SCHOOL OF COMPUTING, ENGINEERING AND DIGITAL TECHNOLOGIES

MSC. DATA SCIENCE WITH ADVANCED PRACTICE

BAAH NKANSAH KWAKU

C2082617

BIG DATA AND BUSINESS INTELLIGENCE

CIS 4008-N

ENGLISH PREMIER LEAGUE 2020/21 ANALYSIS

SECTION 1

BUSINESS INTELLIGENCE REPORT

JANUARY 10, 2023

Contents

Executive Summary

Following the establishment and inauguration of the premier league in the 1992/93 season, the English premier league and the clubs who have participated in the league have attained significant and remarkable success over the years while establishing a formidable global reputation for the best standard and highly entertaining style of football. Conclusively, it could be said that the English premier league 2020/21 season was a defensive season more than an offensive and attacking one and this can be justified by making a comparative analysis between the total number of cards (1070) and the total number of assists (648) and the total number of goals scored (940) as well. Team management and coaches should ensure adequate number of trainings are done particularly by the team players and goalkeepers who are in the bottom half of the league.

Figure 1: card showing total assists

Figure 2: card showing total goals

Also, it could be generally asserted and assumed that players could be said to be in their prime at age twenty-seven (27) and age twenty-eight (28) because players ranging between these ages accounted for the most goals scored as well. Presumably, it could be assumed that the reason

for English players being the most in the English premier league is because it is a league that is being played in their home country although they are very high performing, hardworking and talented professionals who contribute immensely to the performance of their respective teams hence, they do have an upper hand in terms of getting easily scouted than other players who play in different countries.

Data source description and BI Requirements.

Introduction

Established and started in 1992, the English premier league is known to be the topmost domestic football league in the world, having the most cost-effective and high-income broadcasting right deals globally. The league is played annually by 20 teams who face each other twice within a calendar season, that is one home and one away fixture while having the top four on the league table qualify and compete to play in the UEFA champions league. The 2020/21 premier league season was a special, unpredictable and challenging season due to a global pandemic that the world faced, that is the Covid-19 pandemic. Despite all the challenges and hurdles, the teams and players and the premier league as a franchise provided fans and supporters with a fun, entertaining and action-packed season.

This BI project is tailored towards analysing the English premier league 2020/21 season and evaluating how the season played out and analysing club performance as well as player performance and involvement and contributions to a challenging but competitive and fun season.

Data Source and Description

The dataset used for this project is the English premier league 2020/21 which was obtained from Kaggle, a credible and widely accepted website by data analyst and professionals globally: https://www.kaggle.com/datasets/rajatrc1705/english-premier-league202021

The dataset contains one table with eighteen (18) columns and five hundred and thirty-three (533) rows which contains information on crucial stats about the English premier league 2020/21 edition. The dataset contains information on all the players that played in the English premier league and basic and standard stats about these players like the goals they scored, the assists they made, their expected goals and expected assists as well, passes attempted, pass accuracy and the teams these players play for as well as their nationality.

Column name	Description

Name	Name of the player
Club	Club the player plays for
Position	The position a player plays regularly
Nationality	The country the player comes from.
Age	The age of the player
Matches	The number of matches played by each player.
Starts	The number of times the player was named in the starting eleven (11).
Mins	The number of minutes played by the player.
Goals	The number of goals scored by the player
Assists	The number of times the player has assisted other players to score.
Passes attempted	The number of passes attempted by the player
Perc Passes Completed	The number of times that the player accurately passed to his teammate
Penalty goals	The number of penalties scored by the player
Penalty attempted	The number of penalties played by the player
xG	Expected number of goals from the player in a match
xA	Expected number of assists from the player in a match
Yellow cards	The number of yellow cards a player got for indiscipline, technical fouls, or other minor fouls
Red cards	The number of red cards a player got for accumulating two yellow cards or one major foul

Table: English premier league 2020/21 columns and descriptions.

Screenshots of the dataset are presented below:

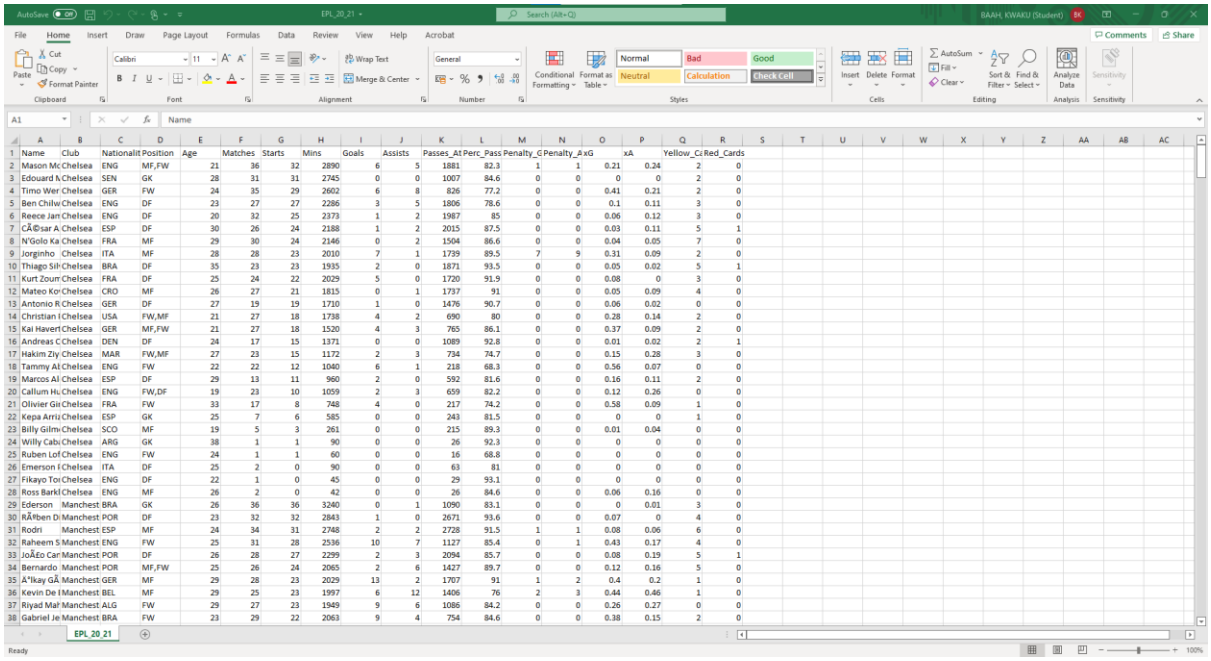

Figure 3: English premier league dataset

Rationale for choosing this dataset

This dataset was chosen for several reason but most importantly it was chosen for the following reasons:

- Challenging of skillsets and knowledge acquired: This dataset was selected because of how it challenged my knowledge acquired over the period of my studies and how it also put my acquired skillsets to the test. The dataset pushed me to limits because it tested my data cleaning and pre-processing skills as well my data analytics skills and most importantly how to analyse big data problems.
- Demonstration of data analytics skills: Another reason why this dataset was chosen was because the size of the dataset that is the number of columns and rows as well as the number of missing values aided in the demonstration of data analytics skills acquired so far. Also, it helped in developing my analytical skills using DAX and M language as well that is, using DAX to create custom columns and new measures.
- Credibility of the source of the dataset: The source of the dataset that is Kaggle, is a globally accepted website, that is acknowledged by millions and

professionals worldwide and in active use in academia and industry as well. So, the credibility of the source of the dataset played a major in the selection of this dataset.

- Interest and passion for subject topic and the accuracy of the dataset due to how recent it is: The interest and passion that I have for football also played a huge role in the selection of this dataset because I am a huge football fan and always take the opportunity to look further and analyse anything that has to do with football, so this played a role. And, how recent the dataset is also played a role. The dataset contains stats about the English premier league 2020/21 season. This means that it displays current information, and it is painting a picture of modern football.
- Easily understandable dataset: For any dataset to be analysed, the understanding of what the dataset entails is very essential to the analysis and addressing the problems to come up with solutions. The dataset gave sufficient information needed for the analysis and also had an extensive number of rows and columns which gave detailed insights and information as to what the whole dataset is about, and this contributed to making the dataset easier to analyse.

BI Requirements\Questions

The project aims to examine strategically the English premier league 2020/21 season while trying to unveil the relationship between the different characteristics of the football matches played and the teams' performances as well as their players contributions to the seasonal performance of the teams**.** The analysis is primarily focused on answering the following questions:

- At what age are players said to be in their prime based on the number of matches they play and their contribution in terms of goals and assists?
- Which teams had the best players in the 2020/21 season per their goals and assists?
- Which nationality had the most players playing in the English premier league 2020/21 season?
- Which positions contributed the most to their respective teams in terms of goals and assists?
- What metrics are used in determining who the best players were for the English premier league 2020/21 season?

Finding based on analysis and evaluation

A variety of different visualization were created, and they were all tailored towards answering the business questions asked earlier on in the report and these visualizations are presented below.

- **At what age are players said to be in their prime based on the number of matches they play and their contribution in terms of goals and assists?**

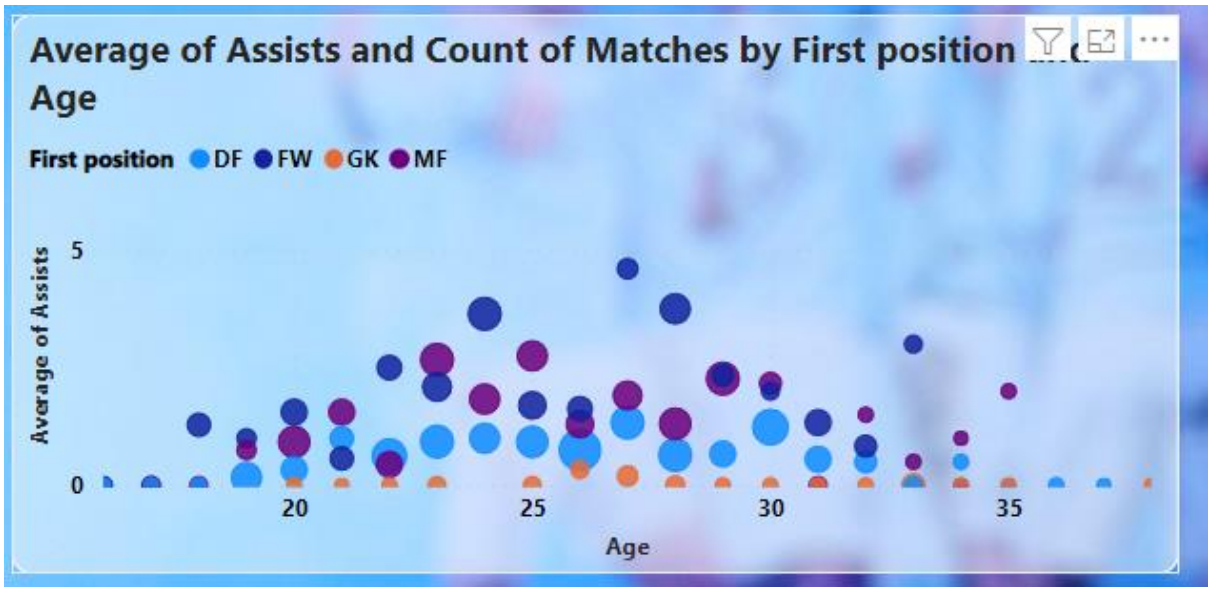

Figure 4: scatter plot from positions dashboard

A scatter chart showing at what age players average the most assist and get the most playing time with matches per the positions they play.

The assists, matches and age column of the EPL 20_21_22 table and positions column of the position column was used.

The scatter plot was used to represent this visual because it best represents two quantitative measures for different categories through the positioning of a single point and to determine if there is a linear relationship between the values.

- **Which teams had the best players in the 2020/21 season per their goals and assists?**

Figure 5: Bar chart for club performance analysis dashboard

A horizontal bar chart showing the number of goals that were scored by various clubs, this chart communicates the fact that Manchester city had the best players for the 2020/21 season judging from the number of goals they scored followed by Tottenham and Liverpool.

The goals column from the EPL_20_21_22 table was used as well as the club's column from the club table.

The bar chart was used because the chart is visualizing a data type that represents quantitative values against categories.

- **Which nationality had the most players playing in the English premier league 2020/21 season?**

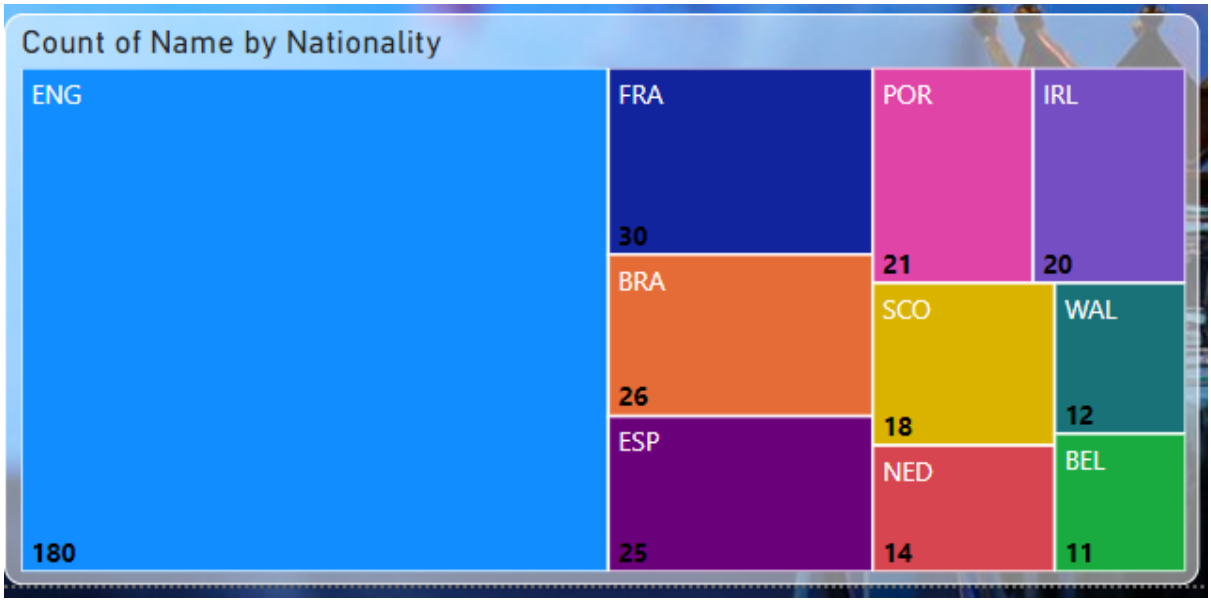

Figure 6: Treemap from Nationality analysis dashboard

A treemap representing the top ten (10) countries with the most players in the league for the 2020/21 season, and it can be inferred from the visualization that England has the most players in the league with one hundred and eighty (180) players.

The name column from the players table together with the nationality column from the nationality table were used in creating this chart.

The treemap was used in visualizing because the dataset represents quantitative values by displaying one constituent part of a whole. Colour has been used to indicate different categorical groupings.

- **Which positions contributed the most to their respective teams in terms of goals and assists?**

Figure 7: Stacked bar chart from the position analysis dashboard

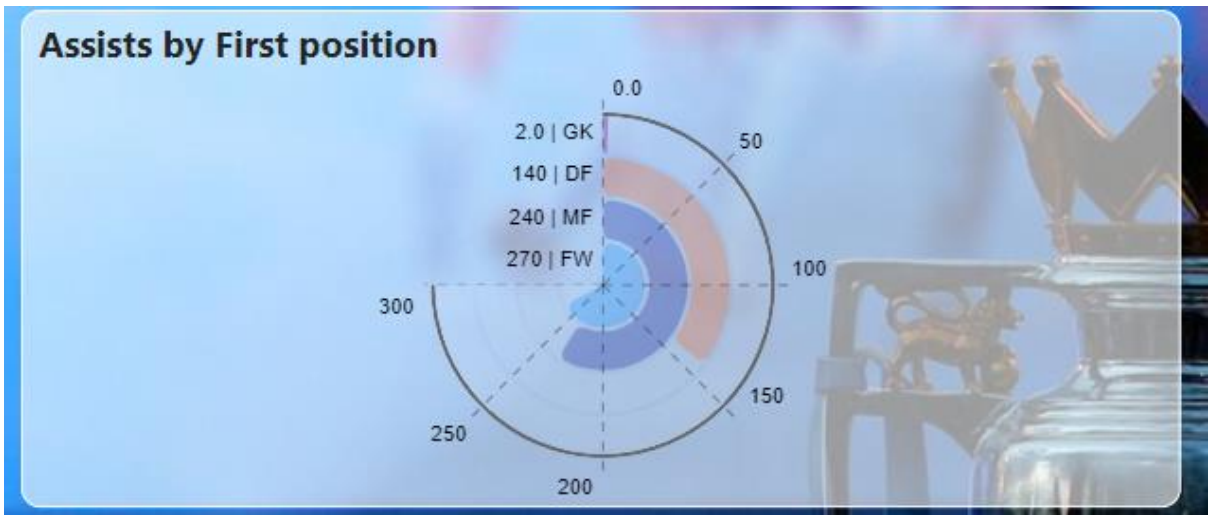

Figure 8:drilldown radial bar chart from positions analysis

Two charts that is the stacked bar chart and the drilldown radial bar chart were used in displaying this information which is communicating the positions that contributed the most to the performance of their respective teams through goals and assists.

The goals column from the EPL_20_21_22 table and the position column from the position table were used for the stacked bar chart. And the assists column from the EPL_20_21_22 table and the position column from the position table were used in the creation of this visualization.

The bar chart was used because the chart is visualizing a data type that represents quantitative values against categories.

Key Findings

- Manchester city had the highest assists at 55, while Sheffield united had the lowest assist at 133. And out of the overall assist's percentage, Manchester city accumulated 8.49% of the general assists, and the assists range between all 20 clubs was 13-55.
- English premier league 2020/21 season winners Manchester city had the most goals with eighty-two (8)2 goals which shows that goals are very crucial in winning the league at the end of the season.
- Sheffield united scored the least goals nineteen (19), which shows they won few games which eventually led to them getting relegated.
- One thousand and seventy goals (1070) yellow cards were issued to players throughout the season which indicated that it was a very aggressive season.
- Players aged 27 and below could be said to be the playmakers of the league because they had the most assist at the end of the season.
- Forwards had the highest number of goals accounting for 59.26% of the total goals scored followed by midfielders.
- England had the most players with one hundred and eighteen (180) players playing in the league and with the most players the English players could be said to be the most aggressive players because they accumulated the most cards three hundred and eighty-eight (388) for the season.

- English players accounted for the 34.80% of the total number of cards issued during the English premier league 2020/21 season.
- There are players from fifty-seven (57) countries who played in the English premier league 2020/21 season.

Recommendations.

- Data on dates could have been provided to allocate a more detailed analysis report. With data on game weeks and matchdays, the analysis of this project could have been drilled down to the number of wins and losses per the games played and a cumulative column could have been created for top half and bottom half of the league table with details on individual match days.
- Teams should focus more on being attacking if they want to win the league or remain in the league because football has evolved, and it has been transformed into an attacking sport rather than a defensive one.
- Team management and coaches should ensure adequate number of trainings are done particularly by the team players and goalkeepers who are in the bottom half of the league.
- Sheffield united and Fulham should endeavour to develop and focus more on attacking prospects to strengthen their goal scoring ratio.
- Metrics used in calculating the existing xA and xG do not favour outfield player that is goalkeepers and defenders hence it should be looked at and revised to accommodate all players of the league.

Conclusion

This project has really helped me in developing and gaining extra knowledge not only on powerBi, but it has also broadened my scope on data analytics, and it has helped me gain much insight and skills in data analytics and pre-processing. Because I did not have prior knowledge and much understanding before this project, but this project has helped me develop and practice many skills that I never knew before.

I encountered multiple problems and challenges right from creating of missing values and nulls values, to data cleaning/pre-processing, and then to creation of new columns and new measures using DAX expressions and M language. During the development and exploration of visualizations, I realized that there were blanks existing in some of my dimension tables when I used a slicer to visualize the dataset, but these blanks were not present when I looked at the dataset in the power query editor and the data view mode on power BI. I removed blanks from rows, removed duplicates and removed errors as well but the blanks were still present. I then went the extra mile to research and watch videos online on how to get rid of these blanks from my dataset, but all of these efforts did not have any effect on the blanks. This is the main and major challenge that I was faced with during the development of my project, and I was unable to rectify this problem. But with the help of my module leader and my lab tutor and some individual research through the internet and YouTube videos, I was able to go through and surpass all the challenges and I look forward to applying and the knowledge gained, and the skills acquired in the coming semesters and during my professional experience as an analyst as I know these skills acquired will go a long way to help me in the industry.

SCHOOL OF COMPUTING, ENGINEERING AND DIGITAL TECHNOLOGIES

MSC. DATA SCIENCE WITH ADVANCED PRACTICE

BAAH NKANSAH KWAKU

C2082617

BIG DATA AND BUSINESS INTELLIGENCE

CIS 4008-N

ENGLISH PREMIER LEAGUE 2020/21 ANALYSIS

APPENDIX: BUSINESS INTELLIGENCE DESIGN

SECTION 2

JANUARY 10, 2023

Data Pre-Processing or Data Cleansing

Loading the data

The first step in the data pre-processing and cleaning stage of this analysis is to load the dataset that was used for the project, in this case the English premier league dataset in Microsoft PowerBI. This can be done primarily in two ways which are, selecting the file type and format from "Get Data" menu and then going ahead to import it or to create a blank query and have M – Language imports the dataset. For the purpose of this project, the dataset was imported by selecting "Get Data".

The Microsoft powerBi tool launched and opened.

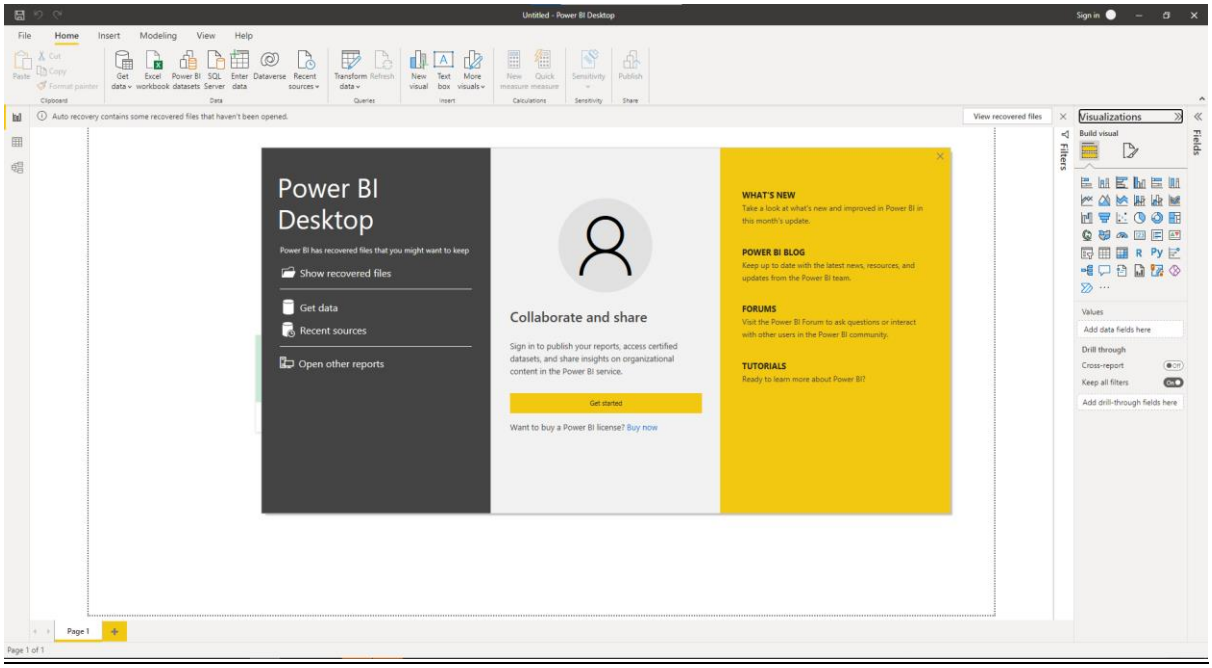

Figure 9: launching PowerBi tool

The data is loaded by clicking on "Get data" from the Home Tab of Microsoft powerBi. After clicking on "Get data", a dropdown containing different formats and

data sources appear. The dataset for this analysis is in the comma separated values (csv. files) format, hence select “Text/csv” from the dropdown that appears.

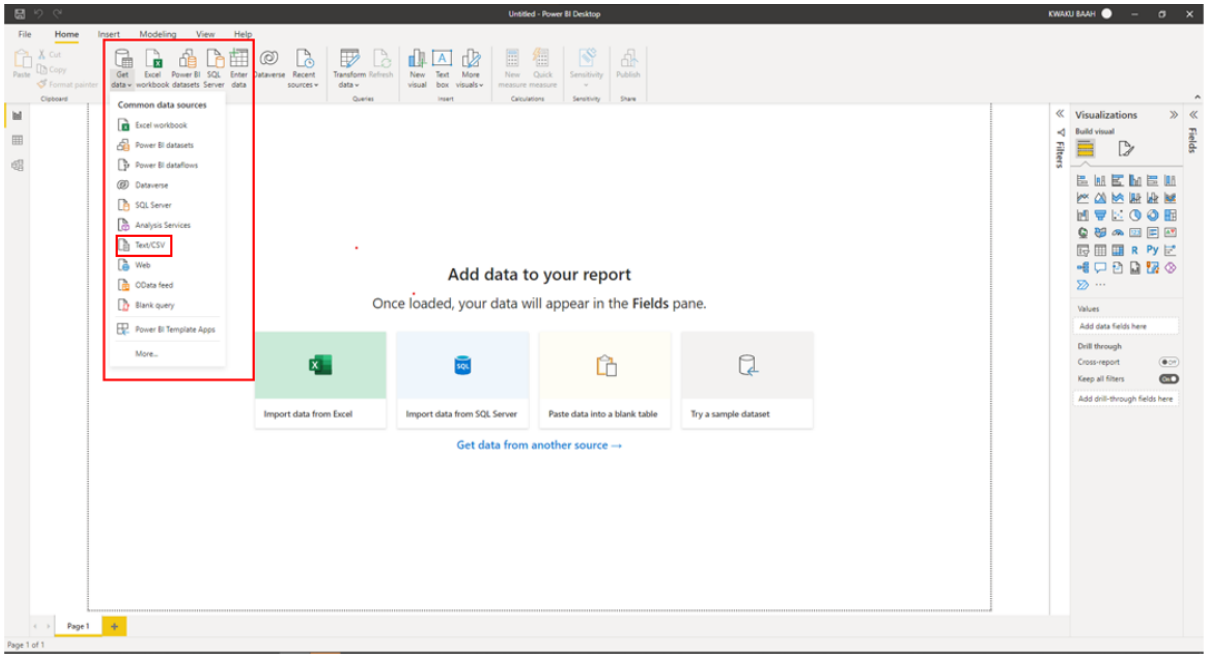

Figure 10: loading the dataset.

This opens a dialogue box that displays the English premier league data with the option to load, transform data or cancel.

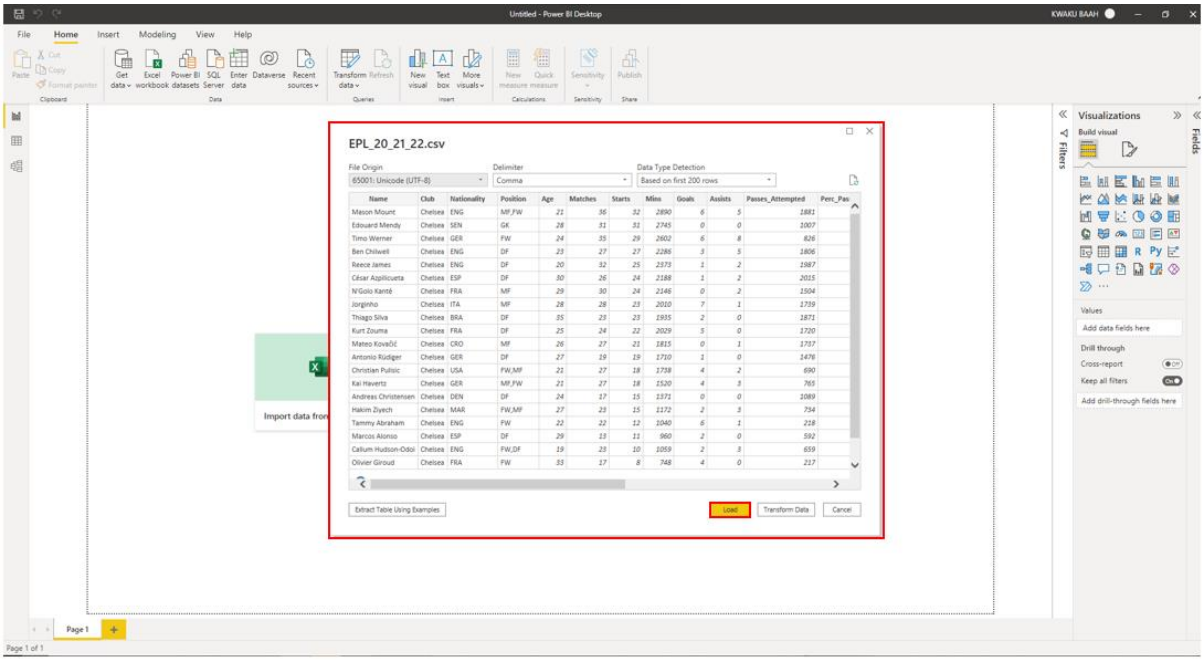

Figure 11: Get data dialogue box.

The process of loading the data was completed after the load option was selected in the previous dialogue box.

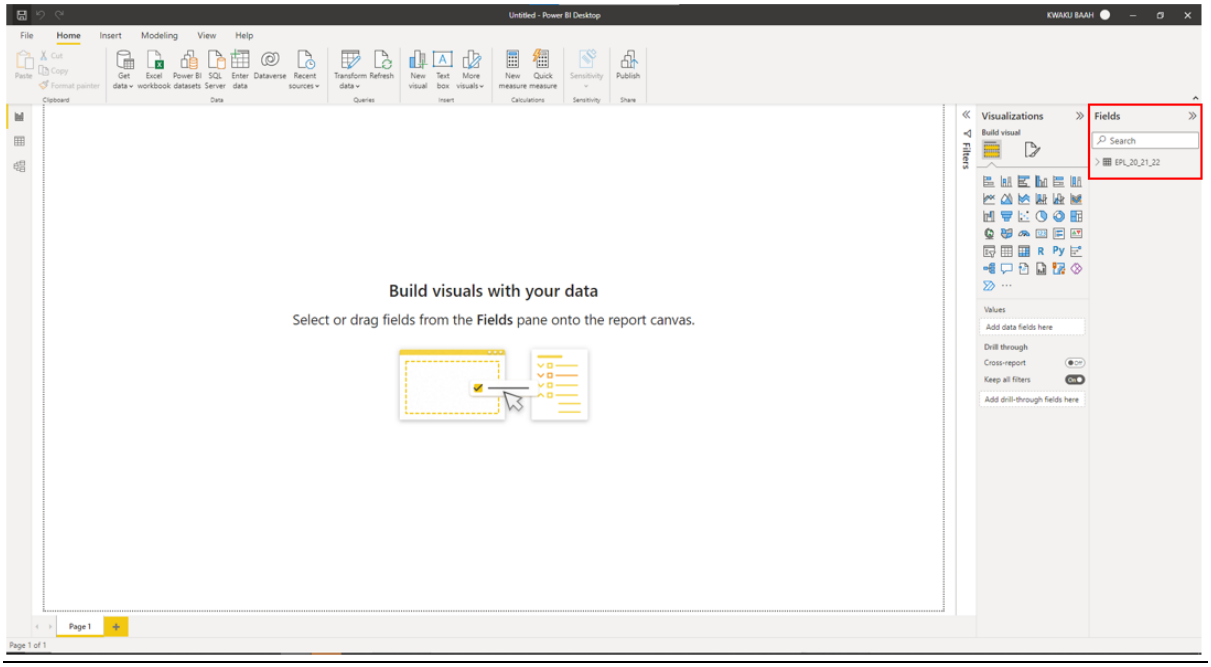

Figure 12: Dataset loaded

Data cleaning and Pre-processing

The next step is the cleaning and pre-processing of the dataset that was just imported and loaded into the Microsoft powerBi tool. The cleaning and pre-processing of the dataset is done using the power query editor. To get the power query editor, click on “Transform data” in the home Tab. Then select “Transform data” from the dropdown that displays

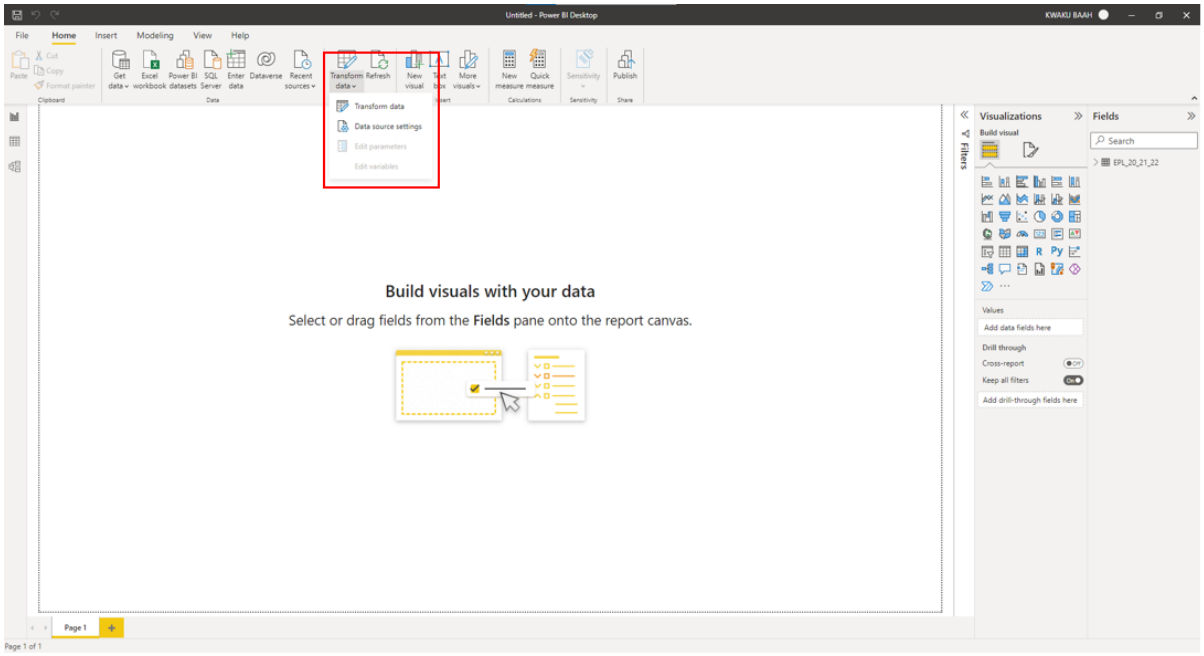

Figure 13: Transform data for power query editor

This opens the power query editor dialogue box.

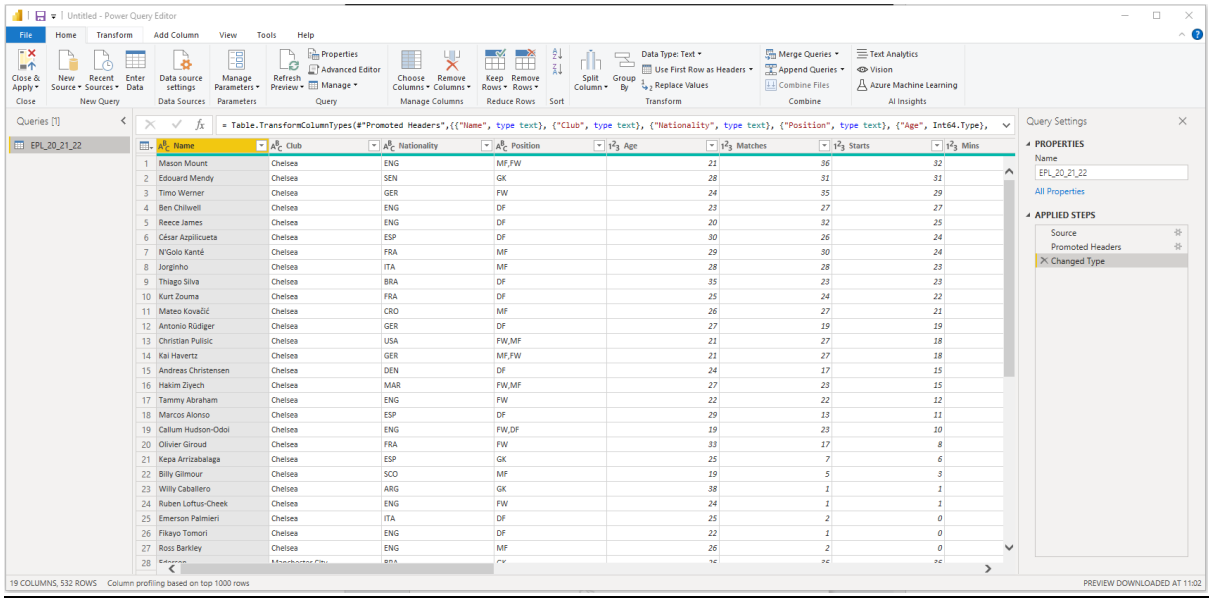

Figure 14: Power query editor dialogue box

Removing Errors/Nulls/Columns

The player ID column was contained errors as initially detected by powerBi. Eventually, the root cause of the error was revealed, and it was because the player ID column contained qualitative values when it should have contained quantitative values ideally and powerBi was unable to rectify these errors automatically.

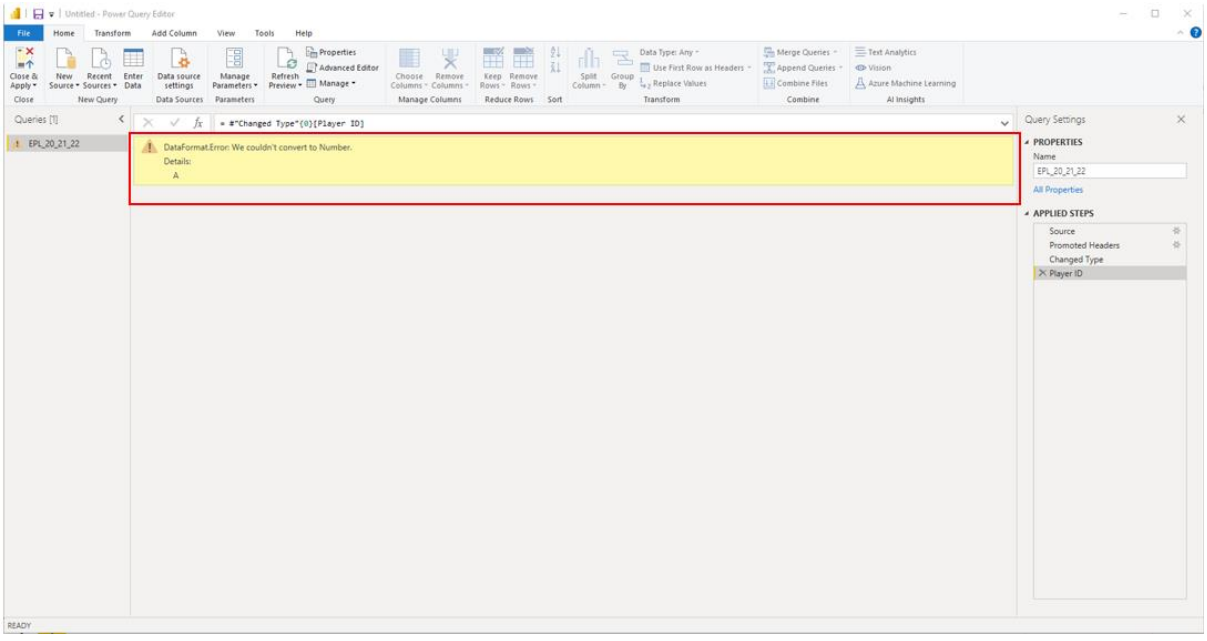

Figure 15: Errors detected by PowerBi.

The column was then removed from the dataset as the absence of this column from the dataset was not going to have a huge impact.

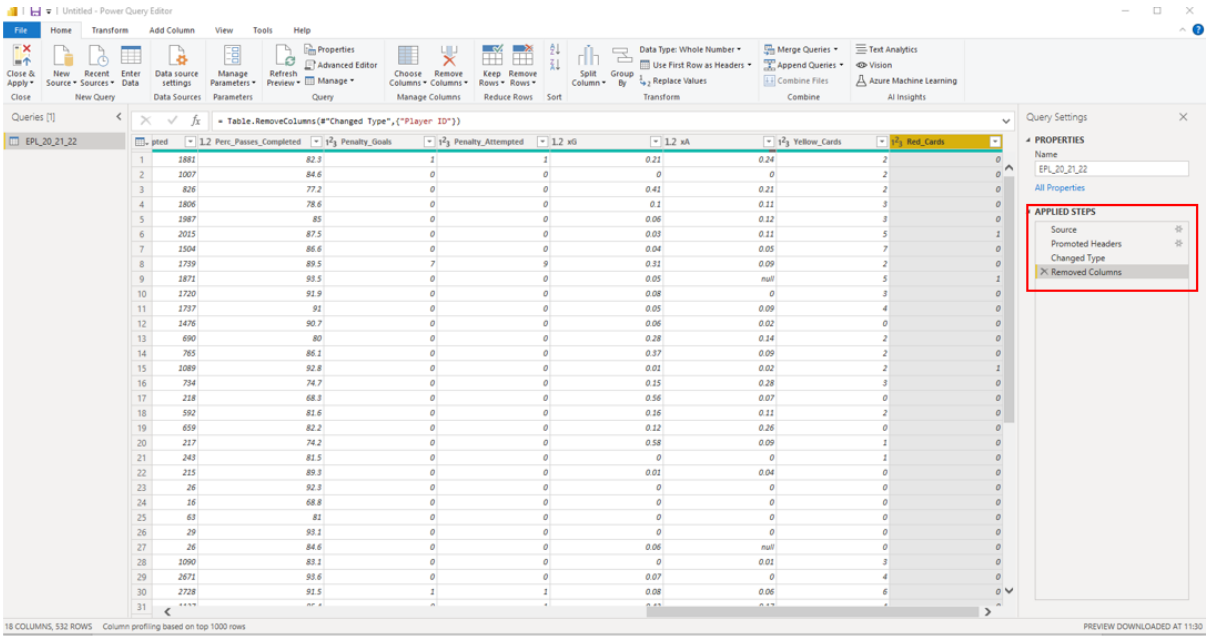

Figure 16: Removing columns

Empty rows as well as rows containing null values and blanks were also removed because the percentages that these null values and empty rows covered in the dataset was not marginal and they were not really essential for the analysis of this project and their presence in the dataset was not going to contribute marginally. Null values in the xA column were removed. This was done by right clicking on the arrow next to the xA column and unticking the “null” values from the dropdown that displayed.

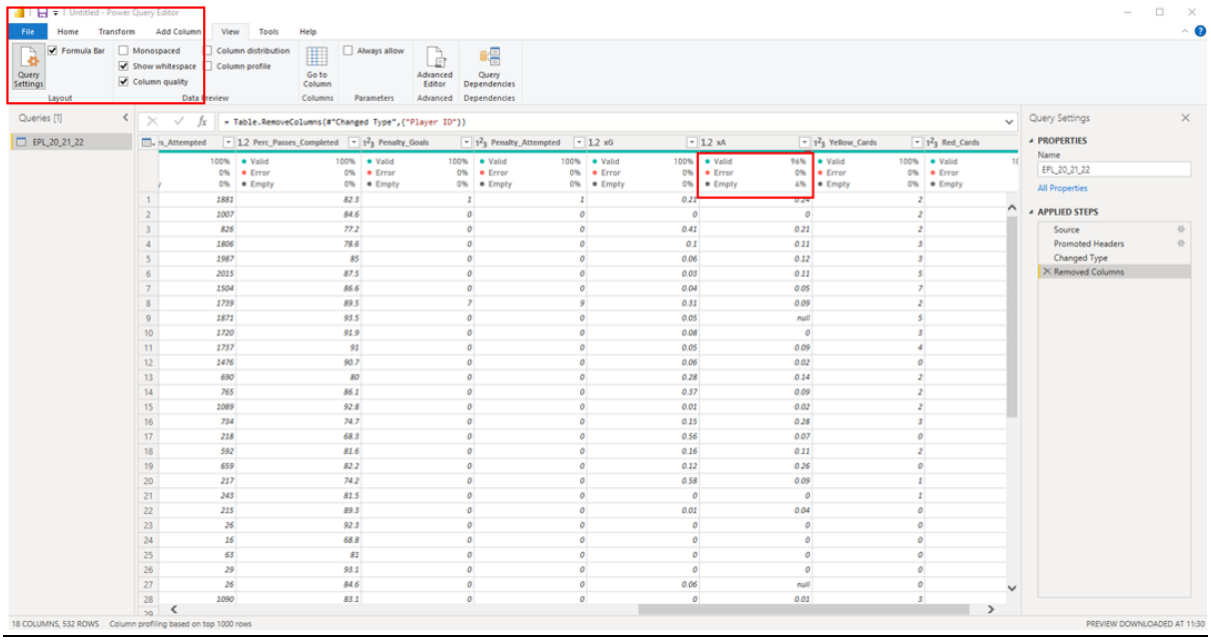

Figure 17: Removing blanks

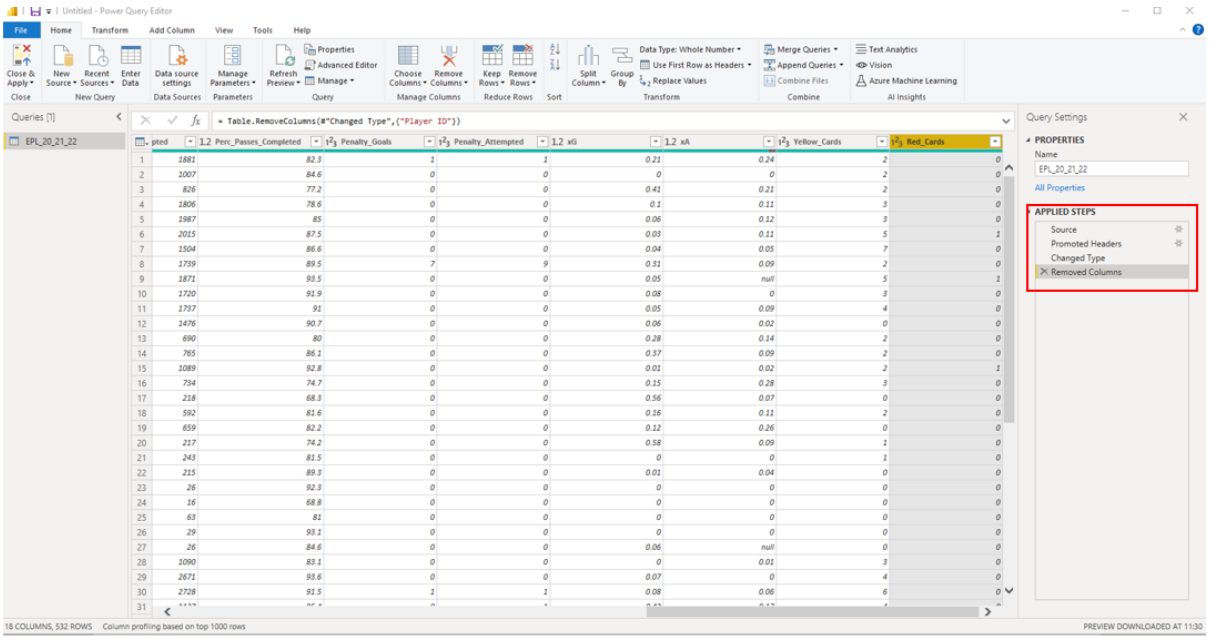

Figure 18: Errors removed

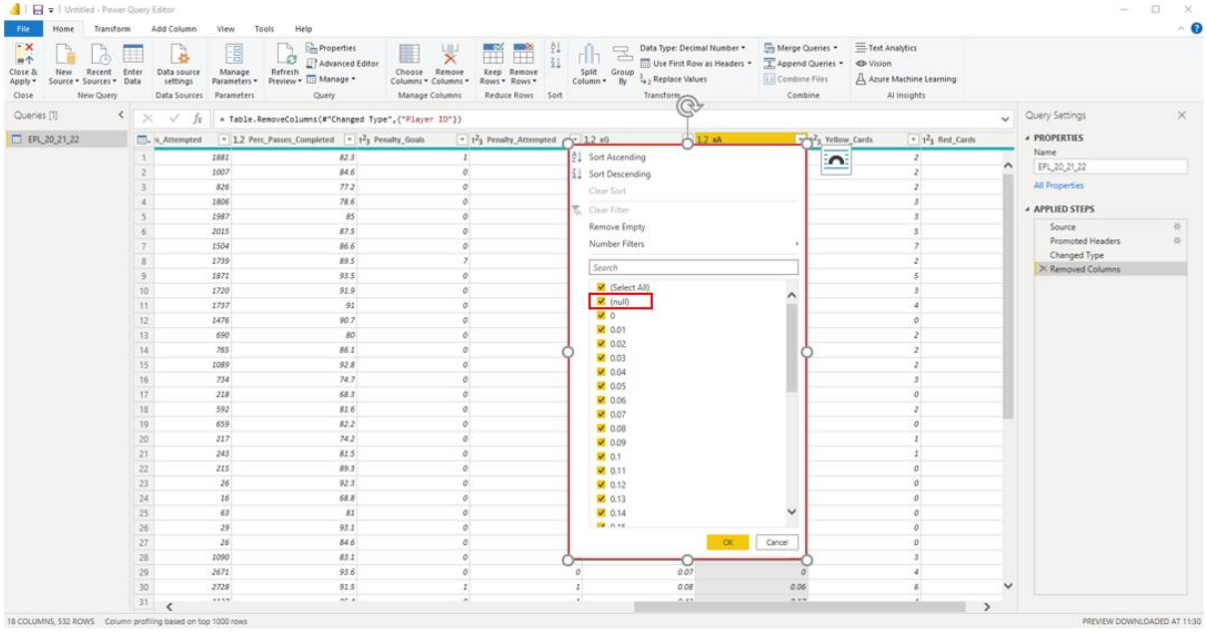

Figure 19: Unticking null values

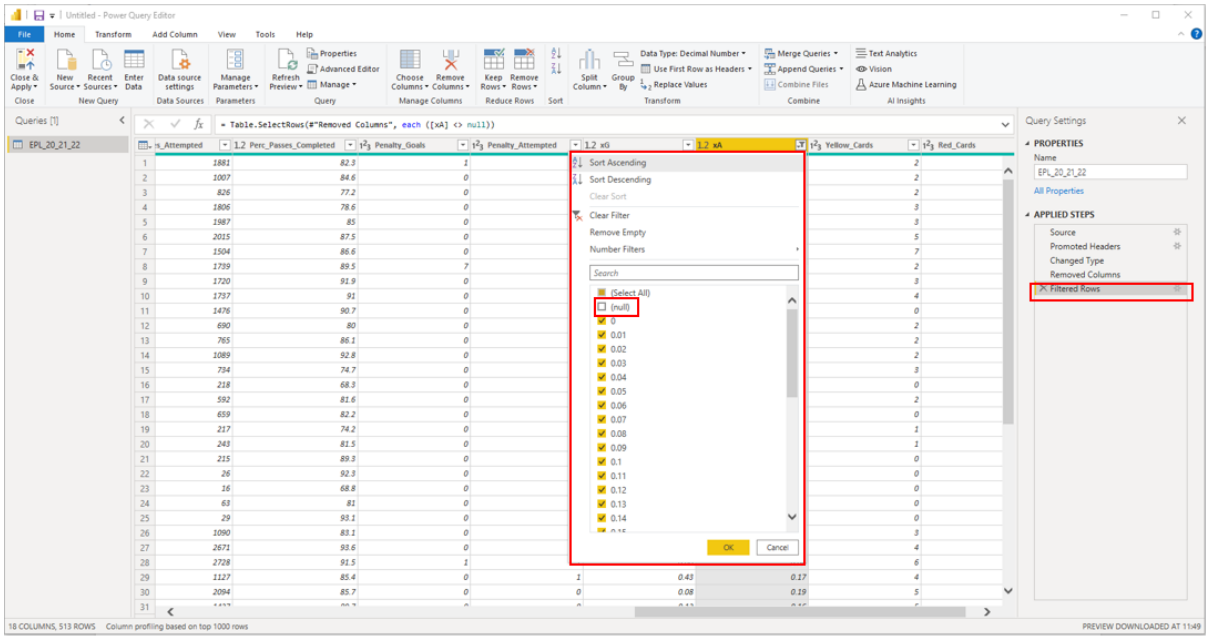

Figure 20: null values removed.

Creating and renaming columns

The index column was created and renamed as ID. This was done by clicking on “Add column” from the home tab and selecting Index column from the Add column tab and choosing “from 1” from the pop-up box that appears. No other columns were deleted as all the columns in the dataset were needed for the analysis

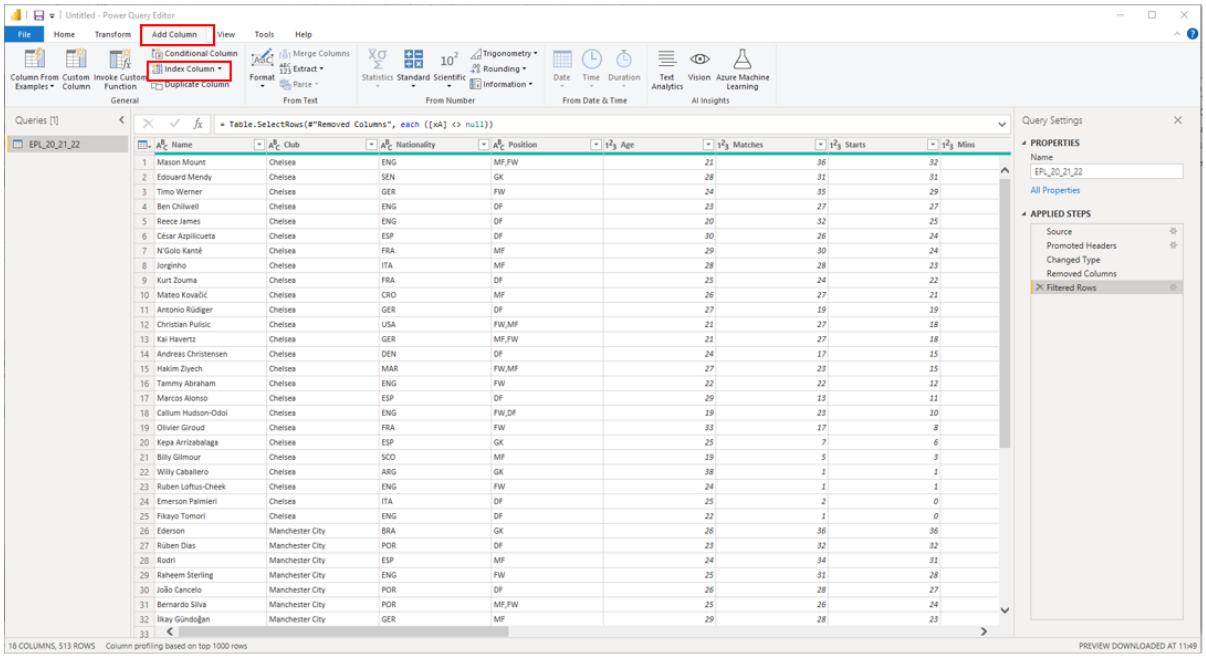

Figure 21: Adding index column

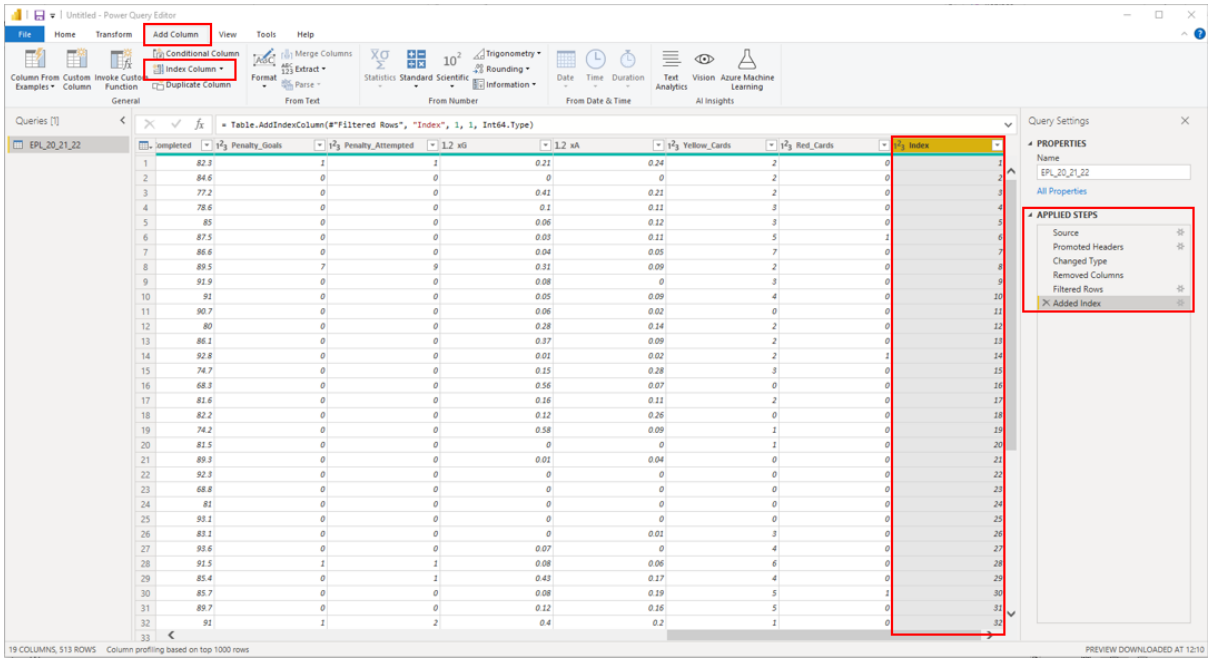

Figure 22: Index column added.

BI Data Modelling via Star Schema - Facts and Dimensions.

After the data cleaning and pre-processing had been done, the next step was to create dimensions tables and establish a relationship between the dimension tables and the fact table**.**

Creating Dimension Tables.

The club dimension table was the first-dimension table that was created. In order to create this, we reopened the power query editor, and duplicate the EPL 2020/21 table. After the table had been duplicated, the duplicated table was renamed “club”, only the club column was to be used for this table hence all the other columns were removed. This was done by selecting the club column and right clicking and choosing to remove other columns.

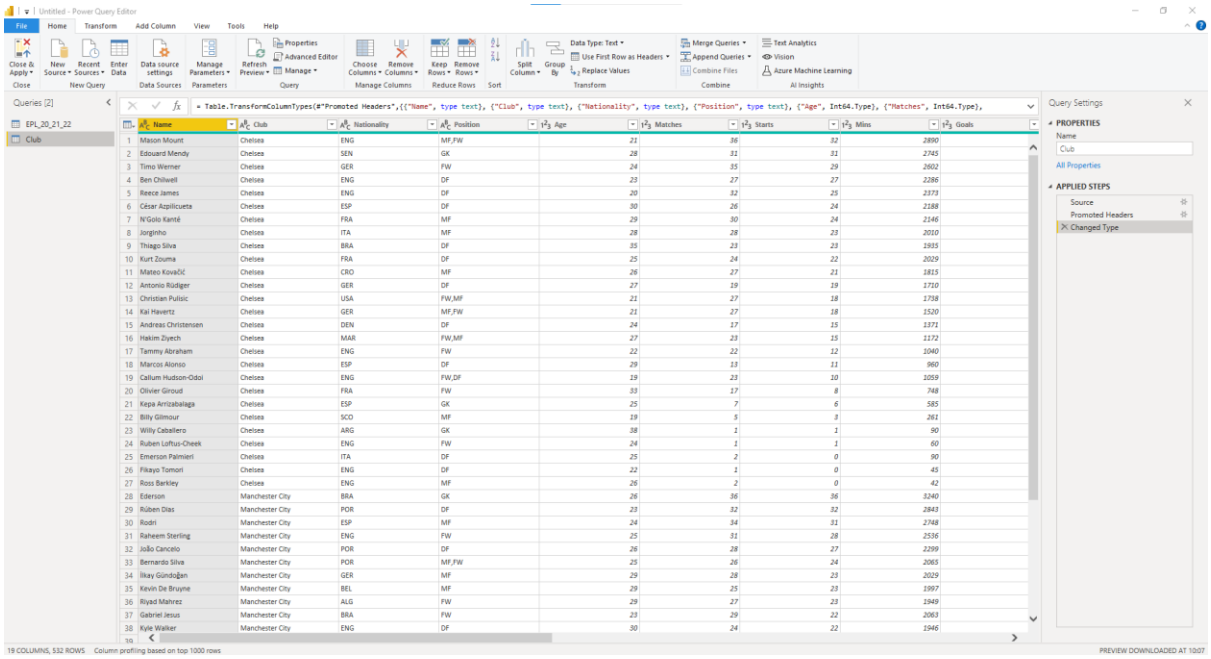

Figure 23: EPL 2020/21_22 table duplicated and renamed

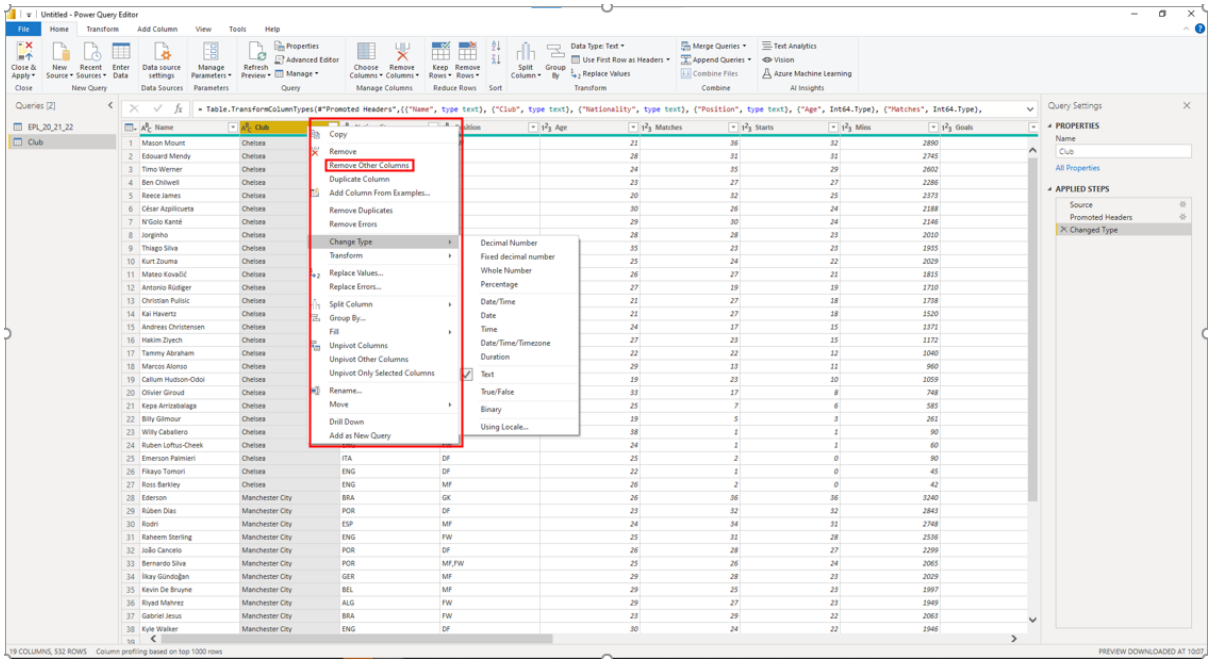

Figure 24: columns removed

The duplicates in the club column were removed by right clicking on club column and selecting “remove duplicates”.

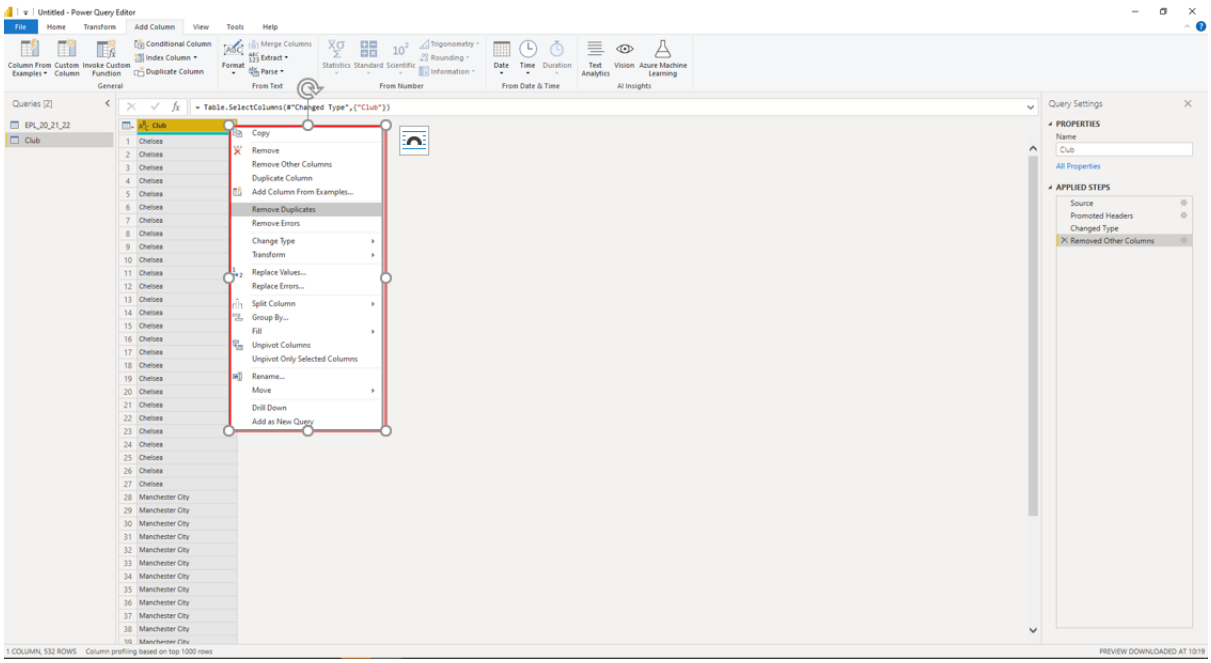

Figure 25: Duplicates removed from the club table

A custom column was created for club ID using the index column. This was done by clicking “Add column” and selecting Index column and choosing “from 1”. This new column was renamed Club ID.

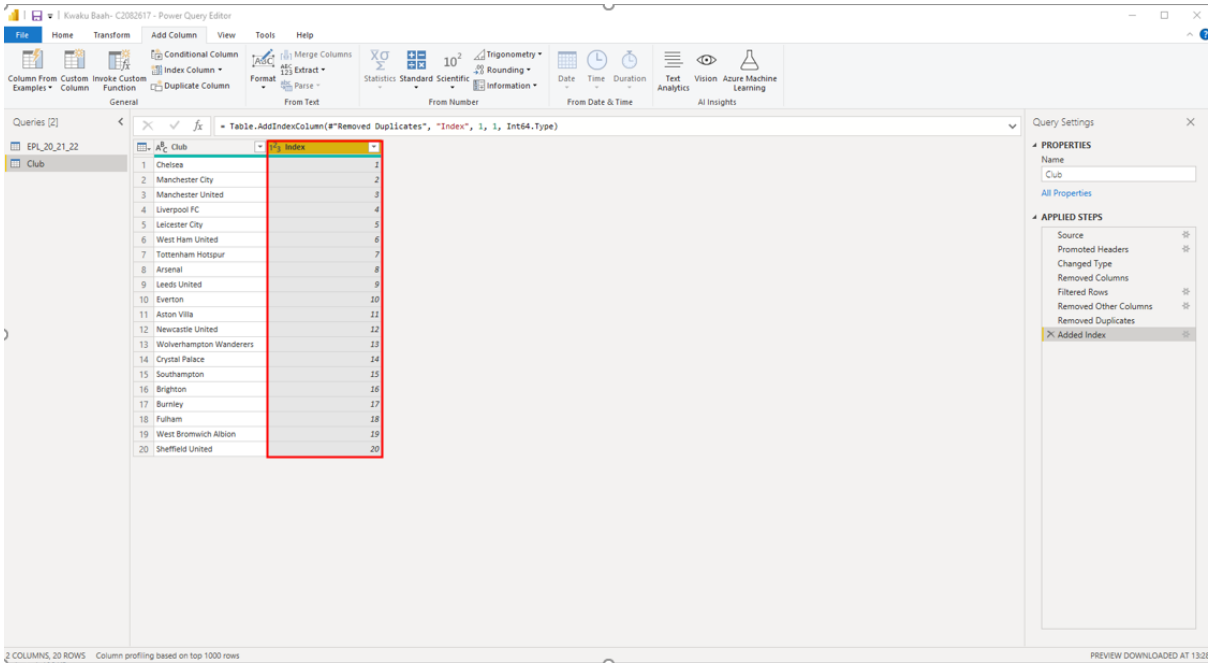

Figure 26: Index column created for club ID

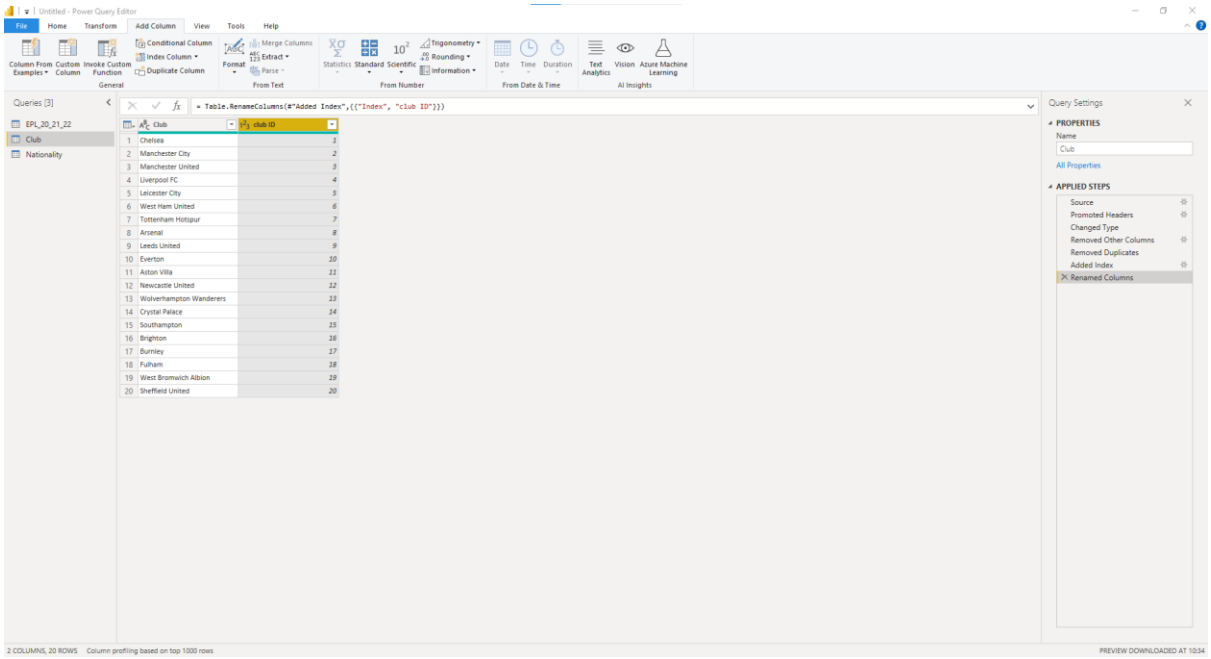

Figure 27: Index column renamed as club ID

The second-dimension table created was the nationality table. This was also done by duplicating the EPL 2020/21 table and renaming it “Nationality”. All columns were removed except for the nationality column because it was the only column needed for this table. This was done by selecting the nationality column and right clicking on it and selecting remove other columns. The duplicates in the nationality column were also removed, this was done by right clicking on the nationality column and selecting remove duplicates.

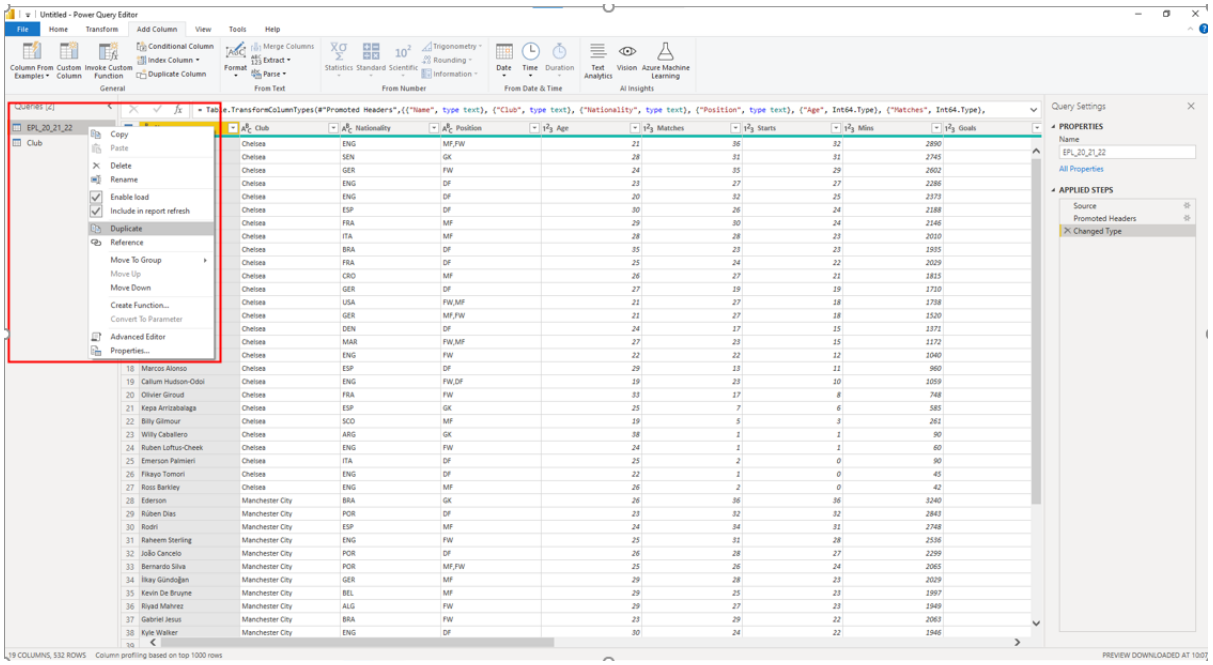

Figure 28: Duplicating the EPL table to create nationality table

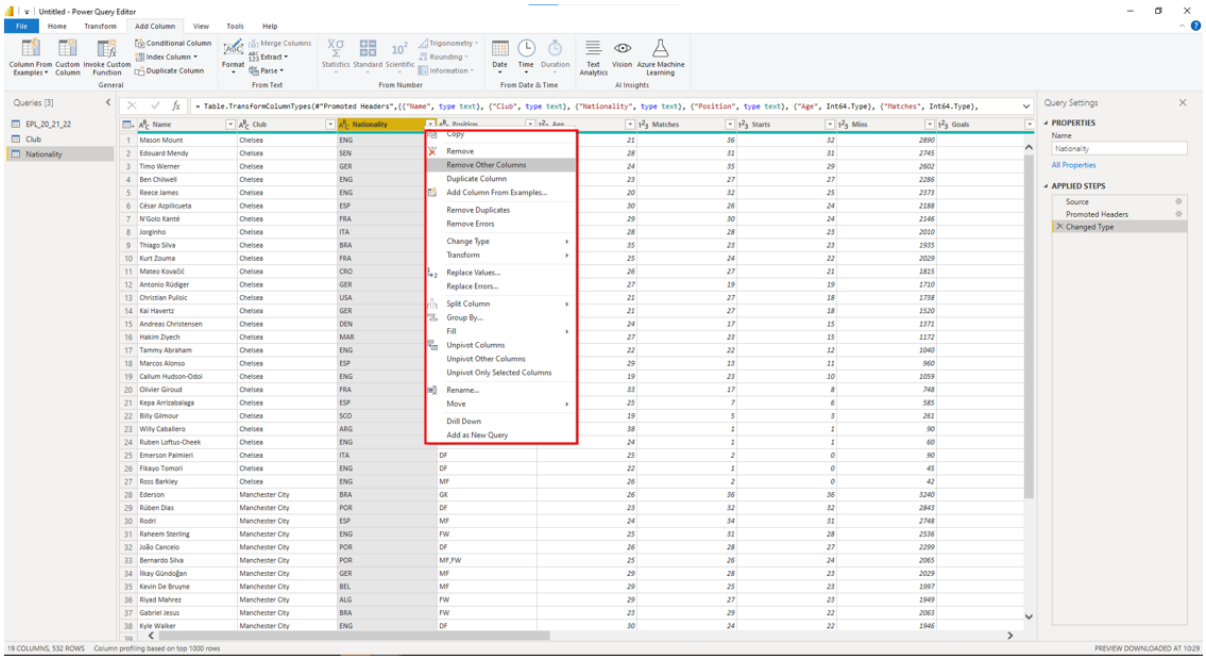

Figure 29:removing other columns

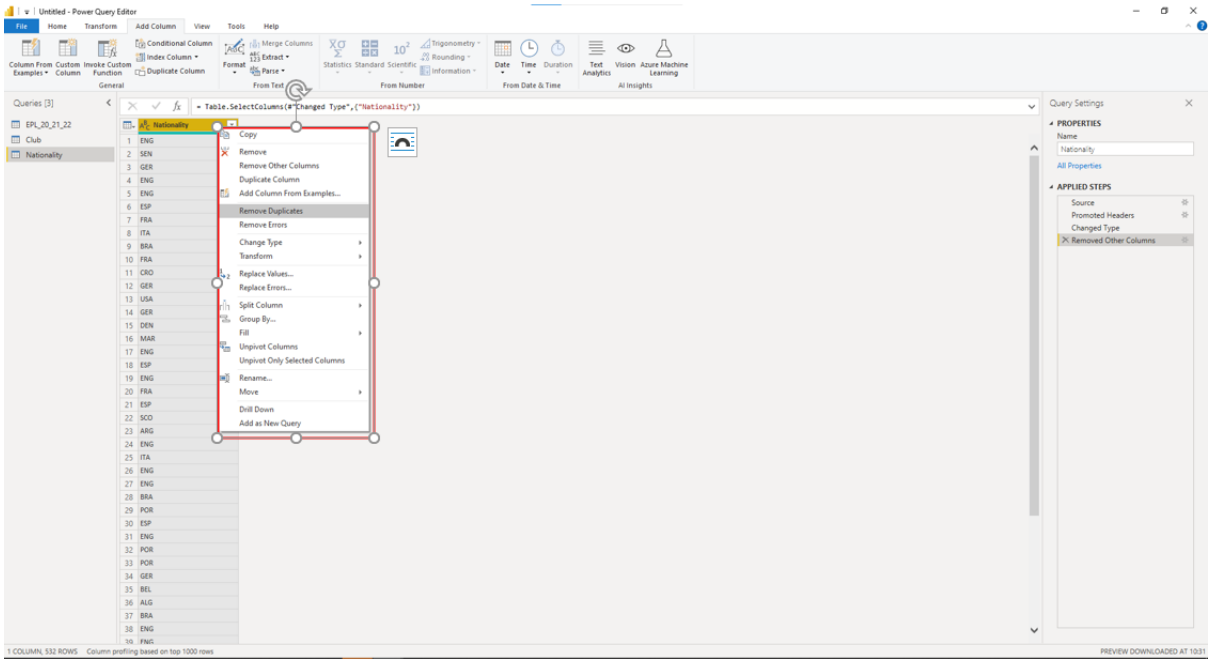

Figure 30: Removing duplicates

An ID column was also added to this table by clicking on “Add column” on the home tab and selecting “index” and choosing “from 1” in the pop-up box that appears. This column was renamed “Nationality Id”.

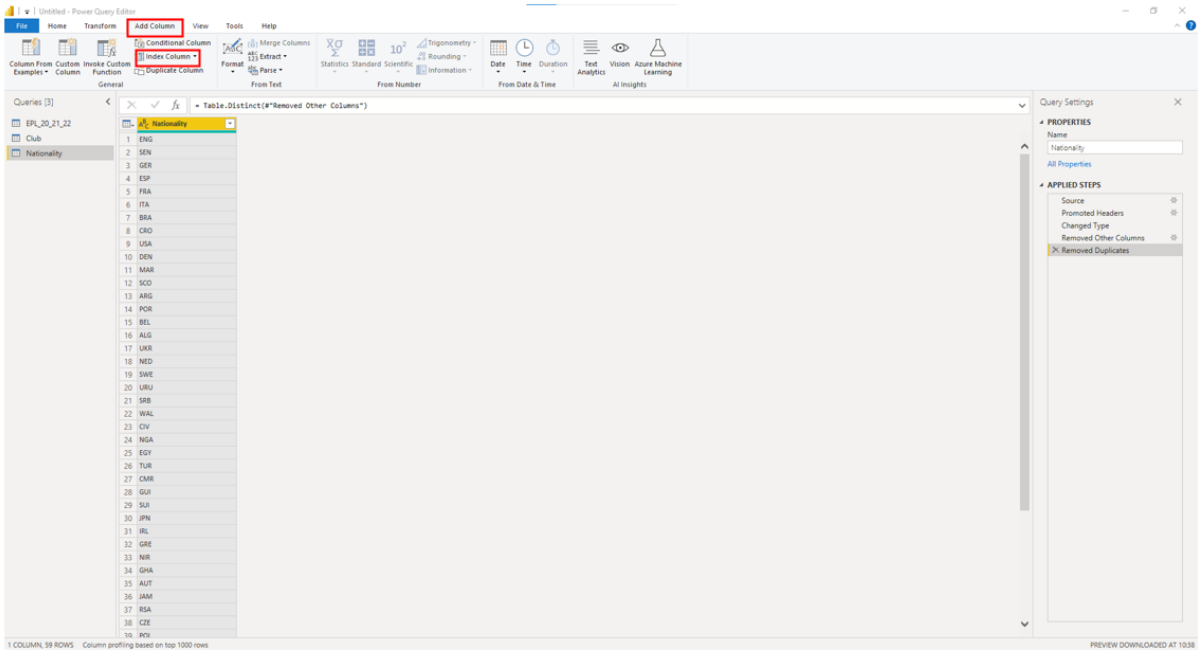

Figure 31: Creating index for nationality ID

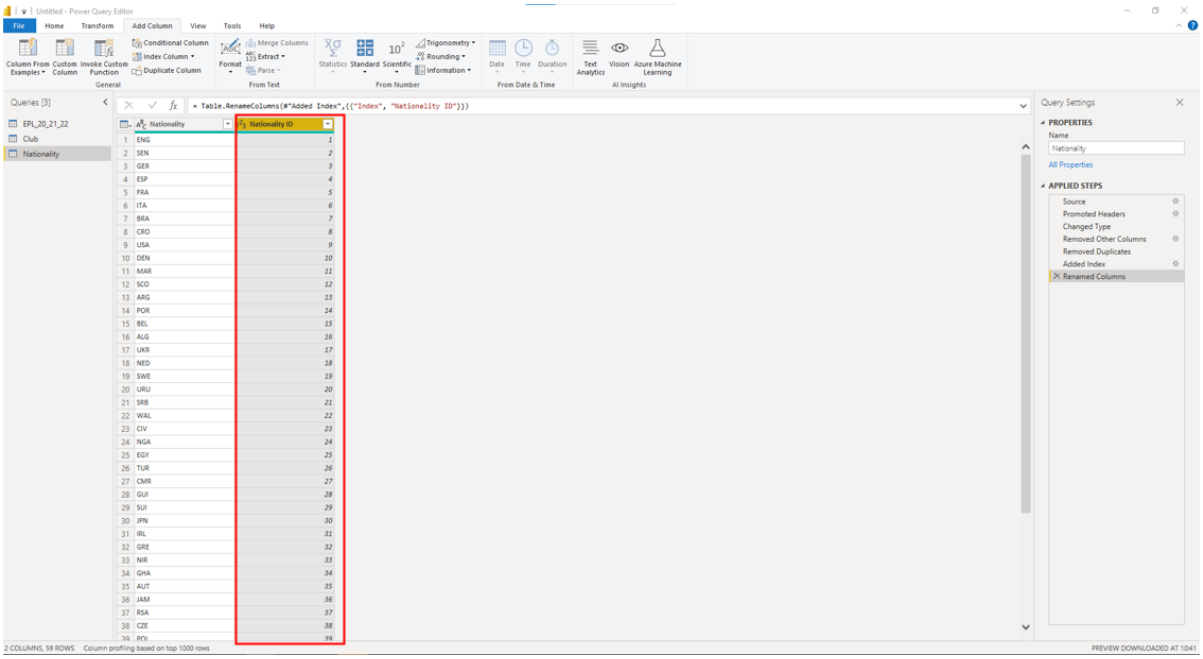

Figure 32: renaming index as club ID

This process was repeated and used to create two more dimension tables, Players, and positions tables.

Adding index columns to fact table.

The index (Id) columns that were created for all the dimension tables were added to the fact table in order to be able to create relationships. The dimension tables had to be merged with the fact table in order to make it possible. This was done by clicking on “Merge Queries” on the home tab and creating a connection between them on their shared columns. for the first-dimension table, that is the club dimension table, the shared column was club, hence the club column was selected on both the dimension table and the fact table using the “left outer” as the join kind. After this had been done, the fact table and club dimension table are merged, by clicking on “ok”. The illustration is detailed below

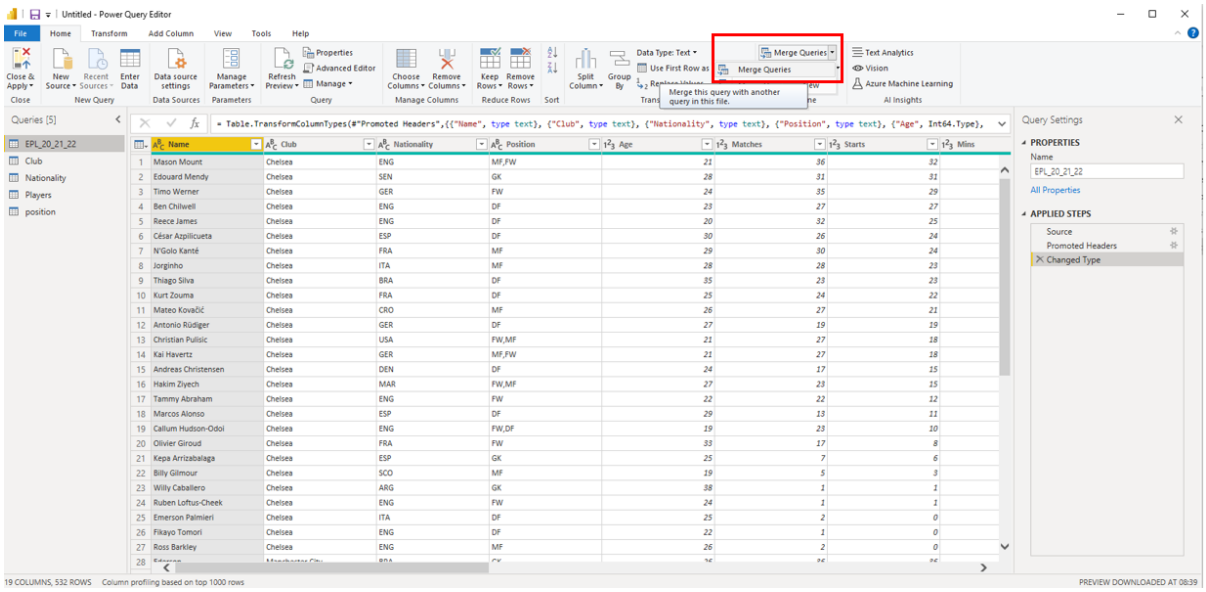

Figure 33: Merging queries for the fact table and dimension table.

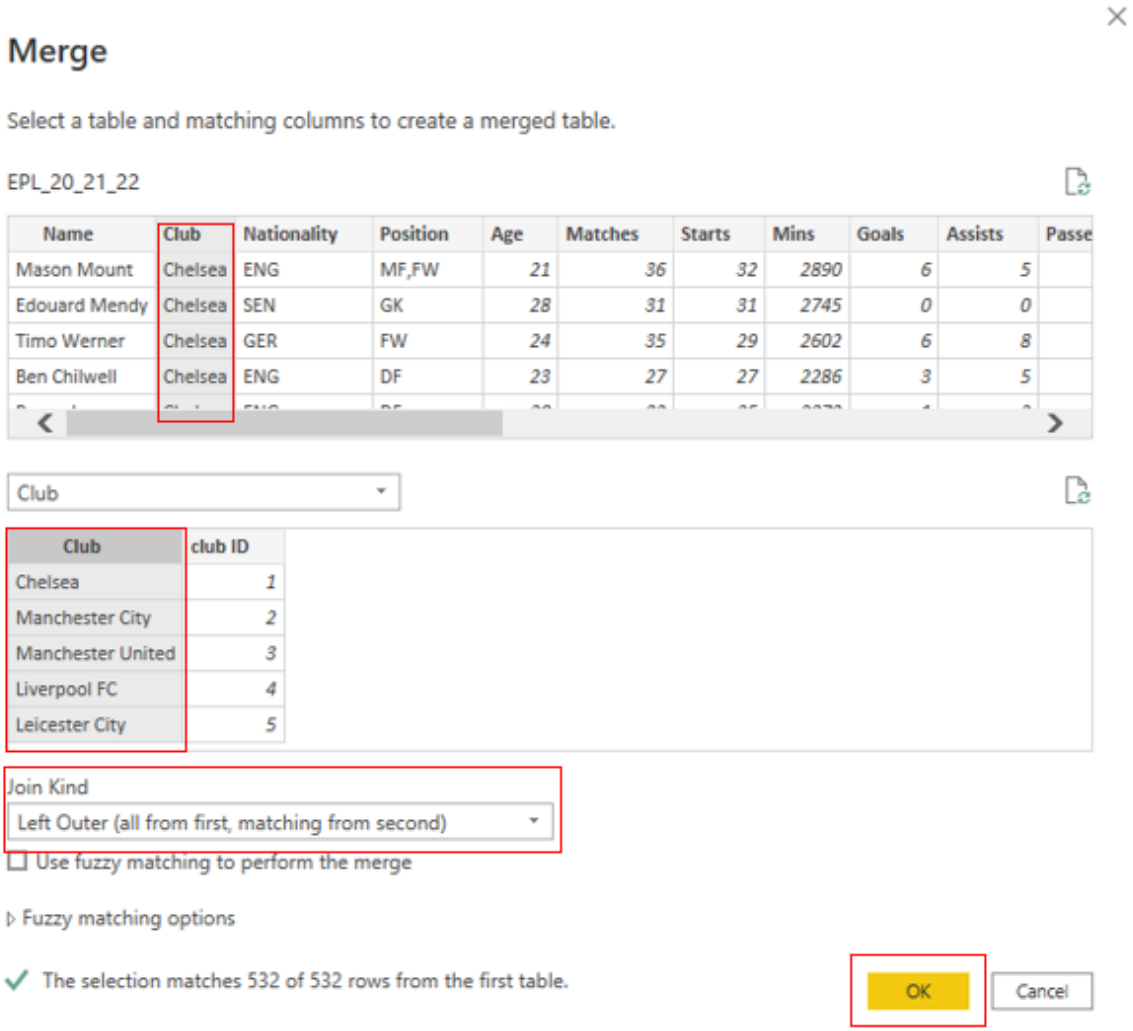

Figure 34: Club column merged for the fact table and the club dimension table.

Now, the column is expanded, and the club column is unticked leaving only the club Id column and this is done by right clicking on the arrow next to the new column. The column is renamed as “club Id”

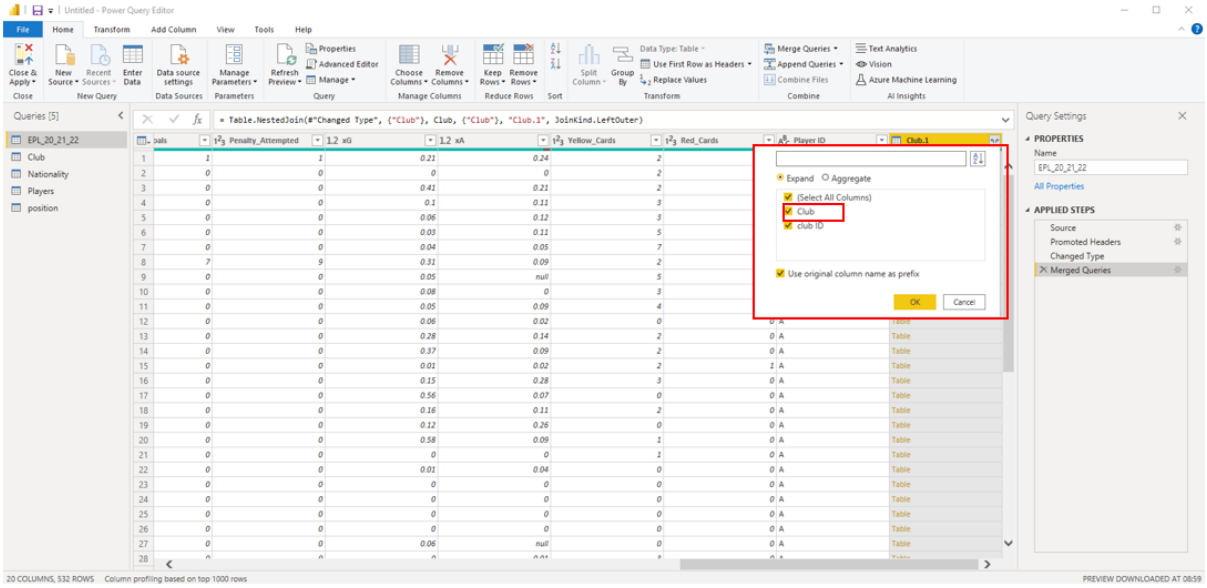

Figure 35: Club ID column expanded

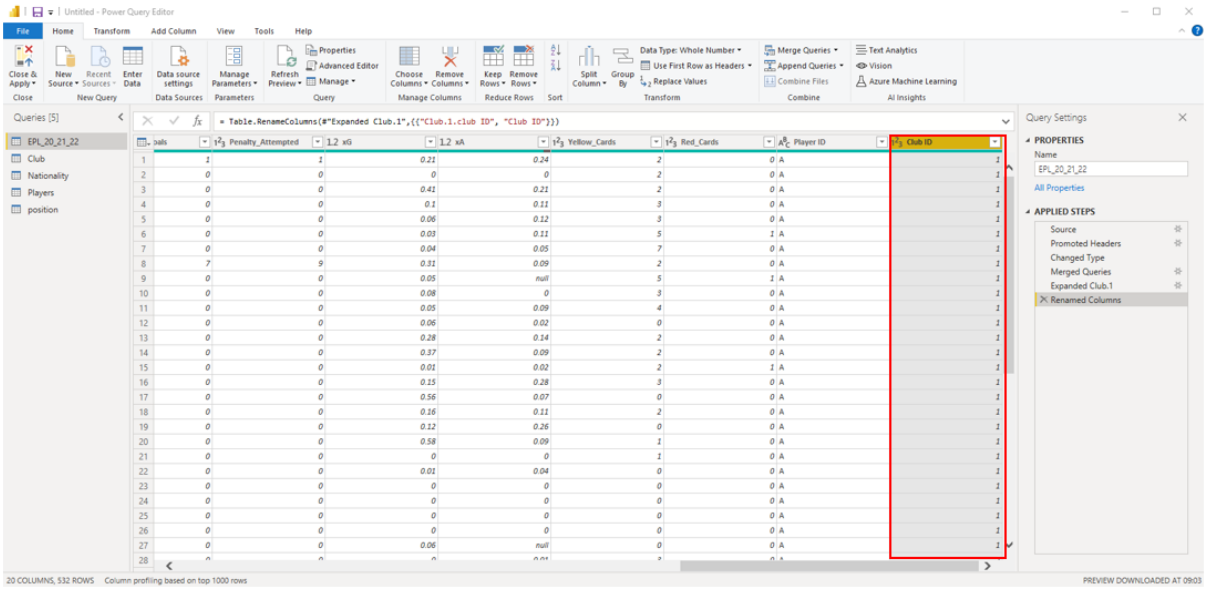

Figure 36: Club ID column renamed.

This process is repeated for all the other three (3) dimension tables.

The next step to do after repeating for all the other dimension tables is, to remove all the columns that are in the dimension tables from the fact table with the exception of the IDs and this is done by pressing and holding the control button and selecting all these columns and right clicking one of the columns and selecting “remove columns” from the dropdown box that appears.

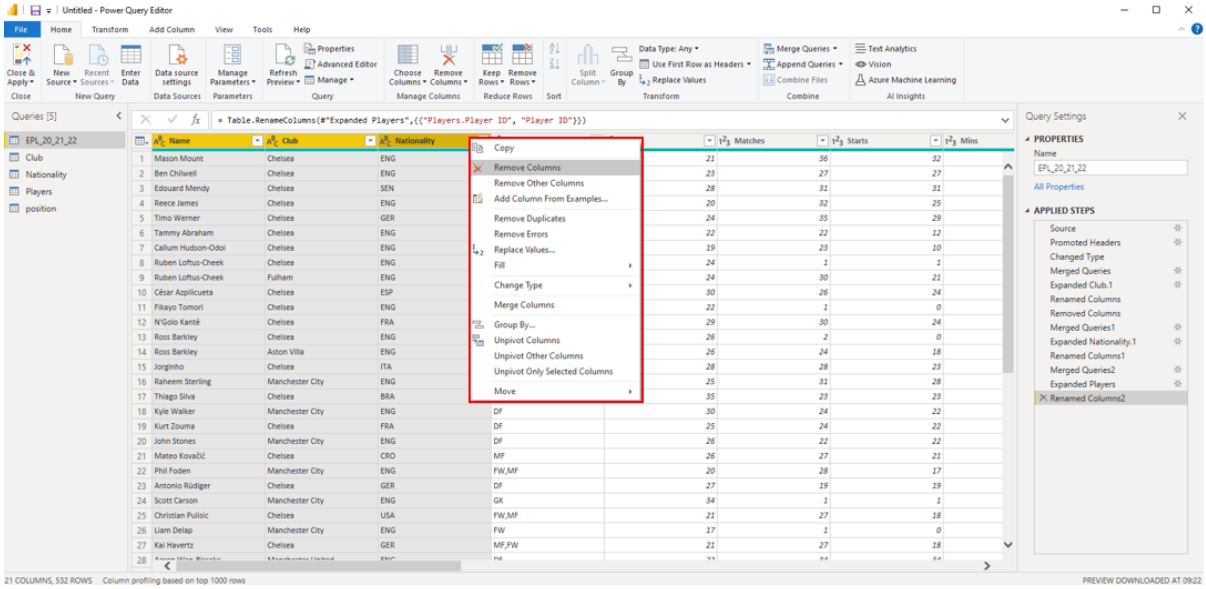

Figure 37: Removing columns that are in the dimension table that is also in the fact table.

Duplicating and splitting columns.

For the positions dimension table, the position column was duplicated by right clicking on the position column and selecting duplicate column from the dropdown box that displays. The column was then split into two using the delimiter, this was done by selecting the "split using delimiter" on the home tab. After the dialogue box opens, select split column using comma, then click "ok". The first position and second position columns were then created.

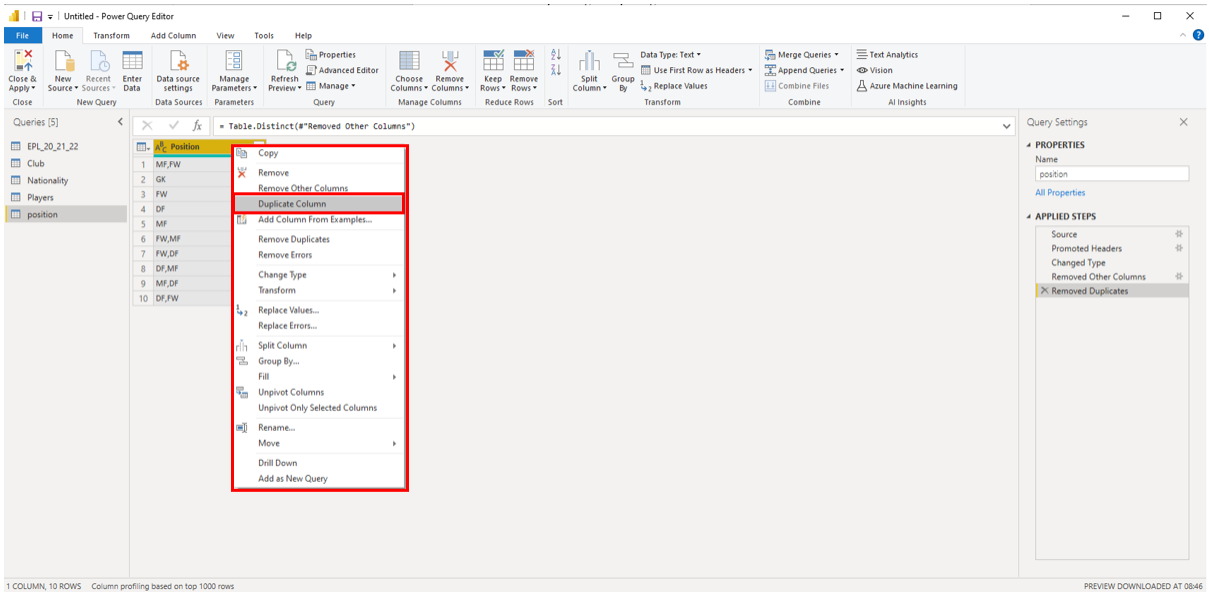

Figure 38:Duplicating positions column

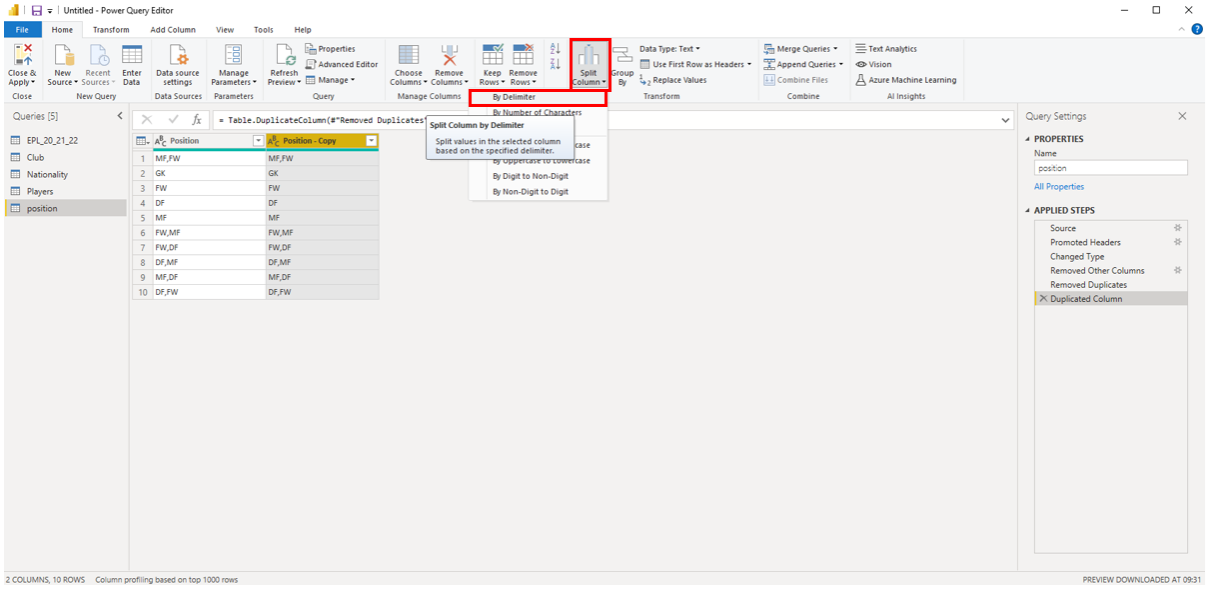

Figure 39: Selecting "split column by" on the Home Tab.

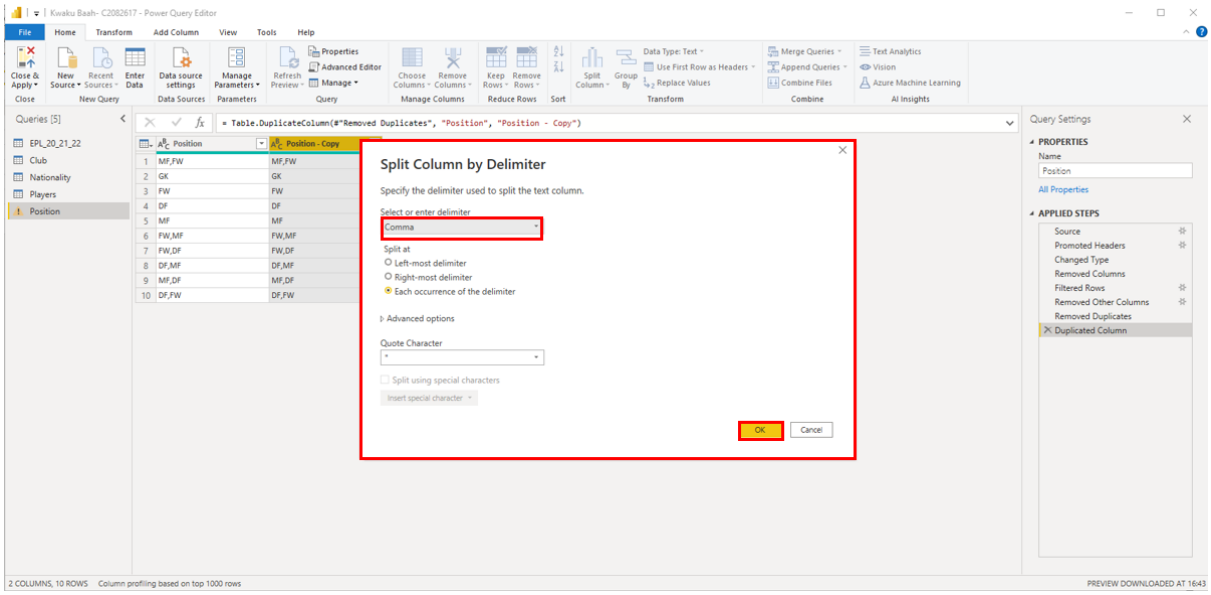

Figure 40: Positions column split using "Split column using delimiter"

For the secondary position column, the null values were replaced with N/A. this was done by right clicking on secondary position column and selecting replace values from the dropdown box that appears. Now, the null values were replaced with N/A in the "replace values" dialogue box that appeared.

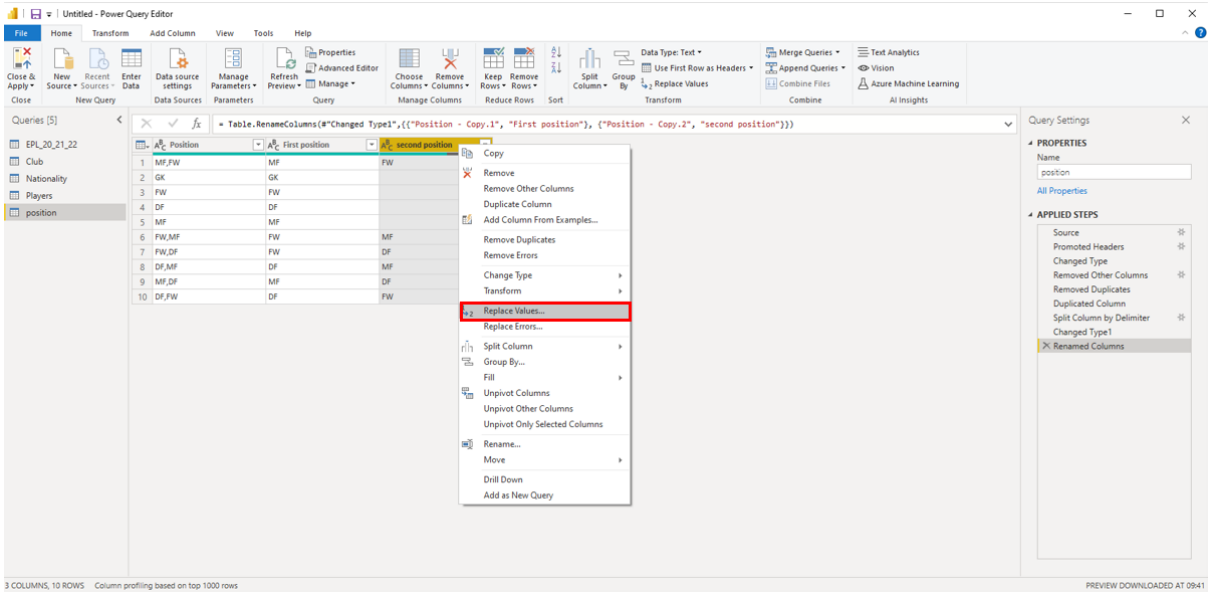

Figure 41:Selecting replace values from the dropdown.

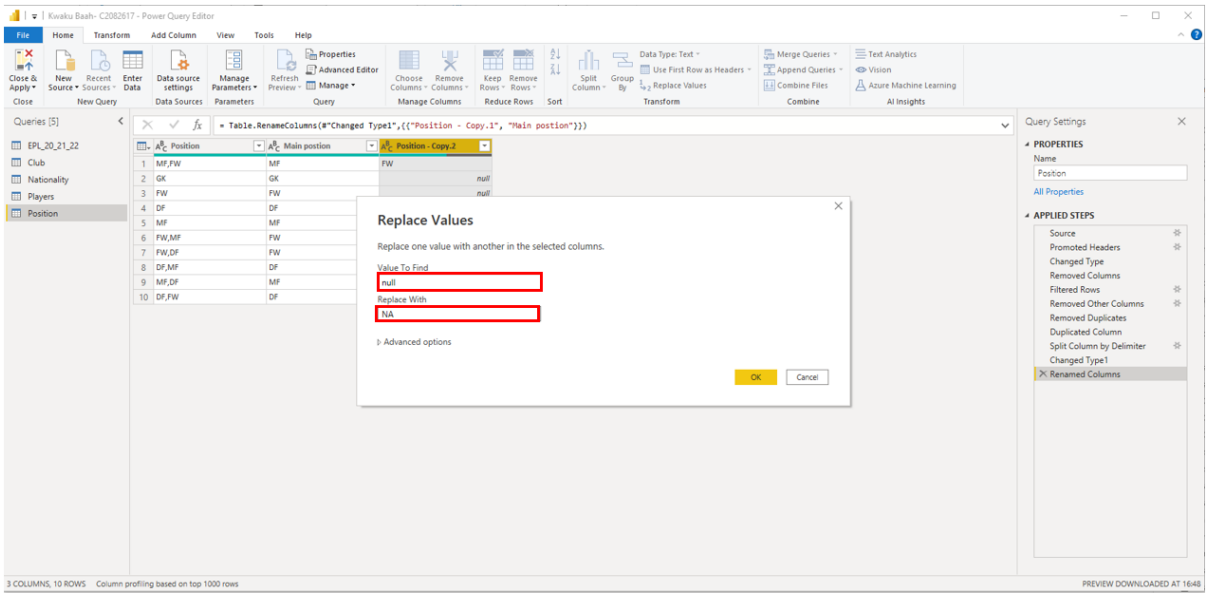

Figure 42:Replacing null values with N/A.

Creating relationships.

In order to properly analyse and visualise this project and meet the requirements of this project, relationships had to be created between the fact table and the dimension tables. This was done by selecting “manage relationships” from the home tab of the model view of the powerBi and “new” is selected from the dialogue box that appears.

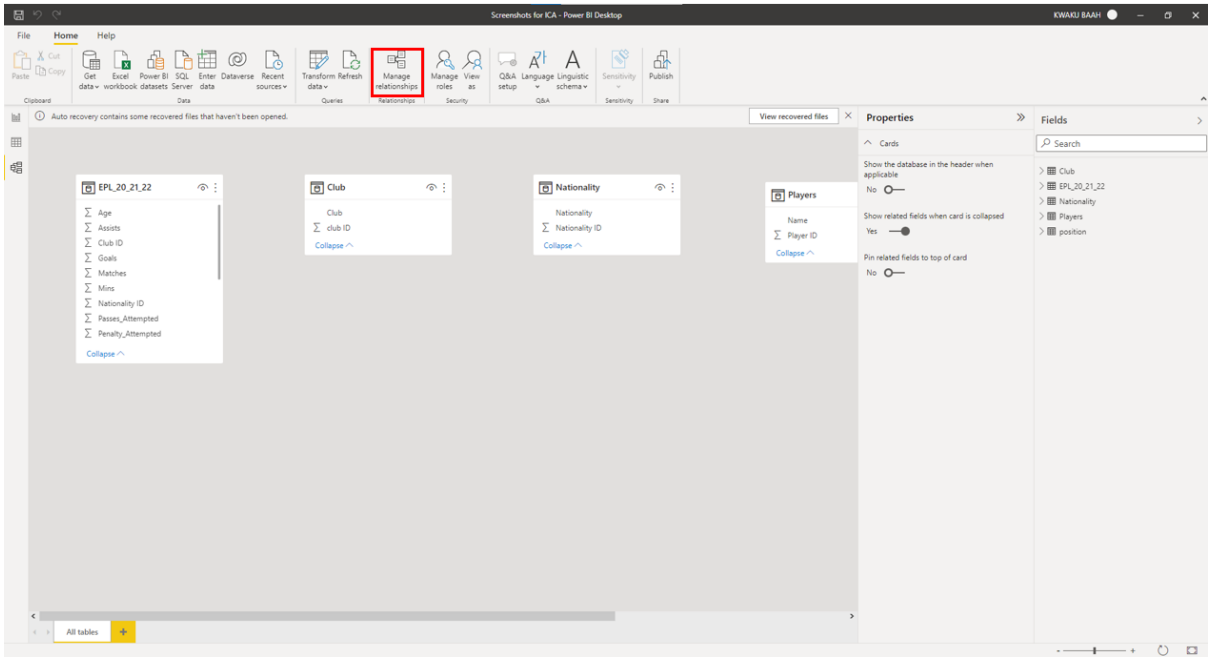

Figure 43:Selecting to manage relationships.

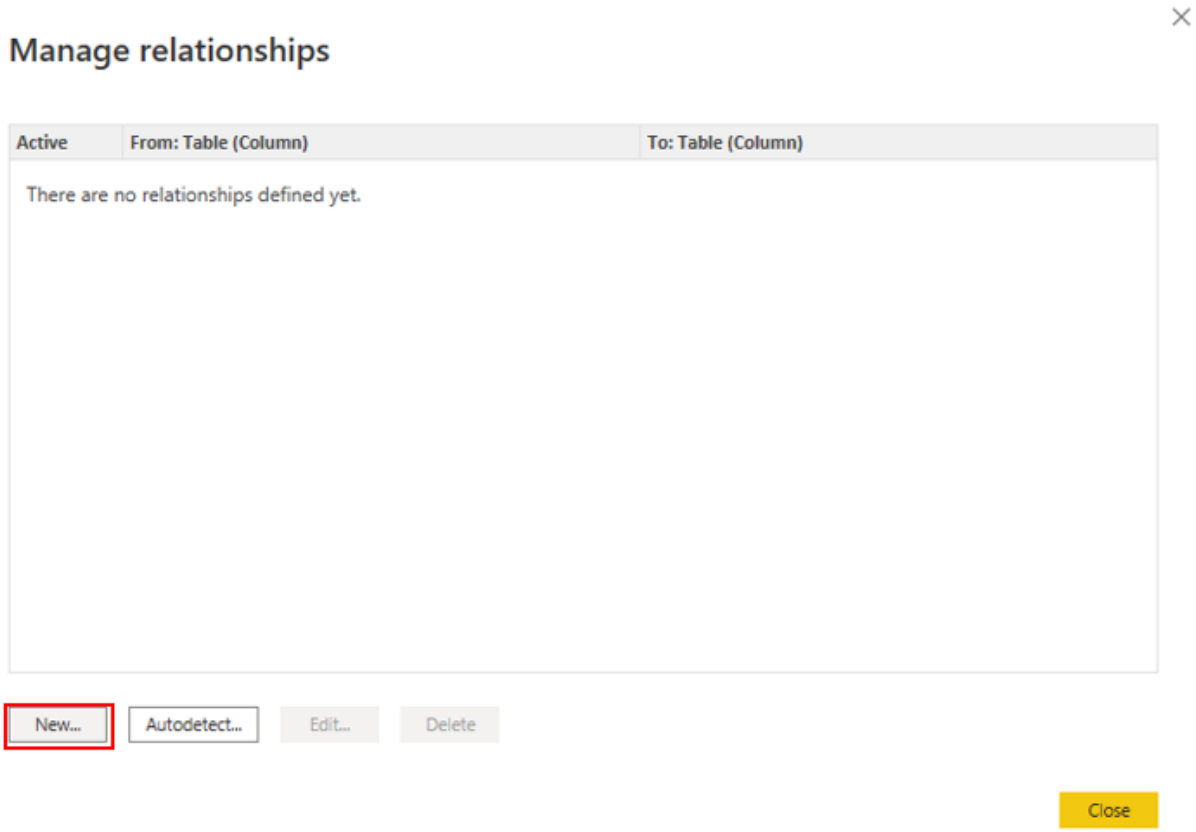

Figure 44:Select "new" from the manage relationships dialogue box.

The first relationship is between the "EPL 2020/21" that is the fact table, and the "club" dimension table is connected by the "club id" column. A relationship with cardinality "many to one" and a "single" cross filter direction is created between the fact table and dimension table.

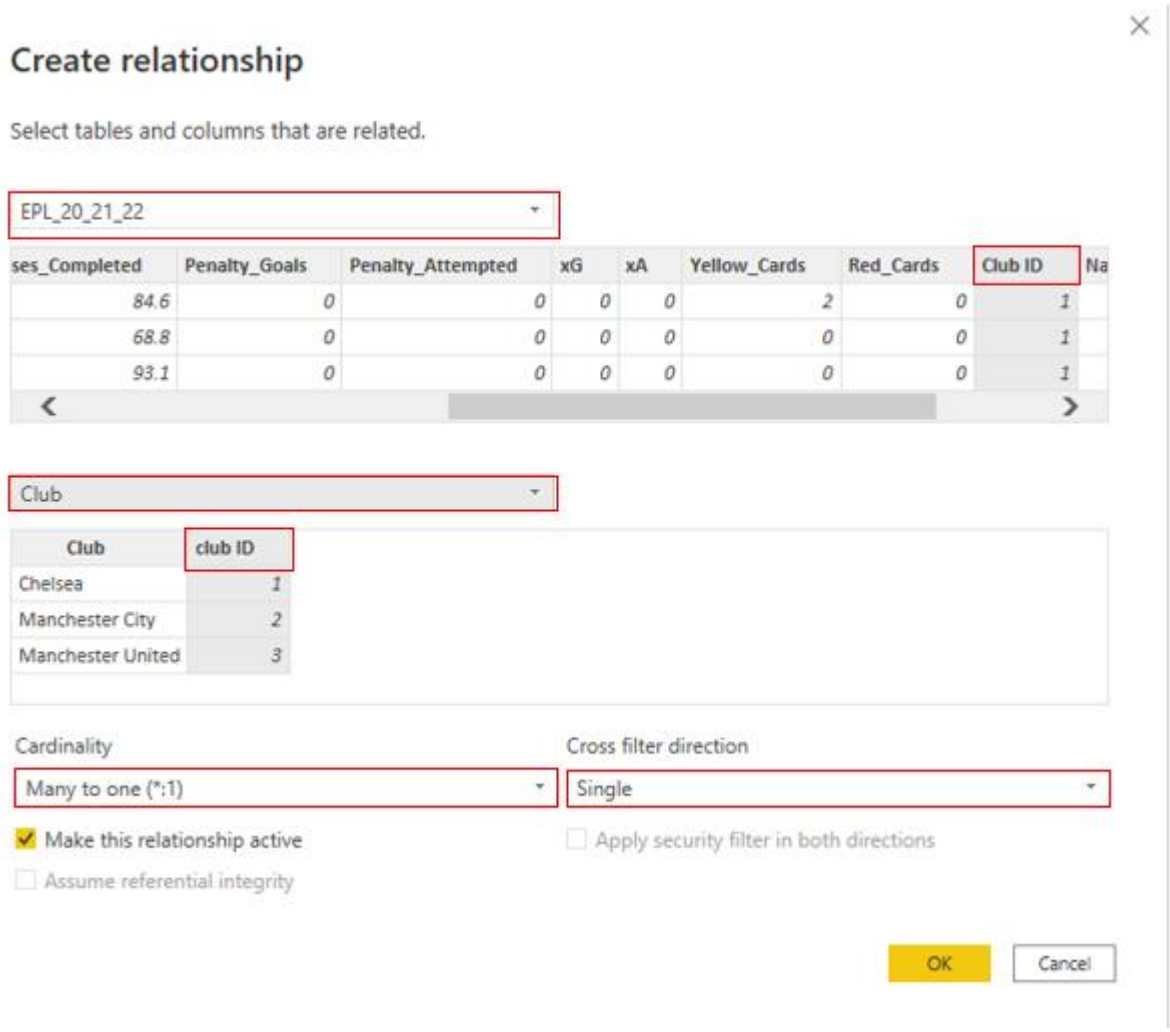

Figure 45:Relationship created between the fact table and club dimension table.

Again, this process is repeated for all the three-dimension tables as well.

A star schema model, which is where one fact table is connected to many dimension tables is the type of data that was developed after all the relationships were created.

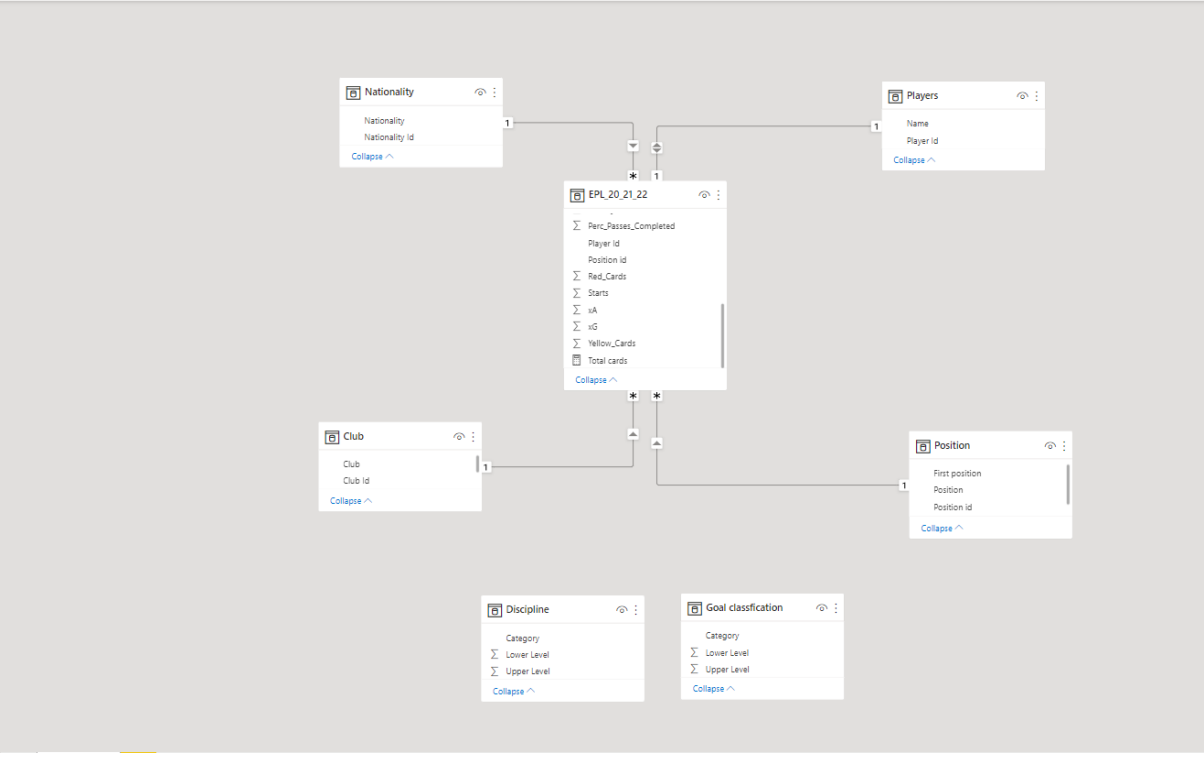

Figure 46:Star Schema model created for relationships.

DAX and M Language

The creation of DAX was the last stage in the data pre-processing stage, and DAX was used to generate two (2) measures and a column that is, Total club goals column, average goals per club measure and Total cards measure. This column and measures were created in order to go a step in simplifying the analysis and making the analysis easier to understand as well using their metrics. M languages were also created using the blank query and these were done so that the goal scorers could be categorised in four types in order to classify and be able to tell who the very best are, and it was also used in generating a table that told a story about how aggressive players were for this particular season judging by the number of cards they accumulated over the season.

The first DAX that was created was the Total club goals column and this was created using the DAX expression below.

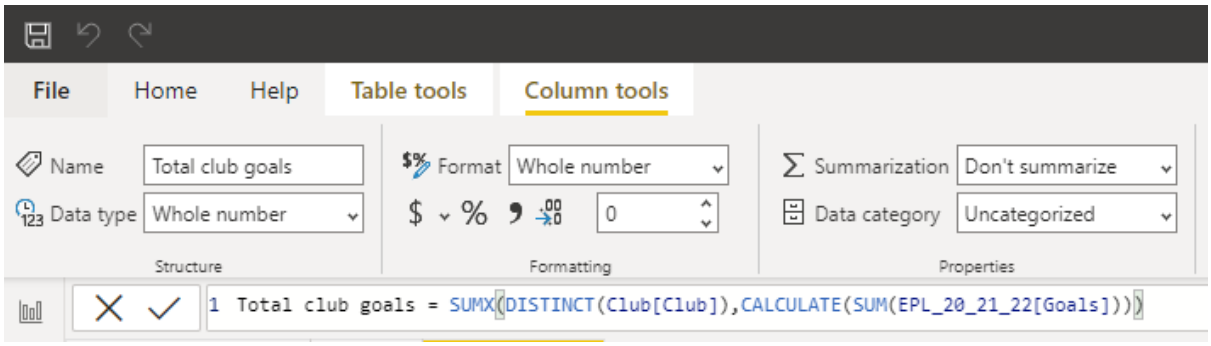

Figure 47:DAX expression for total club goals.

The second DAX that was created was the Average goals per club measure and this was accomplished using the DAX expression below,

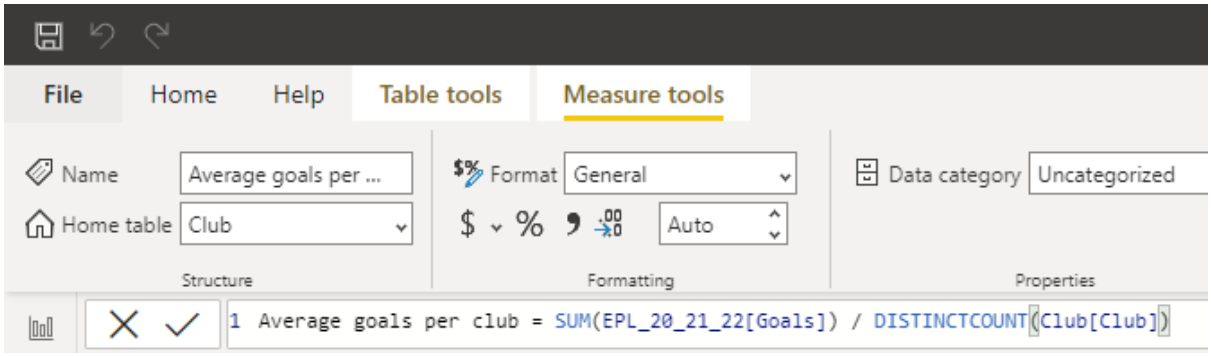

Figure 48:DAX expression for Average club goals.

Finally, total cards measure was the third and last DAX that was created and the expression below was the DAX expression used in creating it.

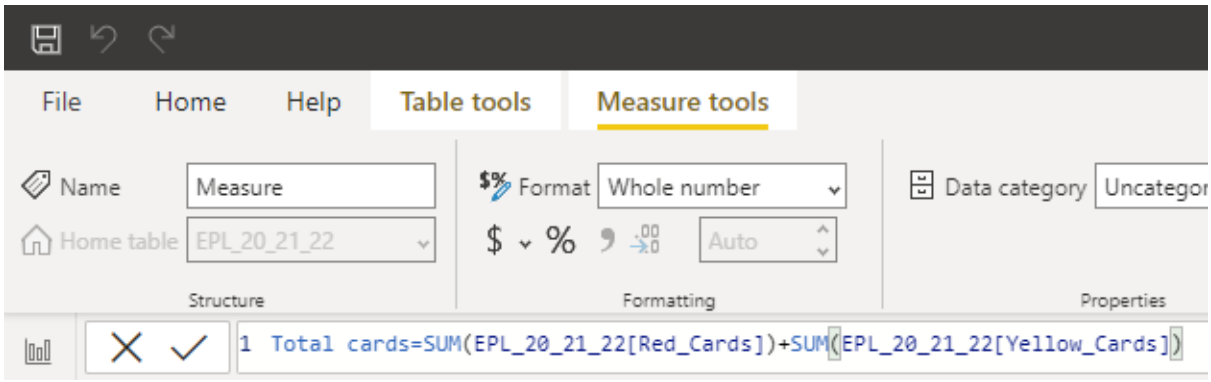

Figure 49: DAX expression for total cards.

Also, an M language was created for goal scorer classification by right clicking on "Get data" and select blank query editor from the dropdown. And the M-language below was used in creating the query,

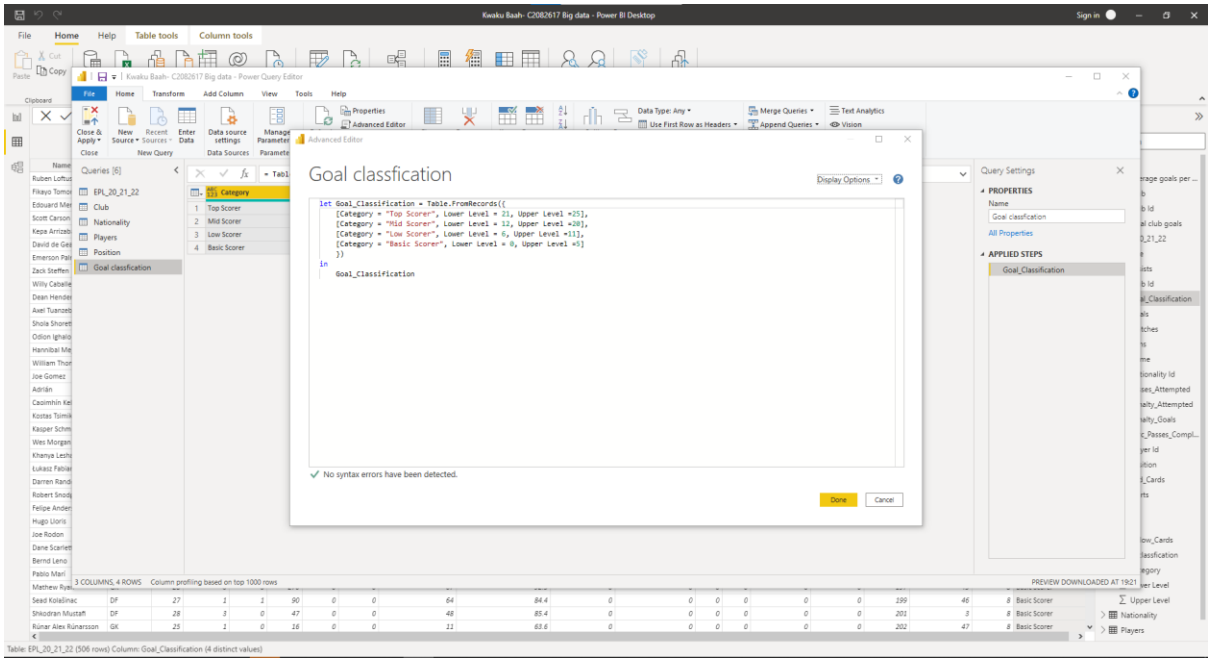

Figure 50:M language for goal classification.

Dashboard

A variety of different visualizations were explored and used in the creation of the dashboards. These dashboards and visualizations were tailored towards answering the business questions that analysis aims to answer hence the particular choice of every chart to visualize particular data types.

Dashboard 1: Homepage

The dashboard being displayed below which is the Homepage was created using a combination of icons, textbox, and blank page navigators. Blank page navigators were used in creating the pages that is, Club Performance Analysis, Player Performance Analysis, Position Analysis and AI Insights. These pages also serve as buttons for page navigators which help in navigating through the pages they are meant to display. A textbox was used for the module name, student name, student ID and course code as well. The home button is a downloaded image which also serves as a page navigator that takes the report back to the homepage. The premier league and Teesside university logos are also downloaded images. The title of the page was also created using the textbox. The image of the first dashboard is displayed below in fig. 51.

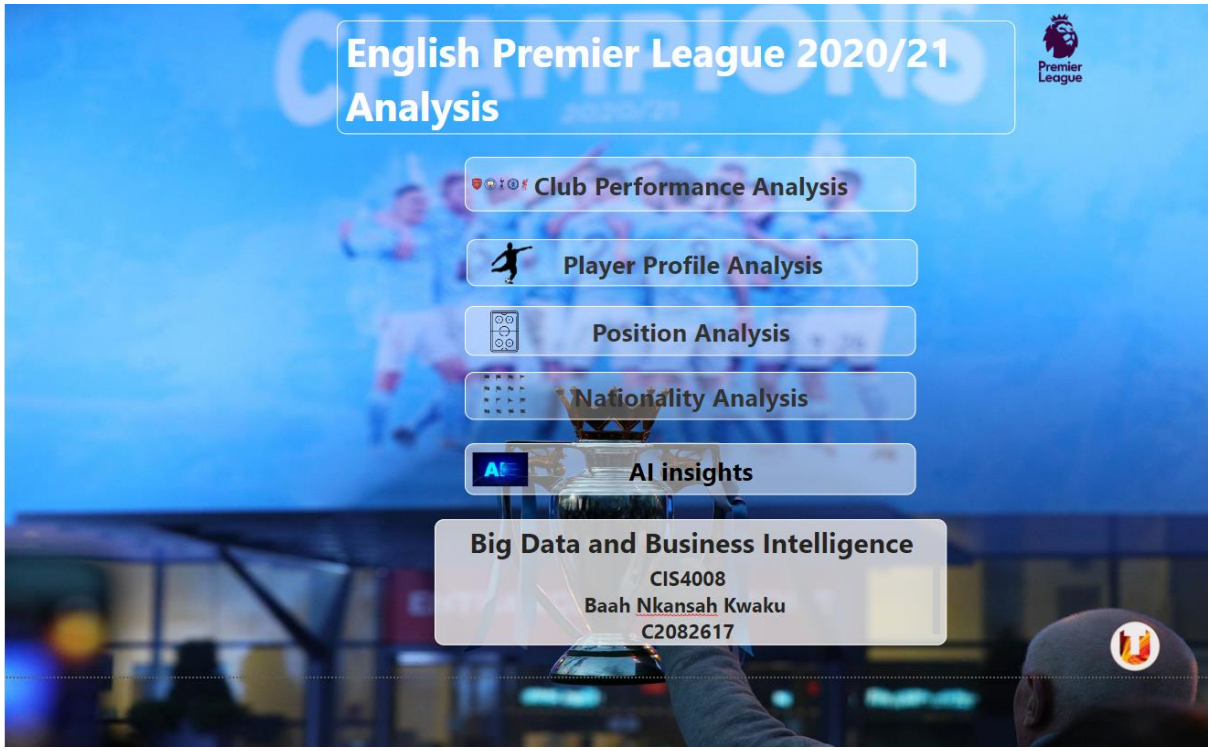

Figure 51:Homepage dashboard

Dashboard 2: Club Performance Analysis.

The below dashboard also displays a combination of different visualizations together with five (5) cards and one (1) slicer. The first chart is a horizontal bar chart which talks about the goals that were scored by the respective teams in the English Premier League during the 2020/21 season. The second chart is an Area chart which discusses the number of assists that each premier league club had for the English Premier League 2020/21 season. The third visualization is an animated chart that looks at the sum of the total club goals scored by each club for this particular season. The Drilldown radial bar chart was used to visualize the data type. A DAX expression was created for a new measure that was used in the creation of this chart. DAX expression created was,

$$Total\ club\ goals = SUMX(DISTINCT(CLUB[CLUB], CALCULATE(SUM(EPL\ 202122[GOALS]))$$

The five cards created were used to give an idea of the number of clubs, the number of matches played, the number of yellow cards, the average goals and the average goals per club, while slicer was created for a dropdown list of the clubs that played the premier league season and this slicer works on the three (3) charts and five (5) cards as well.

The back, home and next button are also page navigators that goes back to the previous page, the homepage, and the next page respectively.

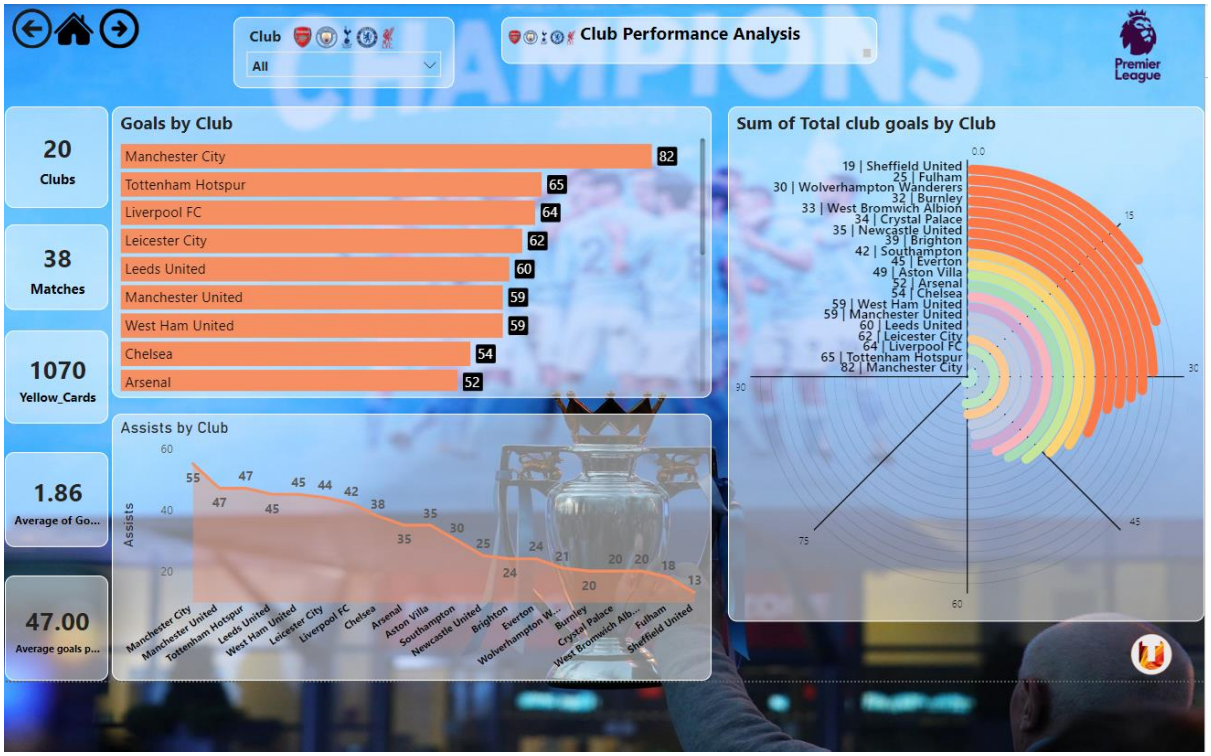

Figure 52:Club Performance Analysis Dashboard

Dashboard 3: Player Profile Analysis.

This is a special and different dashboard that was created with twelve (12) cards and three (3) slicers, that contains details and information about every player. The three (3) slicers which are, players, discipline and goal classification all work on navigating and giving an insight into the cards. The 12 cards contain data on Age of the players, the number of assists they had as well as the number of goals they scored, the passes completed, the number of red and yellow cards, their expected goals, the number of starts and the number of matches they played, and the penalty attempts they had and the number of penalty goals they scored.

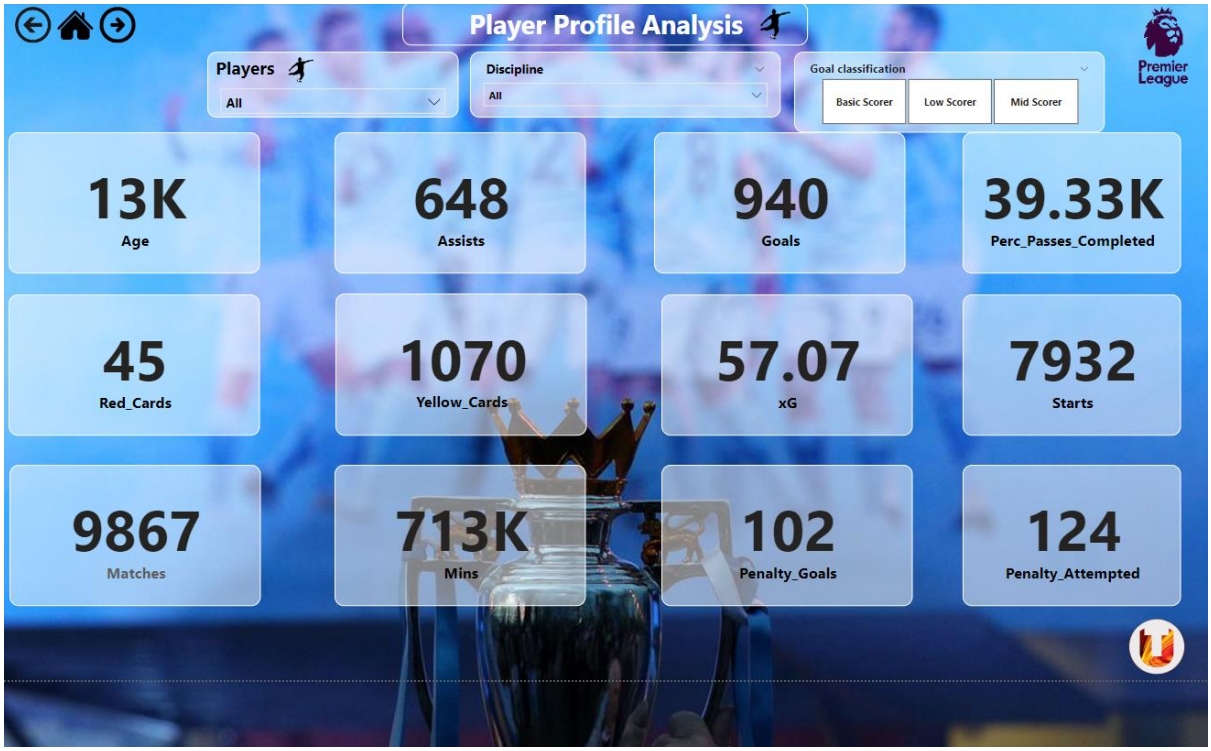

Figure 53:Player Profile Analysis dashboard.

Dashboard 4: Positions Analysis.

This dashboard is made up of a variety of different charts and visualizations. The first chart is a scatter chart that is analysing the average number of assists as against their position and the number of matches played per their ages. This was done to get insight into the age at which players could be said to be in their prime per the position they are playing in their respective teams. The second chart is a stacked bar chart looking at the number of goals scored by various main positions of the players. A card was also created to get a fair idea of the average of goals in a match. The third chart is represented by a drilldown radial bar chart. This visualization takes a look at which position had the greatest number of assists and this was done so as to get insight into which position contributes the most to the performance of the team. An Aster plot was used for the last chart which is visualizing the count of red cards for the respective positions for the 2020/21 season of the premier league. This was also done to gain knowledge about the most aggressive positions for this particular season. A slicer was also created for the various positions and this slicer reacts on the four (4) charts in this dashboard.

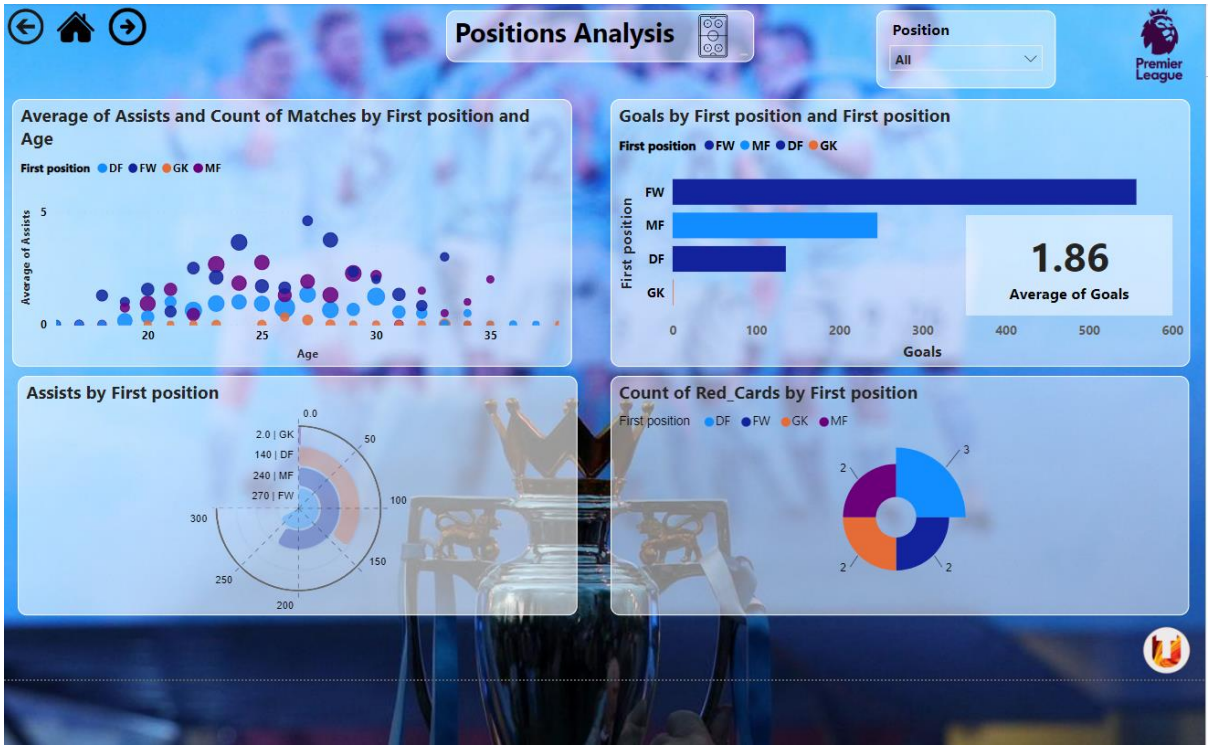

Figure 54:Positions Dashboard

Dashboard 5: Nationality Analysis

The nationality analysis dashboard contains four (4) charts, two (2) cards and two (2). The first chart is a map, and it was used because it is a chart that is representing and visualizing spatial data. The chart looks at the number of goals that were scored by various players through their nationality. The second visualization is an animated bar chart race which is visualizing the total number of cards and the matches played against the nationality of the players, this was done to figure out the nationality with the most aggressive players in the English Premier League 2020/21 season. The third chart is a treemap looking at the top 10 countries with the most players in the premier

league 2020/21 season. The last chart is a stacked bar chart analysing the total number of cards against the nationality of the players.

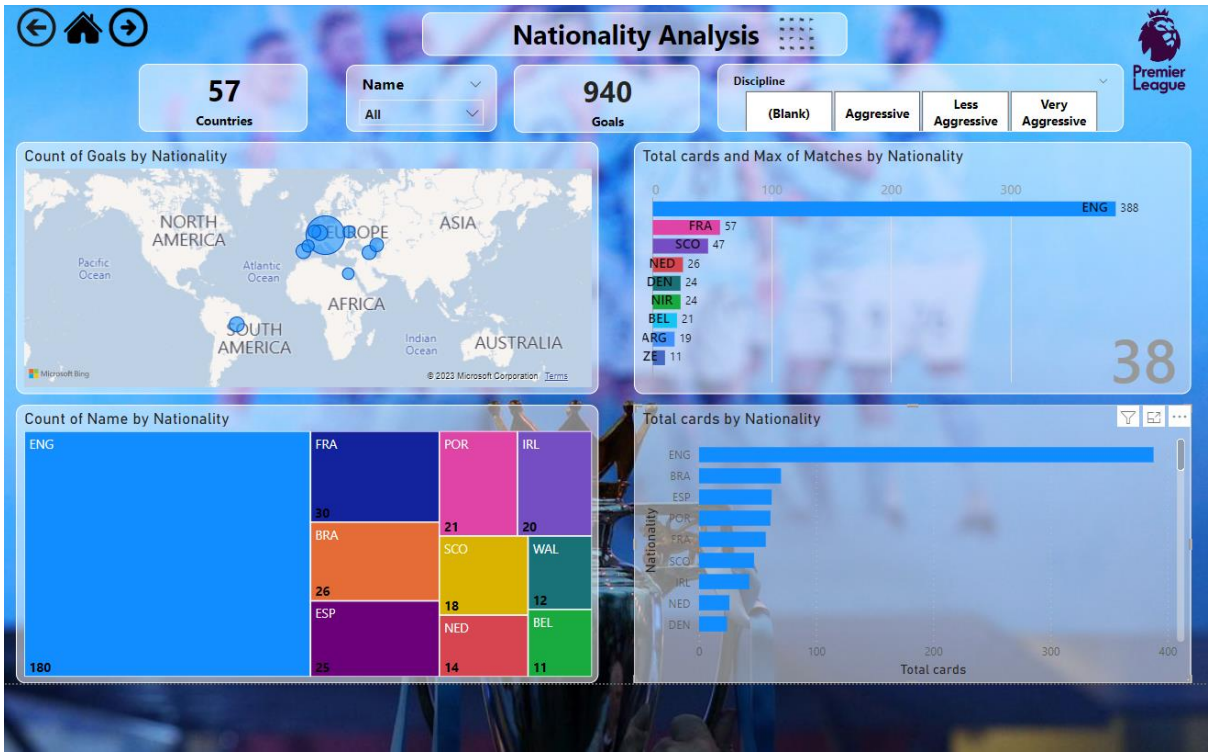

Figure 55:Nationality dashboard

Dashboard 6: AI Insights

This dashboard is a uniquely designed dashboard for insights into artificial intelligence. There were four (4) visualizations that were created for the AI Insights dashboard. The first chart is a decomposition tree that was used to analyse goals based on parameters like positions, club, and Nationality. The second chart is the key influencers chart, this chart was also used to analyse position as well while using goals, assists, matches, mins, penalties scored and age as parameters of explanation and also to analyse what influences decrease or increase in these parameters and also hidden patterns in the data. The third chart was a bar chart but focusing on distribution changes for goals and assists by various position. By right clicking on the bar chart and selecting “Analyze” from the pop-up box, there are charts that appear that help

identify outliers in goals and assists by various positions. The last chart is the Q&A chart, this visualization has certain in-built suggested questions generated by powerBi that comes with answers in the form of charts as well. Also, the "Teach Q&A" function was used in asking questions and determining how the questions are to be answered and submitting them.

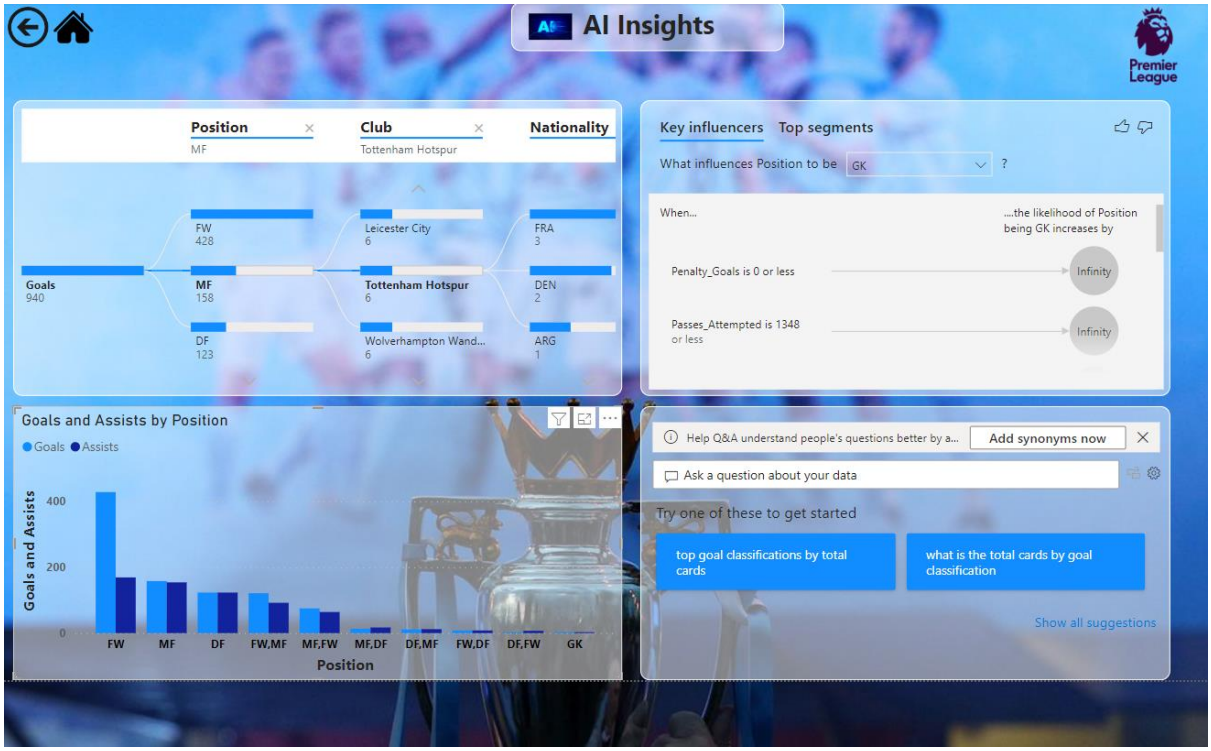

Figure 56:AI Insights dashboard

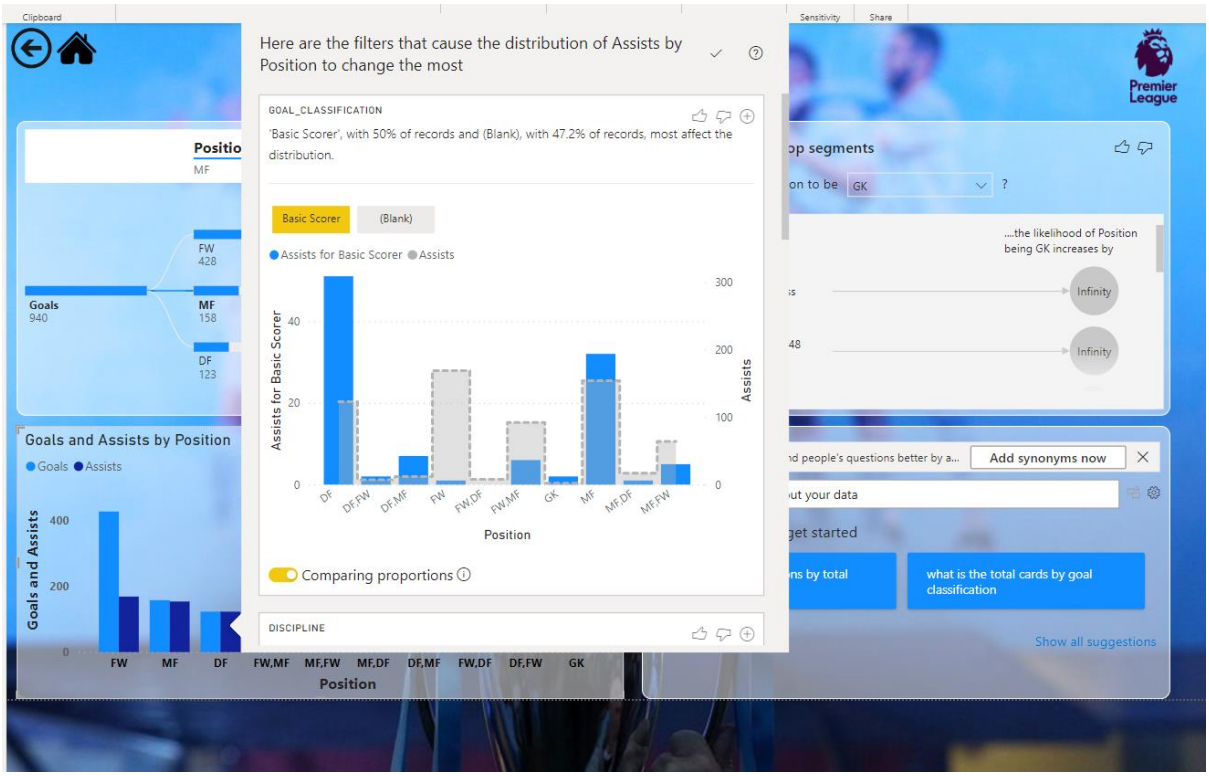

Figure 57:Further step for AI on bar chart.

REFERENCES

(2021) *Premier League Season Review 2020/21*. Available at: https://www.premierleague.com/season-review (Accessed: January 10, 2023).

EY (no date). Available at: https://assets.ey.com/content/dam/ey-sites/ey-com/pt_br/topics/ey-economic-advisory-/ey-premier-league-economic-and-social-impact-january-2019.pdf (Accessed: January 10, 2023).

Mwangi., B. (2021) *Exploring the English premier league 2020–21 season, Medium*. Heartbeat. Available at: https://heartbeat.comet.ml/exploring-the-english-premier-league-2020-21-season-ad10c6c83f1c (Accessed: January 10, 2023).

Use the table below to **self-assess** your work. This will help reflect on your work.

You must keep this table in your report.

Report Section	**Description**	**Grade your work from 0 to 100**
Report Structure	The report is well-written, and it contains all the relevant sections	87
Data Pre-processing and Data Modelling	Many pre-processing steps have been applied. The data model is well-structured	60
Dax and M language	Both DAX and M Language have been **extensively** used in the report	75
Dashboard Design	The dashboard contains a variety of charts, including advanced ones not covered in the module.	82
Average		**Add below the average of the four cells above:** **76**