**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

Table of Contents

[The Analysis of Disabilities in Ireland 1](#_Toc464407445)

[1. Objectives 3](#_Toc464407446)

[2. Background 4](#_Toc464407447)

[3. Technical Approach 6](#_Toc464407448)

[4. Special resources required 8](#_Toc464407449)

[5. Project Plan 9](#_Toc464407450)

[6. Technical Details 10](#_Toc464407451)

[7. Evaluation 11](#_Toc464407452)

# 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 **main 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 **second 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 **final 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 was able to 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

If applicable, e.g., books, hardware, etc.

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.

[**https://www.r-bloggers.com/**](https://www.r-bloggers.com/)**:**

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.

[**https://cran.r-project.org/doc/contrib/Zhao\_R\_and\_data\_mining.pdf**](https://cran.r-project.org/doc/contrib/Zhao_R_and_data_mining.pdf)**:**

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.

# ProjectPlan

|  |  |  |  |
| --- | --- | --- | --- |
| 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(Dashboard) | 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 Dashboard | 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 dashboard & 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** |

# 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.

**Tableau:**

I have set up a student account with Tableau, as I plan to use it to visualize my data due to Tableau having a much cleaner appearance and more attractive charting to better represent my data in the midpoint presentation and the final presentation at the end of the final year project.

# 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