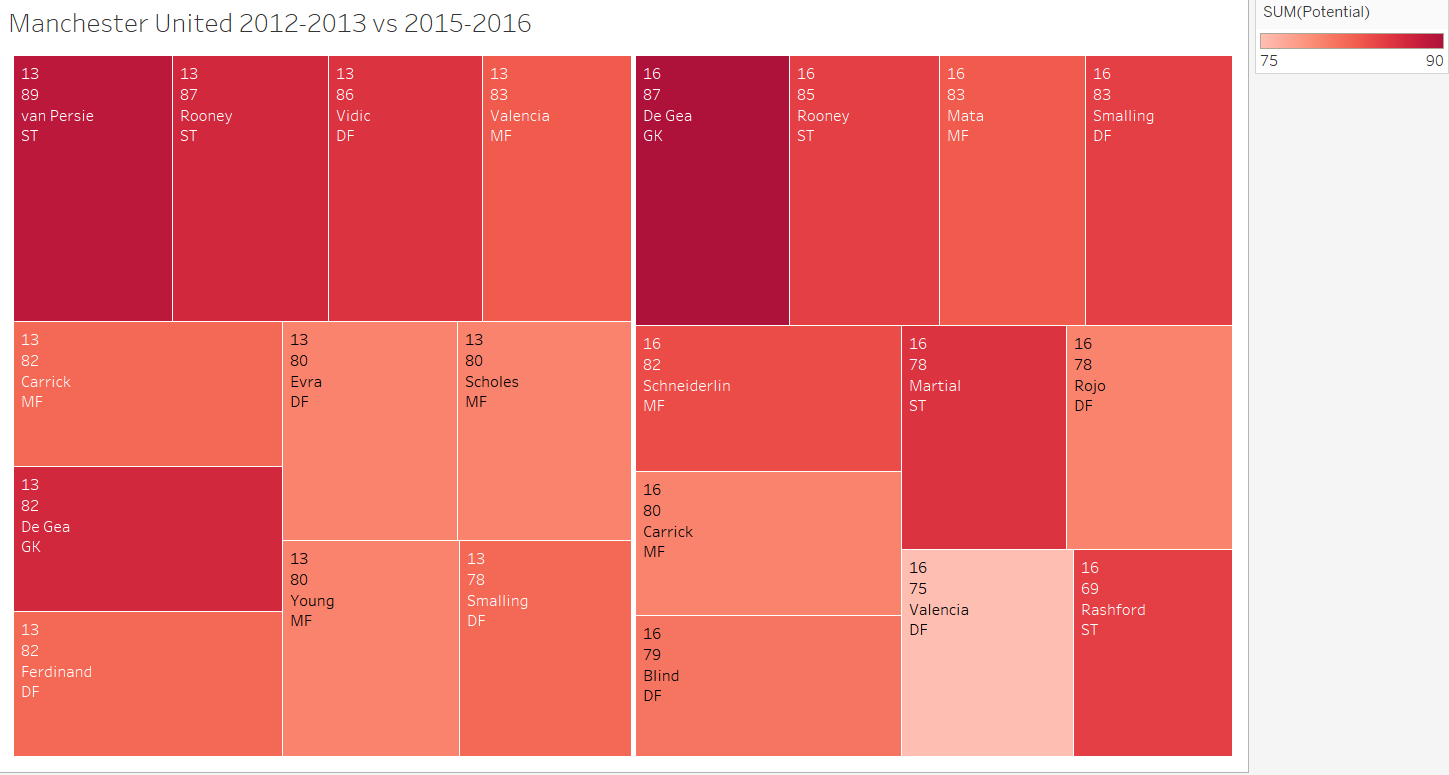
* We decided to visualize how Manchester United had performed as compared to the champions of the seasons in which Manchester United had not performed well.
* We compared the Goals scored and conceded by Manchester United with the goals scored and conceded by the champions of that season.



* It is clear from the visualization that Manchester United conceded almost as many goals as the champions, which suggests Manchester United were solid defensively.
* However, they were not scoring enough goals since the champions scored many more goals than them in all of these seasons.
* Therefore, Manchester United has been in dire need of attacking options according to the above visualization.

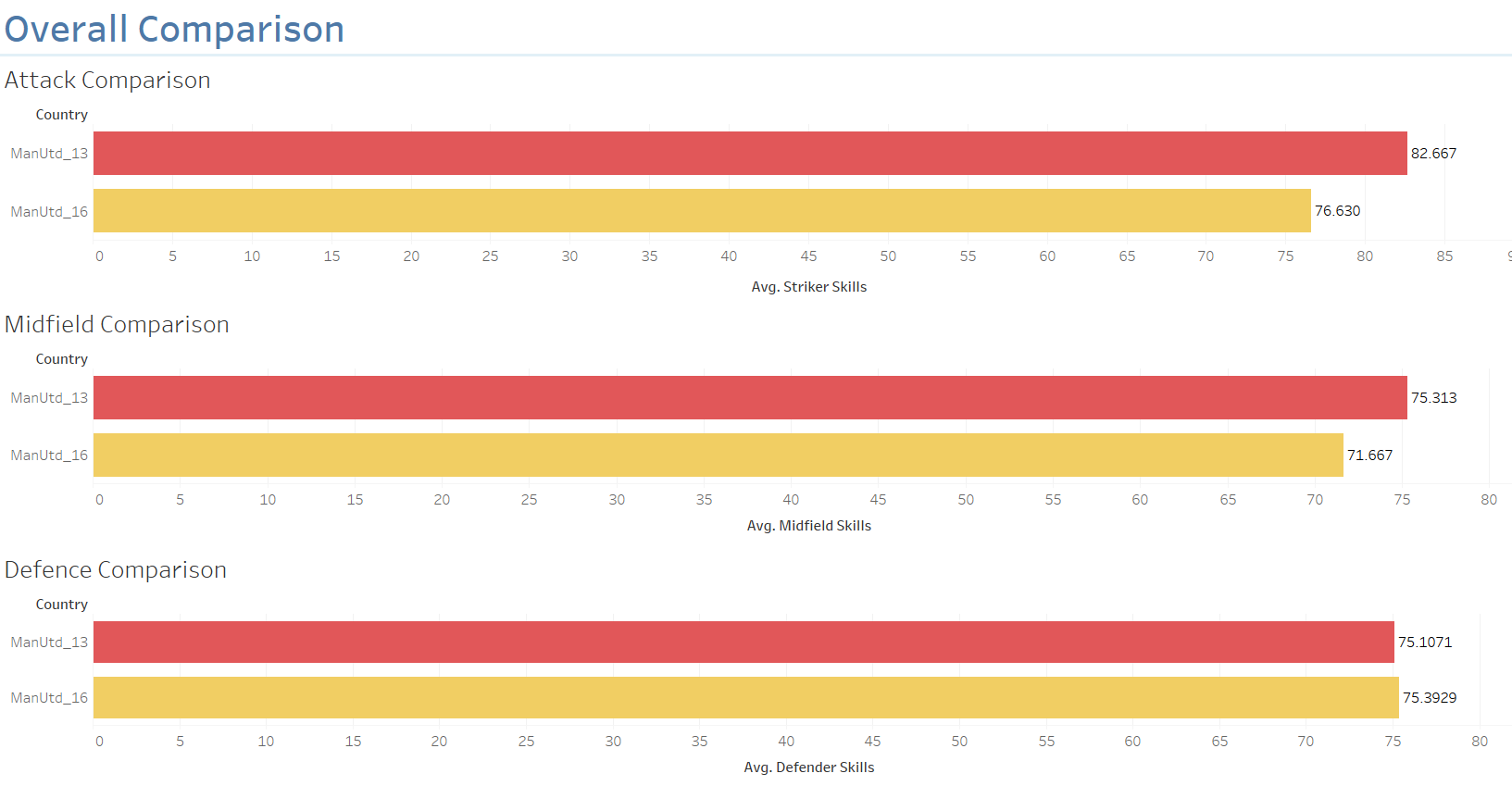
# **INSIGHT 3.b**

* Once we identified our major areas of improvement we decided to compare the stats of Manchester United players between seasons to ensure the validity of above insight.
* We created tree maps of Manchester United players based on respective overall rating and their future potential.
* From the below tree map we can recognize the lack of reliable attacking players among the team’s top performers. Also from the map we can be sure that the team’s defensive options are in better shape compared to rest of the team.



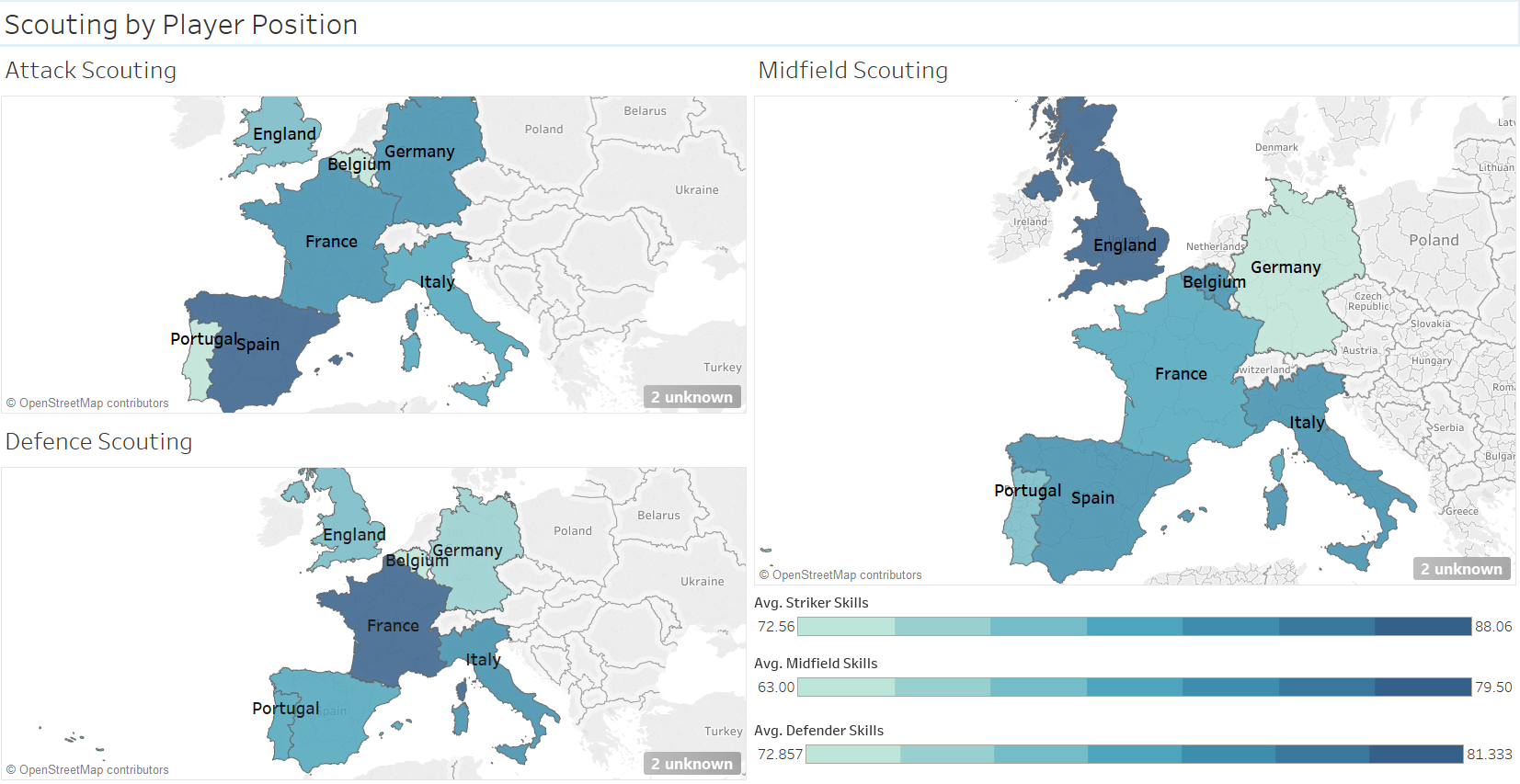
# **INSIGHT 3.c**

* Following that we created a mean of player’s attacking, midfield and defensive skills using calculated fields.
* Next we plotted these fields from team’s best season against current season.
* Based on the graph we can see improvement team’s overall defense performance (also supported goals conceded from insight 3.a). However the both attacking and midfields are in need of upgrade to restore the Manchester United back to top of the table.
* Calculated fields used:
  + Striker Skills = (Acceleration+ Ball Control+ Crossing+ Dribbling+ Finishing+ Heading Accuracy +Shot Power+ Penalties+ Sprint Speed)/9
  + Defensive Skills = (Aggression+ Interceptions+ Jumping+ Marking+ Positioning+ Sliding Tackle+ Standing Tackle)/7
  + Midfield Skills = (Curve+ Free Kick Accuracy+ Long Passing+ Long Shots)/4



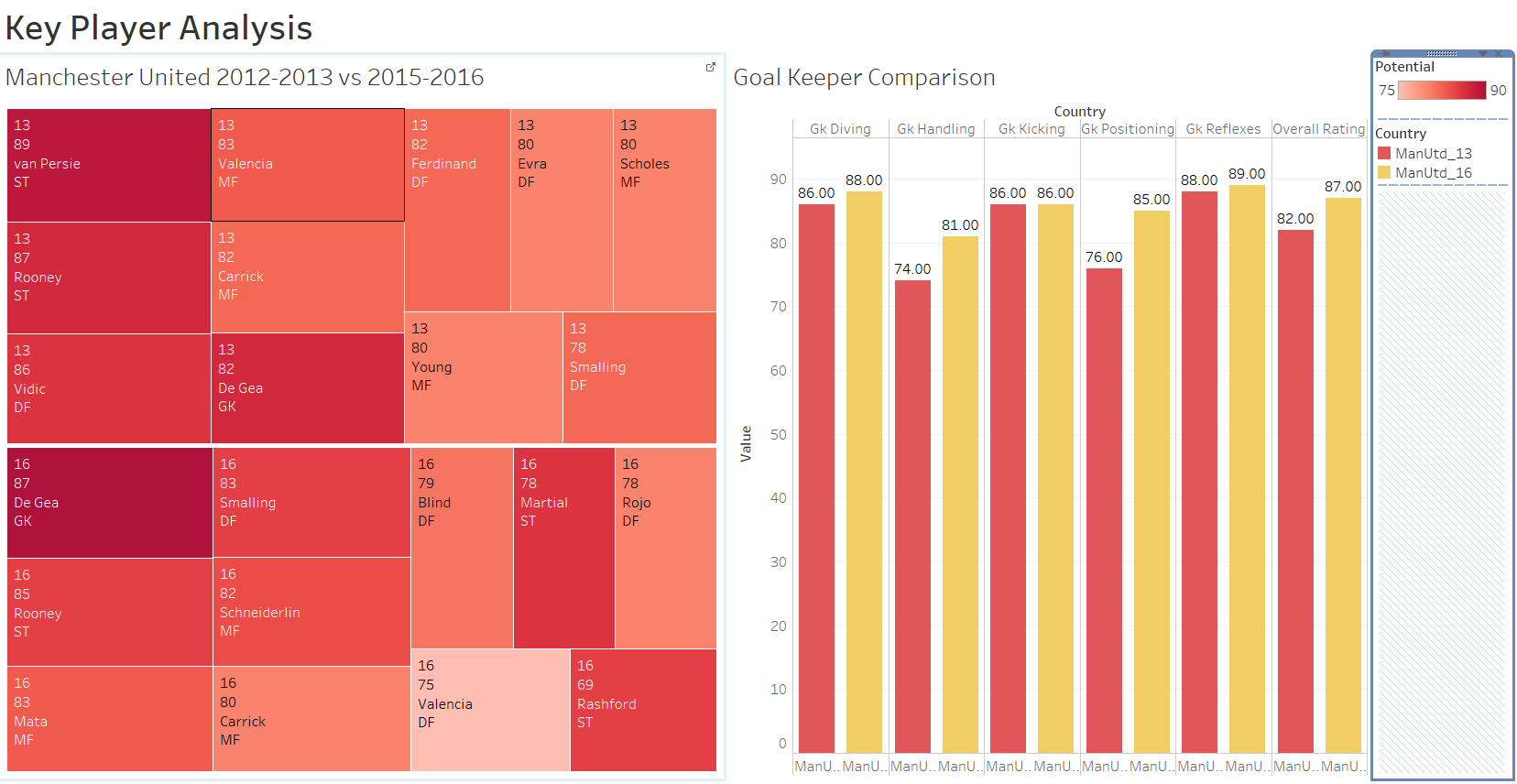
# **INSIGHT 4**

* To build a better set of attacking and midfield options, the management should scout potential players to across leagues who can plug the talent gap which exist between current Manchester United squad and the league champions.
* Instead of sending scouts throughout the continents to look for potential players, we decided to narrow our focus by identifying leagues whose top 4 have better options than United.
* This is achieved calculating the mean of player’s skills (Attacking, midfield and defense) and plotting them across the map.
* From the map we can conclude that Spain, England and France has best attacking, midfield and defense respectively. With this data, the management can plan their scouting expeditions accordingly.



# **INSIGHT 5**

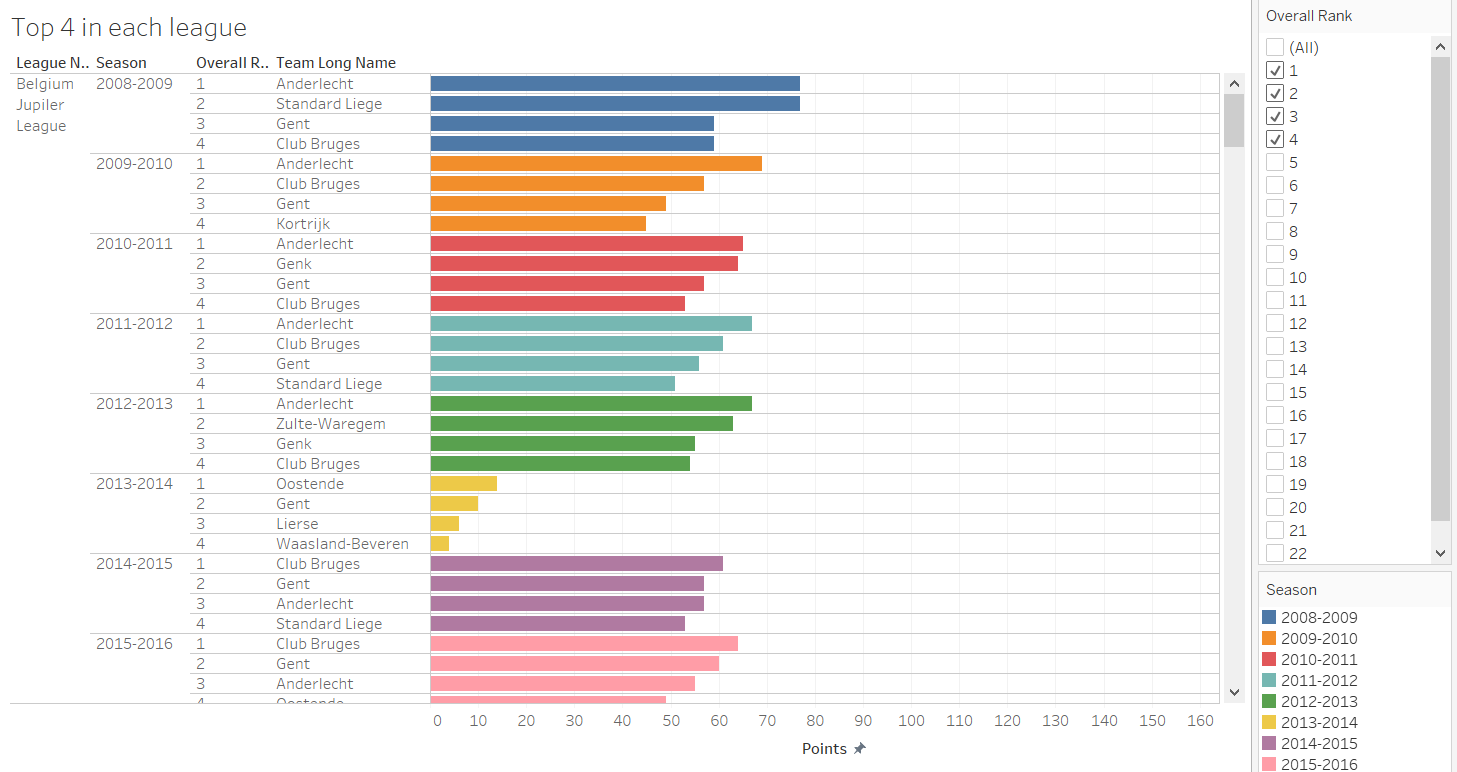
* Before strengthening the squad, Manchester United must retain its current set of best performers with better compensations. So they don’t start to look for better prospects elsewhere.
* From the tree map we can identify “David De Gea” has been a stellar performer for the last season. A quick comparison between his player stats reveal since joining the team he has improved considerably to warrant a better contract.
* Also we can see young players such as Marcus and Martial who are aged 18 and 20 respectively have great overall potential (Both of them 86) which shows they are the players for the future and can offer them lucrative contracts which can keep the rivals at bay.



# **ACTUAL DECISIONS MADE BY MANCHESTER UNITED**

* The team has “Zlatan”, striker, rating 89, from French league (Ranked 2nd among potential scouting locations for attacking options) and “Paul Pogba”, midfielder, rating 86, from Italy league (Ranked 2nd among potential scouting locations for midfield options) which is in accordance with our findings.

# **APPENDIX A**



* We identified top 4 teams in all the 11 leagues through the above graph which helped us kick start our analysis.