Privacy of data visualisation and restriction of data sharing

Author: Ching-Chan Lin

Student Number: 2075360

Supervisor: Dr Parisa Eslambolchilar

School of Computer Science and Informatics

Cardiff University

December 2022

Abstract

In recent years, people have used wearable devices not only to find out how active they are but also to get their personal health information. The personal health information of visualisation allows people to observe their health status and share it on the internet or social media. Yet, it also gives rise to potential issues with the identification of personal information. For example, when people share their health data visualisation, their personal information could be identified by the data trends and analysis. Therefore, the project aims to create a web-based platform to find ways to improve data visualisation and reduce the risk of personal information becoming identifiable. The project creates interactive visualisations on a web platform and invites participants to compare them and fill in a questionnaire. Meanwhile, find out what people think about restricting data sharing options in visualisations and to understand how people want to share data. Finally, from the participant response, we learned that most participants felt that the risk to privacy could be lessened by either reducing the predictability of data visualisations or by allowing individuals to share their health data visualisations in different ways.

Acknowledgments

I would like to thank my supervisor, Dr Parisa Eslambolchilar, for her continuous guidance and great support throughout this project. I would also like to thank her for guiding my mind and attitude in this project which is to plan the project process by myself and keep track of project progress, rather than finish tasks passively.

Finally, I would also like to thank my friends, flatmates and family for their continuous support and willingness to help me through a lot of difficult times.

Contents

1. Introduction

With the boom in wearable devices and their rapid integration into people's daily lives, measuring, self-tracking, and quantifying all aspects of daily life has become easier, and the quantified self has become more common. The quantified self means people can use wearable devices or all kinds of tools to record various data from their daily activities as a basis for future self-improvement or health improvement. For example, sleep quality can be tracked through a wearable device that measures sleep time and physiological situation and then uses data visualisation to understand the self-sleep status. It can not only manage sleep time but also improve sleep quality to keep yourself in the greatest mental condition throughout the day. In addition, the data visualisation provides an overview of the data distribution, regardless of the trend of weight change, the interval of sleep duration, or the relationship between body consumption and intake in the long term. Therefore, data visualisation is a crucial way to help quantify the self and understand the physical condition. Nevertheless, it also exposes us to data privacy issues because there is a possibility that our personal information could be identified when we share our personal health charts with others. Therefore, this project aims to build a platform to evaluate which personal health data visualisations carry the risk of making personal information identifiable. In addition, the project also finds a solution to prevent personal information from being identified in personal health data visualisations.

Tory and Moller (2014) indicate that one effective support people gain from data insight methods is data visualisation. Jesus and David (2015) support that data visualisation helps people to understand and process data easily because it combines human cognition, data mining concepts, and human-computer interaction. At the same time, people usually search for health knowledge or share personal health information on the internet to understand or compare their health status to others. However, data visualisation might lead to personal information becoming identifiable, creating the risk of a leak in personal information, such as stress, emotions, sleep conditions and so on. For example, other people can compare the difference in daily sleep time from data visualisation, leading to the identification of a personal stage of sleep or sleep quality. Therefore, this project focuses on how to prevent the risk of personal

data being identified and leaked when sharing personal data visualisations.

1.1 Motivation

The issue of data privacy has recently come into focus, and people have also started to explore how to protect data, anonymise or encrypt their personal information and so on, to hide sensitive information. However, data visualisation can also identify personal information and cause privacy leaks. In order to verify and investigate personal information leaks, a web platform was built for this research to evaluate if personal information can be identified from data visualisation, and also to collect people's ideas to improve the graph and prevent the personal data identification problem. Also, for any researchers who are interested in improving personal data visualisation, this web platform may possibly help because it has compatibility to collect a wide range of data from different devices to receive users' feedback, thoughts, or ideas.

Thus, this project focuses on building a web platform with the ability to import different datasets. It can then produce different data visualisations from the dataset and invite participants to see which data visualisation they think is effective in preventing data privacy leaks. Moreover, this web platform can test the effectiveness of privacy protection, increase the speed of improvements to data visualisation, create a dedicated data visualisation for each type of data, and understand people's ideas and explore other data sharing methods which could prevent personal data from being identified.

1.2 Problem

Data visualisation can convert large, complex, or numerical data into a more manageable visual representation, allowing people to easily and quickly see and understand what the data represents, how the values relate to each other, and if there are any trends or outliers. Data visualisation can take many forms, such as bar charts, line charts, and so on. It not only presents the meaning

behind the data but also effectively increases the impact of the information, attracting attention and interest.

However, data visualisation can make data easier to decipher and when it is shared, others may be able to infer or identify personal information through trends and correlations or by comparing differences, which can lead to privacy leaks.

1.3 Solution

There are different types of wearable devices on the market, and data is collected from various wearable devices. Each device has its own different data visualisation for presentation. Therefore, building up a web platform for the evaluation of data visualisation can import various kinds of data collected from different devices and then make different types of data visualisation.

Participants were invited to help observe the different data visualisations, and their opinions and feedback were collected to analyse whether the data visualisations could be used to identify personal information. In addition, the web platform was used to analyse how people prefer to share data visualisations, which in turn formed an evaluation platform to explore how data can be visualised and shared to lower the risk of personal information identification. And it may possibly help in evaluating other data visualisations or developing new functions to prevent personal information from being identified in the future.

1.4 Project Aims and Objectives

This project aims to create a web platform to evaluate which data visualisations can prevent personal information from being identified. Furthermore, the project aims to collect people's thoughts and preferences on how to share personal health data visualisations in a way that can reduce the risk of data identification. Their opinions will possibly be used in the future to improve the privacy of

personal health data visualisations. Participants interacted with and compared different data visualisations from the same data set on the web platform to test if improved data visualisations could prevent personal information from being identifiable.

I am attempting to achieve the following objectives in this project:

- Build a web platform and use health datasets to visualise the same data in different ways. Then, investigate the issue of personal health information identification from data visualisation.

- Invite participants to compare and interact with different data visualisations and fill in the questionnaire. After that, analyse the feedback from the questionnaire to understand which data visualisations are effective in preventing personal information identification.

- Invite participants to fill out a form on data sharing restrictions, which is automatically uploaded to the database. Next, participants fill out a questionnaire and feedback is collected. Finally, the results are analysed to understand how people would prefer to share data to reduce the risk that their personal information could be identified.

2. Background

In recent years, wearable devices or tools used for quantifying daily life or monitoring health-related behaviours have become more and more popular. Jennifer et al. (2019) indicates that wearable devices collect data through sensors, for example, gyroscopes, or other devices, like digital scales and thermometers. It can be used to monitor various health behaviours and indicators such as physical activity, heart rate, body weight, body temperature and so on. As a result, the quantified self is becoming increasingly common as it is a method of tracking and improving the self by using various sensors to record personal physical activity status, such as pedometer, weight, blood pressure, mental state, sleep quality, etc.

Furthermore, the data collected from the human body can be used not only to improve health and fitness status but can also alert people when they are ill and track the causes of diseases. However, the massive volume of data from the collection can be difficult for people to process and understand. Thus, data visualisation lets people view trends, outliers, and patterns easily. It also helps people to cross-analyse data and explore data correlations. More and more people are searching for health knowledge on the internet or sharing health data with each other, especially on social networks, to obtain health information (Fox & Duggan, 2013). This is because social comparison is one of the most fundamental ways of understanding oneself. People compare their positions with those of others to make a realistic assessment (Ian et al. 2011). However, when people obtain personal health data from wearable devices, they search and share health information or visualisations on the internet to understand their health and compare it with others. Meanwhile, the data visualisation and the way of sharing might make personal information identifiable and further increase the risk of privacy leaks.

2.1 Data visualisation literature

This section will introduce what data visualisation is through the literature.

Data Visualisation displays data in graphics. In this way, complex data can be simplified into easy-to-absorb content; as Kennedy (2016) indicates, data visualisation is “the representation and presentation of data to facilitate understanding”. Jeremy et al. (2014) supports that data visualisation is the capacity to interpret data through visual charts and use a given data visualisation to explain specified issues in the data. Börner et al. (2016) indicate that visualisation makes it easier to identify patterns, trends, and correlations in data, making it more comfortable to analyse and understand. For example, most people are unfamiliar with statistics and basic statistical methods (mean, median, range, etc.). Therefore, it is hard to see a pattern when looking at statistical data, but the pattern becomes pellucid once it is made visual. Through data visualisation, massive data sets are organised to help people perceive and interpret them. It not only makes data easy to read but also helps

people to comprehend the meaning behind those numbers, enabling them to make the right decisions more quickly.

Moreover, Bertin et al. (1983) pointed out three stages of interpreting visualisation. The first stage is the external identification stage, the framework of visualisation that the reader identifies, such as labels and axes. The second stage is the internal identification stage, the reader identification of visual features or visual patterns. Finally, in the third stage, the reader can extract details from the visualisation analyses of the content. Ultimately, data visualisation should convey the meaning of the data accurately. It should not distort the information. For example, if two numbers are not equal, but look approximately the same in the visualisation, then this visualisation is not proper and even misguides the readers. In addition, data visualisation should be aesthetic because an excellent visual presentation can be more digestible. Visualisation information is often enhanced by good visual representation. In poor visualisation, the viewer will find it difficult to correctly examine images with unbalanced visual elements, incongruous colours, or other distracting features (Claus 2019).

2.2 Studies on how data visualisation can cause and prevent leaks

Data visualisation is a valuable way to explain the meaning of data. However, it can also lead to the identification of personal information and further the leakage of information to third parties, such as insurance companies, the healthcare industry, etc. Therefore, when sharing or using data visualisation, the risk of disclosure is minimised by visualising the data through anonymisation, using different ways of presenting visualisation, etc., to protect individuals' privacy and avoid third party access to personal information (Aritra et al. 2014).

This section presents literature on how data visualisations create identifiable personal information and cause privacy leaks and how to prevent these issues.

Line charts are typically used in data visualisation, and there are two advantages to this. The first advantage is that we can effortlessly catch patterns and trends. The other one is that we can predict future possibilities by analysing past data. However, other people can identify personal information and privacy

by combining patterns or trends from the line chart with external information (Aritra et al. 2014). For example, LifeLines, an electronic health record visualisation system, contains line graphs to represent a patient's medical history from the past to the present and can be combined with external information such as a patient's hypertension appointments, medical prescriptions, and health insurance claims to identify the patient (Plaisant et al. 2003).

Therefore, line charts not only make it easy to observe data trends and make predictions, but others can also do the same thing, leading to data leakage. Aritra et al. (2014) points out that from a privacy perspective, using layered visualisation, such as tree visualisation, is a good option as it hides certain data details. In other words, it makes the data trend or model less noticeable, and it may take more time to identify sensitive trends and infer private information. In conclusion, reducing the trend or predictability of data by different visualisations is a possible way of preventing personal data from being identified.

In terms of bar charts, they can demonstrate aggregated information rather than individual items. From a privacy perspective, while it could be assumed that this kind of aggregated visualisation is intuitively harmless, this hypothesis might not be consistently correct (Aritra et al. 2014). From the literature on hospital visits for people with diabetes, the bar chart in Figure 1shows that diabetes cases are strongly linked to certain characteristics. For example, the group aged 50 to 60 years old in male and female African Americans shows highlighted differences in the correlation between high readmission rates and the number of emergency visits. In addition, another notable case is that of white males weighing between 175 and 200, where the bar chart shows only one category with a readmission frequency more significant than 30. This means that it is easy to infer a person's diabetes status by knowing basic information about them, such as their ethnicity, age, and other quasi-identifiers. Moreover, it was not difficult to identify outliers and correlations between diabetes indicators such as emergency visits and readmission rates, given that only certain classifications in specific age groups had non-zero frequencies. Non-zero frequencies mean that in the same condition, there is only one bar that causes highlighted differences (Aritra et al. 2014). Thus, in some cases, bar charts with highlighted differences can lead to using data in combination with standard identifiers to identify personal data and explore the privacy of individuals.

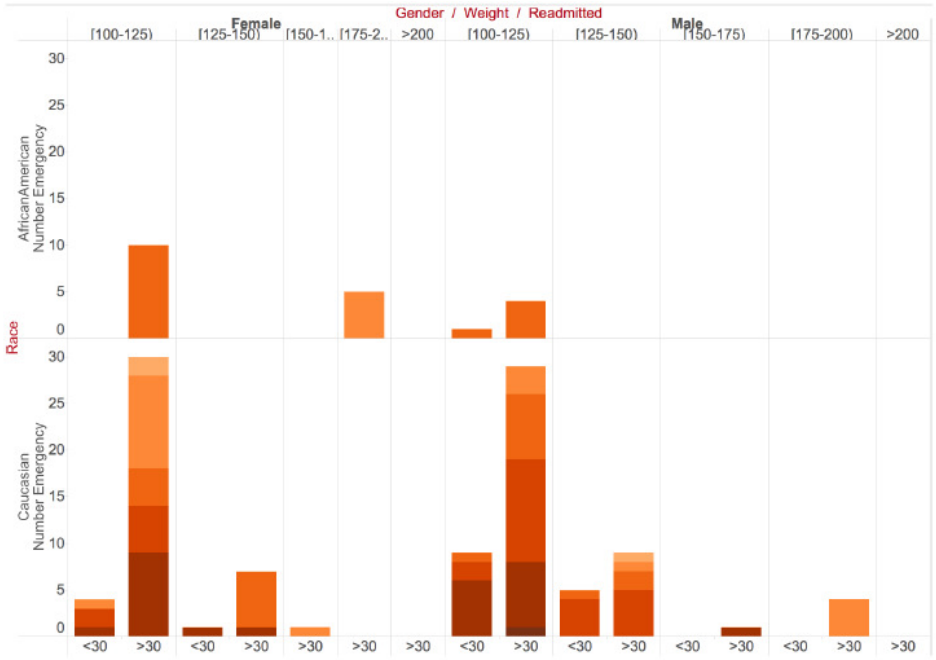

Figure 1. Bar chart representing hospital visits of diabetic patients

The anonymisation methods used in visualisation are mainly divided into data uncertainty and visual uncertainty. One way to use anonymisation in data is to introduce uncertainty into the data space, which ensures that a certain amount of data cannot be differentiated and makes it difficult to obtain sensitive or private information from it (Bhattacharjee et al. 2020). Aritra et al. (2012) supports that introducing additional uncertainty into the visualisation is possible to increase the privacy of the data. However, it also reduces the readability and usefulness of the visualisation, making it difficult for the viewer to understand the meaning that the data visualisation is intended to convey. In another study conducted by Shneiderman (1996), one of the methods applied in the quantification of visual uncertainty is to alter what visualisations can let users access based on the data's vulnerability, disclosure risk, and interaction constraints. In this way, it is able to guarantee a minimum level of privacy. In other words, through the coordinated multi-view and adjusting the parameters related to privacy, it not only helps in limiting the details displayed in the visualisation but also meets the user's needs and objectives. Therefore, it is possible to reduce the combination of data with standard identifiers by adding uncertainty to increase discrepancies and then prevent the identification of personal information in some cases in bar charts. But, what should be considered is whether the readability and correctness will be reduced by doing so.

Bhattacharjee et al. (2020) found that in visualisations, with the labels and value ranges on the axes, as well as the minimum and maximum boundaries of each cluster, etc., at least some information about the data is usually available and it is possible to infer data range or to compare whether the values are normal or not. Another study conducted by Aritra et al. (2012) demonstrated the efficiency and effectiveness of data integration and exploration by intruders as well as legitimate users. This can be achieved through interactive visualisation. Assessing data privacy and risk, therefore, requires consideration of computing resources, as well as human factors. Thus, the details in data can be viewed through interactive visualisation. Doing so reduces the need to view data details through visualisation frames such as labels, axes of coordinates etc. Finally, cross-linking different data or information is also one of the reasons why data visualisation and personal information can be deduced.

In an investigation by Sweeney (2000), attributes such as gender, birthday, and ZIP code were used to correlate the Cambridge, Massachusetts voter registration list with data from the National Association of Health Data Organizations (NAHDO). Six people had the same date of birth as Governor Weld's, and then excluding females by gender, only three of them were male, and finally, the only Governor Weld in the 5-digit postal code be identified. Thus, it was possible to identify Weld's personal health and identifiable information indirectly. Bhattacharjee et al. (2020) supports that personally identifiable information may be identified by cross-linking different datasets, even if the dataset is de-identified. In another study conducted by Motwani and Xu (2007), the authors indicated that sharing data by simply eliminating or reducing personally identifiable information such as name, email address, national insurance number, etc., may still allow for the re-identification of personal information through the link. An attacker could view personal information or privacy by linking two data sources. One is the characteristics of private information which can be acquired from the web or hospital data. The other is public data sources, like voter registration data. Through cross-referencing and linking these two data sources and standard identifiers such as gender, age, postcode, etc., an attacker will acquire personal information easily.

Therefore, the risk of identifiable data can be reduced in two ways: protecting or anonymising data and reducing the connectivity between data. For instance, modifying data attribute values is one of the techniques for anonymising input privacy by protecting attribute links and record links to practice privacy

protection. The methods for modifying data attribute values can be divided into two categories. The first category is called scrambling, which means that noise or meaningless data is added, so that aggregated information can be recovered, but not a single message. The second category is called generalisation and suppression, which means generalising data through aggregation or hierarchy, or by hiding data values.

In summary, this section has discussed four reasons that data visualisation may identify personal information: predictability, salient differences, axes and cross-referencing different datasets. Also, there has been some discussions about how to improve personal data visualisation to prevent privacy leaks in the meantime. The platform developed by this project aims to find the most appropriate data visualisation to reduce the identification of personal information without decreasing the readability and accuracy of the data.

2.3 Literature on data sharing

This section will introduce a common issue when sharing personal data and also provide some ways to prevent data leakage.

As technology improves and our lives become more connected, many people have become more conservative about sharing personal information and more worried about whether it will lead to privacy leaks (Horr and Joe 2017). A study by Morey et al. (2015) shows that many people believe that their privacy is being eroded by personal information requests from marketing environments and businesses. Consumers may feel that they are being unknowingly monitored and their personal information is being collected by companies, especially those who do not explain how their personal data is being used.

Another study conducted by Nam et al. (2006) indicates that many consumers feel they have no control over how government agencies or companies will use the data they share, so many are reluctant to accept requests for personal data sharing. In addition, there has been a long history of human-computer interaction in which systems that provide alerts about potential risks and privacy implications are not effectively communicated to users, leading to a great deal

of research into how to share information about security concerns to users effectively (Cristian et al. 2011). However, users often ignore the prompts given by the system or unwittingly reveal private information. Harbach et al. (2013) indicates that when a user sees a dialogue box from an IT security system about a potential privacy impact or risk, the information is often poorly understood or easily ignored by the user.

Users are often unaware of the risks associated with using social networks to post health information: they misjudge that the information they post online is accessible, they may not be aware of how others may use the information they post, and they are often unaware of what data may be shared with third parties (Williams, J, 2010). As a result, many studies have been conducted on how to protect data privacy when sharing data to prevent personal information from leaking. Barker et al. (2009) indicates that the key elements that make up data privacy are visibility, granularity, retention, etc. Data visibility is defined as who is allowed to access and reuse the data produced by the provider. For example, only doctors should have access to the medical information provided by patients, and it should not be available to third parties (e.g. potential future insurance companies) or be publicly accessible.

This is important for electronic documents because data can be transmitted to others via the network. In addition, limiting visibility reduces accessibility and allows only relevant people to manipulate or view the data. From the data granularity perspective, it is defined as the level of detail that the data owner can choose to provide. In other words, it allows the data owner to decide how much information to reveal in response to a particular query. For example, the granularity could restrict an age range or exact age, such as "child" or "adult", displayed. Thus, it is primarily able to present the level of anonymisation required to respond to legitimate access requests from users for a specific goal. As for the retention aspect, it is defined as the period after which the data cannot be viewed if the expiry conditions are met. The expiry condition is defined here as the time, duration, number of accesses, etc. When this condition is met, legitimate users can no longer access the data. Therefore, for all data collected, the retention period must be clearly stated, for example, how long the data can be used, when expired data must be deleted, or whether the retention period must be renewed by agreement with the data owner (Barker et al. 2009).

In conclusion, data leakage can be reduced by controlling visibility, granularity, and retention restrictions when sharing data. Furthermore, through the platform developed by this project, these methods will be tested to determine whether they are effective and feasible. Further methods of sharing restrictions for the future will also be explored.

3. Methodologies

Firstly, before starting the project, it is vital to understand and evaluate the design requirements and purpose of the overall project.

All the technical issues that may be involved in the development of the system must be considered, and it is crucial to think about how to validate and test the methods to ensure that there are no problems. A timeline must also be planned for each part of the project, knowing exactly when the deadlines are for each part in order to manage and push the project on schedule effectively. This is followed by an analysis of technical feasibility and a detailed understanding of the technical components to identify the most suitable tools or methods for the development of the project, followed by functional design and implementation and testing of the system for proper operation.

In this section, I would like to introduce the project development process and then discuss the purpose of the survey on the web platform.

3.1 Web platform development plan

Many tools were used when building the web platform, and this platform has two important parts. Firstly, it provides different health data visualisations from the same dataset. All personal information in the graph is anonymous, and the license of the dataset is CC0 public domain, which is free to use. The web platform created a function to let users view and even compare each graph. Additionally, the participants could also provide their ideas and preferences on how data can be shared.

Therefore, the purpose of the platform is to gather feedback from the volunteer participants, who evaluate the data visualisation on whether the trend and pattern can be distinguished and even possibly acquire further information by comparing the health data visualisation. The platform is also designed to explore people's preferences on methods to prevent personal data identification, like setting a restriction before sharing.

3.2 Project Management (PDCA)

First of all, this project put the PDCA methodologies into practice.

PDCA was developed by William Edwards Deming and it contains four concepts: Plan, Do, Check and Act. The plan refers to defining the objectives, goals, and targets. Do is the accurate execution of tasks as planned. Check is the process of assessing and evaluating the results of implementation. If there is a gap between the plan and the actual implementation, some suggestions should be raised to improve it. Action is a way of figuring out the improvement proposal based on the Check results. In other words, it is the process of modifying the practices and implementing corrective measures to optimise the system's quality in response to the improvements suggested in the Check stage (Knight and Allen 2012).

Therefore, PDCA makes it easy to plan the framework of the process and the tasks that need to be carried out. Furthermore, it allows researchers to learn from their mistakes, reflect on them, grow from them, and identify what they do well what they need to improve. This enables researchers to achieve the shortest development time, the best system quality, and the most straightforward system maintenance. So, the concept of PDCA was followed in this project.

The first step is Plan, which involves investigating and analysing what functions and requirements are needed and should be developed for the project. Then, a simple sketch design of the system should be created, as well as designs of the interface and functions from the user's point of view. After that, in the stage of DO, the system and UI are designed based on the requirements and plan formulated in the previous step. Finally, the whole project is divided into various

functions and modularised to facilitate and verify until there are no problems within each module. Next is Check, which tests the system's various logical flows, checks for possible bugs, or vulnerabilities in the system, and then makes any necessary revisions. Lastly, in the final step, Action, the module is finalised with various tests, bugs fixes, and model adjustments.

3.3 Introduction to the tools used in this project

The tools used for this project and the reasons why they were relevant will be explained in this section.

- Visual Studio Code

Visual Studio Code is a multi-platform open-source, free code editor developed by Microsoft for Windows, Mac OS, and Linux, which can be used not only to develop HTML, CSS, and Javascript but can also be extended to support Python, C/C++, Java, and other languages.

Therefore, Visual Studio Code was used to develop the web platform because it offers a lot of functionality and can be used with git for version control to increase productivity.

- Git and GitLab

Git is a decentralised version control system. It can keep a history of the process of modifying files. Thus, it also can undo edited files to their previous state and show differences in edited content.

So, in this project, GitLab, which is a Git tool for repository management systems, was used as a code management tool. It tracks the development process and manages code to make development easier.

- Bootstrap

Bootstrap is a front-end framework written in HTML, CSS and JavaScript that includes many pre-made components, such as common buttons, navigation bars, and interactive windows, which can perform many complex functions and styles with a specified short syntax.

So, using Bootstrap not only makes development more efficient and easier to maintain the code but also facilitates the layout of the web page to make it more aesthetically pleasing.

- Flask

Flask is a lightweight web application framework written in Python, mainly consisting of the Werkzeug WSGI toolkit and the Jinja2 templating engine. Flask is highly flexible and can be easily modified with different extensions to increase its functionality.

Therefore, this project used flask as the development framework for the web platform, which allowed flexibility and scalability to modify or extend new features to handle various needs.

- phpMyAdmin

phpMyAdmin is a web-based database management tool for MySQL. In other words, it is a web interface for managing MySQL databases. Therefore, it can manage the MySQL database remotely by using phpMyAdmin and easily create, modify, and delete databases and tables.

- Openshift

OpenShift is a container application platform that provides a cloud application platform with flexible and scalable features for deploying new applications on secure and scalable resources. OpenShift supports web application frameworks in all languages and has friendly and easy-to-use features to help applications complete quickly. Therefore, the tool was applied to this platform for this project.

3.4 Coding style

Good coding style increases the readability and robustness of a program. It not only makes the logic easier to understand, increasing the readability and maintainability of the project but also indirectly speeds up development. In

addition, good naming is like telling a story in code, as it is easy to read and can significantly reduce the time spent by developers writing useless notes.

Therefore, this project uses camel-case naming to make variables and function names more readable and maintainable. In addition, the description of the purpose and meaning of the code was added by writing code comments. Furthermore, the code was also modularized because maintenance and expansion are taken into account; hence, it can increase not only the extensibility but also the readability of the code.

3.5 Requirements & system design

In this section, the functional requirements and the system design method for this project will be introduced.

3.5.1 UI (User Interface)

UI stands for user interface, the design of the functionality on a page, taking into account the ease of use and the aesthetics of the whole design.

In addition, we focus on the presentation of the user interface, emphasising the visual aspect and usability. Therefore, when designing the UI, we not only focus on visual aesthetics and design aesthetics on style and looks but also consider usability and smoothness because the UI plays a vital role in shaping the user's experience.

The platform is used to investigate privacy issues in two aspects: the privacy leak on the visualisation of personal health data and the methods of data sharing restriction when sharing data visualisations. Therefore, the website was designed from the participants' perspective with a human-centred concept, and it should not be too complex to understand. The participants should find it easy to use. Plus, Bootstrap is used to arrange a front-end framework to beautify the web interface to achieve an aesthetically pleasing and straightforward goal in

UI design. Considering the UI design requirements, the overall website design was divided into five pages: Home, HealthData, ShareData, Register, and Login.

Firstly, the purpose and instructions of this survey are introduced in the Home page.

The HealthData page displays the various data visualisations used for this survey.

The ShareData page is used to simulate which restrictions users select when they share data visualisations. This enables users' preferences and thoughts to be investigated. As the user configures the option to share data restrictions, it is uploaded to the database for analysis. Therefore, this page is only displayed after logging in. It not only prevents users from making mistakes and submitting incorrect information but also keeps the scalability, which can be analysed in the future by the submission results which have their gender and age from the registration account.

The Login page will display an alert message when the login is successful or unsuccessful. The user's name will be displayed when the login is successful to remind users of their login status.

The Register page allows participants to create a new account by filling in their username, email, password, gender, and age. When a user enters a value that does not match the format, an error message will be displayed to alert them.

3.5.2 Dataset

This web platform must be able to import a dataset to evaluate different data visualisations. The main reason I chose the dataset from the Kaggle website, called FitBit Fitness Tracker Data, is that all personal data is anonymous, and the dataset is licensed under the CC0 Public Domain, allowing others to reuse the dataset freely. In addition, these datasets are generated by anonymous participants who use Fitbit wearables to collect their daily activities, including physical activity, heart rate, and sleep monitoring. Besides, they agreed to provide their quantified self-data, so this dataset is suitable for the research.

Finally, one of the anonymous participants in the dataset was selected to capture data on daily calories consumed, steps taken, sleep time, time spent in bed, weight, and BMI as a source of data visualisation for approximately three weeks.

3.5.3 Design data visualisation requirement

As mentioned in the background, the currently available solutions for preventing privacy leaks from data visualisation include increasing uncertainty in data visualisation, reducing trends or comparability, increasing efficiency and effectiveness in exploring and integrating data through interactive visualisation, using generalisation and suppression or perturbation in data visualisation, and reducing data detail. Therefore, two different visualisations of the same data set are displayed for presentation.

The first chart contains factors that cause data visualisations to be identified, such as data visualisations having predictability, saliency differences, coordinate axes and cross-referencing different datasets. The second diagram combines existing solutions with new methods for this survey, such as using pie charts or other visualisations to reduce predictability and comparability and replacing raw data with percentages or averages. Participants will compare two charts. This kind of comparison and collecting participants' feedback will be used to test whether the new method can prevent the identification of personal information in data visualisation.

3.5.4 Design restriction of data sharing options

According to the literature in the background section, visibility, data granularity, and retention are the existing methods for restricting data sharing. Therefore, this survey will expand on these methods and work on preventing the risk of personal information being identified from data visualisation.

The personal health data visualisation that can be accessed is subject to the number of views and the chosen expiry date. Which age and gender can access the data depends on the granularity and visibility settings. The participants will

set the values, and the feedback collected from the questionnaires will be used to analyse whether the method effectively improved privacy protection.

3.5.4 Questionnaire design

The questionnaire was divided into two parts.

The first part of the survey was used to investigate people's perceptions of data visualisation by comparing two data visualisations from the same data. In this part, the participants interacted with and compared two different visualisations on a web platform as well as completing a questionnaire to find out which method they think is the best way to stop personal information from being identifiable in data visualisation. In other words, the questionnaire investigated the participants' thoughts about the first data visualisation, then the second data visualisation. Lastly, it asked for their opinion after comparing these two data visualisations.

The second part of the survey investigated participants' views on limiting data sharing options, e.g. limiting the number of views, days of sharing, age and gender, by inviting them to fill in the data sharing options on the web platform and then complete a questionnaire to investigate which method was the most and least effective, and whether they had other ideas for limiting data sharing options.

The questionnaire used the Likert Scale, which takes five levels of options from strongly agree, agree, neutral, disagree and strongly disagree to assess overall attitudes.

3.5.6 The purpose of the survey

The questionnaire for the survey was divided into two parts.

Firstly, this survey aimed to test whether participants could see the trends and highlight differences in the data visualisations or find other elements in the data

visualisations that could identify personal information. Meanwhile, the project investigated whether the improved data visualisation effectively prevented personal information identification.

Secondly, the survey aimed to investigate participants' views about what restrictions they would like to set before sharing their data visualisation, such as limiting the number of views, setting an age or gender limit, etc.

This research project involved using mailing lists or social media posts to recruit participants. The project involved 20 participants and produced 20 effective questionnaire results. All participants understood the number of questions, the purpose of the project, and also how much of their time the project would take through the participant information sheet. In addition, the questionnaires were collected only after informed consent was obtained from the participants. Moreover, all participants were informed that they could leave anytime they wished, and that the data collected was anonymous.

After the participants agreed to get involved in the project, they received the link to access the web platform. Once they accessed the web platform, they saw that each question had different data visualisations from the same dataset. The participants could then compare those charts and fill in the questionnaire. After finishing the first part, the participants were asked to click the ShareData button on the navigator bar to access the second section of the survey about the restriction of data sharing. In this section, the participants were asked to configure the restriction, such as limiting the number of accesses, setting an age or gender limit and so on, depending on their preference. And those instructions were written up in a guidance document, so all participants were able to finish the questionnaire after reading the guidance document.

3.6 Implementation

The previous section organised the web platform into five pages: Home, HealthData, ShareData, Register, and Login. The process of implementing each page will be introduced separately below.

Before starting coding, a repository was created for the project, and it was synced to the local repository as a basis for development. This not only made it easier to debug when problems occurred but also allowed it to compare the file modification process from the history. Furthermore, it also made it easy to compare the differences and then verify the details step by step, maintaining the quality of the code and speeding up development. Since the master branch was used for deployment, it was crucial to ensure that the code in this branch had been tested and no problems existed. When writing code, it should be tested locally before being transferred online, e.g. create a local database and virtual environment, and test it, then transfer the local database to the online database and deploy the whole project to openshift. Finally, the web platform used the Flask framework, which is written in python. The front end was developed by using HTML and Jinja2 with Bootstrap, and the back-end database was supported by phpMyAdmin.

- Home page

First, the design and implementation of the Home page will be introduced.

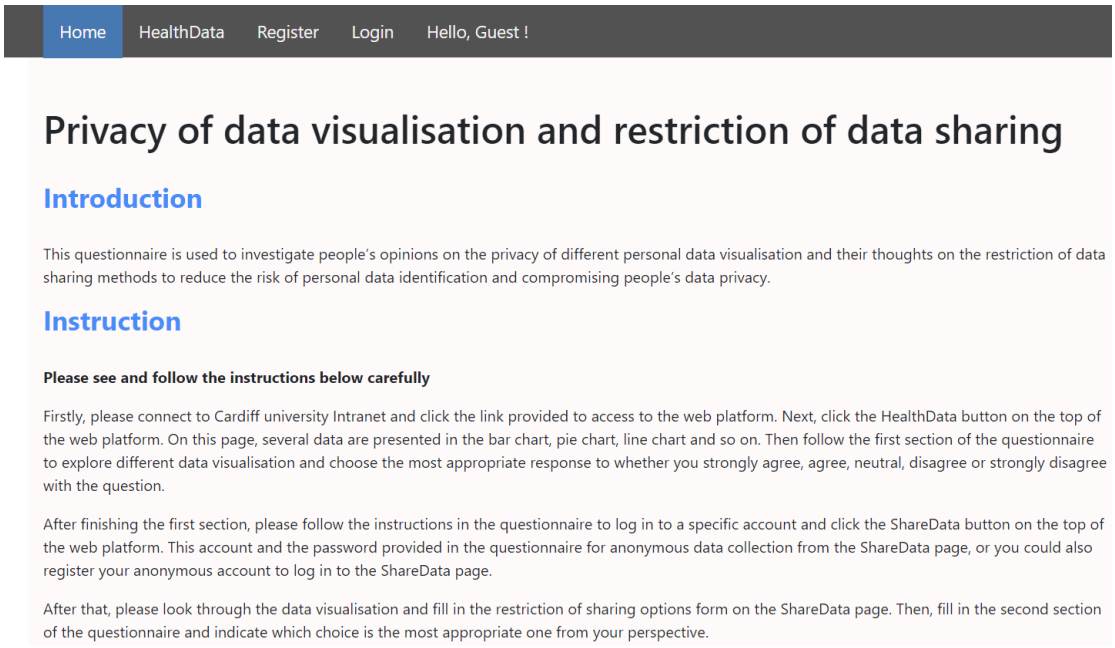

Figure 2. Home page

As shown in Figure 2, the navigation bar helps users to select the page they want to view visually. When the mouse is hovering, the page buttons will display different colours to enhance attention, beautify the picture, and make it easier for users to view the selection. In addition, Home is placed on the far left so that users can return to the Home page anytime. The Home page mainly describes

the topic, introduction, and operation of the survey data visualisation. Therefore, different font sizes and colours are used for layout, and Bootstrap is used for beautification so that users can easily understand the purpose of the survey and the operating instructions. A button at the bottom of the page redirects to the top which makes it easy for users to return to the top of the page.

- HealthData page

Here is the introduction to the HealthData page implementation.

The format and columns of the database should be planned first, and then the data should be uploaded to the database so that the dataset can be used for data visualisation in the future. After that, only the data in a database should be updated. In this way, data visualisation can be continually renewed without changing the code repeatedly. As the data in the dataset covers approximately three weeks, the data obtained from the database is processed and averaged daily before being visualised.

There are many different drawing libraries for Python. Initially, I used matplotlib because it is a common visual drawing library for Python and can draw typical graphs, including line charts, bar charts, etc. However, after testing, it was found that it is impossible to achieve the interactive data visualisation requirements as expected. Therefore, different visualisation libraries were tried, and finally the Plotly visualisation library was used. Plotly is a data visualisation library based on the D3.js framework that generates interactive diagrams that can be displayed in the web form on a browser or saved locally as a figure. Compared to matplotlib, Plotly makes it easier to draw different interactive data visualisation, and it is more aesthetically pleasing. Moreover, it is easy to use and generates frames for various charts, including interactive effects.

The implementation of the data visualisation was then started by using the data of the average daily steps and the number of calories consumed per day, presented in two data visualisations: the first one used a bar chart in figure 3, and the second, figure 4, used a pie chart.

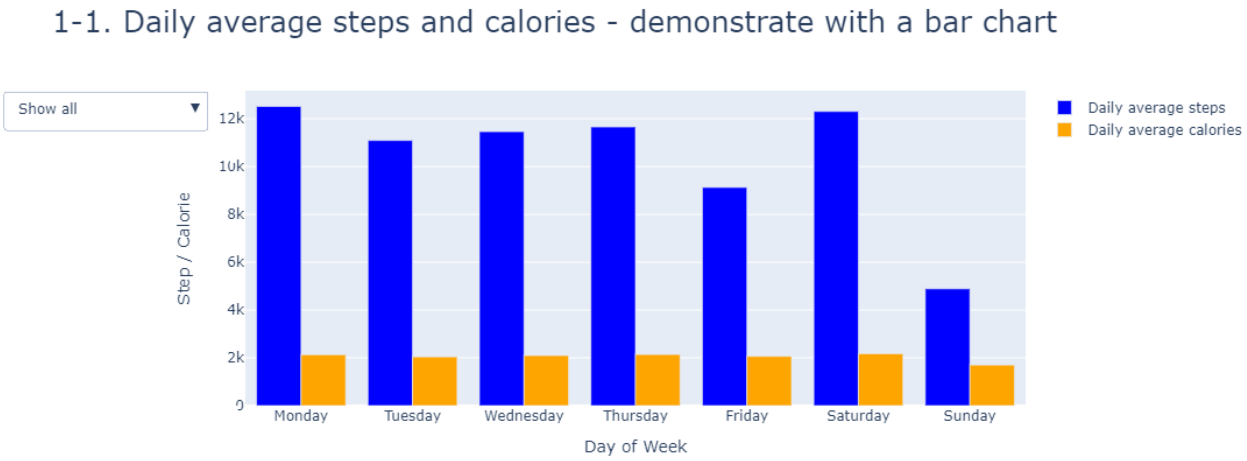

Figure 3. Bar chart

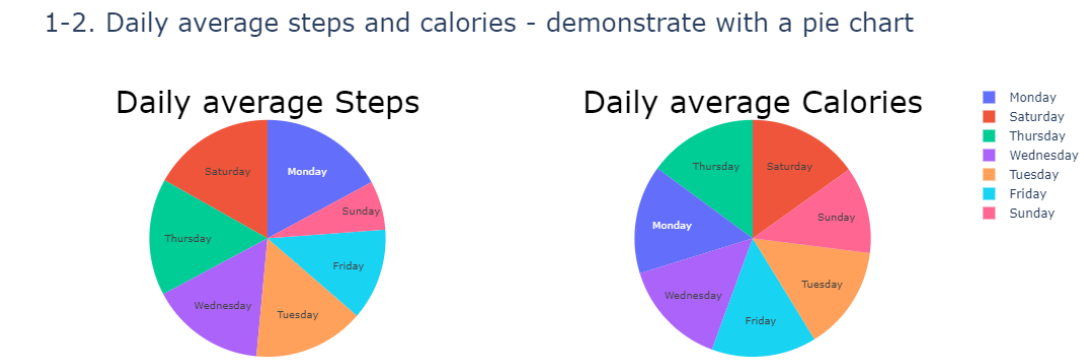

Figure 4. Pie chart

The data on sleep time and time spent in bed was then presented on two data visualisations: the first one uses a line chart with a bar chart in figure 5, and the second uses a violin chart in figure 6.

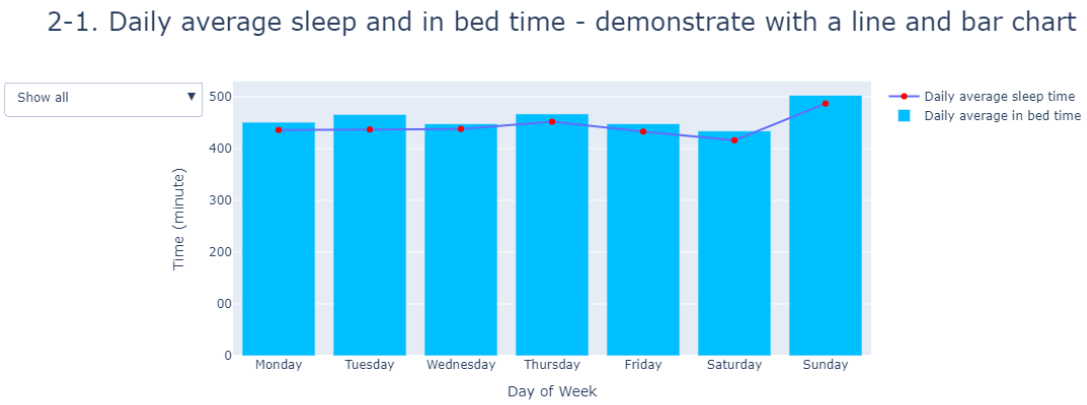

Figure 5. Line chart with a bar chart

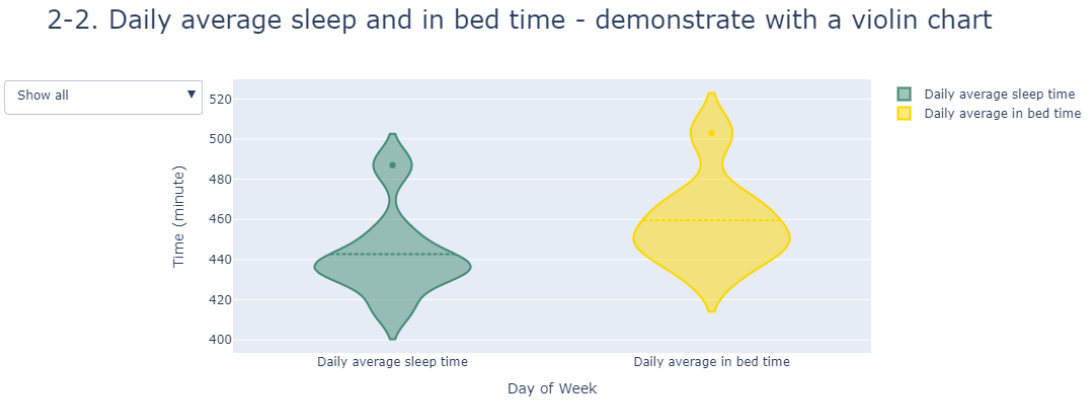

Figure 6. Violin chart

Finally, the weight and BMI data were presented in two data visualisations. The first data visualisation in figure 7 uses a line chart, and the second data visualisation uses a donut chart as shown in figure 8.

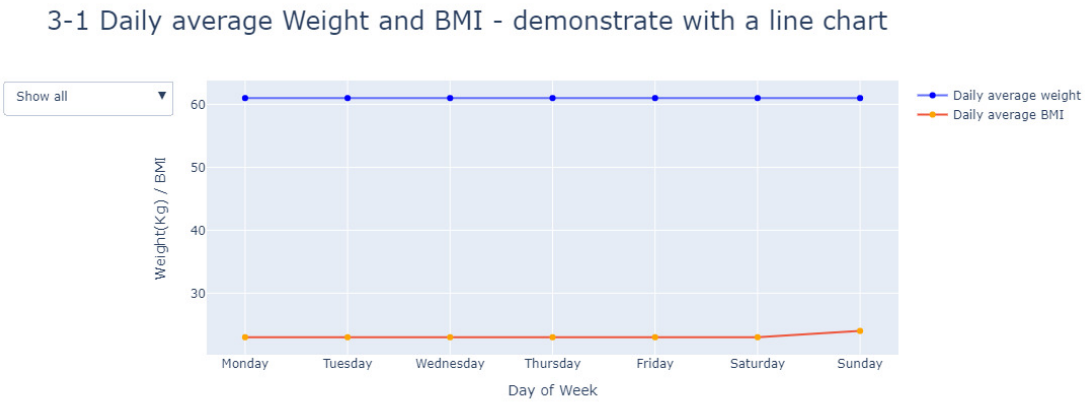

Figure 7. Line chart

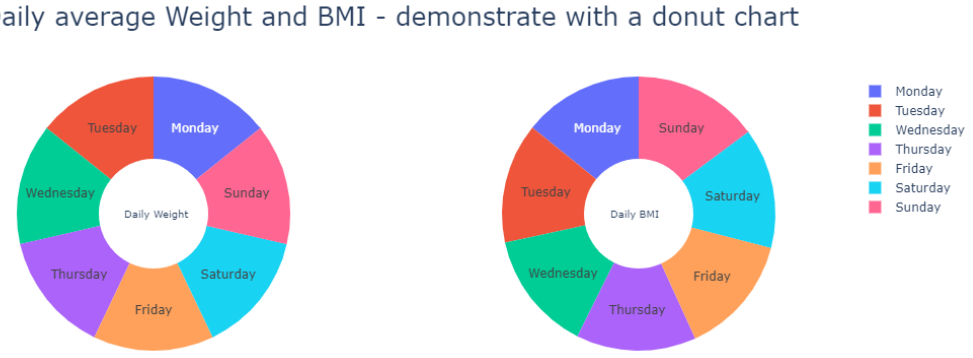

Figure 8. Donut chart

Each data visualisation has interactive features, such as detailed values and

details of the data when the mouse hovers over the chart, a drop-down menu on the left side to switch between different data so that you can look at a single data set or combine them, etc. Finally, the data visualisation was rendered to the front end via Jinja2 and displayed on the web page.

- ShareData page

This page investigates which kind of restrictions users choose when sharing a data visualisation. In addition, this page requires the user to log in. The navigation bar will display ShareData when the users sign in successfully.

This page, figure 9, visualises the average daily calorie consumption and investigates how sharing restrictions are configured when this data is shared visually. The visualisation restriction form sets the sharing expiry conditions for the number of views, days, age, and gender. For example, if you set the number of views to 3, the data visualisation can only be viewed three times. If you set the number of views to 0, the data visualisation can be viewed all the time. The number of days is the number of days the data visualisation will be shared. Age is the age of the person allowed to view the data visualisation. Finally, there is the number of days men, women, or any other gender can view the data visualisation.

Figure 9. ShareData page

In addition, the input format is checked using a regular expression. When the user enters a value that does not match the format, an error message is displayed to inform the user, as shown in Figure 10.

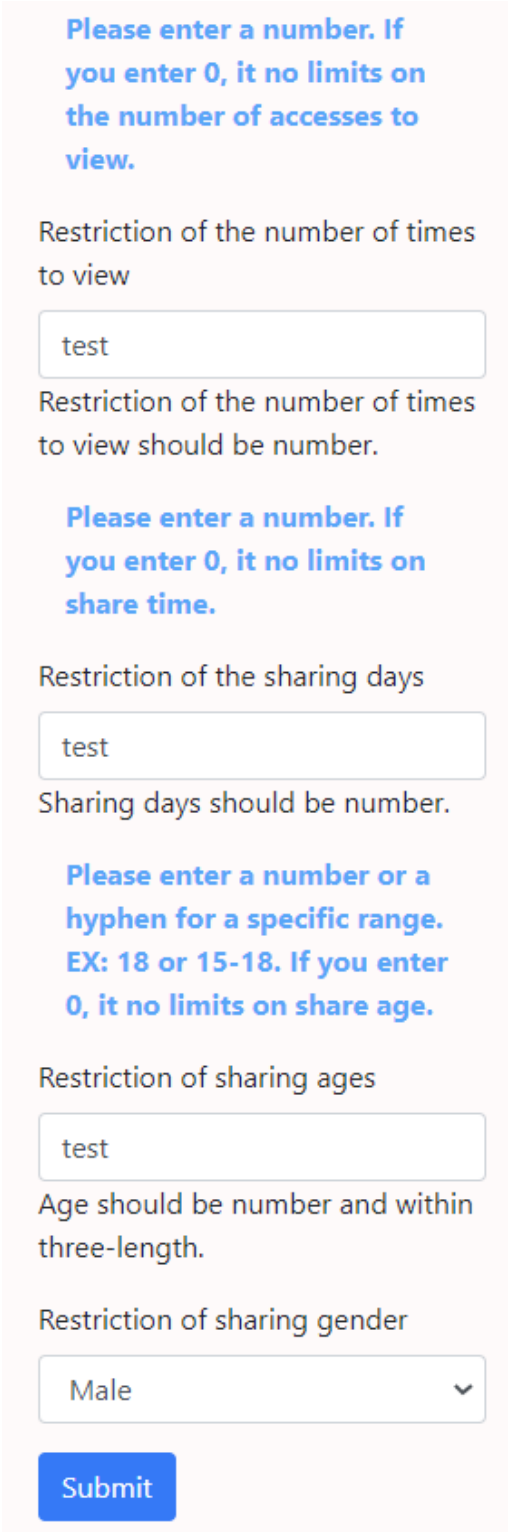

Figure 10. ShareData page with error message

- Register

When a user enters their information for registration, the input format is checked through a formal representation to help prevent errors in each part of the user's input. If it is incorrect or out of form, an error message is displayed to allow the

user to know what is wrong, as shown in figure 11. For example, the username field checks that the input includes less than 20 characters and no special characters. The email field detects the input format and prevents typing errors or formatting problems. The password field checks whether the input is within 20 characters. The age field checks that the number entered is has no special characters. These measures help to promote data security and avoid human error.

In addition, the password must be entered twice to check that there are no errors. In addition, passwords are encrypted with a hash to increase password security. Each field is also protected and validated with a csrf_token to promote safety and enhance user privacy.

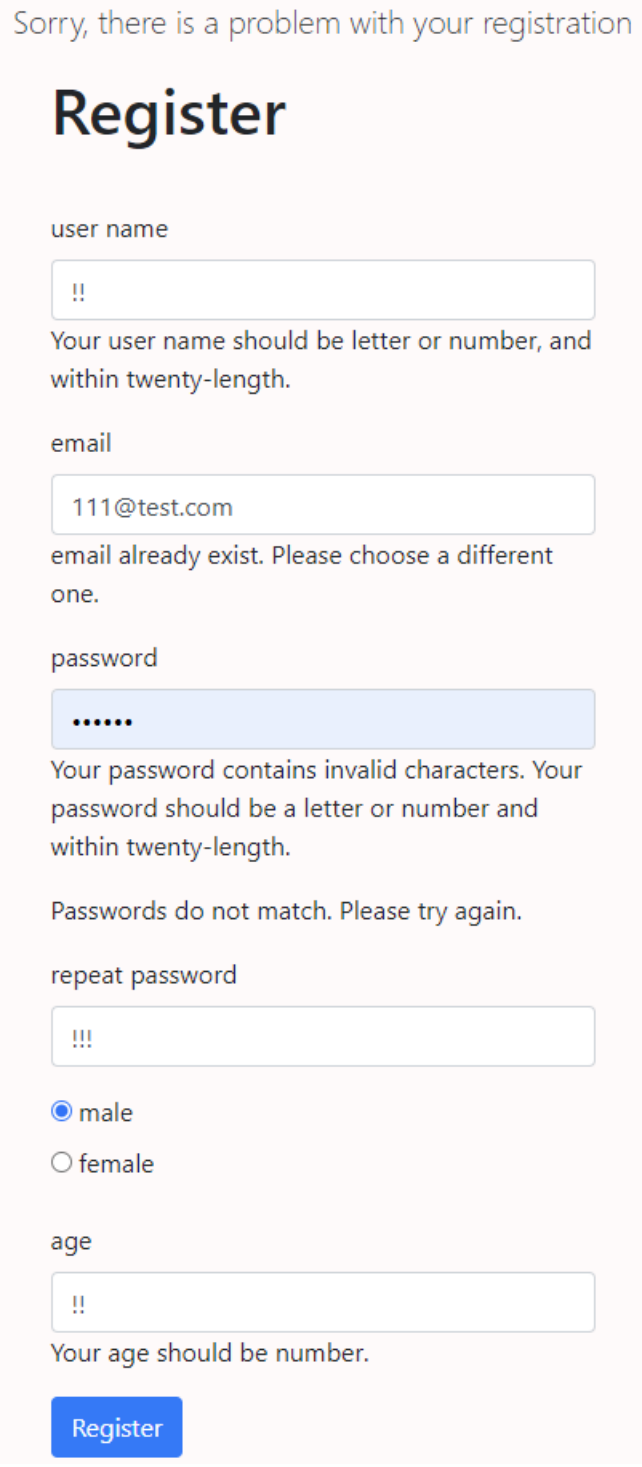

Figure 11. Register page with error message

When registration is successful, a successful registration message is displayed, and users are taken back to the Home page with the username displayed in the navigation bar. Please see figure 12 below.

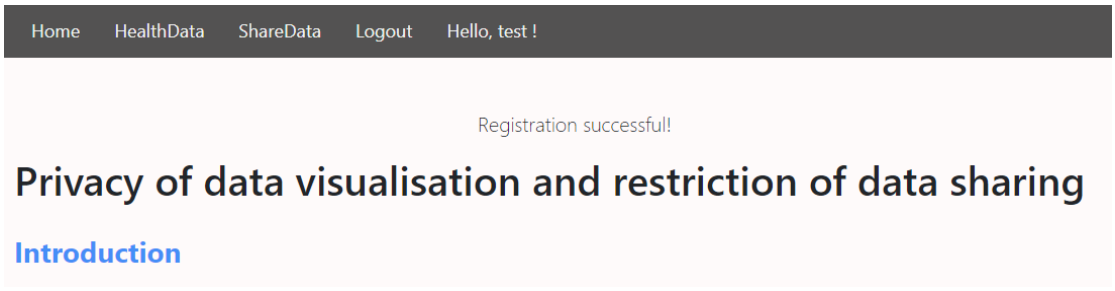

Figure 12. Register successfully

- Login

When a login error occurs, users are directed to an error page, and an error message is displayed to inform them that the email or password provided is incorrect, as shown in figure 13.

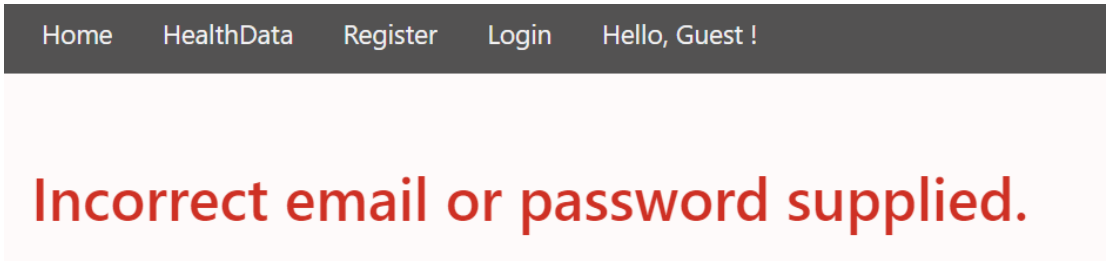

Figure 13. Login fail

When the login is successful, the login success message is displayed, and the user is taken back to the Home page with their username in the navigation bar, like in figure 14.

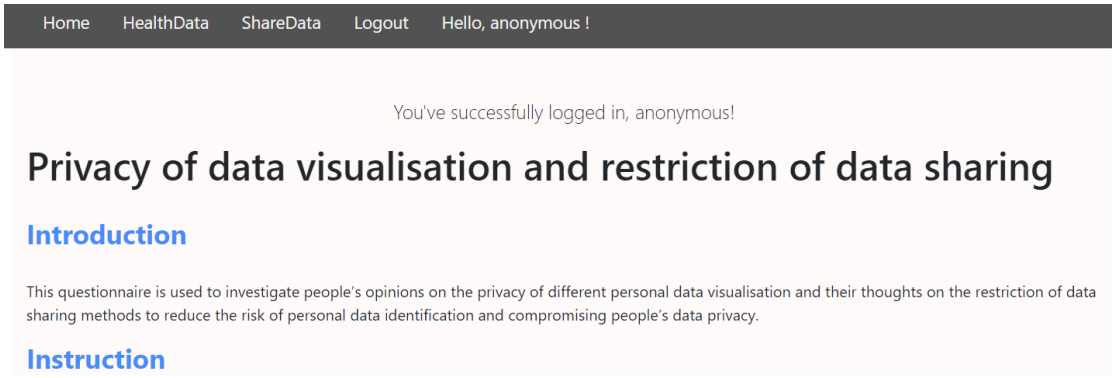

Figure 14. Login successfully

It can be seen in figure 15 that when logging out, the logout message is displayed, and the user is returned to the Home page. Guest is also then displayed in the navigation bar.

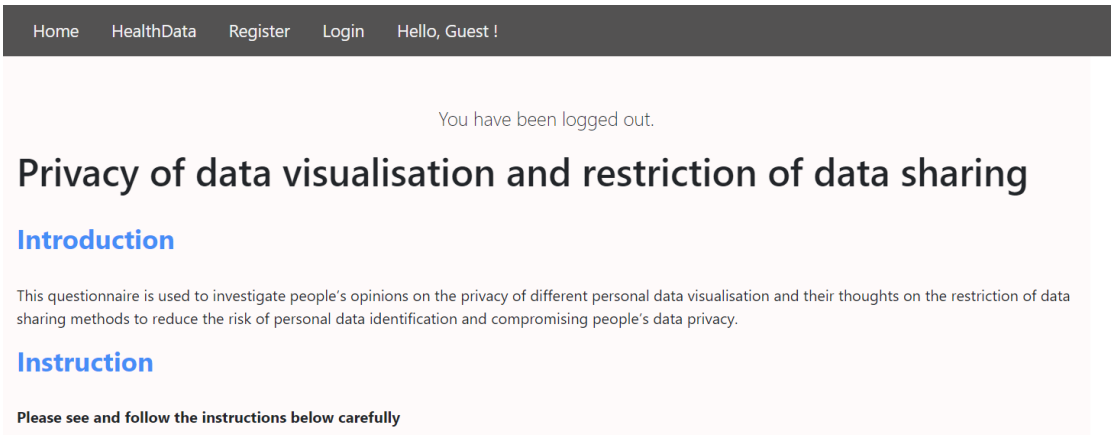

Figure 15. Log out account

4. Evaluation

It was essential to ensure that the code was written to meet all requirements according to the project objectives and functional goals. After local testing, all branches were merged into the main one to create a deliverable version. Next, the code was deployed on openshift for final testing.

The project created a user story (Appendix A) that envisaged all the situations that users may encounter when operating the web platform and provided them with functional requirements. Therefore, the user story not only understood all possible scenarios from the user's point of view but also included additional acceptance criteria, including functional and non-functional requirements, to meet the user's needs. Based on the user story, test cases were also created for users to ensure the usability of the web platform. It not only made sure users could operate it without problems but also checked if the code written had implemented all functional and non-functional requirements (Appendix B). Finally, the test case successfully achieved all the functional goals in the project scope.

Overall, the project was successful in creating a data visualisation evaluation platform and investigating participants' ideas on which data sharing restrictions, like limiting the number of days it would be shared for or limiting the number of times it could be viewed, can be applied to prevent personal data identification when sharing. All in all, the project also achieved all the intended objectives.

4.1 Analysis of questionnaire feedback

In this section, the feedback from the questionnaire is analysed, and the participants' thoughts are summarised.

The participants were asked to finish the questionnaire. There were 20 participants in total who completed the questionnaire, and 20 questionnaires were effective as well. The questionnaire consisted of 37 questions, divided into two parts. The outcome of the feedback was presented with percentages for

each question. All the feedback and responses to the questionnaire are attached in Appendix C.

The first part presented the same data in two different ways. Moreover, the second visualisation was an improved chart to the first visualisation. This part also explored the participants' thoughts on visualisation by analysing the participants' responses and summarising their feedback. The second part explored the methods and preference settings of data sharing restrictions. After participants completed the data sharing restrictions form, the feedback from the questionnaire was analysed and combined with the values achieved by the participants to see how the participants preferred to share data.

The questionnaire for part one will be analysed first. In this section, there were six visualisations and 37 questions, and each question had a 5-point Likert scale, ranging from strongly agree, agree, neutral, disagree and strongly disagree.

The first visualisation, the bar chart in Figure 3, showed the average number of steps taken per day and the number of calories consumed per day. The questions asked were as follows:

Q1 : Do you agree that the bar chart makes it easy to compare differences or highlight numerical correlations between steps and calories, e.g. the number of steps is proportional to calories consumed?

Q2 : Do you agree the bar chart makes it easy to observe the trend of steps or calories?

Q3 : Do you agree the bar chart displaying different unit values on the Y-axis reduces readability?

Q4 : Do you agree the mouse hovering over the chart to show data of values is better than marking the values on the bar chart, which increase the privacy of personal information?

Q5 : Do you agree the chart can explore more details and make a comparison by showing steps and calories separately or together?

Q6 : Do you agree it is better to show average values rather than daily values to reduce data detail and reduce personal data privacy?

Figure 16 is an analysis of user feedback for questions 1 to 6. According to the bar chart in figure 3, up to 90% of the participants felt that they were able to tell the significant difference and the highlight correlation by comparing the data visualisation. Besides, 75% of the participants agreed that the visualisation in the bar chart makes it easy to observe data trends which may make data become predictable. On the other hand, it can be seen that 70% of participants felt that displaying different unit values on the Y-axis would reduce readability. Furthermore, 75% of participants felt that hovering the mouse over the chart to show numerical details would increase privacy. In addition, 85% of participants felt that visualisation could explore and compare more data detail by displaying different data visualisations either individually or together in an interactive way. Lastly, 75% of participants felt that displaying data as an average would reduce data detail.

In conclusion, visualisation has the characteristics of salient differences, predictability, and cross-representation of different data. Therefore, personal information has the potential to be identified from data visualisation.

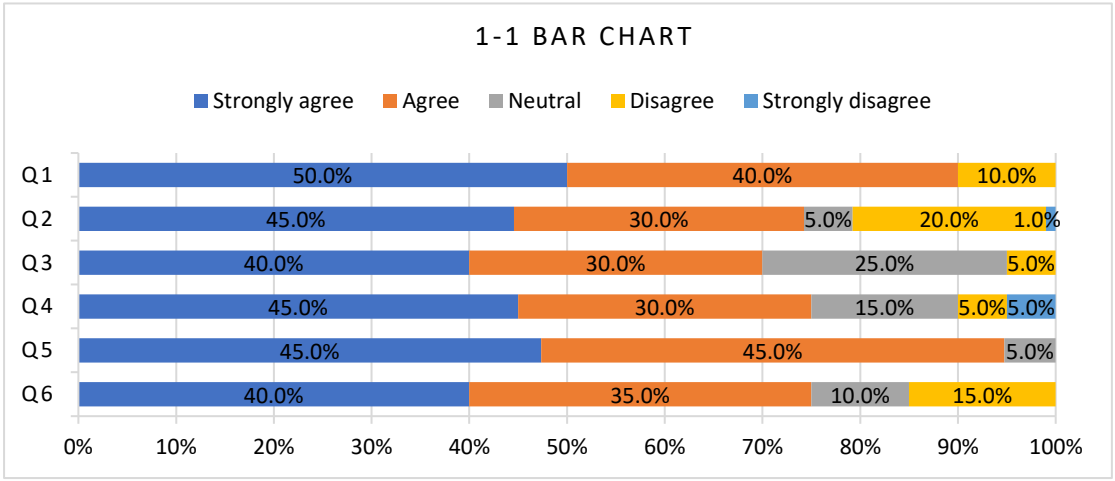

Figure 16. Bar chart feedback analysis

The second visualisation is shown in the pie chart in Figure 4. It is an improved visualisation of the bar chart in Figure 3. The following are the questions for this section.

Q7 : Do you agree the pie chart makes it easy to compare differences or

highlight numerical correlations between steps and calories, e.g. the number of steps is proportional to calories consumed?

Q8 : Do you agree the pie chart makes it easy to observe the trend of steps or calories?

Q9 : Do you agree the mouse hovering over the chart to show data of values is better than marking the values on the pie chart, which increase the privacy of personal information?

Q10 : Do you agree it is better to show average values rather than daily values to reduce data detail and reduce personal data privacy?

Q11 : Do you agree it is better to show percentage values rather than actual values to reduce data detail and reduce personal data privacy?

Q12 : Do you agree it is more difficult to predict trends or differences using a pie chart than the bar chart in Figure 1-1?

Figure 17 is an analysis of user feedback for questions 7 to 12. As shown in Figure 4, 70% of participants felt that they could not distinguish the differences or highlight numerical correlations in the pie chart. Plus, 70% of participants did not consider this visualisation to be predictable enough to observe trends in the data. In addition, 70% of the participants thought that hovering over the graph to show details of values would increase personal privacy. 85% and 80% of the participants thought the data details could be reduced by displaying the average and percentage values. It can be clearly seen that up to 90% of the participants thought that the pie chart in Figure 4 made it more difficult to predict the trend than the bar chart in Figure 3.

To summarise, the pie chart in Figure 4 does not highlight data differences or have predictability. Therefore, the pie chart more effectively protects personal information from being identified visually than the bar chart in Figure 3.

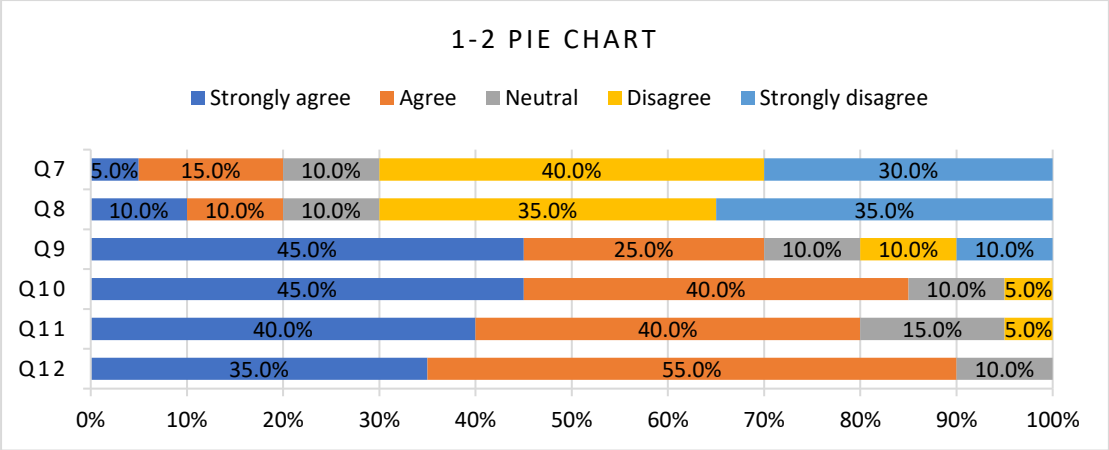

Figure 17. Pie chart feedback analysis

The third chart combines the line chart and bar chart, as shown in Figure 5. The chart demonstrates the average daily sleep time and the time spent in bed with the following questions.

Q13 : Do you agree with the bar chart combined with the line chart makes it easy to compare differences or highlight numerical correlations between sleep time and in the bed-time?

Q14 : Do you agree the bar chart combined with the line chart makes it easy to observe the trend of sleep or in the bed-time?

Q15 : Do you agree the mouse hovering over the chart to show data of values is better than marking the values on the chart, which increase the privacy of personal information?

Q16 : Do you agree the chart can explore more details and make a comparison by showing sleep and in the bed-time separately or together?

Q17 : Do you agree it is better to show average values rather than daily values to reduce data detail and reduce personal data privacy?

Figure 18 is the analysis of user feedback. 90% of the participants thought that the diagram could compare data differences or highlight numerical correlations by merging the line chart and the bar chart in Figure 5. Furthermore, 95% of the participants agreed that the visualisation was predictable enough to observe trends in the data. Nevertheless, 65% of participants felt that the mouse

hovering over the graph to display numerical details would increase privacy. 85% of participants felt that more details of the data could be explored and acquired by cross-referencing those two different data visualisations. For example, by comparing the two graphs interactively, either individually or in combination. 75% of participants felt that displaying data as an average would reduce data detail.

Finally, it can be concluded that the participants' perception of this visualisation is similar to the bar chart in Figure 3. As long as the personal health data visualisation has significant differences, predictability, and details that can be acquired by cross-referencing different data, there is still the potential for personal information to be identified.

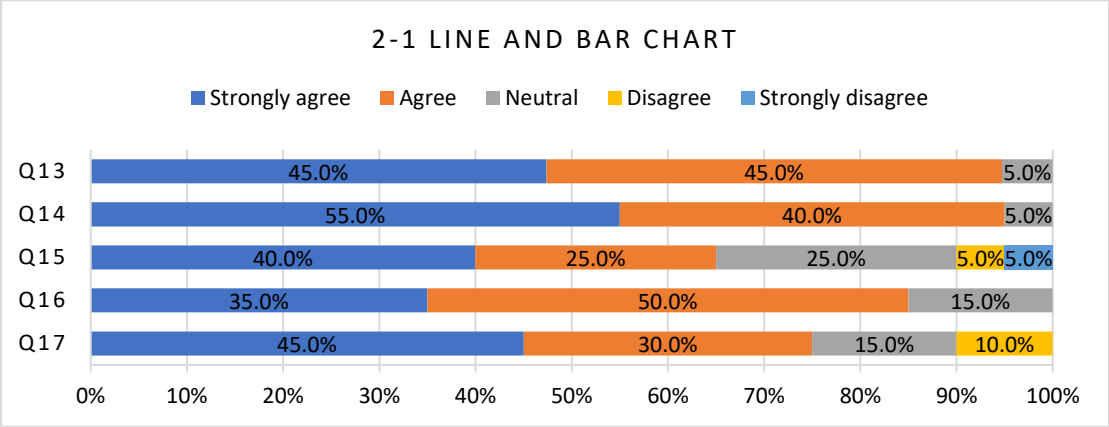

Figure 18. Line and Bar chart feedback analysis

The fourth visualisation is the violin chart, as shown in Figure 6. The fourth chart combines the line and bar chart, as shown in Figure 5.

Q18 : Do you agree the violin chart makes it easy to compare differences or highlight numerical correlations between sleep time and in bed-time?

Q19 : Do you agree the violin chart makes it easy to observe the trend of sleep or in the bed-time?

Q20 : Do you agree the chart can explore more details and make a comparison by showing sleep and in the bed-time separately or together?

Q21 : Do you agree it is more difficult to predict trends or differences using a violin chart than the bar chart combined with a line chart in Figure 2-1?

Figure 19 is an analysis of user feedback which indicated 65% of the participants thought that it was not easy to tell the differences or highlight numerical correlations when data is presented in a violin chart. 65% of participants did not consider the visualisation to be predictable enough to observe data trends. 75% of participants felt that the visualisation was able to explore and compare more details of the data by displaying the different data visualisations either individually or together in an interactive manner. 90% of participants felt that the violin chart in Figure 6 was less predictive of data trends than the line chart combined with the bar chart in Figure 5.

Overall, it may be said the pie chart in Figure 3 does not highlight data differences or have predictability; hence, the violin chart is more effective in protecting and preventing personal information from being identified visually than the line chart combined with the bar chart in Figure 5.

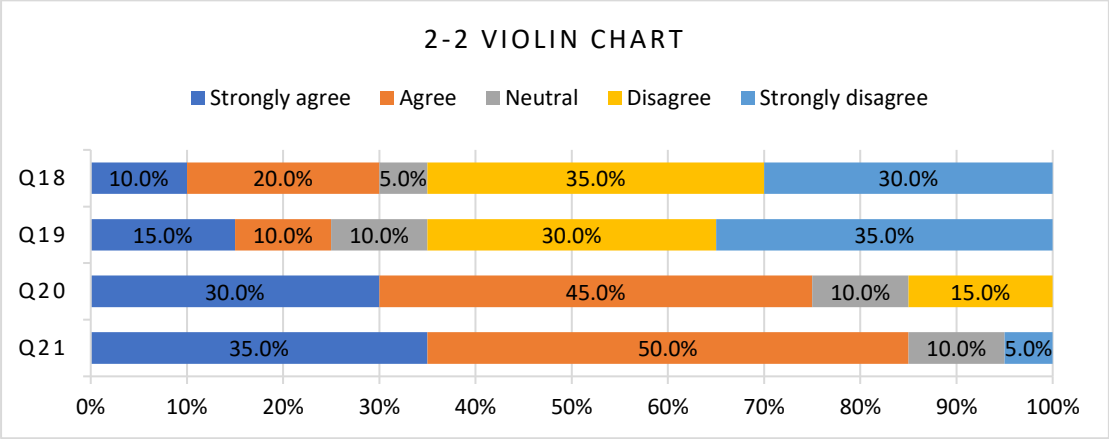

Figure 19. Violin chart feedback analysis

The fifth visualisation is the line chart as shown in Figure 7. The chart showed the average daily weight and BMI data with the following questions.

Q22 : Do you agree the line chart makes it easy to compare differences or highlight numerical correlations between weight and BMI?

Q23 : Do you agree the line chart makes it easy to observe the trend of weight and BMI?

Q24 : Do you agree the line chart displaying different unit values on the Y-axis reduces readability?

Q25 : Do you agree the mouse hovering over the chart to show data of values is better than marking the values on the line chart, which increase the privacy of personal information?

Q26 : Do you agree the chart can explore more details and make a comparison by showing weight and BMI separately or together?

Q27 : Do you agree it is better to show average values rather than daily values to reduce data detail and reduce personal data privacy?

Figure 20 is an analysis of user feedback. 85% of the participants felt that the data differences or highlighted numerical correlations in the line chart in figure 7 could be deduced. And up to 90% of the participants agreed that this visualisation was predictable enough to observe data trends. 85% of participants felt that displaying different unit values on the Y-axis would reduce readability. 75% of participants felt that hovering the mouse over the graph to show details of the values would increase privacy. 85% of participants felt that more data details could be explored and acquired by comparing the two graphs interactively, either individually or in combination. 80% of participants felt that displaying data as an average would reduce data detail.

Overall, based on the participants' feedback, the visualisation is similar to the bar chart in Figure 3 and the line chart combined with the bar chart in Figure 5. It indicated that as long as the personal health data visualisation has significant differences, predictability, and details that can be acquired by cross-referencing different data, personal information may be at risk of identification.

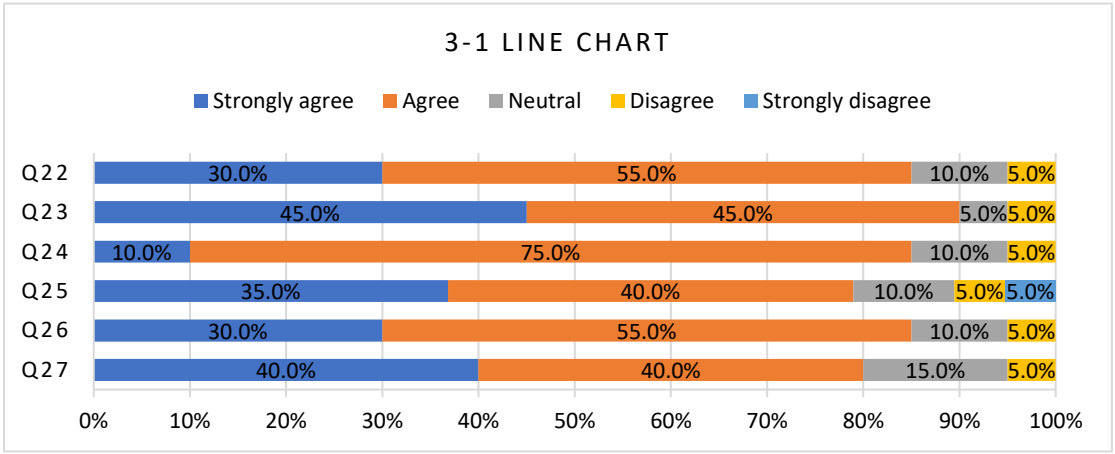

Figure 20. Line chart feedback analysis

The sixth visualisation is the donut chart in Figure 8, an improved visualisation of the line chart in Figure 7. The questions are as follows:

Q28 : Do you agree the donut chart makes it easy to compare differences or highlight numerical correlations between weight and BMI?

Q29 : Do you agree the donut chart makes it easy to observe the trend of steps or calories?

Q30 : Do you agree it is better to show average values rather than daily values to reduce data detail and reduce personal data privacy?

Q31 : Do you agree it is better to show percentage values rather than actual values to reduce data detail and reduce personal data privacy?

Q32 : Do you agree it is more difficult to predict trends or differences using a pie chart than the line chart in Figure 3-1?

Figure 21 is an analysis of user feedback which pointed out that 75% of the participants thought there were no differences or highlight numerical correlations in the donut chart, figure 8. 85% of the participants did not think the visualisation was predictable enough to observe trends in the data. 85% of participants thought the data could be reduced in detail by displaying it in both average and percentage values. 95% of the participants thought it was more difficult to predict the trend in data in the donut than in the line chart.

In conclusion, the donut chart in Figure 8 does not highlight data differences or have predictability. Therefore, it is much more effective than the line chart in Figure 7 in preventing visual identification of personal information and even protecting privacy.

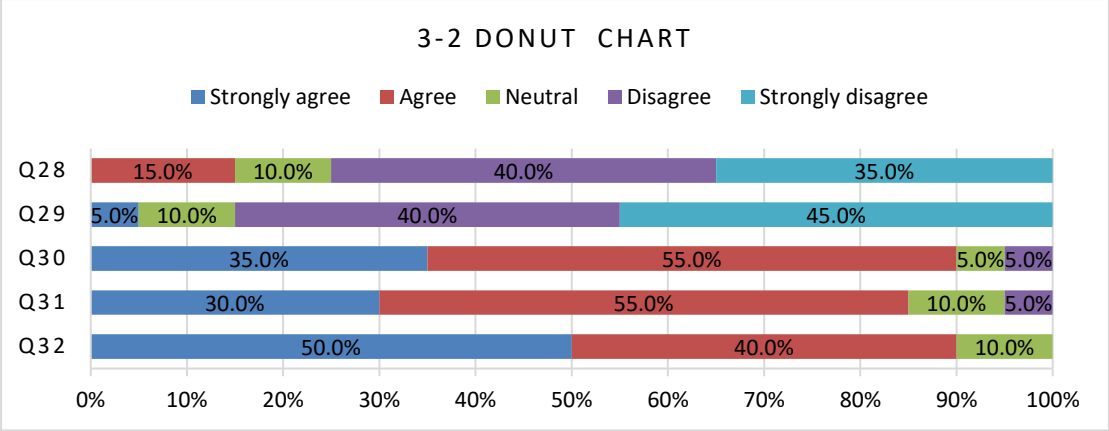

Figure 21. Donut chart feedback analysis

To summarise the analysis in the first part of the questionnaire, it is possible to reduce data trending and predictability and highlight data variability by different visualisations such as a violin diagram, donut chart and so on. Plus, the values displayed as percentages or averages can also prevent visualisations from identifying personal information and even help in protecting privacy.

The second part of the questionnaire consisted of five questions, focused on the effectiveness of data sharing restrictions. There were four options for restricting data sharing methods, namely restricting the number of views, the number of days of sharing, age, and gender.

According to the results of the questionnaire, as shown in Figure 22, 55% of participants considered the most effective method of restricting data sharing is to limit the number of days of sharing. In comparison, 50% of participants considered the less effective way of restricting data sharing to be limiting age, as shown in Figure 23.

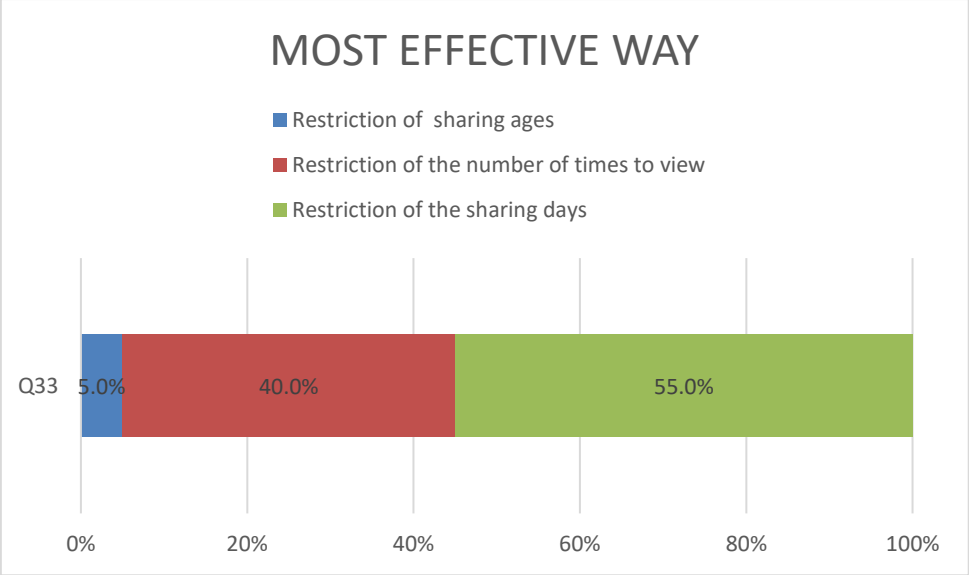

Figure 22. The most effective method of restricting data sharing feedback analysis

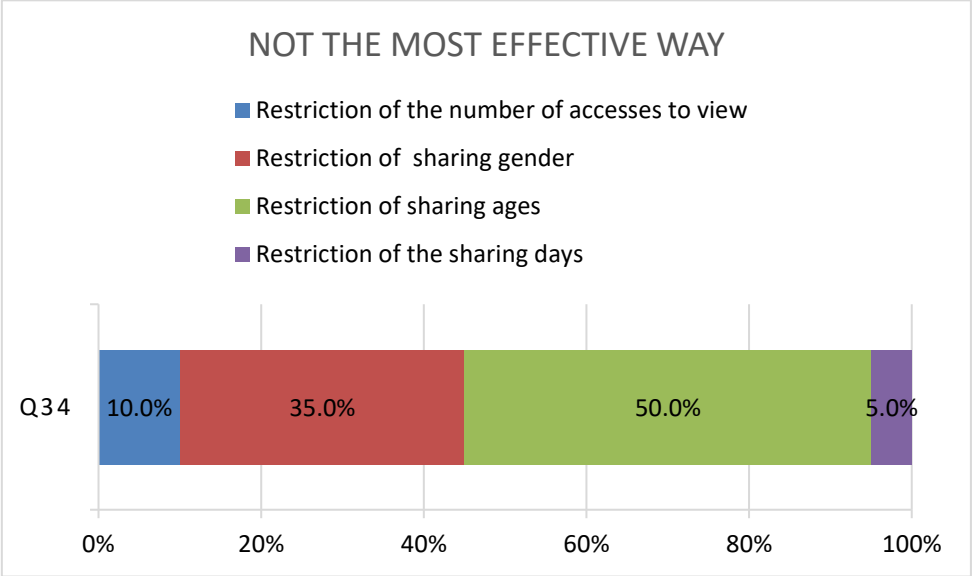

Figure 23. The least effective method of restricting data sharing feedback analysis

Then we got the data sharing restriction values from the database and analysed them to determine the number of sharing restriction days. Most people preferred to share in a day, as shown in Figure 24 (the label value represents the number of days).

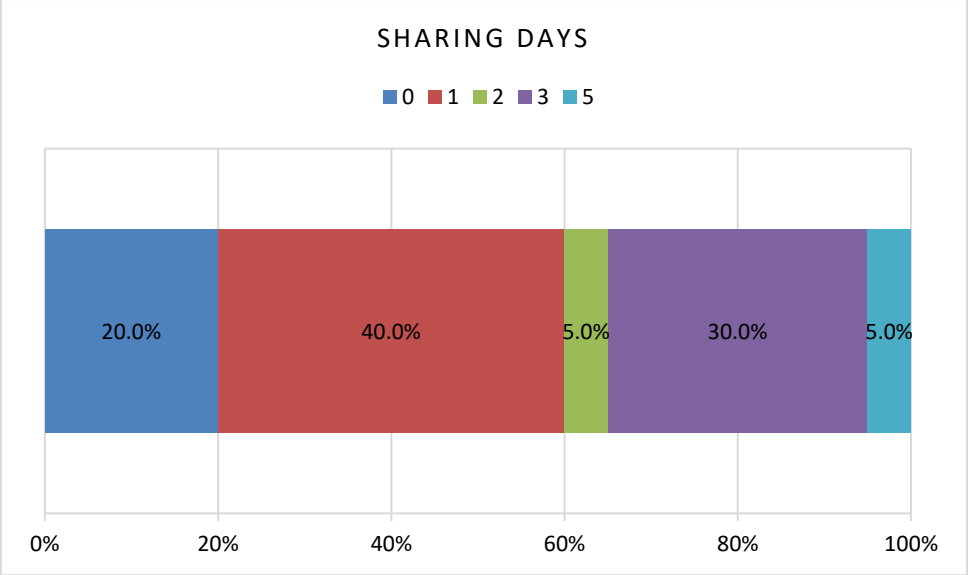

Figure 24. Restriction of the sharing days value feedback analysis

Most people chose not to restrict the age of sharing, as shown in Figure 25.

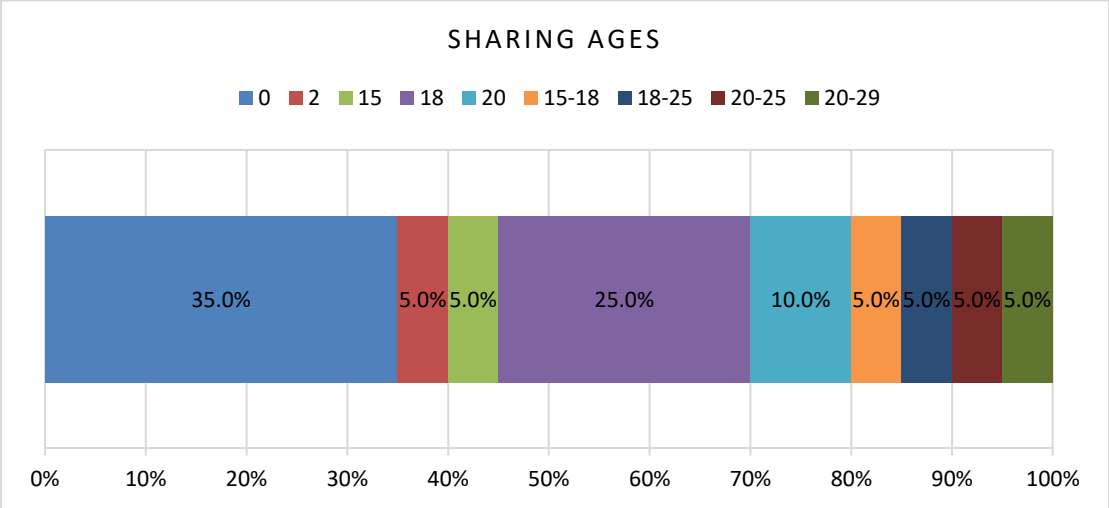

Figure 25. Restriction of the sharing ages value feedback analysis

As for limiting the number of days the visualisation would be shared, 75% of participants felt that more options should be provided for the number of days to be shared, such as hours, minutes, seconds, etc., as shown in Figure 26.

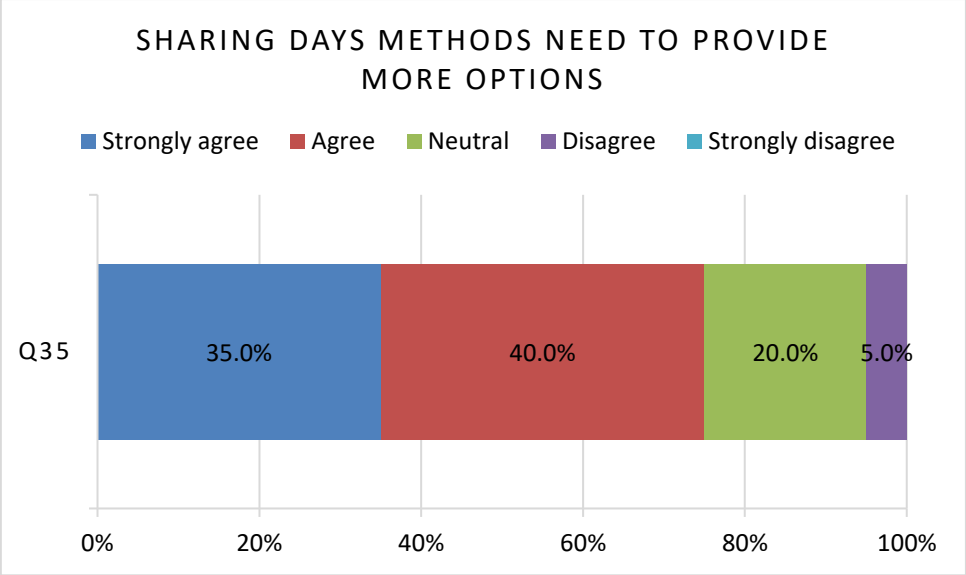

Figure 26. Restriction of the sharing days need more options feedback analysis

One suggestion received for a new method to restrict data sharing was to add an agency option.

The questionnaire feedback will be analysed in this section. It explores the participants' perceptions of different visualisations from the participants' feedback and the results of the questionnaire analysis.

When data is visualised in bar charts or line charts, it is easier to observe trends or compare differences between data; in contrast, most participants felt that pie, violin, and donut charts made it more difficult to see trends and also reduced data variability compared to bar charts or line charts. From the previous literature, it was mentioned that when visualisation is predictable or when data can highlight differences, it is easy to identify personal information, so from the results of the survey, improved visualisation of data not only reduces data trends and thus avoids data predictability, but also reduces differences between data, making it less easy to compare data differences. Most participants felt that using different units for the Y-axis reduced readability and meant that it was not easy to identify more personal information through the coordinates. In addition, most participants would prefer for personal information to appear by displaying detailed values and data details when the mouse hovers over the graph rather than presenting the values on the chart. Moreover, through displaying a single piece of data or combining the visualisation of different data, most participants felt they could explore more details. In other words, personal information could be easily identified.

Finally, by displaying the values as averages or percentages, most participants felt that reducing the detail of the data was sufficient to reduce the risk of identifying personal information. Improved data visualisation removes data trends and variability, increasing privacy protection and reducing the risk of identifying personal information. In addition, visualisation enhances the privacy of personal information by interactively displaying data details and usage averages or percentages.

In the second part of the questionnaire, most participants felt that the most effective way to limit data sharing was to limit the length of data sharing. In other words, when sharing data visualisations, participants prefer to reduce the risk of personal information being leaked to others by limiting the number of sharing days. According to the data on the database, most participants prefer to share for one day only. The results of the survey showed that participants would like the sharing expiry date to be shown when sharing personal information data visualisations. This corresponds to Barker et al's (2009) indication that the retention period must be clearly stated for data privacy purposes. In addition, the unit for limiting the number of days shared should provide more options, such as hours, minutes, and seconds. A participant suggested that data sharing restriction settings should include a new option for agency. This suggestion could be included in future surveys to analyse and evaluate its effectiveness.

In summary, this section discussed the implementation results of the project and the analysis of the questionnaires, as well as the user story scenarios and functional tests. All aspects met the objectives set and improved data visualisations to decrease the chances of people identifying personal information from them.

5. Conclusion

In this project, I succeeded in building a web platform for evaluating data visualisation and improving the problem of identifying personal privacy through visualisation. This platform enabled data collected from devices to be imported and visualised. Furthermore, through testing different data visualisations and collecting feedback from the participants, it was possible to analyse it and figure out how to improve the visualisation and how to present it appropriately. In addition, the restriction of data sharing methods was used to explore ways to reduce the risk of personal information being identified. Different restriction options can be tested in the future to find the most effective way to restrict data sharing.

After collecting feedback from participants, it was confirmed that it is possible to identify personal information by trends, highlight differences, and also, as Bhattacharjee et al. (2020) points out, that personally identifiable information may be identified by cross-referencing different data when people share their health data visualisation. This result matches the research from the background section, which indicated that data visualisation makes personal information identifiable as a result of the following factors: predictability, salient differences, and cross-referencing different data. Moreover, from the feedback, I found if the data visualisation removes data trends and data variability and displays data details in the form of averages or percentages interactively, it reduces the risk of personal information being identified. Moreover, the project also found that people prefer to share personal information for only one day, which could be used in the future to design fixed menus for sharing, such as one day, three days and five days. This could then be and applied to other developments or products.

In the end, the problem of identifying personal information through data visualisation was successfully improved, and the project objectives were achieved. During the development of the platform, I solved problems in writing code and completed the functional and non-functional requirements as planned for the project. Although the platform has solved the problem I wanted to solve, it can still develop other functions like creating the most appropriate data visualisation for the specific group, like age or gender.

6. Reflection

During the development of this project, I have not only learnt about the use of different tools and compilation skills, but I have also improved my time management and planning skills. In this section, I will reflect on these experiences and explore the areas that need further improvement.

- Background research and project management skills

Before implementing this project, I conducted extensive research into the problem it was intended to solve. In the process, I learnt how to use the Cardiff University Library, the ACM Digital Library, and many academic websites to find academic papers and journals on relevant key terms. This knowledge enhanced my academic research skills. In addition, through the PDCA concept, I have learnt how to plan and manage an entire development project and to improve the integrity of the project to reduce the occurrence of errors and shorten the development time. Moreover, setting the timeline for each section of the project allowed me to keep track of the project's progress and avoid exceeding the completion deadline. Undoubtedly, these abilities and skills will be helpful for future projects.

- Learning from theories and the importance of applying them to the project.

During the software engineering course, I learnt how to develop a complete project system from requirements, user stories, and test cases. This experience was beneficial when developing this project. This experience also enabled me to identify potential problems or any potential issues before development began, which not only saved development time but also reduced the number of unexpected errors.

- Learning and selecting the appropriate technology stack

Developing this project allowed me to understand how to design the front-end framework, handle the back-end database, build the entire web platform system, and learn more about deployment and how to deploy the project. In addition, through searching different data visualisation libraries and technical articles and repeatedly testing and reviewing official documentation, I was able to find the most appropriate technology for data visualisation on this platform. Finally, user

stories and test cases were developed to cover all possible user scenarios and to ensure that users operated without problems.

When I designed the questionnaire, I had no previous experience with similar questionnaires, so I observed and compared different questionnaire formats and searched for relevant articles. Finally, I used a Likert scale, commonly used in questionnaires, to create a questionnaire to observe the participants' feedback. Working on this project has allowed me to reflect on my skills, like programming skills and understanding new techniques that will help me comprehend new technologies or work in software development.

In conclusion, this project has exposed me to problems I had not thought of before, and I have learned new techniques and skills. Although I was stuck with programming and technical issues during the process, I eventually solved them through different solutions after discussing with friends and classmates. This made me realise that this is an area where I need to improve, so instead of thinking about solving problems independently, it is better to discuss them with others to generate new ideas. Finally, I have successfully achieved the project objectives through project development and gained different knowledge and experience from the course.

7. References

Aritra, D. Eamonn, M. Alfie, A. and M, C. 2014. Opportunities and Challenges for Privacy-Preserving Visualisation of Electronic Health Record Data. doi: 10.13140/2.1.1291.2642

Aritra, D. Min, C. and Robert, K. 2012. Conceptualizing Visual Uncertainty in Parallel Coordinates. Comput. Graph. Forum 31, 3pt2 (June 2012), 1015–1024. https://doi.org/10.1111/j.1467-8659.2012.03094.x

Barker, K. Askari, M. Banerjee, M. Ghazinour, K. Mackas, B. Majedi, M. Pun, S. and Williams, A. 2009. A Data Privacy Taxonomy. 26 British National Conference on Databases, BNCODth. 5588. 42-54. 10.1007/978-3-642-02843-4_7.

Bertin, J. Berg, W. and Wainer, H. 1983. Semiology of graphics: diagrams, networks, maps. Vol. 1. University of Wisconsin press Madison.

Bhattacharjee, K., Chen, M. and Dasgupta, A., 2020. Privacy‐Preserving Data Visualisation: Reflections on the State of the Art and Research Opportunities. Computer Graphics Forum, 39(3), pp.675-692.

Börner, K. Adam, M. Balliet, R. and Joe, H. 2016. Investigating aspects of data visualisation literacy using 20 information visualisations and 273 science museum visitors. Information Visualisation 15, 3 (2016), 198–213.

Claus, W. 2019. Fundamentals of data visualisation : a primer on making informative and compelling figures. First edition. Sebastopol, CA: O'Reilly Media.

Cristian, B. Lorrie, C. Julie, D. Saranga, K, and Manya S. 2011. Improving computer security dialogs. In Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - Volume Part IV (INTERACT'11). Springer-Verlag, Berlin, Heidelberg, 18–35.

Fox, S, and Duggan, M. 2013. Health online 2013 (p. 55). Washington, D.C.:

Pew Research Center's Internet & American Life Project. Retrieved from http://www.pewinternet.org/files/oldmedia//Files/Reports/PIP_HealthOnline.pdf (Archived by WebCite® at http://www.webcitation.org/6Wsi5u8S9).

Harbach, M. Fahl, S. Yakovleva, P. and Smith, M. 2013. Sorry, I Don't Get It: An Analysis of Warning Message Texts. In Proc USEC.

Horr, T. and Joe, H. 2017. "Data Visualisation Literacy Literature Review: Sharing Personal Information.".

Ian, L. Anind, D. and Jodi, F. 2011. Understanding my data, myself: Supporting self-reflection with ubicomp technologies. Proceedings of the 13th international conference on Ubiquitous computing. 405-414. 10.1145/2030112.2030166.

Jennifer, H. Tim, A. Rok, S. Peter, K. Bojan, B. Abby, K. Jure, L. and Scott, D. 2019. Best practices for analyzing large-scale health data from wearables and smartphone apps. npj Digital Medicine. 2. 10.1038/s41746-019-0121-1.

Jeremy, B. Ronald, R. Enrico, B. and Jean, F. 2014. A principled way of assessing visualisation literacy. IEEE Transactions on Visualisation and Computer Graphics 20, 12 (2014), 1963–1972.

Jesus, C. and David, G. 2015. Visual analytics in healthcare--opportunities and research challenges. Journal of the American Medical Informatics Association : JAMIA. 22. 260-262. 10.1093/jamia/ocv006.

Kennedy, H., & Allen, W. (2016). Data visualisation as an emerging tool for online research. The Sage handbook of online research methods, 307-326.
Knight, J. and Allen, S. 2012. Applying the PDCA Cycle to the Complex Task of Teaching and Assessing Public Relations Writing. International Journal of Higher Education. 1. 10.5430/ijhe.v1n2p67.

Morey, T. Forbath, T. and Schoop, A. 2015. Customer Data: Designing for Transparency and Trust. Harvard Business Review, (May), 1–17. Retrieved from https://hbr.org/2015/05/customer-data-designing-for-transparency-and-trust.

Motwani, R. and Xu, Y. 2007. Efficient Algorithms for Masking and Finding

Quasi-Identifiers. In Proceedings of the Conference on Very Large Data Bases (VLDB) , pp. 83–93. 2

Nam, C. Song, C. and Lee, E. 2006. Consumers' Privacy Concerns and Willingness to Provide Marketing-Related Personal Information Online Changi. Advances in Consumer Research, 33, 212–217.

Plaisant, C. Mushlin, R. Snyder, A. Li, J. Heller, D. and Shneiderman, B. 2003. "Lifelines: Using visualisation to enhance navigation and analysis of patient records," The Craft of Information Visualisation, pp. 308–312. Available at: https://doi.org/10.1016/b978-155860915-0/50038-x.

Shneiderman, B. 1996. The eyes have it: A task by data type taxonomy for information visualisations. In Proceedings 1996 IEEE symposium on visual languages, IEEE, pp. 336–343.

Sweeney, L. 2000. Simple Demographics Often Identify People Uniquely. Health. 671.

Tory, M. and Möller, T. 2004. Human factors in visualisation research. IEEE Trans Vis Comput Grap. IEEE transactions on visualisation and computer graphics. 10. 72-84. 10.1109/TVCG.2004.1260759.

Williams, J. 2010. Social Networking Applications in Health Care: Threats to the Privacy and Security of Health Information. In Proceedings of the 2010 ICSE Workshop on Software Engineering in Health Care (pp. 39–49). New York, NY, USA: ACM. http://doi.org/10.1145/1809085.1809091

8. Appendices

APPENDIX A: USER STORIES

Figure 1: Navigating to the Home Page

Name of Use Case:	Navigating to the Home Page
Description:	The user would like to go the Home Page.
Pre-conditions:	The user can see the navigation bar show Home.
Scenarios:	**SCENARIO 1:** The user is not on the Home page and clicks Home on the navigation bar. Once the user clicks, the Home page will load.
Additional Criteria:	
Homepage, the user should see the following components: • Description of the introduction and instruction on this page. • Page title • Footer	

Figure 2: Navigating to the HealthData Page

Name of Use Case:	Navigating to the HealthData Page
Description:	The user would like to go the HealthData Page.

Pre-conditions:	The user can see the navigation bar show HealthData.
Scenarios:	**SCENARIO 1:** The user is on the Home page and clicks HealthData on the navigation bar. Once the user clicks on HealthData, the HealthData page will load and show data visualisation from the database.
Additional Criteria:	
On the ShareData page, the user should see the following: • Description of the introduction on this page. • Data visualisation • Page title • Footer	

Figure 3: Navigating to the ShareData Page

Name of Use Case:	Navigating to the ShareData Page
Description:	The user would like to go the ShareData Page.
Pre-conditions:	The user is logged in account. Then the user can see the navigation bar show HealthData.
Scenarios:	**SCENARIO 1:** The user clicks ShareData on the navigation bar and submits the share data restrictions form. Once the user clicks, the ShareData page will load, and the user can configure the option to share data restrictions, like the restriction of the sharing days, and then click submit button. After that, it will be uploaded to the database.

	SCENARIO 2: The user clicks ShareData on the navigation bar and fills the share data restrictions form with an incorrect value. The ShareData page will show error messages to notify the user to fill the value in the correct format.
Additional Criteria:	
On the ShareData page, the user should see the following: • Description of the introduction on this page. • Data visualisation of daily average consumes calories. • Page title • Footer	

Figure 4: Navigating to the Login Page

Name of Use Case:	Navigating to the Login Page
Description:	The user would like to go to the Login Page and then log in account.
Pre-conditions:	The user is not logged in account, and the user can see the navigation bar showing Login.
Scenarios:	**SCENARIO 1:** The user clicks on Login on the navigation bar and then logs in to an account. Once the user clicks on the button of Login, the Login page will load, then fill in the email address and password. If the user log in successfully, their username and the logout button will be shown on the navigation.

	SCENARIO 2: The user logs in unsuccessfully. The Login page will navigate to an error page and show the error message to the user. **SCENARIO 3:** The user logs out account. When the user logs in account and wants to log out, then clicks log out on the navigation bar, it will navigate to the Home page and show log out successfully. Moreover, the navigation bar will show Hello, Guest.

Figure 5: Navigating to the Register Page

Name of Use Case:	Navigating to the Register Page
Description:	The user would like to go to the Register Page and then register an account.
Pre-conditions:	The user is not logged in account, and the user can see the navigation bar show Register.
Scenarios:	**SCENARIO 1:** The user clicks Register on the navigation bar and then registers an account. Once the user clicks on the button of Register, the Register page will load, then fill in the username, email address, password, repeat password, gender and age. If the user registers successfully, it will log in and return to the home page automatically, as well as show the username on the navigation bar.

	SCENARIO 2: The user registers unsuccessfully. The Register page will show error messages to notify the user to fill the value in the correct format.

APPENDIX B: TEST CASES

Figure 1: Navigating to the Home Page Text Case

Test Case Id: navigatingToHomePage	**Test Purpose:** To test that that user can go to the Home Page		
Preconditions: The user can see the navigation bar show Home.			
Test Case Steps:			
Step No	Procedure	Expected Response	Pass/Fail
1	Click the Home on the navigation bar.	The user should be taken to the Home Page.	P
Additional Acceptance Criteria:			
• Introduction and instructions for the survey should be displayed • Page title should be displayed • Footer should be displayed			

Figure 2: Navigating to the HealthData Page

Test Case Id: navigatingToHealthDataPage	**Test Purpose:** To test that that user can go to the HealthData Page		
Preconditions: The user can see the navigation bar show HealthData.			
Test Case Steps:			
Step No	Procedure	Expected Response	Pass/Fail

1	Click the HealthData on the navigation bar.	The user should be taken to the HealthData Page and see the data visualisation.	P
2	The user interacts with the data visualisation.	Each data visualisation has a drop-down menu on the left side to switch between different data. Shows the details when the mouse hovers over the data visualisation.	P
Additional Acceptance Criteria:			
• Description of the introduction on this page should be displayed • Page title should be displayed • Footer should be displayed			

Figure 3: Navigating to the ShareData Page

Test Case Id: navigatingToShareDataPage	**Test Purpose:** To test that that user can go to the ShareData Page		
Preconditions: The user can see the navigation bar show ShareData.			
Test Case Steps:			
Step No	Procedure	Expected Response	Pass/Fail
1	Click the ShareData on the navigation bar.	The user should be taken to the ShareData page and user can configure the option to share data restrictions and then submit.	P

2	The user fills the share data restrictions form with an incorrect value.	The ShareData page will show error messages and show error details in the error field.	P
Additional Acceptance Criteria:			
• Description of the introduction on this page should be displayed • Page title should be displayed • Footer should be displayed			

Figure 4: Navigating to the Login Page

Test Case Id: navigatingToLoginPage	**Test Purpose:** To test that that user can go to the Login Page and log in account.		
Preconditions: The user is not logged in account.			
Test Case Steps:			
Step No	Procedure	Expected Response	Pass/Fail
1	Click the Login on the navigation bar then fill in the email address and password.	Show the username and Logout on the navigation bar and navigate to the home page but also display log in successfully message.	P
2	The user logs in unsuccessfully.	The Login page will navigate to an error page and show error message.	P
3	When the user log in account then log out account.	It will navigate to the Home page and show log out successfully. Moreover, the navigation bar will	P

		show Hello, Guest.	

Figure 5: Navigating to the Register Page

Test Case Id: navigatingToRegisterPage	**Test Purpose:** To test that that user can go to the Register page and register an account.		
Preconditions: The user is not logged in account.			
Test Case Steps:			
Step No	Procedure	Expected Response	Pass/Fail
1	Click the Register on the navigation bar then fill in the username, email address, password, repeat password, gender and age.	It will log in automatically, show the username on the navigation bar, and navigate to the home page.	P
2	The user register unsuccessfully.	The Register page will show error messages and show error details in the error field.	P

APPENDIX C: QUESTIONNAIRE FEEDBACK

- Questionnaire feedback

1. Please look at Figure 1-1 on the HealthData page. Do you agree that the bar chart makes it easy to compare differences or highlight numerical correlations between steps and calories, e.g. the number of steps is proportional to calories consumed?

more details

Strongly agree	10
Agree	8
Neutral	0
Disagree	2
Strongly disagree	0

2. Please look at Figure 1-1 on the HealthData page. Do you agree the bar chart makes it easy to observe the trend of steps or calories?

more details In-depth analysis

Strongly agree	9
Agree	6
Neutral	1
Disagree	4
Strongly disagree	0

3. Please look at Figure 1-1 on the HealthData page. Do you agree the bar chart displaying different unit values on the Y-axis reduces readability?

more details

Strongly agree	8
Agree	6
Neutral	5
Disagree	1
Strongly disagree	0

4. Please look at Figure 1-1 on the HealthData page. Do you agree the mouse hovering over the chart to show data of values is better than marking the values on the bar chart, which increase the privacy of personal information?

more details

Strongly agree	9
Agree	6
Neutral	3
Disagree	1
Strongly disagree	1

5. Please look at Figure 1-1 on the HealthData page. Do you agree the chart can explore more details and make a comparison by showing steps and calories separately or together?

more details

Strongly agree	9
Agree	9
Neutral	1
Disagree	0
Strongly disagree	0

6. Please look at Figure 1-1 on the HealthData page. Do you agree it is better to show average values rather than daily values to reduce data detail and reduce personal data privacy?

more details

Strongly agree	8
Agree	7
Neutral	2
Disagree	3
Strongly disagree	0

7. Please look at Figure 1-2 on the HealthData page. Do you agree the pie chart makes it easy to compare differences or highlight numerical correlations between steps and calories, e.g. the number of steps is proportional to calories consumed?

more details In-depth analysis

Strongly agree	1
Agree	3
Neutral	2
Disagree	8
Strongly disagree	6

8. Please look at Figure 1-2 on the HealthData page. Do you agree the pie chart makes it easy to observe the trend of steps or calories?

more details

Strongly agree	2
Agree	2
Neutral	2
Disagree	7
Strongly disagree	7

9. Please look at Figure 1-2 on the HealthData page. Do you agree the mouse hovering over the chart to show data of values is better than marking the values on the pie chart, which increase the privacy of personal information?

more details

Strongly agree	9
Agree	5
Neutral	2
Disagree	2
Strongly disagree	2

10. Please look at Figure 1-2 on the HealthData page. Do you agree it is better to show average values rather than daily values to reduce data detail and reduce personal data privacy?

more details In-depth analysis

Strongly agree	9
Agree	8
Neutral	2
Disagree	1
Strongly disagree	0

11. Please look at Figure 1-2 on the HealthData page. Do you agree it is better to show percentage values rather than actual values to reduce data detail and reduce personal data privacy?

more details

Strongly agree	8
Agree	8
Neutral	3
Disagree	1
Strongly disagree	0

12. Please look at Figure 1-2 on the HealthData page. Do you agree it is more difficult to predict trends or differences using a pie chart than the bar chart in Figure 1-1?

more details

Strongly agree	7
Agree	11
Neutral	2
Disagree	0
Strongly disagree	0

13. Please look at Figure 2-1 on the HealthData page. Do you agree with the bar chart combined with the line chart makes it easy to compare differences or highlight numerical correlations between sleep time and in the bed-time?

more details

Strongly agree	9
Agree	9
Neutral	1
Disagree	0
Strongly disagree	0

14. Please look at Figure 2-1 on the HealthData page. Do you agree the bar chart combined with the line chart makes it easy to observe the trend of sleep or in the bed-time?

more details

Strongly agree	11
Agree	8
Neutral	1
Disagree	0
Strongly disagree	0

15. Please look at Figure 2-1 on the HealthData page. Do you agree the mouse hovering over the chart to show data of values is better than marking the values on the chart, which increase the privacy of personal information?

more details In-depth analysis

Strongly agree	8
Agree	5
Neutral	5
Disagree	1
Strongly disagree	1

16. Please look at Figure 2-1 on the HealthData page. Do you agree the chart can explore more details and make a comparison by showing sleep and in the bed- time separately or together?

more details

Strongly agree	7
Agree	10
Neutral	3
Disagree	0
Strongly disagree	0

17. Please look at Figure 2-1 on the HealthData page. Do you agree it is better to show average values rather than daily values to reduce data detail and reduce personal data privacy?

more details In-depth analysis

Strongly agree	9
Agree	6
Neutral	3
Disagree	2
Strongly disagree	0

18. Please look at Figure 2-2 on the HealthData page. Do you agree the violin chart makes it easy to compare differences or highlight numerical correlations between sleep time and in bed- time?

more details In-depth analysis

Strongly agree	2
Agree	4
Neutral	1
Disagree	7
Strongly disagree	6

19. Please look at Figure 2-2 on the HealthData page. Do you agree the violin chart makes it easy to observe the trend of sleep or in the bed-time?

more details In-depth analysis

Strongly agree	3
Agree	2
Neutral	2
Disagree	6
Strongly disagree	7

20. Please look at Figure 2-2 on the HealthData page. Do you agree the chart can explore more details and make a comparison by showing sleep and in the bed- time separately or together?

more details In-depth analysis

Strongly agree	6
Agree	9
Neutral	2
Disagree	3
Strongly disagree	0

21. Please look at Figure 2-2 on the HealthData page. Do you agree it is more difficult to predict trends or differences using a violin chart than the bar chart combined with a line chart in Figure 2-1?

more details

Strongly agree	7
Agree	10
Neutral	2
Disagree	0
Strongly disagree	1

22. Please look at Figure 3-1 on the HealthData page. Do you agree the line chart makes it easy to compare differences or highlight numerical correlations between weight and BMI?

more details In-depth analysis

Strongly agree	6
Agree	11
Neutral	2
Disagree	1
Strongly disagree	0

23. Please look at Figure 3-1 on the HealthData page. Do you agree the line chart makes it easy to observe the trend of weight and BMI?

more details In-depth analysis

Strongly agree	9
Agree	9
Neutral	1
Disagree	1
Strongly disagree	0

24. Please look at Figure 3-1 on the HealthData page. Do you agree the line chart displaying different unit values on the Y-axis reduces readability?

more details In-depth analysis

Strongly agree	2
Agree	15
Neutral	2
Disagree	1
Strongly disagree	0

25. Please look at Figure 3-1 on the HealthData page. Do you agree the mouse hovering over the chart to show data of values is better than marking the values on the line chart, which increase the privacy of personal information?

more details

Strongly agree	7
Agree	8
Neutral	2
Disagree	1
Strongly disagree	1

26. Please look at Figure 3-1 on the HealthData page. Do you agree the chart can explore more details and make a comparison by showing weight and BMI separately or together?

more details In-depth analysis

Strongly agree	6
Agree	11
Neutral	2
Disagree	1
Strongly disagree	0

27. Please look at Figure 3-1 on the HealthData page. Do you agree it is better to show average values rather than daily values to reduce data detail and reduce personal data privacy?

more details

Strongly agree	8
Agree	8
Neutral	3
Disagree	1
Strongly disagree	0

28. Please look at Figure 3-2 on the HealthData page. Do you agree the donut chart makes it easy to compare differences or highlight numerical correlations between weight and BMI?

more details

Strongly agree	0
Agree	3
Neutral	2
Disagree	8
Strongly disagree	7

29. Please look at Figure 3-2 on the HealthData page. Do you agree the donut chart makes it easy to observe the trend of steps or calories?

more details In-depth analysis

Strongly agree	1
Agree	0
Neutral	2
Disagree	8
Strongly disagree	9

30. Please look at Figure 3-2 on the HealthData page. Do you agree it is better to show average values rather than daily values to reduce data detail and reduce personal data privacy?

more details

Strongly agree	7
Agree	11
Neutral	1
Disagree	1
Strongly disagree	0

31. Please look at Figure 3-2 on the HealthData page. Do you agree it is better to show percentage values rather than actual values to reduce data detail and reduce personal data privacy?

more details In-depth analysis

Strongly agree	6
Agree	11
Neutral	2
Disagree	1
Strongly disagree	0

32. Please look at Figure 3-2 on the HealthData page. Do you agree it is more difficult to predict trends or differences using a pie chart than the line chart in Figure 3-1?

more details

Strongly agree	10
Agree	8
Neutral	2
Disagree	0
Strongly disagree	0

33. Which statement below is the most effective way to restrict sharing data visualisation options? For example, if the restriction of the sharing days is 5, which means the data visualisation you shared will be unavailable in five days. After that, this data visualisation will stop being accessible to others.

more details In-depth analysis

Restriction of the number of tim...	8
Restriction of the sharing days	11
Restriction of sharing ages	1
Restriction of sharing gender	0

34. Which statement below is NOT the most effective way to restrict sharing data visualisation options? For example, if the restriction of the sharing days is 5, which means the data visualisation you shared will be unavailable in five days. After that, this data visualisation will stop being accessible to others.

more details

Restriction of the number of acc...	0
Restriction of the sharing days	1
Restriction of sharing ages	10
Restriction of sharing gender	7
other	2

35. Do you agree restrictions of the sharing days methods need to provide more options about units, e.g. hour, minute, second and so on?

more details

Strongly agree	7
Agree	8
Neutral	4
Disagree	1
Strongly disagree	0

36. Do you have any idea about adding a new restriction method on data sharing?

more details

Yes	1
No	19
other	0

37. If your previous answer is 'Yes', please written down your idea.

more details

1

to respond

latest response

37 . If your previous answer is 'Yes', please written down your idea.

1 to respond

ID ↑	name	to respond	language
1	anonymous	agency	English (United Kingdom)

- From the database to get participants to fill in the Restricted Data Sharing Method value.

	id	authorName	numberAccess	shareTime	shareAge	genderShare
Edit Copy Delete	2	anonymous	5	3	20	male
Edit Copy Delete	3	anonymous	3	1	18	non restriction of sharing gender
Edit Copy Delete	4	anonymous	10	5	15	non restriction of sharing gender
Edit Copy Delete	5	anonymous	5	3	15-18	non restriction of sharing gender
Edit Copy Delete	6	anonymous	0	0	20-29	non restriction of sharing gender
Edit Copy Delete	7	anonymous	5	3	18	non restriction of sharing gender
Edit Copy Delete	8	anonymous	1	1	20	female
Edit Copy Delete	9	anonymous	2	1	18	non restriction of sharing gender
Edit Copy Delete	10	anonymous	3	3	20-25	non restriction of sharing gender
Edit Copy Delete	11	anonymous	1	1	18-25	non restriction of sharing gender
Edit Copy Delete	12	anonymous	3	1	18	non restriction of sharing gender
Edit Copy Delete	13	anonymous	0	0	0	non restriction of sharing gender
Edit Copy Delete	14	anonymous	0	0	0	non restriction of sharing gender
Edit Copy Delete	15	anonymous	3	2	0	non restriction of sharing gender
Edit Copy Delete	16	anonymous	0	1	0	non restriction of sharing gender
Edit Copy Delete	17	anonymous	2	1	18	non restriction of sharing gender
Edit Copy Delete	18	anonymous	0	3	0	male
Edit Copy Delete	19	anonymous	5	1	2	male
Edit Copy Delete	20	anonymous	0	0	0	non restriction of sharing gender
Edit Copy Delete	21	anonymous	0	3	0	non restriction of sharing gender