PROJECT DISSERTATION

ON

# ANALYSIS AND PREDICTION FOR ENGLISH PREMIER LEAGUE

SUBMITTED BY

**MR. NIKHIL MOHAN JOSHI**

UNDER THE GUIDANCE OF

**PROF. MR. ABHIJEET GOLE**

SUBMITTED IN PARTIAL FULFILMENT

OF REQUIREMENT FOR QUALIFYING

MSc. CS [SEMESTER-III]

IN THE YEAR 2019-20

**DEPARTMENT OF**

**COMPUTER SCIENCE AND INFORMATION TECHNOLOGY**

**RAMNARAIN RUIA AUTONOMOUS COLLEGE**

**MATUNGA, MUMBAI-400 019**

# ANALYSIS AND PREDICTION FOR ENGLISH PREMIER LEAGUE

## DECLARATION

I the undersigned, student of **M.Sc. COMPUTER SCIENCE PART II** hereby declare that I have attempted to complete the project on “ **ANALYSIS AND PREDICTION FOR ENGLISH PREMIER LEAGUE** “. The information submitted is true and original to the best of Knowledge.

Signature of the student

Name of the student

**Nikhil Mohan Joshi**

## ACKNOWLEDGEMENT

I acknowledge my sincere gratitude to all those who helped me in my project work of the MSc.CS.

It is a great pleasure that I present my first venture in this application in the form of project work.

I would like to thank my project guide Prof. Abhijeet Gole to have permitted me to go ahead with this project and appreciating my work at every stage.

I would like to thank my Professors Megha Sawant, Rasika Mundhe, Edith Michael and Mahavir Advaya for their valuable inputs and advice; I would also like to thank Vinod Sawant Sir for technical assistance.

I would like to convey my sincere thanks to my parents for their unconditional support throughout and providing me with the best of everything.

I would like to thank all my classmates, friends for all being there boosting my confidence and for their constant encouragement, cooperation and support.

## Table of Contents

## TITLE AND ABSTRACT

### ANALYSIS AND PREDICTION FOR ENGLISH PREMIER LEAGUE

The objective of this study is to analyze the gathered data of the three seasons (2015/16, 2016/17 and 2017/18) of English Premier League and to predict the outcome of the league for the upcoming season(2018/19).

The entire analysis was conducted with the help of datasets and R tool. Datasets are files in CSV format. The analysis is done with the premier leagues data of the seasons 2015/16, 2016/17 and 2017/18.

The English Premier League data consisted the list of teams which played across the three seasons mentioned above. Alongwith the list of teams, data had various statistics which were helpful in determining the performance of each team.

In Football; wins, goals and clean sheets play a major role for forecasting the outcome. Hence, analyzing wins, goals and clean sheets of each team can help in predicting the ranking of teams for the next season.

## BACKGROUND AND LITERATURES REVIEW

The English Premier League is considered as the premium league in the game of Football. It is the most followed football league in the world. It also gains huge attention from the broadcasters of sports from around the world. The teams playing in the league are under constant pressure for performing well and securing a respectable spot in the table.

Predictive analysis is the process of analyzing data using automated statistical processes and summarizing results into useful information.

Analysis of different aspects of the game which are responsible for determining the outcome of the match are helpful for team management which can then take measures to improve in those aspects in which they lack. Improving those can help strengthen the team.

English Premier League also has a strong hold in betting market. Huge number of bets are placed to determine the winning team. Analysis on various aspects of the game can help the betters to place correct bets.

## PROJECT INTRODUCTION AND SCOPE

There are 6 major factors which contribute to the result of the match, they are:-

1. Number of goals scored.
2. Number of clean sheets kept.
3. Shots on target made by the team.
4. Shots taken from inside the box.
5. Goals conceded.
6. Saves made by the goalkeeper of the team.

Ranking Algorithm helps to determine where a team stands comparative to other teams in the above mentioned departments. Use of Ranking Algorithm can also provide the team with their ranking in all these fields individually, so that they can figure out the departments which needs improvement and work on the same.

The final ranking of the teams in the table is determined on number of wins. Generally, team with highest number of wins ends up winning the league. "*Goals*" and "*clean sheets"* are the two factors on which "*wins"* is dependent. Both are directly proportional to wins. I.e. the more the number of goals scored and clean sheets kept, more are the chances of winning.

Furthermore, for determining "*goals"*, "*shots on target"* & "*shots from inside box"* are the two main factors; and for determining "*clean sheets", "goals conceded"* & *"saves"* are the factors. Analysis of all these factors leads us to predict the overall rank of the team, which also can be considered as the finishing position in the table at the end of the season.

## SYSTEM DESIGN AND FLOW

**Dataset**: The datasets used are of xls and csv formats.

**Dataset origin**: www.kaggle.com

**R Studio and R Language:**

In this system the Software used is RStudio:

As RStudio is a free and open-source integrated development environment (IDE) for R, a programming language for statistical computing and graphics

In this System the Technology used is R:

R is an open-source language and environment for statistical computing and data visualization, supporting data manipulation and transformations, as well as sophisticated graphical displays.

**Packages Used:**

**ggplot2**:

ggplot2 is a plotting system for R, based on the grammar of graphics, which tries to take the good parts of base and lattice graphics and none of the bad parts. It takes care of many of the fiddly details that make plotting a hassle (like drawing legends) as well as providing a powerful model of graphics that makes it easy to produce complex multi-layered graphics.

**plotly**:

Plotly provides online graphing, analytics, and statistics tools for individuals and collaboration.

**Dplyr**:

Dplyr is the next iteration of plyr, focused on tools for working with data frames (hence the d in the name).

**Shiny**:

Shiny is a new package from RStudio that makes it incredibly easy to build interactive web applications with R.

**Shiny dashboard**:

Create dashboards with Shiny. This package provides a theme on top of Shiny, making it easy to create attractive dashboards.

**RColorBrewer**:

Provides color schemes for maps (and other graphics) designed by Cynthia Brewer as described at http://colorbrewer2.org.

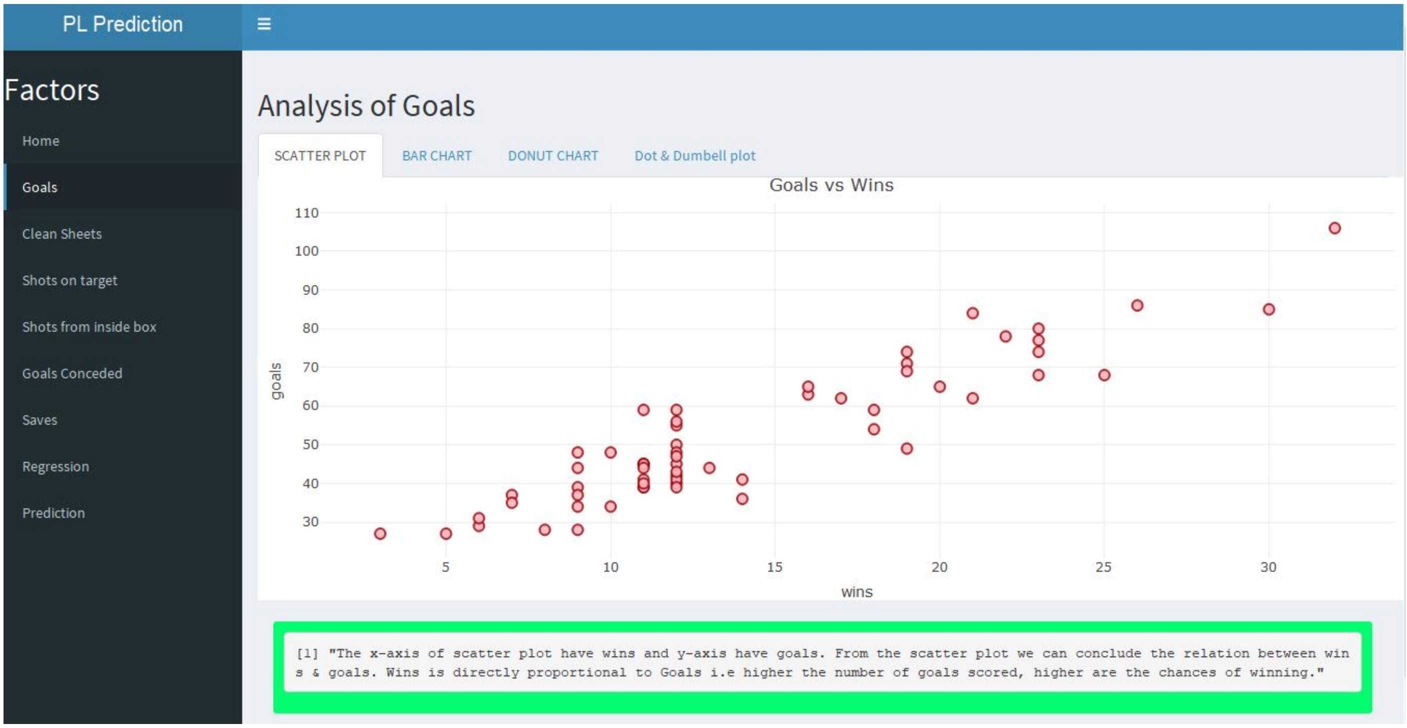
**Session:**

Utility functions for interacting with R processes from external programs. This package includes functions to save and restore session information (including loaded packages, and attached data objects), as well as functions to evaluate strings containing R commands and return the printed results or an execution transcript.

## RESULTS AND INTERPRETATIONS

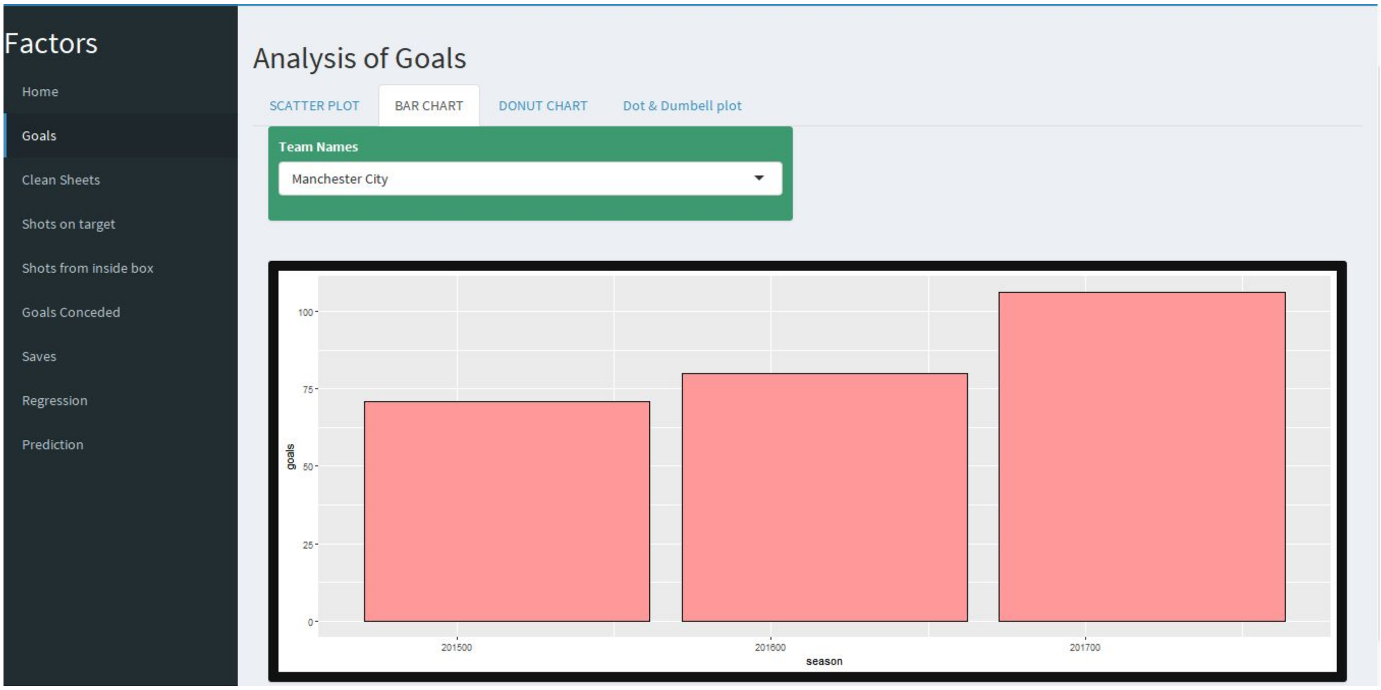
### **A**. *FACTOR - GOALS*

1. Graph showing direct relation between goals and wins.



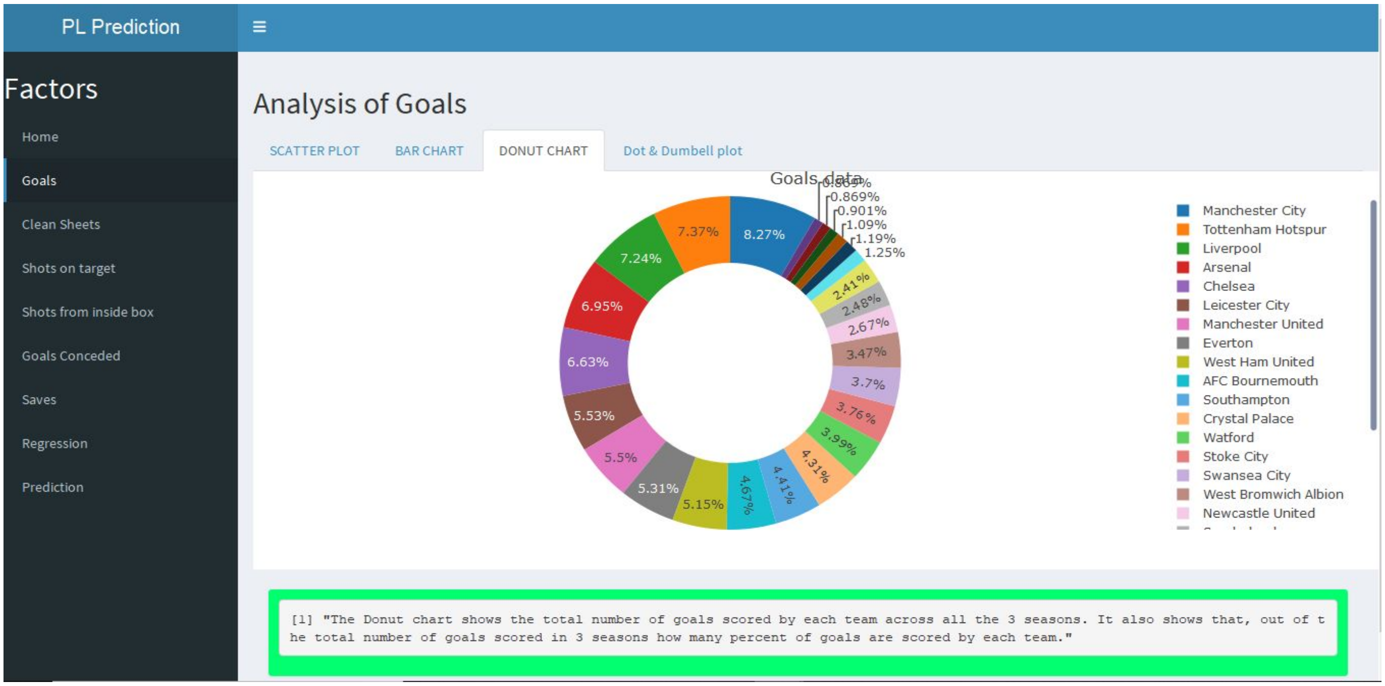
```
[1] "The x-axis of scatter plot have wins and y-axis have goals. From the scatter plot we can conclude the relation between win
s & goals. Wins is directly proportional to Goals i.e higher the number of goals scored, higher are the chances of winning."
```

2. In which season a particular team has scored more number of goals ?



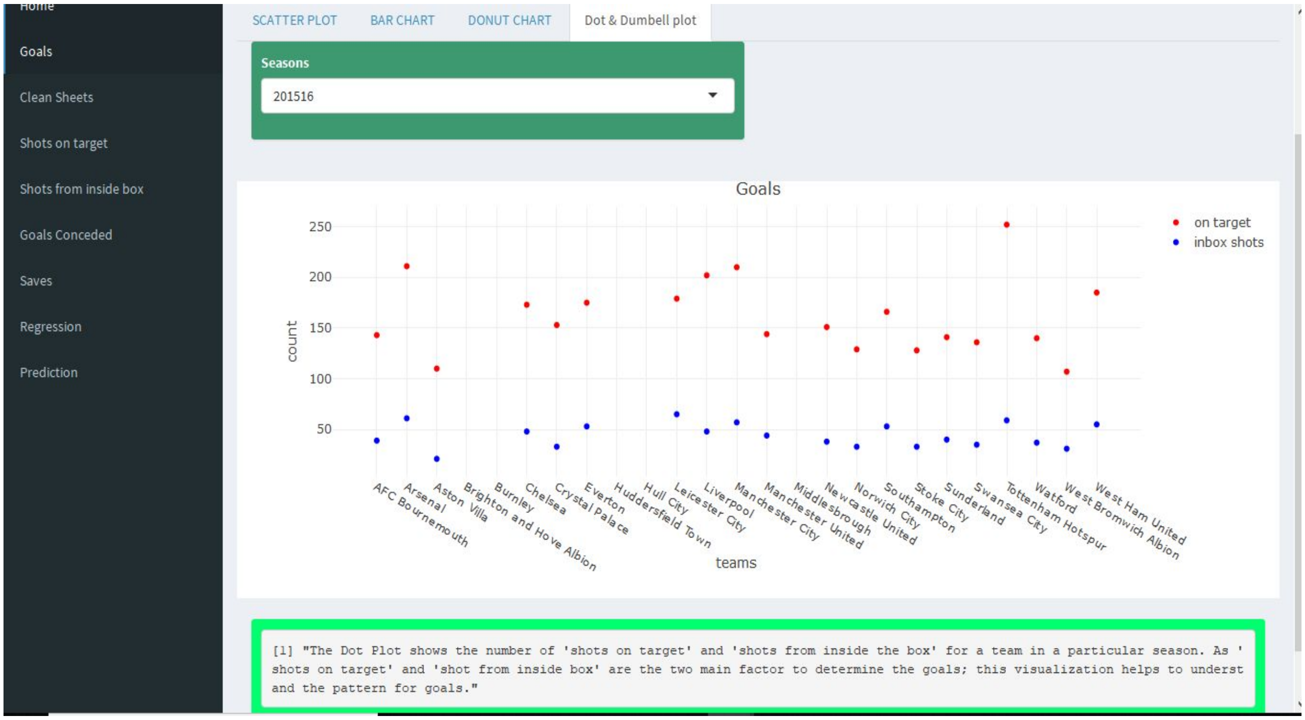
Team Names: Manchester City

3. Which team has scored the highest number of goals in 3 seasons combined ?



```
[1] "The Donut chart shows the total number of goals scored by each team across all the 3 seasons. It also shows that, out of t
he total number of goals scored in 3 seasons how many percent of goals are scored by each team."
```

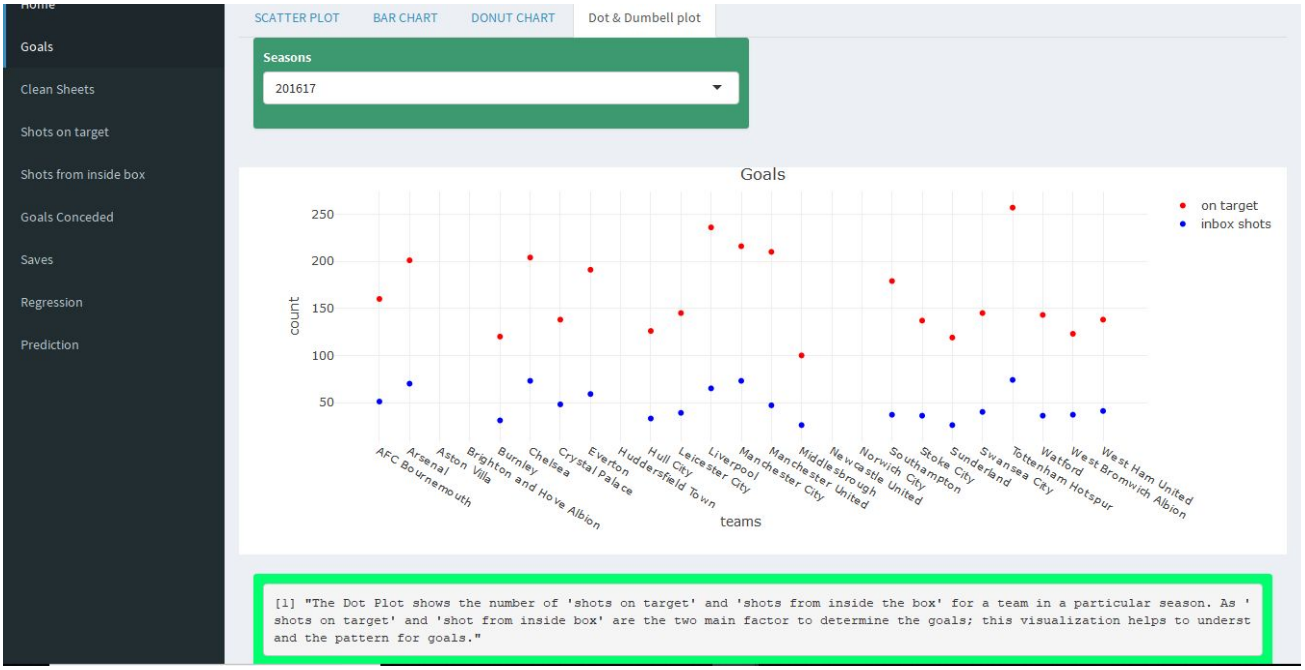
4. Which team had most number of “shots on target” and “shots from inside the box” in season 2015/16 ?



Seasons: 201516

```
[1] "The Dot Plot shows the number of 'shots on target' and 'shots from inside the box' for a team in a particular season. As '
shots on target' and 'shot from inside box' are the two main factor to determine the goals; this visualization helps to underst
and the pattern for goals."
```

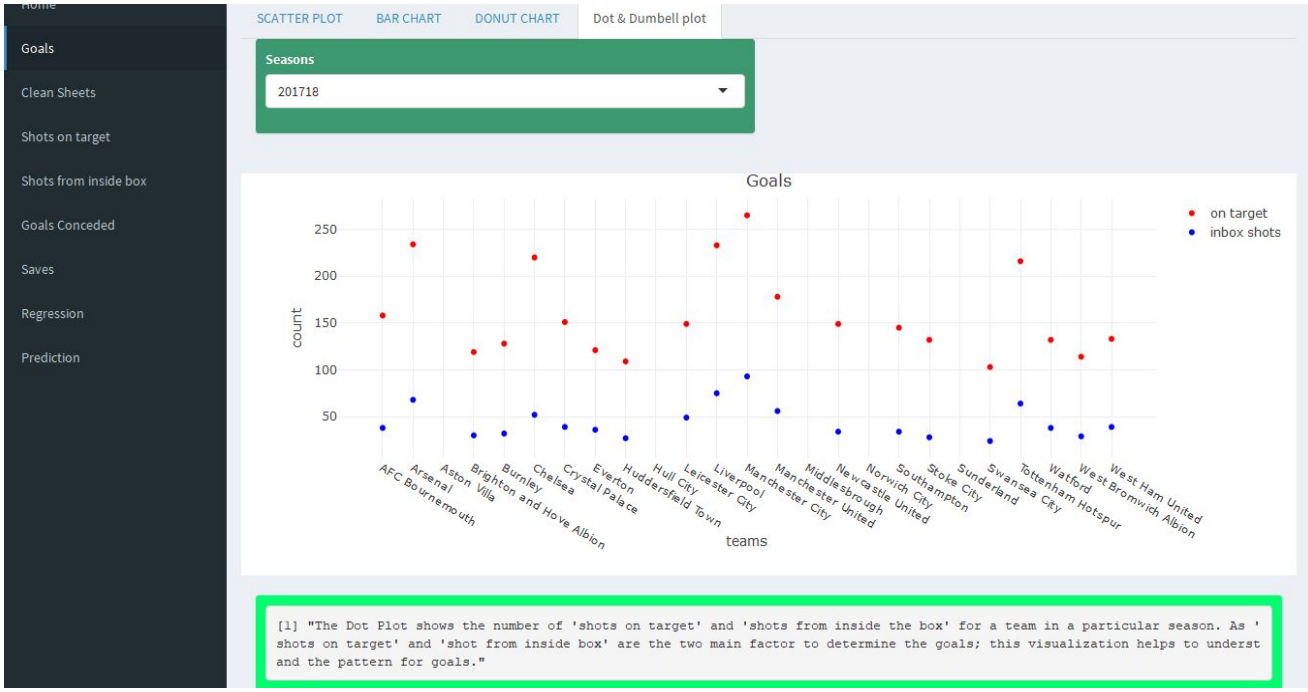
5. Which team had most number of “shots on target” and “shots from inside the box” in season 2016/17 ?



Seasons: 201617

```
[1] "The Dot Plot shows the number of 'shots on target' and 'shots from inside the box' for a team in a particular season. As '
shots on target' and 'shot from inside box' are the two main factor to determine the goals; this visualization helps to underst
and the pattern for goals."
```

6. Which team had most number of “shots on target” and “shots from inside the box” in season 2017/18 ?

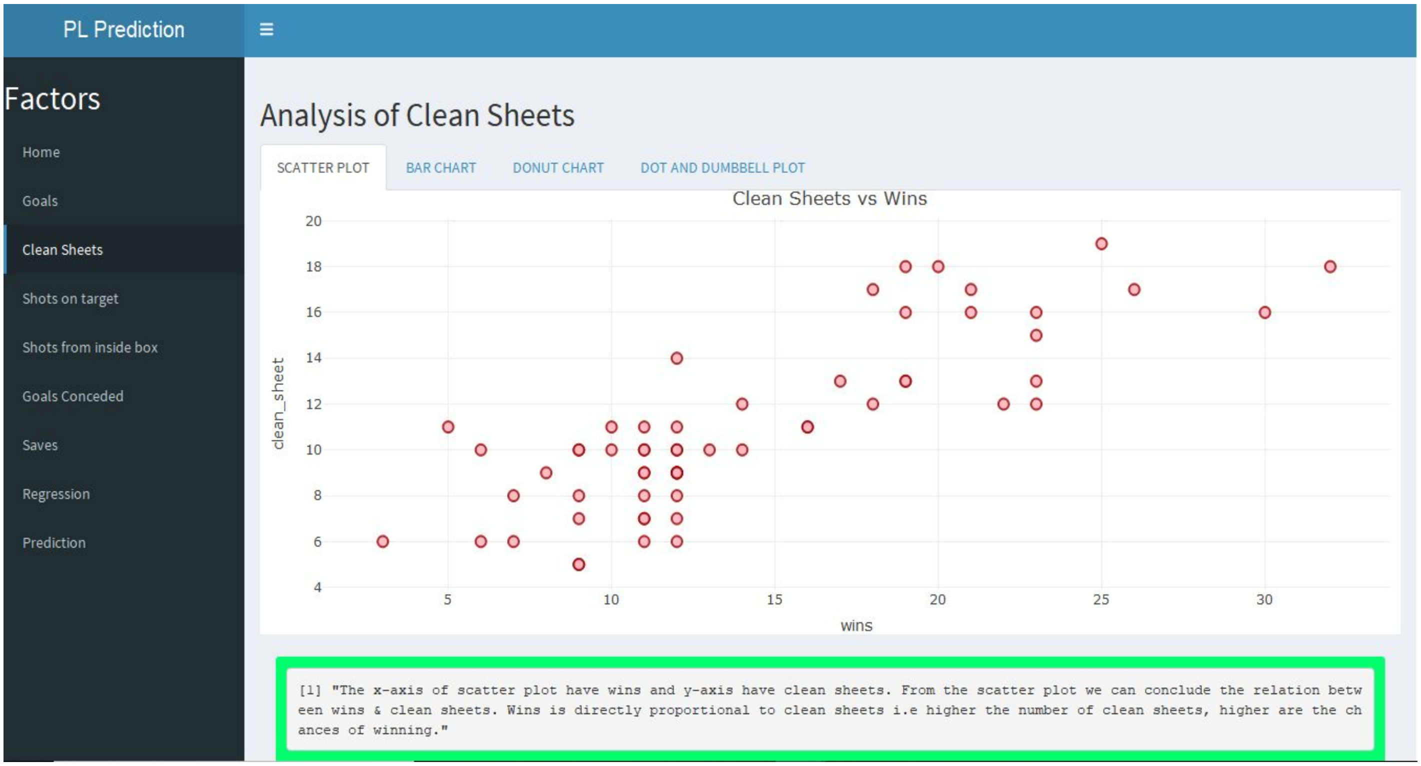


Seasons: 201718

```
[1] "The Dot Plot shows the number of 'shots on target' and 'shots from inside the box' for a team in a particular season. As '
shots on target' and 'shot from inside box' are the two main factor to determine the goals; this visualization helps to underst
and the pattern for goals."
```

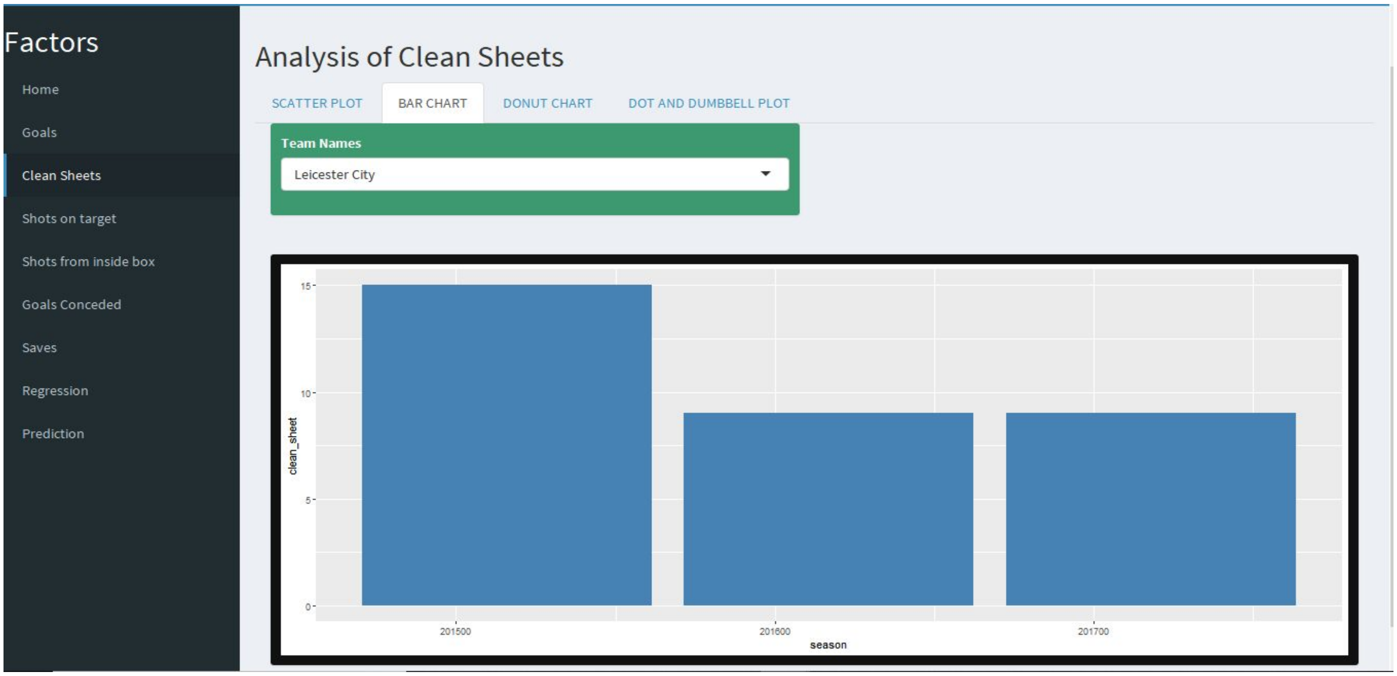
## **B**. *FACTOR - CLEAN SHEETS*

1. Graph showing direct relation between clean sheets and wins.



```
[1] "The x-axis of scatter plot have wins and y-axis have clean sheets. From the scatter plot we can conclude the relation betw
een wins & clean sheets. Wins is directly proportional to clean sheets i.e higher the number of clean sheets, higher are the ch
ances of winning."
```

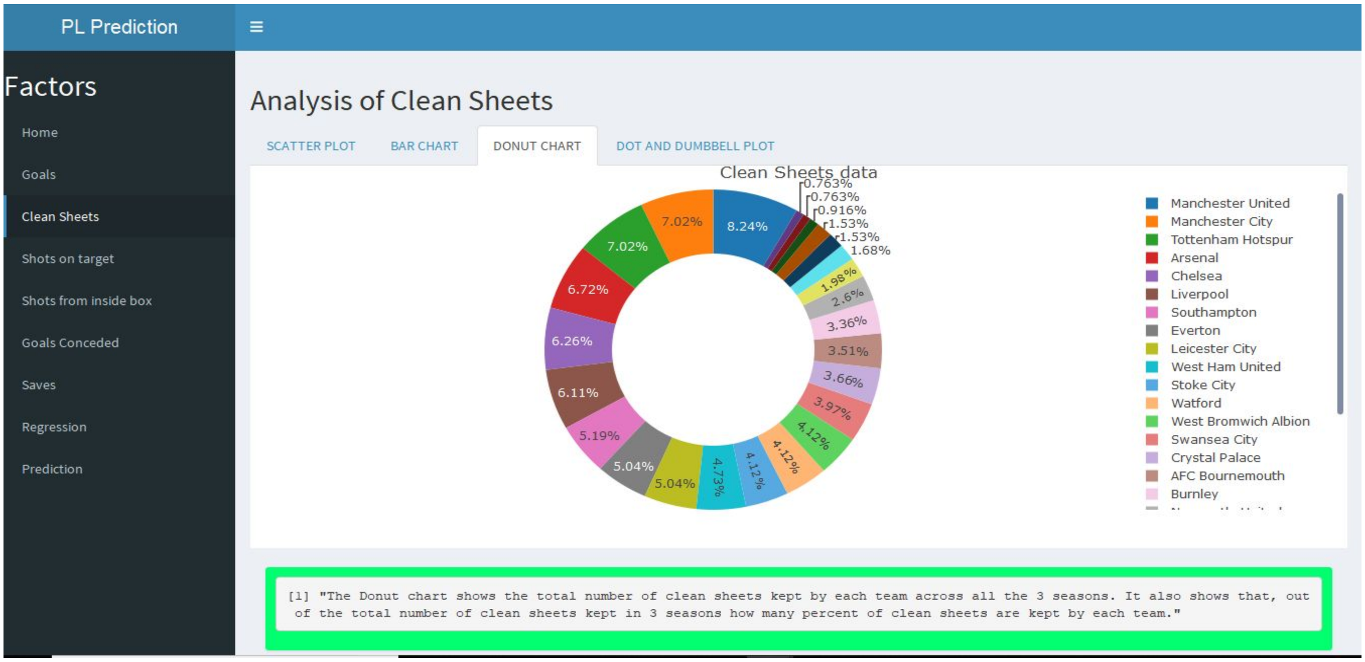
2. In which season a particular team has kept more clean sheets?



Analysis of Clean Sheets

Team Names: Leicester City

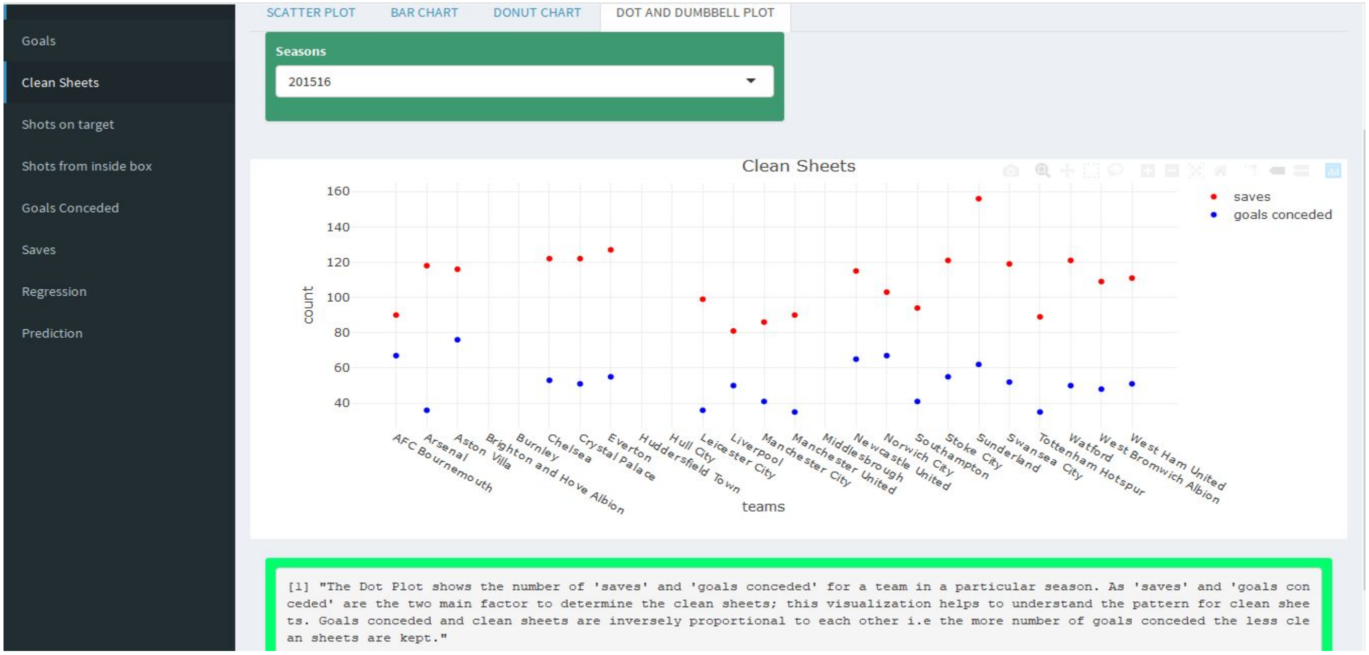
3. Which team has kept the highest number of clean sheets in 3 seasons combined ?



Analysis of Clean Sheets

```
[1] "The Donut chart shows the total number of clean sheets kept by each team across all the 3 seasons. It also shows that, out
 of the total number of clean sheets kept in 3 seasons how many percent of clean sheets are kept by each team."
```

4. Which team had most number of “goals conceded” and “saves” in season 2015/16 ?

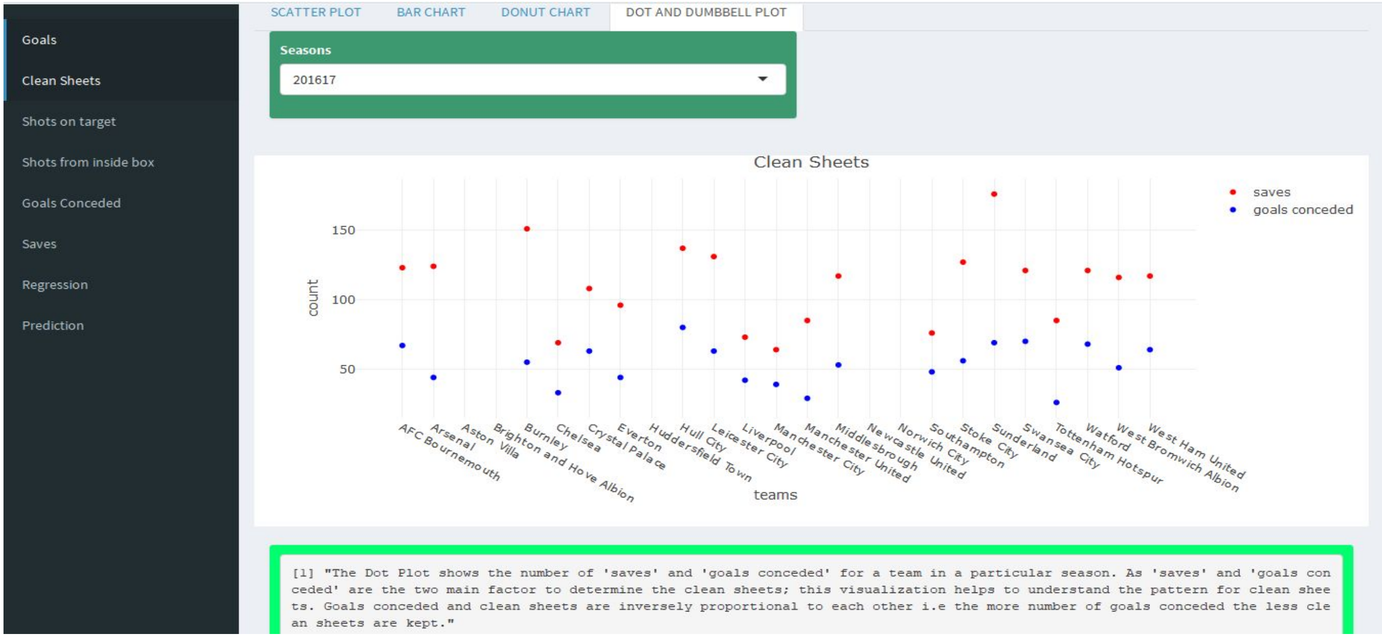


Seasons

201516

```
[1] "The Dot Plot shows the number of 'saves' and 'goals conceded' for a team in a particular season. As 'saves' and 'goals con
ceded' are the two main factor to determine the clean sheets; this visualization helps to understand the pattern for clean shee
ts. Goals conceded and clean sheets are inversely proportional to each other i.e the more number of goals conceded the less cle
an sheets are kept."
```

5. Which team had most number of “goals conceded” and “saves” in season 2016/17 ?

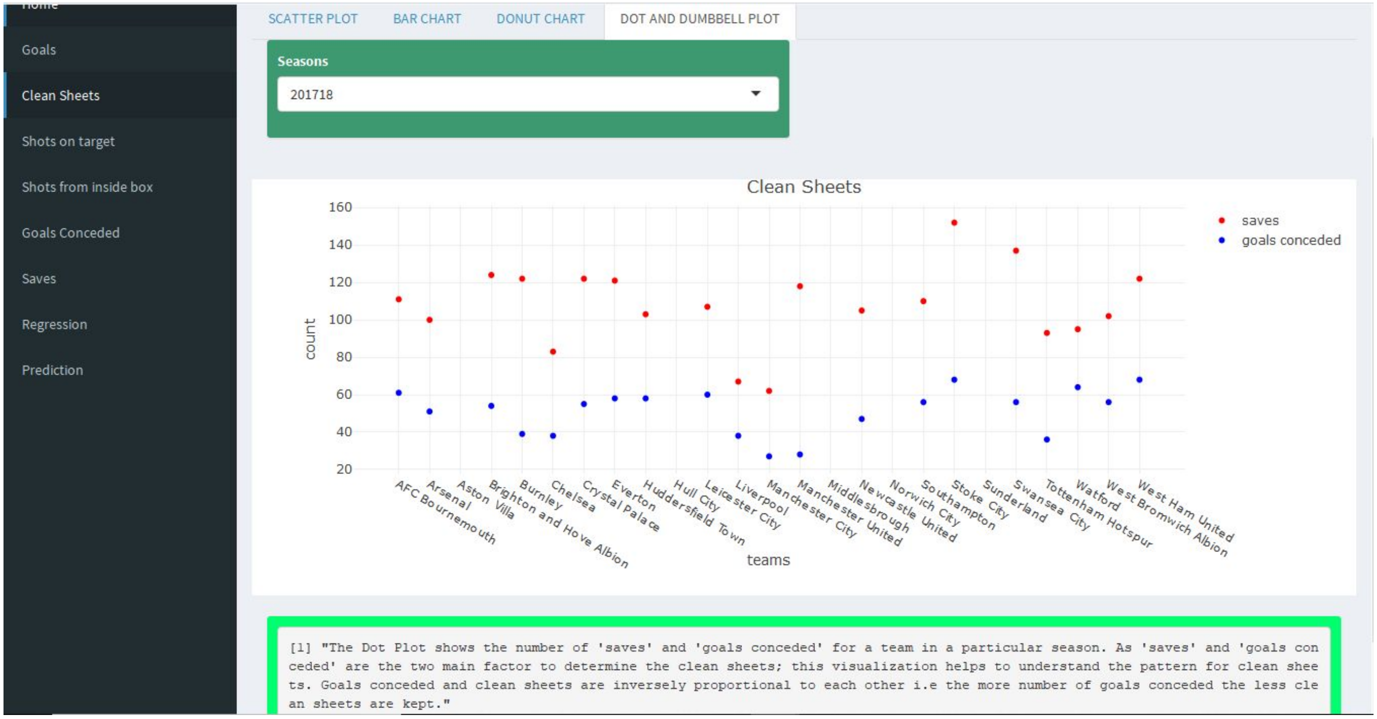


Seasons

201617

```
[1] "The Dot Plot shows the number of 'saves' and 'goals conceded' for a team in a particular season. As 'saves' and 'goals con
ceded' are the two main factor to determine the clean sheets; this visualization helps to understand the pattern for clean shee
ts. Goals conceded and clean sheets are inversely proportional to each other i.e the more number of goals conceded the less cle
an sheets are kept."
```

6. Which team had most number of “goals conceded” and “saves” in season 2017/18 ?

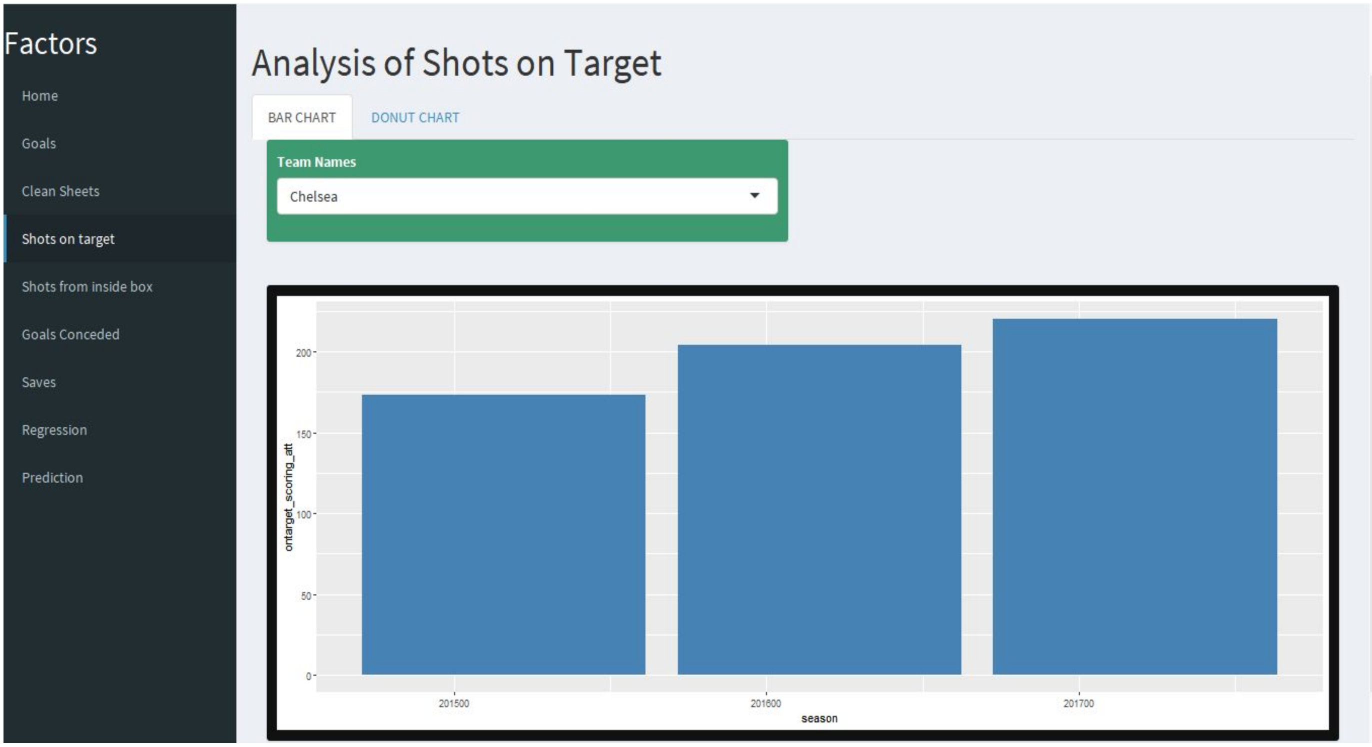


Seasons: 201718

```
[1] "The Dot Plot shows the number of 'saves' and 'goals con
ceded' are the two main factor to determine the clean shee
ts. Goals conceded and clean sheets are inversely proportional to each other i.e the more number of goals conceded the less cle
an sheets are kept."
```

## C. *FACTOR - SHOTS ON TARGET*

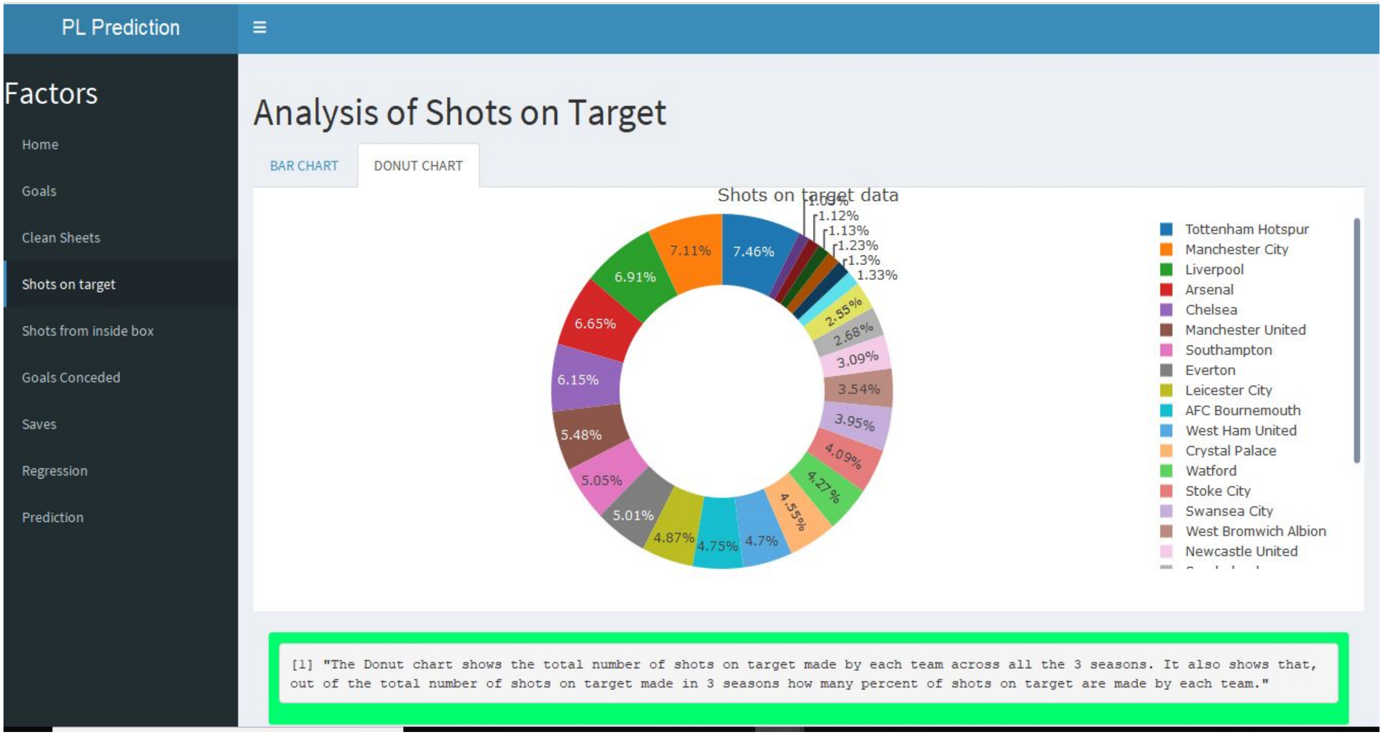
1. In which season a particular team has attempted more shots on target ?



Team Names

Chelsea

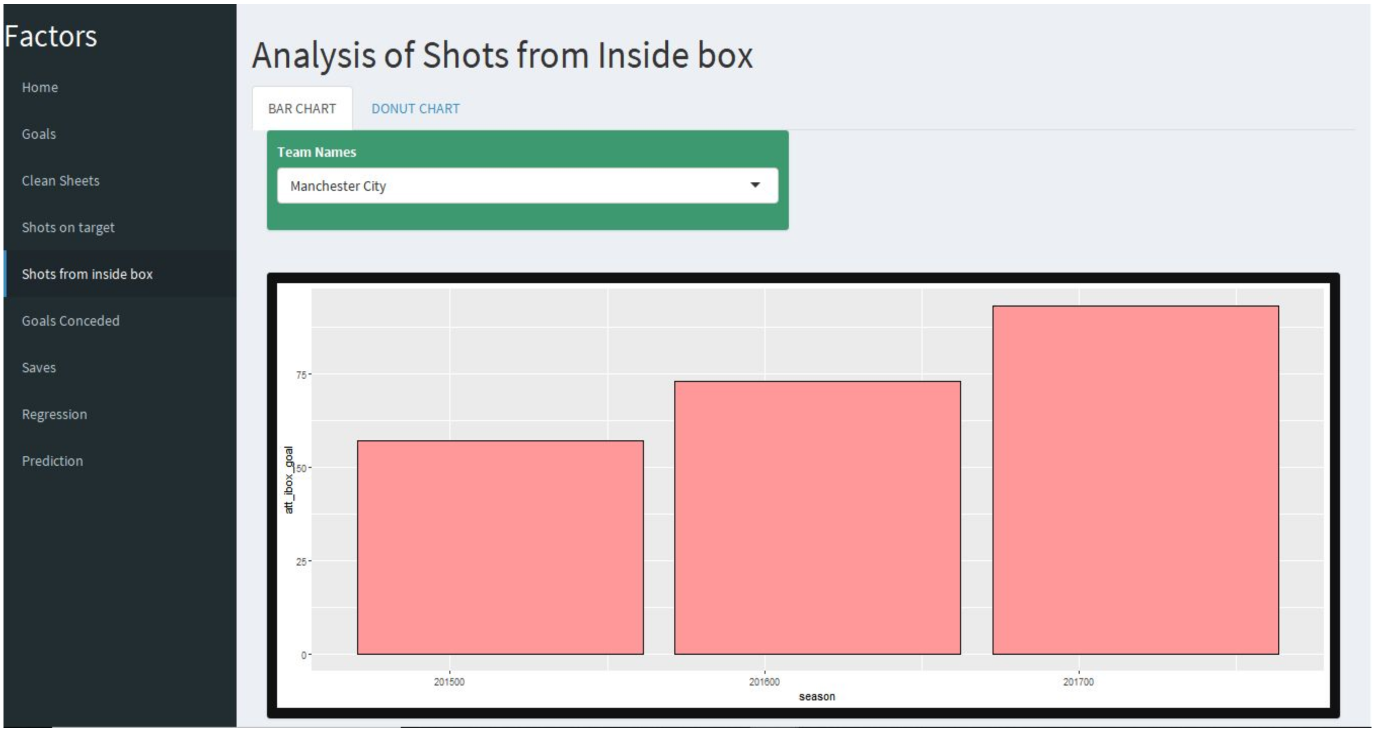
2. Which team has attempted the highest number of shots on target in 3 seasons combined ?



```
[1] "The Donut chart shows the total number of shots on target made by each team across all the 3 seasons. It also shows that,
out of the total number of shots on target made in 3 seasons how many percent of shots on target are made by each team."
```

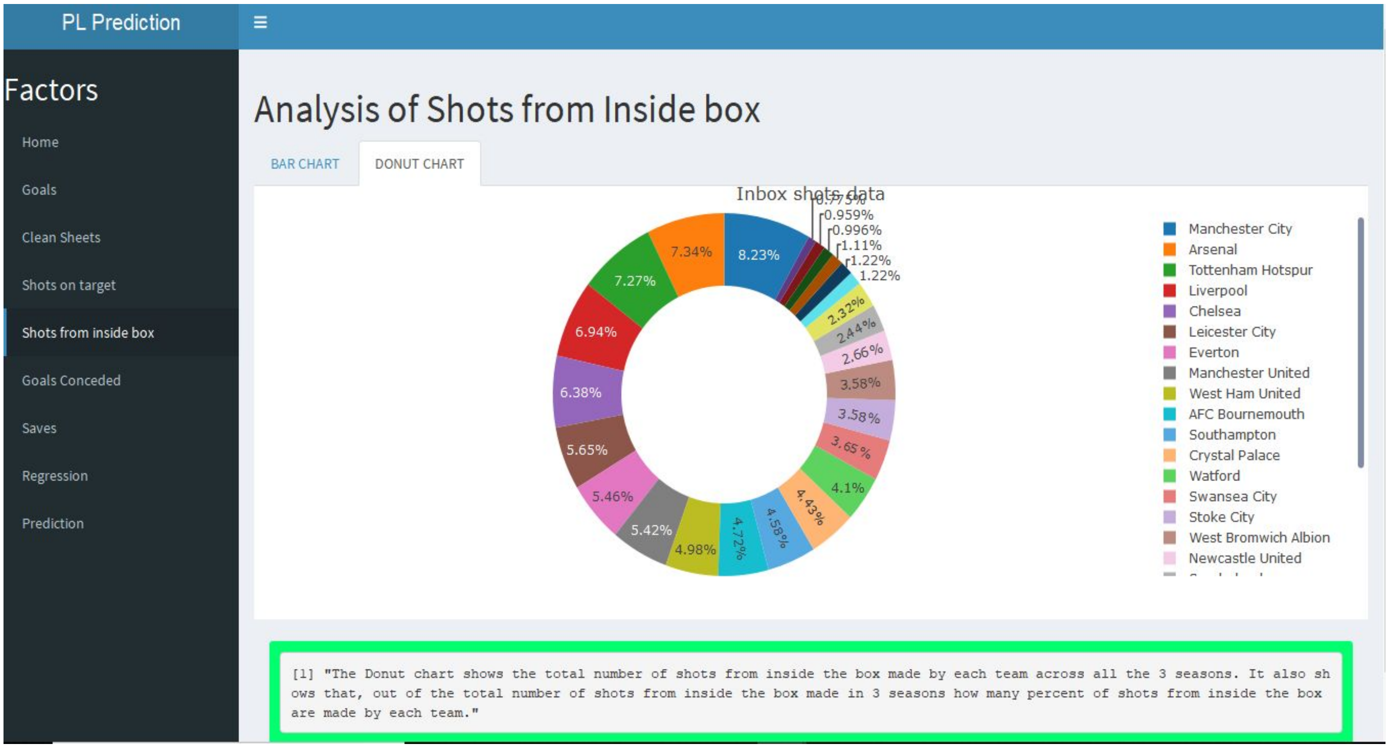
## **D**. *FACTOR - SHOTS FROM INSIDE THE BOX*

1. In which season a particular team has attempted more shots from inside the box ?



Team Names: Manchester City

2. Which team has attempted the highest number of shots from inside the box in 3 seasons combined ?

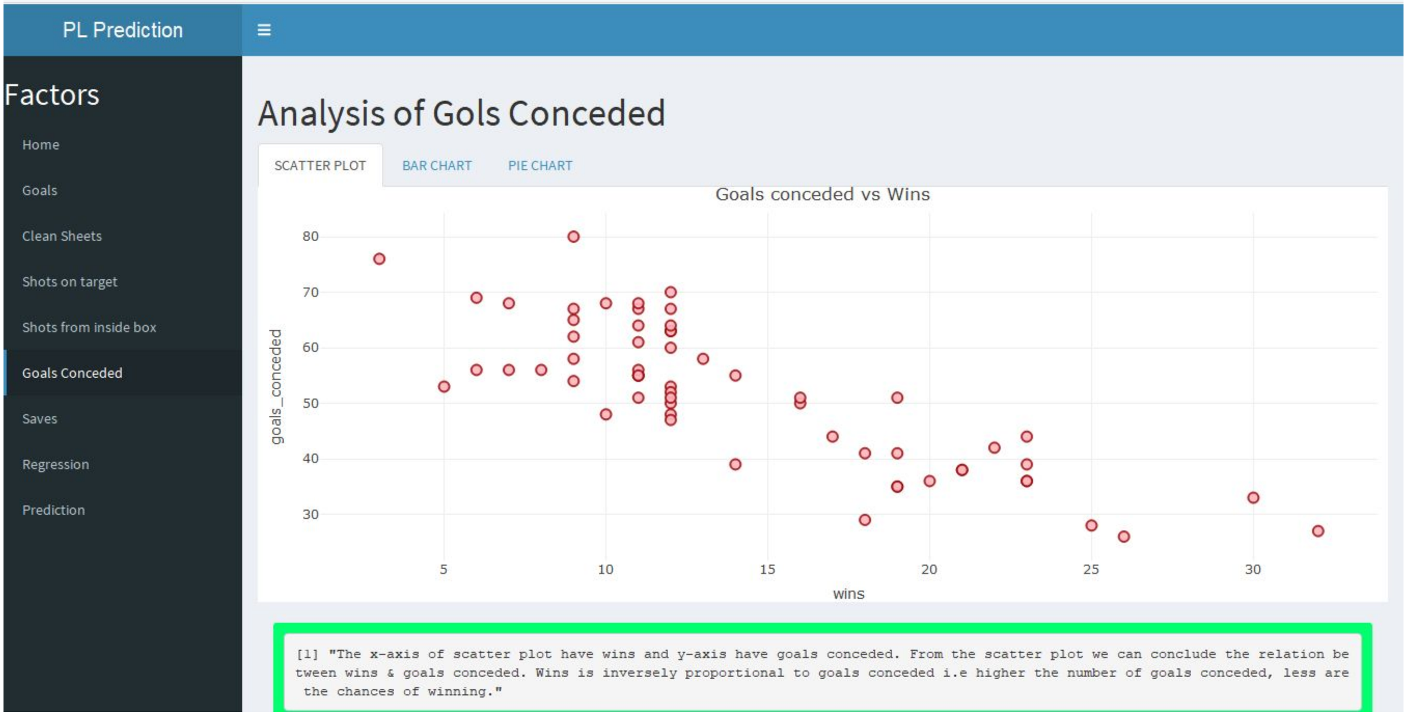


## Analysis of Shots from Inside box

```
[1] "The Donut chart shows the total number of shots from inside the box made by each team across all the 3 seasons. It also sh
ows that, out of the total number of shots from inside the box made in 3 seasons how many percent of shots from inside the box
are made by each team."
```

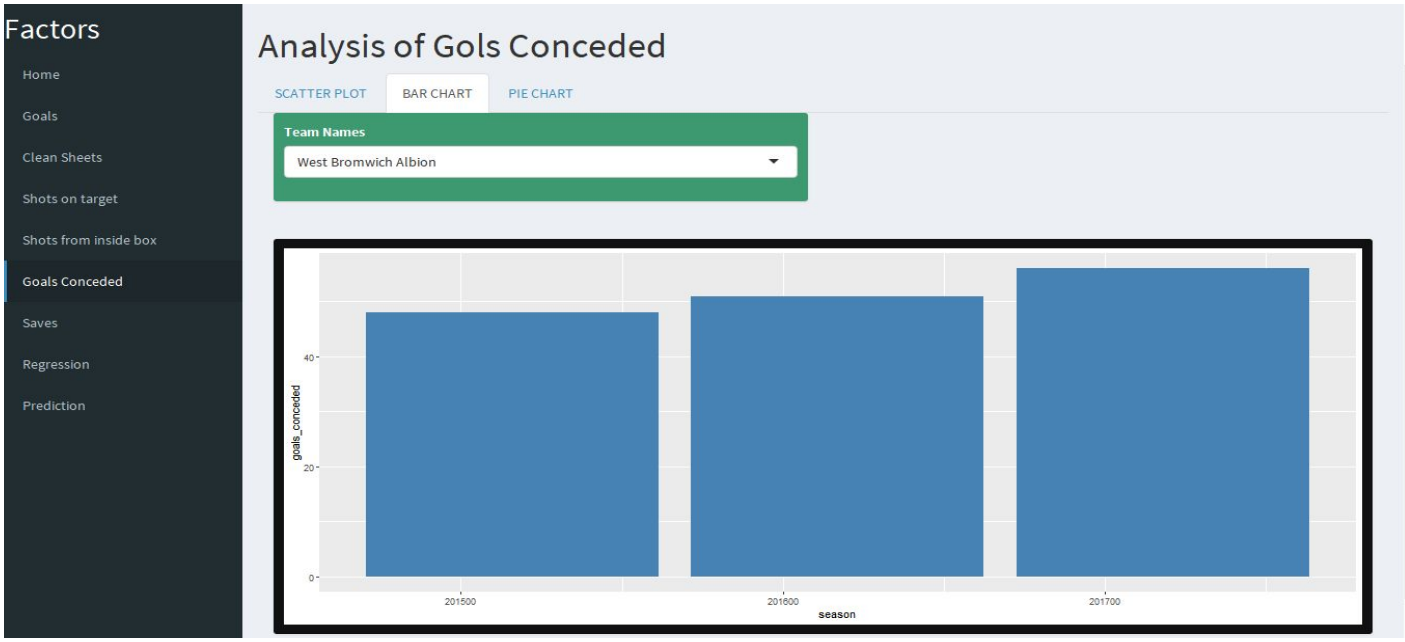
## **E**. *FACTOR - GOALS CONCEDED*

1. Graph showing direct relation between goals conceded and wins.

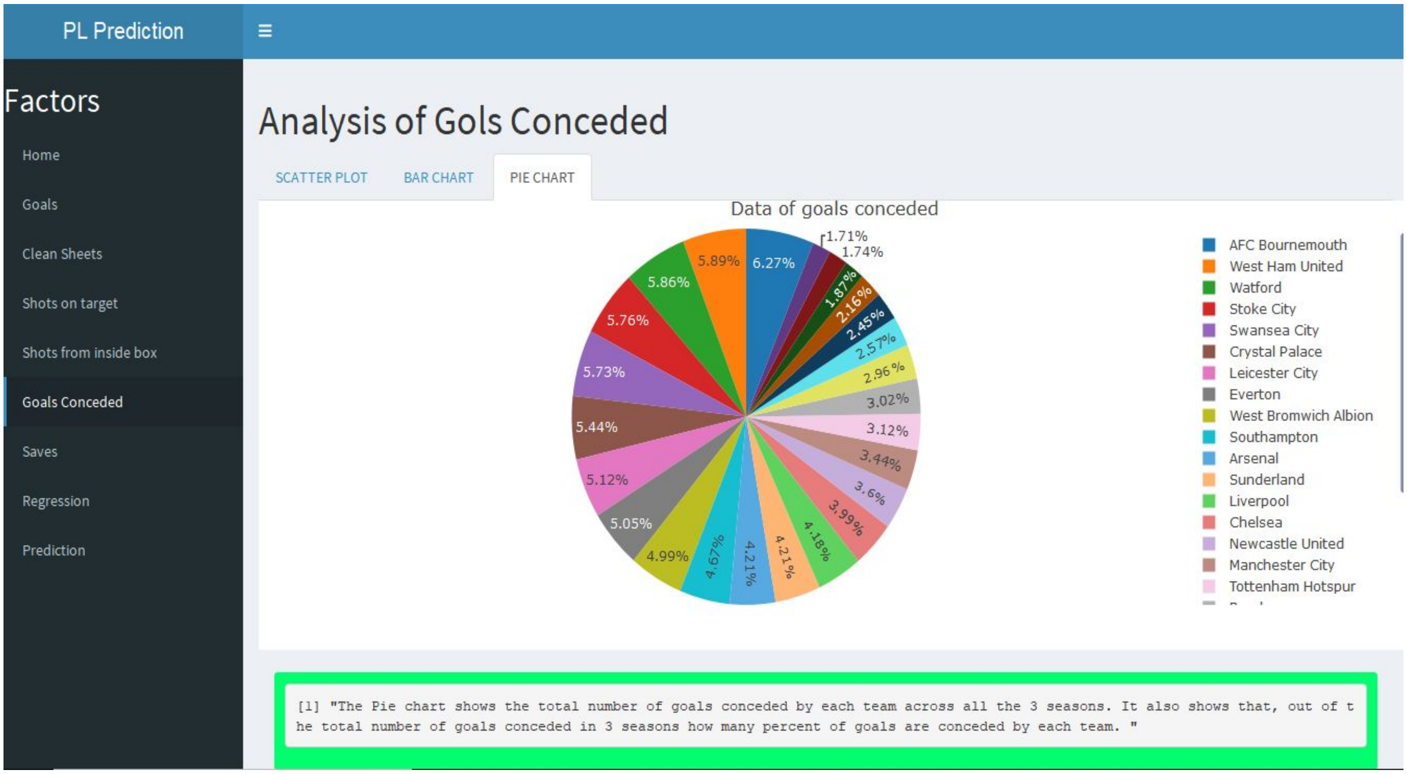


```
[1] "The x-axis of scatter plot have wins and y-axis have goals conceded. From the scatter plot we can conclude the relation be
tween wins & goals conceded. Wins is inversely proportional to goals conceded i.e higher the number of goals conceded, less are
 the chances of winning."
```

2. In which season a particular team has conceded more number of goals ?



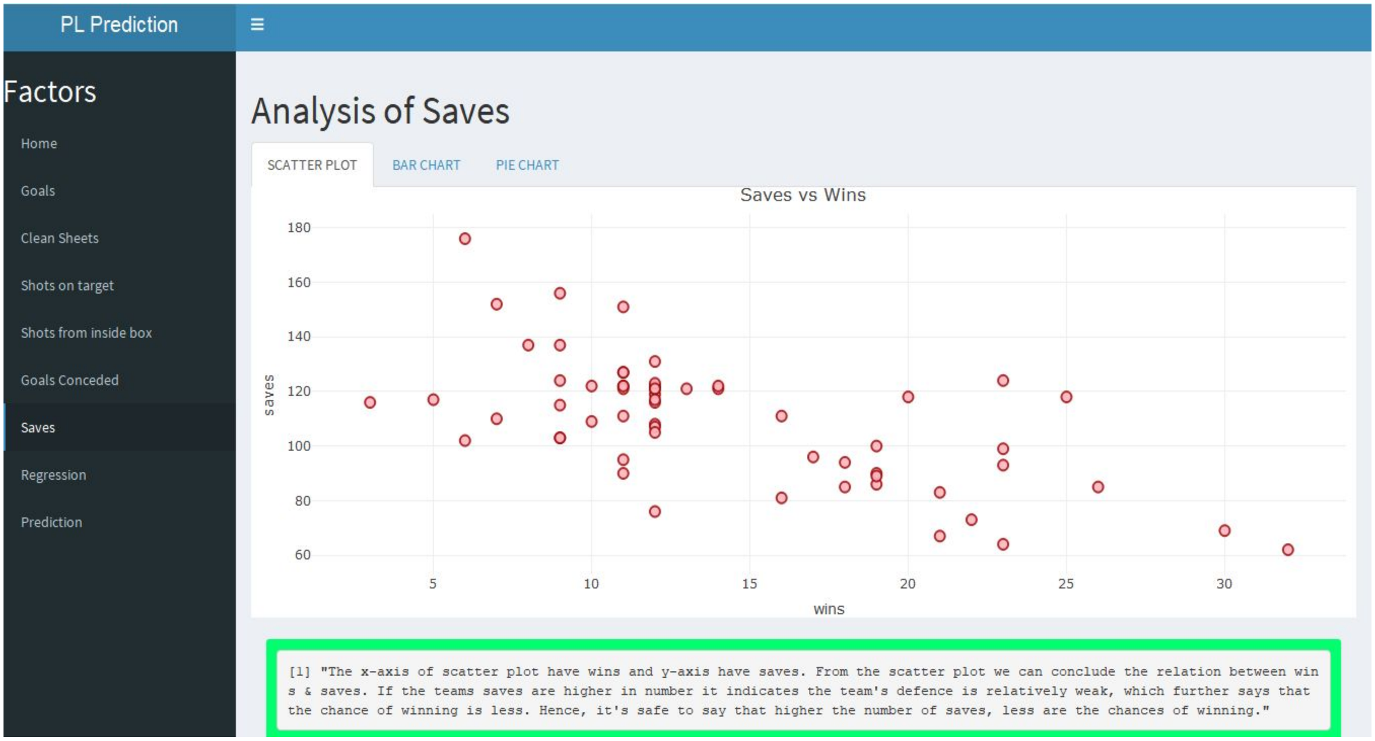
3. Which team has conceded most goals in 3 seasons combined ?



```
[1] "The Pie chart shows the total number of goals conceded by each team across all the 3 seasons. It also shows that, out of t
he total number of goals conceded in 3 seasons how many percent of goals are conceded by each team. "
```

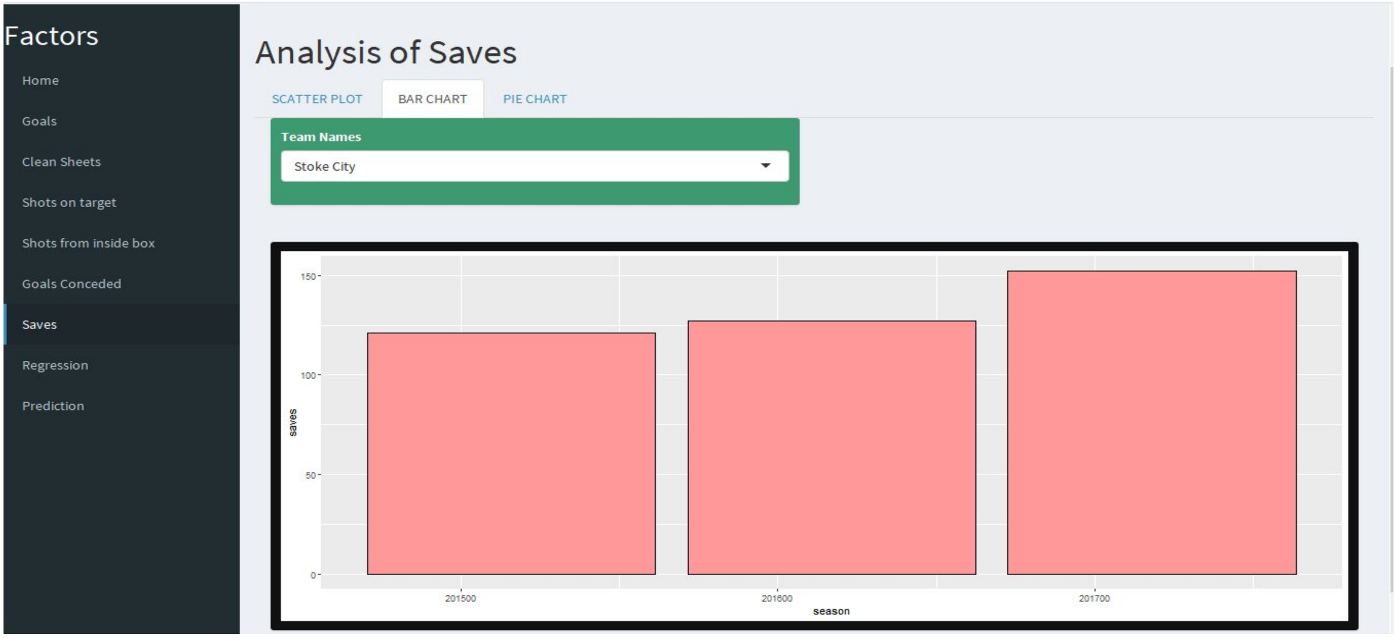
## F. *FACTOR - SAVES*

1. Graph showing direct relation between saves and wins.

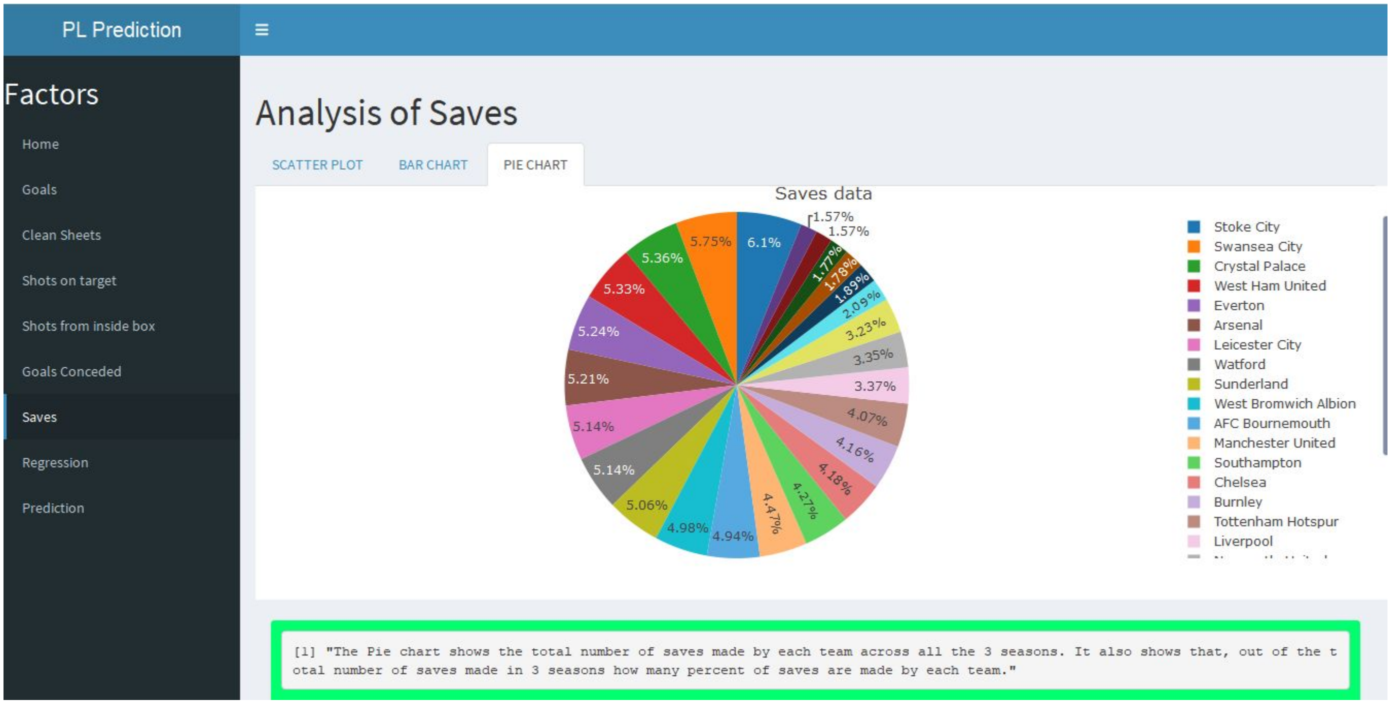


```
[1] "The x-axis of scatter plot have wins and y-axis have saves. From the scatter plot we can conclude the relation between win
s & saves. If the teams saves are higher in number it indicates the team's defence is relatively weak, which further says that
the chance of winning is less. Hence, it's safe to say that higher the number of saves, less are the chances of winning."
```

2. In which season a particular team has made more number of saves ?



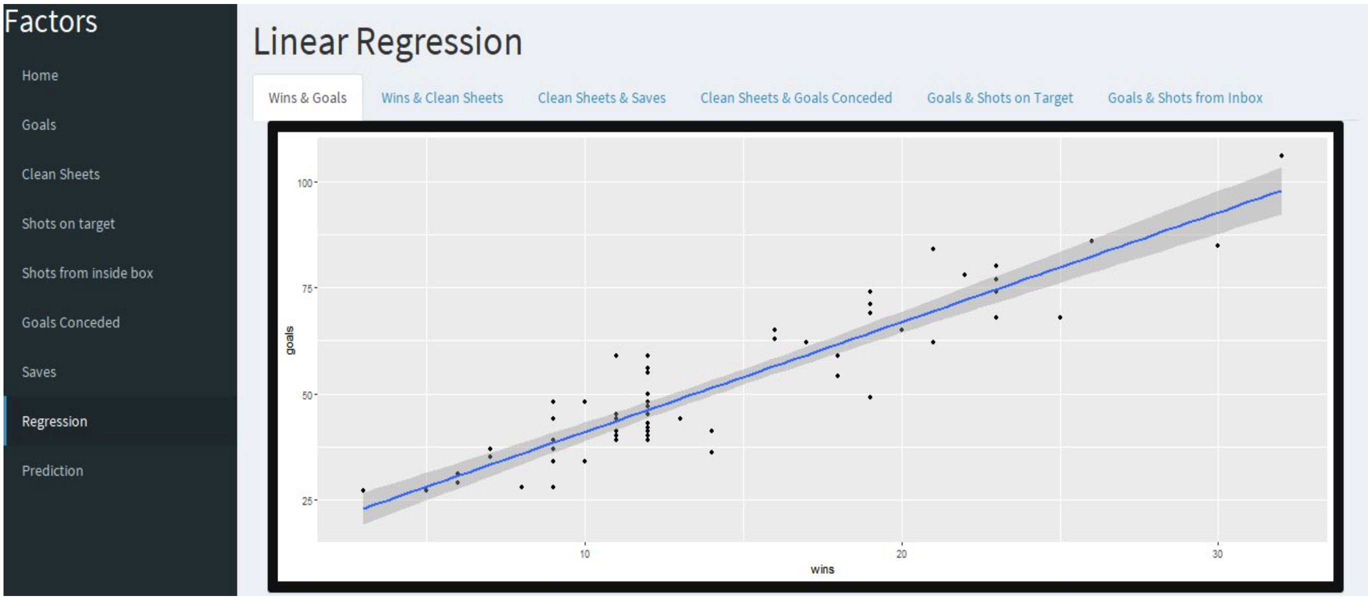
3. Which team has made most saves in 3 seasons combined ?



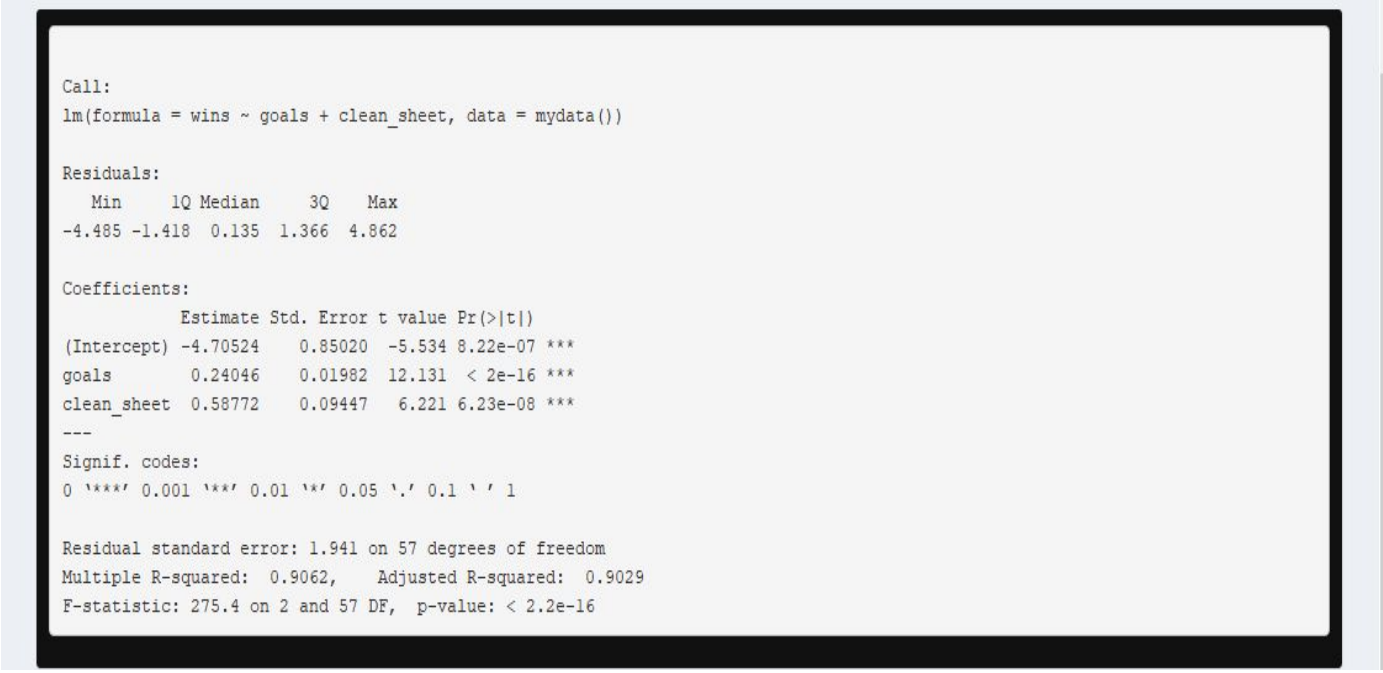
```
[1] "The Pie chart shows the total number of saves made by each team across all the 3 seasons. It also shows that, out of the t
otal number of saves made in 3 seasons how many percent of saves are made by each team."
```

# REGRESSION

1. Graph for - Wins vs Goals

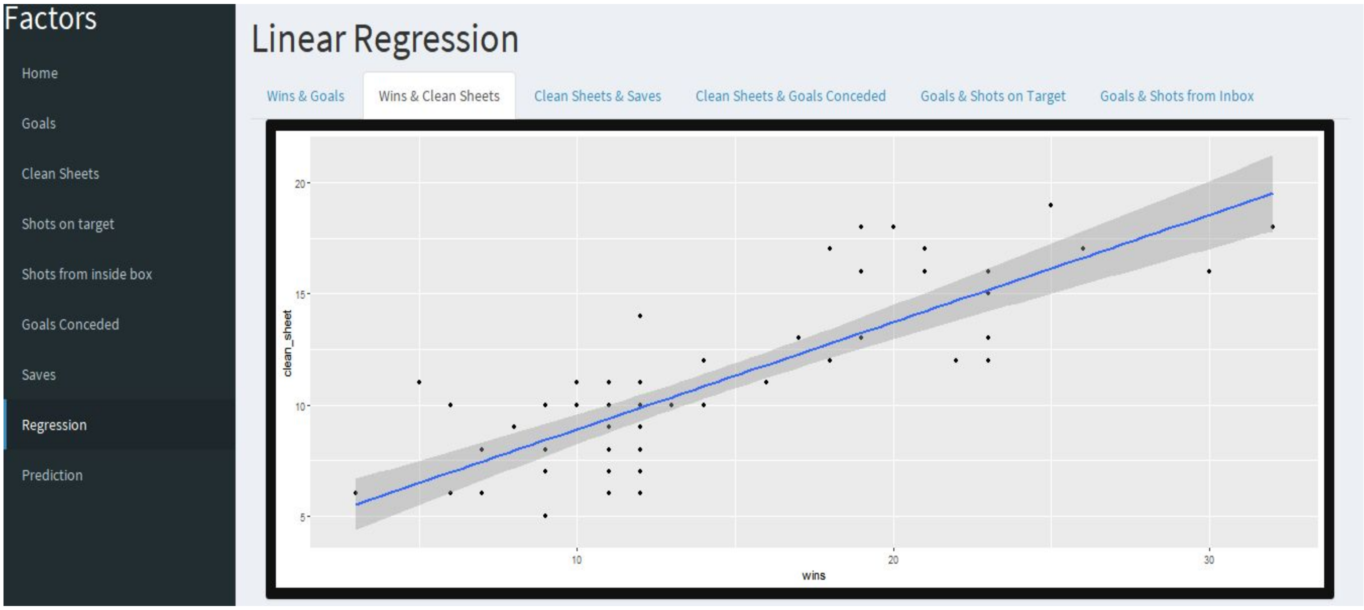


Summary for - Wins vs Goals

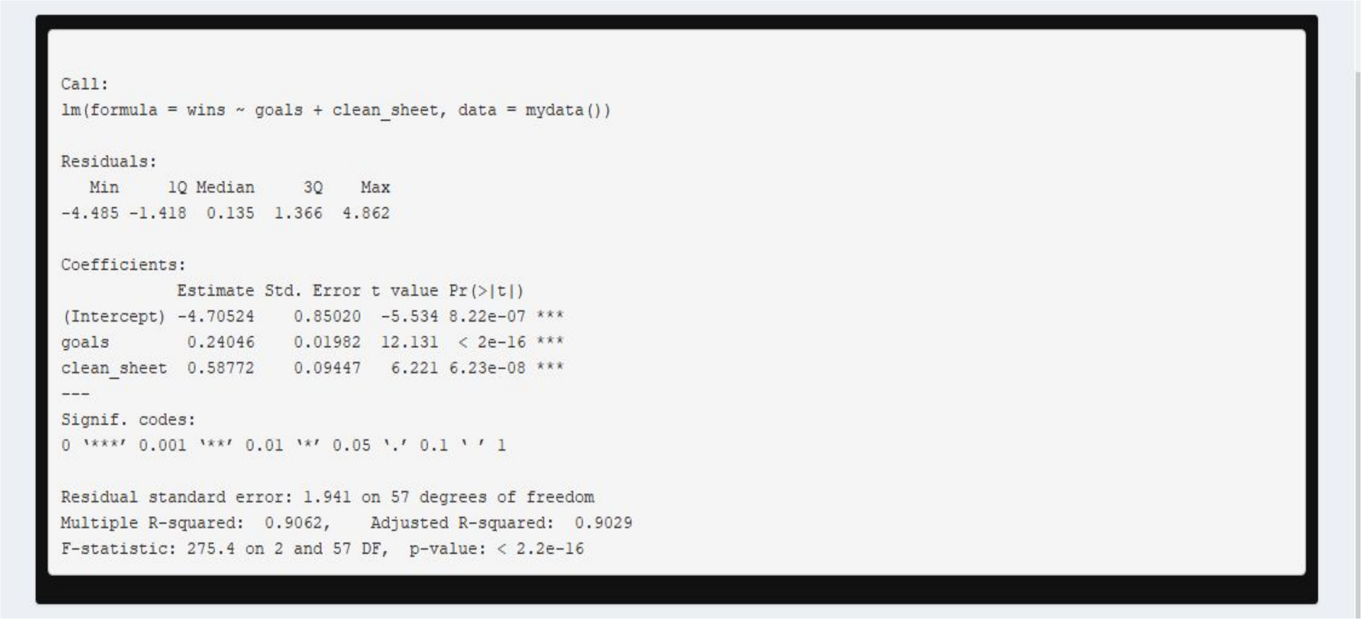


```
Call:
lm(formula = wins ~ goals + clean_sheet, data = mydata())

Residuals:
   Min     1Q Median     3Q    Max 
-4.485 -1.418  0.135  1.366  4.862 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -4.70524    0.85020  -5.534 8.22e-07 ***
goals        0.24046    0.01982  12.131  < 2e-16 ***
clean_sheet  0.58772    0.09447   6.221 6.23e-08 ***
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.941 on 57 degrees of freedom
Multiple R-squared:  0.9062,	Adjusted R-squared:  0.9029 
F-statistic: 275.4 on 2 and 57 DF,  p-value: < 2.2e-16
```

2. Graph for - Wins vs Clean Sheets

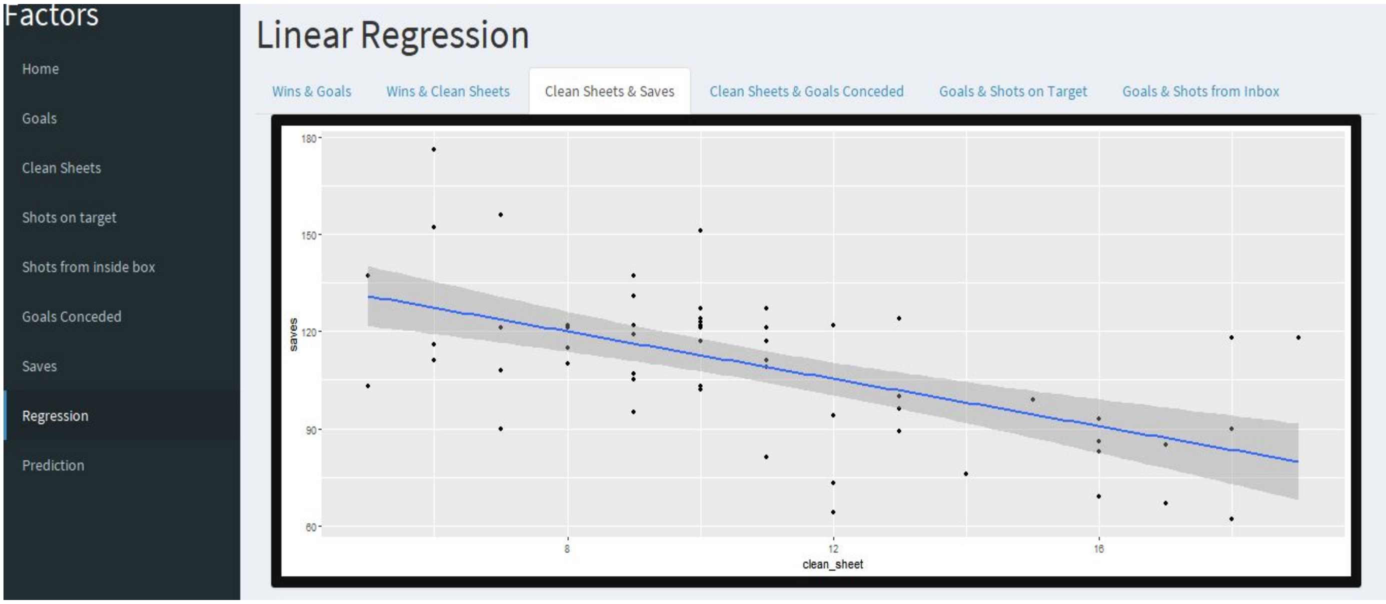


Summary for - Wins vs Clean Sheets

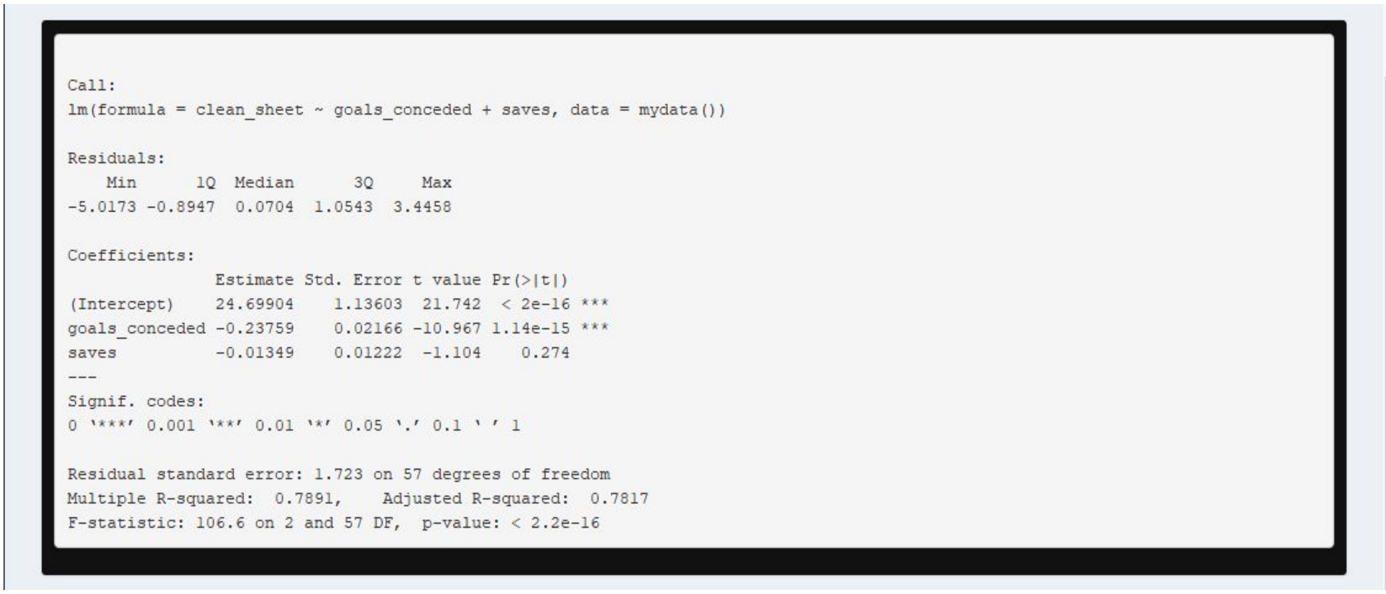


```
Call:
lm(formula = wins ~ goals + clean_sheet, data = mydata())

Residuals:
   Min      1Q Median      3Q     Max
-4.485 -1.418  0.135  1.366  4.862

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.70524    0.85020  -5.534 8.22e-07 ***
goals        0.24046    0.01982  12.131  < 2e-16 ***
clean_sheet  0.58772    0.09447   6.221 6.23e-08 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.941 on 57 degrees of freedom
Multiple R-squared:  0.9062,    Adjusted R-squared:  0.9029
F-statistic: 275.4 on 2 and 57 DF,  p-value: < 2.2e-16
```

3. Graph for - Clean Sheets vs Saves

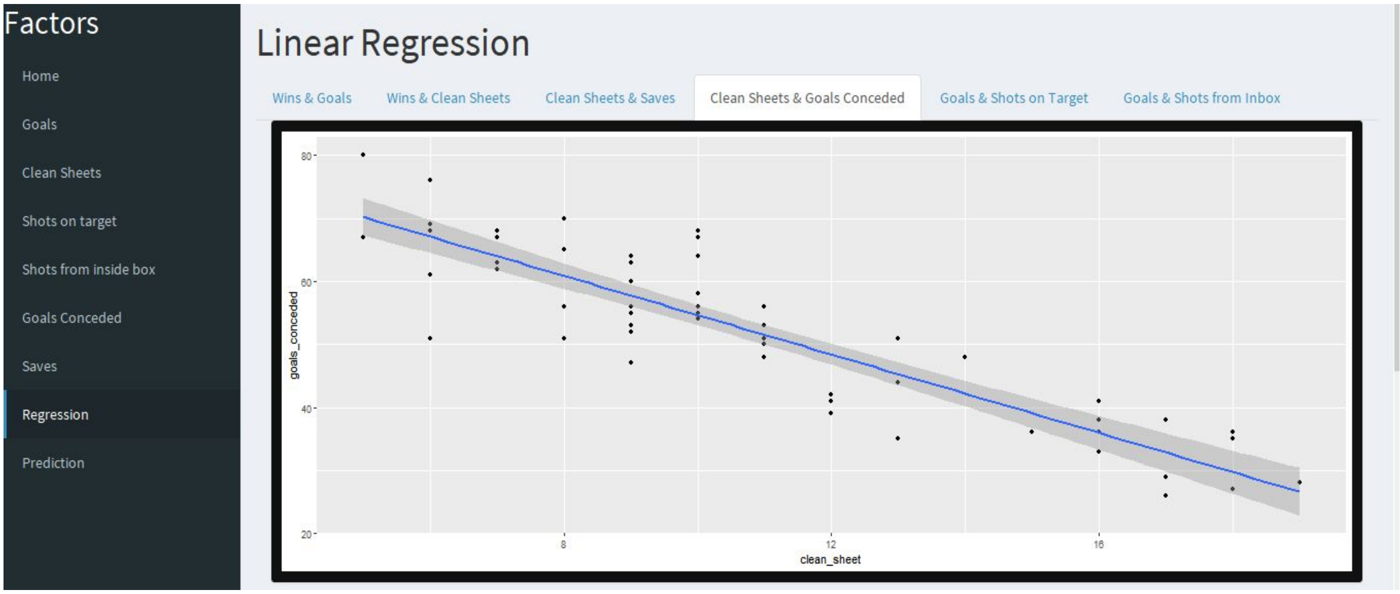


Summary for - Clean Sheets vs Saves

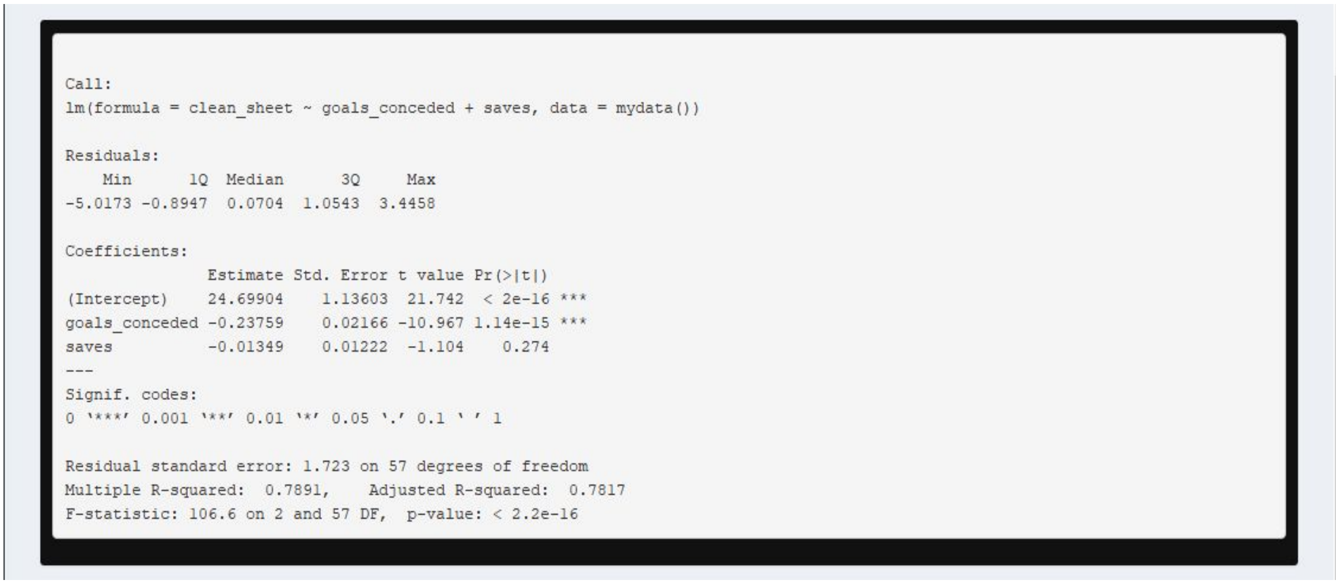


```
Call:
lm(formula = clean_sheet ~ goals_conceded + saves, data = mydata())

Residuals:
    Min      1Q  Median      3Q     Max 
-5.0173 -0.8947  0.0704  1.0543  3.4458 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)    24.69904    1.13603  21.742  < 2e-16 ***
goals_conceded -0.23759    0.02166 -10.967 1.14e-15 ***
saves          -0.01349    0.01222  -1.104    0.274    
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.723 on 57 degrees of freedom
Multiple R-squared:  0.7891,	Adjusted R-squared:  0.7817 
F-statistic: 106.6 on 2 and 57 DF,  p-value: < 2.2e-16
```

4. Graph for - Clean Sheets vs Goals Conceded

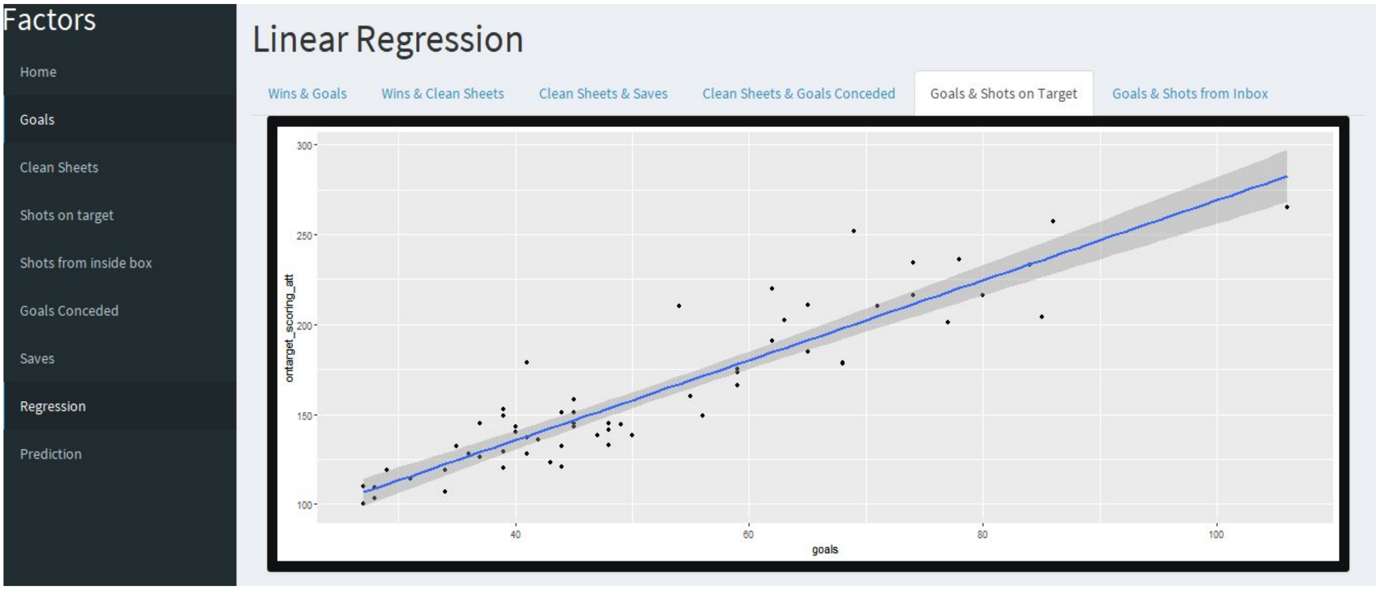


Summary for - Clean Sheets vs Goals Conceded

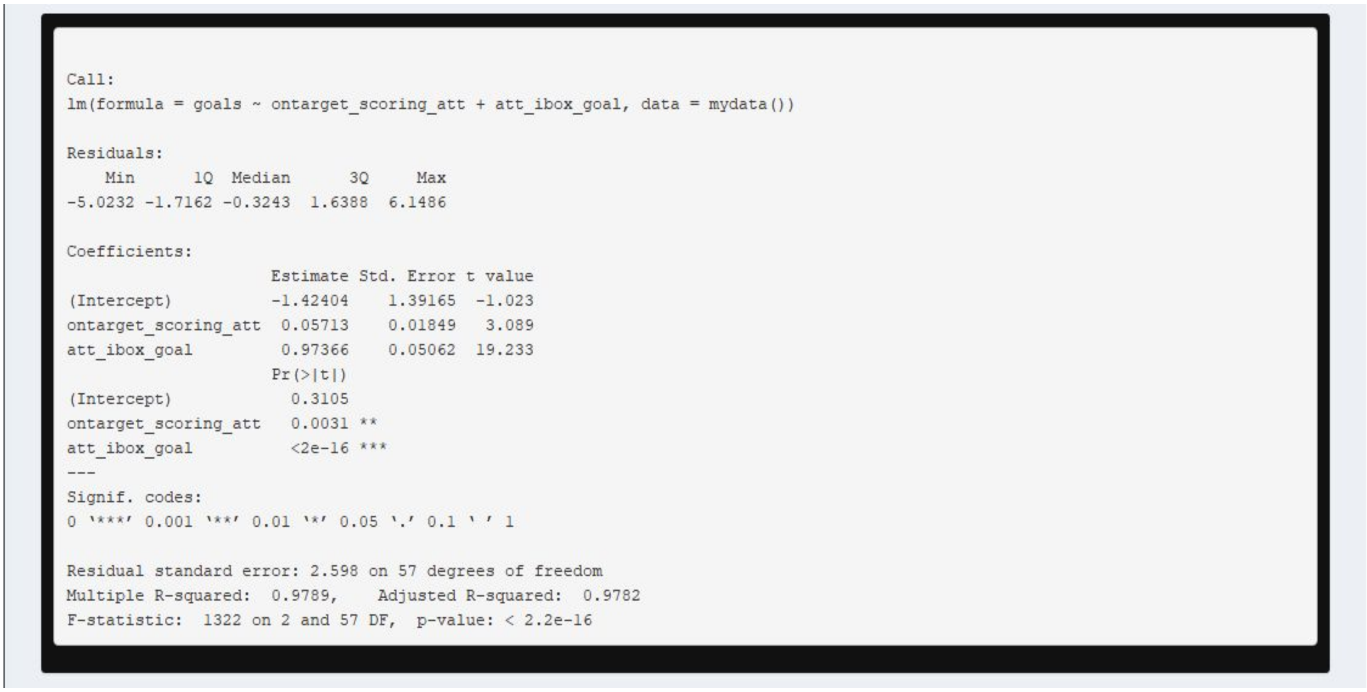


```
Call:
lm(formula = clean_sheet ~ goals_conceded + saves, data = mydata())

Residuals:
    Min      1Q  Median      3Q     Max
-5.0173 -0.8947  0.0704  1.0543  3.4458

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       24.69904    1.13603  21.742  < 2e-16 ***
goals_conceded -0.23759    0.02166 -10.967 1.14e-15 ***
saves             -0.01349    0.01222  -1.104    0.274
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.723 on 57 degrees of freedom
Multiple R-squared:  0.7891,    Adjusted R-squared:  0.7817
F-statistic: 106.6 on 2 and 57 DF,  p-value: < 2.2e-16
```

5. Graph for - Goals vs Shots on Target

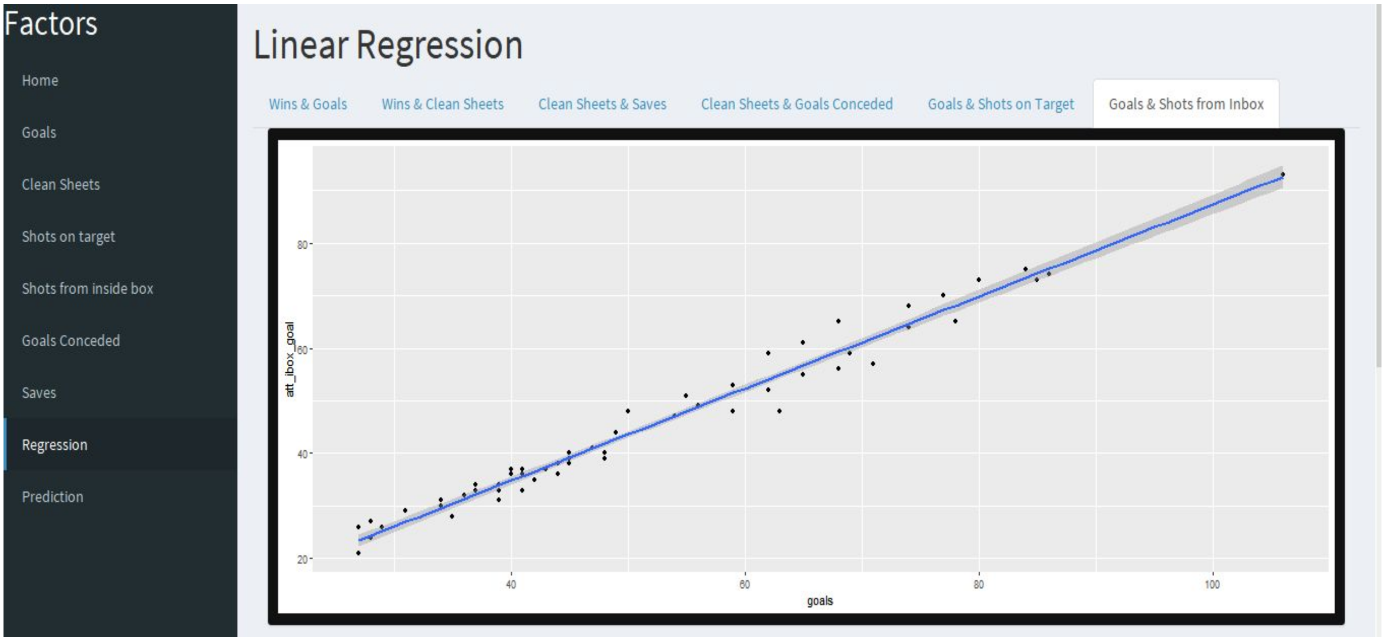


Summary for - Goals vs Shots on Target

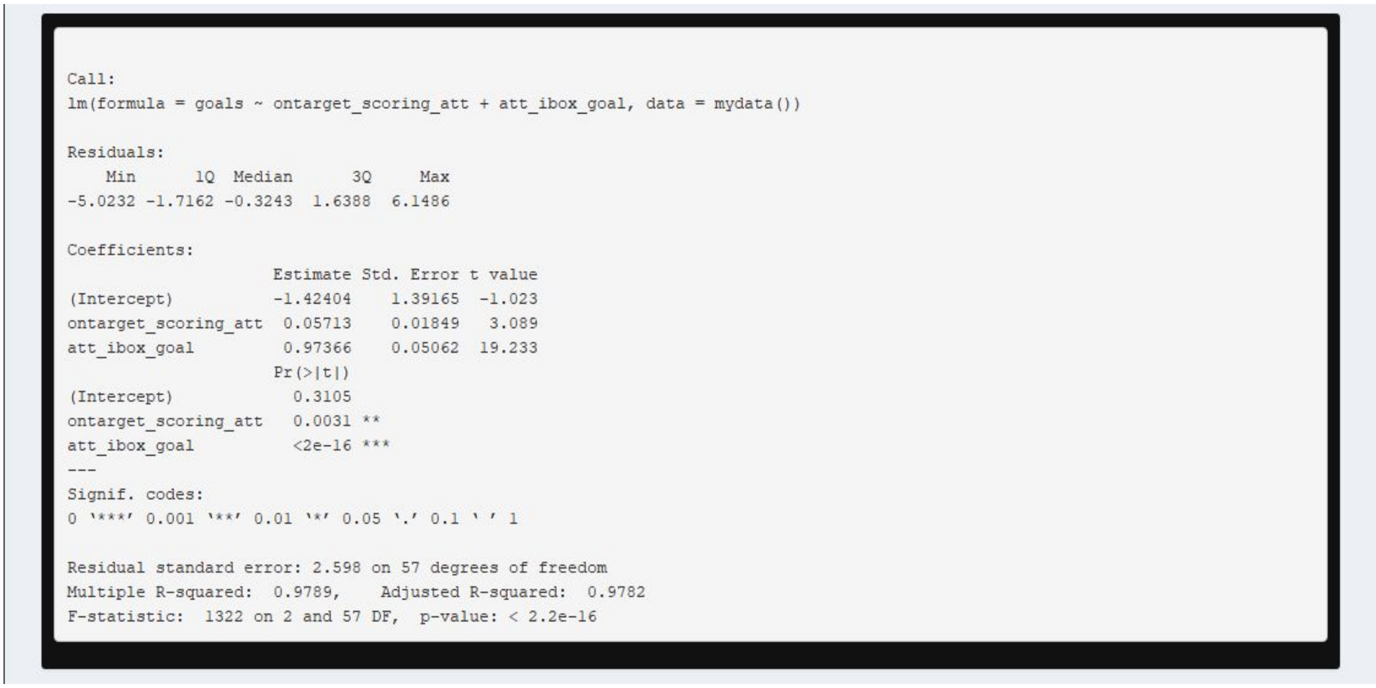


```
Call:
lm(formula = goals ~ ontarget_scoring_att + att_ibox_goal, data = mydata())

Residuals:
    Min      1Q  Median      3Q     Max 
-5.0232 -1.7162 -0.3243  1.6388  6.1486 

Coefficients:
                     Estimate Std. Error t value
(Intercept)          -1.42404    1.39165  -1.023
ontarget_scoring_att  0.05713    0.01849   3.089
att_ibox_goal         0.97366    0.05062  19.233
                     Pr(>|t|)    
(Intercept)            0.3105    
ontarget_scoring_att   0.0031 ** 
att_ibox_goal          <2e-16 ***
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.598 on 57 degrees of freedom
Multiple R-squared:  0.9789,	Adjusted R-squared:  0.9782 
F-statistic:  1322 on 2 and 57 DF,  p-value: < 2.2e-16
```

6. Graph for - Goals vs Shots from inside box

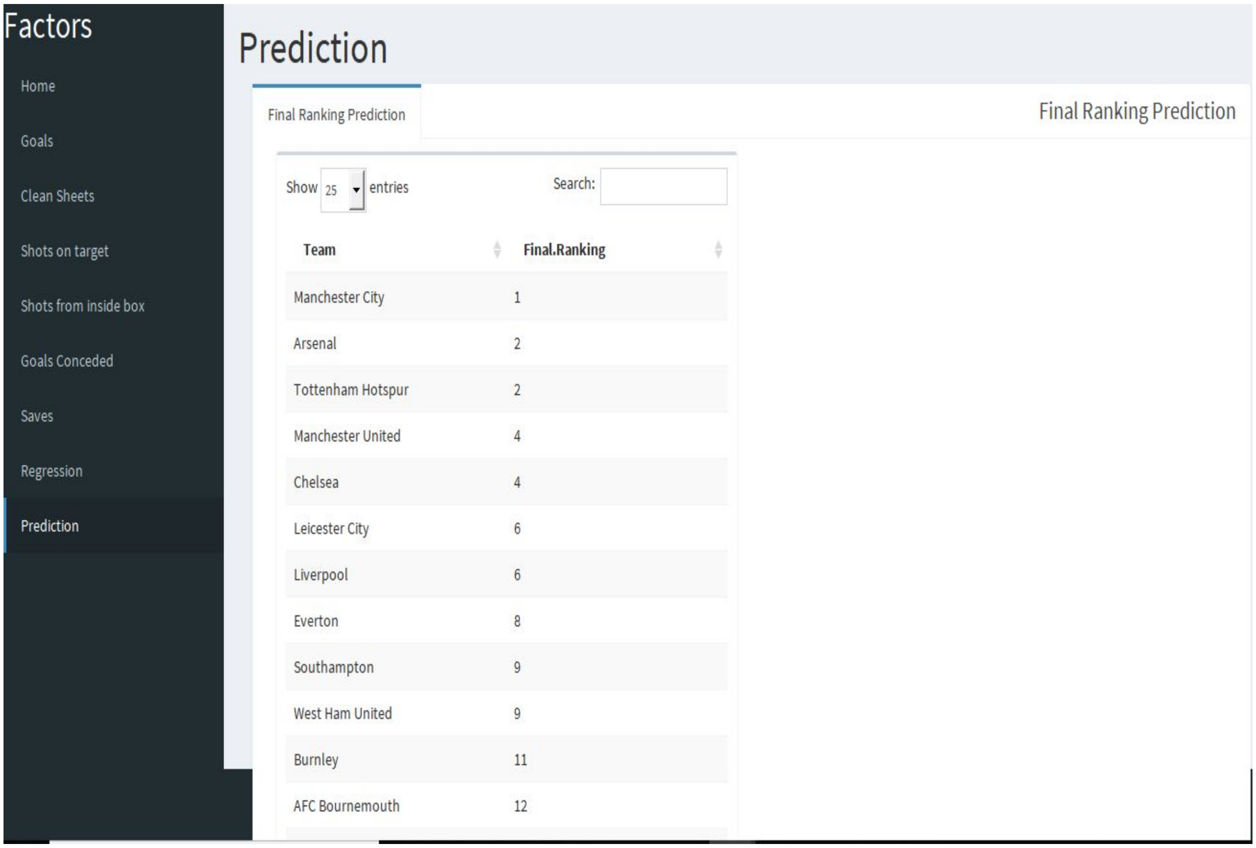


Summary for - Goals vs Shots from inside box



```
Call:
lm(formula = goals ~ ontarget_scoring_att + att_ibox_goal, data = mydata())

Residuals:
    Min      1Q  Median      3Q     Max 
-5.0232 -1.7162 -0.3243  1.6388  6.1486 

Coefficients:
                     Estimate Std. Error t value
(Intercept)          -1.42404    1.39165  -1.023
ontarget_scoring_att  0.05713    0.01849   3.089
att_ibox_goal         0.97366    0.05062  19.233
                     Pr(>|t|)    
(Intercept)            0.3105    
ontarget_scoring_att   0.0031 ** 
att_ibox_goal          <2e-16 ***
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.598 on 57 degrees of freedom
Multiple R-squared:  0.9789,    Adjusted R-squared:  0.9782 
F-statistic:  1322 on 2 and 57 DF,  p-value: < 2.2e-16
```

# FINAL PREDICTED STANDINGS :-



Prediction

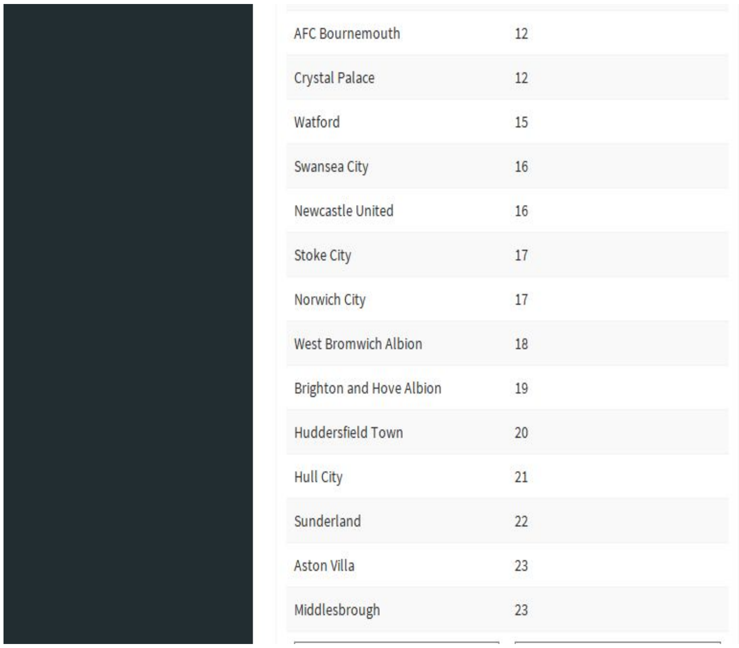
Final Ranking Prediction

Final Ranking Prediction

Show 25 entries

Search:

| Team | Final.Ranking |
|---|---|
| Manchester City | 1 |
| Arsenal | 2 |
| Tottenham Hotspur | 2 |
| Manchester United | 4 |
| Chelsea | 4 |
| Leicester City | 6 |
| Liverpool | 6 |
| Everton | 8 |
| Southampton | 9 |
| West Ham United | 9 |
| Burnley | 11 |
| AFC Bournemouth | 12 |



| Team | Final.Ranking |
|---|---|
| AFC Bournemouth | 12 |
| Crystal Palace | 12 |
| Watford | 15 |
| Swansea City | 16 |
| Newcastle United | 16 |
| Stoke City | 17 |
| Norwich City | 17 |
| West Bromwich Albion | 18 |
| Brighton and Hove Albion | 19 |
| Huddersfield Town | 20 |
| Hull City | 21 |
| Sunderland | 22 |
| Aston Villa | 23 |
| Middlesbrough | 23 |

## CONCLUSIONS

1) The 3 Linear Regression models has R-Squared value of 0.9062, 0.7891 and 0.9789.
2) This models fairly predicts the result of English Premier League for next season.
3) The confidence level for the model is fairly low.
4) The model works satisfactorily.

## FUTURE SCOPE AND DEVELOPMENT

1. Take into consideration players of the team and their performance. Sentimental Analysis can be used to predict whether the player's attitude will affect the overall team's performance.
2. Take into consideration the result of each and every match played between all the teams and then predict in the basis of the Home or Away wins.
3. Take into consideration the list of  impact players who got transferred from one team to another which can affect the team's result relative to previous season.

## REFERENCES AND APPENDIX

### REFERENCES :-

1. **Stern Hal. (1995) *Who's Number 1 in College Football?...And How Might We Decide?*Chance, Summer, 7-14.**

2. **^ Moroney M. J. (1956) *Facts from figures*. 3rd edition, Penguin, London.**

3. **^ Reep C. Benjamin B. (1968) *Skill and chance in association football*. Journal of the Royal Statistical Society, Series A, 131, 581-585.**

4. **^ Hill I.D. (1974), *Association football and statistical inference*. Applied statistics, 23, 203-208.**

5. [a b c d] **Maher M.J. (1982), *Modelling Association Football scores*. Statistica Neerlandica, 36, 109-118**

6. **^ Caurneya K.S. and Carron A.V. (1992) *The home advantage in sports competitions: a literature review*. Journal of Sport and Exercise Physiology, 14, 13-27.**

7. **^ Knorr-Held, Leonhard (1997) *Dynamic Rating of Sports Teams*. (REVISED 1999). Collaborative Research Center 386, Discussion Paper 98**

8. [a b] **Diego Kuonen (1996) *Statistical Models for Knock-out Soccer Tournaments***

9. **^ Lee A. J. (1997) *Modeling scores in Premier League: is Manchester United really the best*. Chance, 10, 15-19**

10. *a b c d* **Mark J. Dixon and Coles S.G. (1997) *Modeling Association Football Scores and Inefficiencies in the Football Betting Market*, Applied Statistics, Volume 46, Issue 2, 265-280**

11. **^ Dimitris Karlis and Ioannis Ntzoufras (2007) *Bayesian modelling of football outcomes: Using the Skellam's distribution for the goal difference***

12. **^ Rue H. and Salvesen Ø. (1999) *Predicting and retrospective analysis of soccer matches in a league*. Technical Report. Norwegian University of Science and Technology, Trondheim.**

13. **^ Marek, Patrice; Šedivá, Blanka; Ťoupal, Tomáš (2014). "Modeling and prediction of ice hockey match results". *Journal of Quantitative Analysis in Sports*. 10: 357–365. doi:10.1515/jqas-2013-0129. ISSN 1559-0410 – via Research Gate.**

14. **^ Famoye, F (2010). "A new bivariate generalised Poisson distribution". *Statistica Neerlandica*. 64: 112–124.**

## APPENDIX :-

***Multiple Linear Regression***

Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the linear relationship between the explanatory (independent) variables and response (dependent) variable.

In essence, multiple regression is the extension of ordinary least-squares (OLS) regression that involves more than one explanatory variable.

The Formula for Multiple linear regression is :-

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \epsilon$$

*where, for i=n observations:*

$Y_i$ = ***dependent variable***

$x_i$ = **explanatory variables**

$B_0$ = **y-intercept (constant term)**

$B_p$ = **slope coefficients for each explanatory variable**

$\epsilon$ = **the model's error term (also known as the residuals)**

The multiple regression model is based on the following assumptions:

- There is a linear relationship between the dependent variables and the independent variables.
- The independent variables are not too highly correlated with each other.
- $y_i$ observations are selected independently and randomly from the population.
- Residuals should be normally distributed with a mean of 0 and variance *σ.*

The coefficient of determination (R-squared) is a statistical metric that is used to measure how much of the variation in outcome can be explained by the variation in the independent variables. $R^2$ always increases as more predictors are added to the MLR model even though the predictors may not be related to the outcome variable.

$R^2$ by itself can't thus be used to identify which predictors should be included in a model and which should be excluded. $R^2$ can only be between 0 and 1, where 0 indicates that the outcome cannot be predicted by any of the independent variables and 1 indicates that the outcome can be predicted without error from the independent variables.

When interpreting the results of a multiple regression, beta coefficients are valid while holding all other variables constant ("all else equal"). The output from a multiple regression can be displayed horizontally as an equation, or vertically in table form.

## RANKING ALGORITHM :-

**Learning to rank** or **machine-learned ranking** (MLR) is the application of machine learning, typically supervised, semi-supervised or reinforcement learning, in the construction of ranking models for information retrieval systems.[2] Training data consists of lists of items with some partial order specified between items in each list. This order is typically induced by giving a numerical or ordinal score or a binary judgment (e.g. "relevant" or "not relevant") for each item. The ranking model's purpose is to rank, i.e. produce a permutation of items in new, unseen lists in a way which is "similar" to rankings in the training data in some sense.

Learning to rank algorithms have been applied in areas other than information retrieval:

- In machine translation for ranking a set of hypothesized translations;
- In computational biology for ranking candidate 3-D structures in protein structure prediction problem.
- In recommender systems for identifying a ranked list of related news articles to recommend to a user after he or she has read a current news article.
- In software engineering, learning-to-rank methods have been used for fault localization.

## INDEX AND ACRONYMS