Exploration of how has money spending on Transfer Market affected English Premier League results between 2000 to 2017

Dhir Nilesh Chandan

dcha0038@student.monash.edu

Date: 27-04-2019

Contents

1. Introduction

1.1 Problem Description: The purpose of this analysis is to explore how the spending on player transfer market has affected the results of English Premier League teams between 2000 and 2017 period. To fulfill this analysis 2 datasets from Kaggle are utilized, one with data related to transfers and another with the final league table for each season during 2000-2017 period. During the course of this analysis, there has been an attempt to answer a number of questions, two of the primary questions are as described below.

1.2 Questions:

a. Which teams have spent/made most Money on transfer market and on which players?
b. How have the Premier League results changed for top spending clubs?

1.3 Motivation: A lot of money is spent each year for getting the best players on the team and a lot of hype is created when such transfers take place. Big clubs often spend big bucks to get big and renowned player into the team and smaller clubs often don't spend as much but may make wise choices, these transfer of players may or may not help with the result of the club during the season depending on how the new player(s) fit into the team. So as a football fan, motivation is to explore if there is an underlying story behind the market spending and the results of the teams in the league.

2. Data Wrangling and Checking

Two main Datasets have been used in this analysis:

1. English Premier League end of season league tables for each season from 2000 to 2017.
    a. Link: https://www.kaggle.com/limmen/premierleague-league-tables-188889-201617
    b. Tabular Data: 1343 rows x 13 cols
2. Top 250 football transfers per season from 2000 to 2018 in Europe.
    a. Link: https://www.kaggle.com/vardan95ghazaryan/top-250-football-transfers-from-2000-to-2018
    b. Tabular data: 4700 rows x 10 cols

First the data was read into a *R project* for Data Wrangling and Cleaning.

Transfers Dataset:

A glimpse of the dataset:

```
> head(trans)
              Name       Position Age     Team_from     League_from      Team_to      League_to    Season Market_value Transfer_fee
1        Luís Figo   Right Winger  27 FC Barcelona          LaLiga  Real Madrid         LaLiga 2000-2001           NA     60000000
2    Hernán Crespo Centre-Forward  25         Parma        Serie A        Lazio        Serie A 2000-2001           NA     56810000
3    Marc Overmars    Left Winger  27       Arsenal Premier League FC Barcelona         LaLiga 2000-2001           NA     40000000
4 Gabriel Batistuta Centre-Forward  31    Fiorentina        Serie A      AS Roma        Serie A 2000-2001           NA     36150000
5   Nicolas Anelka Centre-Forward  21  Real Madrid          LaLiga     Paris SG        Ligue 1 2000-2001           NA     34500000
6    Rio Ferdinand    Centre-Back  22     West Ham Premier League        Leeds Premier League 2000-2001           NA     26000000
```

In this dataset there's data of player transfers from all European teams from 2000 to 2018. But here we are interested in only transfers in and out of English premier League teams and only until 2017 season. Also Market value column is not required for the analysis. So filtering the data to suit the needs of the analysis using following code.

```
# Filtering transfers between 2000/2001 season and 2016/2017 season
trans$Season %>% as.character() -> trans$Season
trans %>% filter(Season <= "2016-2017") -> trans
trans$Season %>% as.factor() -> trans$Season

#  Keeping only the columns required for analysis
trans %>% select(Name,Position,Age,Team_from,League_from,Team_to,League_to,Season,Transfer_fee) -> trans

# Filtering data for Premier League only.
trans %>%
  filter(League_from=='Premier League' | League_to=='Premier League' | League_from==" England" | League_to==" England") -> trans
```

After this step the transfers dataset is left with *1348 rows x 9 cols*.

Premier League Results dataset:

A glimpse of the dataset:

```
> head(plres)
       year                    title              Team  A  D  F GD L  P Pos Pts  W
1 2016/2017 Premier League 2016-2017           Chelsea 33  3 85 52 5 38   1  93 30
2 2016/2017 Premier League 2016-2017 Tottenham Hotspurs 26  8 86 60 4 38   2  86 26
3 2016/2017 Premier League 2016-2017    Manchester City 39  9 80 41 6 38   3  78 23
4 2016/2017 Premier League 2016-2017         Liverpool 42 10 78 36 6 38   4  76 22
5 2016/2017 Premier League 2016-2017           Arsenal 44  6 77 33 9 38   5  75 23
6 2016/2017 Premier League 2016-2017 Manchester United 29 15 54 25 5 38   6  69 18
```

This dataset contains end of season league tables for seasons from 1888 to 2017. Although we need data only for the period of 2000 to 2017. Also the 2nd column is redundant as it has the same value for all rows. So filtering the data to suit the needs of the analysis using following code.

```
# Filtering data from 2000/2001 season till 2016/2017 season for PL results
plres$year %>% as.character() -> plres$year
plres %>% filter(year >= "2000/2001", year <= "2016/2017") -> plres
plres$year %>% as.factor() -> plres$year

#  Keeping only the columns required for analysis
plres %>% select(year,Team,Pos,Pts, W, D, L, F, A, GD) -> plres
```

Upon further checking it was found that there was no data for *2015/2016* season.

```
> levels(plres$year)
 [1] "2000/2001" "2001/2002" "2002/3"    "2003/4"    "2004/5"    "2005/6"    "2006/2007" "2007/2008" "2008/2009"
[10] "2009/2010" "2010/2011" "2011/2012" "2012/2013" "2013/2014" "2014/2015" "2016/2017"
```

So that data was added separately from a freely available source (https://www.kaggle.com/ronman11/english-premier-league-tables#epl20152016.csv). It was integrated into the current dataset using following code:

```
# Checking if there is data for all seasons
levels(plres$year)
## Looks like we donot have data from 2015/2016 season.
# Reading data fro 2015/2016 season
plres_1516 <- read.csv("epl20152016.csv")
# Refomatting before joining
# Changing column names
plres_1516 %>% rename(Pos = X., Pts = P) -> plres_1516
# Creating new columns for year and Goal Difference
plres_1516 %>% mutate(year = "2015/2016") -> plres_1516
plres_1516 %>% mutate(GD = F-A) -> plres_1516
# Selecting columns to join with main results data frame.
plres_1516 %>% select(year, Team, Pos, Pts, W, D, L, F, A, GD) -> plres_1516
# Joining data from 2015/2016 season to original results dataset
plres <- rbind(plres, plres_1516)
```

The factor levels for year (season) were also sorted in order after adding the new season.

```
# Ordering original dataset as per year of season
plres$year %>% as.character() -> plres$year
plres %>% arrange(desc(year)) -> plres
plres %>% mutate(year = factor(year, unique(year))) -> plres
```

Some of the seasons had inconsistent format. For example – 2003/4 , while it actually should look like 2003/2004. So this was also corrected:

```
# Renaming season year values to make it consistent
levels(plres$year)
levels(plres$year)[levels(plres$year)=="2005/6"] <- "2005/2006"
levels(plres$year)[levels(plres$year)=="2004/5"] <- "2004/2005"
levels(plres$year)[levels(plres$year)=="2003/4"] <- "2003/2004"
levels(plres$year)[levels(plres$year)=="2002/3"] <- "2002/2003"
levels(plres$year)
```

In season year, '/' was replaced with '-' to make it consistent with the transfers dataset.

```
# Changing format for how year is displayed to make it consistent with other dataset
plres$year <- as.factor(gsub("/","-", as.character(plres$year)))
```

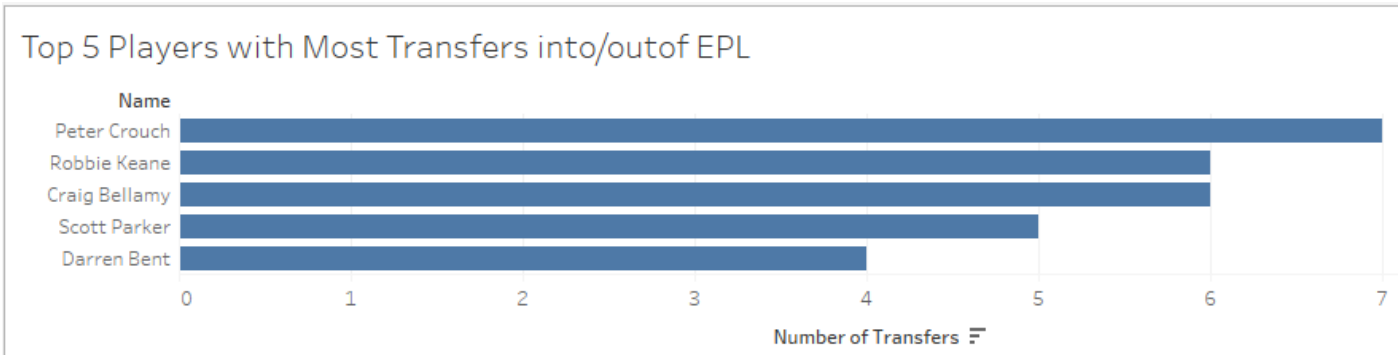
After these changes the premier league results dataset is left with *340 rows x 10 cols*.

3. Data Exploration

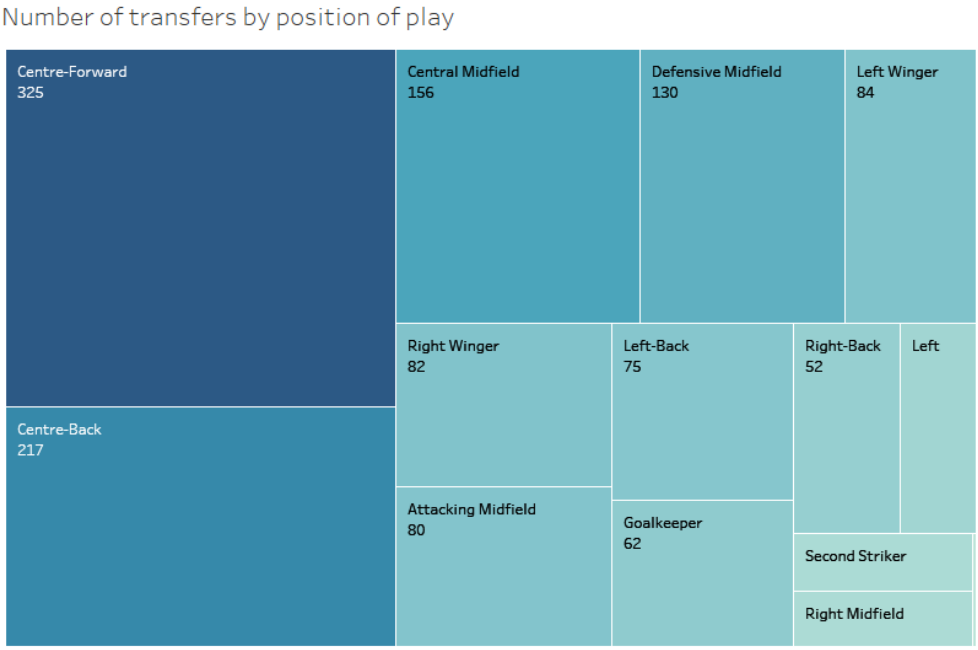
R and tableau have been used for data exploration. R is used for aggregations and descriptive statistics and tableau has been used to create visualizations.

Starting off with Univariate Analysis for **Transfers dataset**.

1. Which **players** have been transferred the most in/out/amongst of Premier League teams?



2. Which **positions** are most in demand in the transfer market?



Looks like players playing in Center-forward position have had the most transfers, followed by Center-back.

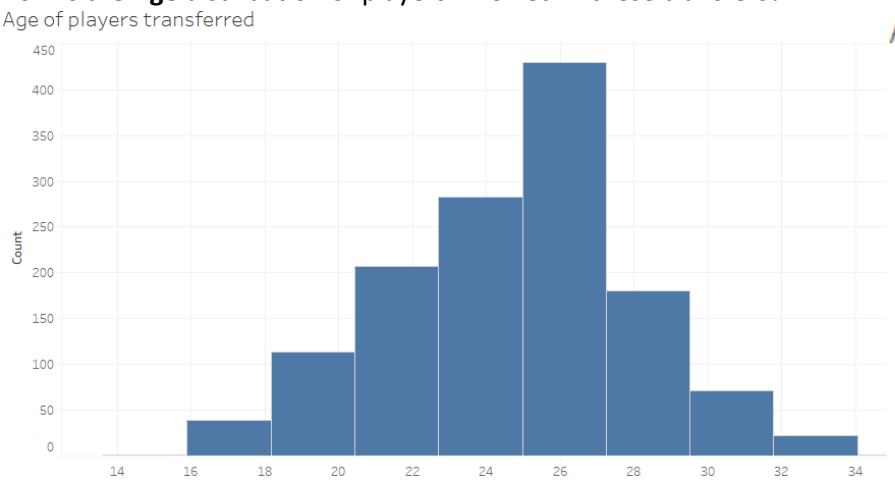
3. Which teams have **bought** the most number of players between 2000 and 2017?

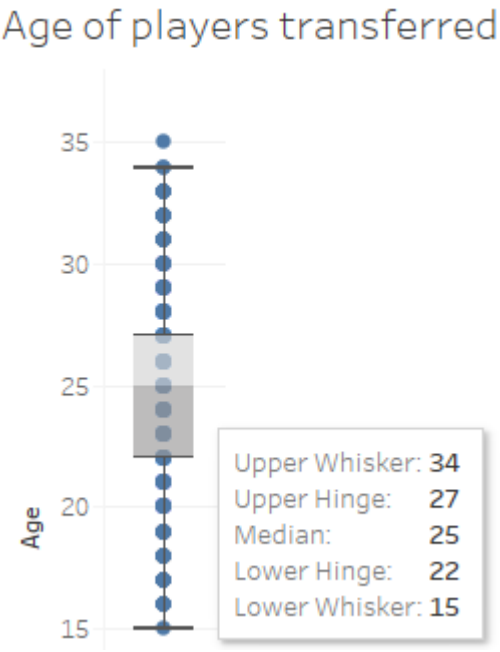
Looks like Spurs, Liverpool, Chelsea, Man Utd and Man City have had the most transfers. Which seems normal as these are some big clubs with big budget.

4. Which teams have **sold** the most number of players between 2000 and 2017?

Looks like a similar trend for club names for most players sold.

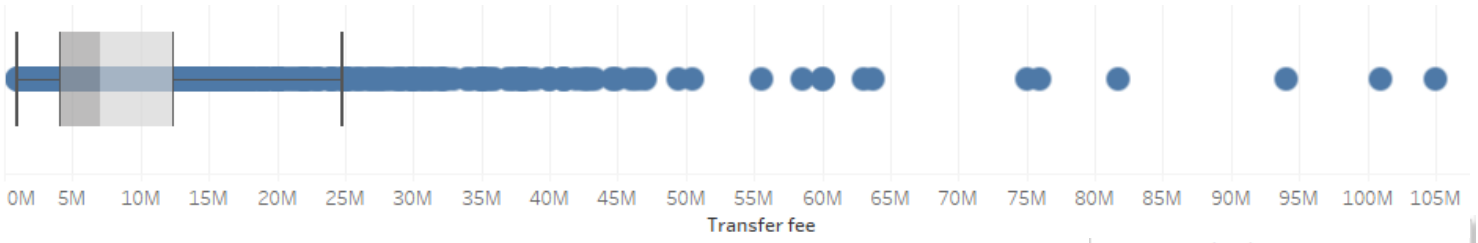
5. How is the **Age** distribution of players involved in these transfers?

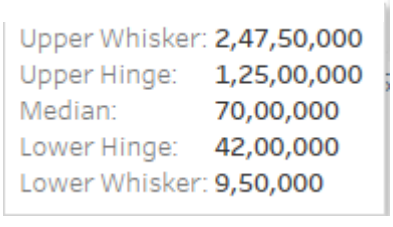




Looks like most players fall in the age bracket of 22-28, which is the prime of any football player, with the median age being 25 and the range being 15 to 34. But there are some player

6. How is the distribution of **Transfer fees**?

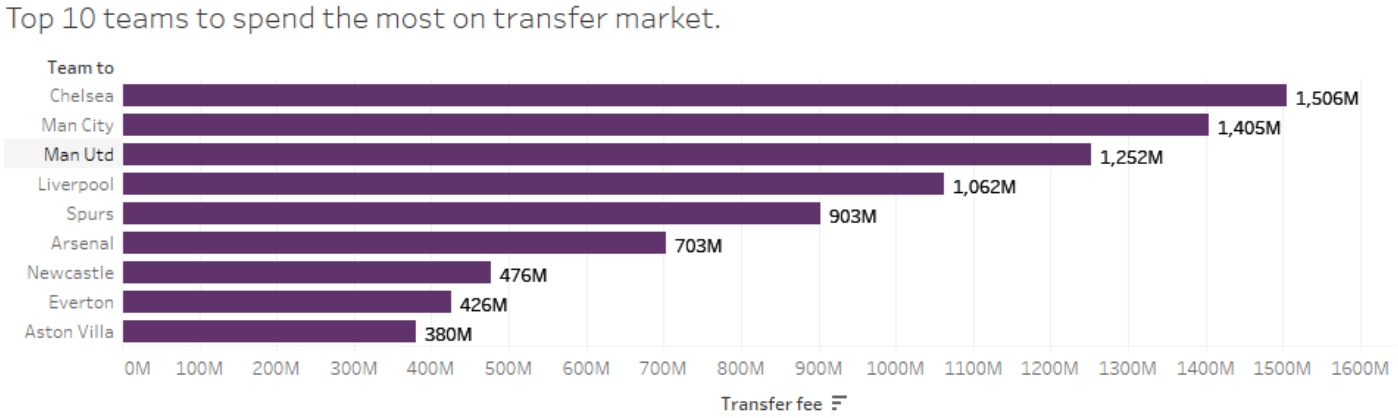




Transfer fees vary from 950K Euros to 100M+ Euros, with a median coming to around 7M Euros. The transfer fees are mainly influenced by the quality of players and the clubs involved in the transfer, hence so much variation, due to this it is noticeable that there are so many outliers. Outliers also exists because once in a while a star players like Cristiano Ronaldo, Neymar, Kylian Mbappe get transferred and due to their exceptional talent and demand, their transfer fees are more than 100M Euros.

Analysing the dataset for the primary questions of this analysis:

1. Which teams have spent most Money on transfer market and on which players?

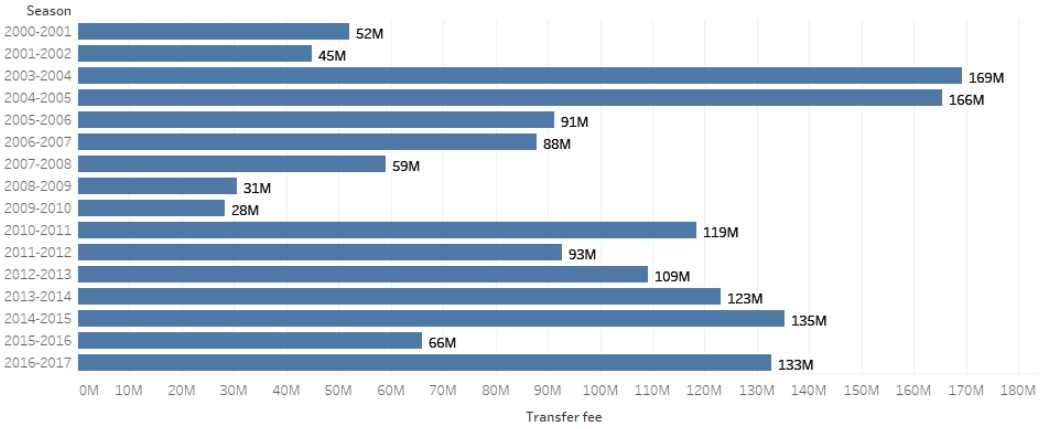


Chelsea is the club which seems to have spent the most at around 1.5 Billion Euros, followed by Man City at 1.4 B Euros, then comes Man Utd at 1.25 B Euros and then Liverpool at 1.06 B Euros. Other teams have spent less than a Billion Euros. These statistics are a total sum of all transfers between 2000 and 2017 seasons.

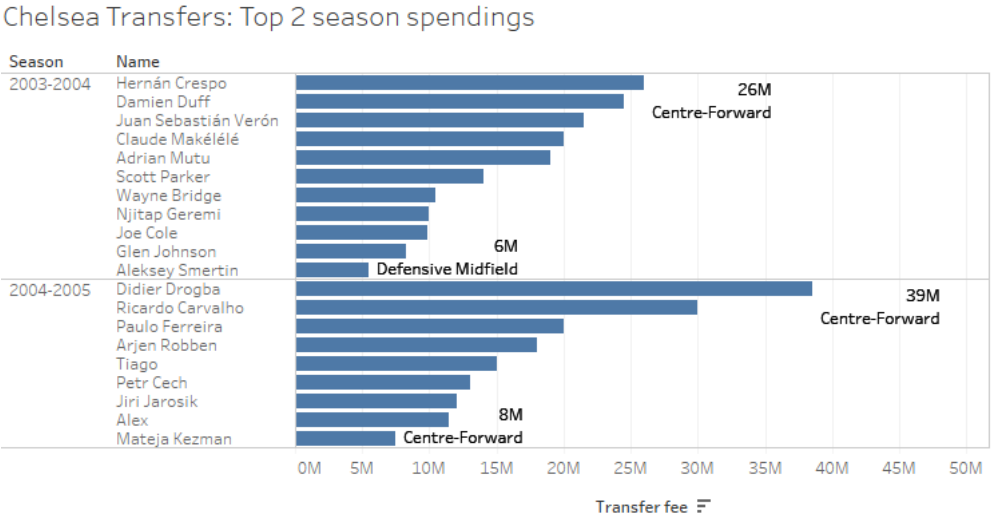
Further analyzing Teams with spending more than 1 Billion on which players they have spent the most and in which season. The further analysis on this report will concentrate on the top 3 spending teams: Chelsea, Man City, and Man Utd.

Chelsea:

Transfer spending for each season:



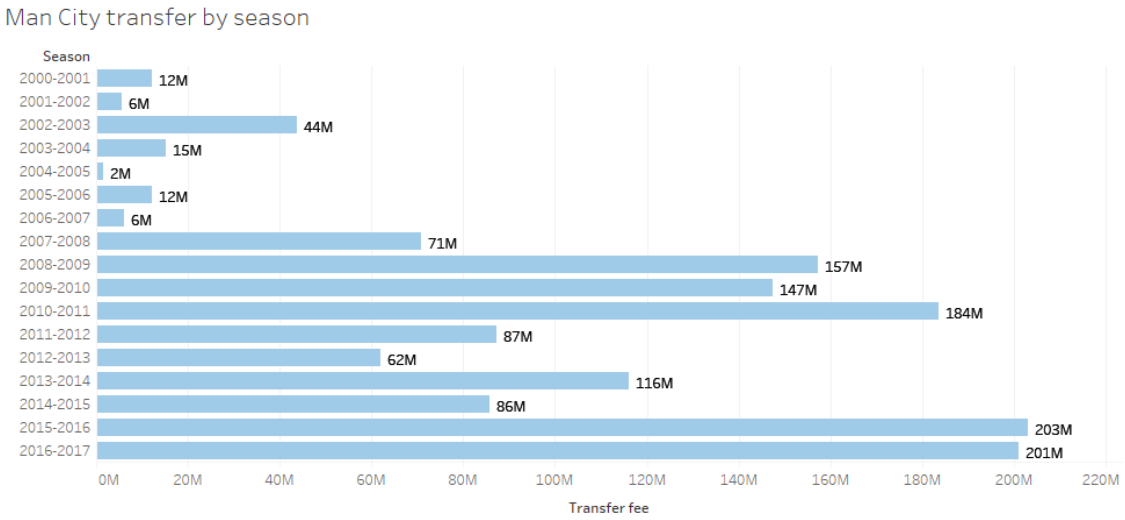
Looks like Chelsea spent the most during 2003-2004, 2004-2005 season. Lets see what they spent on during those seasons.



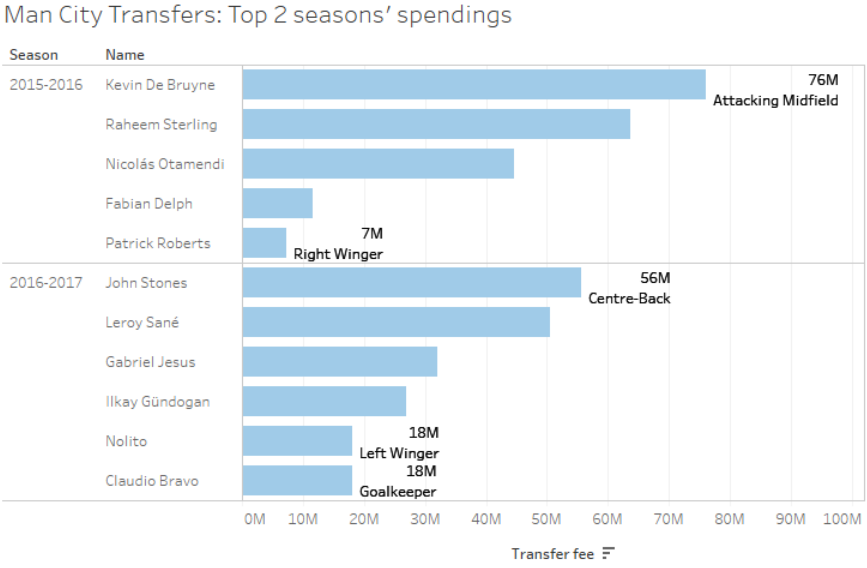
Crepo is the highlight of 03-04 season , and Drogba is for 04-05 season. Both of them were center forwards. *This may be a trend in the future analysis. Forward playing players are sold for a higher price.*

Man City:

Transfer spending for each season:



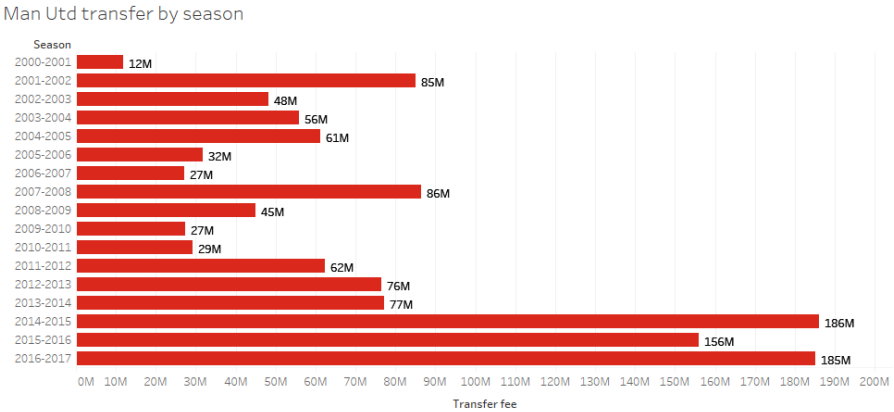
Looks like Man City spent the most during 2015-2016, 2016-2017 season, crossing 200M Euros each time. Lets see what they spent on during those seasons.



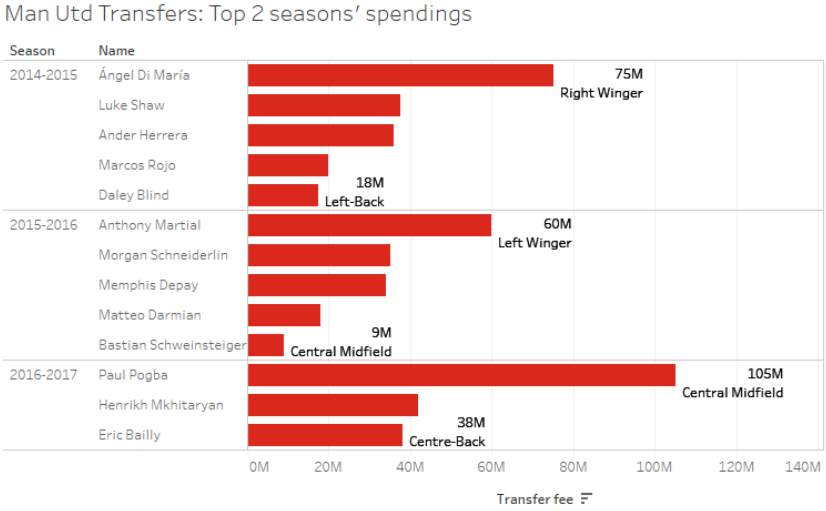
Kevin De Bruyne has been highest spend of 2015-2016 season, and John Stones has been for following season. One Midfielder and one Defender.

Man Utd:

Transfer spending for each season:



Looks like Man Utd spent the most during 2014-2015, 2016-2017 season, crossing 180MEuros each time. Lets see what they spent on during those seasons.



Angel Di Maria was the top spending in 2014-2015 season, and Paul Pogba was the biggest buy in 2016-2017 season for Manchester United, setting a club record.

Lets see how these spending have affected the premier league results by addressing the next primary question of this assessment.

2. How have the Premier League results changed for top spending clubs?

Chelsea: Chelsea has spent the most over the years. Their top 2 seasons where they spent the most on transfers were 2003-2004 and 2004-2005 seasons. Lets see how their premier league results looked for those seasons.

year	Team	Pos	Pts
2003-2004	Arsenal	1	90
2003-2004	Chelsea	2	79
2003-2004	Manchester United	3	75
2003-2004	Liverpool	4	60

year	Team	Pos	Pts
2004-2005	Chelsea	1	95
2004-2005	Arsenal	2	83
2004-2005	Manchester United	3	77
2004-2005	Liverpool	5	58

Looks like for 2003-2004 Chelsea finished 2nd , which is a good result. But it becomes even better during their second speding spree for 2004-2004 season, where they finish at 1st position at the end of the season with a whopping 95 pts. 95 pts is a premier league record for highest points at the end of the season by any club. Chelsea

Manchester City: Manchester City has been infamous for their rich Middle Eastern owners, who don't shy away from pouring money into the club and it can be seen from the previous section that Man City has consistently spend Money over the year, many time over 100M euros in one season. Lets see how their results have been for their top 2 season of spending - 2015-2016 and 2016-2017.

year	Team	Pos	Pts	W	D	L
2015-2016	Leicester City	1	81	23	12	3
2015-2016	Arsenal	2	71	20	11	7
2015-2016	Tottenham Hotspur	3	70	19	13	6
2015-2016	Manchester City	4	66	19	9	10

year	Team	Pos	Pts	W	D	L
2016-2017	Chelsea	1	93	30	3	5
2016-2017	Tottenham Hotspurs	2	86	26	8	4
2016-2017	Manchester City	3	78	23	9	6
2016-2017	Liverpool	4	76	22	10	6

Despite spending so much on those two season, Man City only managed to finsh in the top 4 of the league, once 4th and the following year 3rd. Manchester City has had better results previously and in the following season not a part of this analysis, so this can be considered as a phase where they were building the team.

Manchester United: Manchester United have the most premier league titles as compared to other teams. Lets see how their top 2 spending years (2014-2015, 2016-2017) have had an effect on their results.

year	Team	Pos	Pts
2010-2011	Manchester United	1	80
2011-2012	Manchester United	2	89
2012-2013	Manchester United	1	89
2013-2014	Manchester United	7	64
2014-2015	Manchester United	4	70
2015-2016	Manchester United	5	66
2016-2017	Manchester United	6	69

As it can be seen above in 2014-2015 season, Manchester United finished 4th, and in 2016-2017 season they finished 6th. Hence we can say that the transfers have not been effective for them. There are other factors as well like change of a long time manager which can cause such scenarios. For Manchester United, that was the case when in 2013, their long time manager – Sir Alex Ferguson, retired and team went into a long phase of rebuilding the team under new manager(s). Manchester United, infact, hold the record for most titles in Premier League and that too during the years when they didn't spend as much, possibily due other managerial aspect that went right for them.

4. Conclusion

After Analysing the transfer market trends for English Premier League teams, it can be clearly seen that a lot of money is spent each season by clubs to bring on the most talented players available in the market. Big clubs like Chelsea, Manchester City, Arsenal, Manchester United, Spurs etc. spend around 50-100 M Euros each season to keep competing for the top spots in the league. Some of these transfers do not turn profitable for the clubs in terms of their winning philosophy.

Looking at the premier league results for some of these teams that spend big bucks on player transfers, it shows that clubs like Chelsea were really successful at bringing in talent who could help them win the league. Although it is not only about spending money that can earn league titles like in the case of Manchester City and Manchester United in the recent years, but also about having proper football management in place and other factors which are not analysed as a part of this project.

In conclusion, effect of transfer spending in the premier league can be an influencing factor on the results like its seen with a lot of clubs, but it not the defining factor. There are a lot of other factors that need to go right for a club to get them end of season results, particularly in this modern era of football, where players and managers change a lot of teams.

5. Reflection

While doing this assessment I have learned following things:

- Data Sourcing: Searching for topics of interest and finding an appropriate data set. Framing questions based on the topic of interest.
- Data Wrangling and Checking: Understanding in what form the data may be required for the scope of the analysis and wrangling the data in suitable form. Checking errors in data in data like missing values and outliers and correcting them as required. Also understanding how to plan the data structure for using 2 different datasets. Used R for this.
- Data Exploration and Visualization: Used R and Tableau to explore and visualize data. Understood how different dimensions (numerical, categorical etc) need to be plotted on different kinds of plots and making appropriate use of variety of graphical representations available.
- Structure of Analysis and Report writing: Based on the results obtained from the exploration and visualization steps, understood how to answer the question in a structured way and write it in a report form.

In hindsight, I could've selected a bigger dataset with more variables that can provide more visualization avenues. Although I found the topic very interesting and so it made it easy for me to make and structure the analysis report.

6. References and Bibliography

1. Premier League Data. (2019). Retrieved from https://www.kaggle.com/limmen/premierleague-league-tables-188889-201617
2. Top 250 Football transfers from 2000 to 2018. (2019). Retrieved from https://www.kaggle.com/vardan95ghazaryan/top-250-football-transfers-from-2000-to-2018#top250-00-19.csv
3. The importance of Money in European Football | Kaggle. (2019). Retrieved from https://www.kaggle.com/filotast/the-importance-of-money-in-european-football