INTRODUCTION TO DATA MANAGEMENT

PROJECT REPORT

PREMIER LEAGUE STATISTICS

BY

Srikar Hasthi

2019-2020

Acknowledgement

I am grateful to my instructor Ms. Savleen Kaur Mam. She has truly been a great source of inspiration, constructive criticism, insight and input. The knowledge She has imparted and the patience they have displayed was vital in completing this study. I am thankful for the immense encouragement and morale support that She graciously provided during this study.

I owe much of my academic and personal success to my parents, who, by example, provided me with the motivation and courage to pursue this course in the field of my interest. Special thanks to all my friends, near and far, for their love and support that made me finish my End term project and its report successfully.

Table Of Contents

1. Introduction

The Premier League, often referred to as the English Premier League or the EPL outside England, is the top level of the English football league system. Contested by 20 clubs, it operates on a system of promotion and relegation with the English Football League (EFL).
The Premier League is a corporation in which the member clubs act as shareholders. Seasons run from August to May with each team playing 38 matches.
The Premier League is the most-watched sports league in the world, broadcast in 212 territories to 643 million homes and a potential TV audience of 4.7 billion people.
Here we the data set of the Premier League matches played from 2006-2007 to 2017-2018. It contains the information of all the teams that played against each other and their results. The data set has multiple sheets with each sheet giving insights to different statistics of the team.
In this data set we also have the list of teams in the premier league from 2006 to 2017 with different columns representing their attributes like the result, goals scored, wins, losses etc.
Here are some of the important columns in the Dataset :-

1. **Home Team**- The team which is playing at their home ground.
2. **Away Team**- The visiting team that plays at the ground of the home team.
3. **Home Goals**- Goals scored by the home team.
4. **Away Goals**- Goals scored by the away team.
5. **Result**- The result of the match (Won, Drawn, Lost).
6. **Season-** The year in which the match was held.
7. **Name-** Name of the player in the premier league.
8. **Played-** The number of matches that player played in that season.
9. **Wins-** The number of matches a team won in that season.
10. **Losses-** The number of matches a team lost in that season.
11. **Goals-** The total number of goals a team scored in that season.

2. Objectives

Analysis is done on following Objectives:

1. Estimating the probability of a team winning, losing or drawing the match.
2. Creating the points table of the season.
3. Top goal scorers of the season.
4. Total number of clean sheets each team has during the season.
5. Estimating the number of goals a team scores in each match.

3. Source Of Dataset

The data is collected from the Kaggle website and is transformed according to my need. Though the data collected was not sufficient to full fill the need of all mentioned goals, the data collected are data from other sources like the premier league and the espn website are added into this file. All the data were available in csv format and transformed into excel file by using ETL process.

Link: https://www.kaggle.com/zaeemnalla/premier-league

4. ETL (Extraction Transform and Load) Process

The data is extracted from the kaggle.com. The format of the data available on the website was csv and the data is converted into Excel format to perform analysis. The data was from 2006 to 2017. After the extraction of the data I checked if any null values were present in the data and when there was none I decided to load the data. The identifiers were made according to the need and representation. After extraction, transformed data is loaded into MS Excel and analysis was performed.

5. Analysis On The Dataset

5.2 Match Predictor

1. Introduction

When a match takes place there is a possibility of three outcomes- Win, lose or draw. The result may depend on various factors including the ground in which they are playing. Based on the previous encounters and results of the teams a graph is drawn which tells the probability of the end result i.e win, lose or draw.

2. General Description

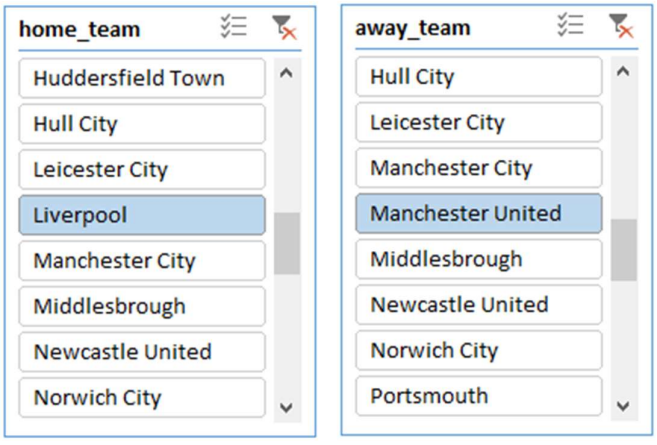
My objective is to identify the probability if a team will win lose or draw the match. Here I created a pivot table with home team and away team as a row and places result in the column as well as in values and the season in the filter. I used slicer to select the home and away team. I used the bar chart to show the probability of the result.

3. Specific Requirements, Function and Formulas:

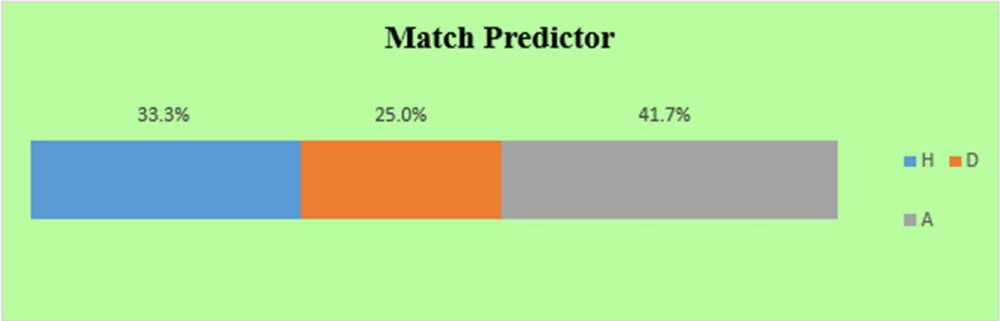
The analysis requires Microsoft excel 2013 or above, requires pivot table and different type of graphs. The function and formula of Count and sum is used for calculation of the number of matches played, won, lost and drawn.

4. Analysis Result

Count of result		result			
home_team	away_team	H	D	A	Grand Total
Liverpool	Manchester United	4	3	5	12



5. **Visualization**



The Probability of home team(Blue) winning is 33.3%

The Probability of the match ending in draw(Orange) is 25.0%

The Probability of away team(Grey) winning is 41.7%

5.3 Creating the points table

1. **Introduction**

   Each season the premier league title is won by whoever is on the top of the table. There is a total of 38 matches played by each team and 3 points is awarded when a match is won, 1 point when the match is drawn and 0 points when they lose the match. A graph is drawn based on the number of matches won, lost and drawn.

2. **General Description**

   I created a pivot table where I places team in rows, season in filter and win, lose, draw in values. I used calculated field to create a points column which calculated the total points a team had at the end of the season. Then I sorted the data from maximum to minimum to see which team won in that season.

3. **Specific Requirements, Function and Formulas:**

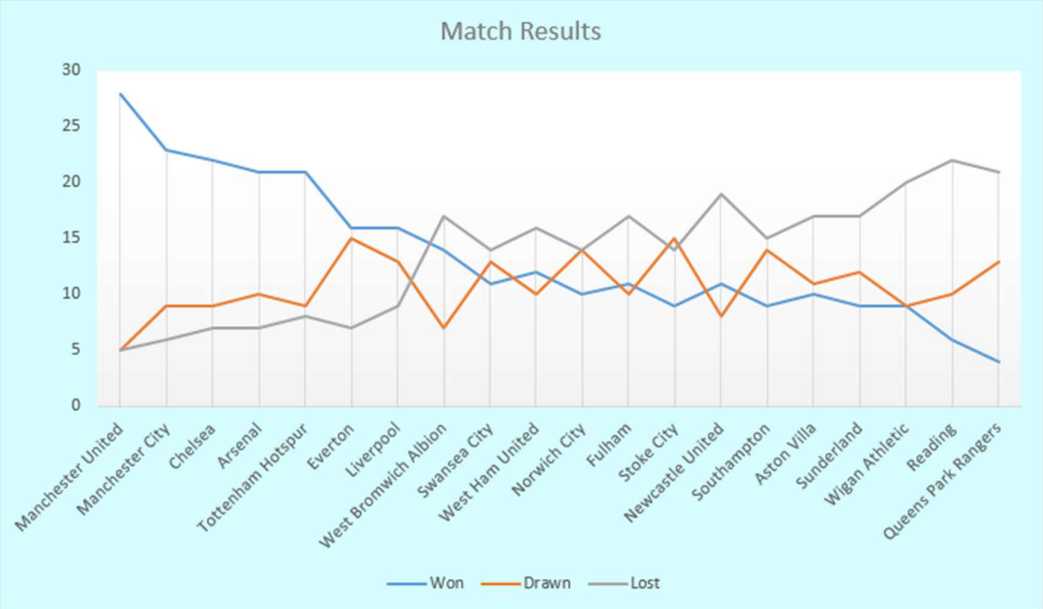
   The analysis requires Microsoft excel 2013 or above, requires pivot table and different type of graphs. The function and formula of sum and calculated field is used for calculation of the total points.

4. Analysis Result

season	2012-2013

Row Labels	Played	Won	Drawn	Lost	Sum of Points
Manchester United	38	28	5	5	89
Manchester City	38	23	9	6	78
Chelsea	38	22	9	7	75
Arsenal	38	21	10	7	73
Tottenham Hotspur	38	21	9	8	72
Everton	38	16	15	7	63
Liverpool	38	16	13	9	61
West Bromwich Albion	38	14	7	17	49
Swansea City	38	11	13	14	46
West Ham United	38	12	10	16	46
Norwich City	38	10	14	14	44
Fulham	38	11	10	17	43
Stoke City	38	9	15	14	42
Newcastle United	38	11	8	19	41
Southampton	38	9	14	15	41
Aston Villa	38	10	11	17	41
Sunderland	38	9	12	17	39
Wigan Athletic	38	9	9	20	36
Reading	38	6	10	22	28
Queens Park Rangers	38	4	13	21	25

5. Visualization



5.4 Top Goal Scorers

1. Introduction

In each season of the premier league there are 38 matches played by each team. Each team have strikers who help their team score goals. A striker is judged based on the number of goals he scored. And every year the top goal scorer is presented with the prestigious golden boot of premier league.

2. General Description

Here I created a pivot table where I placed name in rows, season in filter and goals in values. Then I sorted them in descending order so as to determine the top goal scorer. I then applied a filter to show only top 15 goal scorers. The extracted data is then used to create a column graph which shows the number of goals scored by the top players.

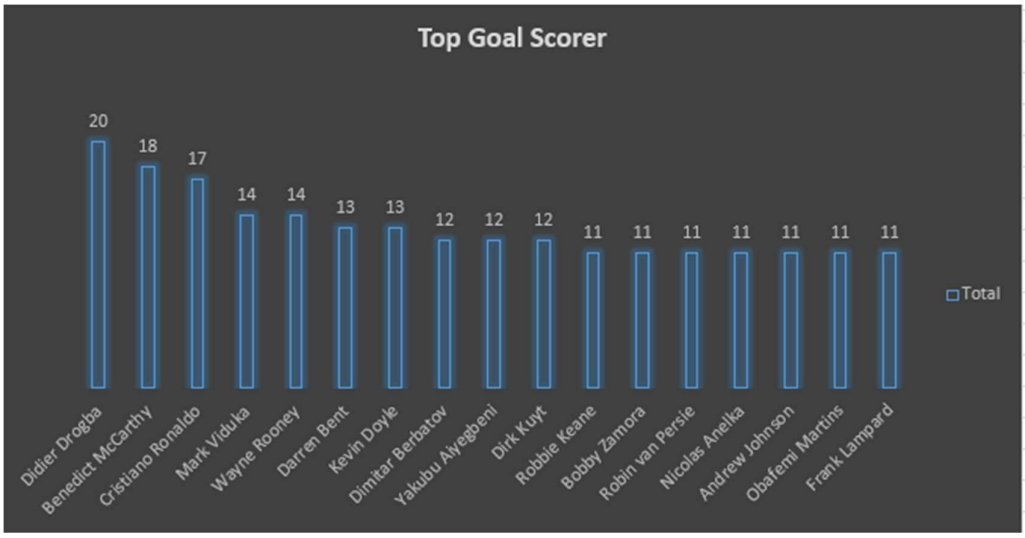
3. Specific Requirements, Function and Formulas:

The analysis requires Microsoft excel 2013 or above, requires pivot table and different type of graphs. The function and formula of sum is used for calculation of the number of goals. And a column graph is used.

4. Analysis Result

Season	2006-2007
Row Labels	**Sum of Goals**
Didier Drogba	20
Benedict McCarthy	18
Cristiano Ronaldo	17
Mark Viduka	14
Wayne Rooney	14
Darren Bent	13
Kevin Doyle	13
Dimitar Berbatov	12
Yakubu Aiyegbeni	12
Dirk Kuyt	12
Robbie Keane	11
Bobby Zamora	11
Robin van Persie	11
Nicolas Anelka	11
Andrew Johnson	11
Obafemi Martins	11
Frank Lampard	11

5. **Visualization**



5.5 Number Of Clean Sheets

1. **Introduction**

   If scoring goals is important for a team to win the match then keeping a clean sheet is also important. A team is given a clean sheet when they do not concede even a single goal. With this the goalkeeper and the team's defense is evaluated. The team with the most number of clean sheets has the strongest defence or the greatest goalkeeper or both.

2. **General Description**

   Here I created pivot table where I have placed team in rows, season in filter and clean sheets in values. Then I sorted the data in descending order to know which team has the most number of clean sheets. I then used the line chart to graphically determine the top teams with most number of clean sheets.

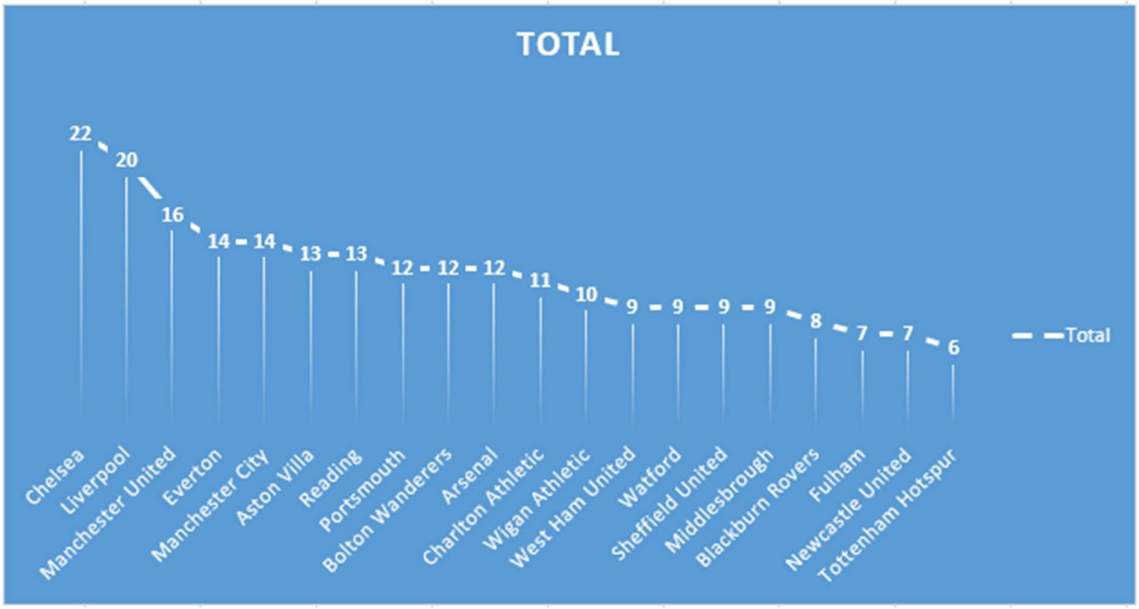
3. **Specific Requirements, Function and Formulas:**

   The analysis requires Microsoft excel 2013 or above, requires pivot table and different type of graphs. The function and formula of sum and sort is used for calculation of the number of clean sheets and sorting the data. And a column graph is used.

4. Analysis Result

season	2006-2007
Row Labels	**Sum of clean_sheet**
Chelsea	22
Liverpool	20
Manchester United	16
Everton	14
Manchester City	14
Aston Villa	13
Reading	13
Portsmouth	12
Bolton Wanderers	12
Arsenal	12
Charlton Athletic	11
Wigan Athletic	10
West Ham United	9
Watford	9
Sheffield United	9
Middlesbrough	9
Blackburn Rovers	8
Fulham	7
Newcastle United	7
Tottenham Hotspur	6

5. Visualization



5.6 Estimating The Goals Scored In A Match

1. Introduction

In every match a team tries to score maximum number of goals. To determine the average number of goals a team scores in each match I divided the total number of goals a team scored in that season with the number of games played that is 38. This analysis helps to determine if a team is attacking team or a defensive team.

2. General Description

Here I created pivot table where I have placed team in rows, season in filter and goals in values. Then I used calculated field to calculate the average number of goals a team scores each match. Then I sorted the data in descending order to know which team scores more goals in a match on average. I then used the area chart to graphically determine the top teams with most goals scored per game.

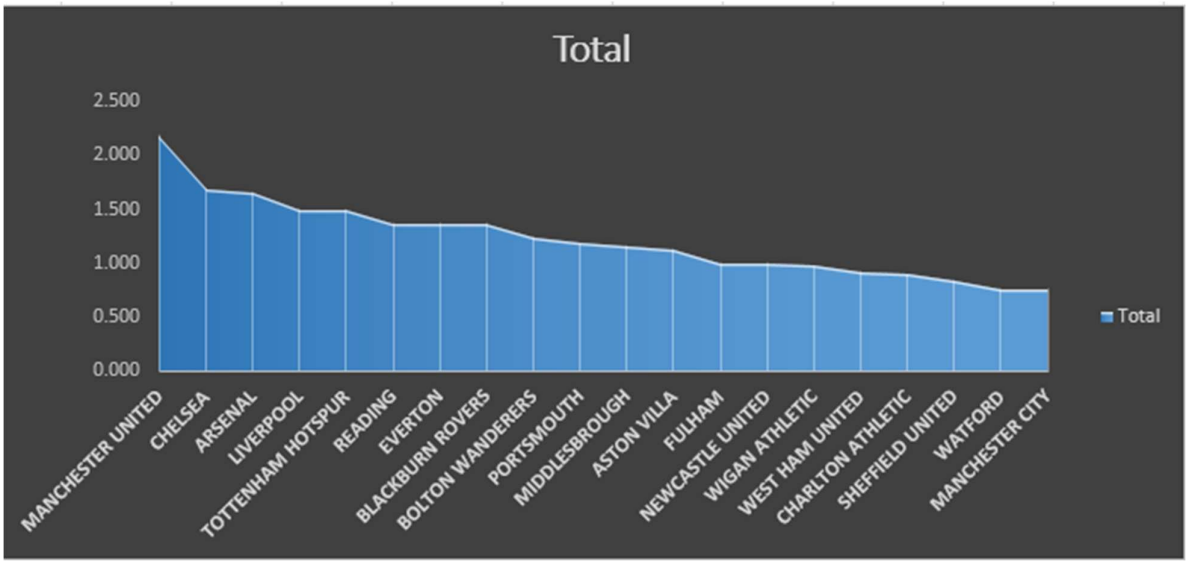
3. Specific Requirements, Function and Formulas:

The analysis requires Microsoft excel 2013 or above, requires pivot table and different type of graphs. The function and formula of sum and calculated field is used for calculation of the average goals.

4. Analysis Result

season	2006-2007
Row Labels	**Sum of Estgoals**
Manchester United	2.184
Chelsea	1.684
Arsenal	1.658
Liverpool	1.500
Tottenham Hotspur	1.500
Reading	1.368
Everton	1.368
Blackburn Rovers	1.368
Bolton Wanderers	1.237
Portsmouth	1.184
Middlesbrough	1.158
Aston Villa	1.132
Fulham	1.000
Newcastle United	1.000
Wigan Athletic	0.974
West Ham United	0.921
Charlton Athletic	0.895
Sheffield United	0.842
Watford	0.763
Manchester City	0.763

5. Visualization



6. List Of Analysis

1. The Probability of a team winning, losing or drawing the match is analysed and a bar graph is drawn. Now the team can see their performance history against that team and prepare better.
2. The points table of each season is created. The manager and the player can see where their team stand in the league and aim for the number one spot.
3. The player with most number of goals is determined. This helps the players to visualize the competition among each other.
4. The team with most number of clean sheets is determined and a line graph is drawn. This will help the manager to see whether he needs to improve the defense or not.
5. The average number of goals each team score in a season is determined.

7. References

https://www.kaggle.com/zaeemnalla/premier-league

https://www.espn.com/soccer/stats/_/league/eng.1

8. Bibliography

https://www.premierleague.com/stats

https://www.google.com/search?q=premierleague&client=firefox-b-d&sxsrf=ACYBGNQtDvlpWzzabgZvOmCoPqXUXgHSgg:1574012372393&source=lnms&tbm=isch&sa=X&ved=0ahUKEwicmNCC5fHlAhXozjgGHUAjBVAQ_AUIEygD

https://www.espn.com/soccer/stats/_/league/eng.1