National College of Ireland

Bachelor of Science in Technology Management

Data Analytics

Academic Year 2020/2021

Ryan O'Toole

X17409626

X17409626@student.ncirl.ie

A Study into the Distribution of Sports Grants by Geography and Category in Ireland

Technical Report

Contents

Executive Summary

The data analytics report focuses on sports capital investments in Ireland from the 8th of December 2014 to the 31st of August 2020. The objective of the report is to investigate the level of sports capital invested in Ireland by sport and by county. The report seeks to identify the correlation between success of a sport and the level of investment received. The report looks at accolades won in a field of sport versus the level of investment of a given county. The level of investment per county based on the population is analysed and the total sum of investment since 2014.

There are four main sections of the report; a) Data; all datasets are discussed and how each dataset was implemented to the project for analysis; b) Methodology; The methodology followed for the report to pre-process the data was the Knowledge Discovery in Databases and all steps are discussed; c) Analysis; The analysis conducted, and the techniques used for the results are discussed; and d) Results, this section of the report concludes the findings and outcomes of the analysis.

Key Conclusion

From conducting the analysis, it was discovered that Gaelic Games received the highest level of investment of all sports in Ireland. It was found that Cork was the county with the highest mean investment. The results section of the report contains visualisations to simplify and highlight key aspects of the data. It was discovered that Gaelic Games has continued to have the highest level of investment for the previous five years in Ireland with more than two times the level of investment than the second-highest invested sport.

1.0 Introduction

1.1. Background

I gained an interest in sport at a young age and began playing and supporting sport when I started primary school and continue to do so today. I have competed in several different sports such as football, rugby, and cross-country running, sports have always been an essential part of my health and well-being and fitness, enabling me to make great friends and the bonus of achieving accolades in both regional and country competition.

Reflecting on my younger years of playing sport, the level of investment within my local football club always seemed to be underfunded, evidenced by local clubs and teams fundraising locally within the community. Clubs were staffed on a voluntarist basis and contribution to running costs sponsored events and local business, this was the primary method of raising funds for the club and purchasing equipment for the teams. Taking this and similar examples from other areas into consideration led me to contemplate sports capital investment within Ireland and how the money is distributed through sports.

In recent years, Ireland has achieved success in sporting events such as rowing, boxing, rugby, etc. The project concept is to analyse the distribution of sports capital and compare the level of investment versus the return. Ireland has seen accolades in both individual and team sports. For example, in 2012, Ireland received 24 Olympic medals, multiple European championships in boxing, open championships in golf and six nations victories in 2013, 2014 and 2018.

Based on the 2016 Irish Census, the population of the Republic of Ireland is now at 4.9m with an increase in families equating to a growth of 3.3%. This project will investigate if the population of a given county correlate with the level of investment in sports capital.

The analysis undertaken will gain insight into whether sports with a lower investment rate are outperforming sports with a higher investment rate on a global stage. The level of participation in the sport must be considered when conducting the analysis of the sport and calculating the level of success.

Is success in a field being recognised and further rewarded, if a sporting team or athlete from Ireland is successful in their field, is this success rewarded to the sport and is the level of investment increased to maintain this level of success achieved.

1.2. Learning Aims

The learning aim from this project is to conduct an analysis of Sports Capital Investment in Ireland from the period 2014 to 2020. I will analyse the data by region and sport, and my goal is to identify investment versus the reward of the specific sports. The data will be analysed, making observations such as investment by country and how many elite athletes/victories come from the specified counties and the sport's success in comparison to investment will also be investigated. Significant events such as the Olympics, European championships and world sporting events will be examined to see how many athletes from respective sports and the allocates won regarding the investment. The findings will be documented, and a similar analysis will be completed and compared to a country with a similar population size to Ireland. The level of capital investment in Scotland was used for a comparison of the performance of Ireland's capital investments.

The project's objectives include discovering data such as the sport with the highest investment, the success of a sport in comparison to investment and contrast in success in sporting events of Ireland and Scotland.

Through this analysis, all components of the project must be considered, along with alternative solutions. Technologies and techniques such as machine learning, data mining, the use of statistics to calculate data, visualisation techniques using R-Studio, Excel and SPSS will be used throughout the project to analyse multiple datasets using different techniques to explore and showcase the data.

1.3. Technology

During the project, there is a wide variety of technologies that will be used to conduct the analysis. The technology's that will be used include R-Studio, Excel, Tableau, SPSS, MS Power BI, GitHub and Canva.

R-Studio

R-Studio is used for multiple different aspects of the project. The first aspect of the project that required R-Studio is the pre-processing of the data. All the data cleansing, preparation and transformation is conducted in R-Studio. The clean data is exported in the format of an excel document.

R-Studio is utilised to create visualisations of the data. Visualisations such as Choropleth graphs is used to demonstrate the distribution of the investment in sports capital within Ireland by the mean investment of each county. Using the leaflet package in R-Studio, an interactive map is created showing each county and the average investment by county. Data frames are created from existing datasets so that new visualisations can be created with specific aspects of the data rather than the dataset as a whole.

Excel

Excel is used to import the data and for the initial observation of the dataset. The datasets are observed and studied in excel. Once the data was explored, and insight into the data was obtained, the data was then imported into R-Studio for pre-processing. Excel is also used to explore the data, statistics and functions are conducted in Excel, and some visualisation techniques are utilised. Excel is also used to create subsets of primary datasets so that in-depth analysis can be carried out on certain aspects of the data.

Tableau

Tableau is used to create visualisations from the datasets. Data was imported into Tableau, and using the interactive dashboard, and queries were run on the data and visualisations explaining the data were exported. The visualisations made in Tableau help give content to the results of the report.

SPSS

SPSS is used for its capability of performing high-level statistics and visualisation techniques. Once the clean data is imported to SPSS, the data was manipulated, and statistics and visualisations carried out. Descriptive statistics are produced using SPSS.

Microsoft Power BI

Microsoft Power BI is used to produce visualisations for the results section of the report. Power BI allows data to be loaded and the data fields to be selected for

investigation. Power BI then allows selection of visualisations that are best suited to the data.

GitHub

GitHub is used throughout to host the project. GitHub is used as a repository to ensure the project is backed up and secured in the event of a disaster and data being lost, or code broken, that there is a backup of the data in the GitHub repository. GitHub allows for the project to be documented throughout its course and tracks changes across the code as changes are made.

Canva

Canva is used to create the project presentation. Through Canva, the presentation slides were prepared and exported as a PDF document.

1.4. Structure

The following is an outline of the information provided in each section.

Section 2.0 - This section addressed in the document is the data section. In this section of the report, the datasets used to complete the objectives of the project are discussed, and an explanation is provided for the purpose of the datasets used. The dataset is broken down, and the critical elements of the dataset are explained.

Section 3.0 - The methodology used within the project is discussed, and the steps are explained on how it is implemented in the project. The six steps of the Knowledge Discovery in Databases methodology (KDD) and its six steps selection, pre-processing, data transformation, data mining, evaluation/interpretation, and knowledge discovery and use are all discussed thoroughly.

Section 4.0 - The analysis section of the report gives insight into how the data is explored and interpreted to complete the project's objectives. The techniques to complete the objectives are explained, and reasoning for specific approaches within the analysis are justified.

Section 5.0 - The results section of the report details the findings by analysing the data and are showcased in the form of statistics, visualisations, interactive graphs, and statistical tests.

Section 6.0 – Once the results observed from the data are discussed, a conclusion is drawn, detailing critical insights obtained and the significance of the analysis is discussed with real-life benefits considered.

Section 7.0 – Future work and further development are discussed, such as the potential for use of additional data and what could be implemented to further enhance the project.

Section 8.0 – All references used to complete the report are attached in this section in Harvard referencing format.

Section 9.0 – The appendix of the report includes content such as the project plan proposal, monthly reflective journal reports, and all other materials used for the completion of the report.

2.0 Data

Within the project, multiple datasets have been analysed. The first dataset analysed was obtained from data.gov.ie. The dataset was titled 'Sports Capital Programme 2014-2016'. The dataset is a free dataset for download on the data.gov.ie website. The functionality on the data.gov.ie website did not support the downloading of the data; therefore, contact was made with the Department of Tourism, Culture, Arts, Gaeltacht, Sport and Media to access the dataset. This department representative responded shortly after the request and the requested data was provided alongside further data. The dataset acquired ranged from 2014-2020.

Once received, the dataset was cleaned, and null values removed from the data. The dataset was then broken down into sub data sets with R-Studio. The data was then further broken down and split by county and divided into 26 separate spreadsheets to analyse each county. The data was split by the grant scheme and the funds that were allocated by sport type. By doing this, each sport can be studied, and each scheme can be analysed in full, and trends can be identified.

The second dataset that was used for the project was obtained from the Central Statistics Office. The data contained information on the demographics of each town in Ireland. This data was split into separate towns by a column labelled GUID. Another dataset labelled “OSi National Statistical Boundaries" was also downloaded. The OSI dataset was a descriptor dataset and provided information such as town, county, and province. The common variable in both datasets was the GUID column. The two datasets were combined using R-Studio by the standard column GUID so that a new dataset was created that had all the information that can be found in both two datasets. The dataset was prepared so that all unrelated columns were removed from the dataset, and only the columns required were kept in the dataset. To pre-process the data, a glossary for the column names was required to understand each column in the dataset. The unnecessary variables were then removed from the dataset. Once pre-processed, the data could then be analysed by each town and broken into further subsets of the data by county. The remaining data after the pre-processing of the data included This data included the total population of each county and the community ages 1 through 24 and observing the sum of people in this age category. To better understand the

data, visualisation techniques were used. Some of the visualisations used were histograms, boxplots, and pie charts.

Data frames were created from existing datasets loaded into R-Studio to analyse specific aspects of the data and visualise subsets of the data.

Datasets were created from research techniques, and the datasets contained accolades by sport. This was conducted on both a national and international level. Once created, the weight of a sport's success was calculated and investment versus success was analysed.

Datasets for Scotland were obtained to compare Ireland and Scotland to weigh the success of each sport and compare the level of investment each country allocates.

3.0 Methodology

To analyse the data, it was first prepared so that the raw data for the analysis and modelling stage of the project were correct. The data was prepared in such a way that the data is meaningful and easily interpreted. There were several different ways in which the data could be prepared, such as the CRISP-DM, but for this project, the methodology selected is the KDD methodology. The KDD methodology is known as the Knowledge Discovery in Databases and can be defined as the process of finding, transforming, and refining meaningful data and patterns from raw data to be used for applications or in specific domains. The steps followed to prepare the data for analysis are:

3.1 Selection

After proposing this research topic and having the concept of the project approved by the college, the first stage of the KDD methodology included selecting the datasets used to research the topic chosen. The goals and the objectives of the project were set, so a primary dataset then needed to be selected to complete the project's goals. To obtain data on sports capital investment sites such as the CSO and Data.gov.ie were visited to identify relevant datasets that would allow the project's objectives to be met.

3.2 Processing/Data Cleansing

The data was pre-processed before use to ensure the data is clean and that noisy data is removed. This includes the removal of all missing value, changing data types and changing column names. The data was first imported to Microsoft excel for an initial analysis to investigate data types, columns, rows etc.

The data set was then imported into R-Studio. The dataset contained multiple different variables, some of which contained high multi-collinearity. Multi-collinearity was tested by conducting a Pearson correlation. Variables that contained high multi-collinearity and deemed unnecessary variables, such as "Tax I.D.", were removed from the dataset using R-Studio. Variable names were lengthy, and some were difficult to understand. The variable names were changed to simpler variable names so that they were easier to understand. For

example, in the primary data set Payments2014-2020 column one was named “Address 1: County (Grant Application Id)". This variable was renamed "CountyID".

After removing unnecessary variables from the data and renaming the variables within the data, the data was checked for missing values within the data. Using the complete cases function within R-Studio, we can see where the missing data values are in the data set. The summary function was then used to figure out the number of null values within the dataset. Once the number of null values in the data set and where the missing values are within the data are discovered, these values were removed by using the omit function within R. Once the N/A values are omitted from the dataset, the data is exported as a new dataset as a clean version. This process is done so that the quality of the data is improved, and the reliability and effectiveness of the data are enhanced.

3.3 Transformation

The transformation of the data involves the process of transforming the data into its appropriate form. The data needed to be classed under its correct data type to be analysed. The process of changing the datatype was completed in R-Studio, and columns were assigned the correct datatype. Columns needed to be changed to string, numerical and character formats.

Two datasets were provided by data.gov.ie, with the second dataset giving additional insight into the investment and the purpose of the grant allocated. The two data files were merged using R-Studio to utilise this additional insight for the report. The datasets were merged in R-Studio by combing the data by a standard row within the two datasets. The row used to combine the two datasets was the "File Reference Number (Grant Application Id)" variable, and by matching these, the two datasets were combined. This process of combining datasets was also carried out for a secondary dataset downloaded from the CSO detailing census data from 2016. The datasets were combined by the standard column within the data, the variable used to combine the data was labelled "GUID", and this gave meaning to the figures in the data.

3.4 Data Mining

Due to the nature of the dataset and the subsets of the data being so small possible data mining methods were discussed with lecturers from the Data Mining module about potential time series forecasting methods. Methods such as the Facebook Prophet model, an Auto-Regressive Integrated Moving Average (ARIMA) method, Holt-Winters method and other methods were discussed as possible methods for forecasting the data.

3.5 Interpretation/Evaluation

Once all previous of the preceding steps of the KDD methodology has been completed, the data must then be evaluated. To help interpret the data and understand the transformed data, exploratory data analysis was conducted on the data. This approach allows for the critical components of the dataset to be summarised and identify the main characteristics of

the dataset. Exploratory Data Analysis (EDA) was used to become familiarised with the data, and the trends and patterns in the data are displayed in the form of visualisations. The visualisations include graphs such as histograms, bar charts, pie charts and many other visualisation techniques. Through this, outliers in the data are identified, underlying assumptions are tested, and the insight into the dataset is maximised. The EDA methodology is not a technique but a philosophy on how data analysis should be conducted.

3.6 Knowledge Discovery and Use

From all the preceding steps, the data was reported on, and results of testing and visualisations are discussed thoroughly in section 5.0 of the report. Knowledge was extracted from the data through visualisations and testing, and this allowed for all outputs from the data to be thoroughly discussed.

4.0 Analysis

Once the data had been pre-processed, the data was exported as a .csv file so that the clean data set could be investigated. The clean data was first opened in Microsoft Excel, and further exploratory data analysis was conducted on the data so that the main characteristics of the data could be identified and further analysed before testing was conducted on the data. The exploratory data analysis techniques that were carried out on the data were:

1. Visualisations – Visualisations of the data will be created to better understand the data and the variables involved. The visualisations will give insight into how the data is distributed and how variables fit with other variables. Visualisations will include
2. Descriptive Statistics - Descriptive Statistics were produced for the data so that the data could be easily interpreted. The descriptive statistics outputted allowed me to gain insight into data such as mean values, ranges, and standard deviations. Summary statistics were used to gain a deeper understanding of the relationship between two variables.

The data was broken down and separated into 26 separate spreadsheets so that each county could be analysed individually. I reviewed the analysis of the data in R-Studio and Excel and decided that this aspect of the task was more suited to Microsoft Excel as it allowed me to analyse and manage the data more efficiently and effectively.

Both Microsoft Excel and SPSS were used to create graphs for the data on sports capital payments in Ireland. The dataset was imported to SPSS. When the chart was imported to SPSS, the graph building function was then utilised. The data was visualised to make the data easier to comprehend and made the data easier to identify trends, patterns, and outliers within the data. From visualising the data, outliers were identified within the data, and the data was easier to understand, allowing these findings to be further analysed.

Statistical analysis was used to calculate the mean investment to each county. An interactive graph was created in R-Studio using the spatial data package in R-Studio and the Leaflet package to present this data. When the mean value of the county was obtained, the longitude

and latitude of the county was added to the dataset and used to map the data to the correct county with tiles showing the County name and the mean investment and the total investment to the county.

Investment in Each County - Age

The level of investment received varies throughout the twenty-six counties. Using both the primary dataset containing investment levels for each county and the dataset obtained from the Central Statistics Office including demographic information for all counties.

Data of the county's population and the population under the age of twenty-four was extracted from the data set and put into a subset of the data. The population under the age of twenty-four was taken as it is under this category in which the highest percentage of the population play sports. This data was then analysed, and using data from the primary dataset, the total investment level per county, investment per citizen in each county and investment per individual under the age of 24 years old per county, could be calculated. Conducting this investigation allowed the identification of the total investments and how much investment is per individual for sporting activities within a given county. The calculations for this output were conducted in Microsoft Excel, and the results were presented in the form of visualisations outputted from Tableau. The results are discussed in full in section 5.0 of the report.

To test for correlation within the data, a Pearson Correlation was conducted on the data. This method for testing if there is a correlation within the data was chosen as the data being tested was count data. This method was determined to be best suited based on the data that was being tested. The Pearson correlation tested to see if there was any correlation between accolades won and the sum of the investment from grants to a county. For the test, a null and alternate hypothesis was stated before testing. The null and alternate hypothesis states

Null Hypothesis (H0): The level of investment has an impact on the number of accolades won

Alternate Hypothesis (H1): The level of investment does not have an impact on the number of accolades won

Significance level (α) 0.05

The test was conducted in SPSS, and the results of the correlation test are discussed in full in section 5.0 with the acceptance of the null/alternate hypothesis stated.

A basic time-series analysis was conducted on the data to predict future values within the data and see what future investment may look like. The time series analysis will showcase how a variable can change over time. In this case, the time series was used to forecast the future sum of grants in Ireland.

5.0 Results

5.1 Total sum of Sports Grants Investment by County

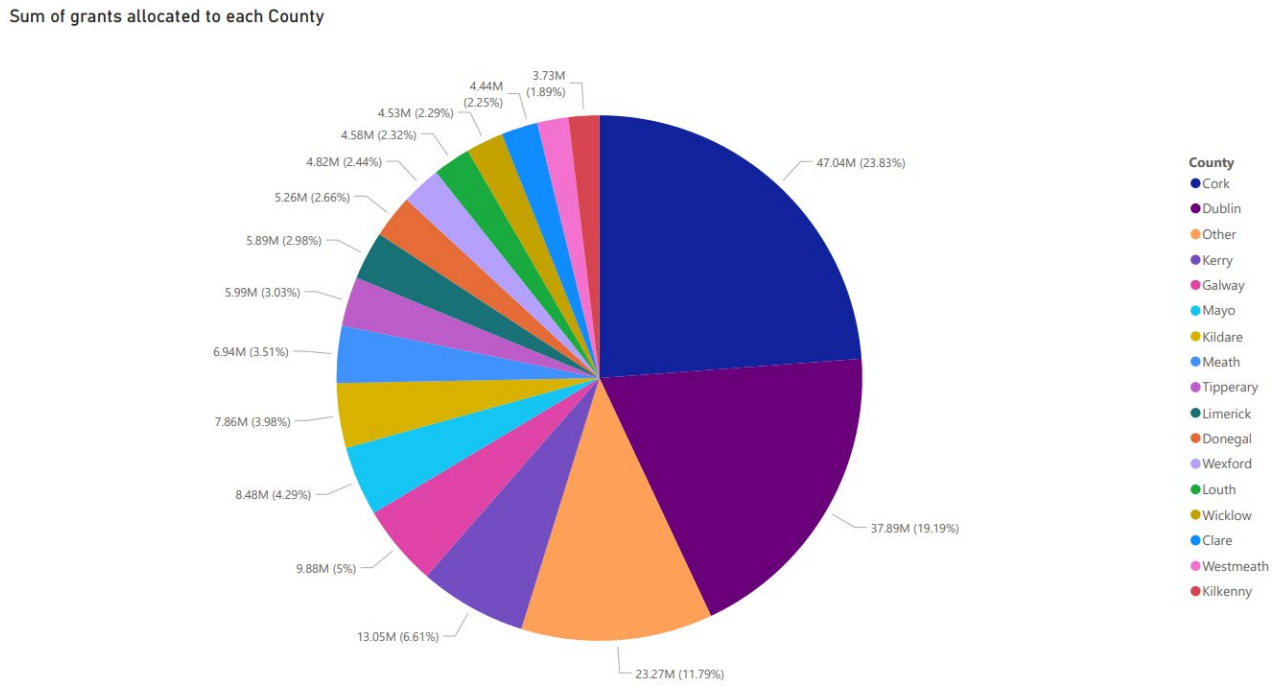


Figure 5.1 – Microsoft Power BI Grants allocated to each county

In the above pie chart, figure 5.1, the total sum of capital received by each county in the form of sports grants is shown. The graph gives a visual representation of how sports capital investments have been divided in each county from the 8th of December 2014 to the 31st of August 2020. The total amount of sports capital invested was €197,422,108 for this period. The key is displayed on the right of the pie chart, with counties represented by a specific colour on the chart.

Looking at figure 5.1, we can see that Cork and Dublin account for a substantial portion of the graph. Both Cork and Dublin account for 43% of all capital investment. The three counties with the highest level of sports capital investment equate to 49.61% of all investment, meaning the remaining 50.39% is distributed between the other twenty-three counties in Ireland. This leaves a sum of €99,481,000.2 between the remaining twenty-three counties. Cork received a total sum of €47,039,235.31 which is 23.83% of the total level of investment.

The ten counties with the least amount of capital investment to sports equate to a total of 11.79% or €23,276,066.5 of the total sum of €197,422,108. These counties are represented within the 'Other' section of figure 5.1 due to the values being so small.

The distribution of the €23,276,066.5 allocated between the ten counties can be seen in figure 5.2 attached below.

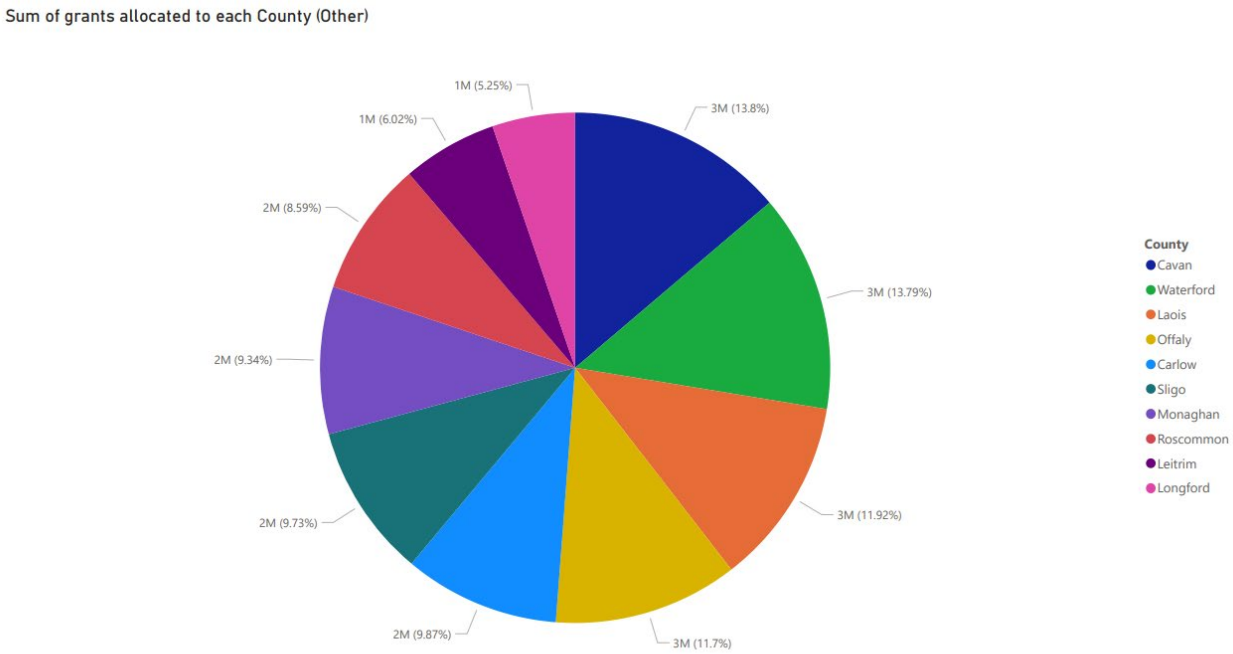


Figure 5.2 Microsoft Power BI Grants allocated to each county (other)

In the above table we can see the counties that were classed underneath the 'Other' heading. As the ten counties accounted for 11.79% or 23,276,066.5, such a small percentage of the total sum it was difficult to interpret the data. The data is shown above and shows how each county split the 11.79% from figure 5.1.

The county that received the lowest capital sum of grants is County Longford. In total, Longford received a sum of 1,220,535 or 0.62% of the total sum. In comparison to Cork's fee of €47,039,235, County Longford received 2.59% of the fee Cork received.

5.2 Average sum of Sports Grants Investment by County

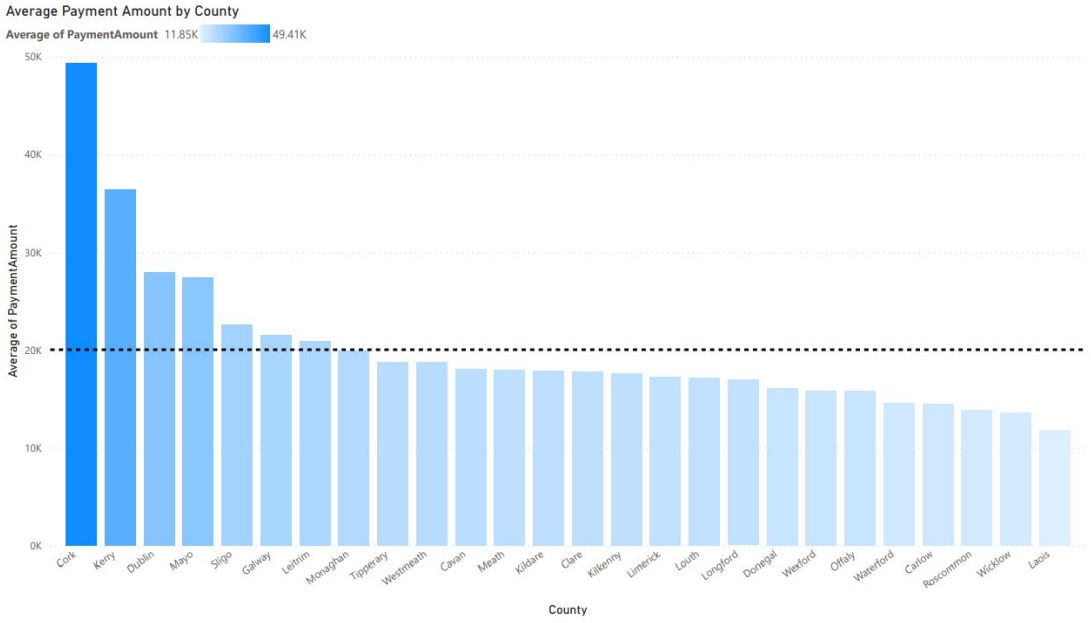


Figure 5.3 – Microsoft Power BI Avg. sum Sports Grants Investment by Country

Figure 5.3 showcases the mean investment throughout the twenty-six counties of Ireland. The mean of the twenty-six counties is shown in the dotted black line for reference when compared to the other counties. The graph shows the mean value of each county in descending order and the colour scheme follows a colour scale ranging from dark to light blue as values decrease in the data.

The bar plot contains data ranging from 2014-2020 with data from all of Ireland's counties and the sports capital investment received. The data was extracted from the primary data excel file containing all the sports capital investment in Ireland and separated in excel into individual excel files for each county. The mean investment was then calculated and input into a new excel file with all the mean investments.

The data shows that Cork has the highest mean investment in Ireland with a mean of €49,410.96. The mean is substantially higher than the second-highest mean investment of €39,459.95 in Kerry. The range of the mean sports capital investment was €37,559.83. Cork has the highest mean investment, and Laois with the lowest level of investment with a mean investment of €11,851.13

The mean investment across the whole twenty-six counties of Ireland was €20,025.60. The mean line for investment can be seen by the black dotted line in figure 5.2. Of the twenty-six counties of Ireland only eight counties are above the mean investment, sixteen counties of Ireland fall below the mean. This means 61% of counties receive less than the national average for sports capital investment.

When the data is analysed, we see Cork's level of investment is 2.46 times greater than the mean value in Ireland. Given the variance between Cork and the man value in Ireland, it must be investigated to identify why the level of investment is so much higher and explore the data to find potential outliers. To identify potential outliers within the data, the Cork data set was imported to SPSS.

In observing the data, it was discovered that Cork had a total of nine hundred and fifty-three investments through the period of 2014-2020. The sum of the investments in Cork was €47,039,235. The most significant investment was allocated on the 19th of January 2017 as part of the 2016 Sports Capital investment scheme. This investment totalled €10,000,00. The range of investment within this period was €9,999,870.

Cork's six highest sports capital investments were allocated to Cumann Luath Chleas Gael to redevelop their sports stadium. The six investments totalled €30,000,000. Páirc Uí Chaoimh's funding for the redevelopment of the stadium spanned over three years and equated for a total of 63.8% of Cork's sports capital investment.

Without the investment in the redevelopment of Páirc Uí Chaoimh and the investment causing outliers within the investment data for Cork, the mean investment would total €18,011.88.
The graph below showcases how Cork would compare to other counties without the high level of investment in Páirc Uí Chaoimh from 2016 to 2019.

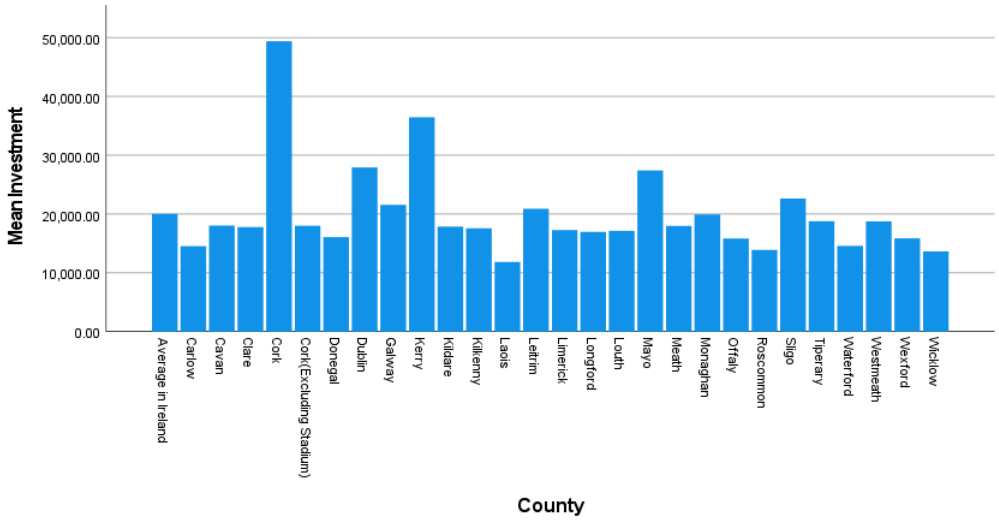


Figure 5.4 – SPSS Output Mean Investment by Country

Should Cork not have received the sports capital grants to Páirc Uí Chaoimh and received the other grants, it would have the 13th highest mean sports capital investment rather than having the highest level of investment in Ireland, as shown in figure 5.4 above. The average in Ireland

is inserted to the histogram as a reference point, the same is done for Cork with and without the capital invested for the development to Páirc Uí Chaoimh.

5.3 Number of Grants received by sport

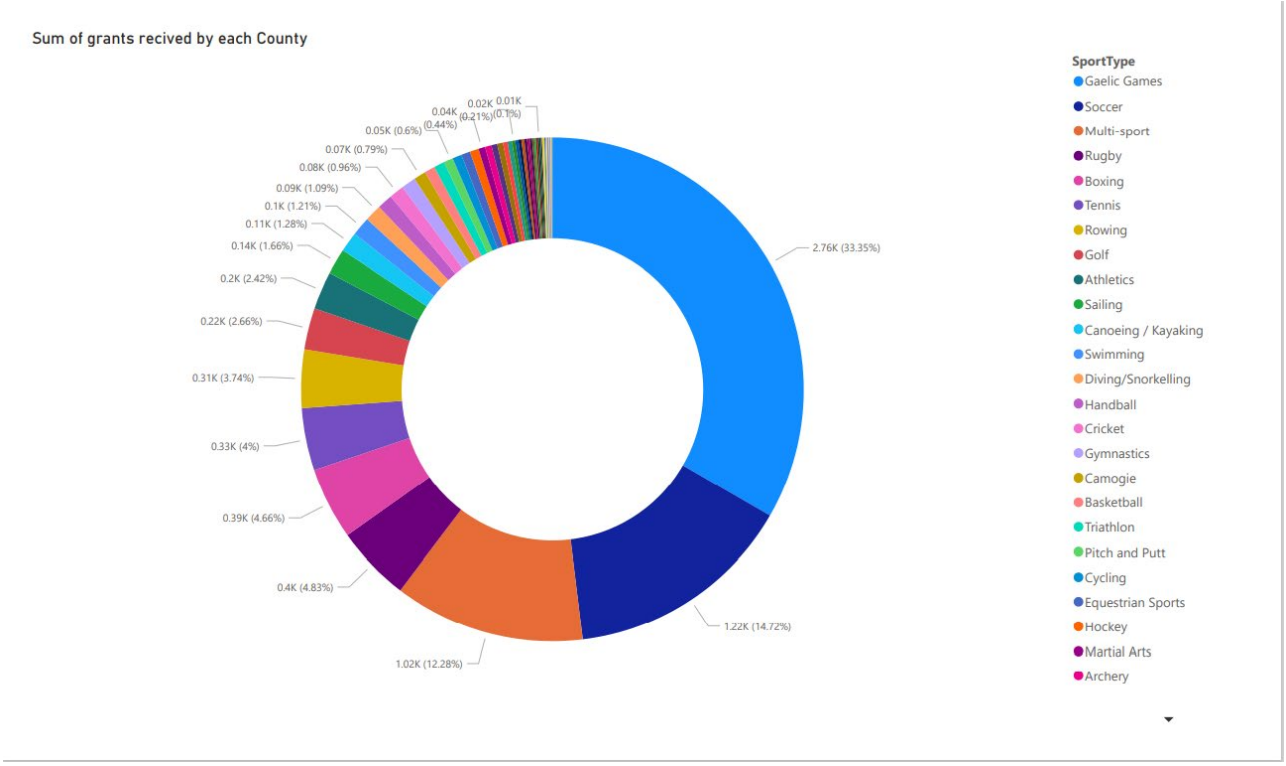


Figure 5.5 – Microsoft Power BI Number Grants received by Sport

The above doughnut chart showcases how many grants each sport received from the period 2014-2020. From observing the doughnut chart, we can see that the Gaelic Games received the highest level of grant investment during this period, receiving 2,757 grants. With a total of 8,267 grants in total, this means Gaelic Games accounts for over one-quarter of the total grants allocated. Gaelic Games accounted for 33.3% of the total grants during this time. The sport receiving the second-highest amount of grants was soccer, with a total of 1,217. This means soccer received 44% of the total amount of grants that Gaelic Games received in this period.

The top three funded sports in Ireland during this period were Gaelic Games, soccer, and multi-sport activities. Between the three sports, the total amount of sports capital grants allocated was 4,989. This equates to 60.3% of sports capital grants in Ireland, meaning the remaining 57 sports received 39.7% of all grants.

It was found that the sum of the total grant payments was €197,422,108.50. Gaelic Games received a total of €85,643,151.40, equating to 43.3% of the grant total. Gaelic Games received 33.3% of all grants throughout this period but received 43.3% of the funds for sports capital. The mean grant amount in Gaelic Games was €31,063.89.

5.4 Sum of payments to sports with the highest sum of grants received

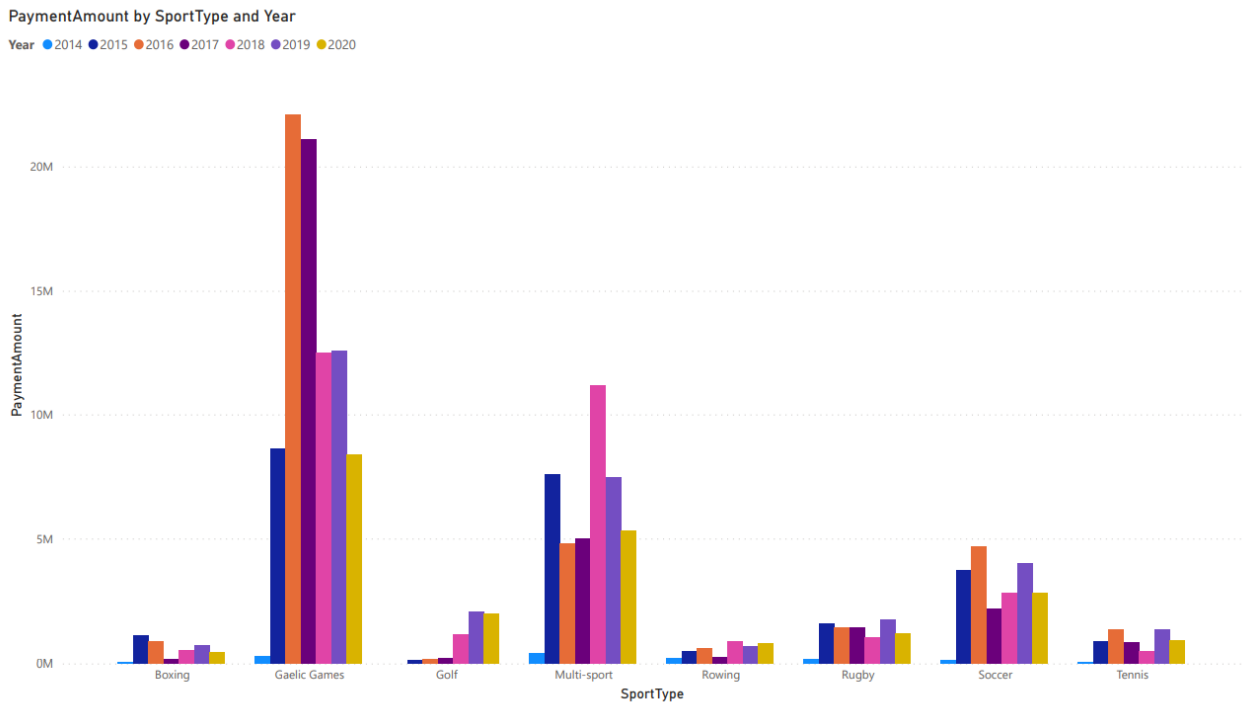


Figure 5.6 – Microsoft Power BI Payment Amount by Sports Type and Year

In figure 5.6, we can see the eight sports with the highest level of sports capital investment in a stacked column chart. Each sport is divided into the capital received each year. As we found through previous results, Gaelic Games has the highest sum of capital for investments and the highest quantity of investments. From observing the graph, we can see the increase and decrease year on year within sports and how capital investment fluctuates in each sport.

In the years 2015 through 2020, we can see that Gaelic Games receives the highest sums for each of these years. For all years, Gaelic Games received the highest level of sports capital, apart from 2014. In 2014 Multi-Sports received the highest level of investment, but this figure may not be accurate due to only 23 days of the year accounted for within the data. For the twenty-three days available in the data, Multi-Sport received the highest amount with €424,328 invested.

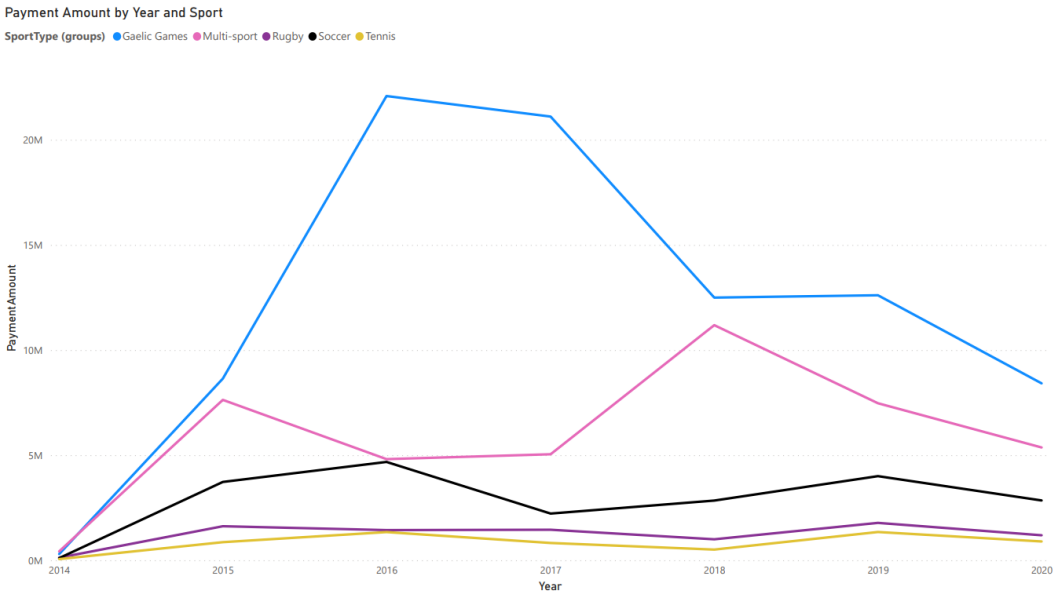


Figure 5.7 – Microsoft Power BI Payment Amount by Year and Sport

In figure 5.7 above, we can see the fluctuations in capital investment for the five sports with the highest investment levels in Ireland. The line chart helps to give a better visual representation of the increase and decrease in investment levels. In 2016 the capital investment for Gaelic Games spiked to €22,081,059 and experienced a gradual decrease in the following years. From 2016 to 2020, Gaelic Games experiences a decrease in the investment of 262%.

For Multi-Sport, we see a continuous fluctuation in the level of capital investment. The level of investment peaks in 2018 with a sum of €11,184,458. The highest increase in the data for Multi-Sport is seen from 2016 to 2017 with a 221.7% increase.

In soccer, the year that received the highest level of investment was 2016. For 2016 soccer received a total sum of €4,679,572. From 2014 to 2016, soccer experiences a large increase in investment. In 2017 the investment levels declined and in 2018 gradually begin to increase again.

Rugby and Tennis both experience similar trends in the fluctuation of levels of investment, with both sports experiencing their highest levels of investment in 2019. In 2019 Rugby received €1,780,994 and tennis receiving €1,346,860. Both sports follow the same trend with declines in 2017 and 2018, peaking in 2019 and a slight decline in 2020.

From this observation, the trends within the sports with the highest level of capital were identified and are used in further stages of the results when analysing sports performance.

5.5 Investment in Each County - Age

For this analysis, the first thing that was identified as the total number of individuals in each county that was under the age of twenty-four years old. Once the data was separated into each county, and the total sum was acquired, a new data set was created with the sum data. With this data Figure, 5.8 was created using Tableau to showcase the population within each county that is under the age of twenty-four years old. The graph is displayed in descending by sum. From the graph, we can see that Dublin has the highest population of this age demographic, with 430,684 people being under this age category; this equates to 32% of Dublin's entire population. Cork has the second-highest population in Ireland from this demographic with a total population of 180,877, equating to 33% of the population of Cork. In the graph, we see that Leitrim have the lowest quantity of this age demographic with a total of 10,097 individuals falling under this category, but this figure can be accounted for by Leitrim having the lowest population in Ireland with 32,044 habitats in total in county Leitrim.

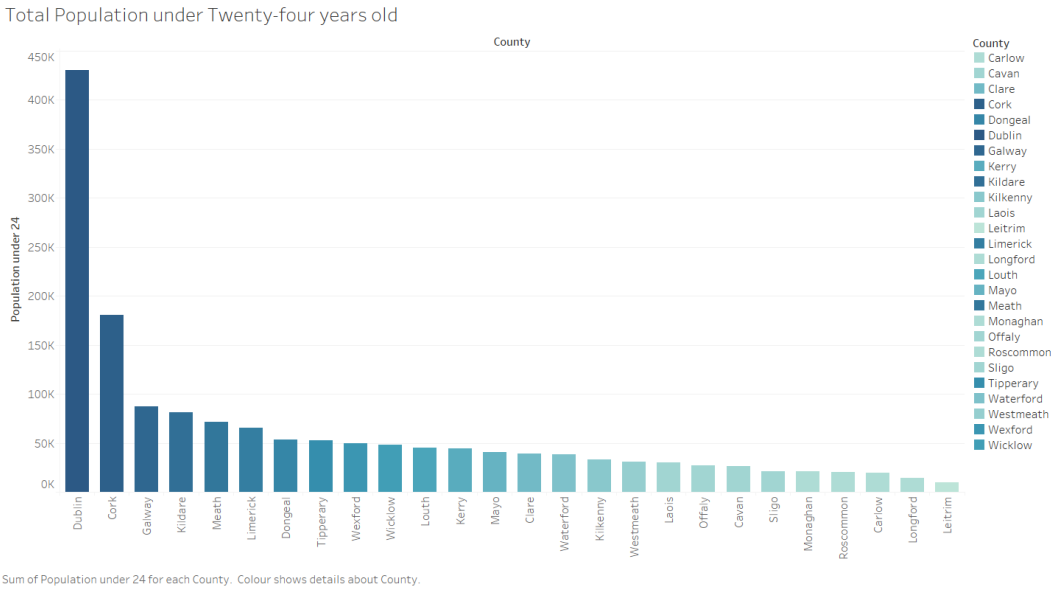


Figure 5.8 – Tableau population under twenty-four years old by County

The data collected to create figure 5.8 with the sum of habitants per county under the age of twenty-four was taken to conduct an analysis on investment each county received from the period of 2014-2020 based on the number of individuals under this age bracket. This was done by taking the total investment of the county, which was obtained through previous analysis and dividing by the sum of habitants in each county from this bracket. To obtain this figure, calculations were conducted using Microsoft Excel, and all figures were then analysed. To better understand this data, it was then imported to Tableau to be visualised and the data could be easier to interpret. The data is presented in the form of a histogram and is formatted in descending order. This data can be seen in figure 5.9 attached below.

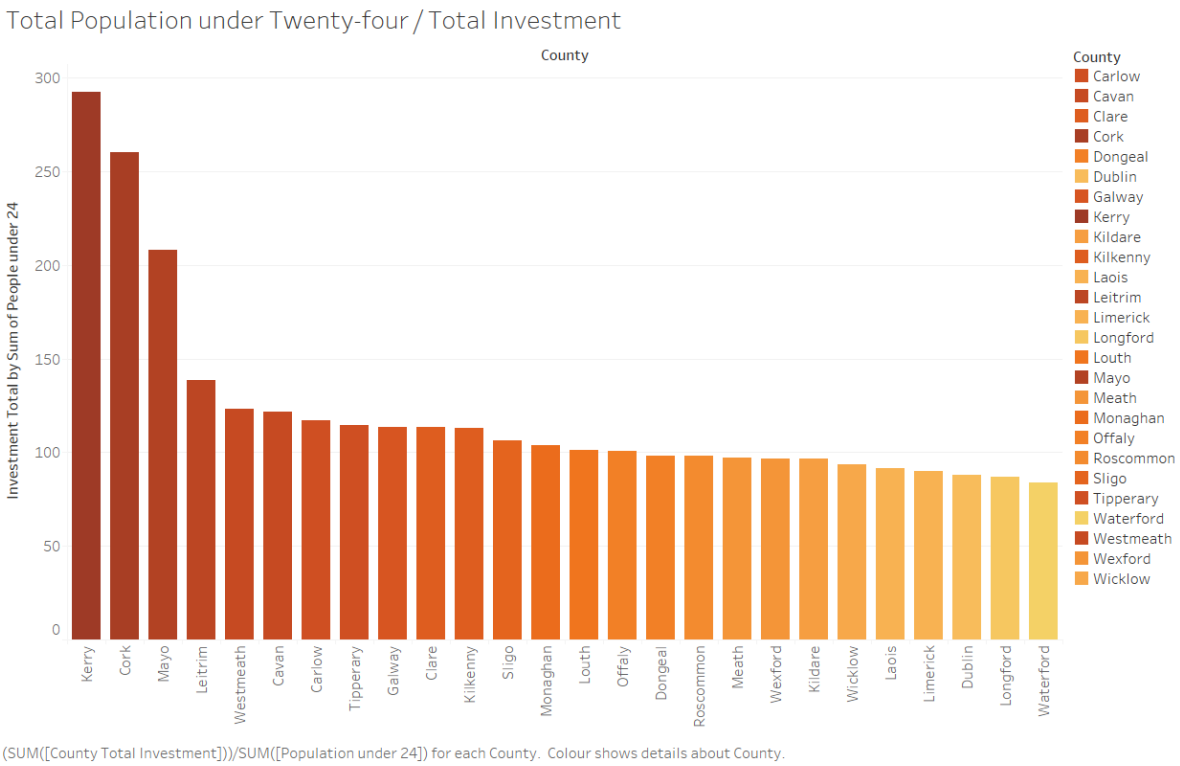


Figure 5.9 – Tableau Population under twenty-four by country and Investment amount

From observing this data, I discovered that of the twenty-six counties of Ireland, Kerry receives the highest sum of investment for each person under the age of twenty-four years old that live in the county. In total, 44,612 individuals reside in Kerry under this age bracket. The mean investment per person under this bracket is €292.58. This is 12.5% greater than the next county.

Cork received the second-highest amount per person under the age of twenty-four. In total, there are 180,877 people under the age of 24 in Cork, of the €47,039,251.31 invested in Cork since 2014, an average of €260.1 is invested per person. This figure is substantially higher than the third-highest investment average per person in Mayo with €208.3. Corks average investment per person is 25% greater than the investment in Mayo. As this figure is an outlier in comparison to the rest of the data, we must see if there are any extreme figures within the data that may cause this significant value. As Cork received an investment of 30,000,00 for the redevelopment of Páirc Uí Chaoimh for Gaelic Games, the figure for Cork can be misconceived. Without this investment, the average investment per person in Cork would total €94.20 per person which would be the lowest figure in Ireland.

We see that Dublin receives the third-lowest investment per person, with a total of €88 being invested per individual under the age of twenty-four. This figure is significant as Dublin has the greatest sum of people within this age category, with 430,684 people falling under this bracket. Dublin receives the second-largest amount in total investments from 2014, receiving €37,877,831. As Dublin receives the second-largest sum of money from investments in this time period, this would be considered a high investment in comparison to other counties.

When broken down further, it is found that Dublin has the third-lowest investment per individual in Ireland. It is only when the population is considered that we see that the level of investment is minor in comparison to other counties.

The county with the lowest level of investment in Ireland per person under the age of twenty-four is Waterford. The total investment for sports in Waterford since 2014 is €3,209,107; this is ranked 18th in terms of investment level in Ireland. When taking the population of Waterford under the age of twenty-four into account, the sum of people in this age category is 38,345. Waterford is ranked 15th

5.6 Accolades won versus investment – GAA Male

To measure the success of the investment for GAA in each county, the accolades for both All-Ireland competitions and Province Championships. The accolades include all competitions won for both Gaelic Football and Hurling. The number of finals reached by each county will also be considered as this also shows how a team performed through the season. The bullet graph labelled Figure 5.10 is attached below and is formatted in descending order from finals of competitions reached high to low. The solid pink bar represents the number of finals a county has reached, and the dotted black line representing the number of trophies won.

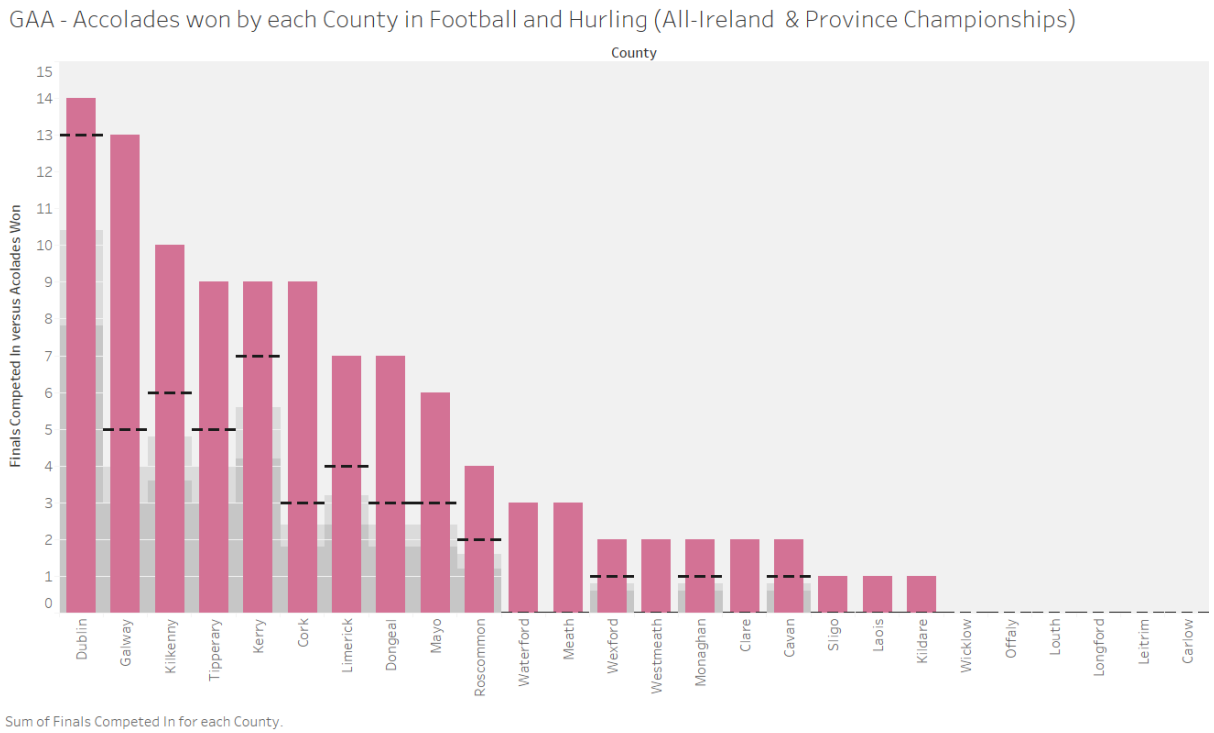


Figure 5.10 – Tableau Number of wins by country in football and hurling

In figure 5.10, we can see from observing the graph that Dublin has experienced the greatest level of success between the period of 2014 to 2020. The graph tells us that Dublin has reached a total of fourteen finals in this time and have been successful in the final thirteen of their fourteen appearances. This means Dublin have a success rate in finals of all competitions of 93%. This is the highest success rate of all counties in Ireland. For Gaelic Football, Dublin

reached a total of thirteen finals, seven province championships and six All-Ireland football championships and found success in all thirteen finals. Through this triumph experienced in Gaelic football, it is this reason that Dublin is the most successful county in GAA.

Galway have the second-highest final appearances of all counties, reaching thirteen finals across Gaelic football and hurling in all competitions. In thirteen finals, Galway found success a total of five times, this means that Galway had a success rate of 35.7% in finals of competitions. Kerry has the second-highest success rate across all counties, with a success rate of 77.7% in finals of competitions. In total, Kerry reached nine finals across Gaelic football and Hurling and was successful in seven of the finals they participated in.

In hurling, the team that has been the most successful in terms of achievements is Kilkenny. Since 2014, Kilkenny has competed in ten hurling finals and was victorious a total of six times. Kilkenny have a 60%-win rate in all finals played. This is the highest number of finals competed in, and the highest number of trophies won by any county in Hurling. The second-highest number of accolades won by a county is held by Limerick and Tipperary. Both counties have won four accolades each, and both competing in seven finals. This means both Limerick and Tipperary have a success rate of 57.1% in finals.

After analysing the success of each county in Gaelic football and Hurling, we must then assess the level of investment allocated to each county and see if there is a correlation between investment levels and accolades won. Since 2014 a total of €85,643,151 has been invested between the twenty-six counties of Ireland. In figure 5.11 attached below, we can see how this money has been distributed across all counties.

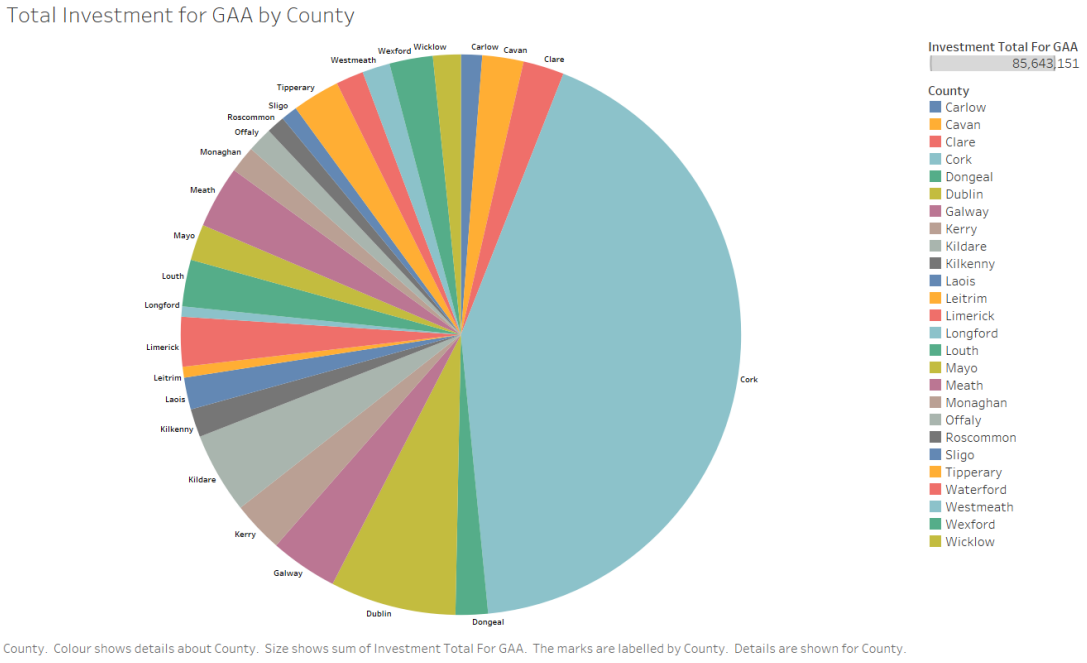


Figure 5.11 – Tableau Total Investment for GAA County

From the €85,643,151 invested into GAA since 2014, we can see from the above pie chart in Figure 5.11 that Cork has received the highest amount of the money invested. For GAA, Cork received a total of €36,412,413. This equates to 43% of the total investment, which is significantly higher than any other county. The next highest investment to a county is Dublin receiving 7.2% of the total investment for GAA. This figure means that Cork received 5.85 times more than the next highest county.

The county that received the lowest level of investment in County Longford, receiving a sum of €495,700. This grant accounts for 0.0057% of all grant allocations for Gaelic Games. The three counties with the lowest level of grants are Longford, Leitrim, and Sligo. From the combined sum of the three counties, this accounts for 2.14% of all grants, whereas the three counties with the most investment in Cork, Dublin, and Kildare account for 54.5% of all grants for GAA.

5.7 Correlation between levels of investment and accolades won

To see if there is any correlation between levels of investment and accolades won by a county, testing was conducted on the data. Three variables were analysed in this correlation test to measure investment success. The three variables tested were the total investment for GAA in a county, the number of finals a county has competed in, and the total accolades a county has won. The test conducted on the data was a two-tailed Pearson Correlation test. A Pearson correlation is a test statistic that measures the relationship or association between variables. This test was chosen as it was the most suitable for this specific data instead of alternative methods. The outputs of the Pearson Correlation are shown below in figure 5.12.

We must first state our null and alternative hypothesis.

Null Hypothesis (H0): The level of investment has an impact on the number of accolades won

Alternate Hypothesis (H1): The level of investment does not have an impact on the number of accolades won

Significance level (α) 0.05

Correlations

		Investment Total For GAA	Finals Competed In	Total Accolades Won - Football & Hurling
Investment Total For GAA	Pearson Correlation	1	.331	.176
	Sig. (2-tailed)		.098	.390
	N	26	26	26
Finals Competed In	Pearson Correlation	.331	1	.903**
	Sig. (2-tailed)	.098		.000
	N	26	26	26
Total Accolades Won - Football & Hurling	Pearson Correlation	.176	.903**	1
	Sig. (2-tailed)	.390	.000	
	N	26	26	26

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 5.12 – SPSS- Pearson Correlation

The results showcase the correlation value for finals competed in, and total accolades won, and the significance level of the correlation.

5.7.1 Finals Competed In

From examining the output of the two-tailed Pearson Correlation test, we can see for finals competed in there is a correlation value of .331. Based on this result, we can determine that there is a moderate/low correlation between investment total and finals competed in. The level of significance returned for the data is .098. As this value is more significant than our Alpha value of 0.05, this means that there is not a significant relationship between finals competed in and levels of investment.

5.7.2 Total Accolades Won – Football and Hurling

When observing the output of the two-tailed Pearson Correlation test, the correlation value between total accolades won by a county, and the investment received by the county is .176. This is a positive correlation, but this value signifies that there is a very low correlation between the data. The significance value for the data is .390. This value is above the alpha value of 0.05, which means the test is not significant. There is no relationship between the level of investment and the number of accolades won by a given county.

As the significance value is more significant than our level of significance of 0.05, we have strong evidence to reject our null hypothesis that the level of investment has an impact on the number of accolades won. This means we have strong evidence to accept our alternate hypothesis that the level of investment does not have an impact on the number of accolades won.

5.8 Comparison in capital investment versus success – Rugby versus Soccer

After completing the analysis for Gaelic Games on a national level and the performance of each county in terms of the level of investment received the next analysis to be conducted was on sport at an international level measuring their performance and analysing their capital investment. The two sports chosen for the analysis were Rugby and Soccer. Rugby and Soccer were chosen as for capital investment they rank third and fourth in Ireland for highest sum received. In the below table, figure 5.13 we can see the yearly investments each sport received and the sum of investments over the total period. There is a total difference of €11,815,873 in investment levels of the two sports.

	Sport	
Year	Soccer	Rugby
2014	104,735	126,489
2015	3,727,759	1,621,230
2016	4,679,572	1,437,751
2017	2,223,583	1,454,757
2018	2,842,065	1,000,411
2019	4,003,911	1,780,944
2020	2,847,392	1,191,562
Total Sum	20,429,017	8,613,144

Figure 5.13 – Excel Spreadsheet – Soccer and Rugby Investment by year

To analyse the performance of the investment factors such as major competitions competed in, major accolades won and win percentages are taken into consideration.

All matches played by the Football Association of Ireland and the Irish Rugby Football Union from 2014 to 2020 have been extracted and put into a clustered bar chart to visualise the outcome of all matches competed in. This data can be seen in figure 5.13 attached below.

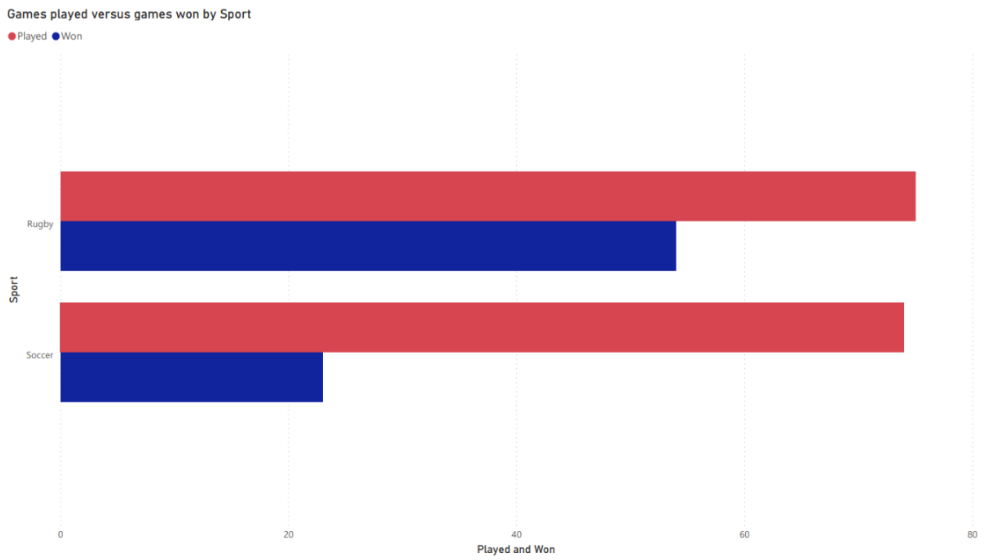


Figure 5.14 – Microsoft Power BI – Rugby & Soccer, Matches played versus matches won

In the graph, Figure 5.14 above we can see that both the IRFU and the FAI have played a similar number of matches from 2014 to 2020. The IRFU competed in seventy-five matches whereas the FAI competed in a total of seventy-four matches. Of the seventy-five matches competed in by the IRFU a total of fifty-four games were won. The Irish national rugby team won 72% of their total matches. For the FAI, the Irish national soccer team competed in seventy-four matches, winning a total of 31% of their matches.

From the data above we can see that the Irish Rugby Football Union has been significantly more successful in sporting events than the FAI. After analysing the matches played and percentage of matches won the qualification for major tournaments is observed and how far the team progressed in the tournament. This data can be seen below in figure 5.14.

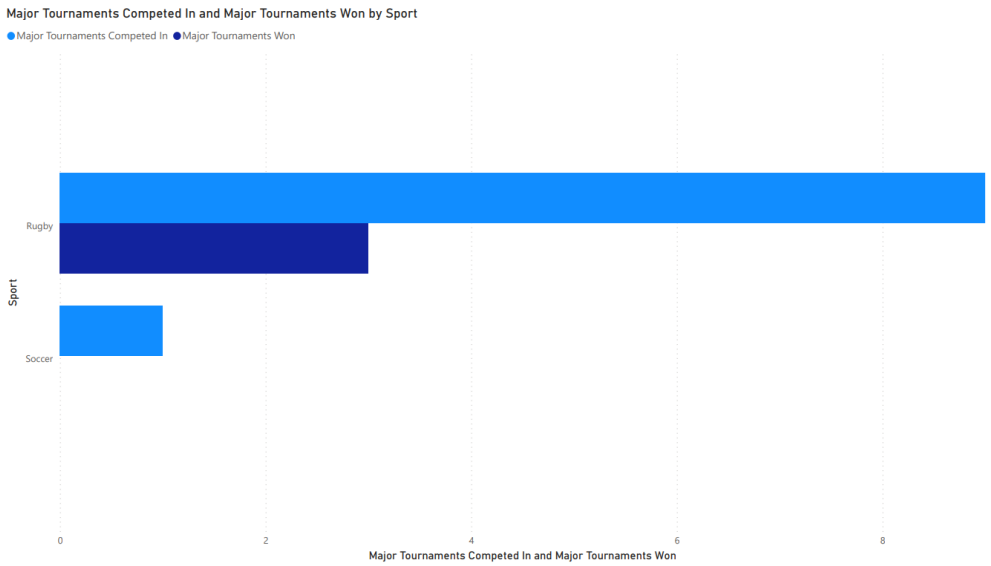


Figure 5.15 – Microsoft Power BI - Rugby & Soccer, Major Tournaments

The Irish Rugby team have qualified for a total of nine major tournaments including the World Cup and the Six Nations. For the Six Nations Ireland receive automatic entry to this tournament. Other than the Six Nations, Ireland has qualified for two major tournaments, both being the World Cup. Of the nine major tournaments competed in the IRFU have won three major titles with success in the 2014,2015 and 2018 Six-Nation's trophies. Ireland Rugby also qualified for two world cups and made it to the quarter finals in both the 2015 and 2019 World Cup. The Irish National rugby team have won 33.3% of all major tournaments competed in since 2014.

In comparison the FAI have only qualified for one major tournament since 2014 of a possible four. The FAI qualified for the European Championships in 2016 held in France and failed to make it out of the group stage of the tournament. From the data above, we can see that rugby in Ireland has a significantly higher win percentage with a 72%-win ratio compared to 31% in soccer. Irish rugby also qualified for and won more major tournaments than soccer despite the €11,815,873 difference in sports capital investment.

6.0 Conclusions

The results showed that Cork receives the highest level of investment of all counties in Ireland receiving €47,039,235.31 which is 23.83% of the total sports capital investment from the 8th of December 2014 to the 31st of August 2020. Cork also has the highest mean investment of all counties with a mean grant of €49,410.96 which is greater than the second highest investment by €9951.01. From observing the mean sum of a grant in each county it was discovered that 61% of counties in Ireland fall below the mean value of €20,025.60. This indicates that investment is not evenly distributed throughout all counties in Ireland.

In the findings of the report the data shows that the ten counties with the lowest level of capital investment receive a total of 11.79% of the sum of investment in Ireland. Geographically, seven of the counties that received the lowest level of investment are clustered. There is a trend of counties that are centralised in Ireland not receiving capital investment, in addition to lack of investment there is no evidence of these counties performing well.

The findings in the report show that there is no correlation between the level of investment in a sport and the success. From our observations in Gaelic Games and the comparison in Rugby and Soccer we see in both cases that the number of accolades won by a team has no link to how much a sport receives. This is seen in Gaelic Games with Cork receiving the highest sum of investment but finding success in only three competitions since 2014 leading to the rejection of the null hypothesis of 'The level of investment has an impact on the number of accolades won' and providing strong evidence to accept the alternate hypothesis of 'The level of investment does not have an impact on the number of accolades won'.

When analysing sports capital investment, the purpose of the investment must be considered and the question of is the level of investment to a county or sport to drive success of a sport or for social benefits, for example to drive involvement within the community. If success is the criteria that investment is measured by there is no evidence of this resulting from the investment as proven in the results.

From the results the observation was made that for individuals under the age of twenty-four Dublin receives the third lowest capital per person in Ireland. Dublin's population under the age of twenty-four is 430,684 which accounts for 32% of Dublin's entire population. The mean investment per individual in Dublin is approximately €88. With Dublin being such an urban area and such a low investment level person factors such as long-term effect must be considered. Given Dublin is the most populous county in Ireland the level of investment is alarming and could having damaging affects to youth and sports development in coming years should this not change.

The level of investment for each county does not follow a normal distribution and is heavily skewed in favour of counties such as Cork. Geographically Cork and Waterford and neighbouring counties and both counties share a border. This would suggest that investment

levels are similar between both counties, but Cork receives €47,039,235.31 in comparison to Waterford's €3,209,197. This suggests that the investment is not thought through and that the investment is ad hoc. The sports grant programme seems to be unstructured, and this could be due to having no dedicated minister for Sport which could suggest that sport investment is not seen as a priority despite reports on obesity in Ireland and teenage disengagement and a more focused approach could have more social benefits in future years.

7.0 Further Development or Research

Throughout the duration of the project, there were many modules that were completed over the course of the academic year. In these modules, methods, techniques, programming languages and software's were showcased and how to utilise this content available to us as explained in the module lectures and labs. As I was learning these methods as the year progressed, it was challenging to implement some methods to the project due to the project being on going and trying to apply the methods I am learning in modules while actively doing the project.

As the project ran alongside other modules and coursework time management was essential for the completion of the project. With proper time management and planning this allowed for other coursework such as continuous assessments and terminal based assessments to completed on schedule, allowing for the final year project to be completed to high standard. If the project were conducted under no time constraints and in my own personal time to progress the project, there are additional topics that I would try to cover.

For further development, the additional topics, with additional time I would like to research are:

1. Additional Data – The primary data set was obtained from Data.gov.ie. On this website, the only data available currently available was the data ranging from the 8th of December 2014 to the 31st of August 2020. To further develop the project and to conduct deeper analysis into sports grants in Ireland, additional data would be required to do so. With additional data, such as data preceding the current data set, a more accurate analysis could be conducted, and further testing could be carried out on the data. Having additional data would allow for a greater detailed time series analysis, and this would allow for a more accurate prediction of future grant levels.
2. A more thorough investigation to additional sports – conduct more thorough research on all sports to assess levels of investment, accolades of the sport and to compare investment versus success at a larger scale.
3. Extensive comparison to other Countries – extend the research into different countries and conduct a similar analysis on investment and sporting accolades, observing investment into specific sports and investments by region. From conducting a similar research, this would mean an accurate comparison can be drawn from the analysis and the performance of Ireland, and the level of grants allocated could be measured against other countries with a similar population density.

4. Lack of investment in inland counties – The inland counties of Ireland receive the lowest level of investment in Ireland, seven of the ten counties with the lowest investment can be found inland. The seven inland counties are Cavan, Laois, Offaly, Carlow, Monaghan, Roscommon, and Longford. These counties account for 8.3% of the total investment. Future research could identify why investment levels are so low in inland Ireland and potentially being disregarded for investment.

8.0 Reference

Census 2016 Small Area Population Statistics - CSO - Central Statistics Office (2016). CSO. Available at: https://www.cso.ie/en/census/census2016reports/census2016smallareapopulationstatistics/ (Accessed: 16 May 2021).

FAI (2021) *Matches | Football Association of Ireland*. Available at: https://www.fai.ie/ireland/senior-men/matches (Accessed: 16 May 2021).

Fayyad (1996) *KDD Process/Overview*, *Advances in Knowledge Discovery and Data Mining*. Available at: http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html (Accessed: 16 May 2021).

IRFU (2021) *Comprehensive Fixtures and Results*, *Irish Rugby*. Available at: https://www.irishrugby.ie/ireland/fixtures-results/ (Accessed: 16 May 2021).

'KDD and Data Mining' (2020) *Data Science Project Management*. Available at: https://www.datascience-pm.com/kdd-and-data-mining/ (Accessed: 16 May 2021).

Roll of Honour (2020) *Six Nations Rugby*. Available at: https://www.sixnationsrugby.com/history/roll_of_honour/ (Accessed: 16 May 2021).

Small Areas Generalised 100m - OSi National Statistical Boundaries - 2015 - data.gov.ie (2015). Available at: https://data.gov.ie/dataset/small-areas-generalised-100m-osi-national-statistical-boundaries-20151 (Accessed: 16 May 2021).

Sports Capital Programme - data.gov.ie (2016). Available at: https://data.gov.ie/dataset/sports-capital-programme (Accessed: 16 May 2021).

Weekly (http://www.worldsportsweekly.com), W. S. (2020) *A List Of Liam MacCarthy Cup Winners*, *World Sports Weekly*. Available at: liam-maccarthy-cup-previous-winners (Accessed: 16 May 2021).

Weekly (http://www.worldsportsweekly.com), W. S. (2020) *A List Of Sam Maguire Cup Winners*, *World Sports Weekly*. Available at: sam-maguire-cup-previous-winners (Accessed: 16 May 2021).

9.0 Appendices

9.1. Project Plan

National College of Ireland

Project Proposal

A Study into the Distribution of Sports Grants by Geography and Category in Ireland

22-12-2020

Technology Management

Data Analytics

Academic Year 2020-2021

Ryan O'Toole

X17409626

x17409626@student.ncirl.ie

Contents

9.1.1 Objectives

The learning objective of this project is to conduct an analysis of Sports Capital Investment in Ireland from the period 2014 to 2020. I will analyse the data by region and sport, and my goal is to identify investment versus the reward of the specific sports. The data will be analysed, making observations such as investment by country and how many elite athletes/victories come from the specified county's and the success of the sport in comparison to investment will also be investigated. Significant events such as the Olympics, European championships and world sporting events will be examined to see how many athletes from respective sports and the allocates won regarding the investment. The findings will be documented, and a similar analysis will be completed and compared to a country with a similar population size to Ireland. The country that the findings will be compared to is Scotland.

Some of the objectives within the project include discovering data such as the sport with the highest investment, the success of a sport in comparison to investment and contrast in success in sporting events of Ireland and Scotland.

Through this analysis, all components of the project must be considered, along with alternative solutions. Technologies and techniques such as machine learning, data mining, the use of statistics to calculate data, visualisation techniques using R-Studio, Excel and SPSS will be used throughout the project to analyse multiple datasets using different techniques to explore and showcase the data.

9.1.2 Background

I gained an interest in sport at a young age and began playing and supporting sport when I started primary school and continue to do so till the present day. I have competed in several different sports such as football, rugby, and cross-country running, so sports have always been an essential part of my health and well-being and fitness, enabling me to make great friends and the bonus of achieving accolades in both regional and country competition.

Reflecting on my younger years of sport, the level of investment within my local football club always seemed to be underfunded, each team would fundraise locally in the community on a voluntary basis through sponsored events and local business, and this was the primary methods of raising funds for the club and purchasing equipment for the teams. Taking this and similar examples from other areas into consideration led me to contemplate sports capital investment within Ireland and how the money is distributed through sports.

In recent years Ireland has found success in sporting events such as rowing, boxing, rugby, etc., and the concept of the project is to analyse the distribution of sports capital and conduct a comparison between the level of investment versus the return. Ireland has seen accolades in both individual and team sports. As of 2012, Ireland has received a total of 24 Olympic medals, multiple European championships in boxing, open championships in golf and six nations victories in 2013, 2014 and 2018.

Based on the 2016 Irish Census, the population of Ireland is now at 4.9m with an increase in family's equating to a growth of 3.3%. The project will investigate does the growth of the population correlates with the level of investment of sports capital.

The analysis will gain insight into whether sports with a lower investment rate outperforming sports with a higher investment rate on a global stage. The level of participation in the sport is something that must be considered when conducting the analysis of the sport and calculating the level of success.

Is success in a field being recognised and further rewarded, if a sporting team or athlete from Ireland is successful in their field, is this success rewarded to the sport and is the level of investment increased to maintain this level of success achieved.

9.1.3 Technical Approach

Once the objectives of the project were outlined clearly, the analysis could then be conducted. The data needed to be identified for the analysis, once the data was identified, it needed to be extracted. The dataset was identified on Data.Gov.ie, and the data was requested for use through contacting the Sports Capital Programme Division in Gov.ie.

Once the dataset was received, it was imported into R-Studio to pre-process the data and cleanse the data. This process involves the removal of unnecessary rows and columns and the removal of null data. Once this process is complete, and all unnecessary elements are removed from the data, the pre-processed data is then exported and is prepared for analysis.

With the clean data set, the analysis can then be conducted. Technologies such as Excel, SPSS and R-Studio will all be used to complete the visualisation of the data.

Additional datasets will be observed and examined thoroughly and will be pre-processed and prepared for use.

With the use of SPSS and Excel, a high level of statistics will be performed on the datasets. Using statistics, the aim is to gather insight into the data, review, analyse and draw conclusions on the data through these findings.

Data Visualization will be carried out on the data to curate the data into a simpler form to highlight trends, outliers and making the data easier to understand. The use of data visualisation enables us to highlight critical information in the data. The data visualisation will be conducted in programmes such as SPSS, Excel, and Tableau.

9.1.4 Special Resources Required

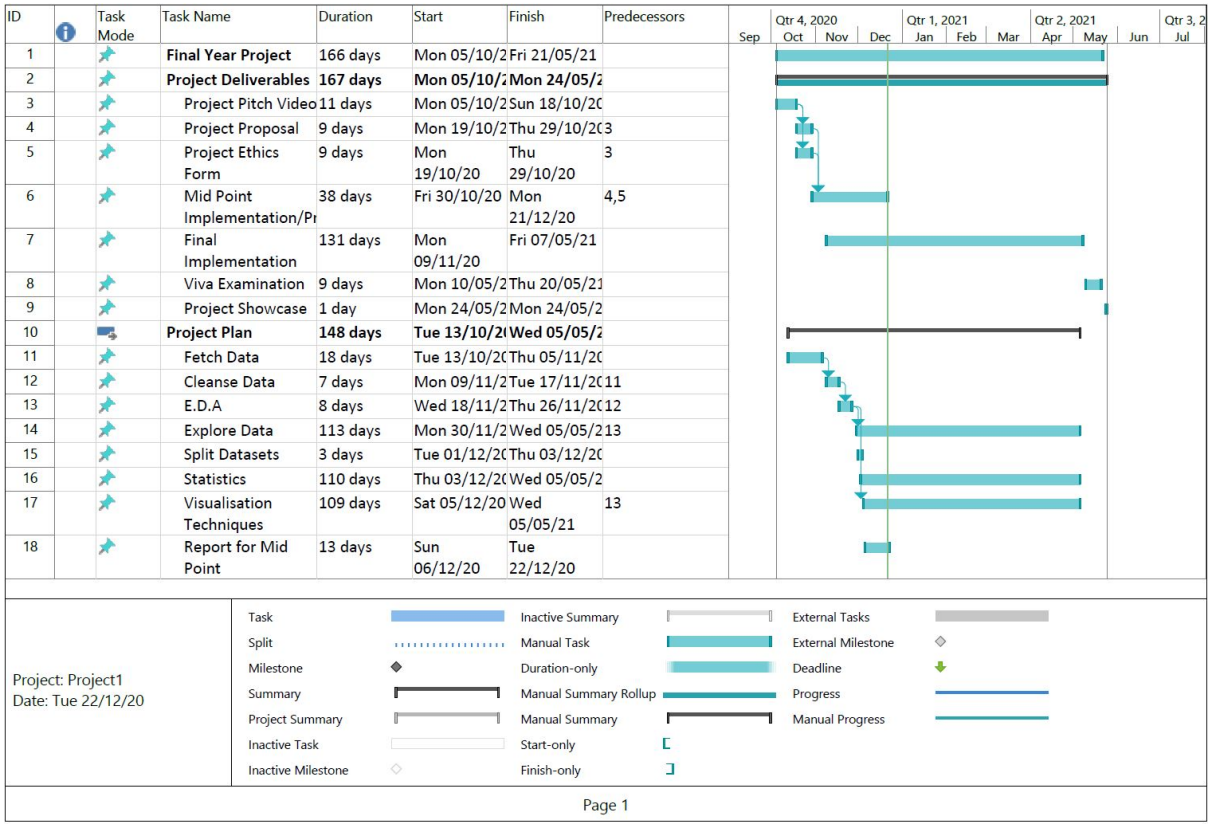
For the project, no special resources are required to complete the data analysis. Contact to external sources to receive datasets for further analysis will be carried out throughout the project. There are no special hardware or literature necessary to complete the project.

Additional learning programmes will be completed through the course of the analysis to gain further insight on how to utilise the software entirely. These programmes will not require special resources.

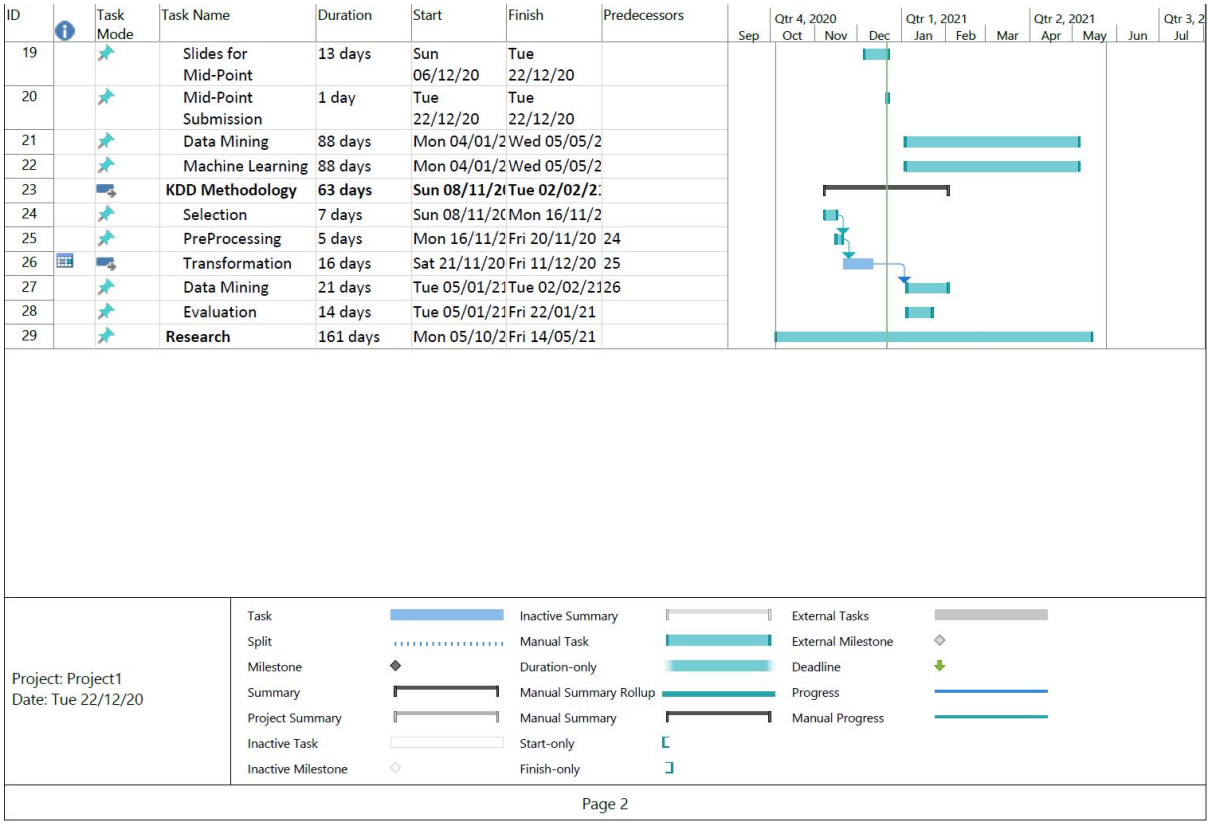
9.1.5 Project Plan

The project plan is outlined clearly in the Gannt chart below. The Gantt chart was created using MSP. The Gantt chart is used to outline all the processes involved in the data analysis in compliance with the time allocated. The Gantt chart below is the proposed timeline in which the criteria of the project will be completed. The Gantt chart was created by highlighting the critical tasks of the project and adding sub-tasks to the key deliverables. The key deliverables were showcased with their submission period, which then enabled sub-tasks of the critical submissions to be created, such as importing the data to R. Due to external factors, some timelines in the sub-tasks may vary through the project. However, the main submissions will remain constant throughout.

The Gantt chart explains the main submissions as the first part and then breaks down the plan of the project in the proceeding sections.



ID	Task Mode	Task Name	Duration	Start	Finish	Predecessors
1		**Final Year Project**	166 days	Mon 05/10/2	Fri 21/05/21	
2		**Project Deliverables**	**167 days**	**Mon 05/10/2**	**Mon 24/05/2**	
3		Project Pitch Video	11 days	Mon 05/10/2	Sun 18/10/20	
4		Project Proposal	9 days	Mon 19/10/2	Thu 29/10/20	3
5		Project Ethics Form	9 days	Mon 19/10/20	Thu 29/10/20	3
6		Mid Point Implementation/Pr	38 days	Fri 30/10/20	Mon 21/12/20	4,5
7		Final Implementation	131 days	Mon 09/11/20	Fri 07/05/21	
8		Viva Examination	9 days	Mon 10/05/2	Thu 20/05/21	
9		Project Showcase	1 day	Mon 24/05/2	Mon 24/05/2	
10		**Project Plan**	**148 days**	**Tue 13/10/2**	**Wed 05/05/2**	
11		Fetch Data	18 days	Tue 13/10/20	Thu 05/11/20	
12		Cleanse Data	7 days	Mon 09/11/2	Tue 17/11/20	11
13		E.D.A	8 days	Wed 18/11/2	Thu 26/11/20	12
14		Explore Data	113 days	Mon 30/11/2	Wed 05/05/21	13
15		Split Datasets	3 days	Tue 01/12/20	Thu 03/12/20	
16		Statistics	110 days	Thu 03/12/20	Wed 05/05/2	
17		Visualisation Techniques	109 days	Sat 05/12/20	Wed 05/05/21	13
18		Report for Mid Point	13 days	Sun 06/12/20	Tue 22/12/20	



9.1.6 Technical Details

The data will be implemented into R-Studio to initialise the data cleansing/pre-processing stage of the data transformation. The procedure of pre-processing the data will follow the KDD methodology. The null data will be removed from the dataset, and the clean dataset will be exported in the format of a CSV file. Once the data has been cleaned, the analysis can then be conducted.

R-Studio will also be used to manipulate the data by creating subsets of the data using data frames so that specific data can be analysed.

Visualisations of the data will be used to simplify the data and showcase the data in a way that is both logical and easy to understand. The tools that will be used to conduct the visualisation of the data will be tools such as Microsoft Excel, SPSS, Tableau and R-Studio. With the use of these tools, the data can be imported and made into graphs to highlight certain aspects of the data.

Rapid Miner will be utilised to conduct data mining techniques. These data mining techniques will include the use of time series analysis. An in-depth analysis will be carried out, and past dates will be studied. From doing this, predictive analysis can be completed, and we are then able to forecast what future allocations may look like based on previous years. From conducting a time series analysis, we can study the changes over time and examine how these changes correlate with other variables.

Elements of machine learning will be used further down the line within the project. Machine learning will be used to identify trends. The technical aspects of machine learning will be discussed further as the techniques of machine learning are carried out.

9.1.7 Evaluation

The datasets for the project will be selected. They will then be processed and manipulated using R-Studio in co-ordinance with the KDD methodology. Once the data is prepared, the data will be used to perform analysis, visualisation techniques and a series of statistics to gain critical insight into the data. Technology tools such as R-Studio, Rapid Miner, SPSS, Excel, and Tableau will be used to explore the data.

The objectives of the project will be met using these tools and the exploration of a series of datasets to retrieve informational insight on the topic. Trends and predictive analysis will also be explored to predict future investment in sports capital.

9.2. Reflective Journals

9.2.1 Reflective Journal 1 - October

Reflective Journal 1

Student - Ryan O'Toole
Student Number - x17409626
Course - Technology Management
Specialisation - Data Analytics

On the 28th of September, we returned to college for our final year. We were introduced to our software project module and the breakdown of what the projects would entail, such as deadlines, percentage weight of submissions and an overview of what the project would require.

We then were told that our project pitch was due on the 18th of October. I began my research and started looking for potential datasets that I could analyse for my data project—searching through websites such as Statista, Data.Gov.ie and the CSO website. Upon browsing the various sites, several data sets stood out in which I could potentially investigate for my project.

Upon my research, I came across two data sets that I felt would allow me to analyse an exciting topic that I am interested in. These data sets from Data.Gov.ie and the CSO website.

The project pitch for my project was the fairness of the distribution of sports capital around Ireland and how well aligned that is over recent years.

Once the pitch was made and submitted, I waited for the pitch to be approved and waited to be assigned a supervisor. The pitch was approved but with amendments to be made to the pitch. Upon receiving this information, my supervisor and I scheduled a meeting to discuss these amendments.

Enda and I will discuss the project amendment on the 2nd of November.

9.2.2 Reflective Journal 2 – November

Reflective Journal 2

Student - Ryan O'Toole
Student Number - x17409626
Course - Technology Management
Specialisation - Data Analytics
Date – 01-12-2020

With the approval of my project concept, I then started the proposal for the project. The initial proposal was discussed with my academic supervisor, and amendments were then made to the pitch to further the scope of the project. The proposal was due on the 8th of November. An ethics form was also required with the upload of the proposal. In the project proposal submission, the project objectives were outlined, the project plan in the form of a Gantt chart, the technical details and approach of the project and the overall evaluation.

On the 26th of November, Enda and I agreed to meet the following day to discuss the overall progression of the project. The discussion was held on Microsoft Teams on the 27th of November, in which the project in its current state was discussed.

The overall progression of the project has been slower than expected due to large amounts of other coursework with high weight due. Allocating time to work on the project has been complicated with the other deadlines, but the project will still be in line with the requirements on the Rubrik.

Although I have not been able to work on the project as much, a similar project was completed through one of my other modules in Data application development. In this project, I became familiarised with the use of R and R-Studio and began to understand R-Studio further. This will benefit me in the project by already having knowledge with R-Studio and using different methodologies such as the KDD methodology.

9.2.3 Reflective Journal 3 – December

Reflective Journal 3

Student - Ryan O'Toole
Student Number - x17409626
Course - Technology Management
Specialisation - Data Analytics
Date – 01-01-2021

In December, the first submission worth a percentage of the final year project's grade was due. The mid-point submission for the project was due on the 22nd of December.

To prepare for the mid-point submission, meetings with Enda were arranged on three occasions to discuss the project's plan and the progress so far.

After the data was processed using R-Studio, the clean data was split into multiple sub data sets. These sub-datasets included the data being divided into categories such as sport, region, province, payment scheme and date of the payment. This allowed me to analyse specific aspects within the data and present the results by dividing the data.

To prepare for this presentation, different technologies were used to analyses the data. These technologies introduced to the project included Tableau, SPSS, Excel, and GitHub.

With these technologies, graphing and statistics were conducted on the data to gain further insight and make the data more readable and easier to understand.

For the midpoint submission, a presentation was created using Canva, which entailed eight slides giving insight into the project's current stage. The presentation included an overview of the project, technologies used in the project so far, preliminary data analysis and a demonstration of the preliminary data analysis/

A video presentation was completed and published on YouTube as an unlisted video to then be marked by academic supervisors. In the presentation, a demonstration of the project was showcased, such as R-Studio code, Visualisation in SPSS, and Tableau.

Multiple graphs were created in SPSS and Tableau to showcase the data, and an interactive graph was also made in R-Studio to demonstrate they mean investment by county.

As the submission point came closer, I became more familiar with the technologies being used within the project, and the learning of new technologies also commenced, such as Python. With this, I plan to further my knowledge and further enhance and implement new techniques to the project to uncover more findings.

The plan in the future with the project is to try to implement Data Mining and machine learning to the project as I continue to learn new tools.

In the research interesting data was uncovered, trends and outliers have been identified, and with some further analysis, the next step in the project is to finish the analysis on Ireland and then conduct a similar study on Scotland to show a comparison.

9.2.4 Reflective Journal 4 – January

Reflective Journal 4
Semester 2

Student - Ryan O'Toole
Student Number - x17409626
Course - Technology Management
Specialisation - Data Analytics
Date – 01-02-2020

In January progress on the software, the project was slow, due to other coursework commitments and terminal-based assessments due from the 1st of January to the 20th of January attention was redirected towards completing these assessments to a decent standard after focusing full attention to the software project mid-point submission.

Once the terminal-based assessments concluded the progress of the project, then recommenced. Planning was put in place of what research would be done and future for the project and how to further the research.

An investigation was conducted to obtain data similar to the current dataset on Ireland currently being analysed. Emails were then sent to Scottish data centres to obtain this data.

New methods of testing the data with statistics were also introduced to predict values.

Overall, the progress of the project was slow due to other coursework due within this period. However, in February, there are plans to make up for this project work to further progress the current state of the project and to further discover different aspects within the data.

Contact has been made with Enda to plan future meetings and discuss the project's current and future state. Feedback is also something that I hope to discuss in the meeting to understand what aspects of the project I performed well in and where I need improvements.

9.2.5 Reflective Journal 5 – February

Reflective Journal 5
Semester 2

Student - Ryan O'Toole
Student Number - x17409626
Course - Technology Management
Specialisation - Data Analytics
Date – 01-03-2020

For the month of February statistical testing was conducted for the project. Data Analysis started by doing testing such as T-Tests, Anova testing and time series analysis.

The project showcase template was completed, and meetings were held with both Siobhan Mockler and Helen Conway to ensure this was completed correctly.

Further splitting of the data was completed for further analysis and contact was made with representatives in Wales to obtain sports capital investment for the comparison between Ireland and another country as data could not be obtained from Scotland.

A meeting will be arranged with Enda in the coming days to discuss the current stage of the project and the future to progress the project.

9.2.6 Reflective Journal 6 – March

Reflective Journal 6
Semester 2

Student - Ryan O'Toole
Student Number - x17409626
Course - Technology Management
Specialisation - Data Analytics
Date – 01-04-2020

Data was obtained from the United Kingdom, Wales, and Scotland on their sports capital investment levels by year. This allowed me to compare investment levels by year for each country against Ireland's level of investment.
As the data is not as detailed as the primary data set obtained from Data.Gov.ie for Ireland, for some countries, an analysis on the level of investment could only be done rather than breaking investment down by sport.

A new data set was created with three columns, including Country, Sport, and year, including medals won by a country. With this data, statistical testing learnt through the Advanced Business Data Analysis module was conducted, and tests such as Wilcoxon Tests, Mann-Whitney U-tests, and Kruskal-Wallis test as the data is non-ordinal data.

Visualisations were created from the testing to further explain the testing results and state null and alternate hypotheses before conducting the testing. This testing was completed using SPSS and R-Studio.

A meeting will be arranged with Enda in the coming days to discuss the project's current stage and the future to progress the project.

9.2.7 Reflective Journal 7 – April

Reflective Journal 7
Semester 2

Student - Ryan O'Toole
Student Number - x17409626
Course - Technology Management
Specialisation - Data Analytics
Date – 01-05-2020

Entering the final month of the project the primary focus was to prepare the word document for the final submission. This included adding all necessary appendices, reformatting aspects of the document and re-touching certain aspects of the report such as methodology, analysis, and the results section.

The project poster was designed based off the headings provided on the template from Moodle. Screenshots of the project and technologies used to conduct the analysis were all included in the poster. The poster was created on Luicdpress.

Additional results were created for the report which included additional visualisations and the use of additional datasets obtained from the CSO. The dataset obtained included census data from 2016 and was used to observe the total population of each county under the age of twenty-four to observe investment levels per individual under this age demographic.

An extra technology was added to technologies used, Microsoft Power BI was used to create a series of graphs and re-create graphs previously made in SPSS. This was done to increase the quality of the output as the visualisations had a large impact on the results.