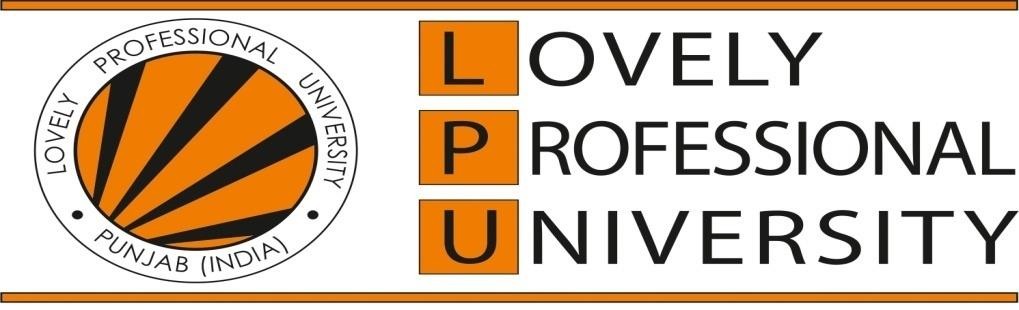
## DATA MANAGEMENT PROJECT REPORT

(Project Semester: September-December 2021)



**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)
   1. Introduction
   2. General Description
   3. Specific Requirements, functions, and formulas
   4. Analysis results
   5. 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 The league, which comprises 20 clubs, superseded the first division of the English Football League (EFL) as the top level of football in England.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.[[](https://en.wikipedia.org/wiki/Premier_League#cite_note-12)  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 off 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, two18-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: Passing accuracy(SucPass/NumofPass)
* SucPass: Total number of successful passes
* NumofPass: Total number of passes
* ShotsAccuracy: Shot accuracy(SucShots/NumofShots)
* SucShots: Total successful shots
* NumofShots: Total number of shots
* SavesAccuracy: Saves accuracy(SucSaves/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**](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.**

**Graphical user interface, text, application

Description automatically generated**

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.

Graphical user interface, application, table, Excel

Description automatically generated

### 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.

Graphical user interface, application, table, Excel

Description automatically generated

### 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.

Graphical user interface, application, table, Excel

Description automatically generated

### 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.

Graphical user interface, application, table, Excel

Description automatically generated

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

Graphical user interface, application, table, Excel

Description automatically generated

# 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:

Graphical user interface, application, table, Excel

Description automatically generated

**Visualization:**

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

Graphical user interface, application, table, Excel

Description automatically generated

1. 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:

**Graphical user interface, chart, bar chart

Description automatically generated**

**Visualization:**

We use the bar graph to visualize the distribution.

Graphical user interface, chart, bar chart

Description automatically generated

1. 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:**

**Graphical user interface, application, table, Excel

Description automatically generated**

Graphical user interface, application, table, Excel

Description automatically generated

### Visualization:

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

Graphical user interface, application

Description automatically generated

1. 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:**

Graphical user interface, application, table, Excel

Description automatically generated

### Visualization:

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

Graphical user interface, application, table, Excel

Description automatically generated

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

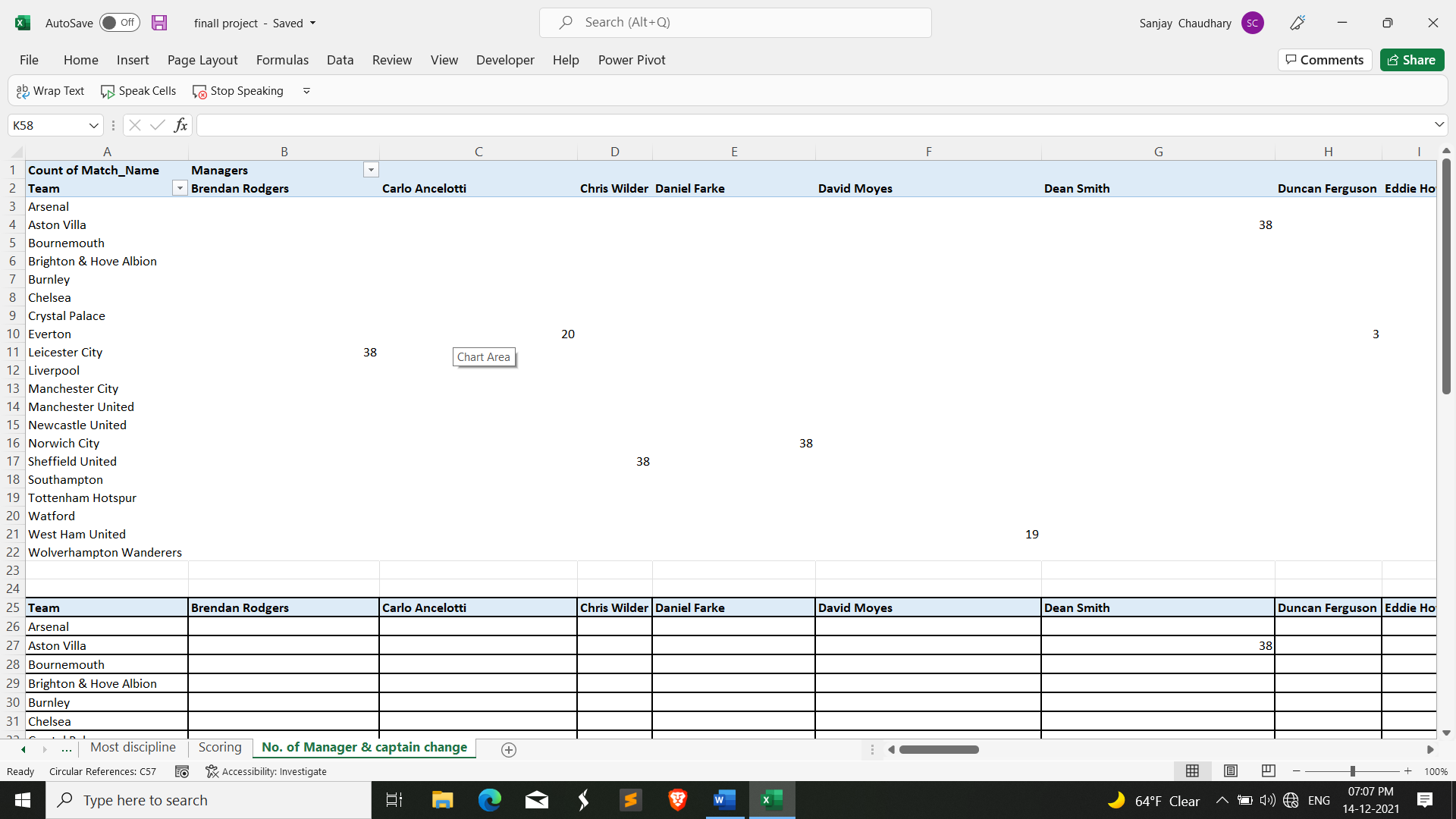
**Description:**

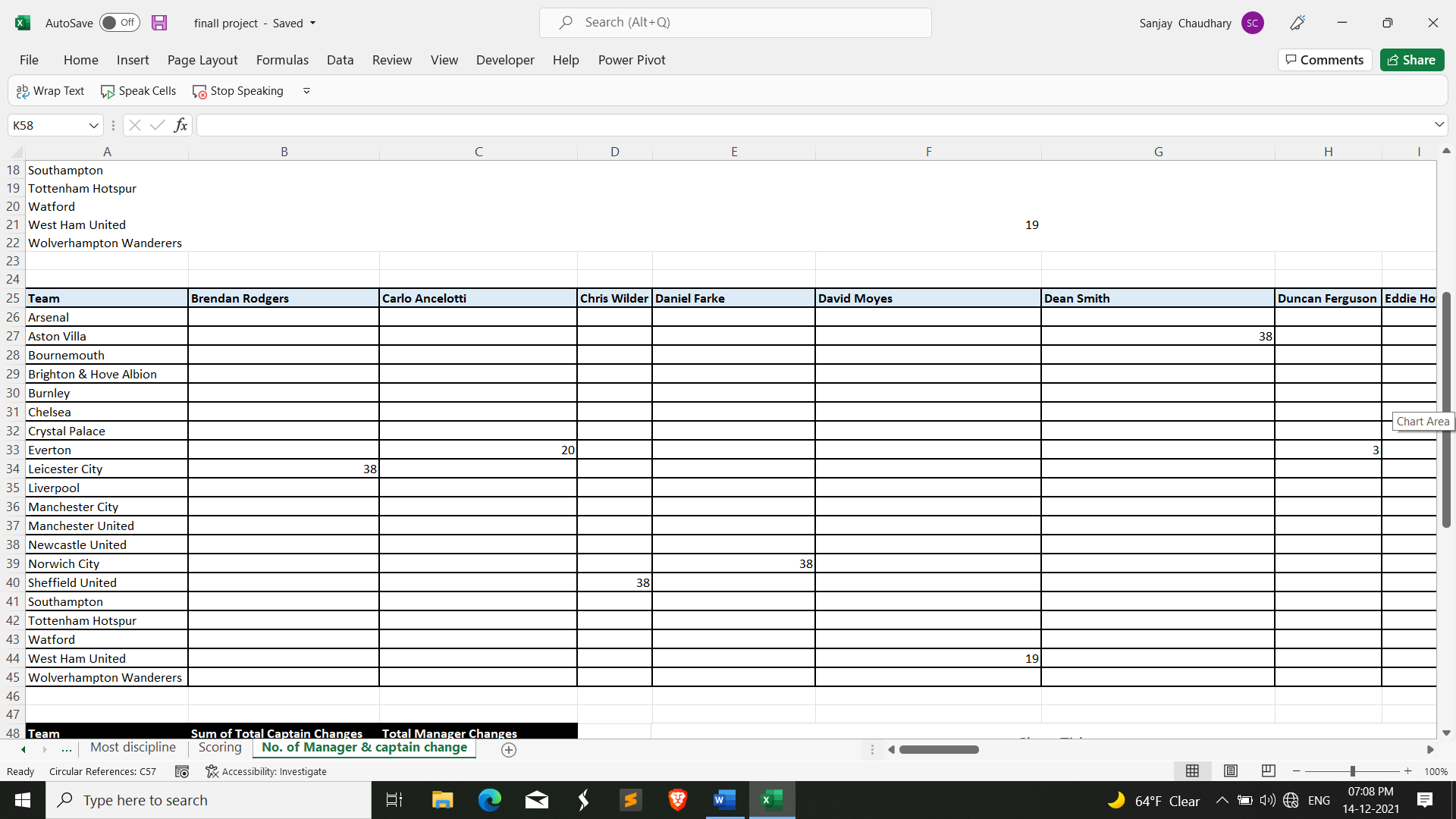
In this objective, we need to count how many times changes a team changes his manager and captain throughout the season.

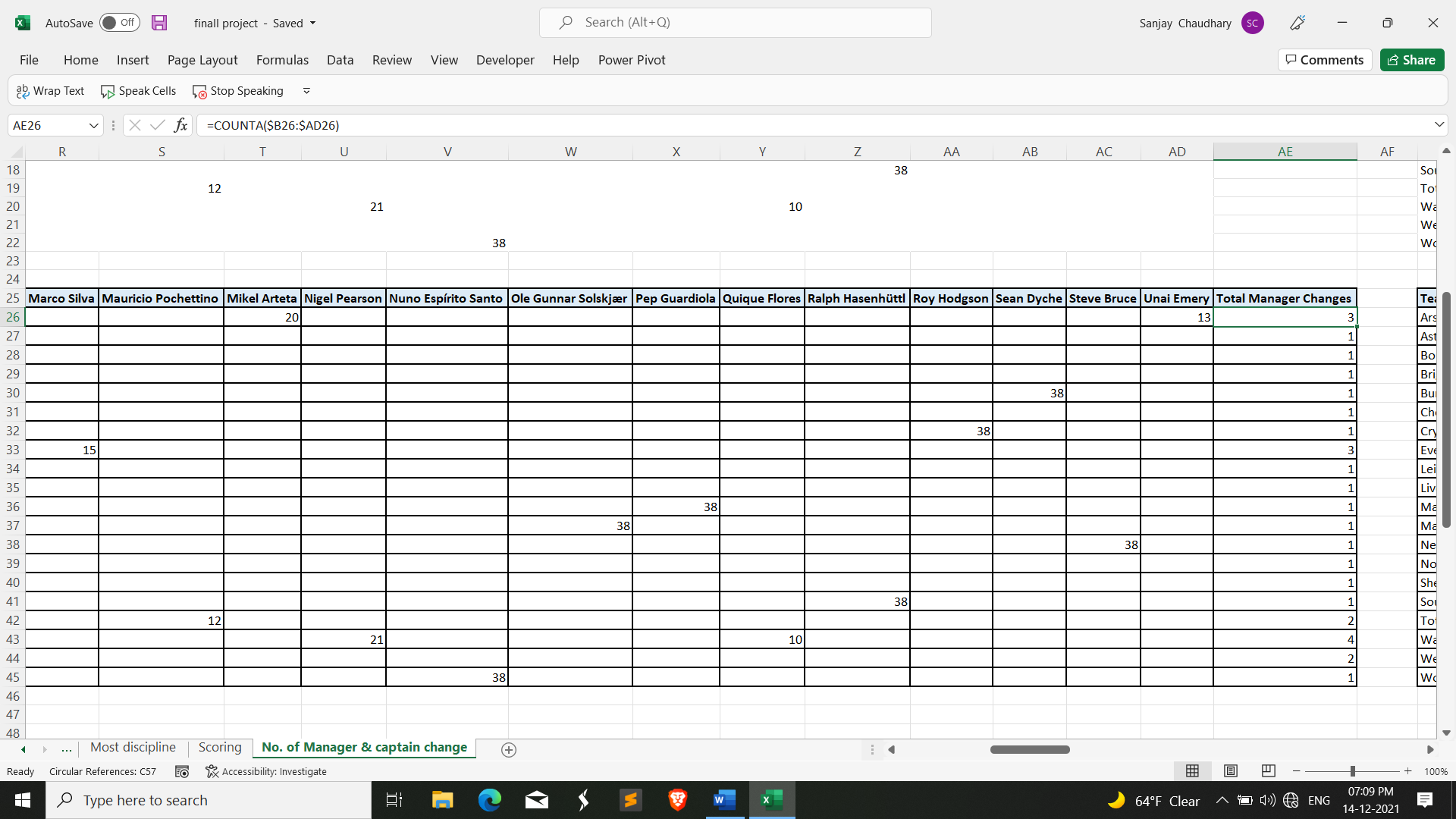
**Specific function and requirements:**

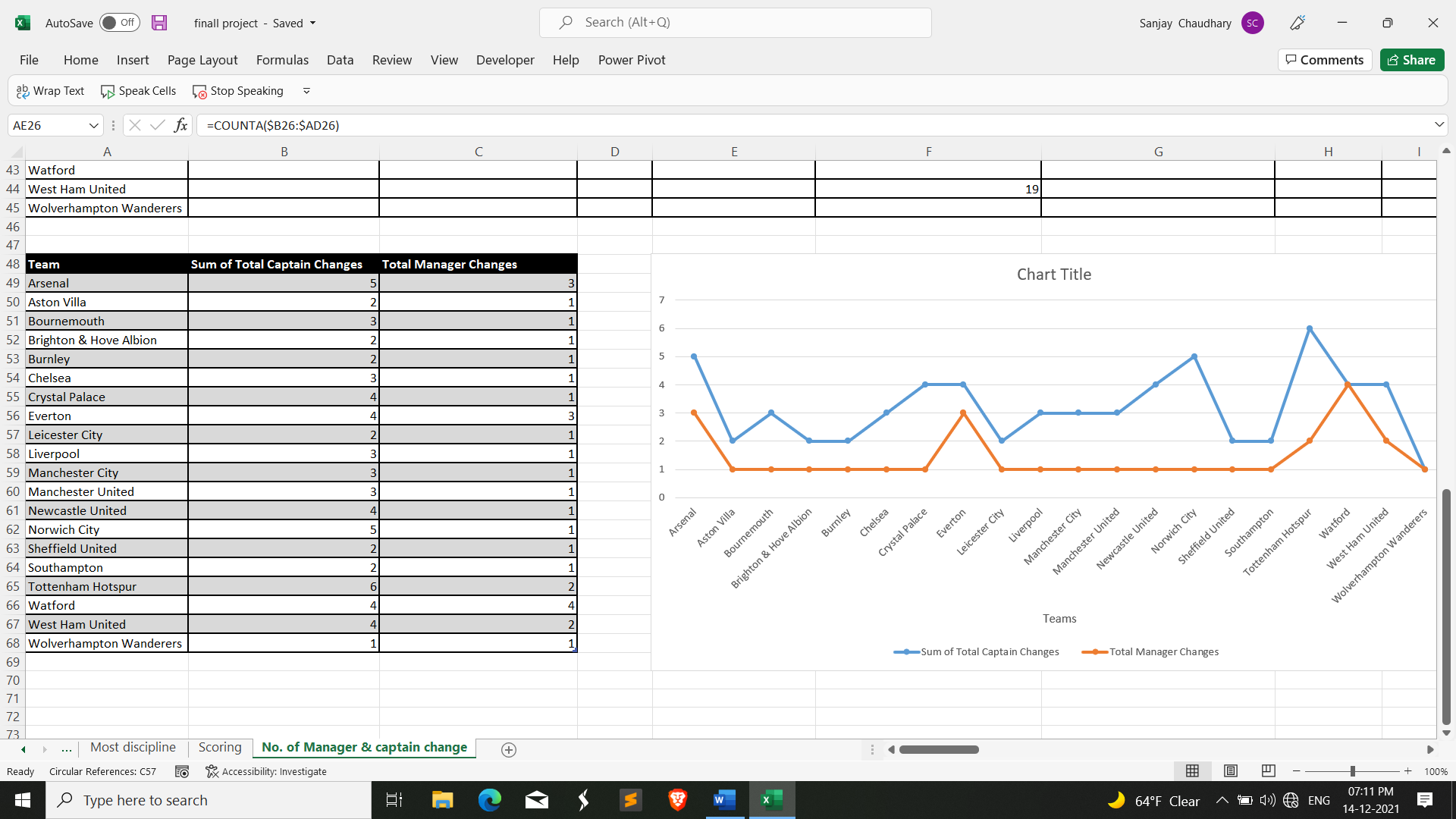
For calculating number of times, a team changes his manager and captain, we need data from team, manager , captain, match\_name column. We create a pivot table with team with row label and manager with column label. Count the number of games played by team with specific manager. Then copy the pivot table 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:**

****

****

****

****

**Visualization:**

The results are visualized in the form of line graph.



1. 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:**

**Graphical user interface, chart, application

Description automatically generated**

### 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.

Graphical user interface, chart, application, Excel

Description automatically generated

**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.

1. 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%.

1. 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.

1. 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.

1. 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 andQuique 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.

1. 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

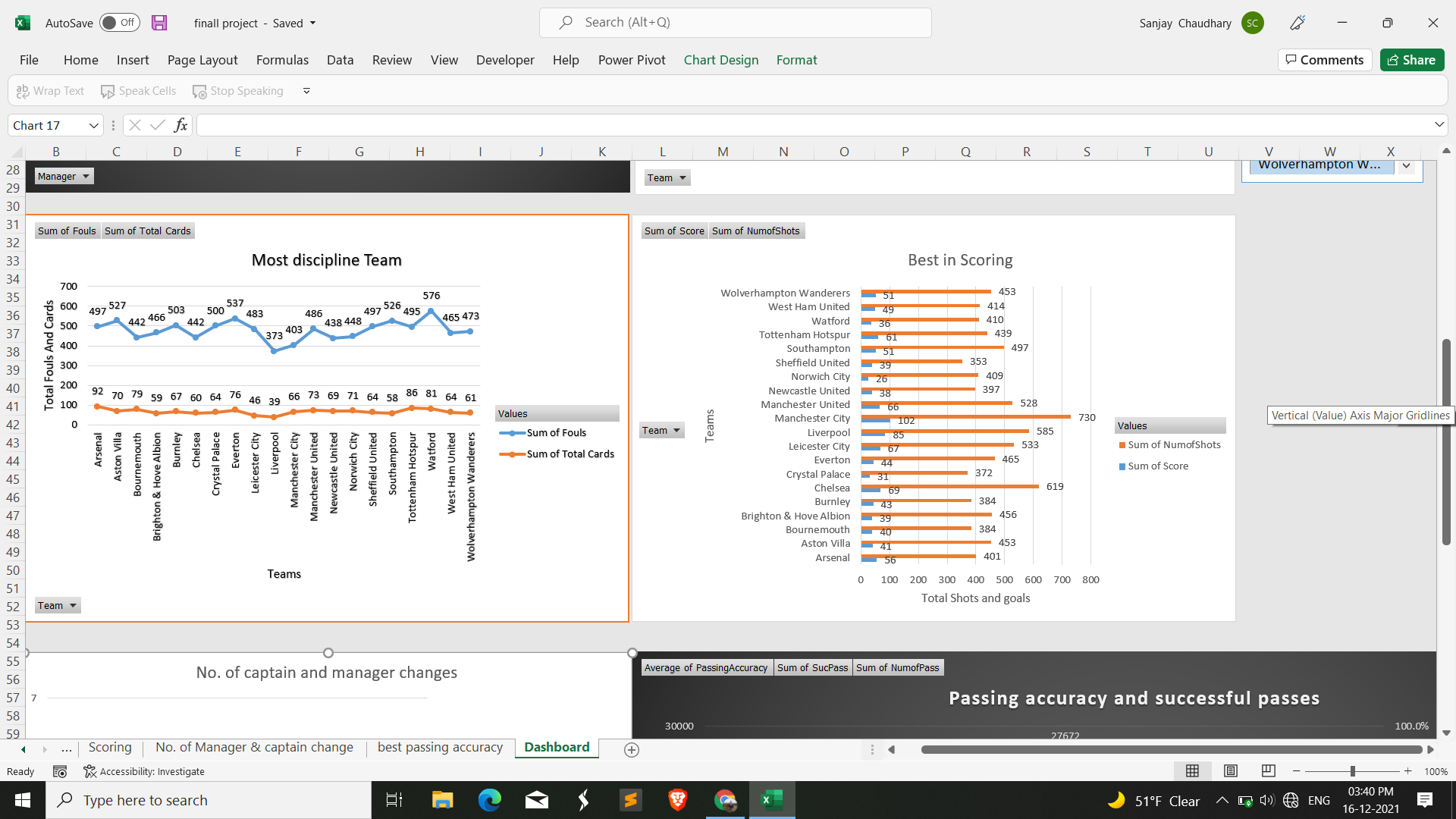
are also at 1st and 2nd in highest number of successful passes.

# Final Dashboard

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

Graphical user interface, application

Description automatically generated



Graphical user interface, application

Description automatically generated

**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