Random Patient Clinical Data Simulator

Karol Groszewski

201507145

# Abstract

Access to real patient data is important for teaching purposes and research. However, due to law this data is protected and can only be used by people who have permission to access it.

# Acknowledgements

# Table of Contents

[Abstract 2](#_Toc14721710)

[Acknowledgements 3](#_Toc14721711)

[Table of Contents 4](#_Toc14721712)

[Chapter 1 7](#_Toc14721713)

[Introduction 7](#_Toc14721714)

[1.1 Objectives 7](#_Toc14721715)

[1.2 Outcome 7](#_Toc14721716)

[1.3 Report Structure 7](#_Toc14721717)

[1.4 Marking Scheme 7](#_Toc14721718)

[Chapter 2 8](#_Toc14721719)

[Background Research/Related Work 8](#_Toc14721720)

[2.1 Simulation 8](#_Toc14721721)

[2.2 Monte Carlo Simulation 8](#_Toc14721722)

[2.3 Data Sources 8](#_Toc14721723)

[2.3.1 National Records of Scotland 8](#_Toc14721724)

[2.3.2 Scottish Government 8](#_Toc14721725)

[2.3.3 ISD Scotland 9](#_Toc14721726)

[2.3.4 Diabetes UK 9](#_Toc14721727)

[2.3.5 ScotPHO 9](#_Toc14721728)

[2.4 Related Work 9](#_Toc14721729)

[2.4.1 generatedata.com 10](#_Toc14721730)

[2.4.2 databasetestdata.com 11](#_Toc14721731)

[Chapter 3 12](#_Toc14721732)

[Data Analysis and Creation 12](#_Toc14721733)

[3.1 Community Health Index (CHI) Number 12](#_Toc14721734)

[3.2 Statistics gathering 12](#_Toc14721735)

[3.2.1 Age Distribution 12](#_Toc14721736)

[3.2.2 Gender Distribution 13](#_Toc14721737)

[3.2.3 Body Mass Index (BMI) Distribution 13](#_Toc14721738)

[3.2.4 Circulatory and Respiratory S&S 15](#_Toc14721739)

[3.2.5 Diseases of the Skin & Subcutaneous Tissue 16](#_Toc14721740)

[3.2.6 General Abnormal S&S NEC 17](#_Toc14721741)

[3.2.7 Digestive/Abdominal S&S 18](#_Toc14721742)

[3.2.8 Diabetes Type 1 19](#_Toc14721743)

[3.2.9 Diabetes Type 2 20](#_Toc14721744)

[3.2.10 Hypertension 21](#_Toc14721745)

[Chapter 4 22](#_Toc14721746)

[Problem Description and Specification 22](#_Toc14721747)

[4.1 Problem Description 22](#_Toc14721748)

[4.2 Where can this software be applied? 22](#_Toc14721749)

[4.3 Functional Requirements 22](#_Toc14721750)

[4.4 Non-Functional Requirements 23](#_Toc14721751)

[4.4 Approach taken to Solving the Problem 23](#_Toc14721752)

[Chapter 5 24](#_Toc14721753)

[Software Description and Specification 24](#_Toc14721754)

[5.1 Software Architecture 24](#_Toc14721755)

[5.2 Data set Design 24](#_Toc14721756)

[5.3 Graphical User Interface Design 24](#_Toc14721757)

[Chapter 6 25](#_Toc14721758)

[Detailed Design and Implementation 25](#_Toc14721759)

[6.1 Choice of Programming Languages Used 25](#_Toc14721760)

[6.2 Choice of Third Part Tools 25](#_Toc14721761)

[6.3 Choice of Development Environment and Tools 25](#_Toc14721762)

[6.4 Challenges of Design and Implementation 25](#_Toc14721763)

[Chapter 7 26](#_Toc14721764)

[Verification and Validation 26](#_Toc14721765)

[7.1 Verification 26](#_Toc14721766)

[7.2 Validation 26](#_Toc14721767)

[Chapter 8 27](#_Toc14721768)

[Results and Evaluation 27](#_Toc14721769)

[8.1 User Evaluation 27](#_Toc14721770)

[8.2 Evaluation of Planning 27](#_Toc14721771)

[8.3 Evaluation of End Product Functionality 27](#_Toc14721772)

[Chapter 9 28](#_Toc14721773)

[Summary and Conclusions 28](#_Toc14721774)

[9.1 Summary 28](#_Toc14721775)

[9.2 Future Work 28](#_Toc14721776)

[9.3 Overall Reflection 28](#_Toc14721777)

[9.4 Final Conclusion 28](#_Toc14721778)

[Bibliography 29](#_Toc14721779)

[Appendix A 30](#_Toc14721780)

[Appendix B 31](#_Toc14721781)

[Appendix C 32](#_Toc14721782)

Chapter 1

## Introduction

Simulation of data can be a huge benefit for teaching and research purposes. In order to achieve data that can be used for that purpose real statics will have to be used with generated patients in order to create this data as real data cannot be used due to law protecting people’s privacy.

### 1.1 Objectives

The main objective of this project is to develop a “random patient data” which then can be used for training, research and teaching purposes. The user will have access to information about various conditions based on Scottish statistics.

The application will allow users to filter through that information in order to analyse the data about these conditions. While filtering through these statistics appropriate graphs are going to be generated to aid ease of reading the data.

The application is going to be web-based so that it can be accessed anywhere and on any type of device. Therefore, the user should be able to access the application on various electronic devices that have internet access.

Sources that have been gathered to produce this application are going to be available to the user. If they would like to find out more information about these conditions or data used to generate these statistics.

### 1.2 Outcome

### 1.3 Report Structure

The report is broken down into “#” different chapters where each chapter focuses on different development stage of the application. The first chapters focus on background research and the problem to be solved by the application. Followed by chapters focused on detailed analysis of design of the application, how implementation process has been executed, what issues have been encountered during the development of the application. The what were the results of testing stage and evaluation of the application. There report will come to an end with a summary and last comments before the final conclusion.

### 1.4 Marking Scheme

The marking scheme for this project is going to be “Experimentation-based with Significant Software Development” this is because before developing the application a lot of research will have to be done beforehand to gather all the requirements in order to generate usable data for the application.

Chapter 2

## Background Research/Related Work

The aim of this chapter is to describe in detail what research has been done in order to prepare for development of “RPCDS”. This section will talk about benefits of simulation in relation to teaching to research and from where statistics were gathered for project.

### 2.1 Simulation

Simulation based learning in a health education environment has some advantages over other teaching methods used in that field depending on context, topic and method. By creating this application, students who are studying to be a health professional can practice analysing patient data and statistics which can result in additional gains in knowledge, critical thinking ability, satisfaction or confidence based on studies.

### 2.2 Monte Carlo Simulation

### 2.3 Data Sources

For “RPCDS” to be useful it needs to have some sort of data which comes from valid statistical sources. In this part of the chapter various data sources are going to be talked about and how they provide valid data for the project.

#### 2.3.1 National Records of Scotland

National Records of Scotland is a non-ministerial department of the Scottish Government which is responsible for civil registration, the census in Scotland, demography and statistics, family history, the national archives and historical records.

Since National Records of Scotland is a government organisation, they provide data which is used by various researchers and this would qualify as a high-quality data source for the application. The statistics that they produce play a vital role in underpinning decisions from national to local level and are a building block in development of economic and social statistics.

#### 2.3.2 Scottish Government

The Scottish Government website provides a range of official statistics about Scotland from a variety of data producers, for information and re-use. They provide public access to data behind our official statistics in linked open data format.

#### 2.3.3 ISD Scotland

Scotland has some of the best health service data in the world. This is because Scotland has information which combines high quality data, consistency, national coverage and the ability to link data to allow patient-based analysis and follow up. The Information Services Division which is a part of NHS National Services Scotland provides health information, health intelligence, statistical services and advice that supports quality improvement in health and care and facilitates robust planning and decision making.

#### 2.3.4 Diabetes UK

Diabetes UK is a charity which is registered Scotland, England and Wales. The charities aim is to tackle the diabetes crisis. They are there to prevent Type 2 diabetes, campaign for and support everyone by diabetes, and fund research that will cure the condition.

Diabetes UK is UK’s leading charitable funder of diabetes research. They help others by pioneering research into all forms of diabetes and diabetes-related complications. The charity releases annual statistics on how and how many people are affected by the condition in the UK.

#### 2.3.5 ScotPHO

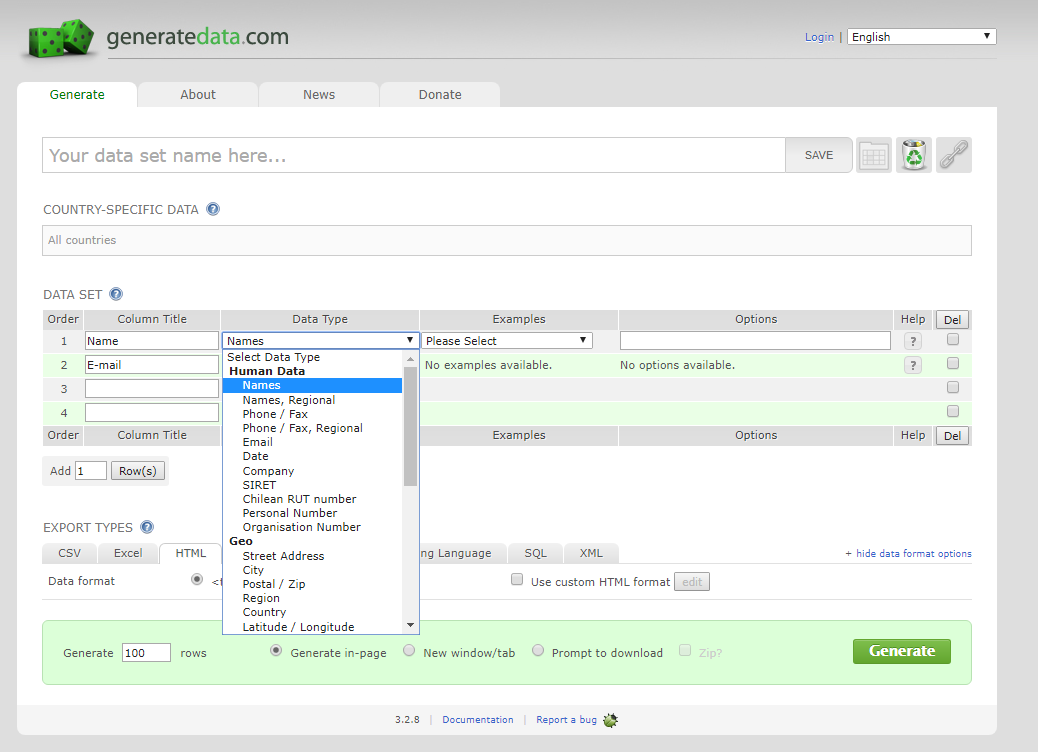
ScotPHO’s aim is to provide a clear picture of the health of the Scottish population and the factors that affect it. They contribute to improved collection and use of routine data on health, risk factors, behaviours and wider health determinants. ScotPHO takes a lead in determining Scotland’s future public health information needs, develop innovations in public health information and provide a focus for new routine public health information development where gaps exist.

### 2.4 Related Work

In this part of the chapter different existing applications are going to be analysed in order to see how other people generate data for the users of their services. From there advantages and disadvantages of each application are going to be listed.

#### 2.4.1 generatedata.com

generatedata.com is a free open-source tool which creates custom formatted samples and test data.



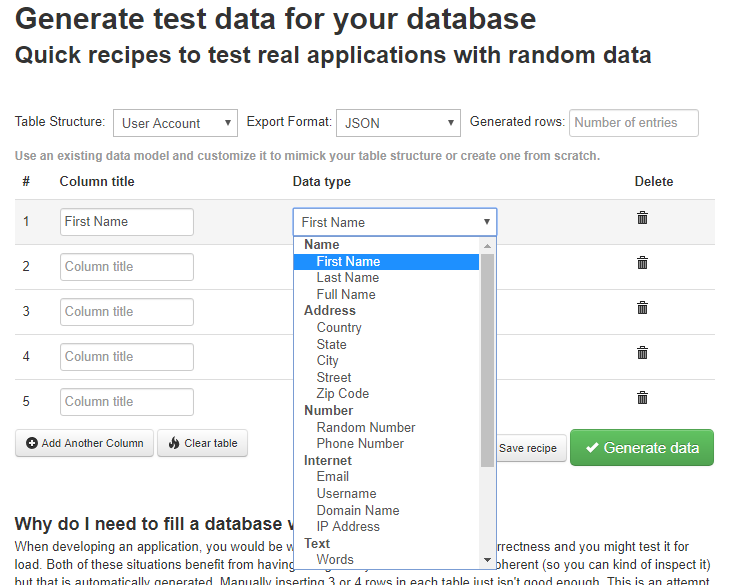
This data tool is a allows users to add various column names and select a data type which tell the website what data to generate for that column. It also has filtering. For example, if you want just to generate a female population you can add that statement to the options tab. After that the user can chose to what file the generated data should be exported to and how much of data should be generated by specifying the number of rows.

The advantage of using this website is that the user has various options in order to generate specific data that they would like. Another advantage is that the user can export the generated data to various formats for their uses. The data can be country specific. If you want to generate people who live in another country and you want to have their addresses, you can add that to the generation by changing that option.

The disadvantages of this website are that the user interface feels complicated to use. For example, if you want to generate just people with random names then you also need a provide an example to the website cause otherwise it won’t know how you would like to generate that data. Another disadvantage is that when you get an error the website won’t tell you what is causing it. When you click on the help button it gives the user a lot of information which is not related to the problem that the user is having.

#### 2.4.2 databasetestdata.com

databasetestdata.com is a free test data generator for databases. The aim of this test data is to test real applications with random test data.



This data tool allows users to generate specific data based on the “Data Types” chosen by the user. The website has some already pre-written styles such as “User Account” or “Shop Product”. The user also has the option to export the generate the data to JSON, CSV and XML. You also have the option to generate various number of entries. After clicking “Generate data” button the website takes the user to another page which shows the generated raw string in a textbox.

The advantages of this website are that the user has a wide range of options that they can generate. Another advantage is that the User Interface is simple to use as all the actions to generate data are simple to execute.

The disadvantage of this website is that the website does not export any of the data to actual files that can be downloaded by the user. Therefore, the user would have to use another service in order to convert the generated data into a file and store it on their system.

Chapter 3

## Data Analysis and Creation

The aim of this chapter is to describe in detail step by step how data was analysed and created for the application.

### 3.1 Community Health Index (CHI) Number

The CHI number is a population register used in Scotland for healthcare purposes. The CHI Number uniquely identifies a person on the index. This number is 10 digit long. The first 6 digits are persons Date of Birth (DDMMYY) followed by 2 random digits. The 9th digit is representing the persons sex which assigns an even number for females and an odd number for males. The final digit is a check digit.

### 3.2 Statistics gathering

Statistics for this project were gathered from official sources for the data to be valid. The statistics were gathered in a format which showed how many people were affected by the condition in different age groups based on gender. Some statistics include signs and symptoms (S&S) that do not necessarily have a confirmed diagnosis, classified according to body system.

#### 3.2.1 Age Distribution

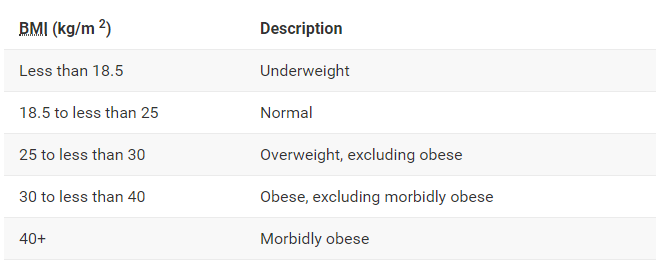
For the software to be useable it will have to have realistic statistics about how the age is distributed in Scotland. This is because conditions to individuals are going to be assigned by their age and gender. Below is a graph included which shows the percentage of age is distributed in Scotland based on Mid-2017 population estimates Scotland.

#### 3.2.2 Gender Distribution

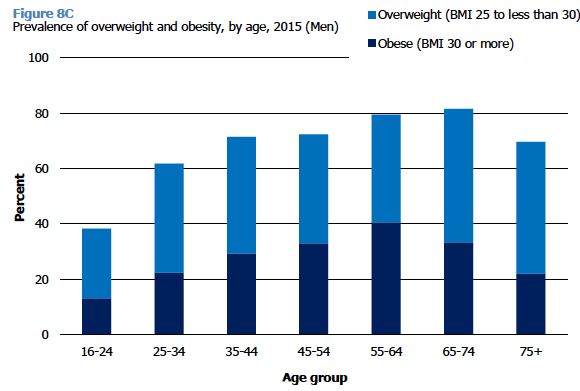
To get accurate results in this project gender will have to be distributed accordingly. This is because if all age groups are split into 50% males and 50% females then wrong data will be outputted. In the graph below it is shown the percentage of males and females at different ages.

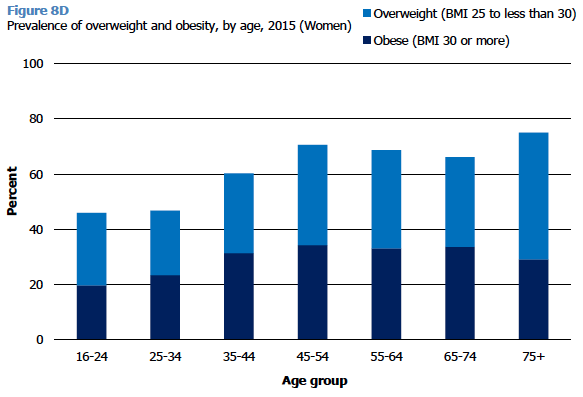
#### 3.2.3 Body Mass Index (BMI) Distribution

BMI is a widely accepted measure that allows for differences in weight based on persons height. This measure is important since people with higher BMI tend to be more affected by conditions such as Diabetes and Hypertension. Below is a BMI table for adults who are age 16 or higher.



Since people with an age lower than 16 are classified differently people who are younger than that will not have a BMI measure in the system. Below are charts which show peoples BMI based on age and gender. For people with a BMI that have lower than 25 will have a BMI which is either normal BMI or underweight BMI will have a random BMI generated between 15 and 25.





#### 3.2.4 Circulatory and Respiratory S&S

Circulatory and Respiratory [5] condition is one most common condition that affects people in Scotland. Below are graphs which show the percentage of males and females that are affected by this condition based on age for a population of a thousand based on Information Services Division Scotland statistics.

#### 3.2.5 Diseases of the Skin & Subcutaneous Tissue

Diseases of the Skin & Subcutaneous Tissue [5] condition is one most common condition that affects people in Scotland. Below are graphs which show the percentage of males and females that are affected by this condition based on age for a population of a thousand based on Information Services Division Scotland statistics.

#### 3.2.6 General Abnormal S&S NEC

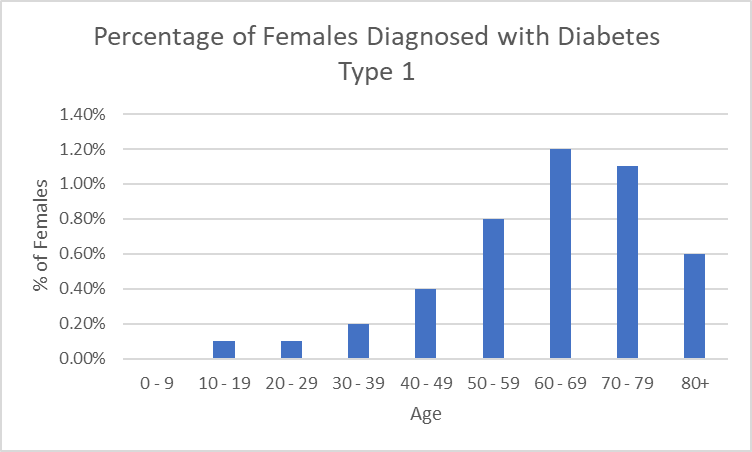
General Abnormal [5] condition is one most common condition that affects people in Scotland. Below are graphs which show the percentage of males and females that are affected by this condition based on age for a population of a thousand based on Information Services Division Scotland statistics.

#### 3.2.7 Digestive/Abdominal S&S

Digestive/Abdominal [5] condition is one most common condition that affects people in Scotland. Below are graphs which show the percentage of males and females that are affected by this condition based on age for a population of a thousand based on Information Services Division Scotland statistics.

#### 3.2.8 Diabetes Type 1

Diabetes is a popular condition in todays society. However, Diabetes Type 1 [6] is less common than Diabetes Type 2. Below are graphs which show the percentage of males and females that are affected by this condition based on age for a population of a thousand based on statistics from Diabetes UK.



#### 3.2.9 Diabetes Type 2

Diabetes Type 2 is more common than Diabetes Type 1 [6] because an obese person has a higher chance of being affected by this condition. Below are graphs which show the percentage of males and females that are affected by this condition based on age for a population of a thousand.

#### 3.2.10 Hypertension

Hypertension [7] condition is one most common condition that affects mainly affects people who are older than 30 years old in Scotland. Below are graphs which show the percentage of males and females that are affected by this condition based on age for a population of a thousand based on ScotPHO statistics.

Chapter 4

## Problem Description and Specification

The aim of this chapter is to describe in detail what the problem is and how it is going to be approached. This will include a complete list of functional and non-functional requirements to be met by the software and how these requirements were identified and approached.

### 4.1 Problem Description

The challenge in Random Patient Clinical Data Simulator will consist of creating a population which will have an age, gender, BMI[[1]](#footnote-1) and conditions based on statistics gathered for the project. Each of the statistics have a number of people that are a certain age, gender, BMI and conditions. These number will have to be converted into percentages to help assign these statistics to the simulated data. These percentages which help generate data will also have to have include an error rate which will make the statistics vary each time the data is simulated. Each person will also require a valid postcode which is in Scotland. Each person will have to have a unique CHI Number[[2]](#footnote-2) where the format will have to be the same for each person. The simulated data will be displayed using appropriate graphs based on the condition chosen by the user.

### 4.2 Where can this software be applied?

This software is going to be designed for researchers and health professionals who want to improve their analytical skills.

### 4.3 Functional Requirements

The software will have to:

* Create a data set which is unique to a user.
* Let the data set to be retrieved to the user that has been simulated previously.
* Simulate data based on statistics gathered.
* Assign postcodes to records in the data set that are valid postcodes in Scotland.
* Create a unique CHI Number for each record in the data set with appropriate format.
* Display the records stored in the data set.
* Display graphical statistics based on each condition to the user.

### 4.4 Non-Functional Requirements

The software will have to:

* Have an easy to use user interface so that they user can access simulated data quickly.
* Contain a user interface that support various screen sizes.
* Load a data set automatically that has been previously simulated by the user on a specific device.

### 4.4 Approach taken to Solving the Problem

The approach taken to solve this problem was to create a website with generated user data and display statistics and records on the website using HTML5, CSS and JavaScript. I have decided to use those programming languages for my website since I am comfortable in programming in those languages. The other reason why I have decided to use those languages since it will allow me to provide the application to a wide range of users since it will be available on the internet.

Before web application development I had to find sources on age distribution in order to know how much percent of people have a specific age have based on gender. I needed this data since otherwise I wouldn’t know how many people I should assign to specific ages. Without this data the application would lose its purpose as the age groups would be all over the places and after applying conditions statistics it would result in unrealistic or in data that would never be possible to achieve.

After getting those statistics I had to get statistics on various conditions that affect population of Scotland. I would only accept data that provides an age distribution and gender in order to keep the generated data usable.

From there the data was cleaned up since I would only get a number of people that are affected by the condition but in order for the data to be useable by the application, I had to convert them into percentages. For example, for one of the conditions the contact rate was per 1000 people was 400 males and that would be converted to percentage and stored in an Excel File created by myself.

Then I have worked on the actual application. The first step was to design the actual system and the way it will work. This has been done by just drawing diagrams and ideas. The next step was to decide on how to store the simulated data so that the user is going to be able to access it multiple times without the need to re-simulate the data. I’ve decided store all the simulated data in local storage. From there it was just the case of programming the application until completion.

Chapter 5

## Software Description and Specification

The aim of this chapter is to describe how the application was designed. This will include what software development process was chosen, how was the structure of the data set was created for the application and how the graphical user interface was designed.

### 5.1 Software Architecture

### 5.2 Data set Design

The initial idea was to use MongoDB to store all the simulated data. However, I wanted to make the application user friendly. Therefore, I have decided to store all the simulated data in the user’s device local storage. This will allow users to access simulated data just by simulating it once and from there they can access it at anytime they would like. They will only have to re-simulate the data if they want a different data set or the local storage has been cleared and there is nothing stored there.

Before any storing I would create an object called person which would follow the following default layout:

|  |  |
| --- | --- |
| **Person Key** | **Person Defualt Key Value** |
| firstName | “” |
| secondName | “” |
| gender | “M” |
| age | age – this is a parameter |
| DoB (Date of Birth) | 000000 |
| postcode | “” |
| SIMD16\_Rank | “” |
| CHINumber | 0 |
| BMI | “N/A” |
| CaR (Condition) | “N” |
| DotSaST (Condition) | “N” |
| GA (Condition) | “N” |
| DA (Condition) | “N” |
| Di1 (Condition) | “N” |
| Di2 (Condition) | “N” |
| HT (Condition) | “N” |

As you can see for the persons key “age” I am passing in a parameter which sets an objects age. This is because this data is being create one by one until a certain number of people is reached with a specific age and then moves one to the next age. This makes data creation simpler as after that step I don’t need to do any changes to the persons age.

For persons key “gender” the default value is “M” standing for male. I have decided to set this as a default value as from there I will just have to apply female gender statistics. This is because let’s say there are 60% Females at an age of 5 and 40% males. I only must change 60% of people with an age 5 to “F” which sets their gender to female. This just makes the simulation cleaner as you only need to work with one gender statistic instead of two at the same time.

The rest of the keys default values are just initializations for the persons keys.

From there all the other statistics would be applied to the person objects.

Finally, after all simulation has been completed. The data set would be converted to a JSON string and later stored in local storage so that data can be accessed by the user at any time.

### 5.3 Graphical User Interface Design

Chapter 6

## Detailed Design and Implementation

The aim of this chapter is to describe in detail what programming languages, tools, environment has been chosen in order to develop the application. In the final topic I will discuss the challenges I had to face in the design and implementation of this application.

### 6.1 Choice of Programming Languages Used

For this application development I have used the following programming languages.

#### 6.1.1 HTML5

Asd

#### 6.1.2 CSS

Asd

#### 6.1.3 JavaScript

fadsf

### 6.2 Choice of Third-Part Tools

For this application I have used the following third-party tools.

#### 6.2.1 Strathclyde DEVWEB

Ads

#### 6.2.2 FileZilla

Asd

#### 6.2.3 ag-Grid

Asd

#### 6.2.4 Chart.js

Dsada

#### 6.2.5 Bootstrap

Fdsaf

#### 6.2.6 jQuery

Asd

#### 6.2.7 Google Chrome

asd

### 6.3 Choice of Development Environment

For this application I have used the following environments.

#### 6.3.1 Operating System: Windows 10

Asd

#### 6.3.2 Mobile Operating System: Android 9

Aasd

#### 6.3.3 IDE: Visual Studio Code

Asd

#### 6.3.4 Version Control: Git and GitHub

asd

### 6.4 Challenges of Design and Implementation

S

Chapter 7

## Verification and Validation

Ads

### 7.1 Verification

Sas

### 7.2 Validation

Sdadf

Chapter 8

## Results and Evaluation

Asd

### 8.1 User Evaluation

Sads

### 8.2 Evaluation of Planning

Asd

### 8.3 Evaluation of End Product Functionality

Asd

Chapter 9

## Summary and Conclusions

Asd

### 9.1 Summary

Asd

### 9.2 Future Work

Asd

### 9.3 Overall Reflection

Asd

### 9.4 Final Conclusion

Asdf

Bibliography

1. Cant, R. P. and Cooper, S. J. (2010), Simulation‐based learning in nurse education: systematic review. Journal of Advanced Nursing, 66: 3-15. doi:[10.1111/j.1365-2648.2009.05240.x](https://doi.org/10.1111/j.1365-2648.2009.05240.x) , [Accessed: 12th March 2019]
2. Information Services Division Scotland, Community Health Index (CHI) Number, [Online], Available from: <https://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Datasets/Patient-Identification-and-Demographic-Information/Community-Health-Index-Number/> , [Accessed: 12th March 2019]
3. National Records of Scotland, Mid-2017 population estimates Scotland, [Online], Available from: <https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates/mid-2017> , [Accessed: 12th March 2019]
4. Scottish Government, Scottish Health Survey 2015 volume one: main report – 8: Obesity, [Online], Available from: <https://www.gov.scot/publications/scottish-health-survey-2015-volume-1-main-report/pages/71/> , [Accessed: 12th March 2019]
5. Information Services Division Scotland, GP Consultations / Practice Team Information (PTI) Statistics, [Online], Available from: <https://www.isdscotland.org/Health-Topics/General-Practice/GP-Consultations/> , [Accessed: 12th March 2019]
6. Diabetes UK, Version 3. Revised March 2014, [Online], Available from: <https://www.diabetes.org.uk/resources-s3/2017-11/diabetes-key-stats-guidelines-april2014.pdf> , [Accessed: 12th March 2019]
7. ScotPHO, High Blood Pressure: prevalence, [Online], Available from: <https://www.scotpho.org.uk/clinical-risk-factors/high-blood-pressure/data/prevalence/> , [Accessed: 12th March 2019]

Appendix A

Appendix B

Appendix C

1. Body Mass Index [↑](#footnote-ref-1)
2. Community Health Index Number [↑](#footnote-ref-2)