National College of Ireland

Computing

Data Analysis

2020/2021

Cian Larkin

X17453136

X17453136@student.ncirl.ie

How has Covid-19 Affected Football Transfers?

Technical Report

Contents

Executive Summary

This document documents the approach the researcher has taken in order to address the subject of the project. The project will look to gather multiple datasets in relation to football clubs' transfer dealings and analyse them in order to see which clubs struggled with the decreased earnings and which clubs could continue as normal, in terms of transfers.

This document intends to help to reader understand what was done for the duration of the project and how the techniques used helped to achieve the project goal.

December 2020 marked the midpoint presentation, delivered virtually, of the project. Some of the techniques needed to fully execute this project would only be taught in semester two which starts in February 2021.

May 2021 marks the completion of the project. Datasets were taken from an open GitHub repository which were used to carry out an analysis of the transfers in the Premier League and the Championship. The project was able to conclude how each of these leagues were affected by Covid in relation to transfers and also make a prediction as how clubs in these leagues will fare for the following season.

1.0 Introduction

1.1. Background

The reasoning for this project comes from my own involvement in the sport. I played as part of a team from age 5 right up until aged 17, and I continue to play occasional 5 a side matches now. Football is the biggest sport in the world with fans everywhere you go and now that COVID-19 is a part of everyone's life I felt that it would be interesting to combine the two into some sort of project, eventually developing into a data analysis project.

Covid has effected all types of business, big and small, so I felt this theme would be appropriate for the project. Football is not exempt from the challenges everyone else has faced so the issue of no fans in the stadiums resulting in less spending power peaked my interest and prompted me to investigate if clubs felt the impact and if so, did the bigger clubs cope better than the smaller clubs as I know, from keeping up to date on footballing matters, that some small clubs have gone out of business since Covid hit.

I felt like the appropriate datasets would be out there and also club transfers are widely available online in case I have to create some of my own datasets. This means that there is a great opportunity to use the ongoing situation to my advantage for this project and conduct an analysis to see how or if football teams have been affected from the pandemic.

As I further got to grips with some of my final year modules such as Business Data Analysis, Advanced Business Data Analysis, and Data and Web Mining I felt like I had the tools necessary to investigate the data available to me. I love football I am constantly keeping up to date with player and team statistics throughout the seasons so I had a lot of motivation to put together some statistics and graphs of my own and carry out tests which would return interesting findings.

1.2. Aims

Aim 1: The first aim is for this project is to find the adequate datasets. Datasets for the professional leagues in England for this year transfers will be needed and then these will be compared to previous years transfers. Different methods of analysis will be carried out which will be taken directly from this year's modules.

Aim 2: The second aim will be to clean the datasets and have them contain the information needed to analyse. This cleaning will be done in RStudio, I will have to fix any errors in the data attributes and remove any columns I feel are unnecessary. The main thing to analyse is the transfer fee of each player so the datasets must be able to represent this.

Aim 3: The third aim will be to look into the data and compare the transfers across the different years and analyse any changes, if any, due to Covid. This will be done through the use of tests and graphs created in a various amount of programs. I want to create data visualisations in different programs such as Excel, SPSS, and RStudio. I want these visualisations to easily represent my findings to the reader.

Aim 4: To complete documentation with the insights of this project, including recommendations on how clubs can cope in the future if such situations similar to Covid-19 arise.

1.3. Technology

R Studio

R Studio will be used to construct this project, which is an open-source development tool for the R language. This will be used alongside excel where the datasets will be stored and then brought into R Studio as they are needed. The datasets I have chosen

for this project will need to be cleaned before I can analyse them, all of this pre-processing and transformation will be done in RStudio.

R Language

R is a programming language used for statistical computing and will be the language used to build this project.

Excel

Excel is a spreadsheet tool which allows a user to manipulate loaded data with its built-in statistical commands. This is where the csv files will be stored before they are transformed in R Studio. Excel will also be used to create graphs when necessary.

Draw IO

This is an online diagram maker, it can be used for making flowcharts, creating UML online, and as an ER diagram tool. For the purpose of this project, it was used to create a diagram which represents the data implementation steps which were carried out in this project.

SPSS

This is one of the most programs used for statistics because it can perform a wide variety of statistical computations and allows the researcher to carry out an analysis with ease.

The project will make use of SPSS by using the cleaned data in RStudio to produce some results.

GitHub

This is a web service for hosting a repository. This will be used to host my code from RStudio.

1.4. Structure

Section 2 describes the data I have chosen for this project, why I chose this specific data, and how I plan on using it for the purpose of this project.

Section 3 details the chosen methodology which will be adhered to throughout the duration of this project.

Section 4 covers the analysis section, this is where I outline how I plan on analysing my data to obtain the necessary results.

Section 5 is the results section where I will show the results I have collected as a result of my analysis. This will be broken down into separate sections for each set of results/type of test.

Section 6 is where I will draw up a conclusion to my findings, talking about the strengths and weaknesses of this project.

Section 7 I will talk about potential future work which could be carried out in this area.

Section 8 this section is for references.

Section 9 this is the appendices section where I will have my project plan and my monthly reflective journals.

2.0 Data

2.1 Dataset Description

As of December 2020, I was at the stage where I was collecting data which was used for the analysis. I was looking at some datasets which had already been created which I would then process and transform in order to suit the needs of this project. There are multiple datasets available across websites such as Kaggle, GitHub and Statista. One of the datasets I found contains a massive amount of football transfers across multiple leagues so this was chosen but I would have to carry out a decent amount of processing to clean it up and suit the project (ewenme, 2020). I would have to look across multiple years for my comparison and this dataset has the years 1992 right the way through to 2020 available so that is more than enough. There was the possibility that I would have to do some web scraping from Transfermarkt.co.uk in order to get some extra transfers that may not be available in readymade datasets like the one I have mentioned above. I expected that this would have to be done for the smaller team's transfer as they are not popular, so it would not be likely that these datasets exist. I expected that the smaller clubs would struggle as naturally, they do not have the same spending power as the larger clubs. The same can be seen in Irish business as The Irish Times reports that small and medium-sized business are facing a revenue shortfall of €15 billion this year (Hamilton,2020). Clubs are basically businesses they are to make money at the end of the day so I believe the relevant data will show a struggle for most of these small clubs. Research was carried out on businesses around the world to show that it is not just football clubs who may struggle financially and that there is a correlation between the football and business world. A survey carried out by Deloitte showed that roughly 49% of CFOs from big UK companies predict a return to pre-Covid levels of business only in the second half of 2021 (Bruce, 2020), This further emphasises the prediction that Clubs will struggle due to Covid.

The collected datasets were pulled from GitHub from a large collection of football transfers which ranged from 1992 up until the current 2020 season. Each dataset contains important details of each transfer such as columns like the fee involved and the clubs involved. This allows me to numerically analyse any changes in the amount of money spent year to year. With the transfer fees having already been collected in datasets this meant I could devote my time on this project to conducting my analysis and not get too tied down on collecting data.

Having collected all of my data a significant amount of cleaning was needed as there was some errors regarding some foreign player and club names. The programs I was using such as RStudio and Excel was not recognising some of the symbols in their names so this meant I had to go into RStudio and change these errors through code as shown below. This was a long and tedious process as I had to do this on each of my 12 datasets, I had to do the same thing in my Data App module in semester 1 so without this I would have had to have turned to Google in order to find some answers. For the fee_cleaned column some of the values returned where NA, this is normal as it is a result of a player moving on loan or on a free transfer, so these were changed to 0 as it is good practice to do so. The fee column was also showing a 'Â' for most rows, so this also had to be removed for each entry.

Arsenal FC	RÃºnar Ale	25	Goalkeepe	Dijon	Â£1.80m	in	Summer	1.8	Premier Le	2020	2020/2021
Arsenal FC	CÃ©dric S	28	Right-Back	Southamp	free transfer	in	Summer	NA	Premier Le	2020	2020/2021

Figure 1: Sample Dataset Before Cleaning

4	Arsenal FC	Ruenar Alex Rúnarsson	25	Goalkeeper	Dijon	£1.80m	in	Summer	1.800	Premier League	2020	2020/2021
5	Arsenal FC	Cedric Soares	28	Right-Back	Southampton	free transfer	in	Summer	0.000	Premier League	2020	2020/2021

Figure 2: Sample Dataset After Cleaning

```
#Replace incorrect characters in the PremierLeage2020 data frame
PremierLeague2020$name <- sub("Ã", "i", PremierLeague2020$name)
PremierLeague2020$name <- sub("i©", "e", PremierLeague2020$name)



#Changing N/A values to 0
PremierLeague2020[is.na(PremierLeague2020)] = 0
PremierLeague2019[is.na(PremierLeague2019)] = 0



#Renaming columns
PremierLeague2020 <- rename(PremierLeague2020, name = player_name)
PremierLeague2020 <- rename(PremierLeague2020, club_from = club_involved_name)
```

The code above are snippets taken from the code which was executed for the cleaning which was carried out for this project. The first shows how the incorrect values were fixed for the player names for the PremierLeague2020 dataset, this code was carried out for all datasets. The second snippet is how I delt with NA values across each of the datasets, and the third is some of the code for renaming columns in my data.

The next thing to do with my data was to create a subset of my data which only contained the details of the players who were signed by clubs, so the players who were sold by clubs were excluded. The reasoning for this was to create graphs depicting the amount spent by each individual club in a particular year. This was done by taking the cleaned data from RStudio, transforming it into a new dataset in Excel, and then re-loading it into RStudio to create my visualisations.

2.2 Data Requirements

The Data Requirements needed in order to carry out this project are the datasets which were pulled from GitHub (https://github.com/ewenme/transfers) this consists of all the transfers necessary. The details of the datasets are shown below:

- ***Club_name:*** This is the name of the main club involved.
- ***Name:*** This is the name of the player in question.
- ***Age:*** This is the age of the player.
- ***Position:*** This is the position the player plays on the pitch.
- ***Club_involved:*** This is the club who our main club is dealing with.
- ***Fee:*** This column tells us the amount the player was bought or sold for or if the player retired, was a loan deal, or was a free transfer.
- ***Transfer_movement:*** This column tells us whether the player came in or out of the club. Returns the values "in" or "out".
- ***Transfer_period:*** This tells us whether the deal was a summer of winter deal.
- ***Fee_cleaned:*** This is the fee after it has been cleaned this is needed to do the analysis. Any free transfers or loans have been changed to 0.
- ***League_name:*** This tells us the league of our main club involved in the deal.
- ***Year:*** This is the year in which the deal was done.
- ***Season:*** This tells us the season of the deal e.g., "2020/2021"

2.3 Data Implementation

Figure 3 illustrates the methodology from an implementation perspective. The original Premier League and Championship datasets were retrieved from GitHub as zipped files. The 12 csv files that comprise the raw data were transformed in r studio. Additional subsets were created based off the main datasets for a particular result shown in the results section below. The transformed datasets then had to have some of the data cleaned as some of the information was not presented correctly. An analysis was then carried out on the cleaned data frames with all of them being compared to each other within RStudio, Excel, and SPSS.

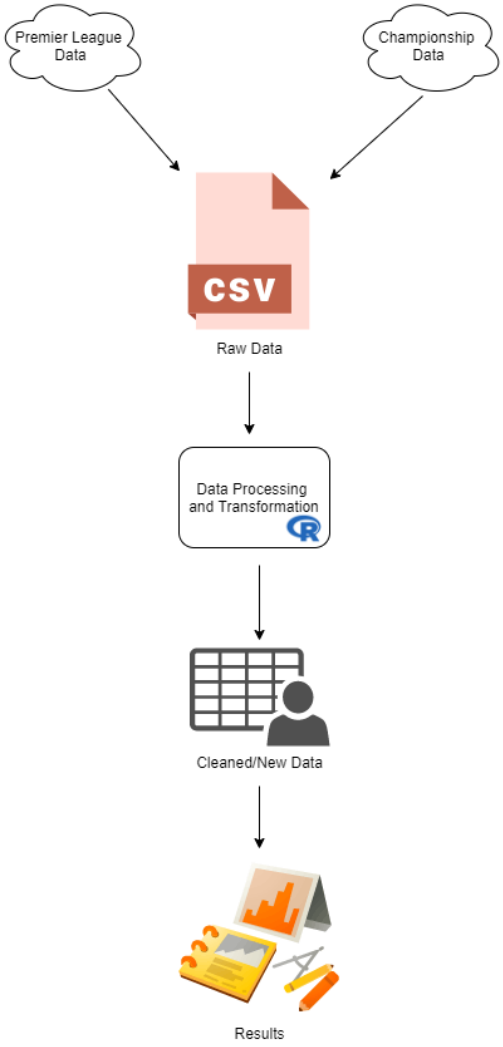


Figure 3: Implementation Architecture

3.0 Methodology

Throughout this project the KDD methodology will be adhered to. An image and an explanation of each section can be seen below.

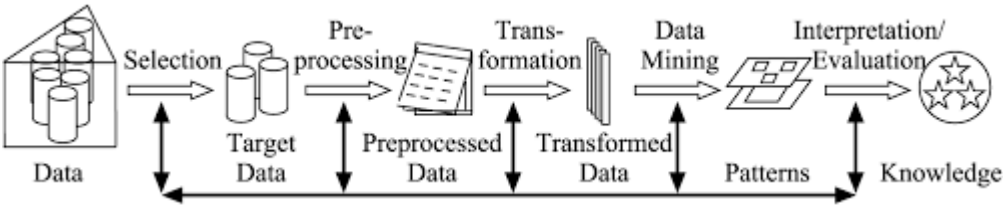


Figure 4: KDD Methodology

3.1 Selection

This part of the KDD is where the person executing the project must go and search any data which will be relevant to the topic of the project. After this is done the data is stored and is ready to be processed. For the purpose of this project some key pieces of information were needed to carry out the objectives of the project. This included information on how much football teams in England have between the years 2015-2020 on transfers, the number of

players they have bought or sold for each of those seasons, and the names of the induvial players which were purchased or sold. All of the datasets were in csv formats when taken from the web with all of them having some missing values, however, these missing values did not affect the outcome of the project as they were negligible pieces of information such as players who were signed or sold as a 'Free Transfer' would default to 'NA' in the 'fee_cleaned' section.

```
#LOADING DATA
PremierLeague2020 <- read.csv("english_premier_league_2020.csv", header = TRUE, sep = ",")
PremierLeague2020PlayersIn <- read.csv("Prem2020PlayersIn.csv", header = TRUE, sep = ",")
PremierLeague2019 <- read.csv("english_premier_league_2019.csv", header = TRUE, sep = ",")
PremierLeague2019PlayersIn <- read.csv("Prem2019PlayersIn.csv", header = TRUE, sep = ",")
PremierLeague2018 <- read.csv("english_premier_league_2018.csv", header = TRUE, sep = ",")
PremierLeague2018PlayersIn <- read.csv("Prem2018PlayersIn.csv", header = TRUE, sep = ",")
PremierLeague2017 <- read.csv("english_premier_league_2017.csv", header = TRUE, sep = ",")
PremierLeague2016 <- read.csv("english_premier_league_2016.csv", header = TRUE, sep = ",")
PremierLeague2015 <- read.csv("english_premier_league_2015.csv", header = TRUE, sep = ",")

Championship2020 <- read.csv("english_championship_2020.csv", header = TRUE, sep = ",")
Championship2020PlayersIn <- read.csv("Champ2020PlayersIn.csv", header = TRUE, sep = ",")
Championship2019 <- read.csv("english_championship_2019.csv", header = TRUE, sep = ",")
Championship2019PlayersIn <- read.csv("Champ2019PlayersIn.csv", header = TRUE, sep = ",")
Championship2018 <- read.csv("english_championship_2018.csv", header = TRUE, sep = ",")
Championship2018PlayersIn <- read.csv("Champ2018PlayersIn.csv", header = TRUE, sep = ",")
Championship2017 <- read.csv("english_championship_2017.csv", header = TRUE, sep = ",")
Championship2016 <- read.csv("english_championship_2016.csv", header = TRUE, sep = ",")
Championship2015 <- read.csv("english_championship_2015.csv", header = TRUE, sep = ",")


#EXPORTING DATASETS SO THEY CAN BE SORTED BY PLAYERS IN
write.csv(PremierLeague2020, "Prem2020PlayersIn.csv")
write.csv(PremierLeague2019, "Prem2019PlayersIn.csv")
write.csv(PremierLeague2018, "Prem2018PlayersIn.csv")

write.csv(Championship2020, "Champ2020PlayersIn.csv")
write.csv(Championship2019, "Champ2019PlayersIn.csv")
write.csv(Championship2018, "Champ2018PlayersIn.csv")
```

The code above shows how each of the Csv files were loaded into RStudio and then exported so that new subsets could be created which only included details on the 'Players In'.

3.2 Processing and Transformation

This part of the KDD is where the project owner has to process their data. The data is cleaned and trimmed down in size so that it is optimal for use in the project. After I was satisfied with the chosen data, I had to start cleaning it which was a tedious process. In the previous section I mentioned 'NA' values in the datasets, these were converted to 0 in RStudio. Also, some of the players' names were outputting incorrectly this is due to them being foreign names so Excel/RStudio struggled to recognise these special characters. The names were transformed to their correct format in RStudio, the only way to do this was to physically find out the players correct name and manually change it through code in RStudio, an example can be seen above in Figures 1 and 2 in section 2 of the report. A similar foreign naming issue came up with some of the foreign teams in my data, the same process used for transforming player names was used for transforming club names. New datasets had to be created from my larger datasets, these new datasets only contained the

players who had the value 'In' in the 'transfer_movement' column. This was done by taking the cleaned data and taking a subset of it in order to carry out data visualisations.

3.3 Data Mining

This part of the KDD is where the project owner will use code to search though the data to interesting insights and patterns in the data. For this project multiple data from different leagues in England will be compared to find some insights into the money spent on transfers. The insights into the data were found through a combination of Excel, SPSS, and RStudio. Anova tests and Forecasting where carried out to analyse the data while also using data visualisations to represent the results.

3.4 Interpretation/Evaluation

This part of the KDD is where the project owner will take their mined data and interpret it visually to represent their insights. The visualisations for this project were created in Excel, SPSS, and RStudio. These programs were chosen as they have been used a lot in my past projects, so I have sufficient experience with them.

4.0 Analysis

The purpose of this project is to carry out an analysis on the transfers done by football clubs in England, I will be using teams in the top 2 leagues. The goal is to analyse their transfers pre Covid and see if Covid has affected them for the 2020 transfer window. The teams were analysed by comparing the amount of money spent on transfers from year to year. For each set of comparisons both leagues were looked at and analysed to see how each league was affected by Covid and which league was more affected. The reasoning for this is because the Premier League is a very wealthy league compared to the Championship so chances are the teams in the Premier League would better cope with a global pandemic.

4.1 Techniques

- Data Visualisation was carried out in order to represent my findings and present them in an easy-to-understand way for the reader. This was done by using a mix of Excel, SPSS and RStudio. A collection of graphs was created for each set of results. These graphs look at aspects of the leagues such as the amount of players being bought and sold, the overall spending of each league year to year, and the individual club spending from year to year.
- Statistical Analysis was carried out in order to collect and interpret my data. I hope to find year by year trends and see if they have been affected by Covid. A form of analysis chosen for this project was an Anova test to assess whether there was a definitive difference in the mean amount of money spent each year on players by clubs in both leagues.

5.0 Results

The results which have been collected in this project will be split into their own sections in order to allow the reader to fully comprehend each piece of work instead of having to read pages of writing. I will title each section in relation to what I am trying to analyse.

5.1 Players In vs Players Out

This section will simply analyse the players in against the players out for each year to see if there are any changing trends. The data will be represented by looking at the total number of players in and comparing it to the numbers of players out. While this is a generalised way of looking at this it gives a simple insight into the changes in transfer movement year to year. After this the spending of each club will be represented through charts produced in RStudio.

5.1.1 Premier League

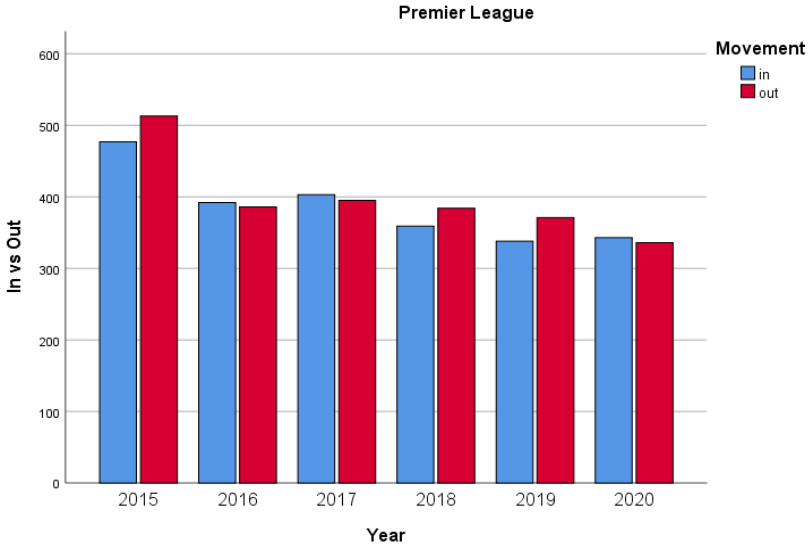


Figure 5: Premier League Players In vs Out

Figure 5 shows the total players in and out for the premier league each year. It is clear to see that year to year there is not a massive difference except for a big spike in 2015. There is a gradual decrease towards 2020 but it is only slight, not enough to say the numbers were affected by Covid. This is not a surprise as the wealth of the clubs in this league is more than enough to sustain them throughout the Covid period. The next section of these results shows the money spent year to year which further backs this up.

5.1.2 Championship

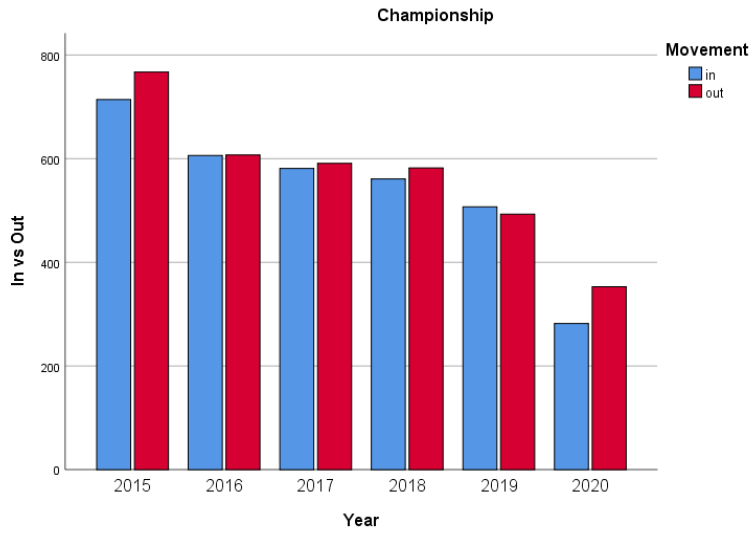


Figure 6: Championship Players In vs Out

Figure 6 is a sharp contrast to figure 4. This shows the players in and out for the championship, this graph is interesting as it shows a similar spike in 2015 to the Premier League and a steady hold 2016-2019. However, there is a massive drop off in 2020 for both players in and players out. This illustrates that the clubs were able to steadily bring in and let go of around 500-600 players each year up until the current season where they only managed a total of 282 in and 353 out. Having seen 5 years of a steady amount of players inbound and outbound to a swift decrease it is safe to say the financial effects of Covid hit the Championship hard. This graph is simply the numbers of players in and out with no regard to the amount of money spent, however, this is explored in the next section below.

5.1.3 Comparison of Yearly Spending

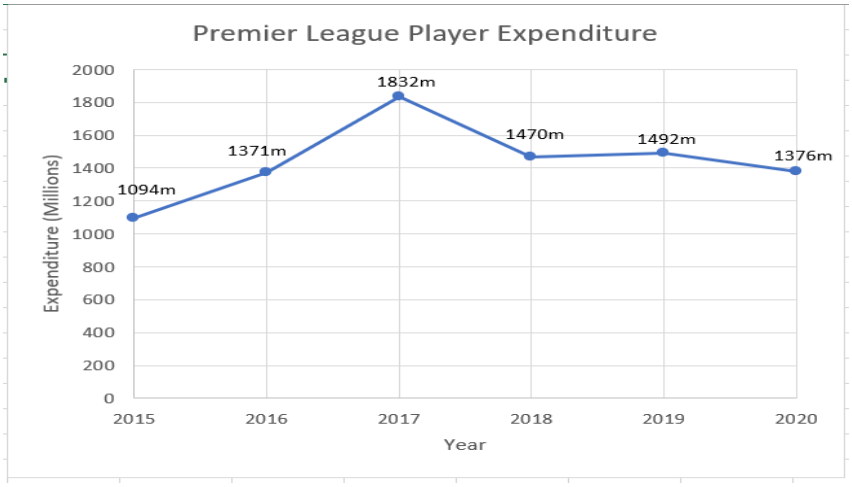


Figure 7: PremierLeague Spending 2015 -> 2020

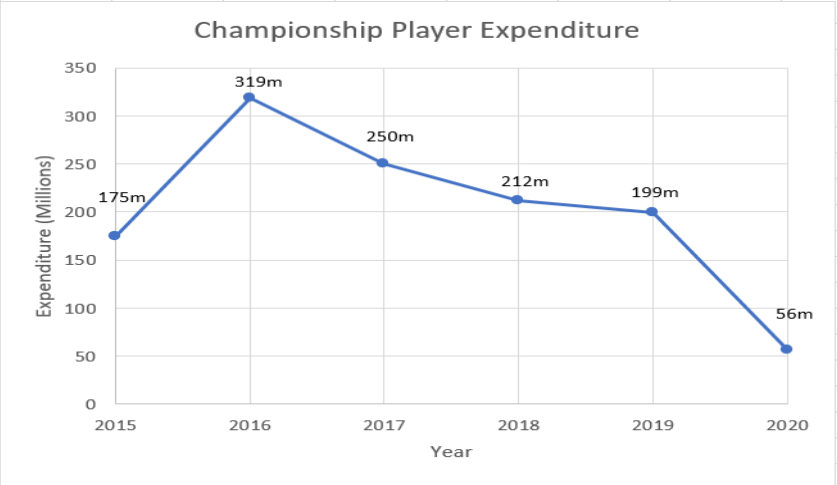


Figure 8: Championship Spending 2015 -> 2020

Figure 7 was created in Excel, it shows the total money spent on transfers in the Premier League for the past 6 years. All 6 of my Premier League datasets were combined by using the 'in' players from the 'transfer_movement' column and adding up all of their fees from the 'fee_cleaned' column. By comparing it to the Ins and Out graph above we can see that the high number of players in and out in 2015 does not directly correlate to a high amount of money spent on transfers as 2015 was the lowest of all 6 years in terms of money spent. However, it does back up the point made above about Covid not having that big of an impact on the Premier Leagues spending. We can see a steady amount spent each season with the lowest being in 2015 and a slight spike in 2017, again the reasoning for this would be the value of each of these clubs. For example, Manchester United, Manchester City and Liverpool were all worth over 2 billion pounds back in 2020 before Covid hit as reported by the Irish Mirror (Polden, 2021). This means that they along with all of the other clubs in the league are more than capable of dealing with a situation such as Covid.

Figure 8 was created in Excel, shows the total money spent on transfers in the Championship for the past 6 years. The same strategy as the Premier League was followed by combining all 6 of the Championship datasets with the 'ins' from the 'transfer_movement' column and added the fees from the 'fee_cleaned' column. Figure 7 follows the same trend as figure 5 with a relatively steady 5 years before a big drop off in 2020. If we compare figure 7 to figure 6, we can see that from 2015 to 2019 they are quite similar in relation to there being a steady level of spending but for the Championship graph once Covid hits in 2020 there is a big drop off in spending. This can be put down to Championship clubs not being worth as much as the Premier League ones if we look at clubs such as Bournemouth and Norwich their teams alone are only worth just over £100 million meaning the likelihood of them spending large amount of money on transfer during a pandemic is low (Transfermarkt, n.d.).

From the graphs above I can concluded that Covid had more of an effect on the Championship than the Premier League with the Championships total spending in 2020 being the lowest it has been in the last 5 years. The decrease in the spending of both the leagues can be attributed to players value decreasing due to the matches being postponed

for so long during 2020. The CIES Football Observatory reported that overall, across the top 5 leagues in Europe player transfer value would decrease by about 28% (Poli, 2020), so this statistic could also be applied to the teams decrease in spending.

5.2 Anova

A One Way Anova test was carried in SPSS out on both the Premier League and Championship Datasets to investigate if there were any differences in the means of the money spent on players.

5.2.1 Premier League

Null Hypothesis (H0): There is no differences in the means of amount spent each year.

Alternate Hypothesis (H1): There is a difference in the means of amount spent each year.

Significance level (α) 0.05

This test was conducted on the 'Amount_Spent' column from the Premier League dataset for each year. An Anova test was carried out to determine if there was any difference in the mean amount spent of each year. The test was performed with a 95% confidence interval.

ANOVA

Amount_Spent

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1222.302	5	244.460	2.568	.025
Within Groups	214666.708	2255	95.196		
Total	215889.009	2260			

Figure 9: Anova Results for Premier League Data

From the results in figure 9 we can see that the degrees of freedom between groups is 5 and within groups 2255, when these values are used in a statistical table for the distribution of F, a critical value of 2.21 is the result which is less than the F value of 2.568 in the table above. The significant P value of 0.025 is less than the declared alpha value of 0.05. By taking both of these results into consideration I can reject the Null Hypothesis.

Multiple Comparisons

Dependent Variable: Amount_Spent

Tukey HSD

(I) Year	(J) Year	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
2015	2016	-1.10776	.66861	.561	-3.0147	.7992
	2017	-2.19833*	.66439	.012	-4.0933	-.3034
	2018	-1.38895	.68288	.323	-3.3366	.5587
	2019	-1.70507	.69734	.141	-3.6940	.2839
	2020	-1.61564	.70964	.204	-3.6396	.4084
2016	2015	1.10776	.66861	.561	-.7992	3.0147
	2017	-1.09056	.69830	.624	-3.0822	.9011
	2018	-.28119	.71591	.999	-2.3231	1.7607
	2019	-.59731	.72972	.964	-2.6786	1.4840
	2020	-.50788	.74148	.984	-2.6227	1.6069
2017	2015	2.19833*	.66439	.012	.3034	4.0933
	2016	1.09056	.69830	.624	-.9011	3.0822
	2018	.80938	.71198	.866	-1.2213	2.8401
	2019	.49326	.72586	.984	-1.5770	2.5635
	2020	.58268	.73768	.969	-1.5213	2.6867
2018	2015	1.38895	.68288	.323	-.5587	3.3366
	2016	.28119	.71591	.999	-1.7607	2.3231
	2017	-.80938	.71198	.866	-2.8401	1.2213
	2019	-.31612	.74282	.998	-2.4348	1.8025
	2020	-.22670	.75438	1.000	-2.3783	1.9249
2019	2015	1.70507	.69734	.141	-.2839	3.6940
	2016	.59731	.72972	.964	-1.4840	2.6786
	2017	-.49326	.72586	.984	-2.5635	1.5770
	2018	.31612	.74282	.998	-1.8025	2.4348
	2020	.08943	.76749	1.000	-2.0996	2.2784
2020	2015	1.61564	.70964	.204	-.4084	3.6396
	2016	.50788	.74148	.984	-1.6069	2.6227
	2017	-.58268	.73768	.969	-2.6867	1.5213
	2018	.22670	.75438	1.000	-1.9249	2.3783
	2019	-.08943	.76749	1.000	-2.2784	2.0996

*. The mean difference is significant at the 0.05 level.

Figure 10: Tukey Test for Premier League

A Tukey Test was carried out alongside the Anova test to clearly represent which years the differences occurred. If we focus on 2015, we can see that 2015 vs 2016 returned a significant value of .561 which tells me that there is not a significant difference between those two years. However, if we compare 2015 to 2017 the significant values come to .012 which tells me that there was a significant difference between those two years.

5.2.2 Championship

Null Hypothesis (H0): There is no differences in the means of each year.

Alternate Hypothesis (H1): There is a difference in the means of each year.

Significance level (α) 0.05

This test was conducted on the 'Amount_Spent' column from the Championship dataset for each year. An Anova test was carried out to determine if there was any difference in the mean amount spent of each year. The test was performed with a 95% confidence interval.

ANOVA

Amount_Spent

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	36.123	5	7.225	4.059	.001
Within Groups	5776.130	3245	1.780		
Total	5812.253	3250			

Figure 11: Anova Results for Championship data

From the results above we can see that the degrees of freedom between groups is 5 and within groups 3245, when these values are used in a statistical table for the distribution of F, a critical value of 2.21 is the result which is less than the F value of 4.059 in the table above. The significant P value of 0.001 is less than the declared alpha value of 0.05. By taking both of these results into consideration I can reject the Null Hypothesis.

Multiple Comparisons

Dependent Variable: Amount_Spent

Tukey HSD

(I) Year	(J) Year	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
2015	2016	-.26678*	.07369	.004	-.4769	-.0567
	2017	-.18389	.07454	.134	-.3964	.0287
	2018	-.11795	.07527	.621	-.3326	.0967
	2019	-.16503	.07748	.272	-.3860	.0559
	2020	.06274	.09384	.985	-.2048	.3303
2016	2015	.26678*	.07369	.004	.0567	.4769
	2017	.08290	.07747	.893	-.1380	.3038
	2018	.14884	.07817	.400	-.0741	.3717
	2019	.10175	.08030	.803	-.1272	.3307
	2020	.32952*	.09617	.008	.0553	.6038
2017	2015	.18389	.07454	.134	-.0287	.3964
	2016	-.08290	.07747	.893	-.3038	.1380
	2018	.06594	.07897	.961	-.1592	.2911
	2019	.01885	.08108	1.000	-.2123	.2501
	2020	.24662	.09683	.111	-.0295	.5227
2018	2015	.11795	.07527	.621	-.0967	.3326
	2016	-.14884	.07817	.400	-.3717	.0741
	2017	-.06594	.07897	.961	-.2911	.1592
	2019	-.04709	.08175	.993	-.2802	.1860
	2020	.18068	.09739	.430	-.0970	.4584
2019	2015	.16503	.07748	.272	-.0559	.3860
	2016	-.10175	.08030	.803	-.3307	.1272
	2017	-.01885	.08108	1.000	-.2501	.2123
	2018	.04709	.08175	.993	-.1860	.2802
	2020	.22777	.09911	.195	-.0548	.5104
2020	2015	-.06274	.09384	.985	-.3303	.2048
	2016	-.32952*	.09617	.008	-.6038	-.0553
	2017	-.24662	.09683	.111	-.5227	.0295
	2018	-.18068	.09739	.430	-.4584	.0970
	2019	-.22777	.09911	.195	-.5104	.0548

*. The mean difference is significant at the 0.05 level.

Figure 12: Tukey Test for Championship

Another Tukey test was carried out this time on the Championship data, looking at the differences from year to year. The major differences occurred between 2015/2016 and between 2016/2020. This data also shows that 2017, 2018, and 2019 are not significantly different to the other years.

5.3 Forecasting

Forecasting is used as a tool for predicting future demand based on past demand for the product or service. This can be applied to my data for the amount spent each year in the Premier League and the Championship to potentially predict whether there will be an increase or decrease in spending for the 2021 season as a result of Covid still being a threat to a club's ability to spend lots of money on players. Forecasts are generally wrong when compared to the actual data but the goal here is not to predict specific numbers it is to simply come up with a prediction for either an increase or decrease in spending.

5.3.1 Premier League

Year	Premier League		Simple Moving Average	Weighted Moving Average	Exponential Smoothing		
2015	1094					Initial Forecast	1400
2016	1371				1277.60	α = 0.4	
2017	1832				1388.40		
2018	1470		1432.33	1555.33	1572.80		
2019	1492		1557.67	1574.17	1428.00		
2020	1376		1598.00	1541.33	1436.80		
2021			1446.00	1430.33	1390.40		

Figure 13: Forecasting table for Premier League

The table above represents three forecasting methods for my Premier League data. For the weighted moving averages, the chosen weights were 3, 2, and 1. For the exponential smoothing the initial forecast was set at 1400 as that number closely represented the data in the 'Actual Amount' column with the smoothing constant being set at 0.4. The exponential smoothing performed the best this could be down to it taking more years into consideration for the calculation. When 'Simple Moving Average' and 'Weighted Moving Average' are compared the results show them both returning similar values with the Simple Moving Average performing slightly better for 2018 and 2019 and the Weighted Moving Average performing better for 2019. All 3 of the calculations are predicting a further downward trend in spending, this has a good chance of happening as Covid is still ongoing so teams may not have enough money to spend on new players. Deloitte predicted that clubs in the Premier League face a loss of around £1billion due to Covid, this is down to the loss of matchday revenue and rebates to broadcasters (Davies, 2020). This prediction further forecasts a continued downturn in spending for the next season.

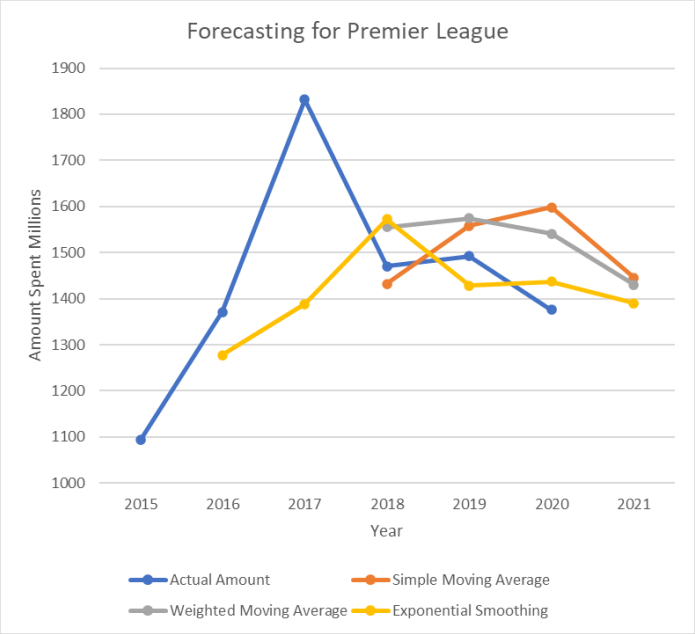


Figure 14: Line graph for Premier League

Figure 14 shows that overall Exponential Smoothing did a good job at representing the Actual Amount data, the returned values were close to the actual data taking the downward trends of the later years of the data into consideration. From 2018 onwards all 3 of the calculations performed well, however, it is worth noting that the 'Weighted Moving Average' was the only calculation which decreased in value for our Covid year of 2020, the other 2 calculations predicted an increase in spending for 2020 from 2019 but this was not the case with the actual data.

5.3.2 Championship

Year	Actual Amount		Simple Moving Average	Weighted Moving Average	Exponential Smoothing		
2015	175					Initial Forecast	200
2016	319				190.00	α = 0.4	
2017	250				247.60		
2018	212		248.00	260.50	220.00		
2019	199		260.33	242.50	204.80		
2020	56		220.33	211.83	199.60		
2021			155.67	129.67	142.40		

Figure 15: Forecasting table for Championship

The table above represent three forecasting methods for my Championship data. For the weighted moving averages, the chosen weights were 3, 2, and 1. For the exponential smoothing the initial forecast was set at 200 as that number closely represented the data in the 'Actual Amount' column with the smoothing constant being set at 0.4. The exponential smoothing performed the best this could be down to it taking more years into consideration for the calculation. When 'Simple Moving Average' and 'Weighted Moving Average' are compared the results show them both returning similar values with the Weighted Moving Average performing slightly better.

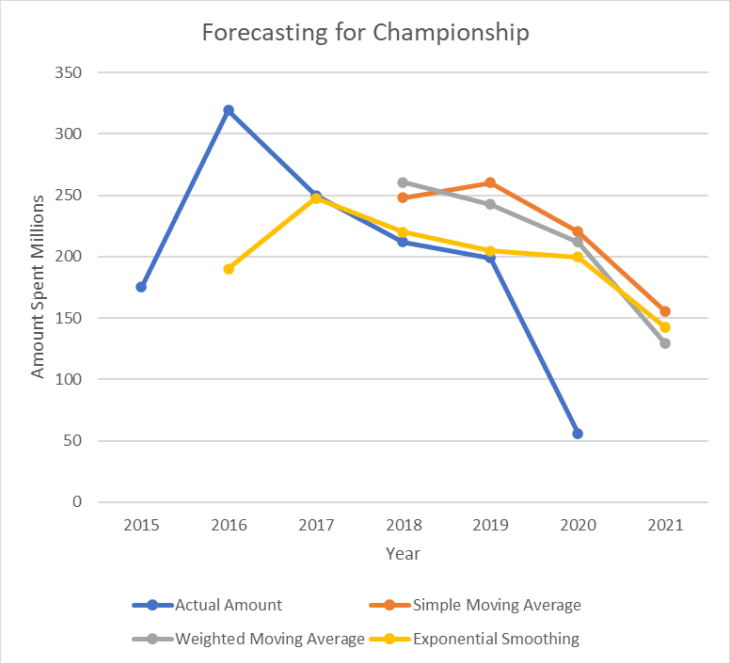


Figure 16: Line graph for Championship

As shown by the results in figure 16 it is clear to see that the results from the exponential smoothing closer represented the results of the years 2017, 2018, 2019, and 2020 when compared to the other 2 averages. For the years 2020 and 2021 all 3 methods followed a downward trend. These 2021 results would have to be compared to real life results in the future when the next set of transfers take place.

5.4 Data Visualisations of Clubs Spending

The following visualisations were executed by means of code in RStudio. They represent the individual clubs and how much they have spent for the years 2018, 2019, and 2020. The point of these graphs is to represent the general difference in club-to-club spending for the Covid year against normal years with some specific clubs being focused on. 2018 and 2019 were chosen because they represent a good baseline to compare the Covid data to, for example, 2017 had a massive spike in spending so this was not chosen for this particular comparison. The code below is what was used to display the graphs in this section, the example here is for the Premier League 2020 data.

```
#PREMIER LEAGUE PLOTS
ggplot(PremierLeague2020PlayersIn, aes(x=fee_cleaned, y=club_name,)) +
  geom_bar(stat = "identity") +
  labs(title="Barplot Of Money Spent By Premier League Clubs in 2020",
       x="Money Spent", y= "Team")
```

5.4.1 Premier League

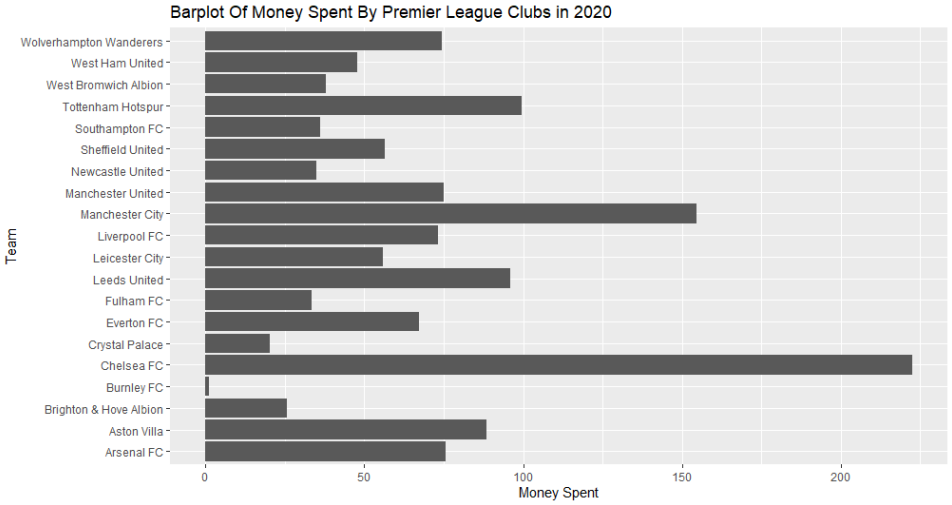


Figure 17: Individual Premier league Club Expenditure 2020

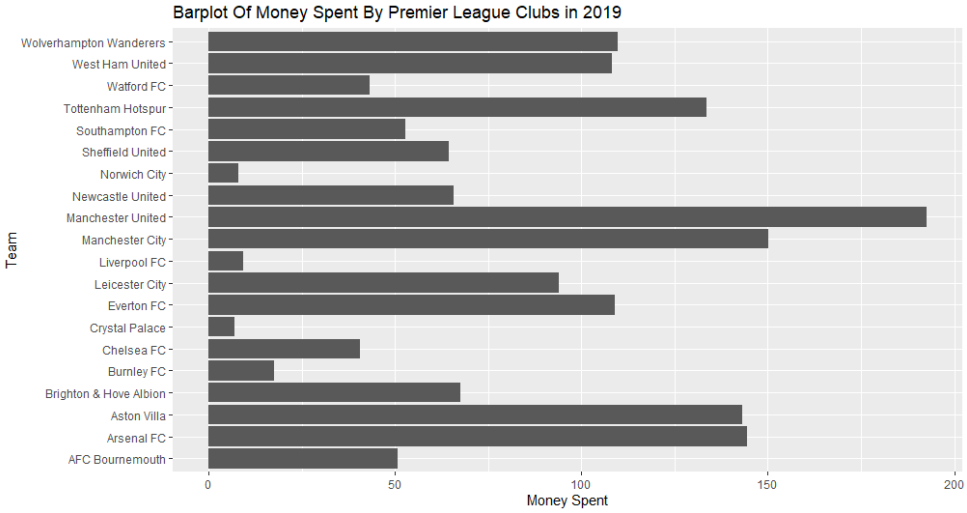


Figure 18: Individual Premier league Club Expenditure 2019

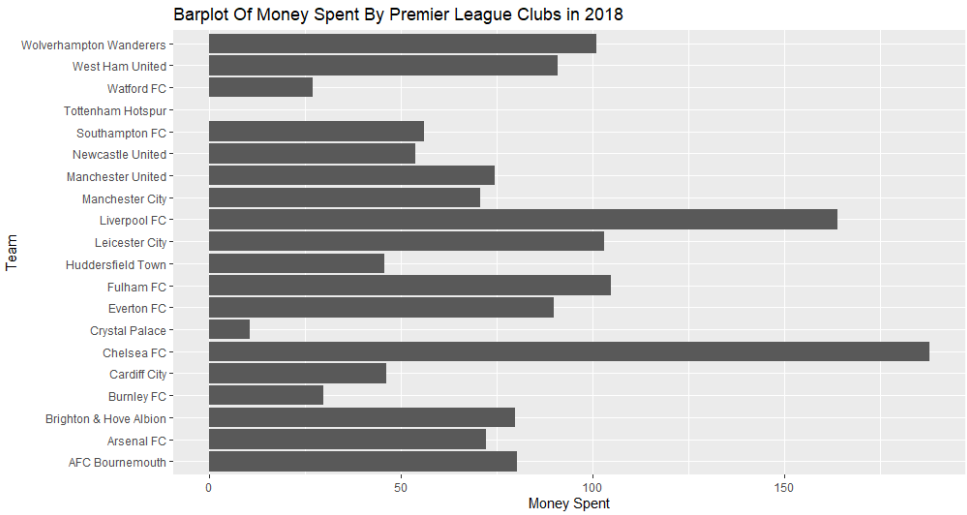


Figure 19: Individual Premier league Club Expenditure 2018

From the boxplots above we can see that the majority of clubs spent more money in the years 2018 and 2019, for both years more teams spent between £50m-£100m and £100m+ than they did in 2020. If we look at Burnley FC for 2018 and 2019, they were able to spend around £20m-£40m for each year but for 2020 they spent next to nothing compared to every other club, this is due to Burnley being a small club so they would have had less money to spend on players due to Covid. The drop off in the number of clubs who spent £100m+ in 2020 was notable, the number went from 5 and 8 in 2018 and 2019 to just 2 clubs in 2020 with most clubs falling in the £50m-£100m range. This tells me that despite Premier League clubs still spending an outrageous amount of money on players, Covid still had an impact on how much they were able to spend.

5.4.2 Championship

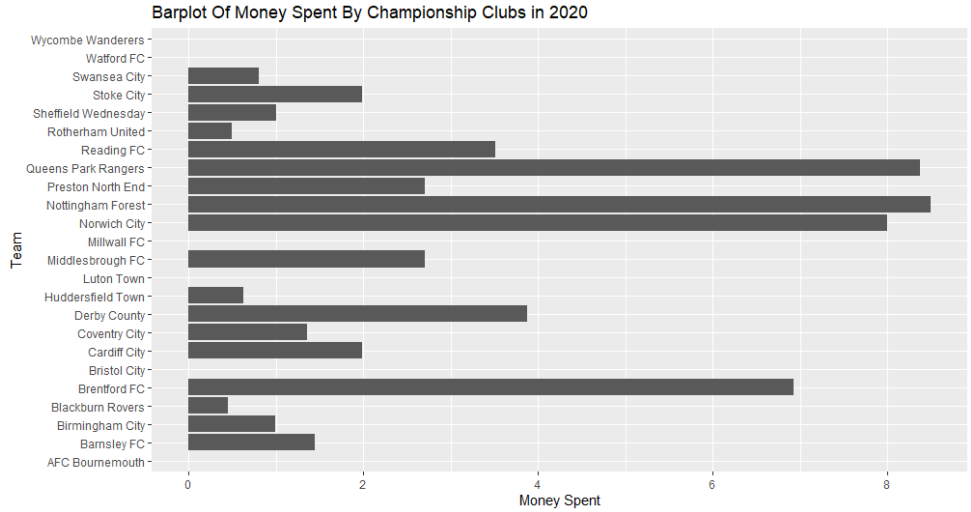


Figure 20: Individual Championship Club Expenditure 2020

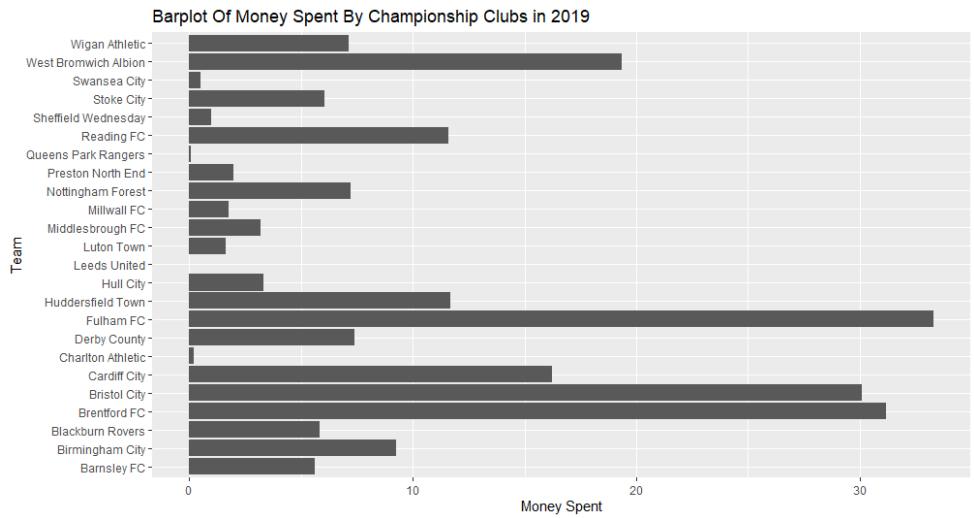


Figure 21: Individual Championship Club Expenditure 2019

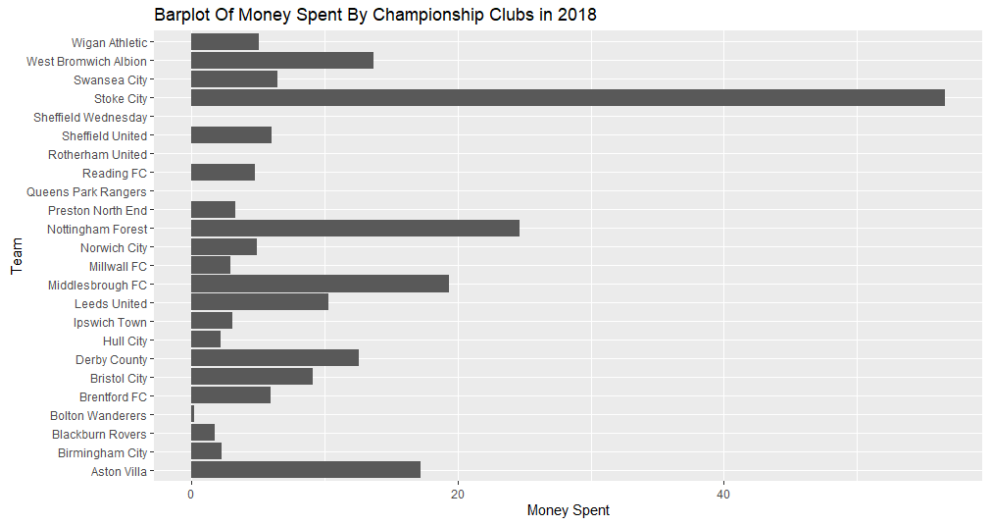


Figure 22: Individual Championship Club Expenditure 2018

From the boxplots above we can see that the majority of clubs spent more money in the years 2018 and 2019, for both years more teams spent between £10m-£20m and £20m+ than they did in 2020, with every single team in 2020 spending less than £10m. Not many clubs stay in this league for long as each season 3 get promoted to the league above (Premier League) and 3 get relegated to the league below (League 1) so I will focus on Bristol City who managed to stay in this league for the 3 chosen years. They were able to spend between £9m-£30m for 2018 and 2019 but in 2020 they were not able to spend anything. Bristol was only 1 of 6 clubs who spent nothing in 2020, this number was only 3 and 1 in the other 2 years, this again shows the impact Covid had on this league.

Both sets of graphs show how Covid has affected the 2 leagues, but they further emphasise how hard it hit the Championship. Clubs here could not manage to spend over £10m whereas teams in the Premier League were able to spend over £50m no problem.

6.0 Conclusions

In conclusion, based on the acquired data this project has found that Covid had an impact on the amount of money spent by clubs in both the Premier League and the Championship with latter being hit the most. This project showed the year-to-year spending of each club and how it dropped off for 2020, while no definitive reasoning was investigated as to why this decrease in spending occurred a strong assumption can be made based off of the researched literature and media reports found online. To recap on the results, they showed that from year to year there was a steady amount of money spent across both leagues with a slight drop off in 2020 for the Premier League and a massive drop off for the Championship, this is simply down to the Championship clubs being worth less than the Premier League clubs. Forecasting was carried out to make a prediction on how the spending will change for 2021, the 3 models all predicted a further decrease in spending. This makes sense because as of May 2021 fans still have not being allowed to attend games meaning every club is still missing out on a massive weekly income. An Anova test was carried out on both sets of data to investigate if there was a difference in the mean amount of money spent each year, with a Tukey test carried out alongside to compare year to year.

The results showed that there was a difference which further backs up the fact that Covid has affected both leagues.

An advantage of this project is that it gives an interesting insight into how each league was affected by Covid. It showed that the Premier League was able to withstand the initial wave of Covid and if there was any downturn in spending it was only slight, whereas the Championship the hit a lot harder with clubs only spending 28% of what they spent in the previous season. Another advantage of this project is the ability to predict a further downturn in spending for next season, when the collected data and researched literature are combined it is easy to conclude that the clubs will more than likely spend less than they did for this season.

A disadvantage of this project is that it only looks at the top 2 leagues in English football, ideally, I would have compared some of the smaller leagues in England given more time as there are not an adequate amount of datasets on these leagues so I would have to create them myself. This was not realistic for this particular project as it would have consumed massive amounts of time which was not available due to having multiple other CA's due this year that needed time and dedication.

7.0 Further Development or Research

Given more time to carry out this project the analysis could be applied to further leagues in the UK, Europe and beyond. I would look to collect data on different clubs and leagues in different countries and see if Covid had different effects on these leagues. Lockdown measures differ from country to country with some countries allowing fans back into stadiums before England did, this means some more earning for these clubs which could be invested into future transfer for them.

Another analysis I would like to carry out would be to see if the earnings for each club had an impact on the spending of clubs. Some clubs may be willing to make a loss and spend more money on players even though their earnings would have decreased compared to pre-Covid levels if it meant that they could sign some really talented players, and other clubs will not be willing to do this. I feel the results would be interesting as not all clubs would earn the same so there is a chance of some surprises coming up. A way of doing this would be by using Pearson's r correlation to see if there is a linear relation between 2 variables, the variables here would be money earned and money spent.

I believe the analysis I carried out could be applied to different sports. An interesting field of study would be American sports as there is massive amounts of money involved. Basketball and American Football would be 2 sports I would like to focus on as I have a personal interest in them and as I said these are very lucrative sports, players earn massive wages and the ecosystem as a whole has lots of money involved, when it comes to the stadiums and facilities, they are state of the art. Back in 2018 Chron reported that the median salary for NFL players was $860,000 with a starting one-year rookie having a minimum income of $435,000 (Woodruff, 2018). This tells me that it is a lucrative sport so it would be interesting to see how it handled the Covid pandemic.

I would like to revisit this project in 1 years' time to again analyse the Premier League and Championship clubs to see how their spending changed for the 2021 season. I would look at the forecasts I made in this report and see if they were correct in predicting a further decrease in spending across both of the leagues. An early indication of this would be the January 2021 transfer window which saw clubs in the Premier League spend just £70m which is the lowest amount spent in this window since 2012, in 2020 just before Covid hit £230m was spent (Ford Rojas, 2021).

8.0 References

- Hamilton, P. (2020). Irish SMEs face revenue shortfall of up to €15bn this year, ESRI warns. Retrieved 21 December 2020, from https://www.irishtimes.com/business/economy/irish-smes-face-revenue-shortfall-of-up-to-15bn-this-year-esri-warns-1.4354750
- ewenme. (2020). Retrieved 21 December 2020, from https://github.com/ewenme/transfers
- Bruce, A. (2020). Big UK businesses predict a slow economic recovery. Retrieved 21 December 2020, from https://www.weforum.org/agenda/2020/07/big-uk-businesses-slow-recovery-deloitte-survey/
- Polden, J., 2021. *Liverpool and Spurs climbed in 2020 richest club rankings before Covid pandemic*. [online] mirror. Available at: <https://www.mirror.co.uk/sport/football/news/premier-leagues-top-six-most-23352954> [Accessed 21 April 2021].
- Transfermarkt.co.uk. n.d. Championship - Club market value. [online] Available at: <https://www.transfermarkt.co.uk/championship/marktwerteverein/wettbewerb/GB2> [Accessed 30 April 2021].
- Poli, R., 2020. Pandemic: 28% loss on players' transfer value. [Blog] CIES Football Observatory, Available at: <https://football-observatory.com/IMG/sites/b5wp/2019/wp289/en/> [Accessed 11 May 2021].
- Woodruff, J., 2018. How Much Money Do Baseball Players Make?. [online] Work - Chron.com. Available at: <https://work.chron.com/much-money-baseball-players-make-14452.html> [Accessed 16 May 2021].
- Davies, M., 2020. Premier League clubs facing £1billion revenue loss due to coronavirus crisis, Deloitte predicts. Evening Standard, [online] Available at: <https://www.standard.co.uk/sport/football/premier-league-clubs-facing-ps1billion-revenue-loss-coronavirus-deloitte-a4465631.html> [Accessed 16 May 2021].
- Ford Rojas, J., 2021. January transfer window spending slumps as COVID pandemic hits football finances. [online] Sky News. Available at: <https://news.sky.com/story/january-transfer-spending-slumps-as-pandemic-hits-football-finances-12206482> [Accessed 16 May 2021].

9.0 Appendices

This section should contain information that is supplementary to the main body of the report.

9.1. Project Plan

9.1.1 Objectives

The objectives of this project are to analyse football transfers from the last few years and compare them to this year. I want to see has Coivid-19 had an effect on clubs and if so I want to try and answer questions such as “Have the bigger clubs had it easier throughout the pandemic compared to the smaller clubs” and “How hard have the smaller clubs been hit by the pandemic”. Some small clubs have had to close down so this has prompted me to look at them and how they have coped.

9.1.2 Background

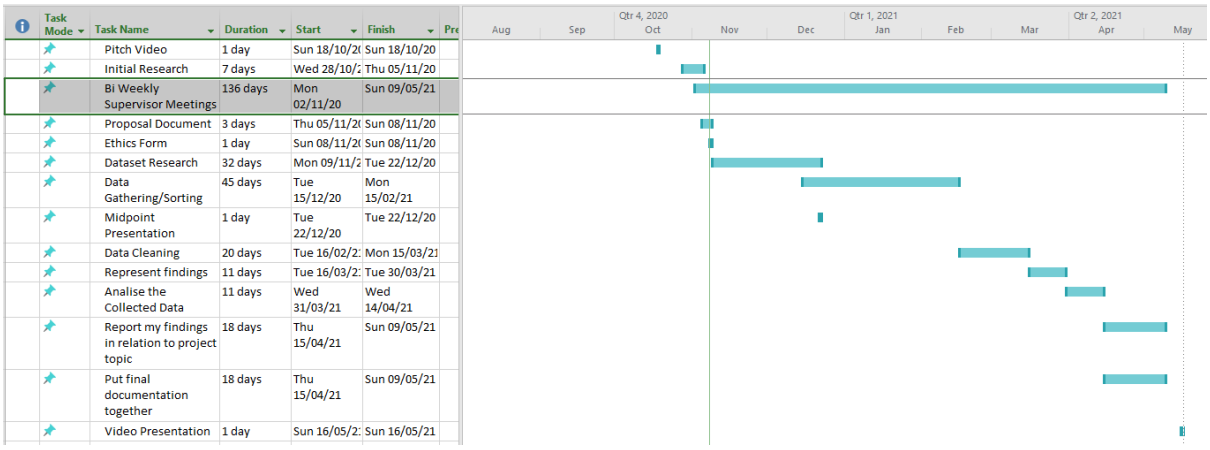
I have always had an interest in football since I was young, so this prompted me into going down this route for the project. When carrying out some research for my choice of topic I came across multiple articles online but a particular one from The Guardian struck me, it talked about how it could take football roughly 10 years to get back to where it used to be. Now obviously this is talking about all aspects of football, but it gave me some inspiration to focus on the transfer side of the game and analyse the changes since Covid-19 struck. There are multiple studies online which look at COVID-19 and football, for example, there is a study on European football after COVID-19 from the University of Reading. Studies like this will give me insights into some areas of the sport which could be relative to my project. As for the specific topic I feel that it will be interesting for everyone involved in the sport to see how the pandemic has effected clubs’ transfer dealings. I think that the idea is fresh in terms of a college project because the pandemic is still quite new for us all, so I am not worried about how unique the idea is. It is well known that all businesses across the world have struggled to cope financially throughout the pandemic and same can be said for football clubs, but I want to look at the actual figures and represent them in this project.

9.1.3 Technical Approach

I will be doing some research in the coming weeks on other projects which are relevant to my subject to help me along with the project, but I also plan on using what I have learned in the Data Application Development module to carry out the bulk of the technical side of the project. I will look at a combination of the languages R and Python the represent the data. At the time of writing this proposal we have not yet been thought Python, just R. We have been using RStudio, but I will look at some other potential programs for this project such as, RapidMiner or Tableau. In order to obtain my data, I will have to take it from multiple websites. Some of the data will simply be downloaded but I plan on using APIs to extract some data and to give my project some more complexity, an example of something like this would be using RStudio to obtain

some simple structured data from static web pages. The information is widely available, but the challenge will be collecting it all and representing it.

9.1.4 Project Plan



Task Mode	Task Name	Duration	Start	Finish	Pre
	Pitch Video	1 day	Sun 18/10/2(	Sun 18/10/20	
	Initial Research	7 days	Wed 28/10/:	Thu 05/11/20	
	Bi Weekly Supervisor Meetings	136 days	Mon 02/11/20	Sun 09/05/21	
	Proposal Document	3 days	Thu 05/11/2(	Sun 08/11/20	
	Ethics Form	1 day	Sun 08/11/2(	Sun 08/11/20	
	Dataset Research	32 days	Mon 09/11/2	Tue 22/12/20	
	Data Gathering/Sorting	45 days	Tue 15/12/20	Mon 15/02/21	
	Midpoint Presentation	1 day	Tue 22/12/20	Tue 22/12/20	
	Data Cleaning	20 days	Tue 16/02/2:	Mon 15/03/21	
	Represent findings	11 days	Tue 16/03/2:	Tue 30/03/21	
	Analise the Collected Data	11 days	Wed 31/03/21	Wed 14/04/21	
	Report my findings in relation to project topic	18 days	Thu 15/04/21	Sun 09/05/21	
	Put final documentation together	18 days	Thu 15/04/21	Sun 09/05/21	
	Video Presentation	1 day	Sun 16/05/2:	Sun 16/05/21	

These are the initial timelines I have set for myself to execute this project. A large amount of time is dedicated to collecting the data and cleaning it up to suit the topic of the project. I have also included any key dates from Moodle. I will meet with my supervisor via Teams biweekly, this has also been included in the timeline.

9.1.5 Technical Details

Various websites to extract data (Kaggle.com would be an example)

R, Python – both will be used in RStudio to carry out the analysis.

API – I hope to make use of APIs to add complexity to the project.

9.1.6 Evaluation

Unit Testing will constantly be carried out within RStudio to make sure the code I am writing is functioning the way I want it to.

Load Testing will be carried out as there is a potential for a lot of data to be involved in this project. For RSudio shiny there is a package called 'shinyloadtest' which is used to generate recordings and analyse results.

Integration Testing will be carried out to make sure the individual pieces or code from the unit tests work well together and with the APIs. I will look at following the Incremental Testing method and combine my chunks of code bit by bit.

I expect to come across different methods of testing throughout this project as I continue my study of this area throughout the year.

9.2. Reflective Journals

9.2.1 October

Reflective Journal

Cian Larkin – x17453136

This month has been different for me and a lot of the other students who spent 3rd year semester 2 on work placement. I have had to get used to the online learning environment as well as get stuck into my new modules. I submitted my project pitch video, which was also a new experience, but it is good to be able to adapt to these situations. My idea did not take me too long the main part was how I go about doing it. I picked a subject that I have an interest in so that I will be enthusiastic for the duration of the project, the project will be to look at how Covid has affected football transfers. I have enjoyed my modules this semester, Business Data Analysis has revolved around statistics and probability and Data Application Development has been good too, I think that both modules will help me with Data Analytics. I feel it is easy enough to follow along with the remote lectures, if one person asks a question in the chat everyone is able to see and the classes are recorded for us to go back on so this will help whenever we are stuck on an assignment and need some help. The labs for each class have been good because the lectures have been using them to allow students to ask questions regarding class material or our assignments. This has been beneficial because the lectures will usually point us in the right direction regarding a solution to whatever the problem is. Ultimately, I feel that I have made a good start to this semester I have been able to get a start of my assignments due to having more free time during lockdown and have been engaged in the classes without any distractions.

9.2.2 November

Reflective Journal - November

Cian Larkin – x17453136

November was a long month with multiple assignments due. At the start of the month, I had my first open book exam for my Business Data Analysis module, it revolved around statistics and probability. I think that it went well but as of writing this journal I have not yet received my results. I had a 2-part CA for my WSAPI module, a split of code and theory. I feel like the theory was fine as it was just 2 exam style questions, but the code was a challenge and I struggled with what had to be done but got there in the end after getting some help. We had our reading week which gave me a chance to catch up on some work that needed to be done. I also enjoyed a good break from the scheduled classes and had time to recharge in time for the final stretch of hard work. I had a pretty sizable project due for my Data App module where I had to conduct an analysis on multiple related datasets and put them into an IEEE report. This was my first time writing in this format, but I felt it went well. The project itself gave me a good idea as to how I want to approach my final year project as it was basically a smaller version of it. I had 2 supervisor meetings with Evgeniia, 1 was in a group and 1 was just us. We discussed what was

expected of me over the next few weeks. Basically, the main focus was to keep doing research online related to my project.

My Reflections

I feel as though this month went well. I was able to keep on top of my classes and the assignments which were due even though it felt like I had something big due each week, I just need to keep up the workflow for the final few weeks and push through until Christmas break.

Changes

I felt as though this month was a solid working month, so I don't plan on making any drastic changes I just need to continue what I've been doing so far.

9.2.3 December

Reflective Journal – December

Cian Larkin – x17453136

December brought semester 1 to a conclusion. I had some tough submissions which were due but, in the end, I am happy with how they turned out. The chess project for our AI module was due this month, it was a 2-part submission in which we had to code up an AI agent to play against us in the game, I am hopeful that I will receive a decent mark. For our Data App module, we had a submission which was due on the 30th. The assignment itself was not too bad but having to do it right after Christmas was tough in terms of cramming the work into a few days and also getting back into the swing of working after a few well needed days off over the Christmas. The biggest submission of this month was my Mid-Point Presentation because it is the highest credited module this year. I am happy with how my presentation video turned out; we have had to record a few videos this semester for other modules so by the time this one came around I felt comfortable about doing it. I had some more meetings with Evgeniia, these helped me to appropriately prepare for my presentation.

My Reflections

I am happy with how this month went. At the beginning I was worried about the number of assignments I had and how I would fit them all in but in the end, everything worked out and I am confident that I will receive good marks for them. The plan for the next week or so is to work on my terminal-based assessments and then take some time off after to prepare for semester 2.

Changes

Similar to last month, I feel my approach to working has been fine as I was well on time with my assignments and didn't feel swamped by the workload.

9.2.4 January

Reflective Journal

Cian Larkin – x17453136

This month composed of TABA submissions and some light project work as I took some time off after a hectic Christmas and New Year. The TABAs went well but there was a lot of work involved for the three of them. Most days I worked from morning to evening on them as there was a good amount of research involved for all of them and I had not got time in December to do them as we had other submissions due. This meant that the workload was crammed into roughly 10 days. After this I took some time off to reset before the start of semester two, I did however continue to do some research for my project.

My Reflections

This was an extremely busy period, but I am happy with the amount of work which I got done. I feel like I did well in the TABA submissions I had so it was worth all of the hours I put it. The break I took after this was needed in order to mentally recover before the final hurdle of semester 2.

Changes

Ultimately, it was another good month time management was a challenge due to having loads of submissions due so maybe I can improve on this for next semester.

9.2.5 February

Reflective Journal

Cian Larkin – x17453136

We started our last 2 modules of the entire programme right at the end of January, Data and Web Mining and Advanced Business Data Analysis. The two modules will be a massive help when it comes to the final project. I feel like there will be skills and techniques learned that I could take into the final project with me. The Data Mining in particular will be interesting as we have not done anything like this so far in the course whereas we have already done Business Data Analysis in the first semester of this year. As far as my Final Project is concerned, I am still in the area of doing

research to try and find some related topics to mine and I am also working through my data and cleaning it as it was messy when I loaded it into RStudio.

My Reflections

I have enjoyed the 2 new modules we are learning I fell like they will contribute to my final project. I am also happy that we only have these 2 modules as it will free up more time to work on the final project.

Changes

My time management has gotten better, I think this is down to getting used to working from home and also only have 2 modules to worry about.

9.2.6 March

Reflective Journal

Cian Larkin – x17453136

This month we got deeper into our 2 modules; I feel like I have learned a decent amount of techniques which can be applied to my final year project. I received my CAs for each of the modules and have already completed the Advanced BDA one with the Data Mining one due in a few weeks. The ABDA CA was fine I spent a week working on it as I did not want to cram it all in to one weekend. I went on some group calls with my friends, and we helped each other do it. This was a massive help as I miss working alongside my friends in the classroom environment. I am confident that I will receive a good mark as I feel the report, I put together was of a good standard and I did what was asked of me. For my project I concluded my research I feel I have enough to go with now and I also cleaned all of my data. The cleaning was a tedious process I am glad it is finished with now.

My Reflections

I am happy with the progress I have made in the 2 modules. It is also good to be finally making good progress with the project with the research I needed being gathered and getting all of my data cleaned in RStudio.

Changes

I do not think I need to make any changes just need to keep up this level of work.

9.2.7 April

Reflective Journal

Cian Larkin – x17453136

This month we finished up our last ever semester of classes in NCI. It was not as busy as some other semesters as I only had 3 assignments due at the end of the month compared to other semesters where I may have had 5 or 6. Despite this I was constantly working on my final project chipping away at the documentation and code so that I am not overwhelmed with work in the final few weeks of May. I'm happy with the three CAs I did, I feel like I did enough work in all of them to get a good grade. I went lots of Teams call with some of the lads in my class and we worked through the questions together like the way we would if we were all in the college together. As for my own final project, I made some significant progress. I got some good results fully wrote up which gave my valuable insights into my data and I also got a lot of the other sections wrote up which sets me up nicely for the final few weeks.

My Reflections

I am happy to finally be finishing up my degree, it has been a long journey, but the end is finally in sight. I feel like I did a good job on the Cas I had due for my 2 modules and again I have made significant progress with the final project. The final few weeks will compile of finishing up my analysis and writing up my findings into the report.

Changes

All I have to do is keep chipping away at the final project and have a strong finish to the year.