

DATA MANAGEMENT PROJECT REPORT

(Project Semester: September-December 2021)



L OVELY
P ROFESSIONAL
U NIVERSITY

Premier League 2019-2020 DATA ANALYSIS

Submitted by: Sanjay Chaudhary
Reg. no : 11913399

Program : B.Tech (CSE)

Section: KM011

Course Code: INT217

Under the Guidance of

Madhuri - 23525

Lovely Professional University, Phagwara

DECLARATION

I, Sanjay Chaudhary, student of B.Tech CSE at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12-Dec-2021

Registration No.: 11913399

Sanjay Chaudhary

ACKNOWLEDGEMENT

I have taken efforts in this course. However, it would not have been possible without the kind support and help of many individuals and University. I would like to extend my sincere tanks to all of them.

I am highly indebted to Ms. Madhuri Ma'am for her guidance and constant supervision as well as for providing necessary information for the course & also for her support in completing the course.

I would like to express my special gratitude and thanks to my faculty for giving me such attention and time.

My thanks and appreciations also go to my batchmates and people who have willingly helped me out with their abilities.

TABLE OF CONTENT

1.Introduction

2. Scope of the Analysis

3.Source of dataset

4.ETL process

5.Analysis on dataset (for each analysis)

- i. Introduction
- ii. General Description
- iii. Specific Requirements, functions, and formulas
- iv. Analysis results
- v. Visualization

6.List of Analysis with results

7.References

8.Bibliography

INTRODUCTION

Data Analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, while being used in different business, science, and social science domains.

Premier League, also called Premiership, English professional football (soccer) league established in 1992. The league, which comprises 20 clubs, superseded the first division of the English Football League (EFL) as the top level of football in England.

During a Premier League season, each club plays one home and one away match with every other team in the league. A match victory gives the winning team three points in the standings, while a draw results in one point for each club. The league has no postseason tournament: the team with the most

points at the end of the season are the Premier League champion. The top four point-scoring teams each year qualify for the next season's Union of European Football Associations (UEFA) Champions League tournament, which pits the most successful domestic teams in European football against each other for the annual title of best European club. Meanwhile, the bottom three clubs of the Premier League are relegated (dropped), and the top three finishers of first division (EFL Championship) teams of the EFL are promoted to the Premier League.

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. For the 2018–19 season, the average Premier League match attendance was at 38,181, second to the German Bundesliga's 43,500, while aggregated attendance across all matches is the highest of any association football league at 14,508,981. Most stadium occupancies are near capacity.[The Premier League ranks first in the UEFA coefficients

of leagues based on performances in European competitions over the past five seasons as of 2021. The English top-flight has produced the second-highest number of UEFA Champions League/European Cup titles, with five English clubs having won fourteen European trophies in total.

Prize Money:

The 2020 season of the premier league offered prize to winning team is 38 million pounds.

The premier league's 20 club are expected to receive 2 million less than upper position club.

Teams set to receive :

1st - Man City - £38m

2nd - Man Utd - £36m

3rd - Liverpool - £34m

4th - Chelsea - £32m

5th - Leicester - £30m
6th - West Ham - £28m
7th - Tottenham - £27m
8th - Arsenal - £24m
9th - Leeds - £23m
10th - Everton - £21m
11th - Aston Villa - £19m
12th - Newcastle - £17m
13th - Wolves - £15m
14th - Crystal Palace - £13m
15th - Southampton - £11m
16th - Brighton - £10m
17th - Burnley - £8.5m
18th - Fulham - £6m
19th - West Brom - £4m
20th - Sheffield United - £2.5m

Rules of Football (Soccer):

1. A match consists of two 45 minutes halves with a 15-minute rest period in between.
2. Each team can have a minimum of 11 players (including 1 goalkeeper who is the only player allowed to handle the ball within the 18-yard box) and a minimum of 7 players are needed to constitute a match.
3. The field must be made of either artificial or natural grass. The size of pitches is allowed to vary but must be within 100-130 yards long and 50-100 yards wide. The pitch must also be marked with a rectangular shape around the outside showing out of bounds, two six-yard boxes, two 18-yard boxes and a center circle. A spot for a penalty placed 12 yards out of both goals and center circle must also be visible.
4. The ball must have a circumference of 58-61cm and be of a circular shape.
5. Each team can name up to 7 substitute players. Substitutions can be made at any time of the match with each team being able to make a maximum of 3 substitutions per side. In the event of all three substitutes being made and a player having to leave the field for injury the team will be forced to play without a replacement for that player.
6. Each game must include one referee and two assistant referees' (linesmen). It's the job of the referee to act as timekeeper and make any decisions which may need to be made such as fouls, free kicks, throw ins, penalties and added on time at the end of each half. The referee may consult the assistant referees at any time in the match regarding a decision. It's the assistant

referee's job to spot offside's in the match (see below), throw ins for either team or also assist the referee in all decision-making processes where appropriate.

7. If the game needs to head to extra time because of both teams being level in a match, then 30 minutes will be added in the form of two 15-minute halves after the allotted 90 minutes.
8. If teams are still level after extra time, then a penalty shootout must take place.
9. The whole ball must cross the goal line for it to constitute as a goal.
10. For fouls committed a player could receive either a yellow or red card depending on the severity of the foul; this comes down to the referee's discretion. The yellow is a warning, and a red card is a dismissal of that player. Two yellow cards will equal one red. Once a player is sent off then they cannot be replaced.
11. If a ball goes out of play off an opponent in either of the side lines, then it is given as a throw in. If it goes out of play off an attacking player on the base line, then it is a goal kick. If it comes off a defending player, it is a corner kick.

Insights of data:

PREMIER LEAGUE ANALYSIS contains the following data fields: -

- Match_Name: Names of teams between match had been played.
- Match_Date: Date of Match
- Team: Name of team
- Team_logo: URL of team logo
- Manager: Manager name of the team
- Captain: Captain name of the team
- Score: Goals scored by the team
- Penalties: Penalties scored by the team
- Fouls: Total Fouls by the team
- Corners: Total Corners taken by the team
- Crosses: Total Crosses given by the team
- Touches: Total touches by the players during the game
- Tackles: Total Tackles by the team
- Interceptions: Total ball intercepted by the team
- Aerials Won: Total Aerial ball won by the team
- Clearances: Ball cleared by the team
- Offsides: Total Offsides during the game
- Goal kicks: Total goal kicks taken by the team
- Throw Ins: Total Throw Ins get the team
- Long Balls: Team gives the long Aerial Passes
- YellowCard: Yellow cards received by the team
- RedCard: Red cards received by the team
- YellowRed: Player received second yellow during the same match, turns out red card
- Possession: Total Physical control of ball
- Passing accuracy: $\text{Passing accuracy}(\text{SucPass}/\text{NumofPass})$
- SucPass: Total number of successful passes
- NumofPass: Total number of passes
- ShotsAccuracy: $\text{Shot accuracy}(\text{SucShots}/\text{NumofShots})$

- SucShots: Total successful shots
- NumofShots: Total number of shots
- SavesAccuracy: Saves accuracy($\text{SucSaves}/\text{NumofSaves}$)
- SucSaves: Total Saves by the goalkeeper
- NumofSaves: Total Shot on Target

SCOPE OF ANALYSIS

This project on Premiere League Statistics of England provides the overall Statistics details of the matches of premier league 2019-2020 season.

Objectives of this project:

1. Which Manager is best at possession play throughout the season?
2. Which team's Goalkeeper is best throughout the season?
3. Which team is most disciplined in premier league football?
4. Which team is more successful in scoring?
5. Which team changes most number of Managers & Captains throughout the season?
6. Which team have highest successful passes and accuracy per game?

Aim of this project is to answer the above objectives in the form of visualization by creating a dashboard to convey the answers effectively and efficiently.

SOURCE OF DATASET

The data is being taken from the Kaggle .Kaggle is the platform for Data Scientists – this is where they spend their nights and weekends. It’s a crowd-sourced platform to attract, nurture, train, and challenge data scientists from all around the world to solve data science, machine learning and predictive analytics problems. It has over 536,000 active members from 194 countries, and it receives close to 150,000 submissions per month. Started from Melbourne, Australia Kaggle moved to Silicon Valley in 2011, raised some 11 million dollars from the likes of Hal Varian (Chief Economist at Google), Max Levchin (Paypal), Index and Khosla Ventures and then ultimately been acquired by the Google in March of 2017. Kaggle is the number one stop for data science enthusiasts all around the world who compete for prizes and boost their Kaggle rankings. There are only 94 Kaggle Grandmasters in the world to this date.

Source link: <https://www.kaggle.com/ekrembayar/premier-league-match-reports-20192020>

ETL PROCESS

In computing, extract, transform, load (ETL) is a process in database usage to prepare data for analysis, especially in data warehousing. Data extraction involves extracting data from homogeneous or heterogeneous sources, while data transformation processes data by transforming them into a proper storage format/structure for the purposes of querying and analysis; finally, data loading describes the insertion of data into the final target database such as an operational data store, a data mart, or a data warehouse. A properly designed ETL system extracts data from the source systems, enforces data quality and consistency standards, conforms data so that separate sources can be used together, and finally delivers data in a presentation-ready format so that application developers can build applications and end users can make decisions.

Precisely, ETL is defined as a process that extracts the data from different RDBMS source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system. ETL stands for Extract, Transform and Load.

Before ETL, the dataset looked like this. **This data is taken from Kaggle.**

The screenshot shows a data viewer application. On the left, a 'Navigator' pane lists 'premierLeague.xlsx [1]' with 'Sheet1' selected. The main area displays 'Sheet1' as a table with 6 columns: Match_Name, Match_Date, Team, Team_Logo, Manager, and Captain. The table contains 20 rows of match data. A message at the bottom of the table states: 'The data in the preview has been truncated due to size limits.' The bottom of the interface has a bar with 'Load', 'Transform Data', and 'Cancel' buttons.

Match_Name	Match_Date	Team	Team_Logo	Manager	Captain
Liverpool vs. Norwich City	09-08-2019 12:00:00 AM	Liverpool	https://d2p3bygnnzw9w3.cloudfront.net/req/202007313/tlogo/fb/822bd	Jürgen Klopp	Jordan Henderson
Liverpool vs. Norwich City	09-08-2019 12:00:00 AM	Norwich City	https://d2p3bygnnzw9w3.cloudfront.net/req/202007313/tlogo/fb/1c781	Daniel Farke	Grant Hanley
West Ham United vs. Manchester City	10-08-2019 12:00:00 AM	West Ham United	https://d2p3bygnnzw9w3.cloudfront.net/req/202007313/tlogo/fb/7c21e	Manuel Pellegrini	Aaron Cresswell
West Ham United vs. Manchester City	10-08-2019 12:00:00 AM	Manchester City	https://d2p3bygnnzw9w3.cloudfront.net/req/202007313/tlogo/fb/b8f40	Pep Guardiola	David Silva
Burnley vs. Southampton	10-08-2019 12:00:00 AM	Burnley	https://d2p3bygnnzw9w3.cloudfront.net/req/202007313/tlogo/fb/943d8	Sean Dyche	Ben Mee
Burnley vs. Southampton	10-08-2019 12:00:00 AM	Southampton	https://d2p3bygnnzw9w3.cloudfront.net/req/202007313/tlogo/fb/33c89	Ralph Hasenhüttl	James Ward-Prowse
Watford vs. Brighton & Hove Albion	10-08-2019 12:00:00 AM	Watford	https://d2p3bygnnzw9w3.cloudfront.net/req/202007313/tlogo/fb/2abfe	Javi Gracia	Troy Deeney
Watford vs. Brighton & Hove Albion	10-08-2019 12:00:00 AM	Brighton & Hove Albion	https://d2p3bygnnzw9w3.cloudfront.net/req/202007313/tlogo/fb/40753	Graham Potter	Lewis Dunk
Bournemouth vs. Sheffield United	10-08-2019 12:00:00 AM	Bournemouth	https://d2p3bygnnzw9w3.cloudfront.net/req/202007313/tlogo/fb/4ba7c	Eddie Howe	Steve Cook
Bournemouth vs. Sheffield United	10-08-2019 12:00:00 AM	Sheffield United	https://d2p3bygnnzw9w3.cloudfront.net/req/202007313/tlogo/fb/1df6b	Chris Wilder	Oliver Norwood
Crystal Palace vs. Everton	10-08-2019 12:00:00 AM	Crystal Palace	https://d2p3bygnnzw9w3.cloudfront.net/req/202007313/tlogo/fb/47c64	Roy Hodgson	Luka Milivojević
Crystal Palace vs. Everton	10-08-2019 12:00:00 AM	Everton	https://d2p3bygnnzw9w3.cloudfront.net/req/202007313/tlogo/fb/d3fd3	Marco Silva	Séamus Coleman
Tottenham Hotspur vs. Aston Villa	10-08-2019 12:00:00 AM	Tottenham Hotspur	https://d2p3bygnnzw9w3.cloudfront.net/req/202007313/tlogo/fb/361ca	Mauricio Pochettino	Hugo Lloris
Tottenham Hotspur vs. Aston Villa	10-08-2019 12:00:00 AM	Aston Villa	https://d2p3bygnnzw9w3.cloudfront.net/req/202007313/tlogo/fb/86022	Dean Smith	Jack Grealish
Newcastle United vs. Arsenal	11-08-2019 12:00:00 AM	Newcastle United	https://d2p3bygnnzw9w3.cloudfront.net/req/202007313/tlogo/fb/b2b47	Steve Bruce	Jamaal Lascelles

Through the process of ETL, we are going to clean the dataset and bring all the entities to their proper data format.

Step 1: Removing the duplicate cells if any from the dataset.

For this, makes the connections with the dataset and open the power query editor, then select the whole dataset. Go to remove rows and select the remove duplicate rows. Then all the duplicate rows are removed if any present.

The screenshot shows the Microsoft Power Query Editor interface. The ribbon at the top includes 'Home', 'Transform', 'Add Column', and 'View'. The 'Remove Rows' button is highlighted in the 'Transform' tab. The main area displays a table with columns: Match_Name, Match_Date, Team, Team_Logo, and Manager. The table contains 28 rows of football match data. The status bar at the bottom indicates '33 COLUMNS, 760 ROWS' and 'Column profiling based on top 1000 rows'. The 'Query Settings' pane on the right shows the 'APPLIED STEPS' list with 'Removed Duplicates' selected.

Step 2: Removing columns which are not properly defined or not crucial to our analysis.

For this we will remove columns by choosing column in power query editor.

Sheet1 - Power Query Editor

File Home Transform Add Column View

Close & Load Refresh Preview Advanced Editor Properties

Choose Columns Remove Columns Keep Rows Remove Rows Split Column Group By Data Type: Text * Use First Row as Headers * Replace Values Merge Queries * Append Queries * Combine Files Manage Parameters Data source settings New Source * Recent Sources * Enter Data New Query

Queries [1] Sheet1

Table.SelectColumns(#"Removed Duplicates",{"Match_Name", "Match_Date", "Team", "Manager", "Captain", "Score", "Fouls", "Offsides",

	Match_Name	Match_Date	Team	Manager	Captain	Score
1	Liverpool vs. Norwich City	09-08-2019 12:00:00 AM	Liverpool	Jürgen Klopp	Jordan Henderson	
2	Liverpool vs. Norwich City	09-08-2019 12:00:00 AM	Norwich City	Daniel Farke	Grant Hanley	
3	West Ham United vs. Manchester City	10-08-2019 12:00:00 AM	West Ham United	Manuel Pellegrini	Aaron Cresswell	
4	West Ham United vs. Manchester City	10-08-2019 12:00:00 AM	Manchester City	Pep Guardiola	David Silva	
5	Burnley vs. Southampton	10-08-2019 12:00:00 AM	Burnley	Sean Dyche	Ben Mee	
6	Burnley vs. Southampton	10-08-2019 12:00:00 AM	Southampton	Ralph Hasenhüttl	James Ward-Prowse	
7	Watford vs. Brighton & Hove Albion	10-08-2019 12:00:00 AM	Watford	Javi Gracia	Troy Deeney	
8	Watford vs. Brighton & Hove Albion	10-08-2019 12:00:00 AM	Brighton & Hove Albion	Graham Potter	Lewis Dunk	
9	Bournemouth vs. Sheffield United	10-08-2019 12:00:00 AM	Bournemouth	Eddie Howe	Steve Cook	
10	Bournemouth vs. Sheffield United	10-08-2019 12:00:00 AM	Sheffield United	Chris Wilder	Oliver Norwood	
11	Crystal Palace vs. Everton	10-08-2019 12:00:00 AM	Crystal Palace	Roy Hodgson	Luka Milivojević	
12	Crystal Palace vs. Everton	10-08-2019 12:00:00 AM	Everton	Marco Silva	Séamus Coleman	
13	Tottenham Hotspur vs. Aston Villa	10-08-2019 12:00:00 AM	Tottenham Hotspur	Mauricio Pochettino	Hugo Lloris	
14	Tottenham Hotspur vs. Aston Villa	10-08-2019 12:00:00 AM	Aston Villa	Dean Smith	Jack Grealish	
15	Newcastle United vs. Arsenal	11-08-2019 12:00:00 AM	Newcastle United	Steve Bruce	Jamaal Lascelles	
16	Newcastle United vs. Arsenal	11-08-2019 12:00:00 AM	Arsenal	Unai Emery	Granit Xhaka	
17	Leicester City vs. Wolverhampton Wanderers	11-08-2019 12:00:00 AM	Leicester City	Brendan Rodgers	Kasper Schmeichel	
18	Leicester City vs. Wolverhampton Wanderers	11-08-2019 12:00:00 AM	Wolverhampton Wanderers	Nuno Espírito Santo	Conor Coady	
19	Manchester United vs. Chelsea	11-08-2019 12:00:00 AM	Manchester United	Ole Gunnar Solskjær	David de Gea	
20	Manchester United vs. Chelsea	11-08-2019 12:00:00 AM	Chelsea	Frank Lampard	César Azpilicueta	
21	Arsenal vs. Burnley	17-08-2019 12:00:00 AM	Arsenal	Unai Emery	Nacho Monreal	
22	Arsenal vs. Burnley	17-08-2019 12:00:00 AM	Burnley	Sean Dyche	Ben Mee	
23	Southampton vs. Liverpool	17-08-2019 12:00:00 AM	Southampton	Ralph Hasenhüttl	Pierre Højbjerg	
24	Southampton vs. Liverpool	17-08-2019 12:00:00 AM	Liverpool	Jürgen Klopp	James Milner	
25	Norwich City vs. Newcastle United	17-08-2019 12:00:00 AM	Norwich City	Daniel Farke	Grant Hanley	
26	Norwich City vs. Newcastle United	17-08-2019 12:00:00 AM	Newcastle United	Steve Bruce	Jamaal Lascelles	
27	Brighton & Hove Albion vs. West Ham United	17-08-2019 12:00:00 AM	Brighton & Hove Albion	Graham Potter	Lewis Dunk	
28	Brighton & Hove Albion vs. West Ham United	17-08-2019 12:00:00 AM	West Ham United	Manuel Pellegrini	Anzelo Orbonna	

Query Settings

PROPERTIES

Name

Sheet1

All Properties

APPLIED STEPS

Source

Navigation

Promoted Headers

Changed Type

Removed Duplicates

Removed Other Columns

21 COLUMNS, 760 ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 12:02 PM

Step 3: Checking columns and validations of data

There is passing accuracy column which shows zero(0) value in whole column. Update and correct the passing accuracy column by taking ratio of SucPass and NumofPass column.

AutoSave final project Search (Alt+Q) Sanjay Chaudhary

File Home Insert Page Layout Formulas Data Review View Developer Help Power Pivot Table Design Query

Calibri 11 Font Wrap Text General Conditional Formatting Format as Table Cell Styles Insert Delete Format AutoSum Fill Sort & Filter Find & Select Analyze Data

Undo Clipboard Font Alignment Number Styles Cells Editing Analysis

M2 =[@SucPass]/[@NumofPass]

	Manager	Captain	Score	Fouls	Offsides	YellowCard	RedCard	YellowRed	Possession	PassingAccuracy	SucPass	NumofPass	Sho
1	Jürgen Klopp	Jordan Henderson	4	10	0	0	0	0	0.569999993	0.785028791	409	521	
2	Daniel Farke	Grant Hanley	1	12	4	2	0	0	0.430000007	0.769620253	304	395	
3	Pep Guardiola	David Silva	5	15	3	2	0	0	0.569999993	0.877659574	495	564	
4	Sean Dyche	Ben Mee	3	6	4	0	0	0	0.460000008	0.61849711	214	346	
5	Graham Potter	Lewis Dunk	3	11	3	1	0	0	0.5	0.795258621	369	464	
6	Roy Hodgson	Luka Milivojević	0	18	4	2	0	0	0.349999994	0.606666667	182	300	
7	Marco Silva	Séamus Coleman	0	13	0	2	0	1	0.649999976	0.791071429	443	560	
8	Mauricio Pochettino	Hugo Lloris	3	14	1	1	0	0	0.699999988	0.863849765	552	639	
9	Manuel Pellegrini	Aaron Cresswell	0	11	2	2	0	0	0.430000007	0.810747664	347	428	
10	Ralph Hasenhüttl	James Ward-Prowse	0	14	1	0	0	0	0.540000021	0.686893204	283	412	
11	Javi Gracia	Troy Deeney	0	15	0	0	0	0	0.5	0.778021978	354	455	
12	Eddie Howe	Steve Cook	1	10	1	2	0	0	0.529999971	0.7731569	409	529	
13	Chris Wilder	Oliver Norwood	1	21	0	1	0	0	0.469999999	0.769392034	367	477	
14	Dean Smith	Jack Grealish	1	11	0	0	0	0	0.300000012	0.76534296	212	277	
15	Nuno Espírito Santo	Conor Coady	0	14	1	2	0	0	0.310000002	0.691358025	224	324	
16	Unai Emery	Granit Xhaka	1	7	3	3	0	0	0.610000014	0.854961832	560	655	
17	Brendan Rodgers	Kasper Schmeichel	0	6	0	0	0	0	0.689999998	0.83427762	589	706	
18	Ole Gunnar Solskjær	David de Gea	4	18	2	3	0	0	0.460000008	0.822175732	393	478	
19	Steve Bruce	Jamaal Lascelles	0	13	1	1	0	0	0.389999986	0.761557178	313	411	
20	Frank Lampard	César Azpilicueta	0	16	0	4	0	0	0.540000021	0.833333333	470	564	
21	Marco Silva	Séamus Coleman	1	13	2	2	0	0	0.439999998	0.708333333	238	336	
22	Daniel Farke	Grant Hanley	3	11	2	1	0	0	0.629999995	0.868719611	536	617	
23	Unai Emery	Nacho Monreal	2	13	5	2	0	0	0.670000017	0.84519573	475	562	
24	Jürgen Klopp	James Milner	2	0	1	1	0	0	0.670000017	0.84519573	475	562	

Queries & Connections

Sheet1

760 rows loaded.

Ready Accessibility: Investigate

Step 4: Improvising Proper Data Formatting

Without proper Data Formatting, proper analysis will not take place. So, we will bring down certain columns to their proper format. For example, the dates should be in the date format.

The screenshot shows the Power Query Editor interface. The main area displays a table with the following columns: Match Name, Match Date, Team, Manager, Captain, and Score. The data is sorted by Match Date. The ribbon at the top includes options for File, Home, Transform, Add Column, and View. The Query Settings pane on the right shows the 'APPLIED STEPS' section with 'Changed Type1' selected.

Match Name	Match Date	Team	Manager	Captain	Score
Liverpool vs. Norwich City	09-08-2019	Liverpool	Jürgen Klopp	Jordan Henderson	
Liverpool vs. Norwich City	09-08-2019	Norwich City	Daniel Farke	Grant Hanley	
West Ham United vs. Manchester City	10-08-2019	West Ham United	Manuel Pellegrini	Aaron Cresswell	
West Ham United vs. Manchester City	10-08-2019	Manchester City	Pep Guardiola	David Silva	
Burnley vs. Southampton	10-08-2019	Burnley	Sean Dyche	Ben Mee	
Burnley vs. Southampton	10-08-2019	Southampton	Ralph Hasenhüttl	James Ward-Prowse	
Watford vs. Brighton & Hove Albion	10-08-2019	Watford	Javi Gracia	Troy Deeney	
Watford vs. Brighton & Hove Albion	10-08-2019	Brighton & Hove Albion	Graham Potter	Lewis Dunk	
Bournemouth vs. Sheffield United	10-08-2019	Bournemouth	Eddie Howe	Steve Cook	
Bournemouth vs. Sheffield United	10-08-2019	Sheffield United	Chris Wilder	Oliver Norwood	
Crystal Palace vs. Everton	10-08-2019	Crystal Palace	Roy Hodgson	Luka Milivojević	
Crystal Palace vs. Everton	10-08-2019	Everton	Marco Silva	Séamus Coleman	
Tottenham Hotspur vs. Aston Villa	10-08-2019	Tottenham Hotspur	Mauricio Pochettino	Hugo Lloris	
Tottenham Hotspur vs. Aston Villa	10-08-2019	Aston Villa	Dean Smith	Jack Grealish	
Newcastle United vs. Arsenal	11-08-2019	Newcastle United	Steve Bruce	Jamaal Lascelles	
Newcastle United vs. Arsenal	11-08-2019	Arsenal	Unai Emery	Granit Xhaka	
Leicester City vs. Wolverhampton Wanderers	11-08-2019	Leicester City	Brendan Rodgers	Kasper Schmeichel	
Leicester City vs. Wolverhampton Wanderers	11-08-2019	Wolverhampton Wanderers	Nuno Espírito Santo	Conor Coady	
Manchester United vs. Chelsea	11-08-2019	Manchester United	Ole Gunnar Solskjær	David de Gea	
Manchester United vs. Chelsea	11-08-2019	Chelsea	Frank Lampard	César Azpilicueta	
Arsenal vs. Burnley	17-08-2019	Arsenal	Unai Emery	Nacho Monreal	
Arsenal vs. Burnley	17-08-2019	Burnley	Sean Dyche	Ben Mee	
Southampton vs. Liverpool	17-08-2019	Southampton	Ralph Hasenhüttl	Pierre Højbjerg	
Southampton vs. Liverpool	17-08-2019	Liverpool	Jürgen Klopp	James Milner	
Norwich City vs. Newcastle United	17-08-2019	Norwich City	Daniel Farke	Grant Hanley	
Norwich City vs. Newcastle United	17-08-2019	Newcastle United	Steve Bruce	Jamaal Lascelles	
Brighton & Hove Albion vs. West Ham United	17-08-2019	Brighton & Hove Albion	Graham Potter	Lewis Dunk	
Brighton & Hove Albion vs. West Ham United	17-08-2019	West Ham United	Manuel Pellegrini	Anelko Oebonna	

After applying all the transform operations, we load the data into table and data table shown in picture below:

The screenshot shows an Excel spreadsheet with the following columns: Match Name, Match Date, Team, Manager, Captain, Score, Fouls, Offsides, YellowCard, RedCard, YellowRed, Possession, and PassingAcc. The data is sorted by Match Date. The spreadsheet is titled 'premier league data'.

Match Name	Match Date	Team	Manager	Captain	Score	Fouls	Offsides	YellowCard	RedCard	YellowRed	Possession	PassingAcc
Liverpool vs. Norwich City	09-08-2019	Liverpool	Jürgen Klopp	Jordan Henderson	4	10	0	0	0	0	0.569999993	0.7
Liverpool vs. Norwich City	09-08-2019	Norwich City	Daniel Farke	Grant Hanley	1	12	4	2	0	0	0.430000007	0.7
West Ham United vs. Manchester City	10-08-2019	Manchester City	Pep Guardiola	David Silva	5	15	3	2	0	0	0.569999993	0.8
Burnley vs. Southampton	10-08-2019	Burnley	Sean Dyche	Ben Mee	3	6	4	0	0	0	0.460000008	0
Watford vs. Brighton & Hove Albion	10-08-2019	Brighton & Hove Albion	Graham Potter	Lewis Dunk	3	11	3	1	0	0	0.5	0.7
Crystal Palace vs. Everton	10-08-2019	Crystal Palace	Roy Hodgson	Luka Milivojević	0	18	4	2	0	0	0.349999994	0.6
Crystal Palace vs. Everton	10-08-2019	Everton	Marco Silva	Séamus Coleman	0	13	0	2	0	1	0.649999976	0.7
Tottenham Hotspur vs. Aston Villa	10-08-2019	Tottenham Hotspur	Mauricio Pochettino	Hugo Lloris	3	14	1	1	0	0	0.699999988	0.8
West Ham United vs. Manchester City	10-08-2019	West Ham United	Manuel Pellegrini	Aaron Cresswell	0	11	2	2	0	0	0.430000007	0.8
Burnley vs. Southampton	10-08-2019	Southampton	Ralph Hasenhüttl	James Ward-Prowse	0	14	1	0	0	0	0.540000021	0.6
Watford vs. Brighton & Hove Albion	10-08-2019	Watford	Javi Gracia	Troy Deeney	0	15	0	0	0	0	0.5	0.7
Bournemouth vs. Sheffield United	10-08-2019	Bournemouth	Eddie Howe	Steve Cook	1	10	1	2	0	0	0.529999971	0.7
Bournemouth vs. Sheffield United	10-08-2019	Sheffield United	Chris Wilder	Oliver Norwood	1	21	0	1	0	0	0.469999999	0.7
Tottenham Hotspur vs. Aston Villa	10-08-2019	Aston Villa	Dean Smith	Jack Grealish	1	11	0	0	0	0	0.300000012	0
Leicester City vs. Wolverhampton Wanderers	11-08-2019	Wolverhampton Wanderers	Nuno Espírito Santo	Conor Coady	0	14	1	2	0	0	0.310000002	0.6
Newcastle United vs. Arsenal	11-08-2019	Arsenal	Unai Emery	Granit Xhaka	1	7	3	3	0	0	0.610000014	0.8
Leicester City vs. Wolverhampton Wanderers	11-08-2019	Leicester City	Brendan Rodgers	Kasper Schmeichel	0	6	0	0	0	0	0.689999998	0
Manchester United vs. Chelsea	11-08-2019	Manchester United	Ole Gunnar Solskjær	David de Gea	4	18	2	3	0	0	0.460000008	0.8
Newcastle United vs. Arsenal	11-08-2019	Newcastle United	Steve Bruce	Jamaal Lascelles	0	13	1	1	0	0	0.389999986	0.7
Manchester United vs. Chelsea	11-08-2019	Chelsea	Frank Lampard	César Azpilicueta	0	16	0	4	0	0	0.540000021	0.8
Everton vs. Watford	17-08-2019	Everton	Marco Silva	Séamus Coleman	1	13	2	2	0	0	0.439999998	0.7
Norwich City vs. Newcastle United	17-08-2019	Norwich City	Daniel Farke	Grant Hanley	3	11	2	1	0	0	0.629999995	0.8
Arsenal vs. Burnley	17-08-2019	Arsenal	Unai Emery	Nacho Monreal	2	13	5	2	0	0	0.670000017	0
Southampton vs. Liverpool	17-08-2019	Liverpool	Jürgen Klopp	James Milner	2	9	1	1	0	0	0.629999995	0.7
Aston Villa vs. Bournemouth	17-08-2019	Bournemouth	Eddie Howe	Steve Cook	2	16	2	2	0	0	0.370000005	0.7
Manchester City vs. Tottenham Hotspur	17-08-2019	Manchester City	Pep Guardiola	Kevin De Bruyne	2	16	2	1	0	0	0.569999993	0.8

ANALYSIS OF DATASET

1. Which Manager is best at possession play throughout the season?

Description:

In this objective we will be finding the best manager to play his team by keeping control on the ball for maximum time. For this analysis, we need a manager column and possession column. By taking the average of possession column under different manager, we will be able to find for throughout the season.

Specific function and requirements

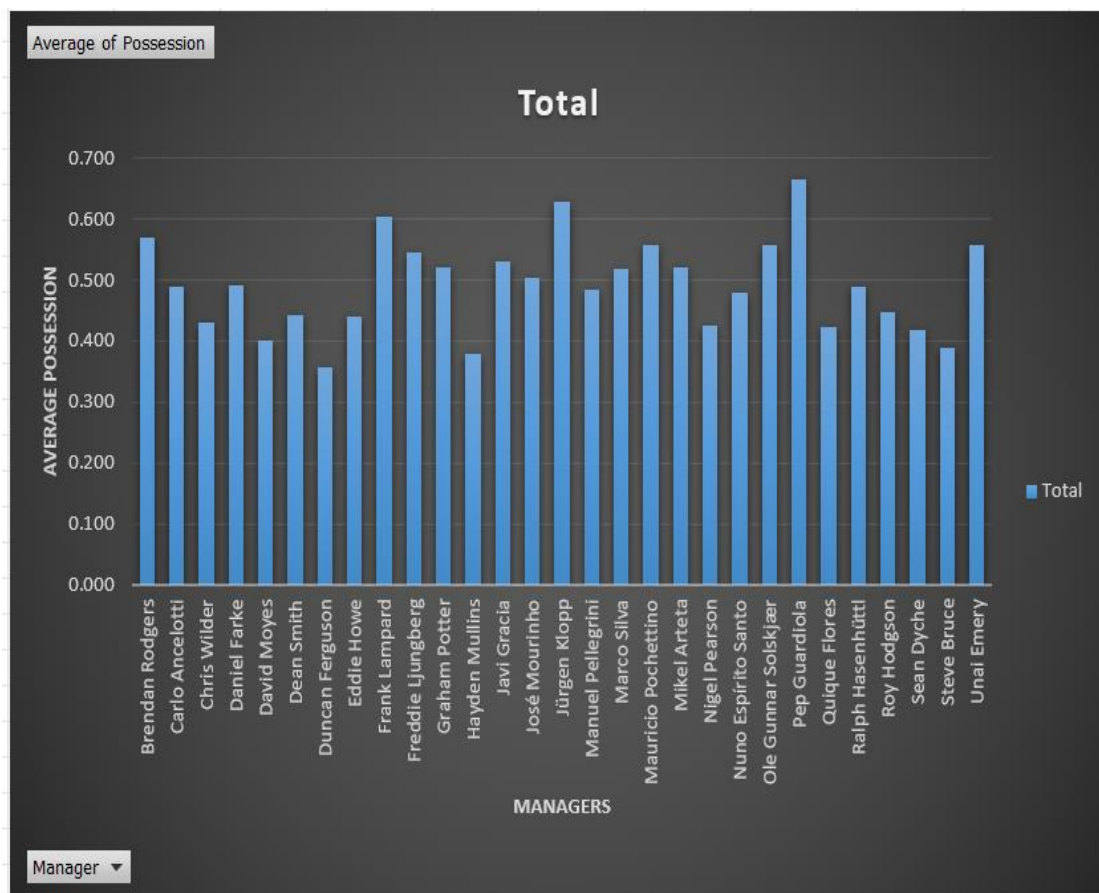
We have to create a pivot table to determine the average possession with different manager and then visualize it on graph.

Results:

2	Managers	Average of Possession
3	Brendan Rodgers	0.569
4	Carlo Ancelotti	0.488
5	Chris Wilder	0.432
6	Daniel Farke	0.492
7	David Moyes	0.402
8	Dean Smith	0.443
9	Duncan Ferguson	0.357
10	Eddie Howe	0.441
11	Frank Lampard	0.604
12	Freddie Ljungberg	0.546
13	Graham Potter	0.522
14	Hayden Mullins	0.380
15	Javi Gracia	0.530
16	José Mourinho	0.504
17	Jürgen Klopp	0.628
18	Manuel Pellegrini	0.485
19	Marco Silva	0.519
20	Mauricio Pochettino	0.557
21	Mikel Arteta	0.520
22	Nigel Pearson	0.425
23	Nuno Espírito Santo	0.481
24	Ole Gunnar Solskjær	0.557
25	Pep Guardiola	0.665
26	Quique Flores	0.424
27	Ralph Hasenhüttl	0.490
28	Roy Hodgson	0.447
29	Sean Dyche	0.419
30	Steve Bruce	0.388
31	Unai Emery	0.558
32	Grand Total	0.500

Visualization:

The results are then visualized in the form of a stacked bar graph for manager and Average possession.



2. Which team's Goalkeeper is best throughout the season?

Description:

We will do analysis on the basis of average saves accuracy per game . So, the required columns are team name and the SavesAccuracy to determine saves accuracy by the goalkeeper.

Specific function and requirements:

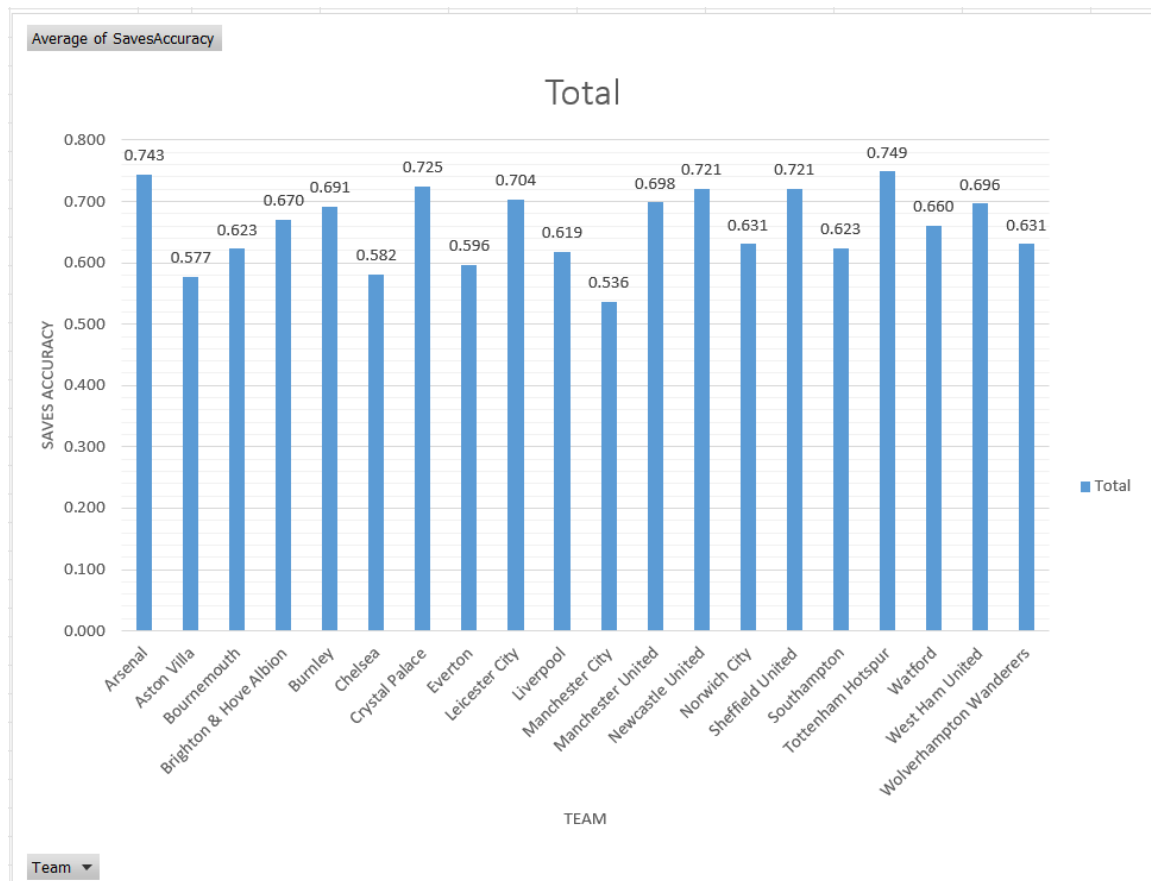
There are 20 clubs played in 2019-2020 season. So, we will be using Team columns as Row labels and the average saves accuracy by each team's goalkeeper.

Results:

Teams	Average of SavesAccuracy
Arsenal	0.743
Aston Villa	0.577
Bournemouth	0.623
Brighton & Hove Albion	0.670
Burnley	0.691
Chelsea	0.582
Crystal Palace	0.725
Everton	0.596
Leicester City	0.704
Liverpool	0.619
Manchester City	0.536
Manchester United	0.698
Newcastle United	0.721
Norwich City	0.631
Sheffield United	0.721
Southampton	0.623
Tottenham Hotspur	0.749
Watford	0.660
West Ham United	0.696
Wolverhampton Wanderers	0.631
Grand Total	0.660

Visualization:

We use the bar graph to visualize the distribution.



3. Which team is most disciplined in premier league football?

Description:

For this objective we will be needing Fouls, yellowcard, redcard and yellowRed columns.

We will find the team with less number of cards received during the season.

Specific function and requirements:

The specific requirements for this objective is use of sum function and to get the number of cards received by the team and then we use the data to create pivot table and plot it by using pivot chart.

Results:

1	Team	Fouls	YellowCard	RedCard	YellowRed	Total Cards
2	Liverpool	10	0	0	0	0
3	Norwich City	12	2	0	0	2
4	Manchester City	15	2	0	0	2
5	Burnley	6	0	0	0	0
6	Brighton & Hove Albion	11	1	0	0	1
7	Crystal Palace	18	2	0	0	2
8	Everton	13	2	0	1	3
9	Tottenham Hotspur	14	1	0	0	1
10	West Ham United	11	2	0	0	2
11	Southampton	14	0	0	0	0
12	Watford	15	0	0	0	0
13	Bournemouth	10	2	0	0	2
14	Sheffield United	21	1	0	0	1
15	Aston Villa	11	0	0	0	0
16	Wolverhampton Wanderers	14	2	0	0	2
17	Arsenal	7	3	0	0	3
18	Leicester City	6	0	0	0	0
19	Manchester United	18	3	0	0	3
20	Newcastle United	13	1	0	0	1
21	Chelsea	16	4	0	0	4
22	Everton	13	2	0	0	2
23	Norwich City	11	1	0	0	1
24	Arsenal	13	2	0	0	2
25	Liverpool	9	1	0	0	1
26	Bournemouth	16	2	0	0	2
27	Manchester City	16	1	0	0	1
28	Tottenham Hotspur	6	0	0	0	0
29	Watford	12	3	0	0	3
30	Burnley	17	1	0	0	1
31	Southampton	12	2	0	0	2
32	Newcastle United	14	3	0	0	3


premier league data

possession

Best goalkeeper

Most discipline

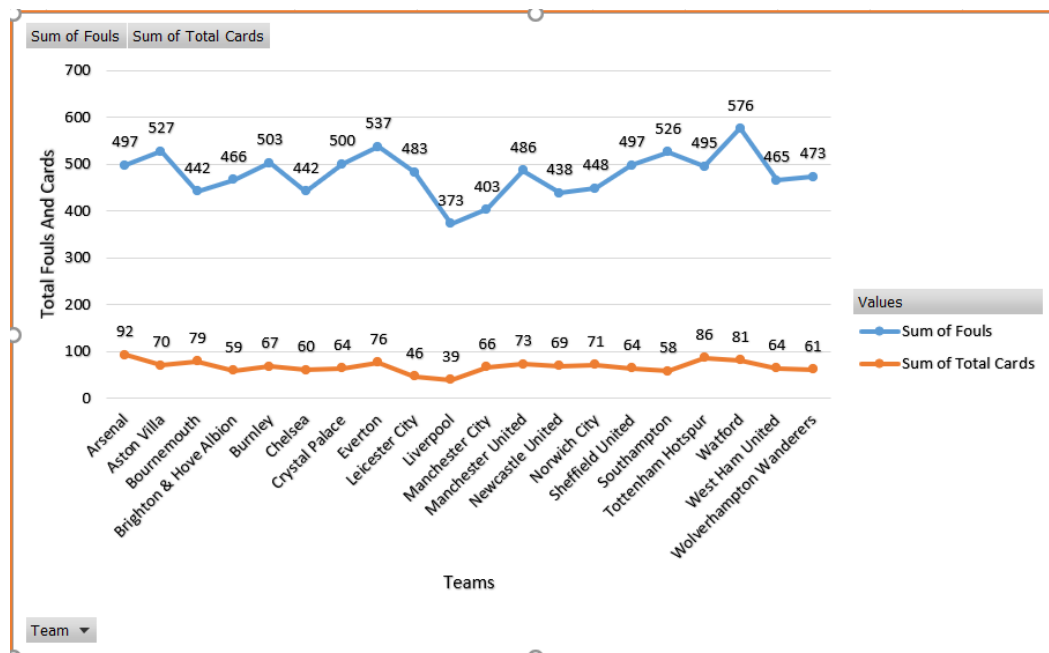
+

Ready  Accessibility: Investigate

Teams	Sum of Fouls	Sum of Total Cards
Arsenal	497	92
Aston Villa	527	70
Bournemouth	442	79
Brighton & Hove Albion	466	59
Burnley	503	67
Chelsea	442	60
Crystal Palace	500	64
Everton	537	76
Leicester City	483	46
Liverpool	373	39
Manchester City	403	66
Manchester United	486	73
Newcastle United	438	69
Norwich City	448	71
Sheffield United	497	64
Southampton	526	58
Tottenham Hotspur	495	86
Watford	576	81
West Ham United	465	64
Wolverhampton Wanderers	473	61
Grand Total	9577	1345

Visualization:

The results are visualized with the help of line graph with total fouls and cards on y axis and teams on x axis.



4. Which team is more successful in scoring?

Description:

In this objective we will be needing the data of total goal scored and number of shots taken by the team.

Specific function and requirements:

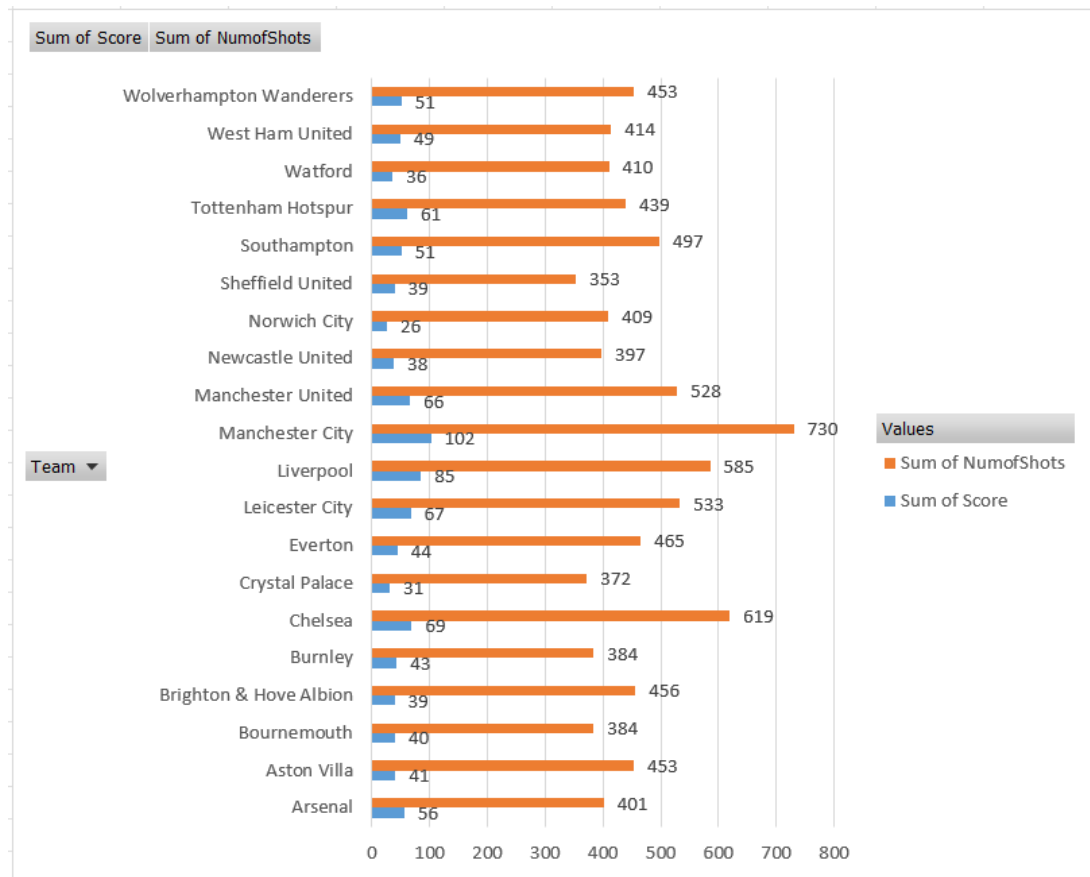
We have to create a pivot table. No specific functions are used. We have put the data of every match and then plot the graph.

Results:

1	Team	Sum of Score	Sum of NumofShots
2	Arsenal	56	401
3	Aston Villa	41	453
4	Bournemouth	40	384
5	Brighton & Hove Albion	39	456
6	Burnley	43	384
7	Chelsea	69	619
8	Crystal Palace	31	372
9	Everton	44	465
10	Leicester City	67	533
11	Liverpool	85	585
12	Manchester City	102	730
13	Manchester United	66	528
14	Newcastle United	38	397
15	Norwich City	26	409
16	Sheffield United	39	353
17	Southampton	51	497
18	Tottenham Hotspur	61	439
19	Watford	36	410
20	West Ham United	49	414
21	Wolverhampton Wanderers	51	453
22	Grand Total	1034	9282

Visualization:

We visualize the above results with the help a bar chart created using pivot table.



5. Which team changes most number of Managers & Captains throughout the season?

Description:

In this objective, we need to count how many times changes a team changes his manager and captain throughout the season.

Specific function and requirements:

For calculating number of times, a team changes his manager and captain, we need data from team, manager, captain, match_name column. We create a pivot table with team with row label and manager with column label. Count the number of games played by team with specific manager. Then copy the pivot table on blank worksheet and apply COUNTA() function which gives number of managers played with specific team. We have also done same to find out number of captains changes during the season.

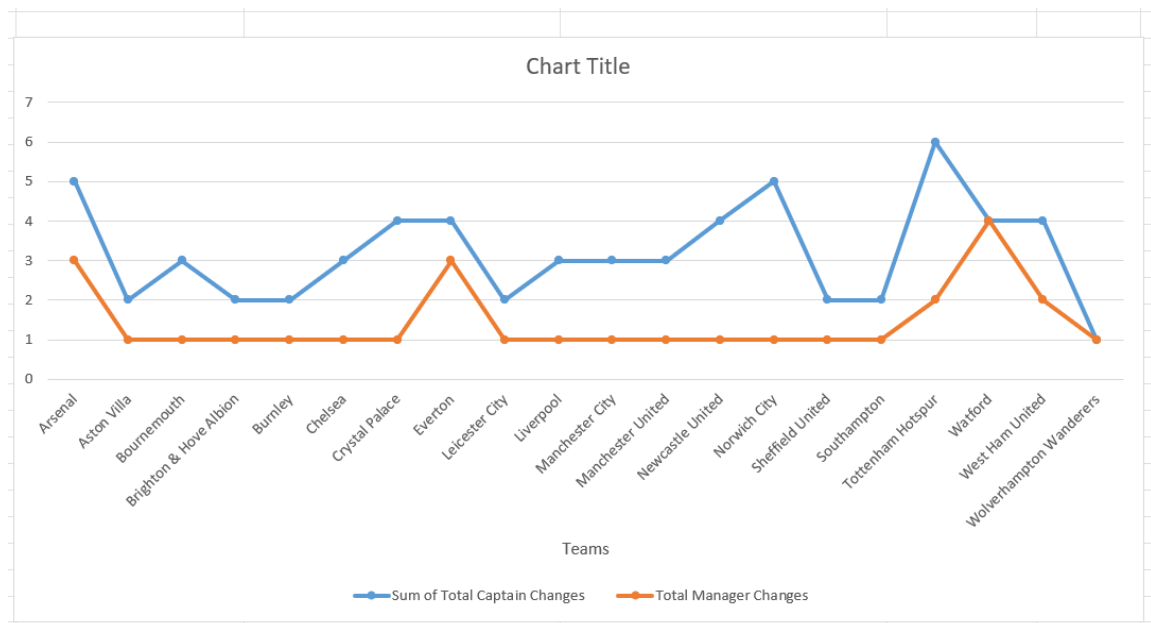
Results:

Count of Match_Name	Managers						
Team	Brendan Rodgers	Carlo Ancelotti	Chris Wilder	Daniel Farke	David Moyes	Dean Smith	Duncan Ferguson
Arsenal							
Aston Villa							38
Bournemouth							
Brighton & Hove Albion							
Burnley							
Chelsea							
Crystal Palace							
Everton			20				3
Leicester City	38						
Liverpool							
Manchester City							
Manchester United							
Newcastle United							
Norwich City					38		
Sheffield United			38				
Southampton							
Tottenham Hotspur							
Watford							
West Ham United						19	
Wolverhampton Wanderers							

Team	Brendan Rodgers	Carlo Ancelotti	Chris Wilder	Daniel Farke	David Moyes	Dean Smith	Duncan Ferguson
Arsenal							
Aston Villa							38
Bournemouth							
Brighton & Hove Albion							
Burnley							
Chelsea							
Crystal Palace							
Everton			20				3
Leicester City	38						
Liverpool							
Manchester City							
Manchester United							
Newcastle United							
Norwich City					38		
Sheffield United			38				
Southampton							
Tottenham Hotspur							
Watford							
West Ham United						19	
Wolverhampton Wanderers							

Visualization:

The results are visualized in the form of line graph.



6. Which team have highest successful passes and accuracy per game?

Description:

We will do analysis on the basis of average passing accuracy per game . So, the required columns are team name, Passing Accuracy, SucPass and numofPass column to determine highest passing accuracy and successful number of passes.

Specific function and requirements:

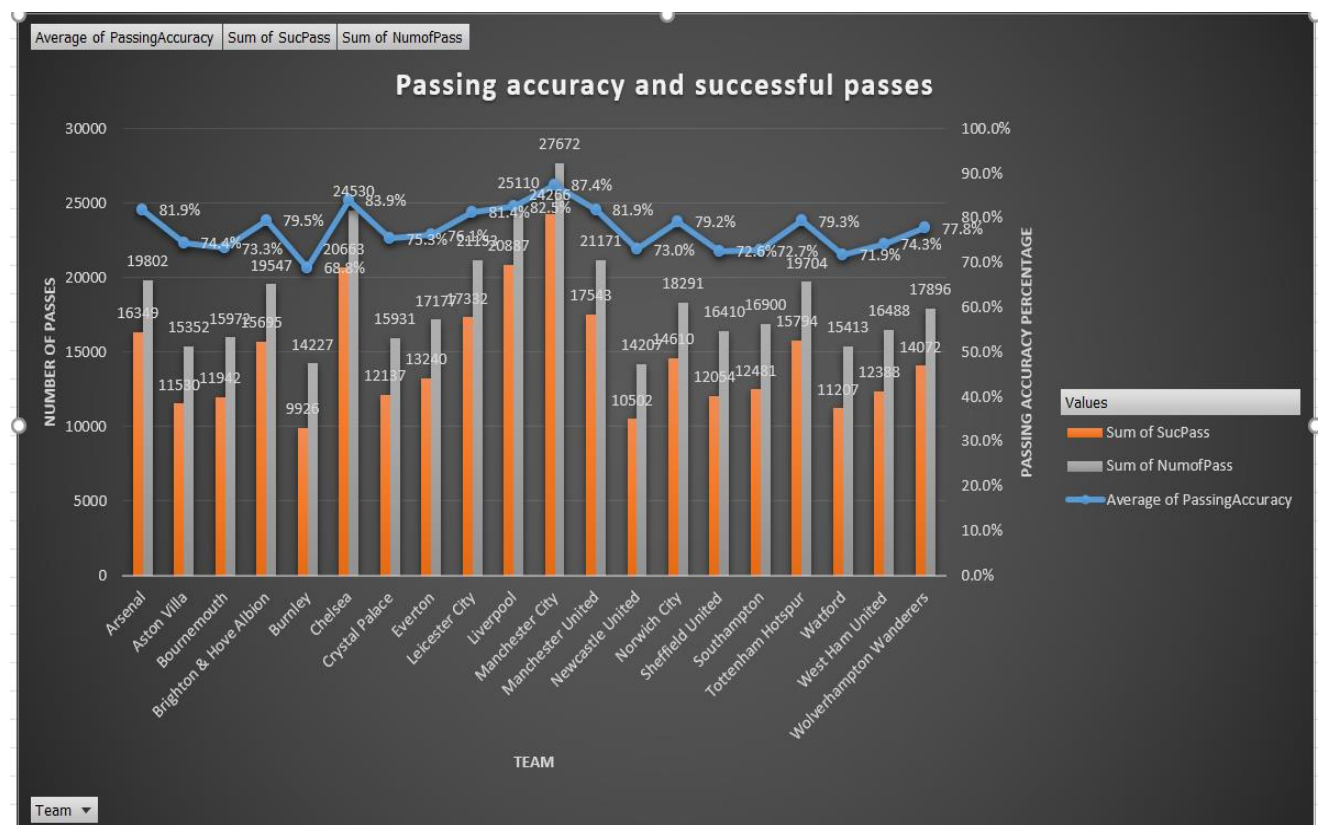
There are 20 clubs played in 2019-2020 season. So, we will create a pivot table and used Team columns as Row labels and the average passing accuracy and highest number of passes by each team.

Results:

Team	Average of PassingAccuracy	Sum of SucPass	Sum of NumofPass
Arsenal	81.9%	16349	19802
Aston Villa	74.4%	11530	15352
Bournemouth	73.3%	11942	15972
Brighton & Hove Albion	79.5%	15695	19547
Burnley	68.8%	9926	14227
Chelsea	83.9%	20663	24530
Crystal Palace	75.3%	12137	15931
Everton	76.1%	13240	17177
Leicester City	81.4%	17332	21153
Liverpool	82.5%	20887	25110
Manchester City	87.4%	24266	27672
Manchester United	81.9%	17543	21171
Newcastle United	73.0%	10502	14207
Norwich City	79.2%	14610	18291
Sheffield United	72.6%	12054	16410
Southampton	72.7%	12481	16900
Tottenham Hotspur	79.3%	15794	19704
Watford	71.9%	11207	15413
West Ham United	74.3%	12388	16488
Wolverhampton Wanderers	77.8%	14072	17896

Visualization:

We visualize the above results with the help of custom combination of cluster column graph with primary axis and line graph for passing accuracy percentage with secondary axis.



ANALYSIS RESULTS

1. Which Manager is best at possession play throughout the season?
 - From our analysis, we would be able to find that Pep Guardiola Manager of Manchester City is the best in possession play with 66.5% average possession per game. Jurgen Klopp manager of Liverpool is the second-best manager in possession play with 62.8% average possession per game.
2. Which team's Goalkeeper is best throughout the season?
 - From the analysis on premier league dataset, Tottenham Hotspur's goalkeeper is best among all the club's goalkeeper with 74.9% Saves accuracy. Arsenal's goalkeeper is not too far behind, is the second-best goalkeeper with saves accuracy of 74.3%.
3. Which team is most disciplined in premier league football?
 - Liverpool is the most disciplined team in premier league 2019-2020 season. Liverpool received 39 total cards which is least number of cards by any team and also attempts least number of fouls on opponents which is 373 in number.
 - Leicester city is second's best team in discipline which received 46 total cards but attempts more fouls than Manchester city. Manchester City is second best in attempting fouls on opponents which 403 in number.
4. Which team is more successful in scoring?
 - From our analysis, we would be able to find that Manchester City is the best Team in scoring goals with 102 goals in the whole season While Liverpool is the second-best goal scoring team with 85 goals.
 - Manchester City leads in number of shot with 730 shots and Chelsea is 2nd team in taking number of shots with 619 shots but they are only able to score 69 goals throughout the season.
5. Which team changes most number of Managers & Captains throughout the season?
 - Watford Changes the most number managers throughout the season. Hayden Mullins, Javi Garcia, Nigel Pearson and Quique Flores are the 4 managers that Watford changes during the season.

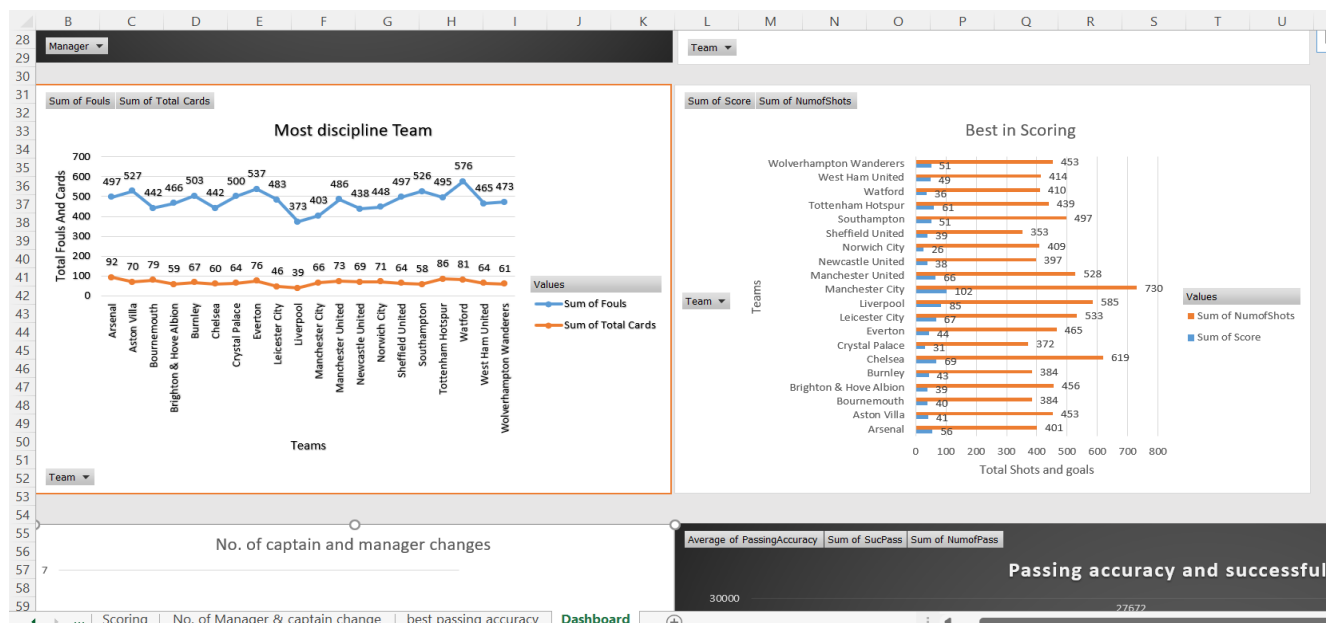
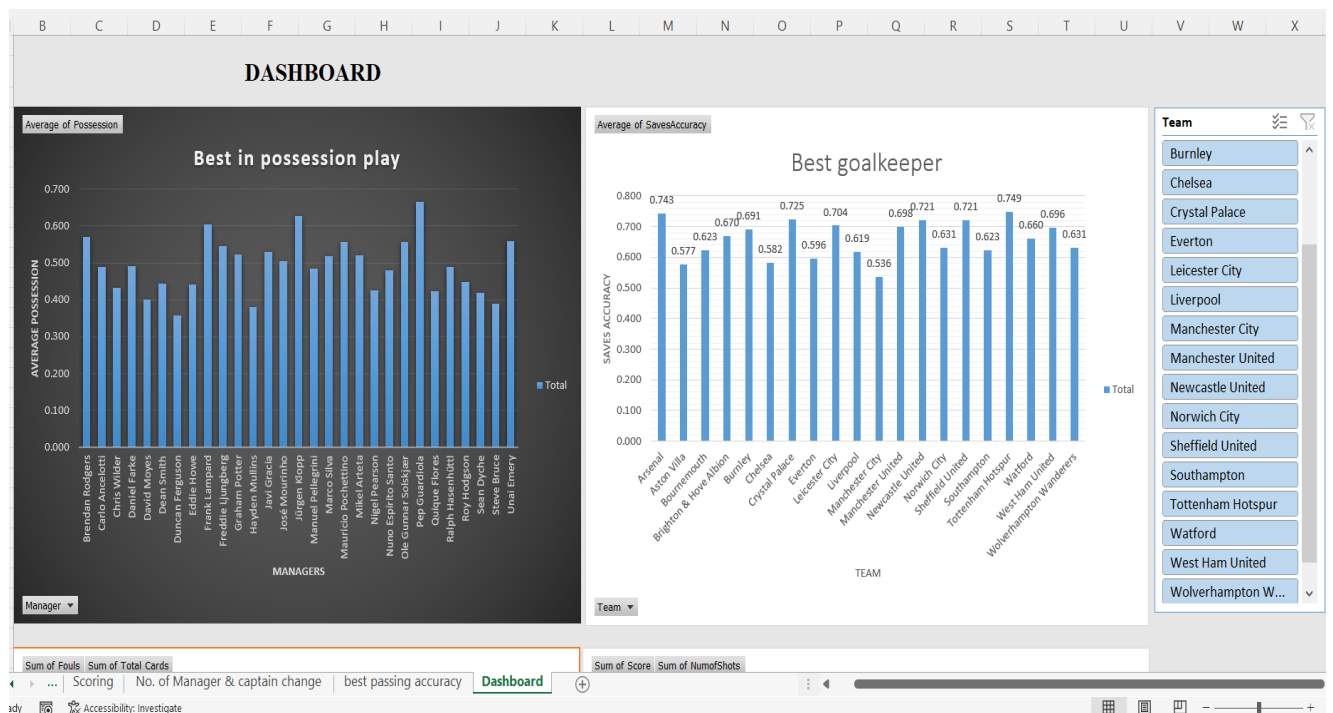
- Tottenham Hotspur changes 6 captains which is most by any team. Arsenal and Norwich city are at 2nd in changing number of captains which is 5 in number.

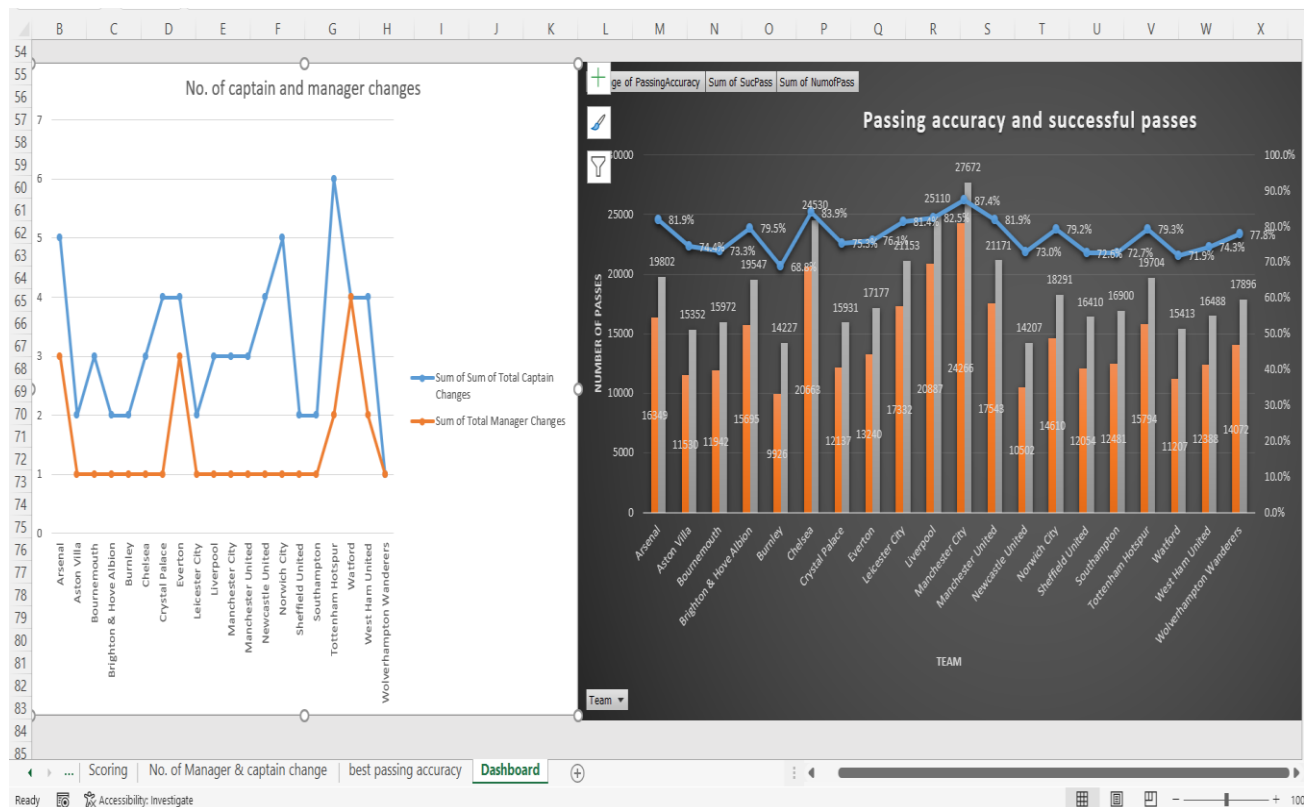
6. Which team have highest successful passes and accuracy per game?

- From our analysis, we would be able to find that Manchester City is the best in passing accuracy with 87.4% average passing accuracy per game. Chelsea is the second-best team in passing accuracy with 83.9% average passing accuracy per game. Both the team are also at 1st and 2nd in highest number of successful passes.

Final Dashboard

Dashboard Will be giving all the brief details of the of work done on the project.





Future Scope

The Premier League is all about the football

Watching fantastic players and exciting matches creates the support that drives the broadcast and the commercial interest that ensures we can re-invest in the game, both domestically and globally.

The principal objective of the Premier League is to stage the most competitive and compelling league with world-class players and, through the equitable distribution of broadcast and commercial revenues, to enable clubs to develop so that European competition is a realistic aim and, once there, they are playing at a level where they can compete effectively.

Many of the most famous clubs in world football play in the Premier League and, thanks to our distribution model, which is the most equitable in top-flight European football, the League is incredibly competitive, unpredictable, and exciting.

In recent seasons, the battles for the title, UEFA Champions League places and to avoid relegation have been the tightest and hardest fought on record.

The Premier League is the world's league. It has become a global phenomenon, inspiring passion, and emotion from Los Angeles to Lagos, from Macau to Melbourne.

Premier league is successful over the years and continually growing year by year. It also wants to continue their legacy in future also.

Reference and Bibliography

- YouTube
- <https://support.microsoft.com/en-us/excel>
- Kaggle
- Wikipedia.com
- www.premierleague.com