ENGLISH PREMIER LEAGUE DATA ANALYSIS

SEASON 2018/2019

by: Dhea Fajriati Anas
Practice Case EDA
Github Link : https://github.com/dhea1323/IYKRA-DheaFajriatiAnas_PracticeCaseEDA.git

SCENARIO

Looking back at the season that was 2018-2019 and looking to delve into sight deeper insights. Using the data to see how clubs are similar stylistically, in the way they pass, attack and score goals.

This data set is wide ranging in the sense it encompasses stats seen on a regular league table but goes beyond looking at how teams pass and keep possession, how they defend, tackle as well as looking at market values of a team and how much money each team was allotted from the TV rights deal. This data was gathered from :

1. BBC Sports Football,
2. Premierleague.com
3. Transfermarkt.co.uk

Data consists of 20 rows and 44 columns . The column of this data are team, category, general league position, finance live games television, finance tv revenue, general matches played, general won, general draw, general lost, attack scored, defence goals conceeded, general goal difference, general points, general squad size , general squad average, age, finance team market, finance market average, attack passes, attack passes through, attack passes long, attack passes back, attack crosses, attack corners taken, attack shots, attack shots on target, attack goals headed, attack goals penalty, attack goals box, attack goals outside box, general card yellow, general card red, attack goals counter, attack goals freekick, defence saves, defence blocks, defence interceptions, defence tackles, defence tackles last man, defence clearances, attack posession, and attack pass accuracy. You can download the data through this link: https://github.com/Syukrondzeko/Fellowship.

DATA PRE-PROCESSING

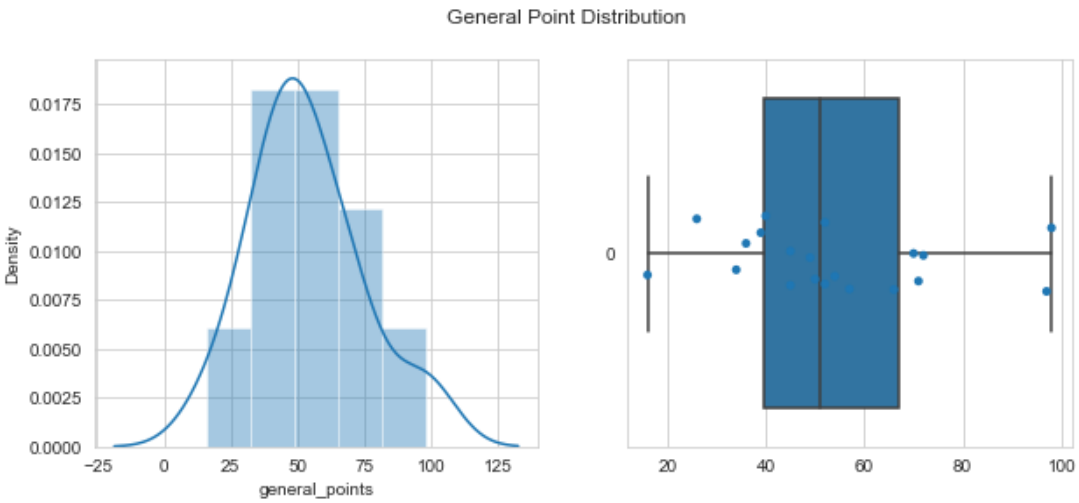
Some important things to do is to check whether there is data that is missing values, duplicates, data types that do not match, and others. There are **no missing values** and **duplicat**e in this data. However, there are mismatched data type which should be numeric/int i.e.,

1. Attack passes
2. Attack passes long
3. Attack passes back
4. Defence clearances

OBJECTIVE

POINT DISTRIBUTION

The figure shows that the general point data **looks normally distributed** and there are **no outliers**. The absolute value of **skewness is 0.6**, which means the distribution is slightly positively skewed and the distribution has more values in the tails compared to a normal distribution. The absolute value of **kurtosis is 0.4**, which means it tends to produce fewer and less extreme outliers than the normal distribution.

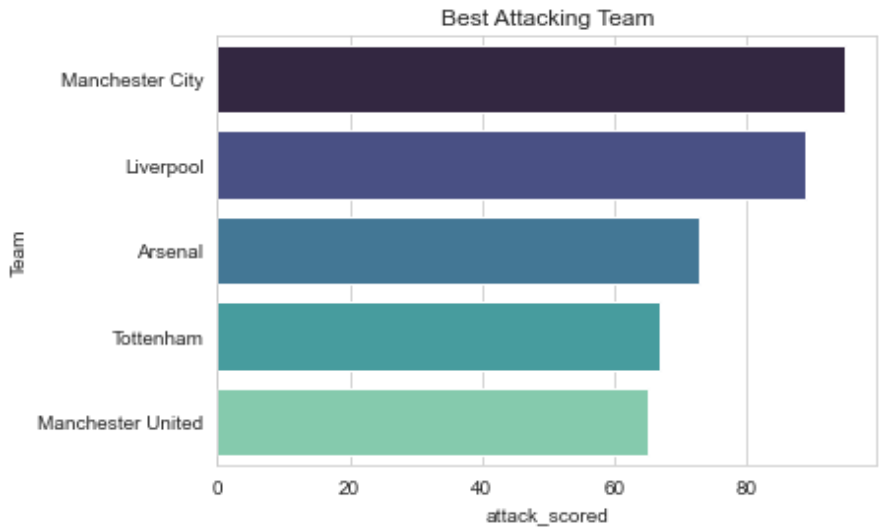


I use boxplot to find out if there are data outliers or anomalies. Statistically speaking, a boxplot provides several pieces of information, two important ones are the quartiles, represented by both ends of the box. The distance between these two quartiles is called the Interquartile Range (IQR).

OBJECTIVE

BEST ATTACKING TEAMS

In my opinion, the best attacking team are based on the number of attack scored or total goals in this season. Here's the result,

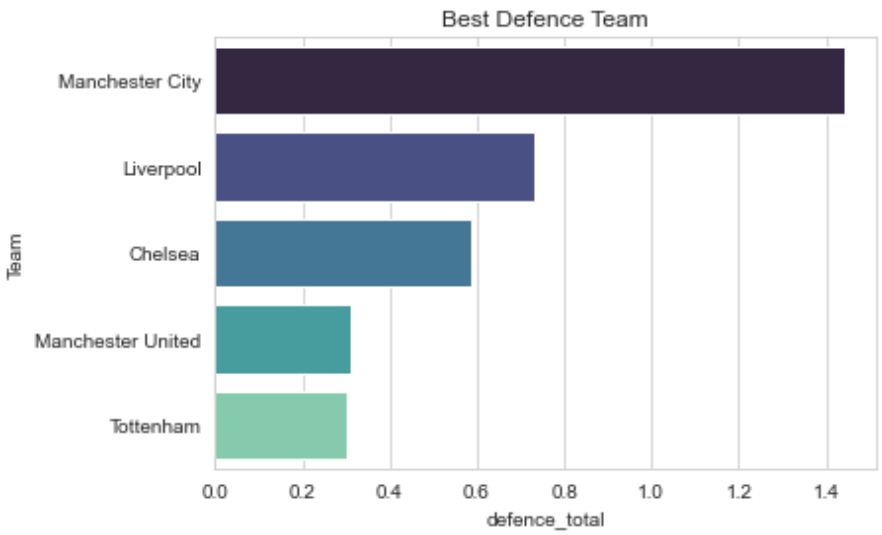


The figure shows that the best attacking team in season 2018/2019 is Manchester City, which is 95, total goals, followed by the next order are Liverpool, Arsenal, Tottenham, and Manchester United.

OBJECTIVE

BEST DEFENCING TEAMS

In my opinion, the best defencing team are based on defence_saves, defence_blocks,defence_interceptions,defence_tackles,defence_tackle_last_man, and defence_clearances. I used MinMaxScaler() to normalize the values, then sum them. Here's the result,



The figure shows that the best defencing team in season 2018/2019 is Manchester City, followed by the next order are Liverpool, Chelsea, Manchester United, and Tottenham. According to https://www.whoscored.com/ , Machester City is the best defencing team in England Premier League Season 2018/2019.

Team	Rating
1. Manchester City	7.09
2. Liverpool	7.01
3. Chelsea	6.89

OBJECTIVE

BEST TEAM IN FINANCIAL ASPECT

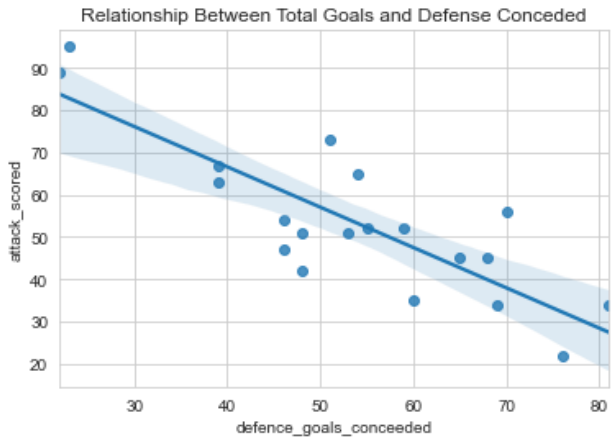
In my opinion, the best team in financial aspect are based on, finance tv revenue, finance team market, and finance market average, then average them. Here's the result,

The figure shows that the best team in financial aspect in season 2018/2019 is Manchester City which is 398M, followed by the fourth after are Liverpool, Chelsea, Tottenham, and Arsenal.

OBJECTIVE

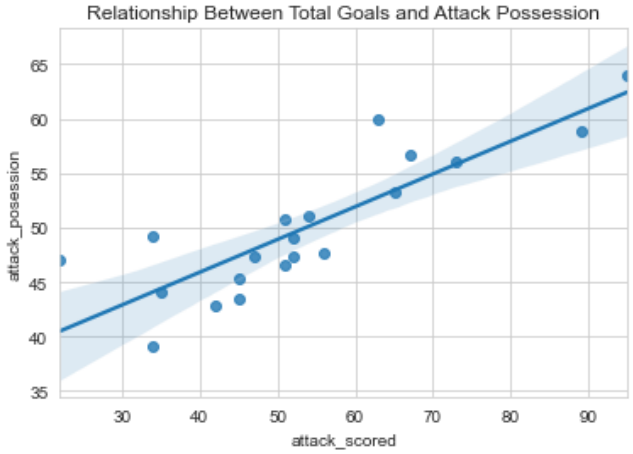
TOTAL GOALS VS DEFENSE CONCEDED

The figure shows that relationship between total goals and defence conceded is negative correlation, which correlation coefficient is -0.83. It means that the more goals scored, the less the team's effort to withstand the opponent's attack.



TOTAL GOALS VS ATTACK POSSESSION

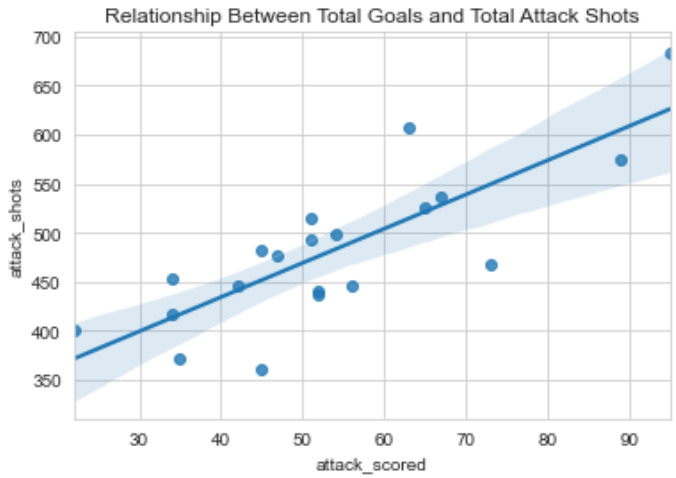
The figure shows that relationship between total goals and attack possession is positive correlation, which correlation coefficient is 0.85. It means that the more teams have possession of the ball, the higher the probability of scoring goals.



OBJECTIVE

TOTAL GOALS VS TOTAL ATTACK SHOTS

The figure shows that relationship between total goals and total attack shots is is positive correlation, which correlation coefficient is 0.8. It means that the more total attack shots, That means the more chances, the more chances to score goals.



TOTAL LOST VS ATTACK POSSESSION

The figure shows that relationship between total lost and attack possession is is negative correlation, which correlation coefficient is -0.79. It means that the less teams have possession of the ball, the higher the probability of lose.