**National College of Ireland**

**BSc(Hons) in Computing(Data Analytics)**

**2016/2017**

Identifying new Health Centre Locations in Ireland





Technical Report

**Student: Declan Barnes**

**Email: Declan.Barnes@student.ncirl.ie**

**Student No: x13114352**

**Supervisor: Paul Stynes**

Table of Contents

[Identifying new Health Centre Locations in Ireland 1](file://localhost/Users/declanbarnes/Documents/FinalYearProject/Technical%20Report.docx#_Toc469078211)

[Executive Summary 4](#_Toc469078212)

[1 Introduction 5](#_Toc469078213)

[1.1 Background 5](#_Toc469078214)

[1.1.1 Research 7](#_Toc469078215)

[1.2 Aims 8](#_Toc469078216)

[1.3 Technologies 9](#_Toc469078217)

[1.4 Structure 10](#_Toc469078218)

[2 System 11](#_Toc469078219)

[2.1 Functional Requirements 11](#_Toc469078220)

[2.1.1 Output Requirements 11](#_Toc469078221)

[2.1.2 User requirements 12](#_Toc469078222)

[2.1.3 Data requirements 12](#_Toc469078223)

[2.1.4 Use Case Diagram- Overall System 13](#_Toc469078224)

[2.1.5 R1: Identify Health Centre locations 14](#_Toc469078225)

[2.1.6 R2: Identify locations of people with specific disabilities 16](#_Toc469078226)

[2.2 Non-Functional Requirements 18](#_Toc469078227)

[2.2.1 Performance/Response time requirement 18](#_Toc469078228)

[2.2.2 Availability requirement 18](#_Toc469078229)

[2.2.3 Recover requirement 18](#_Toc469078230)

[2.2.4 Robustness requirement 18](#_Toc469078231)

[2.2.5 Security requirement 18](#_Toc469078232)

[2.2.6 Reliability requirement 18](#_Toc469078233)

[2.2.7 Maintainability requirement 19](#_Toc469078234)

[2.2.8 Extendibility requirement 19](#_Toc469078235)

[2.2.9 Reusability requirement 19](#_Toc469078236)

[2.2.10 Resource utilization requirement 19](#_Toc469078237)

[2.3 Design and Architecture 19](#_Toc469078238)

[2.4 System Evolution 21](#_Toc469078239)

[2.5 Implementation 22](#_Toc469078240)

[2.6 Graphical User Interface (GUI) Layout 22](#_Toc469078241)

[2.7 Testing(TO BE DONE AT A LATER STAGE) 24](#_Toc469078242)

[2.8 Customer testing(TO BE DONE AT A LATER STAGE) 24](#_Toc469078243)

[2.9 Evaluation(TO BE DONE AT A LATER STAGE) 24](#_Toc469078244)

[3 Conclusions 26](#_Toc469078245)

[4 Further development or research 27](#_Toc469078246)

[5 Motivational Section 28](#_Toc469078247)

[6 Bibliography 29](#_Toc469078248)

[7 Appendix 30](#_Toc469078249)

[7.1 Project Proposal 30](#_Toc469078250)

[7.2 Project Plan 41](#_Toc469078251)

[7.3 Monthly Journals 46](#_Toc469078252)

[7.4 Other Material Used(TO BE DONE AT A LATER STAGE) 49](#_Toc469078253)

# Executive Summary

This report describes the analysis of data to identify health centre locations based on the clustering of disabilities in Ireland.

This analysis is useful for executives in Health Service Executive. It provides them with planning details to optimise where their resources should be used across Ireland to provide better coverage to people with a disability.

For the purpose of this project, new HSE support centres locations will be identified using machine learning, while also predicting what equipment each support centre should have. Users will also be able to see a breakdown of disabilities such as: the types of disabilities(Physical, hearing loss, blindness etc.) the locations of these disabilities shown using a graphical representation.

# Introduction

A major concern in the HSE is finding suitable locations for support centres in the country where there is a lack of disability support services due to budget cuts. For example, John Dolan, CEO of DFI stated on the disability federation website (John Dolan, 2015, www.DisabilityFederation.ie) that cuts of €159 million since 2008 to disability services has left a gaping hole in the support services across the country, disability services such as speech therapy for those who have the likes of down syndrome, physical therapy for those with physical disabilities, learning support for those with intellectual disabilities have all been affected. It is for this reason that the author of this report has chosen to try to identify the locations of new support centres in Ireland, so that more services for people with a disability become available especially in rural areas where the lack of services is worse (NDA, www.NDI.ie). These new health centres will be placed using machine learning to predict the new locations where the centres will be of the most value to those in need of disability support services. As well as locating new health centre locations there will also be predictive analysis conducted with these support centres to predict what necessary equipment these centres will need. Also, as part of the above goals, is the showing relevant information on disabilities in Ireland such as showing the amount of male and females in Ireland who suffer from some form of disability, showing the different age groups suffering with disabilities, and the types of disabilities (Intellectual, physical, hearing loss, blindness, cognitive etc.).

## Background

It is important to try to understand or work out where there are limitations in the services provided to disabled people across Ireland and what’s the best locations for the new HSE support centres and where they will have the most benefit. It is the hope that by doing this project it will show these limitations in a constructive way that will help improve the quality of the services that disabled people rely on daily for their everyday needs. It would be beneficial overall to show where new health centres should be positioned in Ireland as it is unacceptable to expect disabled people in rural areas specifically, to travel a significant distance to get to their nearest support centre, while these same support centres could be failing to provide adequate services/coverage to the people who need them by being out of reach. On the disability-federation website it states the following in regards to disabilities (DFI, 2015, www.DFI.ie): People in Ireland that are living with a disability and parents who have special needs children are at a much higher risk of suffering depression than those who do not. It was realised after reading this that the very idea of predicting new health centres would be a beneficial idea to those disabled people and the parents of special needs children as these health centres could provide very important care and or support to these people to help relieve the pressure that they live with daily by creating more health centres that are much better placed to provide the necessary coverage across the country that they need. Another interesting fact that was noticed through research was the attitudes towards people of a disability. On the national disability authority, it stated that negative attitudes towards disabled people had become a significant barrier to the inclusion of the disabled into various aspects of society and therefore preventing people with a disability from reaching their full potential. The National Disability Authority has conducted a series of surveys at regular intervals on the public’s attitude towards people living with a disability (2001,2006,2011). The NDA’s surveys showed that attitude towards disabled people improved in the years 2001 and 2006, but the year 2011 showed a decrease in attitudes.

### Research

The research conducted involved searching online to see if there is anything similar on the market. After searching online for an extended period, it became apparent that while there was similar analysis done on disabilities in Ireland by government run organisation such as the HSE, there was no machine learning implemented that would identify the locations of where new HSE support centres could be placed in the country. After researching for several more days, it was clear that the machine learning aspect to this project had not being done before by other people or organisations specifically in relation to Ireland.

## Aims

The **aim** of identifying the locations of new HSE support centres in Ireland is to acquire the necessary information needed to complete the project. Several datasets will contain key information such as the description of the support centre, its latitude and longitude. Further data will also be acquired for the prediction of equipment that the support centres should have. The analysis aspect will also need data to display the findings, this data will be like the following: how many people are living with a disability in Ireland, what type of disability do they have, what sex and age are they. The information that will be used will come from the datasets mentioned above that were acquired from public resources such as the central statistics office.

The **first objective** will take the above-mentioned information into account of where current health centres are located and predict where new centres should be built i.e. in areas where the nearest health centre is too far for disabled people to travel to, and then it will be shown in a graphical representation where the new centre should be placed in relation to this information.

The **second objective**, which involves using predictive analysis to predict what necessary equipment is needed at each support centre location. This will ensure that the new centres will have the equipment needed to provide all the services required in each area the support centre covers.

The information that will be shown will be used to represent the key analysis of the disabilities, for example: how many disabled people require homecare in Ireland, how many people with disabilities are not working due to their disability, a graphical representation shown the type of disabilities across Ireland and more.

## Technologies

The technologies that will be used are MySQL, the R programming language, and java. MySQL will be used to store the datasets containing all the variables needed to complete the research that is being set out. The R programming language will be used to create scripts that will identify the health centre locations through machine learning. R will also produce the necessary analysis by pulling the data from the MySQL database and performing data manipulation and data cleaning, Finally, data visualisation will be used to visualize the findings with charts through a web application.

RStudio:

RStudio is a free and open-source integrated development environment (IDE) for R, a programming language for statistical computing and graphics. RStudio also allows the use of packages, each package helps the user perform certain tasks. Some of the packages that will be used for this project are:

* **RMySQL:** database interface and MySQL driver for R.
* **TidyR:** Used for data manipulation and cleaning.
* **Stringr&Magrittr:** Data cleaning.
* **Dplyr:** Also, used for data manipulation, usually used in conjunction with TidyR.
* **GoogleVis:** Data Visualisation.
* **Sqldf:** Provides an easy way to perform SQL selects on R data frames.
* **ggplot2:** Used for Visualizing data.

MySQL:

MySQL will be used for the storage of all the datasets. The reason for using MySQL is due to the author’s familiarity with it. It also has a relatively straight forward interaction with R and RStudio through R’s RMySQL package which will make interacting with my datasets friendlier.

NetBeans:

NetBeans is an open source IDE used to develop desktop, mobile and web applications with multiple programming languages such as Java, C++, HTML5. NetBeans will be used to create the web application portion of this project; this web application will be created in Java and it will be a three-tier architecture system.

## Structure

The structure is as follows:

**Chapter1**: describes the introduction, background, research aims, technologies and structure.

**Chapter 2:** Describes the system, which contains the pages related to the following: functional requirements, output requirements, data requirements, user requirements, use cases and all the non-functional requirements that apply.

**Chapter 3:** Describes the conclusion in relation to this research paper.

**Chapter 4:** Further development or research describes how the research and project could be expanded in the future, if required.

**Chapter 5:** References- contains all the references used for this research paper.

**Chapter 6:** Appendix, the appendix contains the old project proposal, project plan, the monthly journals, and other material used.

# System

## Functional Requirements

This project should identify the locations of where new HSE support centres should be placed in Ireland to provide better support services to people with disabilities. When the new support centres are located, predictive analysis should show what equipment each support centre should have in order to provide the disabled with the equipment they need to help them. The web application should allow a user to search for a specific disability and see a heat map showing the amount of that disability across the country.

### Output Requirements

#### Machine Learning (Location Prediction)

As part of this project, new health centre locations will be predicted. By predicting where these centres should be placed there is hope that they will increase the quality of service to those with disabilities around Ireland by making more of these health centres readily available and reducing the amount of travel the disabled have to do to access these health centre services. This location prediction will be displayed on the dashboard using Tableau and it will be displayed as a spatial map.

#### Prediction Analysis (Equipment Prediction)

The prediction analysis portion of this project will involve having to predict what type of equipment the health centres should have such as, wheelchairs etc., so that the proper disability service and equipment can be provided to the disabled people attending the health centres.

#### Descriptive Analysis (Web Application)

Descriptive statistics, and the summarization of the data with visuals i.e. graphs: pie charts bar charts etc. Explore and analyze disabilities in Ireland using a number of variables such as: male/female, age groups, type of disability. A web application will be used to output the analysis.

### User requirements

The User requirements for this project is that the user should be able to look at all data associated with identifying HSE support centre locations and what disability equipment they should have in stock, the information shown should be easy to understand, while the web application itself will need to be clean and well presented for the user to easily understand and use the application to find the information they are looking for.

### Data requirements

The following data requirements for this project are subject to change throughout the project. For the web application, the data requirements are that the datasets acquired contain relevant information that will allow a decent depth of analysis to be performed, these would include: disability type, age-group, and sex. For the machine learning aspect to this project a dataset containing the latitude and longitude of the current health centres will be needed as well as further information to help with the prediction of new health centre such as: population. Lastly, for the predictive analysis, the data required will include the relevant types of equipment used by disabled people based on their disability type and needs. The following is an example of the variables needed in the datasets that will be used for the project:

**Data**:

**Sex**: Male & female,

**Age** **Group**: 0-14,15-44, 25-44, 45-64, 65+,

**Types of disability:**

Blindness or a serious vision impairment, Deafness or a serious hearing impairment, A condition that substantially limits one or more basic physical activities, an intellectual disability, Difficulty in learning, remembering or concentrating, Psychological or emotional condition, Other disability, including chronic illness, Difficulty in dressing, bathing or getting around the home, Difficulty in participating in other activities.

**Machine Learning:**

**Locations**: Latitude/Longitude

**Population**: Amount by small areas

**Predictive Analysis:**

This dataset has still to be acquired.

### Use Case Diagram- Overall System

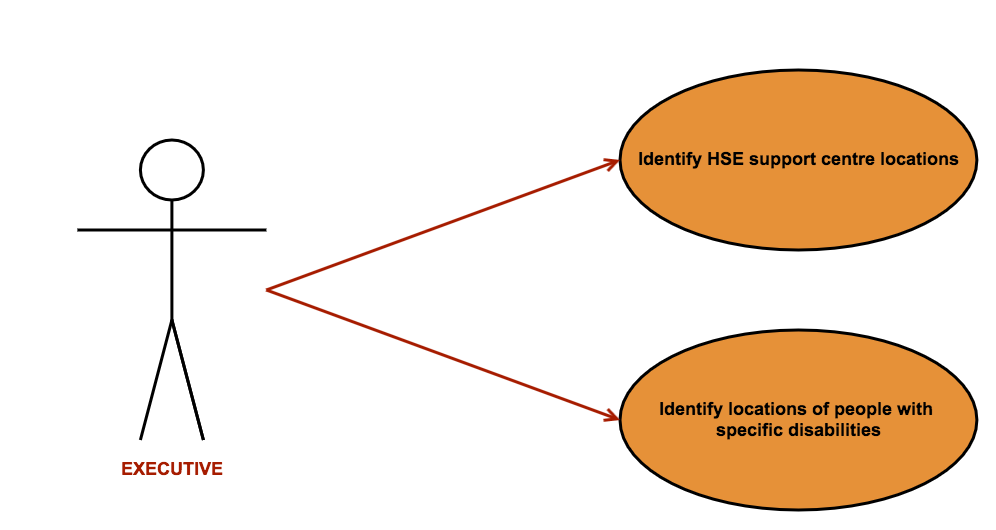


Figure 3 Overall System

### R1: Identify Health Centre locations

#### Description

In this requirement, the locations of where HSE support centres should be placed in the future are decided using machine learning. The <executive> will be able to see where throughout the country these new support centres are to be located, so they can make an informed decision of where to move HSE resources and funds in the country and build new support centres in those areas to provide more adequate services to people with disabilities

#### Use Case

**Scope**

The scope of this use case is identifying the locations of HSE support centres in Ireland that provide services to the disabled.

**Use Case Diagram**

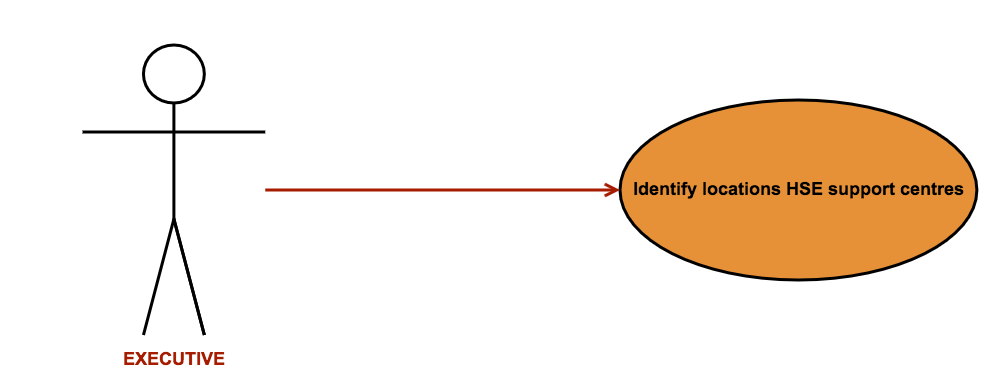


Figure 4 Identify Health Centre Locations

**Flow Description**

**Precondition**

None. This is the first requirement; therefore, it can begin straight away.

**Activation**

This use case starts when the <executive> generates a report that identifies the locations of the support centres.

**Main flow**

1. The <executive> access the web application.
2. The <executive> clicks on a button which shows the location of the new HSE support centres.
3. The <executive> analyses the findings shown.

**Exceptional flow**

E1: < Machine Learning Algorithm Error >

1. The system cannot identify health centre location using the algorithm.
2. The <executive> seeks support to fix the error.
3. The use case continues at position 5 of the main flow

**Termination**

This use case will be aborted if there is no data to be found, and if there is a problem with the database.

**Post condition**

The <executive> can use the information shown to move resources in the country to provide better services to the disabled.

### R2: Identify locations of people with specific disabilities

#### Description & Priority

This requirement will involve users such as an executive in the HSE being able to access a web application to get a breakdown of disabilities in Ireland. The executive will be able to search for a specific type of disability and a graphical representation will be displayed to show the executive the key information associated with that disability including the locations of the chosen disability.

#### Use Case

**Scope**

The scope of this use case is to identify the location of people with specific types of disabilities.

**Use Case Diagram**

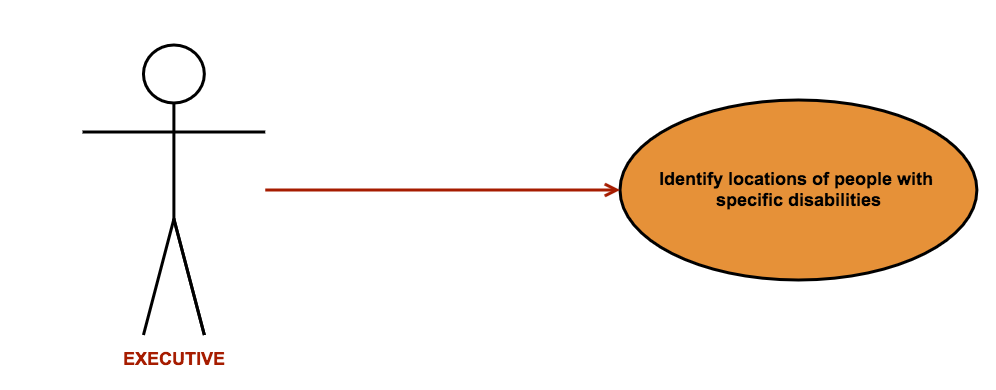


Figure 5 Identify Locations of People with Specific Disability

**Flow Description**

**Precondition**

Adequate data will need to be acquired before this requirement can be started and the web application should be completed. The data must have the locations and type of disabilities.

**Activation**

This use case starts when the <user> has opened the web application.

**Main flow**

1. The <executive> opens web application.
2. The <executive> selects the disability they want to search.
3. The <executive> views a heat map displaying the disability searched.

**Exceptional flow**

E1: <Disability not found>

1. The <executive> searches for a specific disability, but cannot find it on the application.
2. The <executive> searches for a different disability.
3. The use case continues at position 2 of the main flow.

**Termination**

This use case will be aborted if the <executive> cannot find the disability that they are searching for. Or if the application does not provide the necessary results.

**Post condition**

The <executive> closes the application.

## Non-Functional Requirements

### Performance/Response time requirement

Even though performance and response time is an important aspect of many systems, it does not apply to this project as the volume of data being used as part of the data analysis does not need to be analysed at a fast speed or with any real urgency. Therefore, the speed of performance and response time is not an important aspect to this project.

### Availability requirement

The data will remain available to the system throughout.

### Recover requirement

In the event of a hardware or software failure the data and database being used for this system will need to be backed up on a regular basis. The data folders containing the datasets and database schema’s will be backed up on GitHub using version control. The data folder will also be copied and saved to multiple locations such as google drive and Dropbox for extra safety.

### Robustness requirement

The robustness requirement does not apply to this system.

### Security requirement

The data has been acquired from several websites such as the Central Statistics Office which is a public site, the datasets that have been downloaded from the CSO site are all publically available and require no special permissions to use them. The system itself can only be accessed on a personal laptop including the MySQL database which is protected with a username and password. The database is also hosted using gearhost.com which is also being protected by a username and password.

### Reliability requirement

The data that is being used for this system came from the Central Statistics Office which compiles this data on a yearly basis. Therefore, the system is dependent on the accuracy and availability of data produced by the CSO.

### Maintainability requirement

This system does not necessarily need to be maintained once it is created. It may need to be updated with new data in the future, but for this project the data that will be used has already been acquired.

### Extendibility requirement

At the time of this requirement specification, there is no plan to further the study at this, but the methods and processes that are being used could be used again in the future.

### Reusability requirement

This system currently has no requirement for reusability.

### Resource utilization requirement

There is a need for a laptop, and cloud storage for backups. Some of the software that will be used for this system still needs to be installed, they are: RStudio, Tableau, MySQL Workbench, Gearhost, Google Drive and Dropbox.

## Design and Architecture

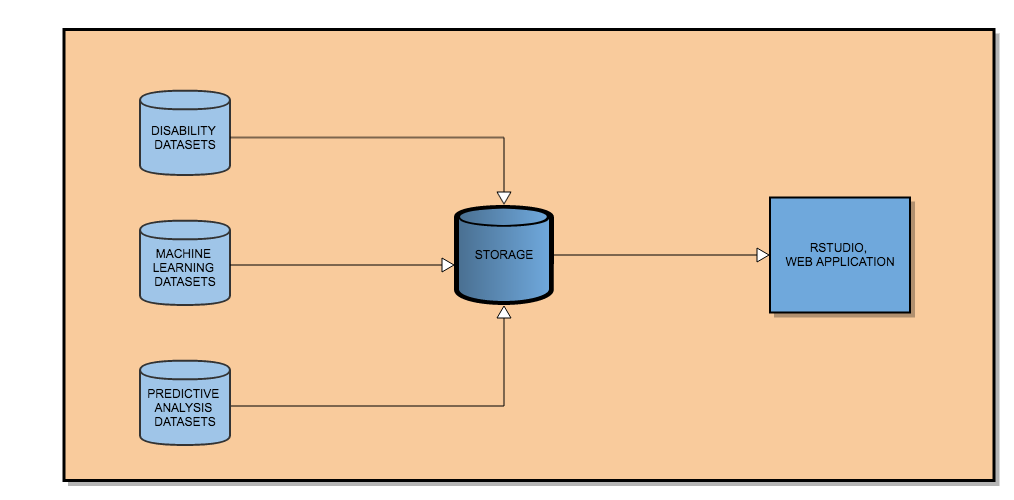


Figure 6 System Architecture

This section describes the system architecture and shows the components that make up the project and show how they interact with each other and how the extraction of the data works. The components that make up this project include an input source which is the datasets mentioned previously, that will be acquired from various data sources such as websites etc. There is a MySQL database component which will be used to store the datasets as tables and RStudios will be used to connect to the database to access the data and conduct all the analysis required for the project. After all the analysis is complete the datasets will be visualized through a java based web application by running the R-scripts against the data in the database. All the components listed above make up the key aspects of the system architecture and they are subject to change as the project progresses.

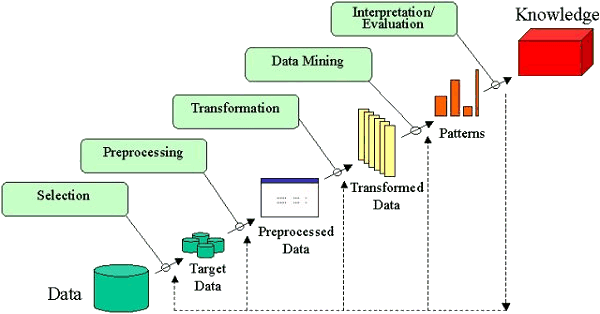


Figure 6 KDD

The data analysis aspect of this project, which will be used with the web application will follow the KDD (Knowledge discovery and databases) process. The KDD will be used to map out the steps to be implemented for the analysis of disabilities.

The steps are:

**Data Cleaning:** In this step, the data will be cleaned by removing inconsistent data. **Data Integration:** In this step, multiple datasets can be combined to get the result needed. **Data Selection:** In this step, the data in the database is retrieved for analysis tasks to be completed. **Data Transformation:** The required data is transformed using aggregation. **Data Mining:** In this step, patterns are extracted. **Pattern Evaluation:** the patterns are evaluated. And **Knowledge Representation:** the knowledge is represented. The web application will follow a three-tier architecture:

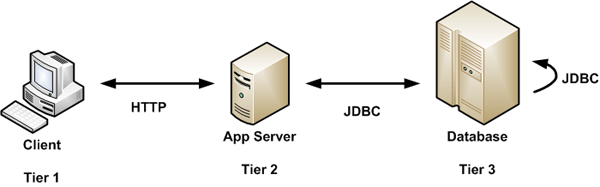


Figure 7 Three-Tier Architecture

The web application will use a GUI, which will be interacting with the database which is storing all the datasets for the project. The GUI will run R-scripts on the data that is pulled from the database and produce the output in a specific form such as a heat map.

## System Evolution

This system could evolve in the future to include further analysis if more datasets are acquired. There could also be further machine learning added, for example a prediction of learning disabilities in school children. Another way this system could evolve is by expanding the geographic focus past Ireland and have it identify where health centres should be placed in order countries or in specific regions i.e. Europe, North America etc. while also conducting analysis of disabilities in those very same locations. By expanding the focus, it will allow further research into health centre locations outside of Ireland, comparing of disability statistics from several different countries or regions to analyze the difference if there is one in the amount of disabilities affecting men, women and children of various types disabilities. The web application could also be expanded later to include more functionality.

## Implementation

This project will be implemented in steps to make it easier to track progress and to manage the work load of the entire project throughout. Firstly, a basic portion of the web application for the mid-point presentation will be implemented. After the mid-point is done the machine learning aspect will be implemented along with the necessary data analysis that is required for this project. The reason that the web application will be postponed until a later stage is due to the application needing R-scripts to produce the output. These R-scripts will therefore be needed first, which will involve completing the machine learning, predictive analysis and data analysis first. The entire machine learning, predictive analysis and data analysis will be implemented using R, while the web application will be implemented in NetBeans using the Java programming language. To identify the locations of new HSE support centres, a machine learning algorithm (cluster) will be found, and customized to show where these new support centres should be placed in Ireland.

## Graphical User Interface (GUI) Layout

The following is a mock-up of how the web application is likely to look like for the mid-point presentation, the designs are subject to change.

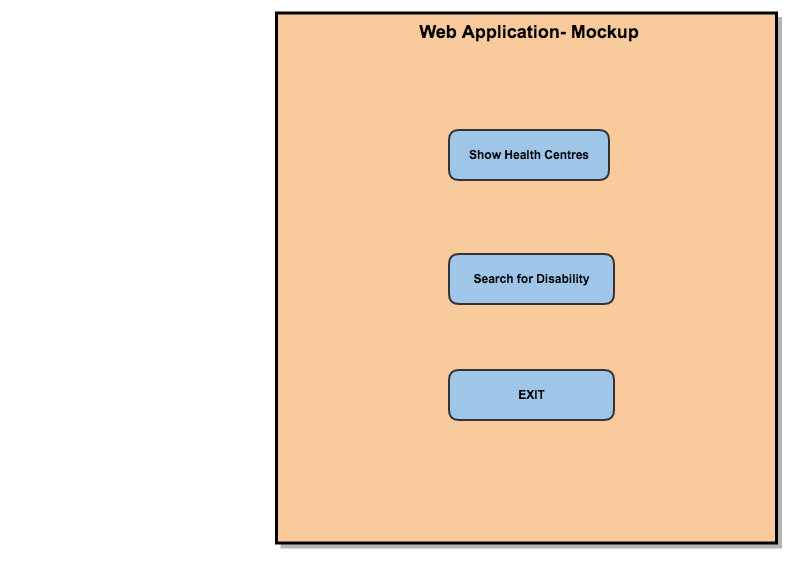


Figure 7 Menu

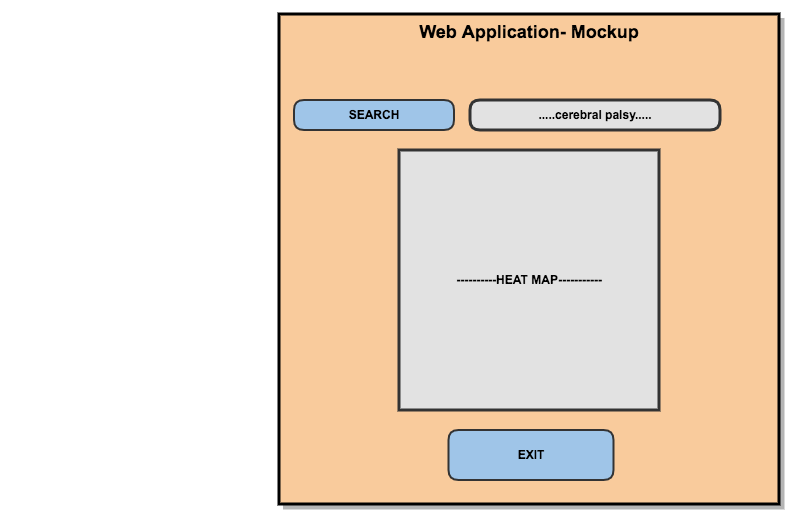


Figure 8 Disability Search

## Testing(TO BE DONE AT A LATER STAGE)

Testing will conducted in order to ensure that all parts of the project is working as planned

## Customer testing(TO BE DONE AT A LATER STAGE)

Provide evidence for and results of customer testing. This may include ratings or quotes from the customer.

## Evaluation(TO BE DONE AT A LATER STAGE)

How was the system evaluated and what are the results? In many cases this will include usage data and user feedback. It may also include performance evaluations, scalability, correctness, etc. depending on the focus of the project.

Quantative results may be reported in tables or figures. Note that tables have their caption above the table and need to be cross referenced in the text. In many cases, tables are better to read if you skip the vertical lines.

Table 1: Performance with and without caching

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Nwithout** | **Nwith** | **Std.-Deviationwith** | **Std.-Deviationwithout** | **p** |
| Records | 100 | 200 | 2.54 | 3.97 | .002 |
| Data (GB) | 100 | 200 | 2.54 | 3.97 | .002 |
| Speed | 100 | 200 | 2.54 | 3.97 | .002 |

Figures have their caption below the figure. Make sure that if you use colour, the figure is still readable when printed in black & white, e.g., by using additional symbols, patterns, etc.



Figure 9: Learning gain across different experimental groups

# Conclusions

The advantages of this project are that it will include the identification of new HSE support centres in Ireland, which after research conducted by the author has been done in other countries but has not been done before in Ireland. This gives the project a unique aspect to it.

The disadvantages of this project are that while these HSE support centres can be identified, it does not necessarily mean that they will be built. The HSE support centres in Ireland are funded by the Irish government, which means there may be a lack of resources and funds to build new support centres in the country. With this lack of funding and resources, identifying where new support centres should be built may not be realistic in Ireland until the government decides to invest more funding into this area of the HSE.

The opportunities that this project provides is that if the Irish government does decide to spend more government resources to build new HSE health centres then this project could give them a head start by identifying the best possible locations to place the health centres, saving valuable time.

The limits to this project is that the author must learn by doing as the author has not learnt significant portions of what will be used to create the project. The areas that the author will learn by doing are: Machine Learning and Predictive Analysis.

# Further development or research

This results of this project could lead to more expansive disability analysis and provide more focus on how badly funded the disability services in Ireland are. With more resources, development and research this project could expand to provide a better analysis on support centres and support services to disabled people in the country. With further research, machine learning could become an even bigger asset to the HSE in Ireland. Machine learning could help predict certain disabilities in children, which will help the HSE plan more effectively in regards to their resource throughout the country.

# Motivational Section

The motivation for this research is to address the challenges faced by people with disabilities having to travel long and difficult journeys to reach vital support services. This research was undertaken for personal reasons due to the fact that the author of this paper was born with a birth defect (born with one hand), While he does not personally consider himself disabled due to the fact that he is not hindered in any way by having one hand as it doesn’t stop or prevent him from doing anything that he want to do, but it did get him thinking about other people around the country who have various disabilities that prevent them from doing certain activities or require them to use certain services that are provided by health centres around the country. While researching the support services in the country, it became apparent the lack of support services in the country so, it was decided to implement a machine learning project predicting where new support centres could be placed so more services could be provided.

# Bibliography

Authority, N. (2015). [online] Nda.ie. Available at: http://nda.ie/Publications/Transport/Transport-Publications/Provision-of-Transport-Services.html [Accessed 9 Dec. 2016].

Cran.r-project. (2016). [online] Available at: https://cran.r-project.org/web/packages/sqldf/sqldf.pdf [Accessed 9 Dec. 2016].

Cso.ie. (2008). [online] Available at: http://www.cso.ie/en/newsandevents/pressreleases/2008pressreleases/nationaldisabilitysurvey2006-firstresults/ [Accessed 9 Dec. 2016].

Data.gov.ie. (2016). [online] Available at: https://data.gov.ie/data [Accessed 9 Dec. 2016].

Disability-federation.ie. (2016). [online] Available at: http://www.disability-federation.ie/index.php?uniqueID=29 [Accessed 9 Dec. 2016].

# Appendix

## Project Proposal

**Project Proposal**

**Identifying new Health Centre Locations in Ireland**

Declan Barnes, X13114352, Declan.Barnes@student.ncirl.ie

BSc (Hons) in Computing

Data Analytics

16/10/2016

**Objectives**

(Max. 1 Page)

The **first part** of this project is to acquire the necessary information needed to complete the project. This information will come in the form of several datasets that will be acquired from public resources such as the central statistics office. Each dataset will contain key information that I will need for my final year project in relation to disabilities in Ireland i.e. how many people are living with a disability in Ireland, what type of disability do they have, what sex and age are they etc.

The **second objective** of this project is to analyse disabilities in Ireland. My aim of doing this is by using several datasets that I have acquired through extensive online research as stated above. The data sets in question will show the statistics of disabled people across Ireland such as: Sex, Age Group, and type of disability. This information will be used to create a dashboard representing key analysis of the disabilities, for example: how many disabled people require homecare in Ireland and how many people with disabilities are not working due to their disability.

The **third objective** of this project will be to map all the health centres in Ireland showing their locations and proximity to each other. These health centres will show us how much of Ireland is ideally covered by health centres and where there is a lack of centres in certain areas (i.e. rural areas) of the country.

The **fourth objective** of this project is to use machine learning to predict where new health centres should be placed in Ireland to provide better services and coverage to disabled people in the country. This objective will take the above-mentioned information (in the second objective) into account of where current health centres are located and predict where new centres should be built i.e. in areas where the nearest health centre is too far for disabled people to travel to, and then it will show on a map where the new centre should be placed in relation to this information.

The **above objectives** will be done using MySQL, the R programming language, and Tableau. MySQL will be used to store the datasets that I will be using for my project, the R programming language will be used pull the data from the MySQL database and perform data manipulation, data cleaning on it, as well as being used for the machine learning aspect of the project. Finally, Tableau will be used to visualize my findings with its professional looking charts.

**Background**

(Max. 2 Pages)

There are a significant number of people in Ireland that are living with some form of disability and they are people of all ages and backgrounds. I thought that it would be of interest to various people and organisations to breakdown or analyse disabilities in Ireland for my final year project.

The reason I chose to do this project was for personal reasons due to the fact that I was born with a birth defect (I was born with one hand), While I don’t personally consider myself disabled due to the fact that I am not hindered in any way by having one hand as it doesn’t stop or prevent me from doing anything that I want to do, but it did get me thinking about other people around the country who have various disabilities that prevent them from doing certain activities or require them to use certain services that are provided by health centres around the country. For my final year project, I wanted to investigate further into this by analysing disabilities across the country in detail.

I wanted to show how many health centres there are and where in the country they are located so I could see if new health centres could be built to provide better coverage of services to disabled people in Ireland by helping to predict where these new health centres should be placed in the country.

While researching information for my project I searched online to see if other people, companies etc. had done something like what I am trying to do in relation to my final year project, after searching for a significant time I was unable to find anything in relation to predicting where new health centres should be built. I could find government run websites that contained statistics about disabilities in Ireland but they didn’t breakdown the statistics the way I am going to show a better picture of disabilities in Ireland.

I believe it is important to try to understand or work out where there are limitations in the services provided to disabled people across Ireland, by doing this project I hope to show these limitations in a constructive way that will help improve the quality of the services that disabled people rely on daily for their everyday needs. I think it would be beneficial overall to show where new health centres should be positioned in Ireland as it is unacceptable to expect disabled people in rural areas specifically, to travel a significant distance to get to their nearest health centre, while these same health centres could be failing to provide adequate services/coverage to the people who need them by being out of reach.

While I was deciding whether to base my project on the analysis of disabilities in Ireland I looked at several informative websites to get a better picture of what was going to be involved with doing this project, for example I researched the disability-federation website to come up with ideas of what type of information would be useful for me to show in my project.

On the disability-federation website it states the following in regards to disabilities:

It states that people in Ireland that are living with a disability and parents who have special needs children are at a much higher risk of suffering depression than those who do not. I realised after reading this that my idea of predicting new health centres would be a beneficial idea to those disabled people and the parents of special needs children as these health centres could provide very important care and or support to these people to help relieve the pressure that they live with daily by creating more health centres that are much better placed to provide the necessary coverage across the country that they need.

Another interesting fact that I noticed through my research was the attitudes towards people of a disability. On the national disability authority, it said that negative attitudes towards disabled people had become a significant barrier to the inclusion of the disabled into various aspects of society and therefor preventing people with a disability from reaching their full potential.

The National Disability Authority has conducted a series of surveys at regular intervals on the public’s attitude towards people living with a disability (2001,2006,2011). The NDA’s surveys showed that attitude towards disabled people improved in the years 2001 and 2006, but the year 2011 showed a decrease in attitudes.

I thought that the information stated above about the attitudes toward the disabled was interesting, so much in fact that as part of my dashboard I plan to include sentimental analysis of twitter to show the positive and negative tweets in relation to disabilities.

**Technical Approach**

Brief description of the approach to be followed (Max. 1 Page), Research, literature review, requirements capture, implementation etc.…

**Research:**

I carried out a great deal of research before starting my project. There were a lot of different factors to be figured out. My main priority was to create a project that was both interesting and enjoyable so I conducted a lot of research into the data analysis of various topics to give me an idea of what I needed to do for my final year project.

During my research, I focused on specific topics such as storing datasets in a MySQL database and if it was possible to pull or access the data from MySQL using RStudio and the R programming language. I also conducted research on data manipulation, data cleaning, data visualization and machine learning. These topics are what I will be using as part of my project so I felt it was best to research into them and find how I could use them in relation to my project.

The aim of researching all these topics was to give me the confidence that I need to start and complete my final year project.

**Literature review:**

As part of my approach to my project I reviewed several websites and books that I thought would be of benefit to me. These websites and books are listed below:

* Machine Learning with R - Brett Lantz
* <https://www.r-bloggers.com/>
* <https://cran.r-project.org/doc/contrib/Zhao_R_and_data_mining.pdf>

**Requirements capture:**

* Acquire needed datasets,
* Dashboard representing the following:
  + -Disability analysis represented on charts,
  + -Predicting location of future health centres.
  + Predictive analysis of equipment needed at each health centre location.

**Implementation:**

I will implement my project using the R programming language which I am currently learning in my final year at NCI. I will use R to pull my data from the MySQL database then perform data cleaning and data manipulation to get the analysis results that I am aiming to show.

On the dashboard, I plan to also show the sentimental analysis of twitter in the R language. This analysis is intended to show where tweets about disabilities are positive or negative in anyway and if they are, how or why are they negative so I can show a better picture of why some people have a negative attitude towards people living with a disability.

I also intend to conduct a word cloud in the R language representing the keywords associated with disabilities, I thought the result might be interesting to see and therefore well worth while investing the time in my project to include it on the dashboard.

I will then use Tableau to import my R scripts and produce a dashboard representing my data. I hope to have the dashboard completed by the end of semester 1.

For semester 2 I will use the R programming language to conduct the machine learning aspect of my project to predict where new health centres should be placed in the country as well as conducting the predictive analysis.

**Special resources required**

I will use the following resources to help me create and complete my project:

**Machine Learning with R - Brett Lantz:**

This is a book that was recommended to me for figuring out how I should go about the Machine learning aspect of my project. I will read this near semester 2 when I intend to start and integrate machine learning into my project.

**R-Bloggers:**

This is a very popular website containing useful information in regards to anything related to the R programming language. I will use this to help me learn R and how to use it for my project.

**Zhao R and data mining:**

This is a PDF of a well-known book that give insight into how to manage data mining and how to use best practices when working with data. I will rely on this to ensure that I am doing everything correctly when I’m working with my data.

**MacBook Pro Retina:**

I will be using my own laptop to do my project instead of the college computers as it will allow me to work from home more easily.

**RStudio:**

Since I will be using R as my programming language, I will therefore be using RStudio. RStudio is a great IDE for R. Its interface is very user friendly and easy to learn.

**Tableau:**

I will be using Tableau to show my data as their charts are more professional looking. I think Tableau will be better suited for presentations due to their attractive charts.

**Technical Details**

I**mplementation language and principal libraries**

The language that I aim to use for my final year project is the R programming language. R is a statistical programming language that has powerful packages for nearly everything that you would want to do with data.

The IDE that I will be using is the standard IDE for R which is RStudio. The R language has great support and a massive community that releases new packages and tutorials on a routine basis.

I am currently learning R on Pluralsight and in college as part of my Data applications module, which will give me the necessary skills that I need for the completion of my project.

**My project will make good use of multiple Packages in R such as:**

* **RMySQL:** database interface and MySQL driver for R.
* **TidyR:** Used for data manipulation and cleaning.
* **Stringr&Magrittr:** Data cleaning
* **Dplyr:** Also, used for data manipulation, usually used in conjunction with TidyR.
* **GoogleVis:** Data Visualisation.
* **Sqldf:** Provides an easy way to perform SQL selects on R data frames.
* **ggplot2:** Used for Visualizing data.

More packages may be used in the future as I work on my project throughout the year.

**MySQL:**

I will be using MySQL for the storage of my datasets. The reason for using MySQL is that I am very familiar with it as I have been using it for the last 3 years in college, it also has a relatively straight forward interaction with R and RStudio through R’s RMySQL package which will make interacting with my datasets friendlier.

**Evaluation**

I will evaluate my data analysis project using testing by getting volunteers to look at my dashboard and the machine learning aspect to my project by checking that everything is displayed properly and professionally.

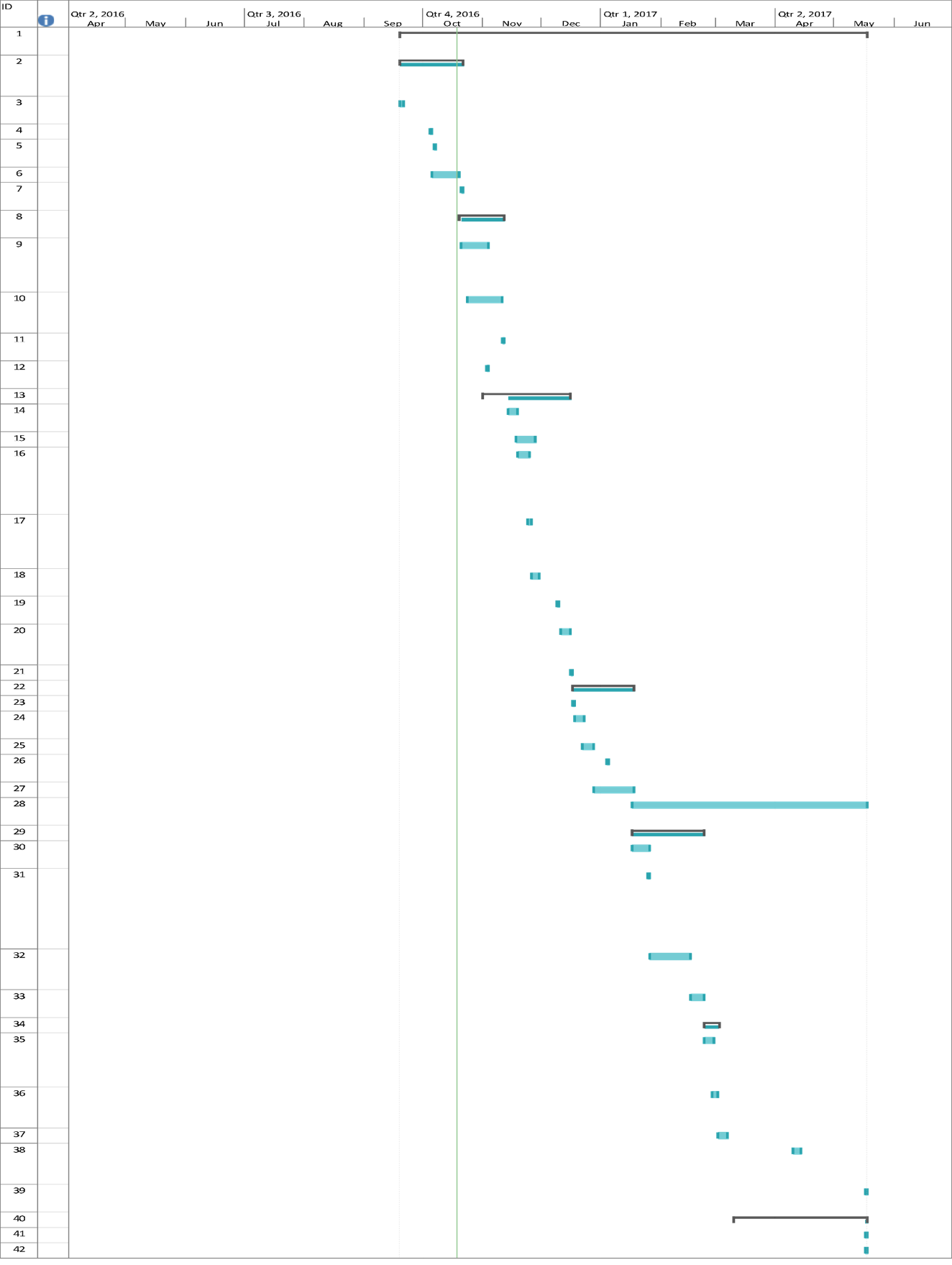
I will also ask for input from certain lecturers and my supervisor in NCI to see if my project is done properly and functioning as it should or if they have any input on how to improve the project further.

I aim to meet with my supervisor on a routine basis to make sure my project is going in the right direction while getting my supervisor to test what I have done up to each meeting to ensure accuracy and consistency.

Declan Barnes

16/10/2016

## Project Plan



|  |  |  |  |
| --- | --- | --- | --- |
| Task Name | Duration | Start | Finish |
| **Project-Disability Analysis** | **173 days** | **Mon 19/09/16** | **Wed 17/05/17** |
| **Project Proposal & Project Plan Documents** | **25 days** | **Mon 19/09/16** | **Fri 21/10/16** |
| Project Brainstorming | 2 days | Mon 19/09/16 | Tue 20/09/16 |
| Project Pitch | 1 day | Wed 05/10/16 | Wed 05/10/16 |
| Reflective Journal(Sept) | 1 day | Fri 07/10/16 | Fri 07/10/16 |
| Project Proposal | 10 days | Thu 06/10/16 | Wed 19/10/16 |
| Documentation Review & Upload | 1 day | Fri 21/10/16 | Fri 21/10/16 |
| **Requirement Specification** | **17 days** | **Thu 20/10/16** | **Fri 11/11/16** |
| Research for Disability Datasets on CSO website | 10 days | Fri 21/10/16 | Thu 03/11/16 |
| Requirement Specification Document | 14 days | Mon 24/10/16 | Thu 10/11/16 |
| Documentation Review & Upload | 1 day | Fri 11/11/16 | Fri 11/11/16 |
| Reflective Journal(Oct) | 1 day | Thu 03/11/16 | Thu 03/11/16 |
| **Project Analysis & Design** | **33 days** | **Tue 01/11/16** | **Thu 15/12/16** |
| Set up Data Warehouse | 5 days | Mon 14/11/16 | Fri 18/11/16 |
| Data Cleaning | 7 days | Fri 18/11/16 | Sun 27/11/16 |
| Filter statistics on male/female and type of disabilities | 5 days | Sat 19/11/16 | Thu 24/11/16 |
| filter on age groups and type of disabilities | 2 days | Thu 24/11/16 | Fri 25/11/16 |
| Create Prototype(Web Application) | 3 days | Sat 26/11/16 | Tue 29/11/16 |
| Reflective Journal(Nov) | 1 day | Fri 09/12/16 | Fri 09/12/16 |
| Prepare for Mid-Point Presentation | 5 days | Sun 11/12/16 | Thu 15/12/16 |
| **Mid-Point Presentation** | **1 day** | **Fri 16/12/16** | **Fri 16/12/16** |
| **Post Mid-Point** | **23 days** | **Sat 17/12/16** | **Tue 17/01/17** |
| Review Feedback from panel and makes changes | 1 day | Sat 17/12/16 | Sat 17/12/16 |
| Finish Web Application | 5 days | Sun 18/12/16 | Thu 22/12/16 |
| Christmas Break | 4 days | Thu 22/12/16 | Tue 27/12/16 |
| Reflective Journal(Dec) | 1 day | Wed 04/01/17 | Wed 04/01/17 |
| Study Break for exams | 15 days | Wed 28/12/16 | Tue 17/01/17 |
| **Project Final Documentation** | **87 days** | **Tue 17/01/17** | **Wed 17/05/17** |
| **Model Finalization** | **27 days** | **Tue 17/01/17** | **Wed 22/02/17** |
| Research Machine Learning(Prediction) | 7 days | Tue 17/01/17 | Wed 25/01/17 |
| Meet with Supervisor about implementing the machine learning aspect | 1 day | Wed 25/01/17 | Wed 25/01/17 |
| Implement Machine Learning | 15 days | Thu 26/01/17 | Wed 15/02/17 |
| Finish Machine Learning | 5 days | Thu 16/02/17 | Wed 22/02/17 |
| **Testing Phase** | **6 days** | **Thu 23/02/17** | **Thu 02/03/17** |
| Test the Web Application & Machine Learning | 3 days | Thu 23/02/17 | Mon 27/02/17 |
| Meet with Supervisor on progress | 3 days | Mon 27/02/17 | Wed 01/03/17 |
| **Final testing phase** | **3 days** | **Thu 02/03/17** | **Mon 06/03/17** |
| **Prepare Showcase Materials** | **4 days** | **Mon 10/04/17** | **Thu 13/04/17** |
| **Upload final code submission** | **1 day** | **Wed 17/05/17** | **Wed 17/05/17** |
| **Technical Report, Findings** | **49 days** | **Fri 10/03/17** | **Wed 17/05/17** |
| Review | 1 day | Wed 17/05/17 | Wed 17/05/17 |
| **Upload Final Documentation** | **1 day** | **Wed 17/05/17** | **Wed 17/05/17** |

## Monthly Journals







## Other Material Used(TO BE DONE AT A LATER STAGE)

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