National College of Ireland

BSHC Computing

Data Analytics

2021/2022

Brandon Caffrey

18514113

x18514113@student.ncirl.ie

A Comparative Analysis on the Trends of Crime in Dublin using Predictive Analytics

Technical Report

Contents

Executive Summary

Dublin is rich in culture and history, but as any other large city, it is also afflicted with high crime rates. Crime rates, however, are widely expected to vary across its four districts (Fingal, Dublin City, South Dublin, and Dun Laoghaire – Rathdown). The purpose of this project is to analyse factors and trends that can lead to crimes being committed in each district of Dublin. These trends include the level of garda personnel, education attainment and income. Focusing on the four districts of Dublin, this project examines how these trends may affect the levels of crime in each district. Crime is measured by the total amount of offences committed in each district ranging from many different offences such as theft and burglary to harming the environment related offences and so on.

This project also uses predictive analytics in the form of time series forecasting to make a prediction on each of the trends stated above. Using the ARIMA model, we have made a prediction on the number of offences committed per region, the amount in second level education per region and the average amount of income in the whole of Dublin.

Many of the results found were expected such as the predictions, however there were some unexpected results when measuring whether the trends have a strong effect on the number of crimes committed. As a society, we naturally assume that the trends that I have previously mentioned will allow for a decrease in the levels of crime. However, this is not the case for some of the different districts of Dublin.

1.0 Introduction

1.1. Background

We chose to undertake this project because I have seen the effects of crime firsthand in the Dublin area. We also have friends and family who have been victims of crime in the area. We want to be able to find many of the root causes of crime and figure out why high levels of crime is occurring in our area. One of the main aims in my project is to promote and raise awareness of how there are multiple factors that are used as a catalyst towards crime.

1.2. Aims

This project aims to showcase how many different factors can be seen as a gateway to crime. It will gather social economic factors such as education attainment and income and showcase how these factors can lead people to committing crimes in order to get by. It also shows how the level of garda personnel can also affect the levels of crime in Dublin. This project also aims to raise awareness on how we, as a country can use these social economic factors to allow people to not commit crimes. This project will also show how the number of employed garda in an area can affect the number of crimes committed.

1.3. Technology

The technologies being used in this project so far are R, R Studio, Microsoft Excel, and SPSS.

R – R is a coding language used in statistical and mathematical computing and graphics. The main body of this project will be computed using R. R allows us to find insights on the data that we have.

R Studio – R studio is the IDE (Integrated Development Environment) where we code R. It is the development environment of R. There are many features within R Studio that assist us in finding insights with our data. It has an easy and open user interface which helps in navigating around it.

Microsoft Excel – Microsoft Excel is part of the Office 365 package. It is a spreadsheet program that allows users to create, organize, store, and transform data.

IBM SPSS – SPSS is a software that is used to conduct statistical analysis. It is useful for creating descriptive statistics in a quick and simple way. It is also used for creating graphs and plots from data that is imported into it.

1.4. Structure

Section 1 (Introduction) – Contains information on what the project is, the aims of the project and tools and technologies used within the project.

Section 2 (Data) – Provides information on the datasets used in this project so far.

Section 3 (Methodology) – Explains what type of methodology is used in this project for analysis and insights.

Section 4 (Analysis) – Contains information on what analysis has been performed so far in this project.

Section 5 (Results) – Provides the results of the analysis with graphics and figures.

Section 6 (Conclusion) – Outlines the strengths, advantages, limitations, and disadvantages of the project.

Section 7 (Further Development or Research) - Provides information on the direction of the project and where I feel the project can go in the future.

2.0 Data

The datasets that have been used for this Software Project can be found below. All of these datasets were sourced via the Central Statistics Office except for the 'dublin population.csv' dataset which was gathered from worldpopulationreview.com. The table below presents a summary of the datasets used before any form of analysis was conducted.

Data File	***#Records***	***Description***	***Attributes***
offences.csv	6397	Contains information on crime offences committed in Dublin from 2008 - 2020	Statistic, Year, Location, Type_of_Offence, Unit, Value
gardaEmployment.csv	79	Contains information on the number of gardai employed in the Dublin regions	Statistic, Year, Garda_Region, Unit, Value
pupils in secondary school 2008-2020.csv	53	Contains information on the number of pupils enrolled in second level education from 2008-2020	Statistic, Year, County, Nationality of Pupil, Unit, Value
dublin population.csv	87	Contains information on the estimates of Dublin population from 1950 – 2035.	Year, Value
population of dublin regions 2016.csv	4	Contains information on the population of the four districts of dublin from the 2016 Census.	Statistic, CensusYear, Sex, County and City, Unit, Value
income	265	Contains information on the average household income	Statistic, Year, Urban and Rural, Unit, Value

Exploratory data analysis was carried out on many of these datasets.

Offences:

For example, there was 1,275,891 offences committed from 2008 – 2020 in Dublin alone. Within the same time period, there was 353,370 members of An Gardai Siochana employed. In terms of the different districts of Dublin, there were 170,055 offences committed in

Fingal over the 13-year period with a garda personnel level of 86,497. Dublin City had 597,855 crime offences reported during this time, with 137,850 members of gardai employed in that area. The South Dublin district saw 208,066 crimes committed from 2008-2020 whilst having a garda personnel strength of 62,470. The district of Dun Laoghaire – Rathdown has a total of 188,303 offences committed with a lower level of garda personnel of 49,497.

As part of the exploratory data analysis, the total number of offences per garda station in each district were found. Please see the tables below.

Fingal

	Garda Station	#Offences
1	63101 Balbriggan, D.M.R. Northern Division	18226
2	63102 Garristown, D.M.R. Northern Division	1002
3	63103 Lusk, D.M.R. Northern Division	9026
4	63105 Skerries, D.M.R. Northern Division	4221
5	63202 Dublin Airport, D.M.R. Northern Division	4988
6	63302 Malahide, D.M.R. Northern Division	12927
7	63303 Swords, D.M.R. Northern Division	33767
8	63402 Howth, D.M.R. Northern Division	10428
9	66101 Blanchardstown, D.M.R. Western Division	75470

Dublin City

	Garda Station	#Offences
1	61102 Irishtown, D.M.R. South Central Division	17044
2	61202 Pearse Street, D.M.R. South Central Division	129989
3	61301 Kevin Street, D.M.R. South Central Division	41245
4	61302 Kilmainham, D.M.R. South Central Division	31994
5	62203 Mountjoy, D.M.R. North Central Division	32645
6	62301 Store Street, D.M.R. North Central Division	118292
7	63201 Ballymun, D.M.R. Northern Division	27075
8	63203 Santry, D.M.R. Northern Division	29611
9	63301 Coolock, D.M.R. Northern Division	37521
10	63401 Clontarf, D.M.R. Northern Division	21556
11	63403 Raheny, D.M.R. Northern Division	12726
12	64302 Terenure, D.M.R. Southern Division	15195
13	66102 Cabra, D.M.R. Western Division	17096
14	66103 Finglas, D.M.R. Western Division	40361
15	66201 Ballyfermot, D.M.R. Western Division	25505

South Dublin

	Garda Station	#Offences
1	61101 Donnybrook, D.M.R. South Central Division	19964
2	64101 Crumlin, D.M.R. Southern Division	18460
3	64202 Tallaght, D.M.R. Southern Division	72414
4	66202 Clondalkin, D.M.R. Western Division	33893
5	66203 Rathcoole, D.M.R. Western Division	8569
6	66301 Lucan, D.M.R. Western Division	20590
7	66302 Ronanstown, D.M.R. Western Division	34176

Dun Laoghaire – Rathdown

	Garda Station	#Offences
1	64102 Sundrive Road, D.M.R. Southern Division	19142
2	64201 Rathfarnham, D.M.R. Southern Division	29531
3	64301 Rathmines, D.M.R. Southern Division	21523
4	65101 Blackrock, Co Dublin, D.M.R. Eastern Division	17766
5	65102 Dundrum, D.M.R. Eastern Division	39130
6	65201 Cabinteely, D.M.R. Eastern Division	14549
7	65203 Dun Laoghaire, D.M.R. Eastern Division	32899
8	65205 Shankill, D.M.R. Eastern Division	13763

From the above tables it can be concluded that the garda station with the most criminal offences reported **(129,989)** was **'61202 Pearse Street'** in the South-Central Division within the Dublin City district. The garda station with the least number of criminal offences **(1,002)** reported was the **'63102 Garristown'** station within the Fingal district.

The plot below also shows the yearly number of offences committed by each region from 2008 – 2020. These yearly figures were compiled using R's aggregate function. As it can be seen the Dublin City region has a more substantial number of offences committed compared to the others.

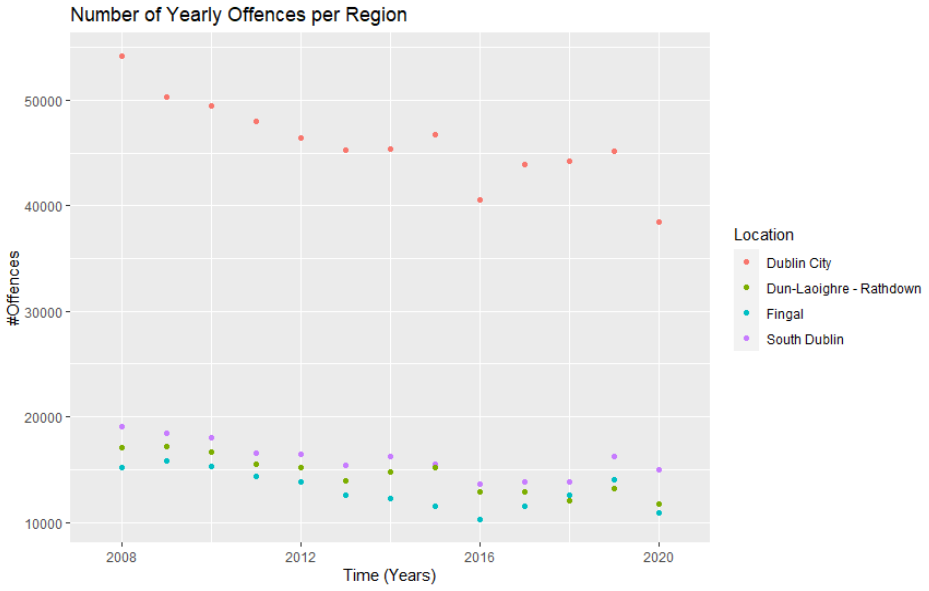


Figure 1 – Number of Yearly Offences per Region

Garda:

The plot below outlines the yearly number of garda personnel for each region in Dublin from 2008 – 2020. The Fingal region has had the highest level of garda personnel closely followed by the Dublin City and South Dublin regions whereas the Dun Laoghaire – Rathdown has a much lower level of garda personnel comparted to the others.

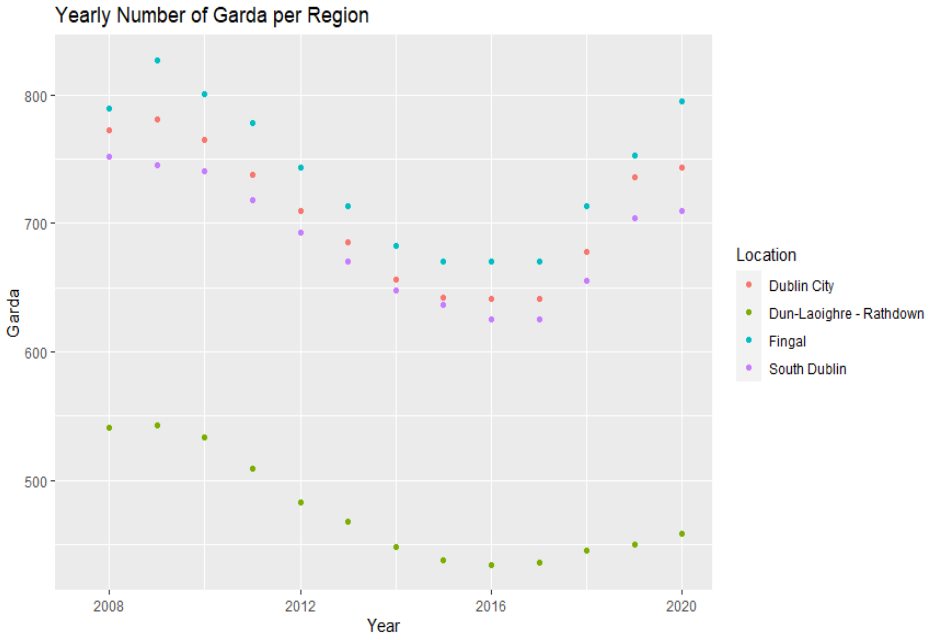


Figure 2 – Yearly Number of Garda per Region

Education:

The plot below displays the total amount of people in second level education for each region from 2008 – 2020. Similar to the offences plot, it can be seen that Dublin City has had the most amount of people in education during this period.

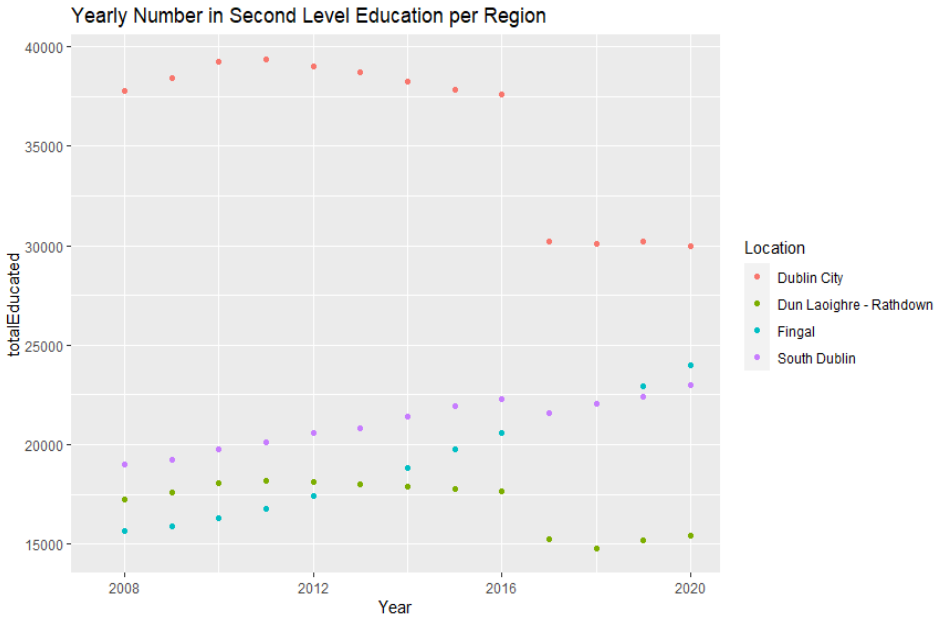


Figure 3 - Yearly Number in Second Level Education per Region

Income:

The plot below shows the average income of Dublin as a whole. As we can see the income levels of Dublin are continuously rising from 2012 onwards. The decrease from 2008 – 2012 may be due to the recession that Ireland went through during that time. It is also worth noting that the datapoint for 2020 is a prediction.

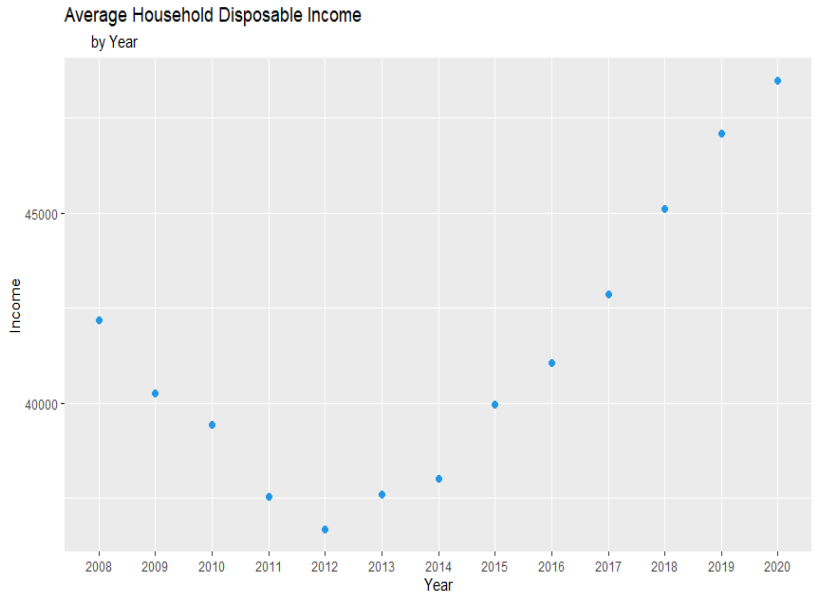


Figure 4 - Average Household Income - Dublin

3.0 Methodology

The methodology that followed for this final year project is the KDD methodology. The KDD methodology stands for Knowledge Discovery in Databases. The KDD process is used to extract knowledge from large databases. It uses Data Mining algorithms to find what's deemed knowledge. The concept of Knowledge Discovery in Databases is a programming approach that involves the study and modelling of vast data sets. This procedure aims at identifying useful and abnormal patterns in the data. The KDD methodology contains multiple steps to be carried out.

Each step in this step-by-step methodology needs to be followed accurately in order to determine the best results for my project.

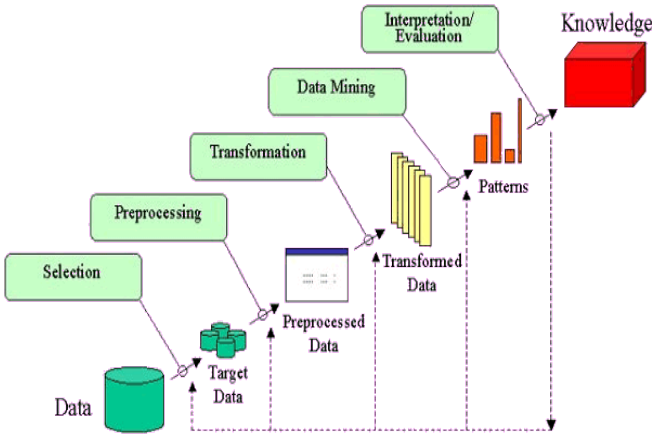


Figure 5 - KDD Process

Source - https://external-content.duckduckgo.com/iu/?u=http%3A%2F%2Fwww2.cs.uregina.ca%2F~dbd%2Fcs831%2Fnotes%2Fkdd%2Fkdd.gif&f=1&nofb=1

Selection:

This step involves selecting the data that will be used for my software project. As mentioned in section 3, there are six datasets used in this project. I feel that these datasets contain the data that I will need in order to carry out my analysis and provide insights.

Pre-Processing:

Out of all of the datasets used within this project, there was only one that contained null values. The dataset that contained null values was the income dataset. The null values were evident in the *'value'* attribute for income for people aged between 0-17 years. To deal with the null values, they were removed in order to assist with the predictions. The yearly values of income were needed, so these null values were of no use. The data frame containing the cleaned data was called '*average_income*'.

Transformation:

There has been large amounts data transformation to assist me with my analysis. In order to combine the offences.csv and the gardaEmployment.csv files mentioned in Section 3, both csv files were exported from R Studio into Excel. In Excel, the data was filtered from the gardaEmployment.csv and added into the offences.csv. The value attribute and the garda region attribute were filtered in order to find the number of gardai employed within the different Dublin regions. This filtering was a lengthy process, to add the data needed which allowed me to create the offencesPerGarda dataset. A summary of this dataset can be found in the table below.

Data File	***#Records***	***Description***	***Attributes***
offencesPerGarda.csv	6397	Contains information on crime offences committed in Dublin and the number of gardai per region at that time from 2008 - 2020	Statistic, year, location, type_of_offence, number_of_gardai, value

This dataset was then filtered for each region, each specific to their regions/districts. For example, the offences_fingal data frame consisted of data relating to the Fingal area. From each data frame the yearly total number of offences for each region was found by using the aggregate function alongside the sum function which summed the number of offences by each year. The yearly total number of offences for each region were combined to create a data frame called *'totalOffences_all_regions'* using the rbind function. The rbind function allows to combine each data frame by their rows.

The next part of the pre-processing stage was to create a data frame that consisted of the number of offences and the number of gardai for each region from 2008 – 2020. After creating a new data frame called *'garda_all_regions'* by finding the yearly number of garda employed within each region, it was combined with the *'totalOffences_all_regions'* to create *'offences_garda_all_regions'* by the cbind function. Similar to the rbind function the cbind function allows me to combine the data frames by column rather than row. The *'offences_garda_all_regions'* now contained the yearly amount of garda and offences per region.

Using the aggregate function to get the yearly amount of average income for Dublin for the time period from 2004 – 2019 proved beneficial. Using the ARIMA model, a time series forecast was made for the average income for 2020. This forecasted value was then added to the original data frame containing the yearly average incomes. As the main timeline is from 2008 – 2020, the values from 2004 – 2007 were removed. Now, the yearly average income in Dublin from 2008 – 2020 had been found.

For the Education variable, the aggregate function was used to get the total yearly number in second level education for each region in Dublin from 2008-2020. When the total yearly number in second level education for each region were gathered, they were combined using the rbind function. This created a data frame called *education_all_regions.*

The project needed the population for each region from 2008 – 2020, however the only data that existed regarding this was from the 2016 census. Using the population values for each region in 2016 and dividing by the total population of dublin, how much each region makes up Dublin was found. For example, the population of Fingal in 2016 was 296020, when this figure was divided by the total population of Dublin in 2016, a value of 22% occurs. This assumes that in 2016, 22% of the Dublin population was made up of Fingal citizens. This approach was followed for the other regions in Dublin also. When how much each region makes up Dublin was found, population estimates of Dublin were retrieved from the worldpopulationreview.com. Each yearly population value of dublin was then multiplied by the percentage of which a region makes up, which gave the population values of each region. E.g., in 2008 the estimated population of Dublin was 1073906. To find the population of Fingal in 2008, multiply that figure by 0.22 as Fingal makes up 22% of Dublin's population. This approach was followed for each year (2008 – 2020) and for each region of Dublin. The population of each region from 2008 – 2020 was now found. The population of each region was then added to the *education_all_regions* data frame mentioned above by using the cbind function.

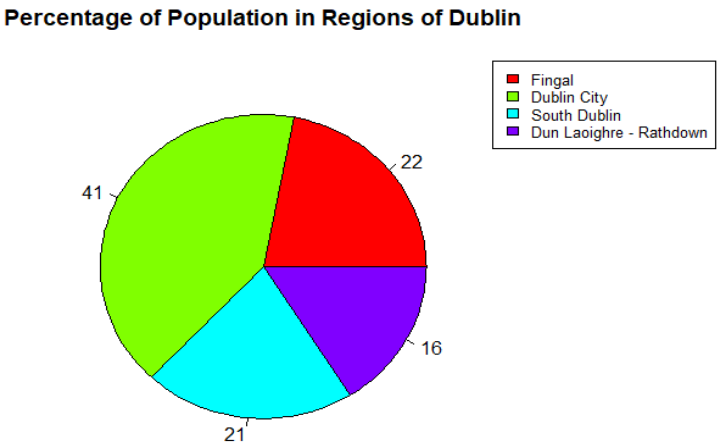


Figure 6 - Percentage population of each region - 2016

As mentioned previously, this project aims to show how Income may affect criminal offences in Dublin. Using the aggregate function, the yearly average income for the entirety of Dublin from 2008 – 2020 was found. This allowed for an easy join with the *education_all_regions* data frame by using the cbind function.

At this stage there were two data frames that needed to be joined together. The two data frames that needed to be joined were the *offences_garda_all_regions* and the *education_all_regions.* Using the rbind function, a data frame called *offences_garda_education_income_all_regions* was created. This data frame consisted of all of the relevant data that was needed to show how factors such as education attainment, income, and the number of garda personnel impact crime.

As the population of each region had been found, it was necessary for this project to find the crime rate and education rate for each region and add them to the *offences_garda_education_income_all_regions* data frame. The formula to find the yearly crime rate for each region was offences/population * 100. This returns a value which is the crime rate per 100 people. Similarly, the formula to find the yearly education rate was number in education/population * 100. The crime rate and education rate were both added to the *offences_garda_education_income_all_regions* data frame. Finally, a data frame that had the following columns was created.

Year	Garda	Location	TotalOffences	Population	Education	Income	CrimeRate	EducationRate

Data Mining:

There are two data mining techniques used within this project. The main data mining technique used for this project is time series forecasting. Time series forecasting uses historical data in order to make predictions on that data. Time series forecasting is used to predict the number of offences per region, the number in second level education per region, the number of garda personnel per region and the average income of Dublin. The time series forecast makes a prediction for 2020 on income and makes a prediction for 2021 for the other trends that have been outlined previously. Time series forecasting would be a great technique for this project as many of the datasets had a timeline within them, it would be beneficial for the project to provide some form of predictive analysis.

Another data mining technique used in this project is clustering. The clustering technique that is used with our data is K-means clustering. In simpler terms, clustering is splitting data into groups based off their similarity. In this project, clustering was applied to the yearly income data. The clustering technique that was applied to the income data was k-means. In order to do this, the number of clusters for the analysis were selected. When the number of clusters were chosen, the starting centroids otherwise known as 'seeds' were chosen. It is chosen at random. Each data point is then assigned to a starting centroid; this is measured by its distance to the centroid known as the Euclidean distance. Each data point is then measured based on their distance to each starting centroid. Applying k-means clustering to

my data will allow me to see what datapoints are similar to each other in terms of how they are grouped together.

Interpretation and Evaluation:

This final step of the KDD process is for providing patterns/results based off your analysis. It is the step before it can assume that the insights found can be considered as factual knowledge. It takes in all previous steps of the KDD process in order to provide conclusions and evaluations. Following the KDD process allowed for interesting insights and patterns to be found with my data. The interesting insights found within this analysis were visualized within R Studio. In section 5 of this report, you will find the insights that were discovered and the results of the analysis.

4.0 Analysis

The main approach that for this project was predictive analysis. Predictive analysis was employed because it is an interesting type of analytics as it allows us to look into the future based off historical data. This approach was chosen because it tied in well with how the data was structured. The datasets contain time attributes such as year, these attributes can be used to be able to perform predictive analysis using time-series forecasting.

Time series forecasting is the method of exploring and analysing time-series data recorded or collected over a set period of time. This technique is used to forecast values and make future predictions (Biswal, 2021)

The predictions were made using time series forecasting with the ARIMA package. ARIMA model – ARIMA stands for Auto Regressive Integrated Moving Average. This is a linear regression model that explains a given time series based off its past values. Using ARIMA, a predictive model can be created to predict Xt when there are past values. This is denoted as p (Xt | Xt – 1…., X1). The forecasts used in this project use the auto. arima function to estimate the missing values within the predictions.

This analysis is used to predict the following:

- The number of offences committed per region
- The level of garda personnel per region
- The number in second level education per region
- The average income of Dublin

Each dataset regarding the following above, have a time attribute which will be used for the time series forecasting. The time series forecasting analysis forecasts yearly values for the above. Therefore, it is crucial that each of the datasets are transformed into yearly values ranging from 2008 – 2020. The aggregate function allowed for the yearly number of offences, people in second-level education, level of garda personnel, and average income of Dublin per region to be found. As the yearly number of each were found, time series forecasting on the yearly values can be performed.

The time series forecast is started by creating our time series object. For each region, each time series object was named appropriate to what is being predicted. For example, for the offences prediction each time series was called *'region_offences'*. This allowed for the time series object to be distinguished easier. Within the time series object, the start and end date are specified. The start date was 2008 and the end date was 2020. However, for the Income prediction, the start date was 2004 and the end date was 2019. Within the time series object, the frequency at which the predictions are made must also be specified. The time series forecast uses yearly frequency which means we give frequency a value of '12' as there are 12 months within a year. As we can see the ts function is applied to the 'fingal_offences' values as that is the attribute that is used for the prediction.

```
fingal_offences <- ts(totaloffences_fingal_yearly$Totaloffences,
                      start=c(2008), end=c(2020), frequency=1)
```

After the time series object is created, the forecast object is created. The arima objects were named similar to the time series objects. Each arima object for each region was named as 'fit_*region*'. Within the forecast object we specify our ARIMA model. This is done by using the auto. arima function with the time series object as its argument. Essentially, we apply the auto. arima function to the time series object.

```
fit_fingal <- auto.arima(fingal_offences, trace = T)
```

Following the arima objects the forecast function was applied to our arima objects. This is done by the following code:

```
forecast_fingal_offences <-forecast(fit_fingal, 1)
```

The arima objects were passed as the argument of the forecast function and specify the length of the forecast. With a value of 1 also passed as an argument, to forecast the value for 2021. This returns a yearly prediction of the number of offences committed in 2021. This method was applied to each region to predict the number of offences, the level of garda personnel and the amount in second level education. The same method was also applied to predict the average income in Dublin.

Also, within the analysis, trends such as the level of garda personnel per region, the education attainment per region and the average income of Dublin were explored to see how they can affect the number of offences committed. As a society, we naturally assume that these factors would influence the number of offences committed in a positive way. For example, the more garda employed in a particular region suggests that there would be a reduction in the number of criminal offences. It was essential to have one data frame consisting of all relevant attributed and values in order to explore the relationships that exist between the variables/attributes. Data visualization tools such as the ggplot package in R was used to visualize the relationships that existed.

Another data mining technique used in this project is the use of clustering analysis. Clustering analysis is performed on the yearly income values to determine what incomes are similar to each other from 2008 – 2020. The clustering analysis uses k-means clustering to determine the clusters within our income values. The cluster analysis was visualized using fviz_cluster in R. The kmeans clustering analysis is performed using the kmeans function in R. The number of clusters that is used in the cluster analysis is also passed as an argument within the kmeans function.

5.0 Results

Please find the results of this analysis below.

Offences:

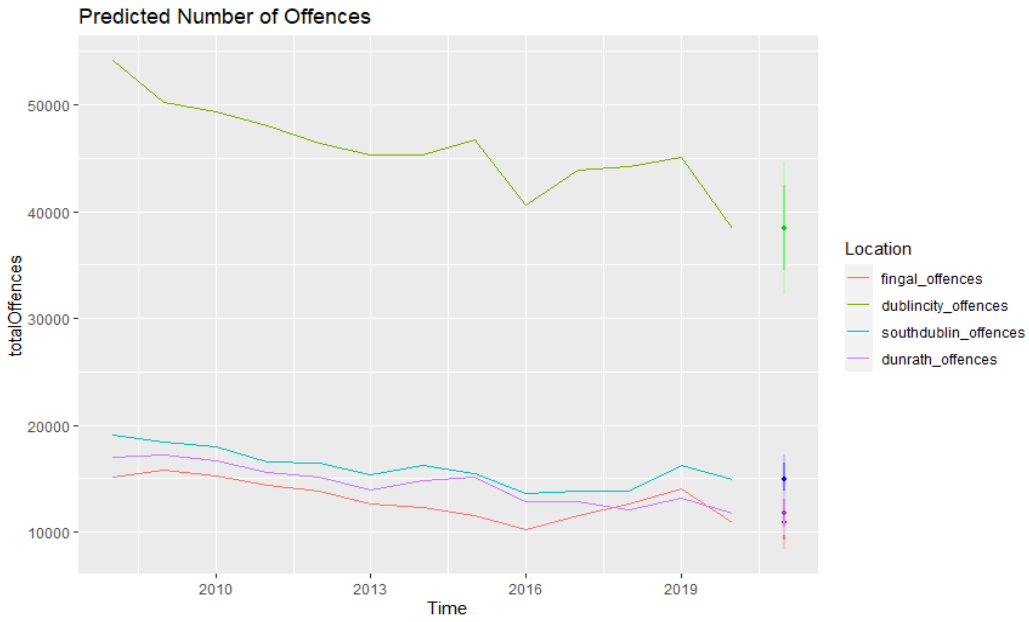


Figure 7 - Predicted Number of Offences

Fingal:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021	10892	9238.042	12545.96	8362.489	13421.51

Dublin City:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021	38465	34453.52	42476.48	32329.97	44600.03

South Dublin:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021	14977	13501.28	16452.72	12720.08	17233.92

Dun Laoghaire – Rathdown:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021	11759	10403.02	13114.98	9685.201	13832.8

The plot above highlights the 2021 predicted number of offences for the four districts of Dublin. Dublin City has the greatest number of predicted offences whilst the Fingal area has the least number of predicted offences. Each prediction shaded area is colour co-ordinated to the line plot. Below the graph are the forecasted values for each area. The forecast is made up of:

- Forecast – The average forecasted value
- Lo 80 – Lower 80% confidence interval
- Hi 80 – Higher 80% confidence interval

- Lo 95 – Lower 95% confidence interval
- Hi 95 – Higher 95% confidence interval

The 80% intervals suggest that the forecast is 80% confident that the prediction lies within the value associated with them whereas the 95% intervals suggest that the forecasts is 95% confident that the prediction lies within the values associated with them. For example, looking at the Fingal forecast, the forecast is 95% confident that the prediction is between 8362 offences and 13422 offences. Similarly, the forecast is 80% confident that the prediction is between 9238 offences and 12546 offences.

Level of Garda Personnel:

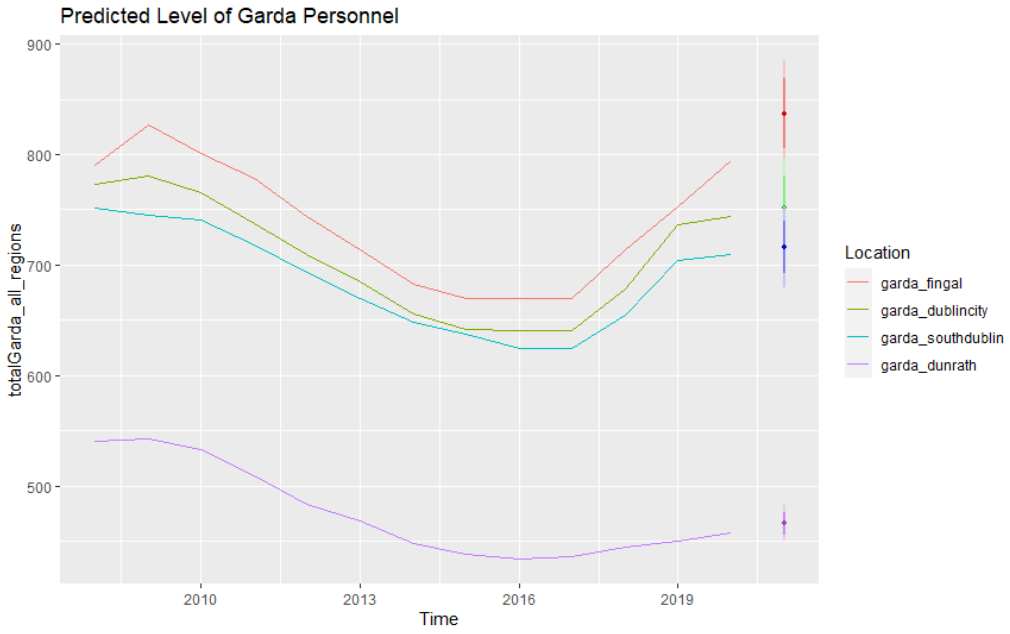


Figure 8 - Predicted Level of Garda Personnel

Fingal:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021	837	805.6559	868.3441	789.0634	884.9366

Dublin City:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021	752	723.5288	780.4712	708.457	795.543

South Dublin:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021	716	692.3184	739.6816	679.782	752.218

Dun Laoghaire – Rathdown:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021	466	455.513	476.487	449.9616	482.0384

The graph above shows the predicted level of garda personnel for the four districts of Dublin. The graph displays that the Fingal area has the greatest prediction in terms of level of garda personnel whereas the Dun Laoghaire – Rathdown area has the lowest prediction. A difference of 371 exists between the greatest and lowest forecasts. Looking at the Dublin City area we can see that the forecast is 95% confident that the level of garda personnel for that area is between 680 and 752 whereas the forecasts is 80% confident that the level of garda personnel is between 724 and 780. It is worth noting how close the forecast above for the Fingal and Dublin City areas are quite close, however there is a huge difference between the two areas with regard to the number of offences committed in each area.

Education Attainment:

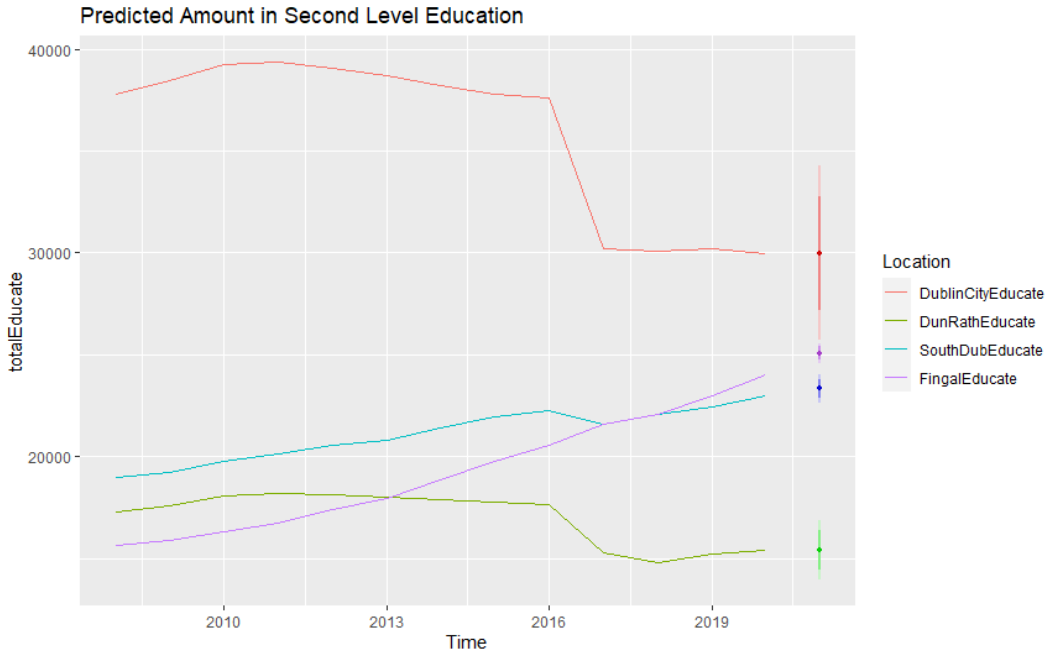


Figure 9 - Predicted Amount in Second Level Education

Fingal:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021	25063	24736.91	25389.09	24564.29	25561.71

Dublin City:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021	29979	27186.15	32771.85	25707.7	34250.3

South Dublin:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021	23322.5	22876.92	23768.08	22641.05	24003.95

Dun Laoghaire – Rathdown

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021	15402	14450.67	16353.33	13947.07	16856.93

The plot above illustrates the forecasted amount in second level education across the four districts in Dublin. The plot shows that the Dublin City area has the greatest prediction with regard to the number in second level education whilst the Dun Laoghaire – Rathdown area has the lowest forecast in terms of the number in second level education. Observing the South Dublin area, the forecast indicates that it is 95% confident that the number in second level education lies between 22,641 and 24,003. The forecast also indicates that it is 80% confident that the number of second level education falls between 22,876 and 23,768.

Income:

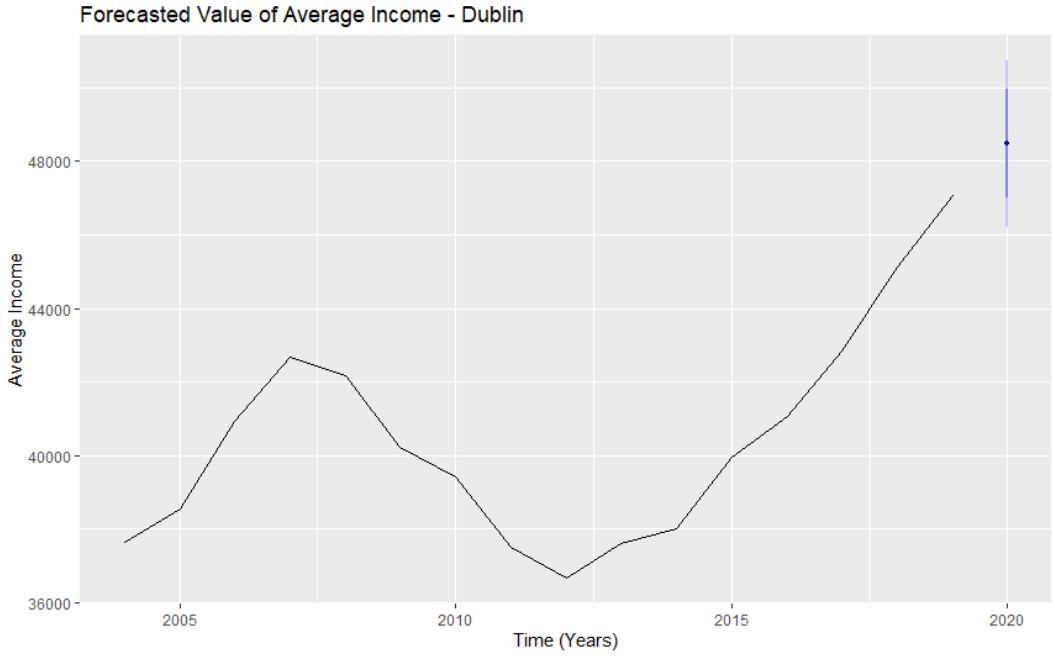


Figure 10 - Forecasted Value of Average Income

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2020	48471.14	46994.85	49947.43	46213.35	50728.93

As mentioned previously, the timeline for the Income data is from 2004 – 2019. Therefore, to get the most accurate forecast, the forecast is for the next consecutive year, 2020. The plot above indicates to us that the average income for 2020 is €48,471.14. The 80% confidence intervals are between €46,994.85 and €49947.43 whereas the 95% confidence intervals lie between €46,213.35 and €50,728.93.

There are many relationships that exist between the trends discussed previously.

Is there a direct relationship between the level of Garda personnel and the number of offences for each region in Dublin?

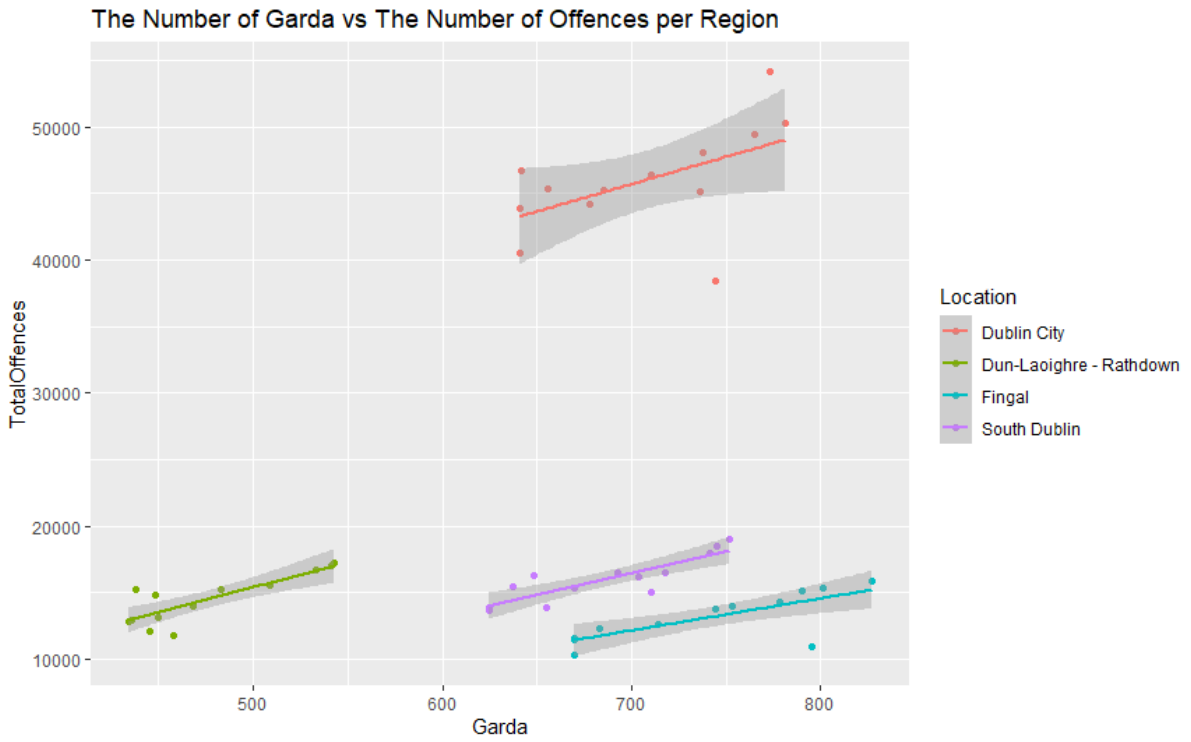


Figure 11 - The Number of Garda vs The Number of Offences per Region

The plot above demonstrates the relationship that exists between the level of garda personnel and the number of offences committed for each location. For each region, as the level of garda personnel increases, the number of offences committed increases or perhaps because of each region's high number of offences, they keep increasing garda personnel. As a society, it is naturally asummed that the more garda there are employed in a specific area, the less offences will be committed. However, generally, this is not the case for the different districts of Dublin. There is a positive linear relationship between the level of garda personnel and the number of offences committed for each.

Is there a link between the crime rate and the number of garda personnel in each region of Dublin?

Crime Rate is measured by the total amount of offences divided by the population. Crime Rate was calculated for each region of Dublin. As we can see from the plot below, there is a steady increase in the crime rate for each region as the number of garda increases. Again, we see a result that does not meet society's expectations.

Fingal and Dublin City regions have roughly the same amount of garda employed for each year, however the crime rate of Dublin City is nearly twice the crime rate of Fingal. As Crime Rate takes in the size of population within its calculations, we can determine that the Population of an area plays a detrimental factor in defining the crime rate.

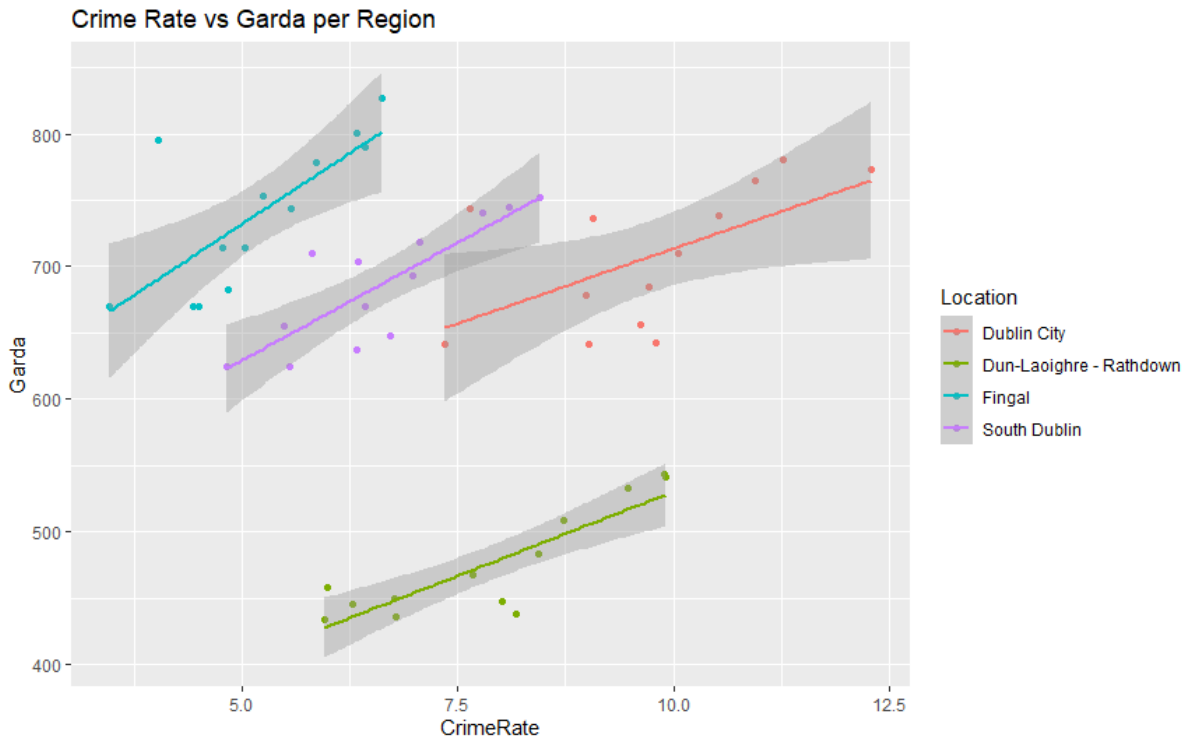


Figure 12 - The Crime Rate vs The Number of Garda per Region

What relationship exists between the education rate and the crime rate per region in Dublin?

Education Rate is measured by the number of people in second level education divided by the population within its region.

Looking at the plot below, we can see that for Dublin City and Dun Laoghaire – Rathdown, the rise in education rate causes a rise in the crime rate. Naturally, we would assume that the more educated a region is, the less chance there would be of crimes to be committed. Could this be due to white collar crimes?

For the region of South Dublin, we can see a miniscule increase in crime rate as its education rate increases which again could suggest that people who are within second level education are committing crimes in the streets of South Dublin.

The region of Fingal follows the assumption that is, an increase in education rates allows for a decrease in the crime rate.

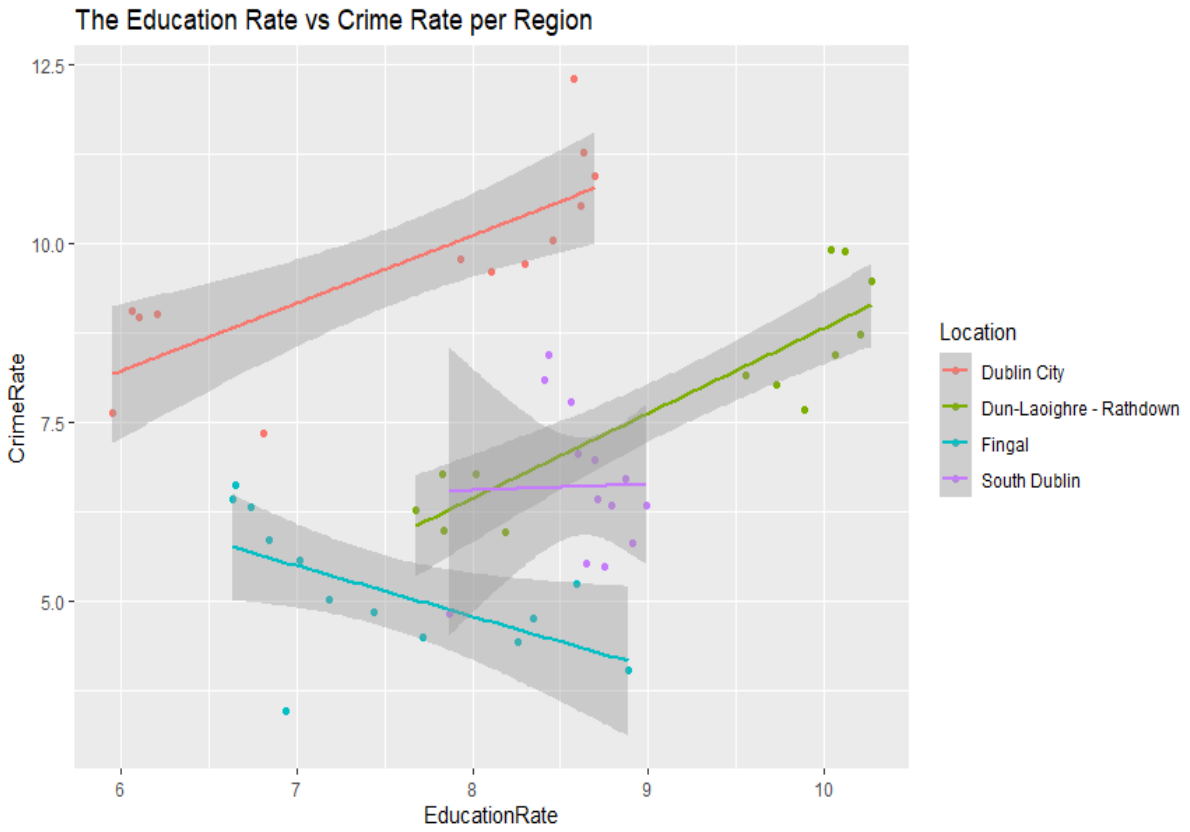


Figure 13 - The Crime Rate vs The Education Rate per Region

Does the number of offences increase in regions of lower education attainment?

As we can see, the region with the lowest second level education attainment rate is Dublin City followed by Fingal. We can see for Dublin City, that generally, the number of offences increases as the education attainment rate increases suggesting that the number of offences increases in areas of lower education attainment.

Fingal has a slightly increased education attainment rate to Dublin City however as we can see the number of offences for Fingal declines as the education rate increases, which shows that the number of offences decreases as the level of education attainment rate increases.

South Dublin and Dun Laoghaire – Rathdown have relatively high education rates, however it is worth mentioning that for the Dun Laoghaire – Rathdown region, as the level of education attainment increases, the number of offences increases also.

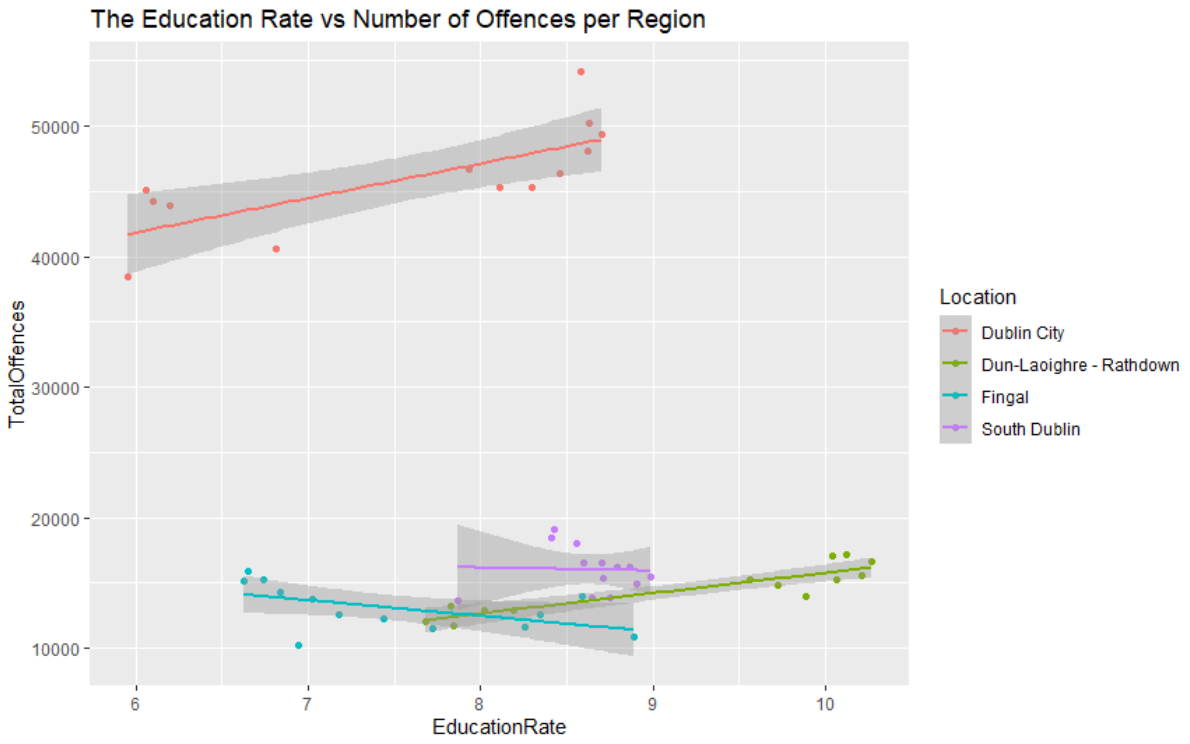


Figure 14 - The Education Rate vs The Number of Offences per Region

Is there a relationship between the number of offences in each region of Dublin and the average income of Dublin?

Using the average income of the whole of Dublin, it is interesting to see if it has an effect on the number of offences committed per region. As we can see from the plot below, as the income levels increase, the number of offences decreases slightly for each region.

In terms of average income, it was assumed that a much stronger result for the regions of Dun Laoghaire – Rathdown, Fingal and South Dublin would be found. For those regions we can see that there is not a strong relationship but quite a weak one between the average income and the number of offences committed.

From the plot below, it may be assumed that the rise in income causes a decrease in the total number of offences committed.

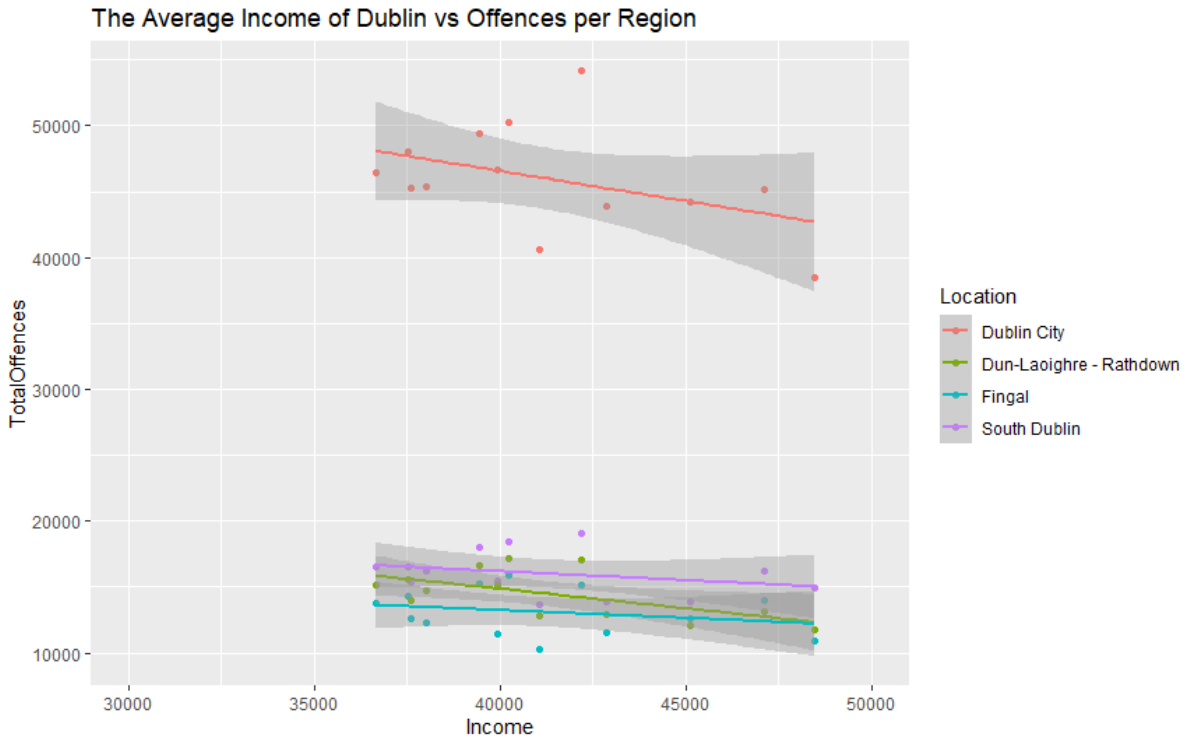


Figure 15 - The Average Income of Dublin vs Offences per Region

K-means Clustering

The image below shows the k-means cluster analysis of the income data values from 2008 – 2020. It can be observed that the data points 5 – 14 are assigned to cluster 2 whereas data points 15-17 are assigned to cluster 1. The graph below suggests how many of the income values found are of similar value between the two clusters. Notice how within each cluster there is a centre point, each data point is assigned to the centre based on their Euclidean distance.

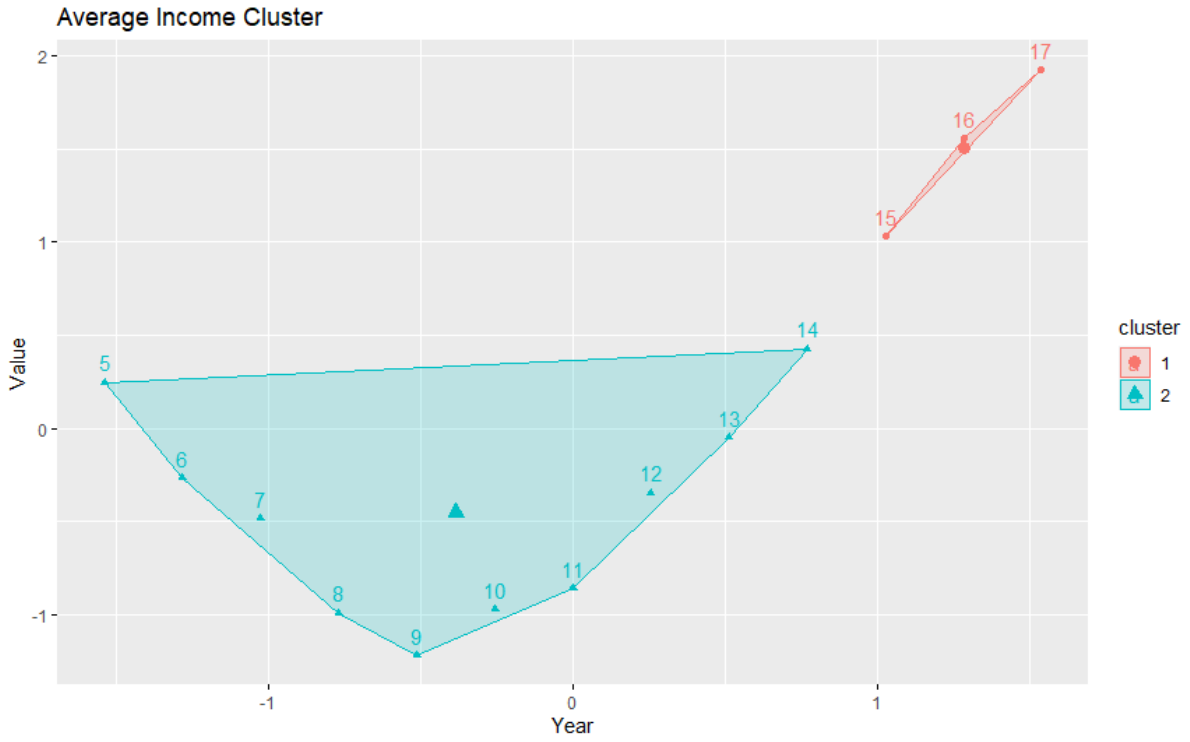


Figure 16 - K-means Clustering on Income

6.0 Conclusions

Diverse factors affecting crime levels were analysed for the Dublin area. This project highlights that there are many factors that can be seen as a gateway to crimes being committed. It can be seen as an advantage as it shows how factors such as education and level of garda personnel have had little effect in some areas across Dublin. It also shows how income levels are the most detrimental factor to reduce crime. By reading and examining this project, external representatives on behalf of An Garda Siochana can try to tackle the issue of crime throughout Dublin. Perhaps they could put new initiatives in place such as shifting their workforce from low criminality regions to the ones with high levels

Another contribution of this research is that it may be seen as an eye opener on the problem of crime throughout Dublin. There are many intriguing results found in this project which may also be of use to external departments such as the Department of Education. With education as a factor that can affect crime levels, perhaps the Department of Education could introduce new schemes to promote second level education for the Dublin citizens.

There are also disadvantages and limitations within this project. The main problem faced within this project was finding the necessary data needed to carry out this investigation. It was difficult to find datasets that all aligned with the same timeline. For example, the income dataset used in this project has a timeline of 2004 – 2019 whereas all of the other datasets used have a timeline of 2008 – 2020. Another limitation within this project was finding data to measure the accuracy of our prediction. The Central Statistics Office have yet to release the crime statistics for 2021 alongside the second level education and garda personnel statistics.

Overall, this project compares how multiple factors can influence the level of crime throughout Dublin. It compares these factors against the four districts of Dublin. There are different results for different areas which also made this project continuously intriguing.

7.0 Further Development or Research

With additional time and resources, this project can continue to analyse how there are numerous trends/factors that can influence the level of crime throughout Dublin. If there was more time available for this project, it would be beneficial to perhaps find another dataset that can complement this investigation. For example, implementing a homelessness/poverty dataset would show whether and to what extent the number in extreme poverty/homelessness causes a shift in the number of crimes committed throughout Dublin.

A different approach that could be taken is using a different type of data mining technique for predictive analytics. One approach that could be taken, is the use of regression models. Regression models could be used to make a prediction on variables such as the number of offences and the level of garda personnel.

8.0 References

References

Biswal, A., 2021. *A Guide to Time Series Forecasting in R You Should Know.* [Online]
Available at: https://www.simplilearn.com/tutorials/data-science-tutorial/time-series-forecasting-in-r
[Accessed 23 November 2021].

https://data.cso.ie/table/CJA07

https://data.cso.ie/table/ED133

https://data.cso.ie/table/SIA13

https://data.cso.ie/table/E2001

https://worldpopulationreview.com/world-cities/dublin-population

9.0 Appendices

9.1. Project Proposal

Objectives

My project is “Analysing the Causes of Crime in Dublin”. This project sets out to achieve how multiple variables can be examined and analysed to measure the causes of crime within the Dublin region. My aim is to perform a thorough analysis on each data set to allow for insights to be made. I hope to be able to show audiences, how many external factors can be seen as a gateway to crime. These factors include but are not limited to unemployment, education attainment, poverty and more. I hope to use the factors I have named but also use order factors that I find during further research. I hope that this project will be able to highlight how these factors can I have such a negative effect on society and individuals. Crime is a huge issue within the Dublin region not only to society but to individuals who are performing these crimes. I may also compare the crime rates between the northern region of Dublin and the southern region of Dublin to highlight if there are any similarities and differences between the two regions in terms of crime.

Background

I chose to undertake this project because I have seen the effects of crime firsthand in the Dublin area. I also have friends and family who have been victims of crime in the area. I want to be able to find many of the root causes of crime and figure out why high levels of crime is occurring in our area. One of the main aims in my project is to promote and raise awareness of how there are multiple factors that are used as a catalyst towards crime. In my opinion there are many factors that lead to a gateway towards crime. I also want to be able to highlight how young people in Dublin can commit crimes and what leads young people to commit them. For example, I'm going to try to gather child attainment data within a secondary school and show how this can lead to crime. In order to meet my objectives that I've set out in section 1.0, it is important for me to gather the datasets needed to examine and analyze. I hope to gather datasets regarding employment, poverty, crime rate, and child attainment in school. My plan also is to compare these factors with the Northern Dublin region and the Southern Dublin region and note the similarities as well as the differences between them both.

State of the Art

In the past there have been others who have conducted similar analysis on what I plan to analyze in the future. For example, there are numerous diverse types of projects affiliated with crime that pupils in the National College of Ireland have carried out. Although some of my analysis may be similar, I do strongly believe that my work will stand out on its own. I plan to use different datasets that my past peers have not used. Of course, some of the datasets that I use in my analysis may have been used in the past by pupils but there are also datasets that I feel have not been used in the past. What I also believe will make my work stand out is the actual analysis itself. I plan to use a different type of analysis that has not been done in the past to my knowledge. I hope that this will give me a competitive advantage in comparison to previous projects that been carried out in the past.

Data

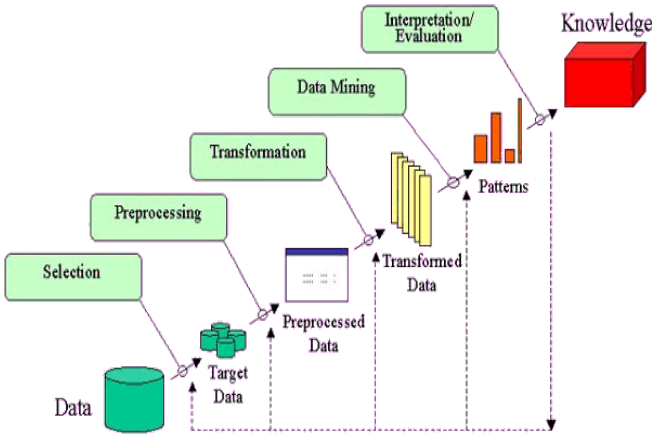
This project will consist of a lot of data and multiple different datasets. As my project covers the topic of crime, it is essential for me to gather the necessary crime data before I perform any analysis. Recently I have gathered a data set that contains the necessary crime data of Dublin, dating as far back as 2003 to present. The data is split up by quarterly and consists of columns such as “Garda Division”, “Type of Offence” and “Number of Occurrences”. I believe that this dataset will be the “backbone” dataset of my project. This is quite a large dataset consisting of 5255 rows of data. In my opinion, it is good that this is such a large data set as it means there is lots of information within the data. This will also allow me to do much more with the data. I also have numerous other datasets that I feel relate to crime in Dublin. These datasets vary in size but are still over 100 rows long. Other datasets that I have gathered include the employment rate within the Dublin region, the number of Gardai employed within the Dublin region, children referred to the Garda Diversion Program, poverty in Dublin and more. I believe that these social economic factors will assist my project in analyzing how they contribute to crime. Although I may not use each and every dataset that I come across, I feel that they could be useful to keep on file in case I do decide to use some. So far, every data set I have gathered has been gathered from the Central Statistics Office. The Central Statistics Office is Ireland’s national statistical office which contains multiple statistics of the people of Ireland. The datasets stored and provided by the Central Statistics Office are known as public datasets which means that members of the public have free access to it. I'm currently going through the process of getting myself more familiar with the data that I'm planning on using. This involves performing some basic descriptive statistics on the data such as the mean, median etc. Whilst getting myself more familiar with the data, I am ensuring that the data is clean and able to be used. This is a crucial part of the data preprocessing process and is essential to be carried out.

Methodology and Analysis

The methodology that I will follow for my final year project is the KDD methodology. The KDD methodology stands for Knowledge Discovery in Databases. The KDD process is used to extract knowledge from large databases. It uses Data Mining algorithms to find what's deemed knowledge. The concept of Knowledge Discovery in Databases is a programming approach that involves the study and modelling of vast data sets. This procedure aims at identifying useful and abnormal patterns in the data. The KDD methodology contains multiple steps to be carried out. These steps include:

1) Domain and understanding of KDD Goals.
2) Choosing data that investigation will be carried.
3) Pre-processing and cleaning data
4) Data transformation
5) Predict and description
6) Selecting a data mining method
7) Implementation of data mining
8) Evaluation

Each step in this step-by-step methodology needs to be followed accurately in order to determine the best results for my project.



https://external-content.duckduckgo.com/iu/?u=http%3A%2F%2Fwww2.cs.uregina.ca%2F~dbd%2Fcs831%2Fnotes%2Fkdd%2Fkdd.gif&f=1&nofb=1

I plan to break down my project in a weekly basis. Each week, I hope to be able to progress with my project as the weeks go on. My academic supervisor and I plan to meet on a weekly basis also, this meeting allows me to understand what I need to do next for the following week. I feel as though these regular meeting will allow me to stay focused and keep on track of my project. Each week I will set short term goals for my project and will plan to reach these goals by the week. I am a firm believer in continuous goal setting and believe it allows us to stay motivated and on track. At the midpoint of my project, I will re-evaluate this process.

Technical Development

There is a lot of technical development involved in my project. Firstly, the software language where a lot of my analysis will be carried out is known as R. This is my first year ever using R so as you can imagine there is a lot of continuous learning as well as self-learning involved within this project. I currently am studying a module known as Data Application Development. This is the module where we learn R and Python and how we can use these software languages to analyze datasets and interpret results. As I mentioned before, I hope to be able to develop many skills within the data analysis area from this module and implement these skills into my final year project.

Other technical development under consideration for this project is the use of predictive analysis. At the minute this is under consideration as perhaps I could use prediction analysis to predict the crime rates in the Dublin region for the next two years etc. Predictive analysis is a type of data analysis that focuses on coming up with predictions based on the data collected. It is very useful for analysing various types of data. It uses various statistical models and techniques to allow to predict the future. There are a variety of data analysis models that belong to the category of predictive analysis. Almost all

of these are regression models, which means they try to identify the relationship between two or more variables. By identifying the relationship between these variables, they can help predict the value of an unknown variable because the value of a known variable (such as time) changes. Some of these include:

- Linear Regression – This allows us to assume an unknown value of a variable based off a known value of a different variable.
- Random Forests – These are used to model regression by the use of decision trees based off multiple datasets consisting of large amounts of data.

I hope to gain more familiarity with the two models mentioned above to allow for me to make predictions on the data gathered for my project.

Project Plan

Task Name	Assigned To	Start Date	End Date	Duration
Creating a Project Idea	BC	27/09/2021	20/10/2021	30 days
Finding relevant data to be used for my project	BC	27/09/2021	16/11/2021	76 days
Meet with academic Supervisor	BC/GE	05/11/2021	05/11/2021	< 1 day
Discuss project idea and implementation with academic supervisor	BC/GE	05/11/2021	05/11/2021	< 1 day
Identify what tasks need to be carried out for next week	BC	05/11/2021	05/11/2021	< 1 day
Gain more familiarity with datasets gathered	BC	05/11/2021	12/05/2021	7 days
Perform some minor descriptive statistics with data	BC	05/11/2021	12/05/2021	7 days
Find regions that Garda stations cover in Dublin	BC	05/11/2021	12/05/2021	7 days
Present visuals of data so far in Excel	BC	05/11/2021	12/05/2021	7 days
Research knowledge of predictive analysis	BC	05/11/2021	12/05/2021	7 days
Upload Ethics and Project Proposal	BC	07/11/21	07/11/21	< 1 day
Meet with academic supervisor	BC	12/11/21	12/11/21	< 1 day
Re-format Data	BC	01/12/2021	01/12/2021	5 days
Find total offences North and South	BC	04/12/2021	04/12/2021	< 1 day
Plot total offences North and South	BC	04/12/2021	04/12/2021	<1 day
Meet with academic supervisor	BC/GE	10/12/2021	10/12/2021	<1 day
Find predicted values of offences	BC	13/12/2021	14/12/2021	2 days
Plot predicted values	BC	15/12/2021	15/12/2021	<1 day
Mid-Point Report	BC	15/12/2021	22/12/2021	7 days
Gain more familiarity with new datasets to be added	BC	15/12/2021	15/01/2022	1 month
Show how trends can effect # of crimes committed such as garda employment	BC	01/02/2022	01/02/2022	1 day
Analyse new datasets and find a method to add them to current data	BC	17/02/2022	21/02/2022	1 week

Use predictive analysis to analyse values in new datasets – regression techniques	BC	24/02/2022	28/02/2022	1 week

9.2. Ethics Approval Application (only if required)

9.3. Reflective Journals

October:

Reflective Journal – October

Supervision & Reflection Template

Student Name	Brandon Caffrey
Student Number	X18514113
Course	BSHCDA4

Month: October

What? Reflect on what has happened in your project this month? Over the month of October, I was introduced to the module "Software Project" which will be the module that covers all aspects of my final year project. We were introduced to our lecturer of this module, Frances Sheridan. In week one, we had an introductory lecture. The agenda for this introductory module was to explain to use the main components of our final year project such as the milestones throughout the year, the marking scheme, and the rules for the project. Frances took us through the major milestones throughout the year with regard to submissions. We have quite a large marking scheme which allows us to be accessed throughout the whole year rather than one submission at the end of the year. We were provided with great advice on how to keep on top of our submissions during a busy year for us all. We were also given an exercise to help us create our monthly journals. I found this exercise very helpful with regard to reflection and how we should reflect on the month prior to our written journal on it. This month we also uploaded our project pitch video. This three-minute video consisted of recording ourselves outlining our project and discussing why we want to go ahead with our project idea. Prior to this recording, I conducted some research on my project idea and searched for datasets that I think will be helpful for myself and my project. I gathered this data through the Central Statistics Office (CSO). I discussed how these datasets will help me with my investigation and analysis. As this is a "Project Pitch Video" we have to wait and see weather our idea is suitable in terms of effort and difficulty. This will be judged by our academic supervisors.
So What? Consider what that meant for your project progress. What were your successes? What challenges still remain? At the moment I am still waiting for my project idea to be approved by my academic supervisor. Once this is approved, I can then arrange a meeting with my academic supervisor to discuss what

to do next with my idea. I have also gathered the datasets in which I hope to use for my project. However, these datasets may change as I progress with my project. My successes so far have been coming up with an idea for my project, identifying multiple datasets that I may use for my project and creating and uploading my “Project Pitch Video”. The challenges that remain still are meeting with my academic supervisor, identifying more datasets that I could use for my project and getting familiar with the data itself. We recently had a workshop of discovering data and how we can “play around” with data before we actually analyze it. It is extremely important for us to get familiar with our data and our data sets before we implement any analyzation with the data.	
Now What? What can you do to address outstanding challenges? The next step for me is to become familiar with the data that I have at the minute. This may take time, but it is critical that I do it. I must also research and see if there is more data that would be beneficial for my project. As well as this, I have to arrange a meeting with my academic supervisor so we can discuss what I have already and what I can do in the future regarding my project. I am also doing further research on how I can analyze the data and find relationships between the data. I hope to use the skills that I'll acquire from my Data Application Development module to assist me with this. In this module we use the language R to analyze data. I hope to complete these steps by the end of the month.	
Student Signature	*Brandon Caffrey*

November:

Reflective Journal – November

Supervision & Reflection Template

Student Name	Brandon Caffrey
Student Number	X18514113
Course	BSHCDA4

Month: November

What? Reflect on what has happened in your project this month? Over the month of November, I continued to attend the "Software Project" module. Over the weeks I attended useful workshops that will assist me with my project. These workshops included "Finding Good Data" by Michael Bradford and "Report and Referencing" by a member of staff for the NCI library. These workshops provided insights on how I can find relevant data related to my project as well as conducting my reports in the correct and concise manner. I hope to use the insights I have gained from these workshops in my future reports and analysis. In the month of November, my project was approved by my project supervisor. Once my project was approved, I met up with my project supervisor to have further discussions on my project. We discussed where the project can go and what will be needed to ensure my project is successful. We met up face to face for the first couple of weeks however due to covid cases rising we now meet up online each week. In terms of technical work, I am in the process of importing my datasets into R studio. I still have some work to find another complimentary dataset that I can combine with my current one. At the moment, I am working on my current one and getting more familiar with it.
So What? Consider what that meant for your project progress. What were your successes? What challenges still remain? At the moment, I am in the process of analysing one of my datasets in my project. It is going smoothly at the minute in terms of results etc. I have found some insights already however there is still a lot of work to be done to the data. My reports and submissions are up to date so far which is good as it allows me to work on my mid-point presentation which is due in the month of December.

The challenges that still remain is finding another complimentary dataset that I can use in my project. This will hopefully be completed by the first week of December.
Now What? What can you do to address outstanding challenges? The next move for me is to continue working on the dataset I have and to try complete as much work as possible before the mid-point presentation. Also, I will need to prepare my mid-point presentation as early as possible in order to have some time left over if I need to make any changes. I will also continue to meet with my project supervisor.

Student Signature	*Brandon Caffrey*

December:

Reflective Journal – December

Supervision & Reflection Template

Student Name	Brandon Caffrey
Student Number	X18514113
Course	BSHCDA4

Month: December

What? Reflect on what has happened in your project this month? Over the month of December, I continued to attend the "Software Project" module. As my mid-point presentation is fast approaching, I have been working towards getting everything prepared for it. With regard to my project, I feel as though I am making good progression with regard to my analysis. I have continued to implement the auto. arima function to my data which allows me to forecast a prediction for 2021. I have also applied exploratory analysis to my data whereby I have found the total amount of crimes committed in Ireland. Another dataset that I have added to this analysis, is the garda employment dataset. In order to join this with my current offences' dataset, I found the average number of garda per location and added these values onto my offences dataset under a new column called 'garda', This variable shows the average amount of garda per year for each location.
So What? Consider what that meant for your project progress. What were your successes? What challenges still remain? My successes this month was finding more data that I can use towards my project. A challenge that still remains is exploring the relationships between the number of offences and the number of garda.
Now What?

What can you do to address outstanding challenges? To address the outstanding issue, it is important for me to examine the number of offences committed vs the number of garda employed and try to find if there are any patterns between the two. I will also continue to meet up with my supervisor and submit my mid-point presentation	
Student Signature	*Brandon Caffrey*

January:

Reflective Journal – January

Supervision & Reflection Template

Student Name	Brandon Caffrey
Student Number	X18514113
Course	BSHCDA4

Month: January

What? Reflect on what has happened in your project this month? Over the month of January, I have continued to work on my project throughout the break. I submitted my mid-point presentation also. I learned a lot from my mid-point presentation especially in terms of my presentation skills. I was quite shocked to see how much goes into presenting our project. It can be quite nerve wrecking even if you do not have an audience. I also think that it can be difficult to fit every detail into such a short presentation. I hope to be able to take the experience gained from my mid-point presentation into my final presentation in May. Also, this month, I am continuously meeting with my supervisor virtually on a weekly basis. Meeting on a weekly basis is beneficial and it allows me to show my supervisor the progress I have made each week whilst taking in his feedback on it. I have also been looking for more datasets that I can incorporate into this project such as education data and income data.
So What? Consider what that meant for your project progress. What were your successes? What challenges still remain? My successes this month was learning from my mid-point presentation and taking this experience into my final presentation in May. I have also researched on exercises that can assist with a presentation such as reading my presentation out aloud many times before starting to record it.

The challenges that still remain is identifying further datasets that I can implement into this project.
Now What? What can you do to address outstanding challenges? To address the outstanding issue, it is essential for me to search on the Central Statistics Office or elsewhere for data that can be beneficial for this project such as education attainment data and income data.
Student Signature | *Brandon Caffrey*

February:

Reflective Journal – February

Supervision & Reflection Template

Student Name	Brandon Caffrey
Student Number	X18514113
Course	BSHCDA4

Month: February

What?
Reflect on what has happened in your project this month?
Over the month of February, I have continued working on my project. My project so far consists of the total number of offences from 2008-2020 as well as a predicted values for the total offences for 2021 and 2022 for the north and south of Dublin. This time series analysis was forecasted using the auto. arima function. I continue to use this function to make other predictions with my other data. This month, I have started working on new data that I wish to incorporate into my project. This new data is the number of guards employed within Dublin.
In order to make this data useful for my project I must be able to find the number of guards employed in north and south Dublin for the years 2008 – 2020. This requires some data cleaning. To find the yearly figures of Garda employed in north and south Dublin, I use the aggregate function to find the mean yearly figure. The aggregate function calculates the mean figure for each year within my offences_nDublin data and my offences_sDublin data. The code for this function is:

- totalGarda_nDublin_yearly <- aggregate (offences_nDublin$`No ofGardai`, by = list(offences_nDublin$Year), FUN = mean)
- totalGarda_sDublin_yearly <- aggregate (offences_sDublin$`No ofGardai`, by = list(offences_sDublin$Year), FUN = mean)

These return a list that contain the average amount of guards for north and south Dublin. I then applied some feature engineering and created a column for both lists called ‘Location’. The values

within ‘Location’ were either north or south. I then join these by rows using ‘rbind’ to create ‘averageGarda’
In order to create my time series object for the north side of garda and the south side of garda, I joined my ‘averageGarda’ and ‘TotalOffences’ data frames. Once this was complete, I was able to perform my time series forecast using the ARIMA model. The results from my time series forecast were successful.

So What?
Consider what that meant for your project progress. What were your successes? What challenges still remain?
In terms of my project progress this is a good start since my break over January. I was successful in what I planned out to do. I was successful in getting familiar with this new data and was able to perform some time series forecasting using the ARIMA model. I was also able to plot these forecasts so the viewers can get a better graphical understanding of the prediction. The challenges that remain ahead is finding new data to make more predictions and being able to implement this into the data story that I am trying to tell.

Now What?
What can you do to address outstanding challenges?
My next objective for this project is to find data that is relevant to education attainment in order to show how education attainment can have an effect on crime and the number of crime offences. I plan to use this data and again apply the ARIMA model in order to make a prediction on the number of people within our education system for both north and south Dublin.

Student Signature	*Brandon Caffrey*

March:

Reflective Journal – March

Supervision & Reflection Template

Student Name	Brandon Caffrey
Student Number	X18514113
Course	BSHCDA4

Month: March

What?
Reflect on what has happened in your project this month?
Over the month of March, I continue to make progress on my final year project. I was able to find more data that I feel would be useful for my project. This dataset is called 'pupils in secondary school 2008 – 2020'. It contains data showing the number of pupils in secondary school across Dublin.
In order to use this data to make a prediction on the total amount in secondary school education, I had to get the total yearly values of pupils in secondary school education for 2008 – 2020. This was performed by using the aggregate function. The aggregate function calculates the sum of pupils in secondary school education for each year. The code for this function is below:

- total_educated_north_yearly <- aggregate (northDub$Value, by = list(northDub$Year), FUN = sum)

This line of code returns a list with the total values of pupils in secondary school education during the years 2008 – 2020.
Once I have gathered this list, I then used it to forecast values for 2021, 2022 and 2023. This is a prediction of the total pupils in secondary school education for those years using the ARIMA model.

So What? Consider what that meant for your project progress. What were your successes? What challenges still remain? I was successful in what I planned out to do. I was successful in getting familiar with this new data and was able to perform some time series forecasting using the ARIMA model. I was also able to plot these forecasts so the viewers can get a better graphical understanding of the prediction. The challenges that remain ahead is trying to find another machine learning model that would fit into this project. At the moment, I am thinking of applying clustering to one of my already gathered datasets. We have learned clustering in one of our other modules and I feel it could add to this project.
Now What? What can you do to address outstanding challenges? My next objective for this project is to find another machine learning model that I can use for this project. As mentioned before, I wish to apply a clustering model to show the total number of offences for the different areas across Dublin

Student Signature	*Brandon Caffrey*

April:

Reflective Journal – April

Supervision & Reflection Template

Student Name	Brandon Caffrey
Student Number	X18514113
Course	BSHCDA4

Month: April

What?
Reflect on what has happened in your project this month?
Over the month of April, there have been many changes to my project. Rather than comparing north vs south Dublin, I am know going to be comparing between the four districts of Dublin. These are:

- Fingal
- Dublin City
- South Dublin
- Dun Laoghaire - Rathdown

I have reformatted all previous work that I have done before to fit into this new approach with the four districts of Dublin. I have also found an income dataset where I have made a prediction on the income level of Dublin.
I have nearly completed my project and have therefore found many results with my project. My project shows how multiple trends such as level of garda personnel, education attainment and income affect the number of crimes committed in throughout each region/district of Dublin. My project also uses predictive analytics on each trend using the ARIMA model. The time series forecast is made for the next consecutive year from when the data timeline stops.

My results are plotted within R studio.

So What?

Consider what that meant for your project progress. What were your successes? What challenges still remain?

My success in April was finalising my project in terms of code. The challenge that still remains is creating my project report. I hope to have this completed a day before the submission so I can read over it etc. It is important for me to have my results formatted neatly and explained well. Another challenge that still remains is creating my showcase poster. My showcase poster will be created using Canva and will outline my project objective etc

Now What?

What can you do to address outstanding challenges?

To address the outstanding issues, it is important for me to get started on my report as soon as possible. I think getting a section done each day will be sufficient time to complete it before the deadline. This also gives me time to proofread the report before the submission.

To address my poster issue, it is vital for me to set up a Canva account and start creating a nice poster. I want my poster to be visually appealing but also want to have vital information about my project on it. I will also look at previous posters for inspiration on how I should create mine.

Student Signature	*Brandon Caffrey*

9.4. Invention Disclosure Form (Remove if not completed)

Please fill in the following sections, if you think your idea is innovative:

1. Title of Invention

2. Inventors

Name	School/Research Institute	Affiliation with Institute (i.e., department, student, staff, visitor)	Address, contact phone no., e-mail	% Contribution to the Invention

3. Contribution to the Invention

Each contributor/potential inventor should write a paragraph relating to his/her contribution and include a signature and date at the end of the paragraph.

4. Description of Invention

(Please highlight the novelty/patentable aspect. Attach extra sheets if necessary, including diagrams where appropriate). What is novel, the 'inventive step'? For more information on patents, please look at http://www.patentsoffice.ie/en/patents.aspx

5. Why is this invention more advantageous than present technology?

What is its novel or unusual features? What problems does it solve? What are the problems associated with these technologies, products or processes? Explain how this invention overcomes these problems (*i.e.*, what are its advantages).

6. What is the current stage of development / testing of the invention?

7. List the names of companies which you think would be interested in using, developing or marketing this invention

8. Funding Partner(s)

Government Agency & Department	
% Support	
Contract/Grant No.	
Contact Name	
Phone No.	
Address	

Industry or another Sponsor	
% Support	
Contract/Grant No.	
Contact Name	
Phone No.	
Address	

9. Where was the research carried out?

10. What is the potential commercial application of this invention?

11. Was there transfer of any materials/information to or from other institutions regarding this invention?

If so, please give details and provide signed agreements where relevant.

12. Have any third parties any rights to this invention?

If yes, give names and addresses and a brief explanation of involvement.

13. Are there any existing or planned disclosures regarding this invention?

Please give details.

14. Has any patent application been made? Yes/No

If yes, give date: _____ Application No.: _____

Name of patent agent: _____

Please supply copy of specification.

15. Is a model or prototype available? Has the invention been demonstrated practically?

I/we acknowledge that I/we have read, understood and agree with this form and the Institute's *Intellectual Property and Procedures* and that all the information provided in this disclosure is complete and correct.

I/we shall take all reasonable precautions to protect the integrity and confidentiality of the IP in question.

Inventor: _____ _____

Signature Date

_____

Signature

9.5. Other materials used

Any other reference material used in the project for example evaluation surveys etc.