2017

Analysis on Automated External Defibrillator (AED) Distribution in Leinster, Ireland

BSC (HONOURS) IN COMPUTING 2016/2017

CHRISTOPHER DORAN
DATA ANALYTICS
X12724145

National College of Ireland

Declaration Cover Sheet for Project Submission

SECTION 1 *Student to complete*

Name: Christopher Doran
Student ID:x12724145
Supervisor: Michael Bradford

SECTION 2 Confirmation of Authorship

The acceptance of your work is subject to your signature on the following declaration:

I confirm that I have read the College statement on plagiarism (summarised overleaf and printed in full in the Student Handbook) and that the work I have submitted for assessment is entirely my own work.

Signature: **Christopher Doran** Date: **10/5/2017**

Table of Contents

Executive Summary

This project was compiled for the BSc (Honours) in Computing program, at the National College of Ireland. The purpose of this project is to analyse Automated External Defibrillator (AED) distribution in every village/town across Leinster, Ireland. The result of the project identifies high priority areas in Leinster that with the inclusion of an easily accessible AED could help reduce mortality rates. High priority areas are identified by analysing the characteristics of each area that include Population, the breakdown of the Population, the distance to the nearest emergency service station in kilometres, the number of heart related deaths and the number of existing AED's in the area.

The system for the statistical analysis consists of a database for storing and analysing real data retrieved from the National Ambulance Service (NAS), Dublin Fire Brigade (DFB) Helpinghearts.ie and Central Statistics Office (CSO). The data was analysed using R programming language in RStudio. Finally, the system provides the results of the analysis through Visualisations and Representations for the end-user to consume.

1 Introduction

1.1 Background

Emergency Service Response times in Ireland have always been a hot topic politically and socially especially when fatalities occur. Consuming the content various news outlets produced relating to the subject and subsequently reading about people suffering from sudden arrhythmic death syndrome and more specifically young sports athletes suffering from cardiac arrest and unfortunately emergency services not responding in time to administer early defibrillation the interest for this project was born. Recognizing that AEDs are very expensive and cannot be on every street corner, there is still a significant demand that AED's should be distributed adequately across Ireland that have an emergency service response time greater than eight minutes. AED Ireland [1] outlines that early defibrillation to cardiac arrest sufferer's saves lives and involves trained personnel delivering an electric shock to the patient's chest to help reinstate the normal function of the heart. It is the link in "The Chain of Survival" that increases survival rates. The ability to possess a defibrillator on site within two minutes provides an 80% chance of survival. Every subsequent minute survival rates decrease by 7-10%. Further studies show when early defibrillation that is administered within three minutes there is a 74% chance of survival. Furthermore, if time to defibrillation is greater than ten minutes there is a 0% chance of survival without cardiopulmonary resuscitation (CPR) although when CPR is used survival rates increase to between 10% and 20%. Considering the above statistics, the National Ambulance Service response times in Ireland are very topical and in October 2015, the HSE published a report named the "National Ambulance Service of Ireland, Emergency Service Baseline and Capacity Review". The report commissioned by the Health Service Executive and carried out by Lightfoot Solutions UK [8]. The report found that 26.6 % of ambulances fail to reach the incident within eight minutes of all life threatening calls including Echo calls (Life threatening cardiac or respiratory arrest).

Table 1: ECHO/DELTA Current and Future Response Times (2014) [7].

	8 minute ECHO/DELTA				19 minute ECHO/DELTA			
	Major Urban	**Minor Urban**	**Rural**	**Total**	**Major Urban**	**Minor Urban**	**Rural**	**Total**
Current	36.7%	45.6%	6.6%	**26.6%**	87.0%	76.1%	44.0%	**67.2%**
Future	85%	73.2%	32.9%	**60.6%**	95.5%	84.1%	69.3%	**81.5%**

When considering having a defibrillator on site within two minutes, there can be an 80% chance of survival and as of 2014, only 6.6% of the population in rural Ireland possess a response time within eight minutes. Based on the above statistics shown on Table 1 and there is a necessity for a different approach regarding the availability and or distribution of AED's in Ireland.

National Ambulance Service

The National Ambulance Service is the statutory public ambulance service in Ireland. The service operates under the National Hospitals Office of the Health Service Executive (HSE). Up until the establishment of the HSE by the Health Act in 2004, ten regional Health Boards ran Ireland's Ambulance Service. Subsequently the establishment of the NAS as part of the HSE in 2005 occurred consisting of ninety-four Ambulance Stations and a staff of 1400 [11].

Dublin Fire Brigade

Dublin Fire Brigade is Ireland's largest full-time brigade and employs over 1000 people. There are fourteen stations, twelve of which are full-time and two are part-time (retained). DFB operates an Emergency Ambulance Service as well as a Fire, rescue and emergency service for Dublin City and County [5].

Helping Hearts

Helping Hearts founded in 2001, provides CPR and First AID Training Courses to the public across Ireland. The Helping Hearts websites contains a map of AED locations across Ireland. Helping Hearts offices are in Galway [6].

Pre-Hospital Emergency Care Key Performance Indicators

The Health Information and Quality Authority (HIQA) published Pre-Hospital Emergency Care Key Performance Indicators on Emergency Response times in January 2011. One of the Key Performance Indicators (KPIs) being recommended by the HIQA was that Echo calls (Life threatening cardiac or respiratory arrest) are responded to within 8 minutes for 75% of all cases [9].

HSE's Performance Assurance Report (PAR)

The HSE's Performance Assurance Report (PAR) provides an overall analysis of key performance data from Divisions, such as Acute, Mental Health, Social Care, Primary Care, Health and Wellbeing as well as Finance and HR. The activity data reported consists of Performance Activity and Key Performance Indicators outlined in the current National Service Plan. The report was collated by the Planning and Business Information Unit (PBI), [14].

National Ambulance Service of Ireland, Emergency Service Baseline and Capacity Review

The National Ambulance Service of Ireland, Emergency Service Baseline and Capacity Review published by the HSE and commissioned by the Health Service Executive and carried out by Lightfoot Solutions UK was created to determine the underlying capacity required to deliver the standards for response performance across Ireland [8].

1.2 Aims

The overall objective of the project is to develop a system that analysis Automated External Defibrillators (AED) distribution throughout Leinster, Ireland. Identifying high priority areas in need of easily accessible AEDs to help decrease mortality rates is a key end goal. The following is a list of specific aims that will help achieve the overall objective:

Aim 1: The first project aim is to target data that would help the main objective of this project fulfilled. In order to find the appropriate data, the NAS and DFB require contacting as both organizations provide a large proportion of Ambulance Services in Ireland. Data regarding the locations of AED's in Leinster is essential and is required for this project.

Aim 2: All retrieved datasets will require Pre-processing in order to analyse the correct information for example, the AED's dataset may require augmenting if the co-ordinates of each address does not exist within the dataset.

Aim 3: Plotting the locations of all recorded AED's, Ambulance Stations and Firestations in Leinster onto a map for a visual representation will be required for analysis purposes.

Aim 4: The fourth aim will entail further analysis conducted using programming and data mining techniques.

Aim 5: The final aim will entail constructing a full report that documents the results of the study. Finally, visualizations and representations of the results demonstrated through a data visualization tool is essential to fulfil all project aims.

1.3 Technologies

R:

R is a programming language and environment for statistical computing and graphics. R proved the perfect fit to conduct in-depth analysis with the selected data sets that will analyse Automated External Defibrillators (AED) distribution throughout Leinster.

RSutdio:

RStudio is an integrated development environment (IDE) for R programming language. RStudio will be the platform used to conduct the above in-depth analysis on the data sets.

Microsoft Excel:

Microsoft Excel is a spreadsheet program created by Microsoft. The data sets acquired are in comma separated values file (.csv) format. Microsoft Excel was used to pre-process the data and loading the data into a MySQL database.

SQL:

Structured Query Language is a computer programming language for querying relational database systems. SQL is the perfect language to query the created relational database.

MySQL:

MySQL is an open source relational database management system. MySQL was the platform used to create the required database that allows queries on the stored data for analysis purposes.

Tableau:

Tableau is data visualization software that allows the user to represent their findings in a clear and concise way. Once the analysis is completed, Tableau is the preferred platform to represent all findings.

1.4 Structure

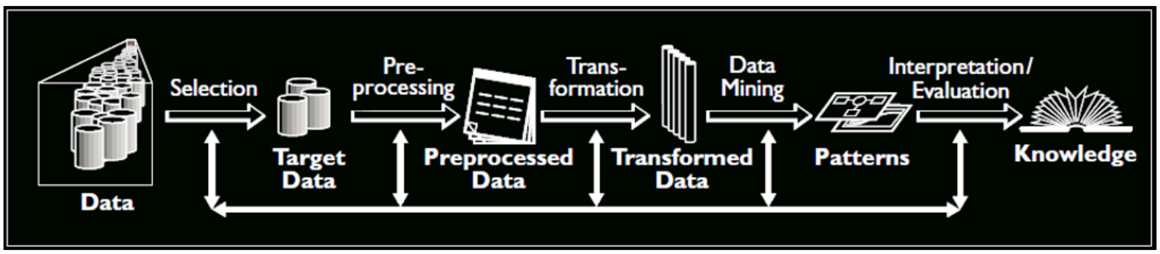


Figure 1: Knowledge Discovery in Database Diagram

Throughout the entirety of the project the core data analytics methodologies; knowledge discovery in databases (KDD) was adopted and applied to the project appropriately. The KDD methodology has a series of fundamental stages that require serious attention;

1. **Data Selection:** The data selection stage focuses on acquiring the data set or data sets that require analysis to discover useful information relating to the chosen topic. When applied to this project carefully selecting appropriate datasets from several sources proved fundamental in fulfilling the projects objectives.
2. **Pre-Processing:** The pre-processing stage requires the cleansing of data. Data cleansing is the process of detecting, correcting or removing corrupt or inaccurate data records from a dataset or database. Pre-Processing in this project entailed augmenting some of the datasets to make the data richer and removing unnecessary data from certain datasets. The AED locations dataset required

cleaning to represent only AED locations in Leinster, while redundant records required deleting i.e. replicated records existed and the distribution of each address spread across four columns and did not contain the geographic co-ordinates for each address. To solve this problem, the merging of all columns into one address column was necessary and enabled the possibility to use Google Maps Application Programming Interface (API) that connects to a google server to provide the geographic co-ordinates of any address by using pre-determined functions. A similar process was required when plotting the locations of all Ambulance and Fire Stations onto an Interactive map in RStudio. The Dublin Fire Brigade dataset contained 20,000 records and was reduced to 3,503 records in order to possess a dataset that represented only ECHO emergency calls.

3. **Transformation:** The transformation stage requires methods that include dimension reduction and record sampling. The Transformation process implemented for the project included preparing the data for K-means clustering by transforming the data into a z-score format. This is a necessary step as some variables contain values much higher than other variables. Transforming the data to z-score enables us to calculate the probability of a score occurring within normal distribution that ultimately provides the ability to compare scores from different distributions. Another example of data transformation included converting specific coordinates into European Petroleum Survey Group (EPSG) format in order to convert a data frame into a SpatialPointsDataFrame that ultimately enabled the counting of AEDs in every settlement in Leinster. Section 2.3.9 outlines this process in detail.
4. **Data Mining:** The data mining stage searches for data patterns that may be of interest, the patterns extracted come in various forms depending on the method used. Machine Learning Algorithms are of this kind, and used to discover relationships between variables or predict outcomes by analysing all data applied to the algorithm. In relation to this project, the K-means clustering algorithm was implemented to discover if accessible AED's tend to reside in larger populations. A Regression Decision Tree was also built to predict Heart Related Deaths in Leinster.

5. **Interpretation/evaluation:** The final stage of the KDD involves evaluating the discovered knowledge. The main aim for applying the Interpretation/evaluation stage to the project will be using a visualization software application that allows the user to represent their findings in a clear and concise way.

1.5 Research

Research in preparation for the practical aspect of this project entailed studying both the Dublin Fire Brigade and National Ambulance Services at great length. The NAS.ie, HSC.ie and DFB.ie websites provided a good starting point in guiding me in the right direction when trying to understand the Ambulance set-up throughout Ireland and Leinster in particular. Various websites including the community first responders and helping hearts sites also provided initial valuable information. Several reports and documents published by the Health Information and Quality Authority and HSE were examined to gain insights into the structure and expected performance levels of the industry. For example, in 2015 the HSE published “The National Ambulance Service of Ireland, Emergency Service Baseline and Capacity Review”. The report commissioned by the Health Service Executive and carried out by Lightfoot Solutions UK [8]. The report found that 26.6 % of ambulances fail to reach the incident within eight minutes of all life threatening calls including Echo calls (Life threatening cardiac or respiratory arrest). The analysis conducted in this project differs, as the focus is effective distribution of accessible AED’s rather than response times of Emergency Services.

1.6 Acronyms, Definitions and Abbreviations

1.6.1 Acronyms

NAS	National Ambulance Service
HSE	Health Service Executive
DFB	Dublin Fire Brigade
CSO	Central Statistics Office
AED	Automated External Defibrillators
KDD	Knowledge Discovery in Databases

1.6.2 Definitions

Automated External Defibrillator

AED is a small portable piece of equipment that can deliver an electric shock to a person in order to convert a cardiac arrhythmia (ventricular fibrillation) into a natural rhythm.

National Ambulance Service

The National Ambulance Service is the public ambulance service in the Republic of Ireland.

Health Service Executive

The HSE provides all of Ireland's public health services in hospitals and communities across the country.

Dublin fire Brigade

Dublin Fire Brigade provides a fire and emergency response service to Dublin City and County.

Central Statistics Office

Government body that compiles official statistics in Ireland [3].

Stakeholder

A person with an interest in the project

Echo

Clinical Status 1 ECHO, life threatening emergency (cardiac/respiratory arrest).

Delta

Clinical Status 1 DELTA, potentially life threatening emergency (not cardiac/respiratory arrest), calls responded to by a patient carrying vehicle in 18 minutes 59 seconds.

Machine Learning

Rob Schapire Lecturer at Princeton University when speaking on Machine Learning; "Machine learning studies computer algorithms for learning to do stuff. We might, for instance, be interested in learning to complete a task, or to make accurate predictions, or to behave intelligently. The learning that is being done is always based on some sort of observations or data, such as examples direct experience, or instruction. So in general, machine learning is about learning to do better in the future based on what was experienced in the past” [12].

Knowledge Discovery in Databases

Knowledge Discovery in Databases (KDD) is the process of implementing several fundamental techniques including data mining to discover knowledge in data.

1.7 Project Restrictions

Data:

All response times analysed arouse from historical data. This report acknowledges NAS and DFB response times regarding Echo Cat 1 calls can change over time. A request to acquire data relating to ECHO calls responded to by the National Ambulance Service was denied by Desmond Kelly, NAS community first responder coordinator, contacted by email on 24/10/2016.

Time:

The project timeline spans from September 19th 2016 to May 17th 2017.

Cost:

The possible cost of hosting the systems could arise.

Software:

Student resources and open source platforms provide the necessary tools, although unforeseen additional software may be required that could come at a cost.

Legal:

Datasets in relation to the DFB Ambulance Calls 2013-15 are open source and provided by the Dublinked website [7]. Helping Hearts provided the dataset in relation to AED Locations via their website [6]. The Central Statistics Office provided datasets relating to the demographics of settlements in Ireland, [3].

2 System

2.1 Requirements

2.1.1 Data requirements

The three main data requirements to fulfil the projects objectives are as follows:

The dataset for DFB Ambulance Calls 2013-15 is open source and was acquired from the Dublinked website [7]. Dublin Fire Brigade is Ireland's largest full-time brigade and employs over 1000 people. There are fourteen stations, twelve of which are full-time and two are part-time (retained). DFB operates an Emergency Ambulance Service as well as a Fire, rescue and emergency service for Dublin City and County.

Table 2: Dublin Fire Brigade Dataset

Name	Type	Description
Date	date	Date of incident
Station	varchar(255)	Location of incident by fire station area
Clinical Status	varchar(255)	Critically of medical incident
TOC	time	Time of Call
ORD	time	Time 1st appliance ordered
MOB	time	Time 1st appliance mobile (to incident)
IA	time	Time 1st appliance in Attendance (at scene)
LS	time	Time 1st appliance leaves scene
AH	time	Time 1st appliance books at Hospital
MAV	time	Time last appliance books Mobile and Available
CD	time	Time last appliance close down at station - Incident Closure

Helping Hearts provided the dataset regarding the locations of all recorded AED's on their website [6]. The dataset contains street addresses of all the AED's in Leinster, Ireland. Augmenting the dataset with the Co-ordinates of the AED Locations will be a key requirement.

Table 3: Helping Hearts Data Dictionary

Name	Type	Description
title	varchar (255)	Name of the location where the AED is
address1	varchar(255)	Street Address of the AED
address2	varchar(255)	An extension to address 1
city	varchar(255)	Name of the town where the AED can be located
county	varchar(255)	Name of the County where the AED is locate
The following variables were added to the dataset for analysis purposes.		
latitude	decimal	The exact latitude of each AED
longitude	decimal	The exact longitude of each AED

Please note: A request to acquire data relating to ECHO calls responded to by the National Ambulance Service was denied by Desmond Kelly, NAS community first responder coordinator, contacted by email on 24/10/2016. He outlined the only available data relating to ECHO calls available to the public already resided on their website. Unfortunately, the available data on the NAS website is minimal and does not provide meaningful information relating to ECHO calls response times. In order to obtain meaningful information relating to ECHO calls responded to by the NAS, data was manually extracted from the National Ambulance Service of Ireland emergency service baseline and capacity review Undertaken by Lightfoot Solutions UK Ltd Commissioned by the Health Service Executive [8] and the Health Service Performance Assurance Reports [14].

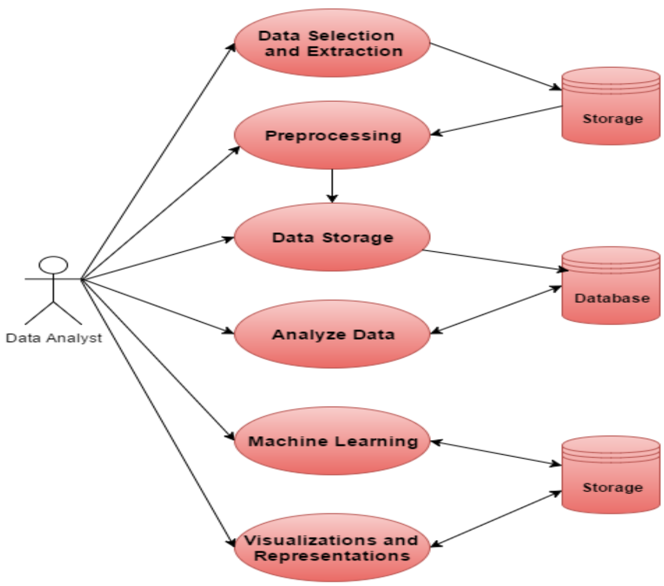
Further datasets were acquired from the Central Statistics Office (CSO) relating to the demographics of Leinster and mortality rates relating to the circulatory system [3].

2.1.2 Functional requirements

The Functional requirements within this project were achieved by adopting and applying the Knowledge discovery in Databases (KDD) methodology. The fundamental stages when creating a data analytics project include; data selection, Pre-Processing, data transformation, data mining and finally evaluations.

2.1.2.1 Use Case Diagram

The following use case diagram provides an overview of all functional requirements.



2.1.2.2 Use Case Priority Table

The following priority table determines the meaning of each priority ranking associated with each Use Case.

Table 4: Priority Table

Priority 1	Considered a must have and vital to the scope of the project.
Priority 2	Also considered a must have but alternatives are viable.
Priority 3	Not considered a must have although enhances the possibilities of analysis.

2.1.2.3 Requirement 1: Data Selection and Extraction

Description & Priority

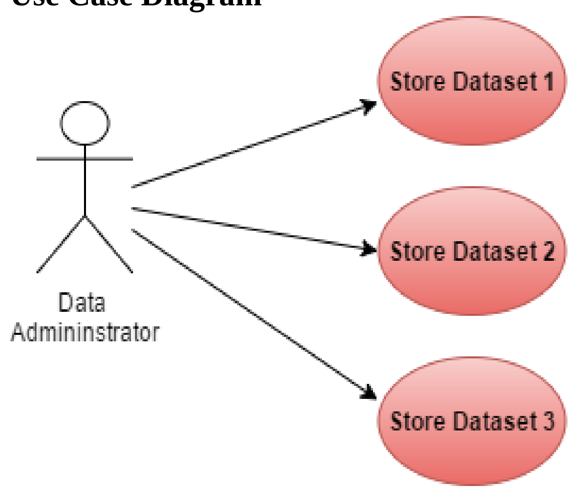
The data selection requirement is the initial requirement. This requirement is essential to the project and ranked priority 1. Without selecting the appropriate data, all other requirements are impossible.

Use Case

Scope

Selected data is extracted and downloaded. Subsequently the data is stored in a secure location.

Use Case Diagram



Flow Description

Precondition

The user must have access to their email account and relevant website to download the required data.

Activation

This use case starts when the data administrator retrieves

Main flow

1. The Data Administrator (DA) logs onto email account and downloads data to secure location.
2. The DA accesses the relevant webpage and downloads the dataset to a secure location.

3. The DA checks data integrity by opening and reading files in relevant software.
4. Data is stored to a secure location.

Exceptional flow

1. The relevant webpage is down.
2. Email account is not accessible.
3. The datasets are corrupted.

Termination

The process of selecting, extracting and storing the data is complete. Therefore, termination of the process occurs.

2.1.2.4 Requirement 2: Pre-Processing

Description & Priority

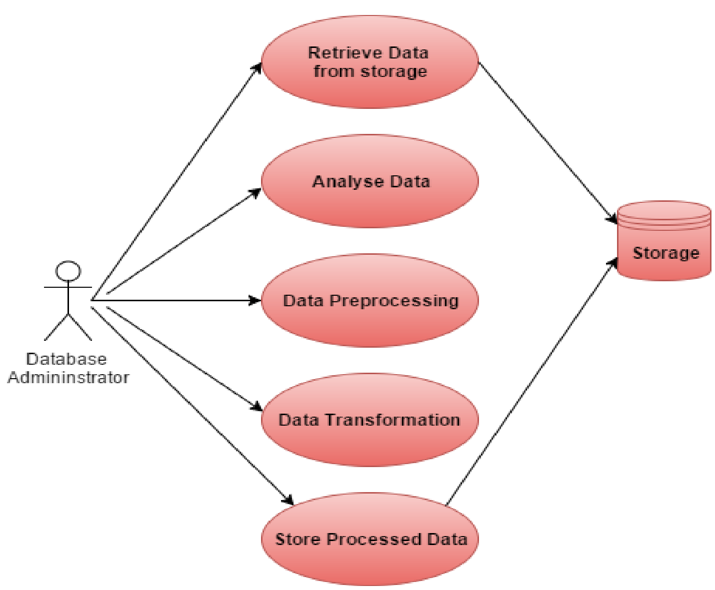
The Pre-Processing use case entails augmenting the dataset, which will enable the possibility of in depth analysis at a later stage. The Pre-Processing requirement is of priority 1.

Use Case

Scope

Pre-processed data enhances the ability to conduct the appropriate analysis.

Use Case Diagram



Flow Description

Precondition

The data must be accessible.

Activation

This use case commences when retrieval of the data occurs.

Main flow

1. The Data Administrator retrieves the data.
2. The data is analysed and cleaned.
3. Pre-Processing commences.
4. Transformation of the data occurs if required.
5. The processed data is stored.

Alternate flow

Various platforms have the ability to apply this use case.

Exceptional flow

The data is not retrievable.

Termination

The Pre-Processing use case terminates when the data is cleansed, transformed and Pre-Processing is completed.

2.1.2.5 Requirement 3: Data Storage

Description

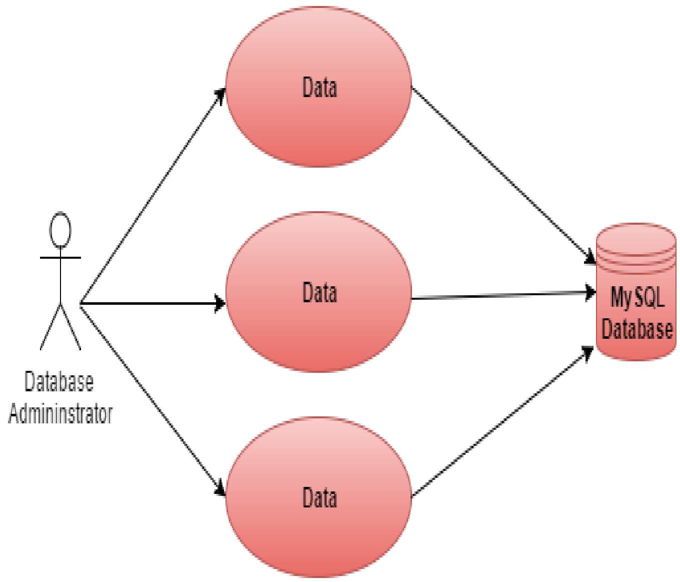
The data storage requirement entails storing the datasets in database tables to ensure the data is secure and accessible at all times. The data storage requirement is ranked priority 3.

Use Case

Scope

Development of a Database with specific tables to store the data occurs.

Use Case Diagram



Flow Description

Precondition

The data must be Pre-processed and Transformed to the correct format before it can be stored in a database.

Activation

Activation of the use case commences when the data administrator makes the data available.

Main flow

1. The Database Administrator (DBA) retrieves the data.
2. The DBA creates suitable tables within the database that will store the data.
3. The data is loaded into database tables.

Alternate flow

Alternative storage is a viable option.

Exceptional flow

Retrieval of the data is not possible.

Termination

The termination of the data storage requirement occurs when the data is successfully stored in a database.

Post Condition

Data is accessed while residing within the database.

2.1.2.6 Requirement 4: Analyse Data

Description & Priority

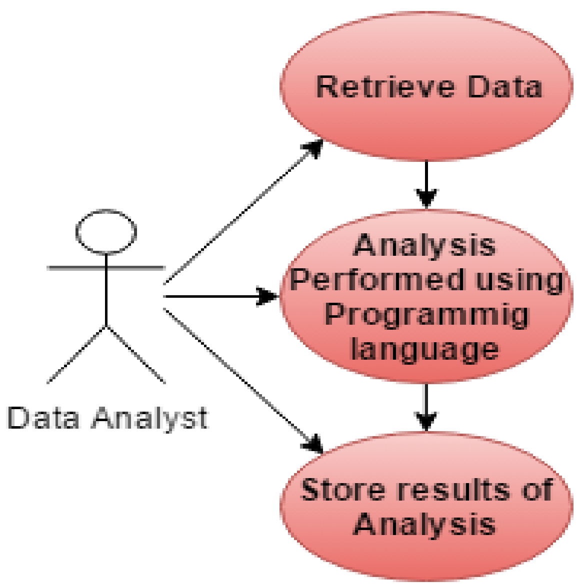
The Analyse Data requirement is a fundamental requirement to the project and ranked priority 1. The analysis of the datasets requires significant attention to detail and will help achieve the project goals.

Use Case

Scope

Data is retrieved from the database where analysis is performed using several programming scripts. Subsequently evaluation and interpretation of the results occur. Results are then saved.

Use Case Diagram



Flow Description

Precondition

The data must be residing in a database and be accessible to the data analyst.

Activation

This use case starts when the data analyst calls the data from the database.

Main flow

1. The Data Analyst retrieves the data from a database.
2. Data is analysed using specific programming language.
3. Interpretations of the results commence.
4. The results are saved to a secure location.

Exceptional flow

Retrieval of the data from the database is unsuccessful.

Termination

The use case terminates when the analysis is complete.

Post Condition

The data is residing in a database waiting for further analysis.

2.1.2.7 Requirement 5: Machine Learning

Description & Priority

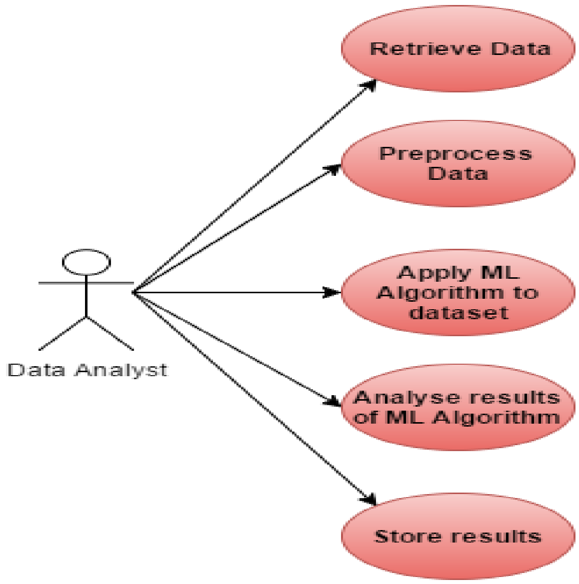
The Machine Learning Requirement is an exploratory requirement and ranked priority 2. The datasets will run through Machine Learning Algorithms to discover patterns and comparisons between the datasets.

Use Case

Scope

Datasets are retrieved and Pre-processed. When preparation is complete, the data runs through the Machine Learning (ML) algorithm. The results are analysed and finally the results will be stored.

Use Case Diagram



Flow Description

Precondition

Data is retrievable.

Activation

The use case activates when retrieval of the data occurs.

Main flow

1. The Data Analyst retrieves the data.
2. The data analyst Pre-processes the data.
3. The data runs through Machine Learning Algorithms.
4. Results are analysed.
5. Results are stored.

Exceptional flow

The problem is unsolvable.

Termination

The Machine Learning use case terminates when the results are stored.

2.1.2.8 Requirement 6: Visualizations and Representations

Description & Priority

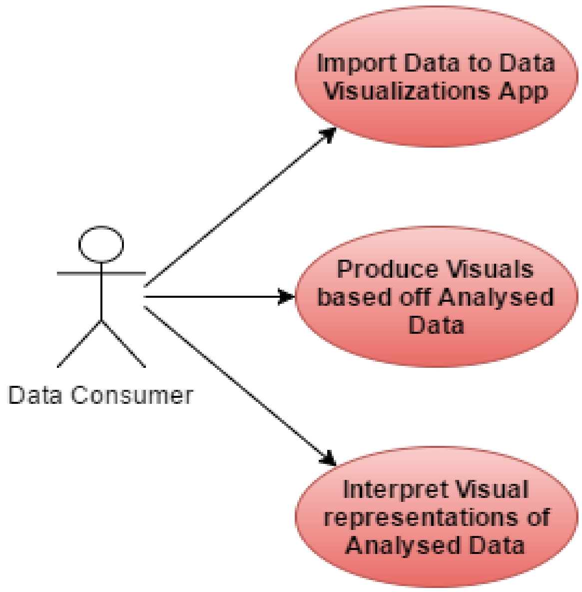
Visualizations and Representations is the final requirement within the scope of this project. Visualizations and Representations is ranked priority 2 as it is important to show the results of the analysis in a clear meaningful way.

Use Case

Scope

Import analysed data into data visualizations application. Visual representations produced then help the data consumer interpret the results of the analysis.

Use Case Diagram



Flow Description

Precondition

All analysis of the datasets must be complete at this stage in order to create final visualizations and representations of the data.

Activation

This use case starts when the data administrator accesses the internet.

Main flow

1. The analysed data is imported to a data visualizations application
2. Create final visualizations and representations of the results.
3. Visualizations and representations are available to consume.

Exceptional flow

Imported the data to the visualizations application is not possible.

Termination

The use case terminates when all visualizations and representations are complete.

2.1.3 Non-Functional Requirements

2.1.3.1 Security requirement

Security is a fundamental aspect to the system. The data provided will be securely stored in a private account hosted on a secure server. The Data Administrator will have full rights to the account protected by strong passwords and will only authorize access to the account when suitable.

- The system shall be secure.
- Authorization shall be granted if necessary.
- A protected file system shall be designed.

2.1.3.2 Availability requirement

Data shall remain available to the system throughout the project scope.

2.1.3.3 Integrity requirement

The datasets shall contain all attributes and values in order for the correct analysis to be performed resulting in accurate outcomes.

2.1.3.4 User requirements

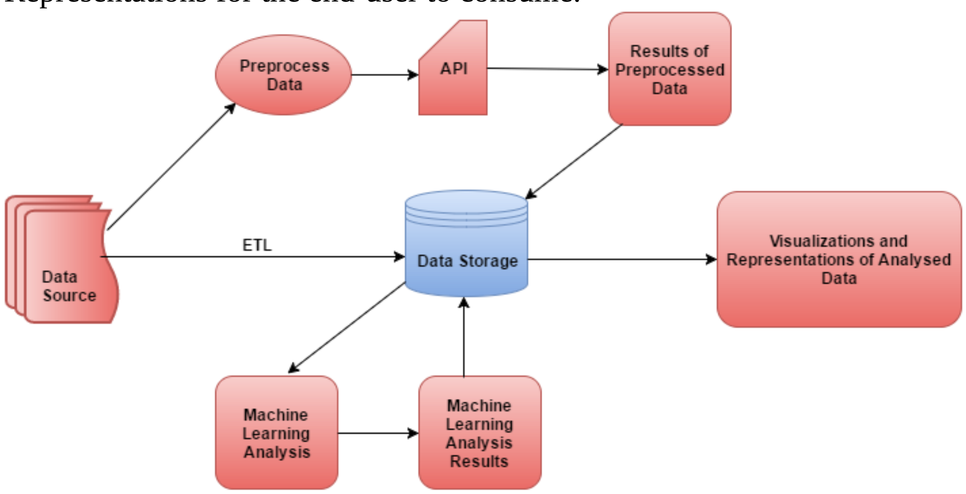
The User Requirements Definition defines the objectives and requirements for the development of a system that through in depth analysis either verifies or rejects the idea that Automated External Defibrillators (AED) should be easily accessible to the public in every village/town across Ireland to help decrease mortality rates.

2.1.3.5 Application Programming Interfaces (API)

The initial dataset Helping Hearts provided containing the locations of all recorded AED's on their website does not contain the co-ordinates of any address. To solve this problem The Google Maps Geocoding API will be required to retrieve the co-ordinates for each provided address via a remote call.

2.2 *Design and Architecture*

The following diagram is a high-level view of the system architecture that contains various components. The storage component is central to the system and interacts with all components. Once the data is retrieved and stored, analysis of the datasets can commence by applying programming techniques to the data and running the data through Machine Learning algorithms to find patterns and insights within the data that will indicate if more accessible AED's in Leinster will help decrease mortality rates. The system will than have the ability to provide the results of the analysis through Visualisations and Representations for the end-user to consume.



2.3 *Implementation*

This section outlines any significant analysis conducted throughout the project lifespan. All of the R code used in the analysis and implemented in RStudio is attached and available with this submission. Please proceed to the next page to view a horizontal low-level view of the system architecture.

2.3.1 Low Level System Architecture

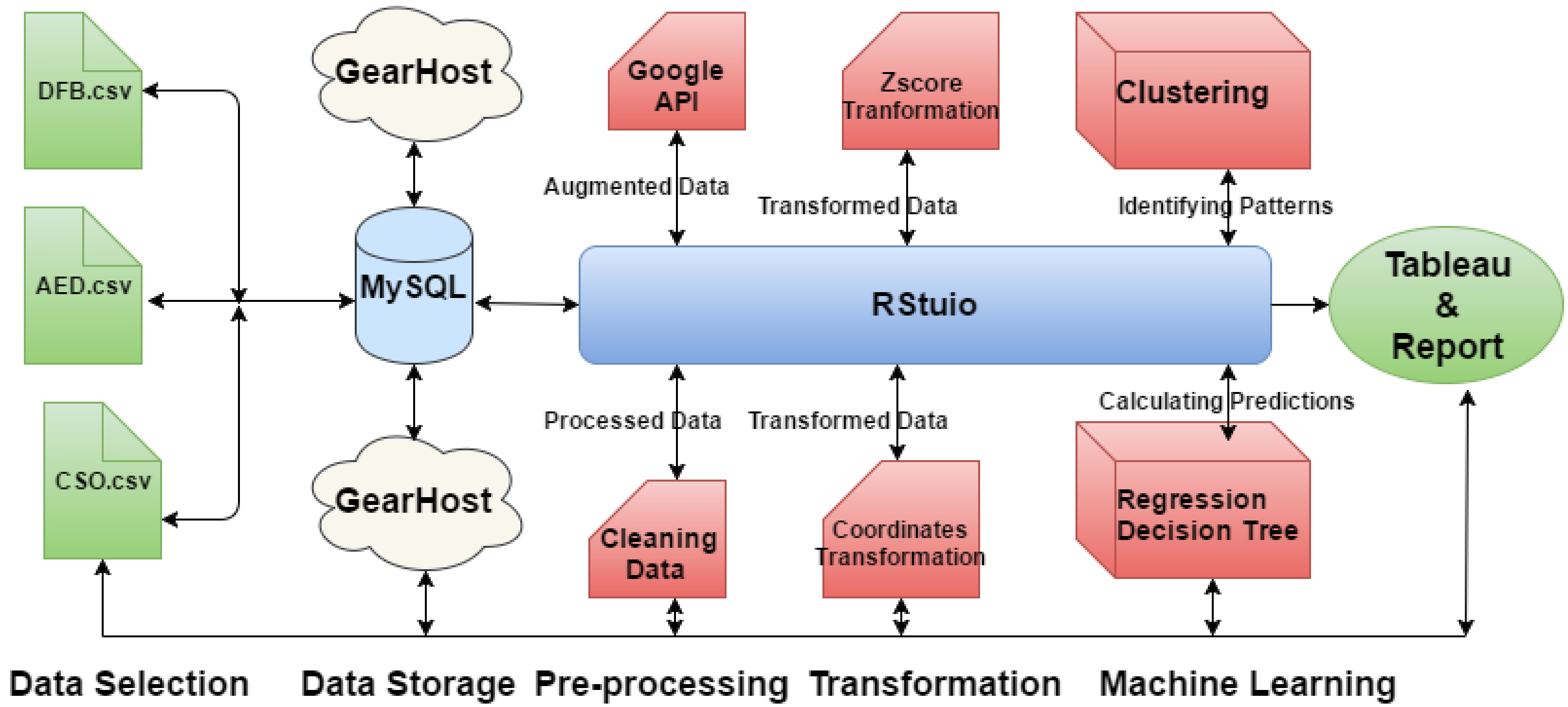


Figure 2: Displays Low-Level View of System Architecture

A low-level view of the system architecture is shown in Figure 2. The system architecture displays the workflow throughout the project life span. The KDD methodology provided the perfect base when planning and strategizing the steps required to fulfil the projects goals. The graph displays key phases of any data analysis project that include, data selection, cleaning/pre-processing the data and transformation of the data in preparation for Machine Learning Algorithms.

2.3.2 *Important R functions used*

- mean()
- sd()
- str()
- summary()
- POSIXct()
- earth.dist()
- subset()
- table()
- over()
- set.seed()
- readOGR()
- sptransform()
- data.frame()
- cor()
- asNumeric()
- asfactor()
- printcp()
- plotcp()

2.3.3 *Important R packages used*

- kmeans
- Rpart
- rattle
- RColorBrewer
- C50
- rgeos
- fossil
- plyr
- rgdal
- CrossTable
- leaflet
- ggplot2
- ggmap
- sp

2.3.4 *Initial Analysis*

The dataset that represents AED Locations in Ireland required Pre-processing before the dataset was suitable for analysis. Records required deleting due to duplication. The address for each AED spread across four columns and did not possess the co-ordinates for any of the addresses. To solve this problem, I merged all the columns into one address column and used a Google Maps Application Programming Interface (API) to connect to a Google server that provides the co-ordinates of an address by using pre-determined functions. See Figure 3 and 4 that illustrates the function used to call the co-ordinates for any address.

Figure 3: Displays the function used to call the latitude for a specific address.

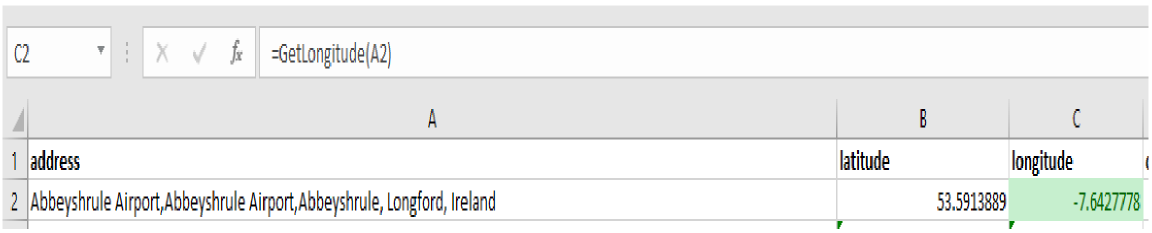


Figure 4: Displays the function used to call the longitude for a specific address.

The main storage component for the system is a relational databse created in MySQL Workbench. Figure 5 illustrates the schema used to create a table that stores all AED locations in Leinster, Ireland. Subsequently, the database was then used to construct various other tables that represented other datasets used for analysis in this project.

```
create database project;
use project;

DROP TABLE IF EXISTS `aed_locations`;
/*!40101 SET @saved_cs_client     = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `aed_locations` (
  `title` varchar(255) DEFAULT NULL,
  `address1` varchar(255) DEFAULT NULL,
  `address2` varchar(255) DEFAULT NULL,
  `city` varchar(55) DEFAULT NULL,
  `county` varchar(55) DEFAULT NULL,
  `latitude` decimal(10,8) DEFAULT NULL,
  `longitude` decimal(11,8) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;
```

Figure 5: The schema used to create the aed_locations table in MySQL Workebench.

Once the co-ordinates of each address were obtained, the next step was to plot the locations of all recorded AED's in Leinster onto a map for a visual representation.

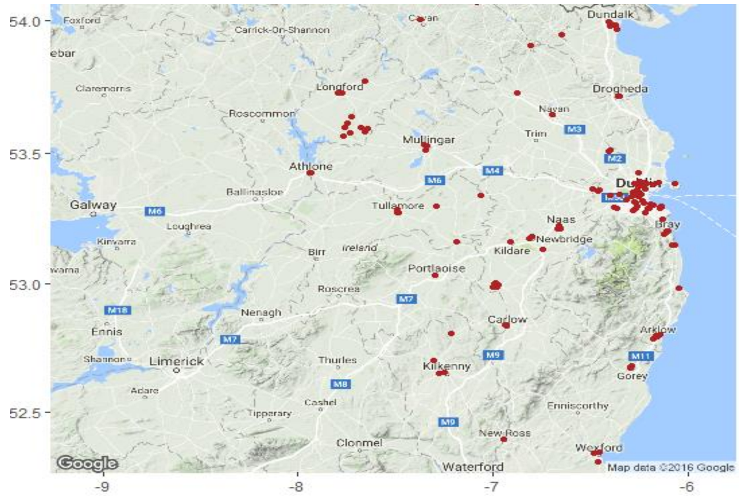


Figure 6: Visual Representation of the AED locations in Leinster, Ireland.

Figure 6 displays a visual representation of all AED locations in Leinster. When visually inspecting the map, we can see clusters appearing in Dublin and other larger populations around Leinster. Further investigation is required as examining the demographics of each town and analyzing the response times from the National Ambulance Service and the Dublin Fire Brigade Ambulance service is necessary to identify areas of concern.

2.3.5 Dublin Fire Brigade Response Time Summary

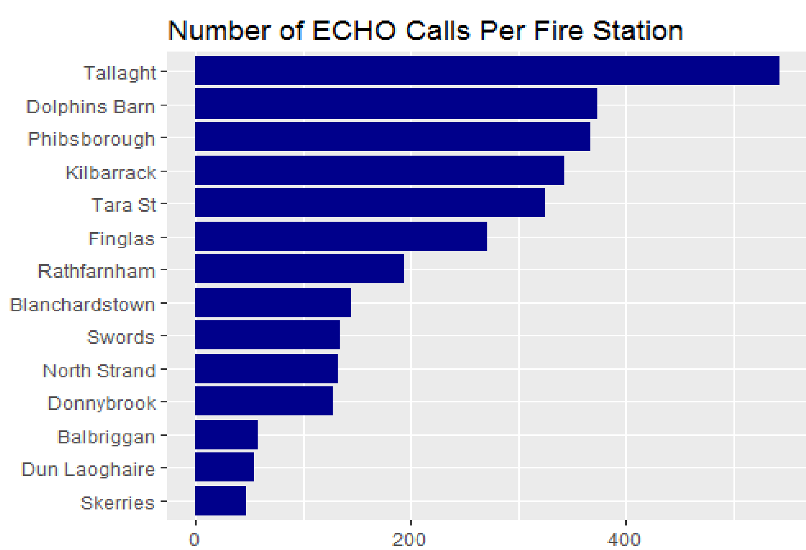


Figure 7: Number of ECHO calls recorded per Fire Station.

The number of ECHO calls for all Fire Stations in Dublin from 2013 – 2015 is shown on Figure 7. Tallaght was the busiest station with over 500 ECHO calls and Skerries was the station that experienced the least amount with less than 50 ECHO calls during the same time-period.

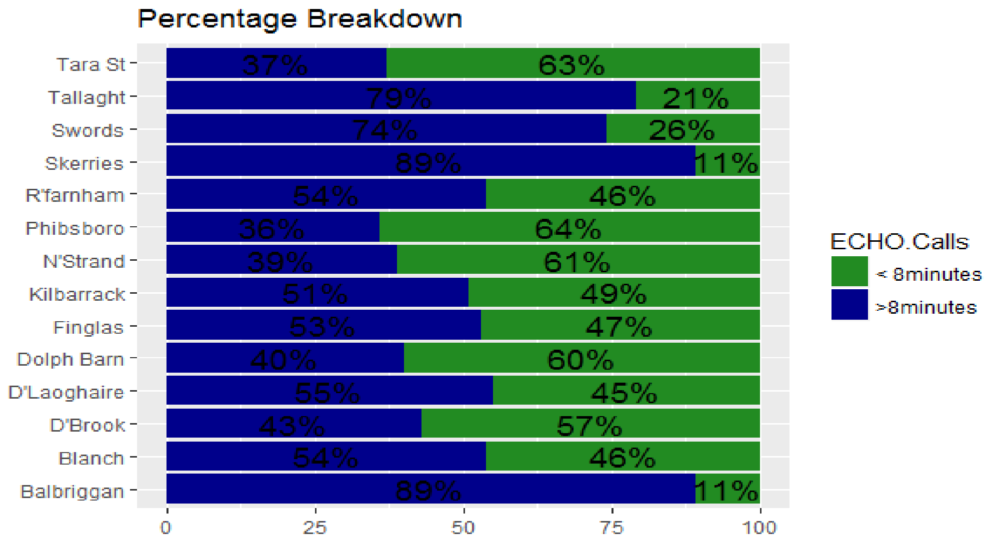


Figure 8: Percentage breakdown of ECHO Calls per Station

When breaking down the response times of ECHO calls per fire station Figure 8 shows that none of the fire stations adhere to the Key Performance Indicators (KPIs) recommended by the HIQA that Echo calls (Life threatening cardiac or respiratory arrest) are responded to within 8 minutes for 75% of all cases are met. Figure 8 also displays alarming percentages for Ballbriggan, Skerries, and Tallaght fire stations where more than 75% of all ECHO calls are responded to after the eight minutes requirement. Phibsborough and Tara St perform the best with over 60% of ECHO calls responded to within eight minutes.

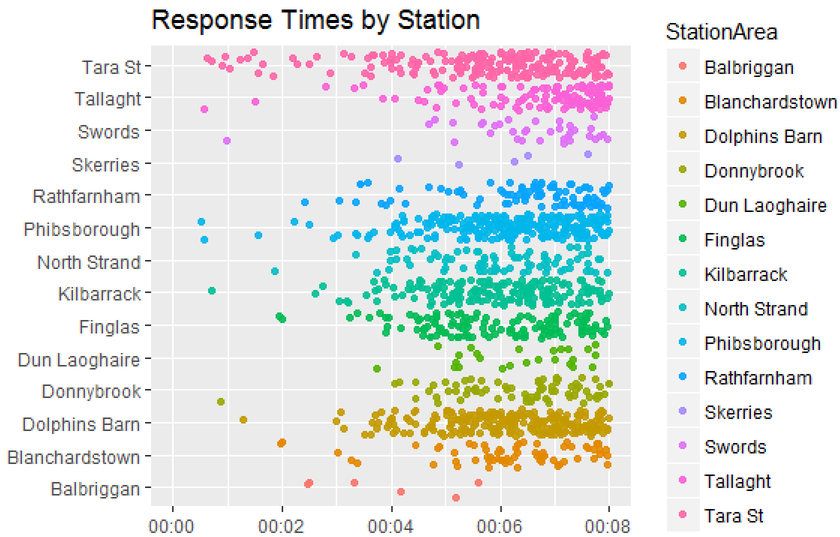


Figure 9: Breakdown of ECHO calls responded to within 8 minutes.

Figure 9 visually breaks down the number of ECHO calls responded to within eight minutes per fire station. Although Skerries, Balbriggan and Dun Laoghaire received the least amount of ECHO calls in comparison to other Fire Stations the graph clearly displays a low number of calls responded to within eight minutes for said stations. Tara St and Phibsborough response times are the most impressive with Tara St in particular responding to a number of calls less than four minutes, which ultimately increases the chance of survival. Although Tara St and Phibsbourough perform well against the other stations, over 10% of received ECHO calls are responded to outside of the KPI's recommendation. Thus, the need for easily accessible automated defibrillators is a necessity especially in poor performing areas.

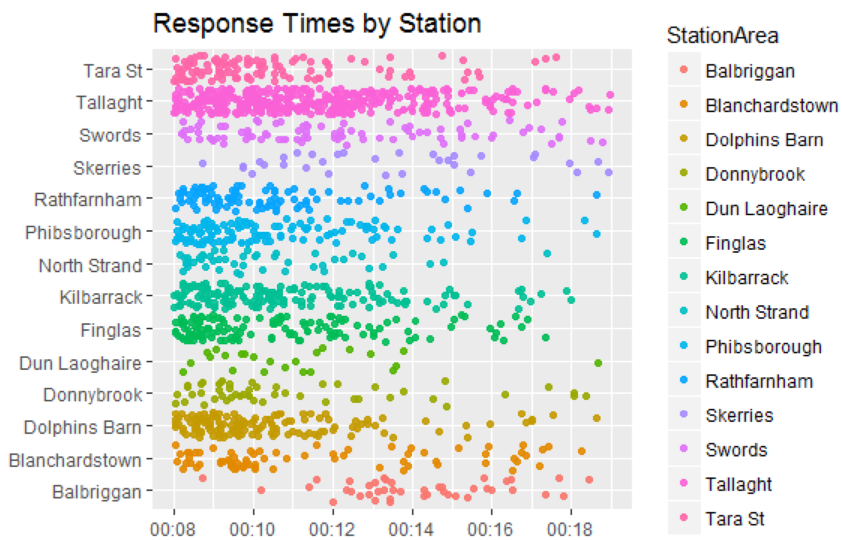


Figure 10: Breakdown of ECHO calls responded to greater than 8 minutes and less than 19 minutes.

From examining Figure 9 and Figure 10, the fire stations response times largely fall between five minutes and twelve minutes for most fire stations. Unfortunately, the graphs also show all stations record response times as high as nineteen minutes. Figure 10 displays Balbriggan performing poorly again with the majority of ECHO calls within the displayed time period greater than twelve minutes.

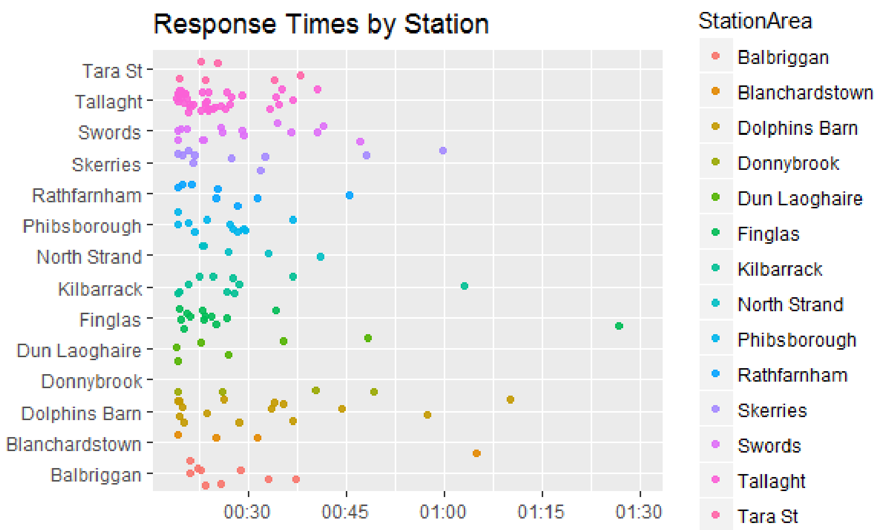


Figure 11: Breakdown of ECHO calls responded to greater than 19 minutes.

Although Figure 11 displays a small number of ECHO calls responded to greater than nineteen minutes per fire station, the chances of survival if suffering from cardiac arrest and time to defibrillation is greater ten minutes without CPR is 0%. Studying the statistics displayed on Figure 7 through to Figure 11 confirms the need for more easily accessible AEDs in all areas Dublin Fire Brigade respond to; furthermore, fire stations that perform extremely poorly i.e. Balbriggan Skerries and Tallaght are high priority areas.

2.3.6 National Ambulance Service Response Time Summary

Due to the National Ambulance Service and their unwillingness to provide data for this project and the lack of in-depth available data regarding response times per station, the following data was obtained from a report published by the HSE named the "National Ambulance Service of Ireland, Emergency Service Baseline and Capacity Review". The obtained data represents the NAS response times from August 2014 – August 2015 of twenty-seven towns across Leinster with an ambulance station operating from 23 of the 27 town's recorder in the report.

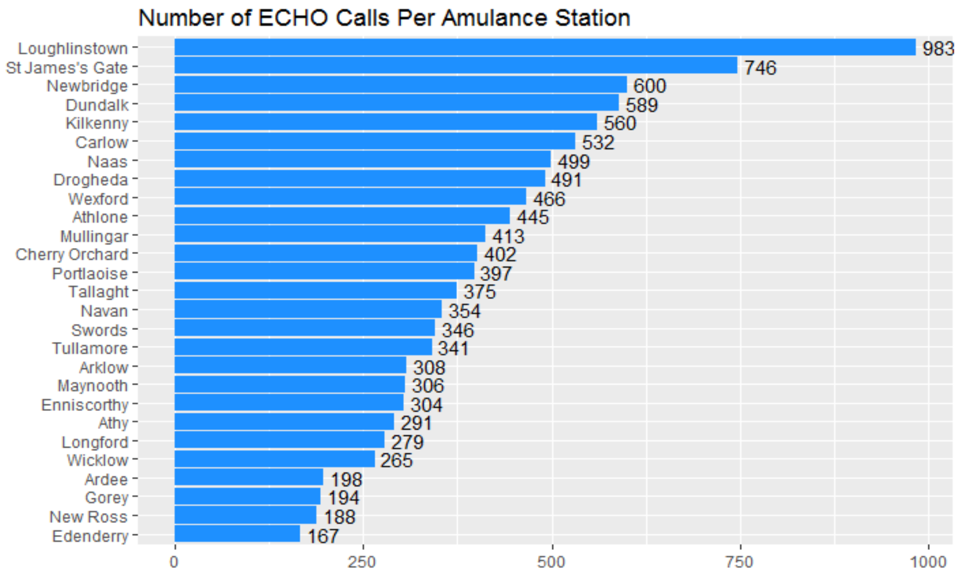


Figure 12: Number of ECHO calls recorded per town and responded to by the NAS.

Figure 12 displays the number of ECHO calls per town in Leinster (chosen by the HSE) and responded to by the NAS from August 2014 – August 2015. Loughlinstown was the busiest area with 983 ECHO calls and Edenderry experienced the least amount of calls with only 167 ECHO calls during the same time-period. The average number of ECHO calls received by the NAS during a one-year period across the chosen towns is 409. In comparison, the average number of ECHO calls received by the DFB during a two-year period is 209. As expected the NAS, which covers the whole of Leisnter including parts of Dublin is the busier of the two services relating to ECHO calls.

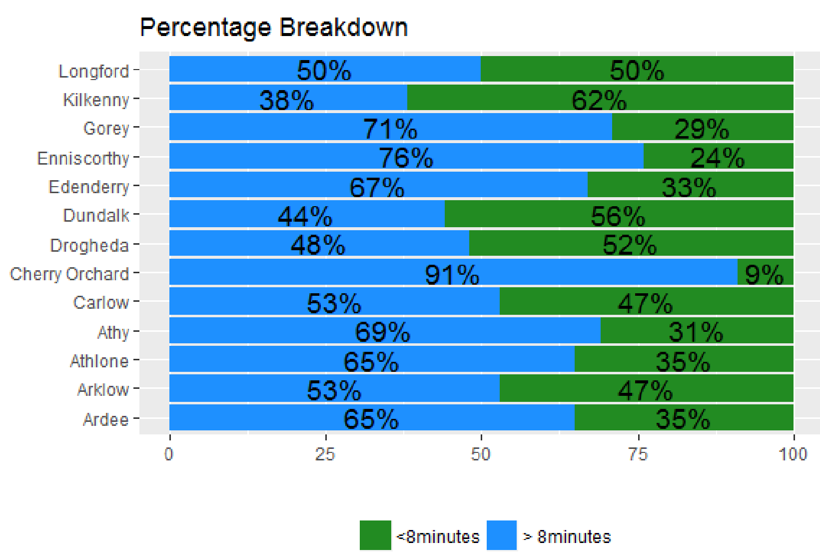


Figure 13: Percentage breakdown of ECHO Calls per Station.

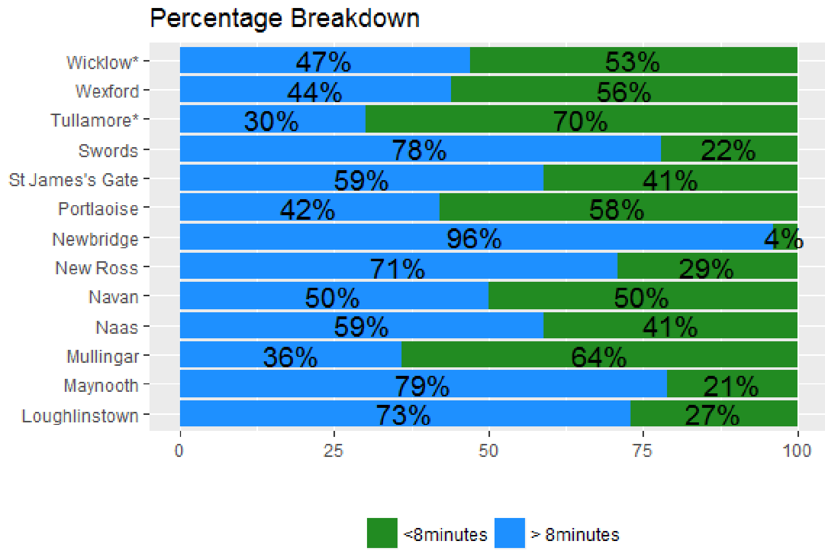


Figure 14: Percentage breakdown of ECHO Calls per Station.

Since there are twenty-six areas provided by the NAS, the areas are split into two graphs for optimal inspection. Figure 13 and 14 both show the NAS Percentage breakdown of ECHO Calls per Station and similarly to the DFB they are not meeting the Key Performance Indicators (KPIs) recommended by the HIQA, that Echo calls (Life threatening cardiac or respiratory arrest) are responded to within 8 minutes for 75% of all cases are met. Towns fixed with an asterix indicate a rapid response vehicle (RRV) operates in that area. Figure 14 displays Wicklow and especially Tullamore performing extremely well with an RRV available. Maynooth and more concerningly Newbridge are both towns without a RRV and/or Ambulance Station and response times less than eight minutes relating to ECHO calls are twenty-one and an alarmingly low four percent in Newbridge. Figure 13 shows Cherry Orchard also performing poorly with only 9% of all ECHO calls responded to within eight minutes. Other areas i.e. Gorey, Swords, New Ross, Enniscorthy and Loughlinstown all contain 70% of ECHO calls responded to in over eight minutes. All of the above towns with concerning response times are considered high priority for easily accessible AED’s.

2.3.7 Interactive Map

Progressing from a static map seen in the initial analysis Section 2.3.4, the creation of an interactive map using the Leaflet package in RStudio proved more intuitive.

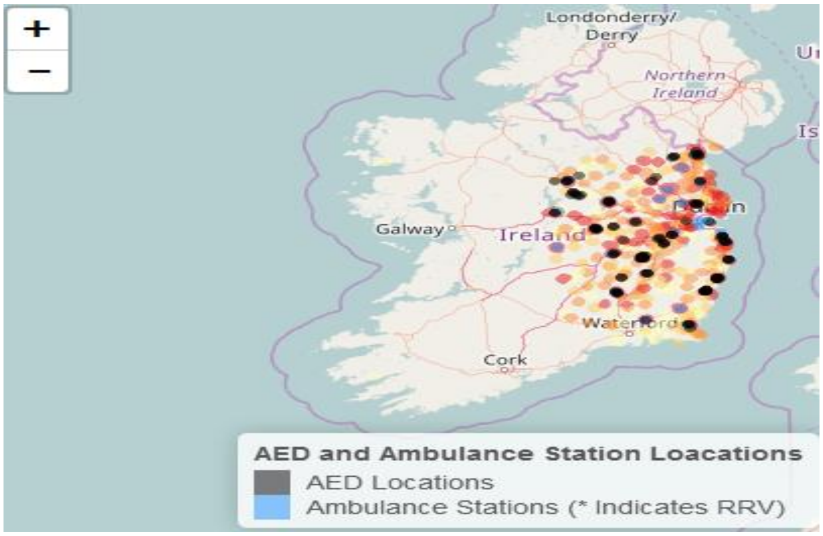


Figure 15: Interactive map of Ireland

Th interactive map seen in Figure 15 displays AED locations, all Ambulance Stations and every settlement in Leinster excluding Dublin City Centre and its suburbs. This was achieved by downloading a shape file from the CSO website relating to the 2011 census containing the polygons of every settlement in Ireland was subsequently plotted along with the AED and Ambulance station locations. The yellow to red colour quantile represents the population for every settlement; yellow represents small population settlements and red represents larger population settlements.

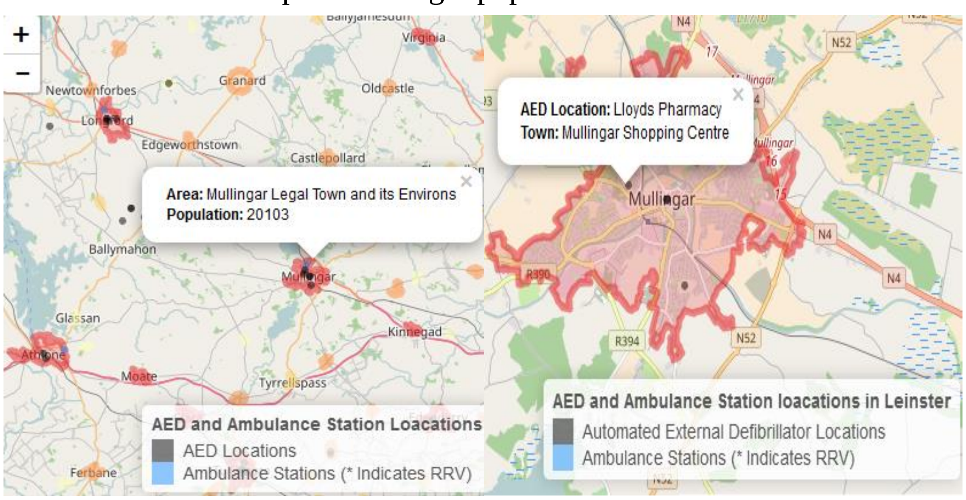


Figure 16: Mullingar slightly zoomed in on the left and more zoomed in on the right.

Figure 16 displays the Map popups programmed for every settlement and AED location. The address of each AED location, the name of the town, population of the area and the name of each Ambulance station is available via popups when clicked. When visually analysing the graph we can see the larger population tend to have AED's and Ambulance stations as opposed to the smaller populations seen in orange tend to have neither.

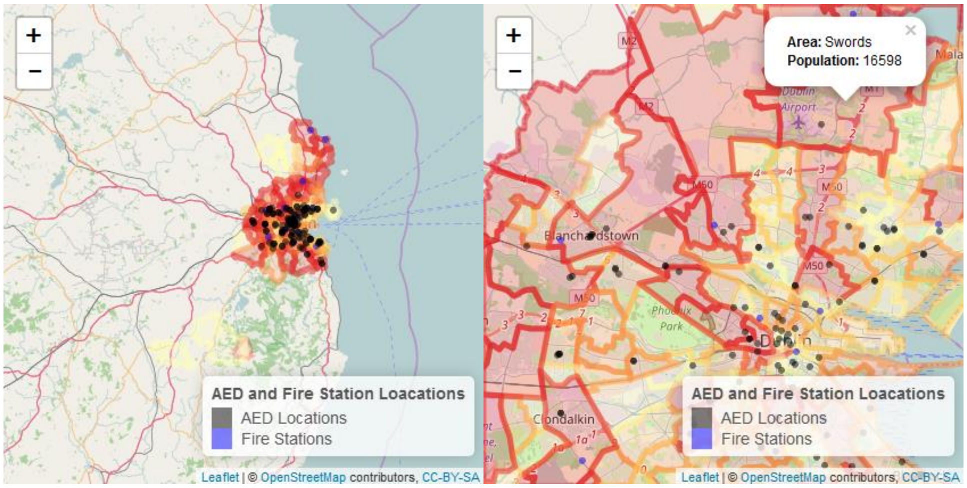


Figure 17: Map of Dublin on the left and zoomed in on the right.

The same procedure implemented for Leinster seen in Figure 16, was also used for Dublin City Centre and its suburbs seen in Figure 17. The graphs display a significant amount of areas with neither an AED nor fire Station.

2.3.8 Distance to Nearest Ambulance/Fire Station

Following on from visually analysing every settlement in Leinster through the interactive map, the next step in the process entailed calculating the distance from every settlement to the nearest station. The process included extracting the centroids (centre of every polygon boundary) and using the NearestL function seen in Section 2.4.2 to look for the closest station. The NearestL function entailed binding the Station coordinates and centroid data frames together by the Lat/Lon variables and using the earth.dist function from the fossil package to calculate the distance. The output of the function shown in Figure 18 includes the name and population of each settlement, coordinates of each centroid and the distance to the closest station in kilometres. The output provides another extremely important variable in preparation for Requirement number 5 seen in Section 2.1.2.7.

Lat	Lon	Area	Population	Distance to Station
53.22885	-6.358254	Bohernabreena	7425	9.1015826
53.23186	-6.129871	Shankill	10102	6.0392305
53.23843	-6.217122	Sandyford	22288	7.9120497
53.24717	-6.128389	Loughlinstown	3665	4.6226457
53.24825	-6.291021	Ballyboden	9611	5.6464071
53.25049	-6.157074	Cabinteely	10038	3.5446044
53.25853	-6.121495	Ballybrack - Killiney	8954	4.2153861
53.26039	-6.461967	Saggart	13124	10.1709826
53.26307	-6.139131	Johnstown - Killiney	4313	2.4972694
53.27055	-6.255662	Ballinteer	8547	2.5471567
53.27130	-6.388869	Killinarden	8955	4.0214300

Figure 18: Output from the NearestL function.

2.3.9 Counting Number of AED's in each Settlement.

Counting the number of AEDs per settlement was another fundamental process required. The process entailed transforming the AED locations dataset into a SpatialPointsDataFrame and subsequently overlaying the on the shape file containing all the settlements in Leinster. The Over function from SP was then used to iterate through each settlement and return a list with a count of the number of AED's in each settlement.

SETTLEMENT	SETTL_NAME	GEOGID
35016	Portarlington	T35016
35010	Athlone Legal Town and its Environs	T35010
35010	Athlone Legal Town and its Environs	T35010
35009	Drogheda Legal Town and its Environs	T35009
35009	Drogheda Legal Town and its Environs	T35009
35008	New Ross Legal Town and its Environs	T35008
35007	Bray Legal Town and its Environs	T35007
35007	Bray Legal Town and its Environs	T35007
35007	Bray Legal Town and its Environs	T35007
35007	Bray Legal Town and its Environs	T35007

Figure 19: Output from the Over function.

Figure 19 displays the output from the Over function from the SP package in R. The output displays the number of AEDs in a particular settlement. By examining the output, we can see that there are two AED locations in Athlone, two in Drogheda, four in Bray etc.

2.3.10 Machine Learning Algorithms

Machine Learning Algorithms are an important requirement for this project. All collected, analysed and generated data using various techniques has led the project to this point. The two main Machine Learning Algorithms implemented in this project are as follows:

- Clustering: An Unsupervised Machine Learning task that automatically groups data, based on similarity. For the purpose of this project, we hope to find clusters relating to population, number of AEDs and distance/time to station.
- Decision Tree: Supervised Machine Learning method that applies a strategy of dividing the data into smaller portions to identify patterns used for prediction.

2.3.11 Clusters in Dublin

The K-means Clustering Algorithm implemented here and designed to produce clusters with the minimum difference within each cluster and the maximum difference between each cluster. The dataset used for clustering in this section contains 167 observations and 20 variables relating to each settlement in Dublin and its suburbs. The variables include; distance to closest station, time it takes to each settlement when driving 80kmph, number of AEDs in each settlement, Population and the breakdown of population by sex and age intervals. All variables required transformation in the form of a z-score format because the variables consisted of different distributions. The algorithm uses Euclidean distance to calculate the distance from each data point to each K (cluster centers) and assigns the data point to the nearest cluster. The centroids are recalculated using the current class membership until the users stopping criteria is met. K is defined by the user and can be altered as many times as necessary. After testing K = 3, 4, 5 and 6 separately, K was ultimately chosen as K= 4 as it proved the best fit for the data.

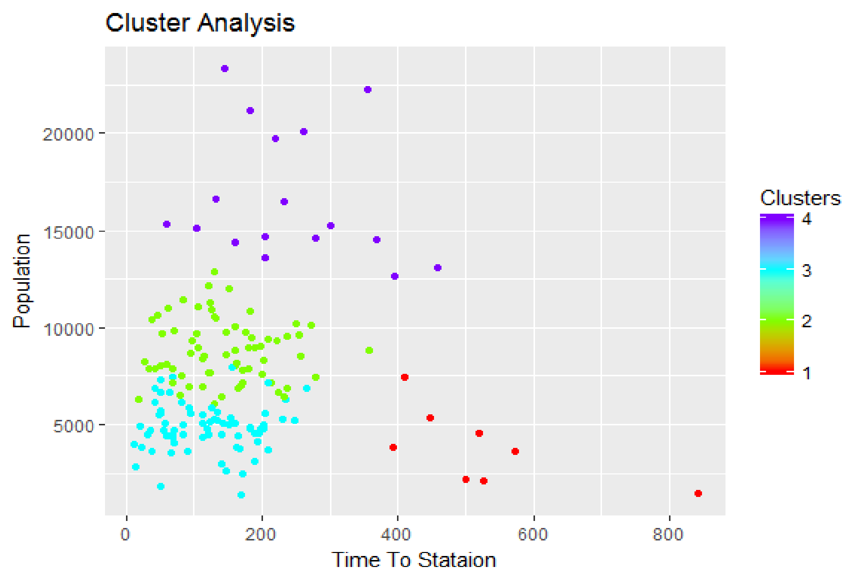


Figure 20: Population and Time to Station Clusters (Dublin)

The results of the K-means Clustering Algorithm seen on Figure 20 display four clearly formed clusters relating to population and time it takes to drive to each station when driving 80kmph (represented in seconds). Since the time it takes to drive to each settlement from the nearest station and the distance to the nearest station is highly correlated with a correlation coefficient = 0.99 (very strong positive relationship) there is no need to display the population and distance to station graph as it displays the exact same results.

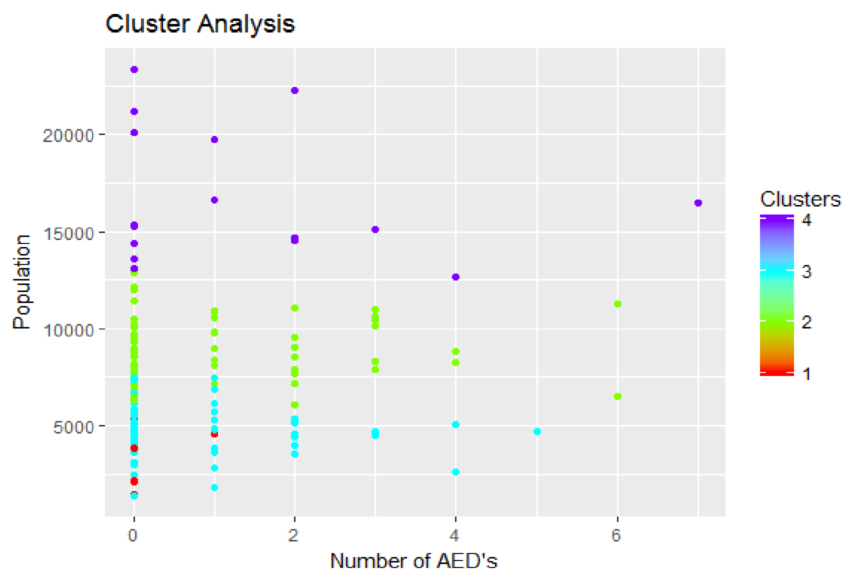


Figure 21: Population and Number of AEDs Clusters (Dublin)

The Clusters displayed in Figure 21 represent the population and number of AEDs in all settlements in Dublin. We can see Clusters 1 and 3 overlap while Clusters 2 and 4 clearly differentiate from all other clusters by population.

2.3.11.1 Clusters Conclusion (Dublin)

From examining Figure 20 and 21 the K-means Clustering Algorithm with K = 4 has grouped data based on similarity. Cluster 1 data points seen in red consist of populations less than 7,500, with a driving time to the nearest station greater than six minutes when driving 80kmph and finally with one or less available AED. Cluster 2 data points seen in green show some overlap with Cluster 2 and consist of populations ranging from 5,000 to around 12,500, AEDs ranging from none to six and less than six minutes driving time to the nearest station. Cluster 3 data points seen in light blue share a very similar population range with Cluster 1 with less than 7,500 but differ as driving time to the nearest station is less than five minutes and the number of AEDs in Cluster 3 range from zero to five. Finally, Cluster 4 data points seen in purple consists of a population greater than 12,500, a driving time to the nearest station of less than eight minutes and AEDs ranging from zero to six. I can conclude Cluster 1 data points that contain minimum AEDs and with small population and have the longest time to wait for emergency services are high priority for easily accessible AEDs. Other data points with no AEDs and large populations are also high priority areas.

2.3.12 Clusters (Rest of Leinster)

The Rest of Leinster dataset contained the identical twenty variables as the Dublin dataset but contained 317 observations. The same preparation and transformation of the dataset occurred for the Rest of Leinster as it did for Clusters in Dublin explained in section 2.3.11 in preparation for the K-means Algorithm.

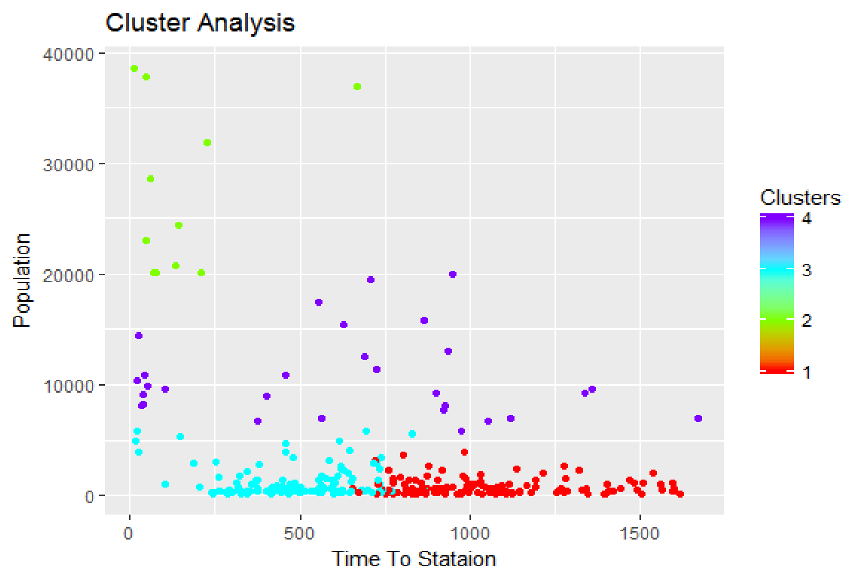


Figure 22: Population and Time to Station Clusters (Rest of Leinster)

The results of the K-means Clustering Algorithm seen on Figure 22 display four clearly formed clusters with only slight overlap between Cluster 1 and Cluster 3. The Clusters relate to population and time it takes to drive to each station when driving 80kmph (represented in seconds). Since the time it takes to drive to each settlement from the nearest station and the distance to the nearest station is highly correlated with a correlation coefficient = 0.99 (very strong positive relationship) there is no need to display the population and distance to station graph as it displays the exact same results. As was the case for the Clusters calculated in Dublin, K = 4 was chosen as it fits the data appropriately.

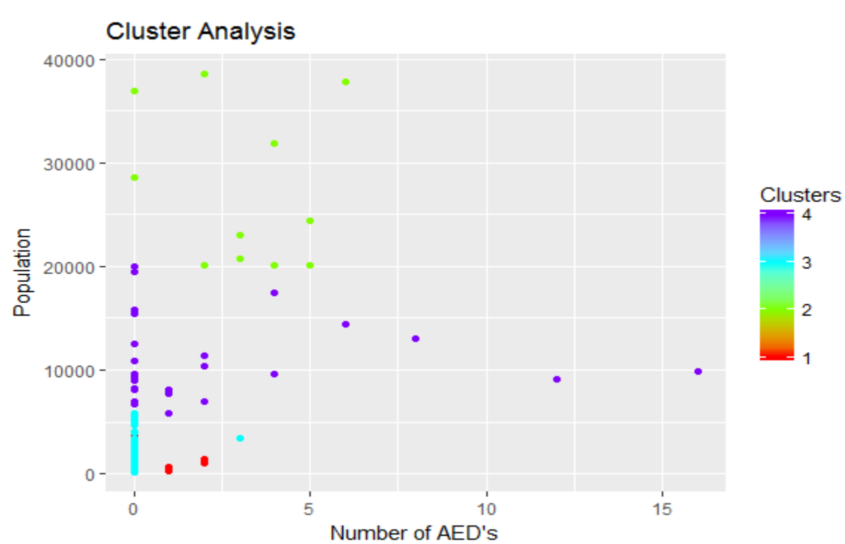


Figure 23: Population and Number of AEDs Clusters (Rest of Leinster)

The Clusters displayed in Figure 23 represent the population and number of AEDs in all settlements in Leinster except Dublin City and its suburbs. Figure 23 also shows Cluster 1 overlapping a significant amount of Cluster 3 data points while Clusters 2 and 4 clearly differentiate from all other clusters by population.

2.3.12.1 Clusters Conclusion (Rest of Leinster)

From examining Figure 22 and 23 the K-means Clustering Algorithm with K = 4 has grouped data based on similarity. Cluster 1 data points seen in red consist of populations less than 5,000, with a driving time to the nearest station greater than nine minutes when driving 80kmph and finally less than three available AEDs. Cluster 2 data points seen in green consist of populations greater than 20,000, AEDs ranging from zero to six and less than six minutes driving time to the nearest station apart from one potential outlier. Cluster 3 data points seen in light blue share a very similar population range with Cluster 1 with 6,000 or less but differ as driving time to the nearest station ranges between one minute and around eleven minutes and the number of AEDs for data points in Cluster 3 are predominately zero. Finally, Cluster 4 data points seen in purple consists of a population greater ranging from 7,500 to 20,000, and contain AEDs and driving time to nearest station ranging across the board. I can conclude Cluster 1 data points that contain minimum AEDs and with small population and consist of the average longest time to wait for emergency serves are high priority areas for easily accessible AEDs. Cluster 3 data points mainly consist of areas with no available AED’s and all areas in Cluster 2 with an estimated response time greater than eight minutes are also high priority areas. All other areas with no available AEDs are also high priority.

2.3.13 Decision Tree

The Decision Tree seen in this section is a Regression Tree built using the Rpart package in RStudio. Considering Decision Tree’s possess the ability to identify patterns in data to predict outcomes, the purpose of the Decision Tree in this case is to predict mortality rates related to the circulatory system (heart) for every settlement in Leinster. Due to the lack of available data relating to this matter, a request to the CSO on 17/04/2017 seeking to obtain said data occurred. The request was successful but only contained fifty-eight records spread over Leinster. Subsequently the obtained data, appropriately added to the previously developed data containing variables that include; distance to closest station, time it takes to each settlement when driving 80kmph, number of AEDs in each settlement, Population and the breakdown of population by sex and age intervals. The data used for the Decision Tree when cleaned (unnecessary variables were removed) contained two datasets, one with 58 records and 19 Variables including the mortality rates per settlement, used as the training/test data. The second dataset used contained 418 observations and 19 variables.

2.3.13.1 Decision Tree Training and Testing the Model

When training the model the dataset containing the mortality rates relating to the circulatory system required several splits between the training and test datasets in order examine the Cross Validation Error (XError). Testing various splits ensued and by using the printcp() and plotcp functions the Complexity Parameters of the various splits could be examined through a table and visually through a graph. Figure 24 shows the 90/10 split between the training and test datasets and the split contains the lowest XError of all splits. Figure 24 also displays the percentage of variable importance attributes when constructing the tree. Measuring the performance of the algorithm by calculating the Mean Absolute Error using a created function also occurred for every split and the lowest MAE value was derived from the 90/10 split. The MAE takes the mean of absolute differences between the actual values against the predicted values. The following function produced an MAE value of 12.22:

- MAE <- function(actual, predicted) (mean(abs(actual - predicted)))

```
          CP nsplit rel error    xerror      xstd
1 0.53708457      0 1.0000000 1.0487801 0.3134579
2 0.12924236      1 0.4629154 0.7362050 0.2101664
3 0.01973667      2 0.3336731 0.5917135 0.1811478
4 0.01000000      3 0.3139364 0.5709292 0.1707537

Variable importance
    Population     TotalMales   TotalFemales    X20_39Males    X20_39Total     X0_19Males
            18             18             14             13             13             12
X65PlusFemales    X40_64Males    X65PlusTotal   X65PlusMales    X40_64Total
             4              3              3              2              1
```

Figure 24: Complexity Parameter Table

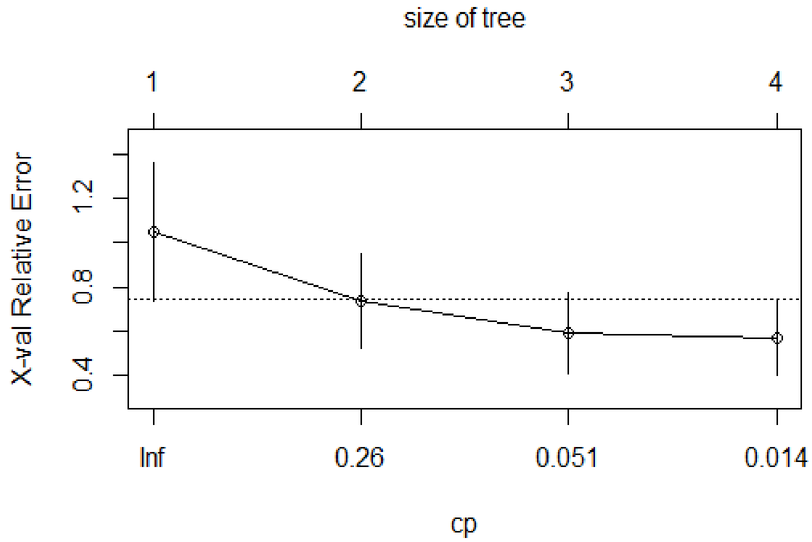


Figure 25: Complexity Parameter Table

The Plotcp() function provides a visual representation of the Cross Validated Error (XError) summary. Figure 25 shows the results of CP values plotted against the geometric mean to show the deviation until the minimum value is reached. The graph displays that the tree is most optimal at four terminal nodes and since the tree already contains four terminal nodes there is no need for Pruning. Please see Figure 26 on the next page for a visual representation and breakdown of the regression decision tree.

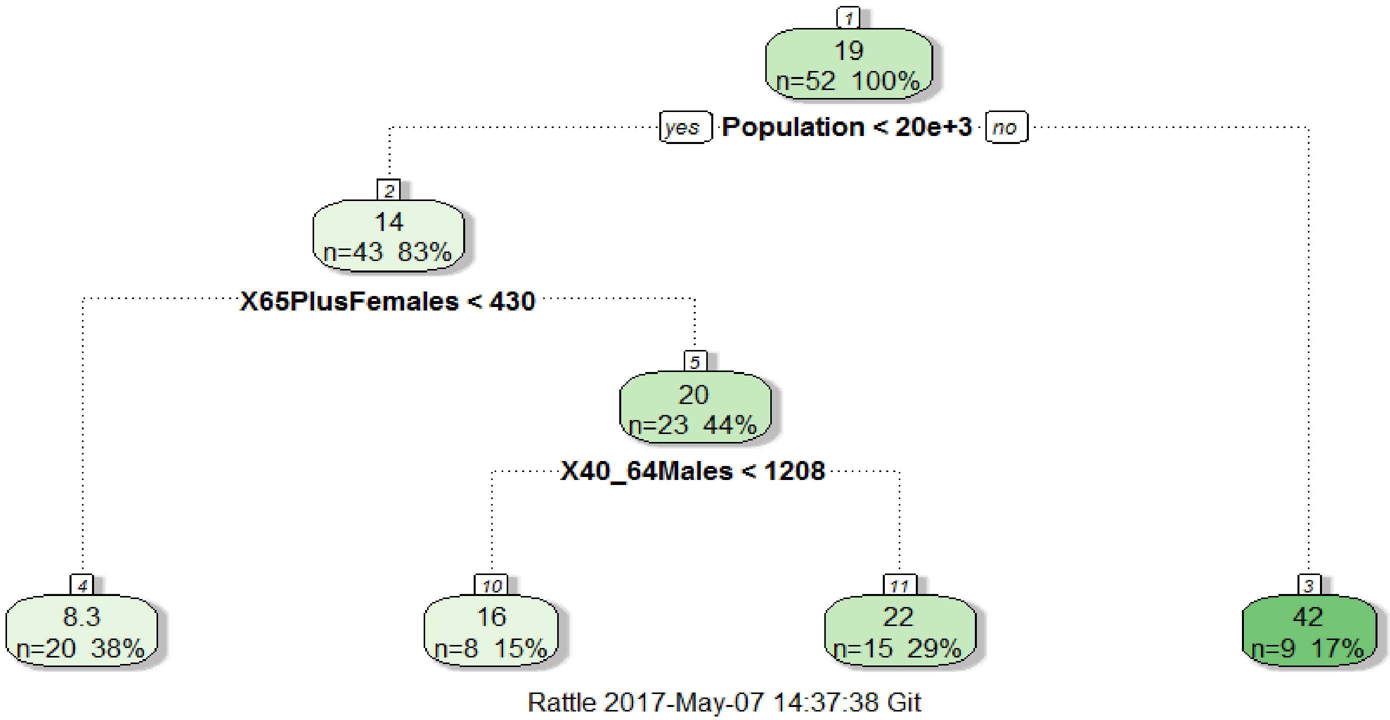


Figure 26: Regression Decision Tree

The Tree Structure shown on Figure 26 derived from the Rattle package in RStudio visually breaks down how the tree was constructed. The Tree displays how the less than operator is used to make decisions and split the data. The Root node is the first node displayed at the top of the graph and displays that the Population variable was the most contributing variable and was split by data points less than 20,000 and anything equal to or greater than 20,000. The nodes at the bottom are leaf nodes also referred to as terminal nodes and contain the predicted value and the percentage of data points within the data that meet the decision criteria. For example, we can see that 38% of the data points contain a Population less than 20,000, and consist of less than 430 Females over 65.

2.3.13.2 Augmenting the Dataset

Finally the dataset containing no Heart Related Deaths per population was augmented with the data returned by the predict function from the Rpart package. The function contained the model and the new data and returned a vector for all predicted values.

2.4 Testing

2.4.1 Unit Testing

Unit Testing is a vital step when creating a project and ensuring individual components perform as intended. Each unit test conducted in RStudio using the testthat package consists of a table that explains the purpose of the function as well as the expected and actual outcomes. See Section 2.4.2 and 2.4.3 for screen shots of the unit test and the performance and explanation of each function.

2.4.2 NearestL Function

Table 5: NearestL unit test table

Function	**Purpose**	**Expected Outcome**	**Actual Outcome**	**Solution**
NearestL	To calculate the distance from x data points (centre of every town in Leinster) to the nearest station.	Data frame containing 3 variables. The variables include the latitude and longitude of town centres and the distance to the nearest station.	As expected	N/A

The NearestL function unit test conducted in RStudio using the testthat package results displayed on Table 5 outlines the purpose of the function, the Expected and Actual outcomes. The expected outcome column details the necessary output for the function to work as intended. Figure 27 displays the output of the unit test and as no errors were recorded the function executed as expected. Figure 28 displays the actual output where tmp1 is the calculated distance to the nearest station. Finally, Figure 29 displays an unsuccessful unit test for the NearestL Function. The error message is a result of purposely editing the expect_output function in order to make sure an error message appears when incorrect output produced by the NearestL function occurs. The NearestL function is a fundamental component in achieving the projects main objectives and enables Requirement 5: Machine Learning to produce informative results.

```
> test_that("distance to station is calculated", {
+   N <- NearestL(1)
+   expect_equal(3.544374,3.544374)
+   expect_output(str(N), "1 obs. of  3 variables")
+   expect_is(earth.dist,"function")
+ })
```

Figure 27: Unit test for NearestL Function

```
> NearestL(1)
      tmp1 firecoords.Lat firecoords.Lon
1 3.544374       53.25049      -6.157074
```

Figure 28: Expected and Actual Output for NearestL Function

```
> test_that("distance to station is calculated", {
+   N <- NearestL(1)
+   expect_equal(3.544374,3.544374)
+   expect_output(str(N), "1 obs. of  2 variables")
+   expect_is(earth.dist,"function")
+ })
Error: Test failed: 'distance to station is calculated'
* str(N) does not match "1 obs. of  2 variables".
Actual value: "'data.frame':\t1 obs. of  3 variables:\n $ tmp1
 -6.16"
```

Figure 29: Error message returned from NearestL Function unit test

2.4.3 Testing Data Output

Considering the analysis throughout the project consisted of manipulating data frames and calling data from database tables and comma separated value files a function was created to test the output of a given data frame. The assertr package was used to conduct these tests in RStudio.

```
data <- read.csv("distance2FS.csv")
data %>%
  assertr::verify(nrow(.)>166)
...
```

Figure 30: Function Testing the Output of a Data Frame

Figure 30 shows the function created to test the output of a given data frame. In the above case the expected rows in the data frame is greater than 166 and no errors occurred.

```
> data %>%
+   assertr::verify(nrow(.)>167)
verification [nrow(.) > 167] failed! (1 failure)

Error: assertr stopped execution
```

Figure 31: Error from Function Testing the Output of a Data Frame

Figure 31 displays a returned error from the test. There are 167 observations in the data frame and the function was manipulated to test for more than 167 resulting in the returned error. This function proved useful throughout the project.

2.5 Evaluation and Recommendations

The results of all analysis conducted in this project have been presented and discussed thoroughly in Section 2.3. The KDD methodology, adopted and applied throughout the project lifecycle to great effect played a significant role during the entirety of the project. The workflow by adhering to the phases of the KDD methodology ensured the implementation of the project followed a structure that included selecting the appropriate data, pre-processing the data in preparation for in depth analysis and transforming the data to the correct format when running the data through specific machine learning algorithms. Finally, the analysis concluded and the results required interpretation, evaluation.

2.5.1 Key Analysis

Key analysis included implementing the use of an interactive map shown in Section 2.3.7 and plotting the shape files containing the settlement boundaries of every population in Leinster. This allowed for great visuals when plotting the AED, Fire and Ambulance Station locations. One of the main functions executed was the NearestL function. The function entailed binding the Station coordinates and centroid data frames together by the Lat/Lon variables and using the earth.dist function from the fossil package to calculate the distance from each settlement in Leinster to the nearest emergency service station. The output of the function shown in Figure 17 includes the name and population of each settlement, coordinates of each centroid and the distance to the closest station in kilometres. Calculating the distance from each population to the nearest station proved decisive when implementing the K-means Clustering Machine Learning Algorithm. The results of the K-means algorithm explained in Section 2.3.11 and Section 2.3.12 provide four Clusters with the DistanceToStation variable proving a key differentiator between the four clusters.

Following on from the Clustering analysis, a Regression Tree built to predict 418 Heart Related Deaths based of the 58 observations provided by the CSO. The total sum of the predicted Heart Related Deaths for 2011 is 5,060 and the actual number of Heart Related deaths in Leinster for 2011 is 4,440. When taking this into consideration the overall sum of overall predicted Heart Related Deaths produced from the Regression Tree is 88% accurate. The results of the regression tree were augmented to the already collected data containing the population, population breakdown, distance to the nearest emergency services station and the number of accessible automated external defibrillators in every settlement in Leinster.

2.5.2 Recommendations

Now possessing a rich dataset containing all the required information, a clear and concise decision relating to what settlements require AEDs is achievable. Taking into consideration the breakdown of the Ambulance and Fire Station Response Times Summary seen in Section 2.3.5 and 2.3.6, all the collected data that includes Population, Distance to Station, Number of AEDs, and Heart Related Deaths in every settlement in Leinster the final dataset was broken into three subsets based on certain conditions. The subsets represent the suggested areas for easily accessible automated external defibrillators by different priority levels. High Priority areas are settlements that contain less than three accessible AEDs, contain a driving time greater than seven minutes to the nearest emergency service station when driving 80 kmph, have a Population greater than 2500. Areas that contain 70% of ECHO calls responded to in over eight minutes are also considered high priority. There are 37 high priority areas. Medium Priority areas consist of Population less than 7500, AEDs less than 3, a population percentage of heart related deaths greater than 2% and a driving time greater than seven minutes to the nearest emergency service station when driving 80 kmph. There are 79 areas marked Medium Priority. Finally, Low Priority areas consist of Populations greater than 2500, a driving time greater than two minutes to the nearest emergency service station when driving 80 kmph and AED's less than 5. There are 109 areas marked Low Priority. In total, there are 458 settlements in Leinster and 225 have been marked with different Priority Levels from the results of in depth analysis conducted in RStudio using several programming techniques and the implementation of Machine Learning methods.

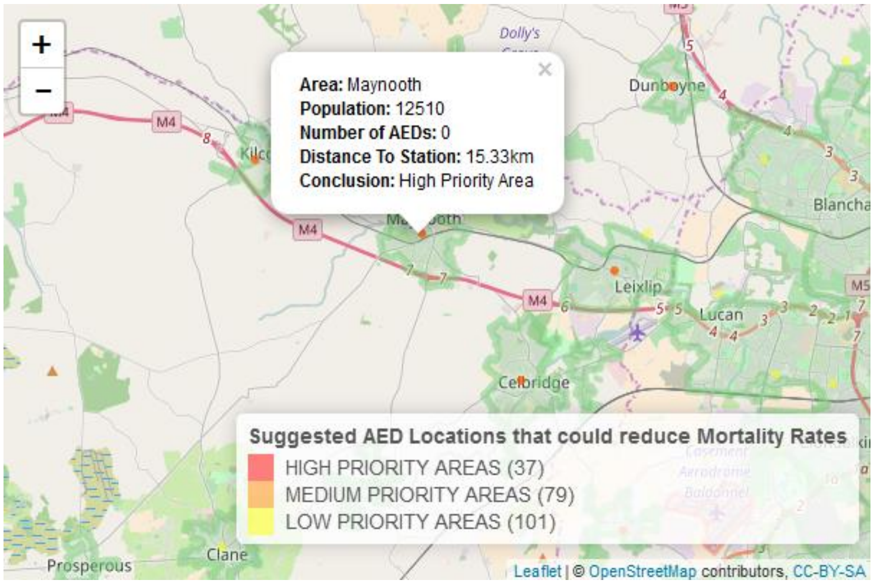


Figure 32: Recommended AED Locations by Priority

The recommendations for easily Accessible External Defibrillators shown on Figure 32 highlight Maynooth and the characteristics of the area. Maynooth consists of a relatively large population with a distance of 15.33km to the nearest emergency service station and contains no accessible AEDs. Therefore, Maynooth is a high priority area for easily accessible AEDs.

2.5.3 Key Findings

1. 49% of all settlements in Leinster are considered priority areas for the installation of easily accessible Automated External Defibrillators.
2. Only 19% of settlements in Leinster contain accessible Automated External Defibrillators.
3. The distance from the centre of half the settlements in Leinster to the nearest emergency service station is 10 kilometres or more.
4. Key Performance Indicators (KPIs) recommended by the Health Information and Quality Authority (HIQA) that Echo calls are responded to within 8 minutes for

75% of all cases are not been met by any of the Ambulance or Fire Brigade Stations.

5. A Regression Tree consisting of a Mean Absolute Error value of 12.22 and built with 88% accuracy was used to calculate 418 Heart Related deaths based of 58 provided instances.
6. A K-means clustering algorithm identified four clusters within the data. The clusters formed based of attributes consisting of Population, Distance to nearest emergency service station and number of AED's in each settlement.

2.5.4 Project Changes

When evaluating the entire project lifecycle certain analysis conducted have not been included as part of this report. For example, before the interactive map was included the AED, Fire and Ambulance Stations were plotted onto a static map. Following on from here two and four kilometre radiuses around each station provided a view of the areas a station could respond to in a matter of minutes. The radiuses also provided a visual representation of how many AEDs fell under the two and four kilometre jurisdiction of an emergency service station. Realising a more precise approach was required that not only included all AED, Ambulance and Fire Station Locations but also the characteristics of every settlement in Leinster, a different approach was undertaken.

3 Conclusions

The findings of the analysis conducted in this project outline the need for easily accessible AEDs across settlements in Leinster, Ireland. Considering there are 458 settlements in Leinster and 225 (49%) have been marked as high, medium or low priority areas and the population total for those areas is 1.2 million, a significant amount of people could benefit from the installation of easily accessible Automated External Defibrillators in the centre of every settlement. When analysing the performance of the DFB and NAS, and realising the Key Performance Indicators (KPIs) recommended by the Health Information and Quality Authority (HIQA) that Echo calls are responded to within 8 minutes for 75% of all cases are not been met by any of the 23 Ambulance Stations or 14 Fire Brigade Stations in Leinster is alarming. KPIs not be met alone is of great concern to the general public but when you factor in some of the stations respond to less than 12% of all ECHO calls within eight minutes and the response times of the NAS for settlements without an Ambulance Station is non-existent the grim reading of this analysis could be worse. Furthermore, the distance from the centre of half the settlements in Leinster to the nearest emergency service station is 10 kilometres or more. Considering all the above factors and all analysis conducted during the study it reaffirms the idea that AEDs ought to be available to the public especially in settlements identified as priority areas in this study.

The advantages of working on a project of this kind included gaining experience of having on-going correspondence with people in industry form the National Ambulance Service, Central Statistics Office and Helping Hearts website. It has provided an insight into how organizations and businesses operate when dealing with requests for sensitive data. Working on real data from Helping Hearts and Dublin Fire Brigade has also been an advantage and has provided the opportunity to analyse datasets of significance.

A major disadvantage during this study was the unsuccessful attempts to acquire significant real data from the National Ambulance Service.

4 Further development or research

Phase two of this project will entail the development of a software application that geolocates the positions of all recorded AEDs in Leinster Ireland. A Search functionality shall also be implemented to enable the user find the nearest available AED by the location of said user.

Further research could be conducted with the most recent census datasets and the most recent National Ambulance and Dublin Fire Brigade Services data to find out if the Key Performance Indicators (KPIs) recommended by the HIQA are closer to been met. Comparisons could also be made from this study against the new data to see where improvements have occurred if any.

5 References

[1] Aedireland.ie. (2016). *AED Automatic External Defibrillator Ireland*. [online] Available at: http://www.aedireland.ie/chainofsurvival.html [Accessed 11 Oct. 2016].

[2] *Cardiac First Responder Guide*. 1st ed. Health Service Executive, 2008. Web. 4 Oct. 2016.

[3]"CSO - Central Statistics Office". *Cso.ie*. N.p., 2017. Web. 8 Nov. 2016.

[4] "Development Of Intermediate Care Service (ICS) - HSE.Ie". *Hse.ie*. N.p., 2016. Web. 18 Oct. 2016.

[5] "Dublin Fire Rescue And Emergency Ambulance Service | Dublin City Council". *Dublincity.ie*. N.p., 2016. Web. 30 Sept. 2016.

[6] "Helping Hearts". *Helpinghearts.ie*. N.p., 2016. Web. 18 Nov. 2016.

[7] "Home - Dublinked". *Dublinked*. N.p., 2016. Web. 1 Oct. 2016.

[8] *National Ambulance Service Of Ireland Emergency Service Baseline And Capacity Review*. 1st ed. Lightfoot Solutions UK Ltd, 2016. Web. 26 Sept. 2016.

[9] "Pre-Hospital Emergency Care Key Performance Indicators For Emergency Response Times | Hiqa.Ie".*Hiqa.ie*. N.p., 2016. Web. 10 Nov. 2016.

[10] "Stack Overflow". *Stackoverflow.com*. N.p., 2017. Web. 5 Oct. 2016.

[11] "The National Ambulance Service - HSE.Ie". *Hse.ie*. N.p., 2016. Web. 21 Sept. 2016.

[12] *Theoretical Machine Learning*. 1st ed. Princeton: Rob Schapire, 2008. Web. 9 Nov. 2016.

[13] Walsh, Sinead. *Sineadwalshdfbincidentactivity*. 1st ed. Dublin: National College of Ireland, 2017. Web. 3 Oct. 2016.

[14] "2016 Performance Reports - HSE.Ie". *Hse.ie*. N.p., 2016. Web. 29 Nov. 2016.

[15] wark, e. "Calculate The Distance Between Two Points Of Two Datasets (Nearest Neighbor)". *Stackoverflow.com*. N.p., 2017. Web. 2 Feb. 2017.

6 Appendix

6.1 Project Proposal

6.1.1 Objectives

I plan to create a project that either verifies or rejects the idea that Automated External Defibrillators (AED) should be easily accessible to the public in every village/town across Ireland to help decrease mortality rates. My project will also consist of creating a software application that geolocates the positions of all recorded AEDs in Leinster Ireland, for the application, I will base the AED geolocations from the Helping Hearts website [3] who kindly shared their data with me.

Achieving the above goals will require analysing a considerable amount of data. For example, the analysis of data from the National Ambulance Service (NAS) who cover most of Ireland, and the Dublin Fire Brigade (DFB) who cover most parts of Dublin is crucial in gaining a true reflection when determining what towns require AEDs.

My findings will be based on several datasets recorded by the DFB and NAS for the year 2014.

6.1.2 Motivation

AEDs are expensive and ought to be distributed fairly to areas of Ireland that have an emergency service response time greater than eight minutes, AED Ireland [1] outlines that Early defibrillation to cardiac arrest sufferer's saves lives and involves trained personnel delivering an electric shock to the patient's chest to help restore the normal function of the heart. It is the link in "The Chain of Survival" that is most likely to improve survival rates. Having a defibrillator on site within 2 minutes, there can be an 80% chance of survival. Every minute that passes survival rates are reduced by 7-10%. Studies show survival rates as high as 74% can be achieved if defibrillation is administered within 3 minutes. Furthermore, if time to defibrillation is greater than ten minutes nobody survives without cardiopulmonary resuscitation (CPR) although when CPR is used survival rates increase to between 10 and 20%. Considering the above statistics, the National Ambulance Service response times in Ireland are very topical and the HSE has recently published a report of the National Ambulance Service of Ireland, Emergency Service Baseline and Capacity Review. The report was commissioned by the Health Service Executive and carried out by Lightfoot Solutions UK. The report found that 26.6 % of ambulances fail to reach the incident within eight minutes of all life threatening calls including Echo calls (Life threatening cardiac or respiratory arrest).

Table 1: ECHO/DELTA Current and Future Response Times (2014). [2]

	8 minute ECHO/DELTA				19 minute ECHO/DELTA			
	Major Urban	Minor Urban	Rural	Total	Major Urban	Minor Urban	Rural	Total
Current	36.7%	45.6%	6.6%	26.6%	87.0%	76.1%	44.0%	67.2%
Future	85%	73.2%	32.9%	60.6%	95.5%	84.1%	69.3%	81.5%

When considering having a defibrillator on site within 2 minutes, there can be an 80% chance of survival table 1 is grim reading for rural Ireland where only 6.6% of the population have a response time within eight minutes.

6.1.3 Background

The National Ambulance Service is the statutory public ambulance service in Ireland. The service is operated by the National Hospitals Office of the Health Service Executive (HSE). Up until the HSE was established by the Health Act, 2004 and came into operation on January 1st, 2005 Irelands Ambulance Service were run by ten regional Health Boards. The NAS was established as part of the HSE in 2005 consisting of 94 Ambulance Stations and a staff of 1400.

Dublin Fire Brigade is Irelands largest full-time brigade and employs over 1000 people. There are fourteen stations twelve of which are full-time and two are part-time (retained). DFB operates an Emergency Ambulance Service as well as a Fire, rescue and emergency service for Dublin City and County.

Helping Hearts was founded in 2001 and provides CPR and First AID Training Courses to the general public across Ireland. Helping Hearts is based in Galway.

The Health Information and Quality Authority (HIQA) published Pre-Hospital Emergency Care Key Performance Indicators o Emergency Response times in January 2011. One of the Key Performance Indicators (KPIs) being recommended by the HIQA was that Echo calls (Life threatening cardiac or respiratory arrest) are responded to within 8 minutes for 75% of all cases.

6.1.4 Technical Approach

Data Analysis:

When analyzing the data, I will use the core data analytics methodology: knowledge discovery in databases (KDD). KDD has a series of fundamental stages that require serious attention;

- **Selection:** The selection stage focuses on acquiring the data set or data sets that will be analyzed to discover useful information relating to the chosen topic.
- **Pre-Processing:** This stage requires the cleansing of data. Data cleansing is the process of detecting, correcting or removing corrupt or inaccurate data records from a dataset or database.
- **Transformation:** The transformation stage requires methods that include dimension reduction and record sampling.
- **Data Mining:** The data mining stage searches for data patterns that may be of interest. This stage also extracts the said data patterns.
- **Interpretation/evaluation:** The final stage is based on the knowledge discovered and how it was interpreted.

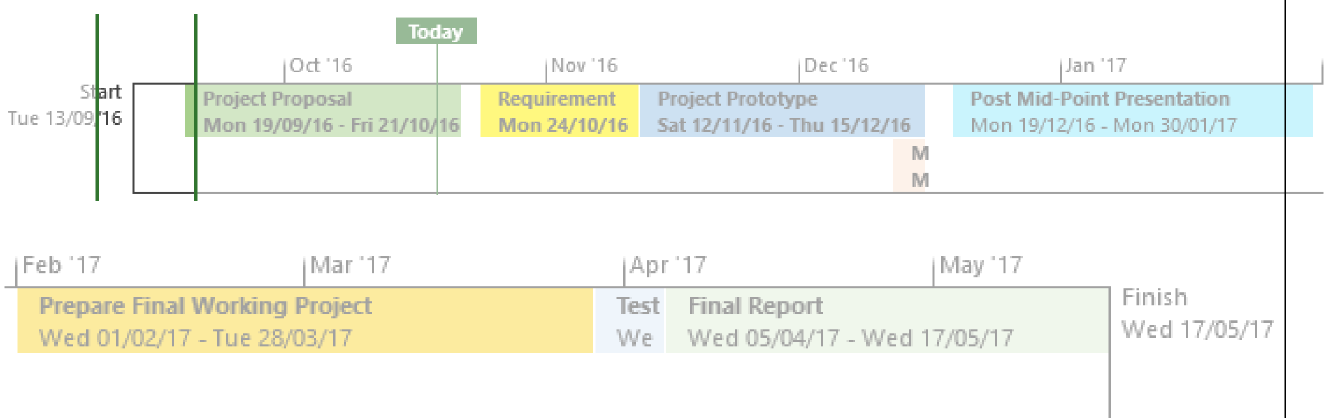
Software Application:

I plan to develop the software application for my project in Android Studio using java programming language. The application will consist of a map of Leinster, Ireland where I will input the co-ordinates of all AEDs recorded by Helping Hearts website. I will investigate if a search feature where the user will choose a town/village from a menu and the distance to the nearest AED will be shown back to the user.

6.1.5 Special resources required

As of this moment I require no special resources.

6.1.6 Project Plan



Task Mode	Task Name	Duration	Start	Finish	Predecessors	Resource Names
Manually Scheduled	**Project Proposal**	**25 days**	**Mon 19/09/16**	**Fri 21/10/16**		
Auto Scheduled	Brainstorming Project Ideas	5 days	Mon 19/09/16	Fri 23/09/16		
Auto Scheduled	Email Lecturer for feedback on Idea	1 day	Mon 26/09/16	Mon 26/09/16	2	
Auto Scheduled	Project Pitch	1 day	Wed 05/10/16	Wed 05/10/16		
Manually Scheduled	First Monthly Journal Entry	1 day	Fri 07/10/16	Fri 07/10/16		
Auto Scheduled	Project Proposal Document	12 days	Thu 06/10/16	Fri 21/10/16	4	
Manually Scheduled	Meeting with Supervisor	1 day	Tue 13/09/16	Tue 13/09/16		
Manually	Meeting with Supervisor	1 day	Tue	Tue		

Scheduled			20/09/16	20/09/16		
Manually Scheduled	Proof read Project Proposal Document and Upload	1 day	Fri 21/10/16	Fri 21/10/16		
Manually Scheduled	**Requirement Specification**	**15 days**	**Mon 24/10/16**	**Fri 11/11/16**		
Auto Scheduled	Acquire all data sets for the project	5 days	Mon 24/10/16	Fri 28/10/16		
Auto Scheduled	Requirement Specification Document	15 days	Mon 24/10/16	Fri 11/11/16		
Manually Scheduled	Meeting with Supervisor	1 day	Thu 03/11/16	Thu 03/11/16		
Manually Scheduled	Requirements Specification Review & Upload	1 day	Fri 11/11/16	Fri 11/11/16		
Auto Scheduled	Second Monthly Journal Entry	1 day	Fri 04/11/16	Fri 04/11/16		
Manually Scheduled	**Project Prototype**	**25 days**	**Sat 12/11/16**	**Thu 15/12/16**		
Manually Scheduled	Data Cleansing	5 days	Mon 07/11/16	Fri 11/11/16		
Manually Scheduled	Start to create prototype for Application in Android Studio	14 days	Sat 12/11/16	Wed 30/11/16		
Manually Scheduled	Meeting with Supervisor	1 day	Thu 17/11/16	Thu 17/11/16		
Manually Scheduled	Third Monthly Report	1 day	Fri 09/12/16	Fri 09/12/16		
Manually Scheduled	**Mid-Point Presentation**	**4 days**	**Mon 12/12/16**	**Thu 15/12/16**		
Manually Scheduled	Meeting with Supervisor	1 day	Thu 15/12/16	Thu 15/12/16		
Manually Scheduled	Finish Prototype and Prepare for Presentation					
Manually Scheduled	**Post Mid-Point Presentation**	**31 days**	**Mon 19/12/16**	**Mon 30/01/17**		
Manually Scheduled	Process feedback from mid-point presentation and start to make alterations	5 days	Mon 19/12/16	Fri 23/12/16		
Manually Scheduled	Christmas Break	5 days	Sat 24/12/16	Thu 29/12/16		
Manually Scheduled	Fourth Monthly Report	1 day	Fri 30/12/16	Fri 30/12/16		
Manually Scheduled	Continue to develop prototype	10 days	Sun 01/01/17	Thu 12/01/17		
Manually Scheduled	Meeting with Supervisor	1 day	Thu 26/01/17	Thu 26/01/17		
Manually Scheduled	Analyse Datasets and arrive to conclusions	12 days	Fri 13/01/17	Mon 30/01/17		

Manually Scheduled	**Prepare Final Working Project**	**40 days**	**Wed 01/02/17**	**Tue 28/03/17**		
Manually Scheduled	Design my Application	5 days	Wed 01/02/17	Tue 07/02/17		
Manually Scheduled	Fifth Monthly Report	1 day	Fri 03/02/17	Fri 03/02/17		
Manually Scheduled	Meeting with Supervisor to discuss final stages and implementations	1 day	Thu 09/02/17	Thu 09/02/17		
Manually Scheduled	Finish Analysis of datasets	17 days	Wed 08/02/17	Thu 02/03/17		
Manually Scheduled	Completion of application and data analysis of data-sets	14 days	Thu 09/03/17	Tue 28/03/17		
Manually Scheduled	**Testing**	**5 days**	**Wed 29/03/17**	**Tue 04/04/17**		
Manually Scheduled	Create and Implement Test Plans for final Application	3 days	Wed 29/03/17	Fri 31/03/17		
Manually Scheduled	Create and Implement Test Plans for findings in data-sets	3 days	Sat 01/04/17	Tue 04/04/17		
Manually Scheduled	**Final Report**	**31 days**	**Wed 05/04/17**	**Wed 17/05/17**		
Manually Scheduled	Prepare Final Document					
Manually Scheduled	**Upload Software**	**1 day**	**Wed 17/05/17**	**Wed 17/05/17**		
Manually Scheduled						

6.1.7 Technical Details

R:

R is a programming language and environment for statistical computing and graphics. I will use R to conduct in depth analysis with the target data sets that will analyse Automated External Defibrillators (AED) distribution throughout Leinster.

RSutdio:

RStudio is an Integrated development environment (IDE) for R programming language. I will use R in RStudio to create scripts to analyse the data sets.

Java:

Java is a programming language used to create applications that can run in various environments. Java is the programming language I will use to develop my software application in Android Studio.

Android Studio:

Android Studio is an IDE for the development of android Applications. Android studio is the platform I will use to develop the software application that geolocates the recorded available AEDs in Leinster. The application will be developed using the java programming language.

Tableau:

Tableau is a data visualization software application that allows the user to represent their findings in a clear and concise way. Once I have completed the analysis of my data sets in RStudio using R programming language I will use Tableau to represent my findings.

Microsoft Excel:

Microsoft Excel is a spreadsheet program created by Microsoft. The data sets I have acquired are in comma separated values file (.csv) format. I will use excel to play initially pre-process the data and eventually loading the data into a MySQL database.

SQL:

Structured Query Language is a computer programming language for querying relational databases systems. I will use SQL to query the relational database I create.

MySQL:

MySQL is an open source relational database management system that is based around the (SQL). I will use MySQL as my platform to create a database that allows me to use SQL to perform various queries on to analyze the data.

6.1.8 Evaluation

I will implement several test plans throughout the project and upon completion of final products. I will follow the main testing methodologies when implementing my test plans.

- Unit Testing: testing individual components of my software.
- Integration Testing: Testing of how all the components perform when integrated.
- System Testing: The phase that tests the system as a whole for the first time in the project lifecycle.
- Acceptance Testing: During the acceptance phase the finish product will be tested by external sources or the intended end user to determine whether the product is ready for distribution.

Christopher Doran x12724145

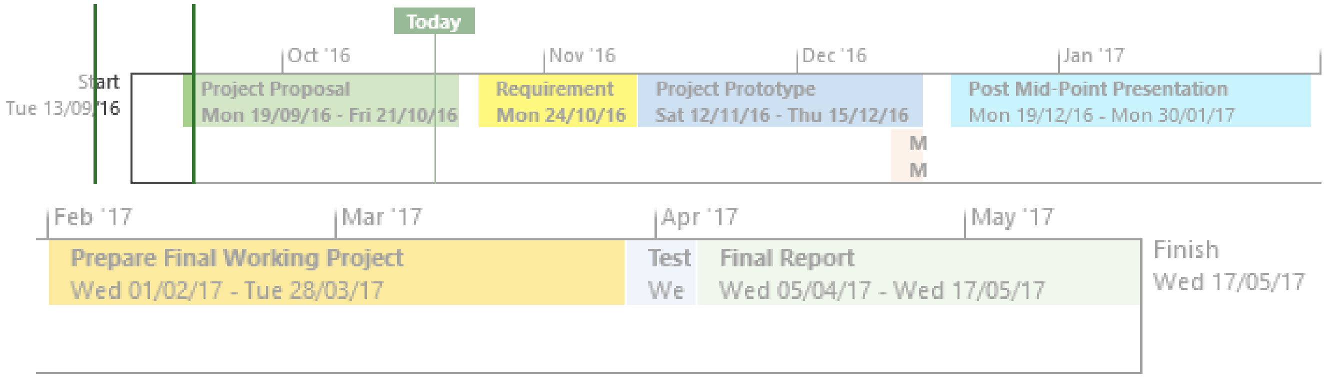
6.1.9 References

[1] Aedireland.ie. (2016). *AED Automatic External Defibrillator Ireland*. [online] Available at: http://www.aedireland.ie/chainofsurvival.html [Accessed 11 Oct. 2016].

[2] *National Ambulance Service Of Ireland Emergency Service Baseline And Capacity Review*. 1st ed. Lightfoot Solutions UK Ltd, 2016. Web. 26 Sept. 2016.

[3] Helpinghearts.ie. (2016). *Helping Hearts*. [online] Available at: http://helpinghearts.ie/ [Accessed 28 Sep. 2016].

6.2 Project Plan



Task Mode	Task Name	Duration	Start	Finish	Predecessors	Resource Names
Manually Scheduled	**Project Proposal**	**25 days**	**Mon 19/09/16**	**Fri 21/10/16**		
Auto Scheduled	Brainstorming Project Ideas	5 days	Mon 19/09/16	Fri 23/09/16		
Auto Scheduled	Email Lecturer for feedback on Idea	1 day	Mon 26/09/16	Mon 26/09/16	2	
Auto Scheduled	Project Pitch	1 day	Wed 05/10/16	Wed 05/10/16		
Manually Scheduled	First Monthly Journal Entry	1 day	Fri 07/10/16	Fri 07/10/16		
Auto Scheduled	Project Proposal Document	12 days	Thu 06/10/16	Fri 21/10/16	4	
Manually Scheduled	Meeting with Supervisor	1 day	Tue 13/09/16	Tue 13/09/16		
Manually Scheduled	Meeting with Supervisor	1 day	Tue 20/09/16	Tue 20/09/16		
Manually Scheduled	Proof read Project Proposal Document and Upload	1 day	Fri 21/10/16	Fri 21/10/16		
Manually Scheduled	**Requirement Specification**	**15 days**	**Mon 24/10/16**	**Fri 11/11/16**		
Auto Scheduled	Acquire all data sets for the project	5 days	Mon 24/10/16	Fri 28/10/16		
Auto Scheduled	Requirement Specification Document	15 days	Mon 24/10/16	Fri 11/11/16		
Manually Scheduled	Meeting with Supervisor	1 day	Thu 03/11/16	Thu 03/11/16		
Manually Scheduled	Requirements Specification Review & Upload	1 day	Fri 11/11/16	Fri 11/11/16		
Auto Scheduled	Second Monthly Journal Entry	1 day	Fri 04/11/16	Fri 04/11/16		
Manually Scheduled	**Project Prototype**	**25 days**	**Sat 12/11/16**	**Thu 15/12/16**		
Manually	Data Cleansing	5 days	Mon	Fri		

Scheduled			07/11/16	11/11/16		
Manually Scheduled	Start to create prototype for Application in Android Studio	14 days	Sat 12/11/16	Wed 30/11/16		
Manually Scheduled	Meeting with Supervisor	1 day	Thu 17/11/16	Thu 17/11/16		
Manually Scheduled	Third Monthly Report	1 day	Fri 09/12/16	Fri 09/12/16		
Manually Scheduled	**Mid-Point Presentation**	**4 days**	**Mon 12/12/16**	**Thu 15/12/16**		
Manually Scheduled	Meeting with Supervisor	1 day	Thu 15/12/16	Thu 15/12/16		
Manually Scheduled	Finish Prototype and Prepare for Presentation					
Manually Scheduled	**Post Mid-Point Presentation**	**31 days**	**Mon 19/12/16**	**Mon 30/01/17**		
Manually Scheduled	Process feedback from mid-point presentation and start to make alterations	5 days	Mon 19/12/16	Fri 23/12/16		
Manually Scheduled	Christmas Break	5 days	Sat 24/12/16	Thu 29/12/16		
Manually Scheduled	Fourth Monthly Report	1 day	Fri 30/12/16	Fri 30/12/16		
Manually Scheduled	Continue to develop prototype	10 days	Sun 01/01/17	Thu 12/01/17		
Manually Scheduled	Meeting with Supervisor	1 day	Thu 26/01/17	Thu 26/01/17		
Manually Scheduled	Analyse Datasets and arrive to conclusions	12 days	Fri 13/01/17	Mon 30/01/17		
Manually Scheduled	**Prepare Final Working Project**	**40 days**	**Wed 01/02/17**	**Tue 28/03/17**		
Manually Scheduled	Design my Application	5 days	Wed 01/02/17	Tue 07/02/17		
Manually Scheduled	Fifth Monthly Report	1 day	Fri 03/02/17	Fri 03/02/17		
Manually Scheduled	Meeting with Supervisor to discuss final stages and implementations	1 day	Thu 09/02/17	Thu 09/02/17		
Manually Scheduled	Finish Analysis of datasets	17 days	Wed 08/02/17	Thu 02/03/17		
Manually Scheduled	Completion of application and data analysis of data-sets	14 days	Thu 09/03/17	Tue 28/03/17		
Manually Scheduled	**Testing**	**5 days**	**Wed 29/03/17**	**Tue 04/04/17**		
Manually Scheduled	Create and Implement Test Plans for final Application	3 days	Wed 29/03/17	Fri 31/03/17		

Manually Scheduled	Create and Implement Test Plans for findings in data-sets	3 days	Sat 01/04/17	Tue 04/04/17		
Manually Scheduled	**Final Report**	**31 days**	**Wed 05/04/17**	**Wed 17/05/17**		
Manually Scheduled	Prepare Final Document					
Manually Scheduled	**Upload Software**	**1 day**	**Wed 17/05/17**	**Wed 17/05/17**		
Manually Scheduled						

6.3 Monthly Journals

6.3.1 September

Student name: Christopher Doran x12724145

Programme: BSc in Computing (Data Analytics)

Month: September

My Achievements

Initially I was quite worried coming back to NCI for my fourth and final year as I didn't quite have an idea for my final year software project. Once we were informed about the project pitch, a process where students had to pitch their idea to computing lecturers it gave me the kick start I needed to turn my thoughts into a solid idea suitable for a level 8 honours degree in computer science. Thankfully I acted quite quickly and I concentrated on gathering enough information to realize I had a solid idea and the information/data I required was out there and accessible.

My Reflection

The project pitch itself went well and I was glad to have my idea accepted. I had positive feedback from the three lectures that were present relating to my idea and how I expect to implement it.

Intended Changes

For the month of October, I will be preparing my project proposal and I will hope to have all the data sets I require for my project.

Supervisor Meetings

Date of Meeting: To be confirmed.

Items discussed: To be confirmed.

Action Items: To be confirmed.

6.3.2 October

Student name: Christopher Doran x12724145

Programme: BSc in Computing (Data Analytics)

Month: October

My Achievements

Once my idea was accepted I started preparing my project proposal that was due on the 21st of October. Preparing the project proposal was interesting as it really made me sit down and think about what my project was about, what direction I could take it and what technologies I would use to fulfil the project goals. Creating a gaant chart for my project plan also made me realise that setting goals and deadlines throughout my project timeline will be essential come May 2017.

During the month of October, I also acquired two very important data sets. Immediately I started cleaning my data sets especially the one I received from helping hearts.ie. The data set consists of all the recorded locations of Automated External Defibrillators (AED's) on helping hearts.ie. Although the data set had the addresses it did not have the latitude and longitude of those addresses. I used Visual Basic in Excel to install a macro that connected to the google server and when in excel I used the installed functions to generate the latitude and longitude of all my addresses.

My Reflection

Although I am happy with my progress so far one elusive data set remains a priority that I hope to acquire in the next few weeks.

Intended Changes

For the month of November, I will be continuing to play with my data sets and creating my technical document.

Supervisor Meetings

Date of Meeting: 20-10-16

My supervisor and I discussed my project proposal deliverable. He also tasked me with finding the average number of AED's per kilometer in a particular area. I am currently working on this and hope to have a solution for our next meeting: 10-11-16.

6.3.3 November

Student name: Christopher Doran x12724145

Programme: BSc in Computing (Data Analytics)

Month: November

My Achievements

During the month of November, I worked on the Requirements Specification and started to construct my Technical report that requires submission before the mid-point presentation. The technical report consists of the project proposal, requirements specification and project plan and is a living document that may change throughout the lifespan of the project.

I also started to create a prototype to demonstrate at the mid-point presentation.

My Reflection

I'm feeling the time constraints at this stage as a lot of work went into the Requirements Specification and the Technical Report and a prototype still has to be created and presented.

Intended Changes

I plan to start developing the project into a working prototype and the project will start to takes shape in the coming months so I expect to see considerable changes.

Supervisor Meetings

Date of Meeting: 24-11-16

My supervisor and I discussed the Requirements Specification Document.

6.3.4 December

Reflective Journal

Student name: Christopher Doran x12724145

Programme: BSc in Computing (Data Analytics)

Month: December

My Achievements

During the month of December, I completed the Requirements Specification and constructed my Technical report and submitted both for the mid-point presentation. The technical report consists of the project proposal, requirements specification and project plan and is a living document that may change throughout the lifespan of the project.

I also created a prototype to demonstrate at the mid-point presentation.

My Reflection

There was ultimate relief after submitting both the requirements specification and technical report but subsequently it left little or no time to perfect the prototype for a 1:1 standard. Ultimately I had to make time which was not easy as there were several other project deliverables in other modules all at similar times. I received a grading of 68.8 which I am still mulling over, one hand it is a good mark but upon deep reflection I could have easily produced a 1:1 standard given even a few extra days.

Intended Changes

Plenty of work is to be done as the fun starts in semester two, I will actually have a considerable amount of time to work on my project alone (who'd have thought that giving the students time to work on their projects would be a good idea☺).

Supervisor Meetings

Date of Meeting: 7-12-16

My supervisor and I discussed preparing for technical document upload and mid-point presentation.

6.3.5 January

Reflective Journal

Student name: Christopher Doran x12724145

Programme: BSc in Computing (Data Analytics)

Month: January

My Achievements

As the month of January consisted of sitting four exams it is safe to say the project took a back seat for the first half of the month. Once the exams concluded I re-assessed my project goals and got back to action.

I successfully extracted required data from several NAS Data management reports to fulfil the projects data requirements.

I started to explore several machine learning algorithms with the idea of gaining unseen insights from the datasets.

Augmented the DFB dataset with station AED locations.

My Reflection

I am optimistic with fulfilling the projects objectives with considerable more time this semester than the previous semester.

Supervisor Meetings

Date of Meeting: 02-02-17

My supervisor and I discussed Machine Learning Algorithms.

6.3.6 February

Reflective Journal

Student name: Christopher Doran x12724145

Programme: BSc in Computing (Data Analytics)

Month: February

My Achievements

February consisted of getting stuck into certain aspects of the project for example I implemented a new interactive map that is much more intuitive than the previous map developed which enabled me to apply 2, 4 and 6km radius zones around each fire and ambulance stations. The interactive map also enabled me to label data points and zoom in and out when required. I also collected significantly more data to make my data richer, i.e. collected the population and co-ordinates of towns in Leinster. I also plotted all the response times against the respective station using ggplot to visually examine the data. Finally, I manipulated the DFB and NAS attributes to prepare it for certain algorithms.

My Reflection

Time is becoming a precious entity so from now on I expect to see the projecting piecing together and hopefully I will start to come to final conclusion.

Supervisor Meetings

Date of Meeting: 06-02-17/20-02-17

My supervisor and I discussed Machine Learning Algorithms in depth and talked about running the data through various algorithms to find specific information out.

6.3.7 March

Reflective Journal

Student name: Christopher Doran x12724145

Programme: BSc in Computing (Data Analytics)

Month: March

My Achievements

In the month of March I was able to extend on the leaflet interactive map that was created in February. I found various shape files on CSO.ie that enabled me to plot all settlement town/village boundaries in Ireland. This is a cool new feature that will help me loop through every settlement when building machine learning algorithms. I also designed a decision tree with 100% accuracy that predicts what towns required AED's. The decision was based on numerous variables. Although the decision tree was 100% accurate it was based only on towns the NAS provided data on. When deciding AED locations for the rest of Leinster a new algorithm will be built with the help of the data within the settlement shape files.

My Reflection

Lots done Lots more to do

Supervisor Meetings

Date of Meeting: 22-03-17/20-02-17

My supervisor and I discussed Machine Learning Algorithms in depth and talked about how I will implement main algorithm that decides whether a town or village in leinster is in need of an accessible AED.