



The Investigation Of Users' Perception Of Computer Viruses
And Dangers Of Public Wi-Fi Particularly, For Social Media
Use And Whether Their Obsession Of Platforms Causes
Ignorance To Potential Malicious Scams And Links.

Dissertation submitted for the Degree of Bachelor Of Arts With
Honours (Combined) - Information Technology And Media &
Communication

Ellen Gault

17001172

May 2020

Word count: 10,802

Research Adviser:

Neil Buckley

DECLARATION OF ORIGINALITY

**I declare that this is an original study based on my own
work and that I have not submitted it for any other course
or degree.**

Acknowledgement

I would like to give my most sincere thanks to my friends and family who consistently supported me throughout my time at university, especially when researching and conducting this study. I would like to express my deepest gratitude to Neil Buckley and Anthony Ridge Newman for going above and beyond to guide me to complete this study and for their support during my time at university. Furthermore, I would like to thank the participants who took part in making this possible, and for that, I am extremely grateful.

Abstract

Objective

This research study aimed to critically examine the impact and knowledge of computer viruses on end-user devices, whilst exploring their knowledge on public Wi-Fi and their use of it. Examining if users have an obsession of social media and if their obsession caused ignorance in detecting malicious features and scams. Significant results were discovered during the experiment. Their knowledge was measured through the survey and their ability to detect maliciousness of links was carried out using the quiz. The hypothesis stated that there would be low-level knowledge among all participants, with no awareness of public Wi-Fi dangers and social media scams.

Research method

Thirty-five males, forty-one females and one 'other' aged over 18 all participated. The use of an online survey was used to establish the level of knowledge participants had among all variables whilst the quiz recognised participants' ability to differentiate malicious features among links and emails. Age, gender and occupation were all explored to determine if they had an impact.

Results

The main findings within the study found that all participants who did not have an occupation within computing had a ranking of 1 or 2; no knowledge or low-level knowledge among all variables investigated. There was a significant rise in knowledge among the seven participants who had higher knowledge with rankings of 4 and 5. It was discovered that social media obsession does cause users to become ignorant of the malicious features and public Wi-Fi is used for convenience over consequences.

Conclusion

The findings suggest that there has been no education to participants on computer viruses and public Wi-Fi dangers. Nor have they been acknowledged on the ways to spot a malicious link or email. However, there was no apparent impact of age on the results, although it was expected that younger participants would have more knowledge and it was only males who had the occupations within computing. This research provides insight into the users' perception of said variables and obsession of social media.

List of figures

Figure 1: Creating the database	10
Figure 2: Creating the quiz	11
Figure 3: Quiz completed message	12
Figure 4: Future attack encounter.	17
Figure 5: The platforms and apps users believe viruses get onto the device.	19
Figure 6: Virus infections encounter.....	19
Figure 7: Virus attack causes.....	20
Figure 8: Anti-virus software counter	20
Figure 9: The termination methods undergone.....	21
Figure 10: Emotions on virus attacks.....	22
Figure 11: Security features.....	22
Figure 12 Common responses from quiz question 2.....	23
Figure 13: Common answers amongst question 6 in the quiz.....	23
Figure 14: Those who use public wifi.....	28
Figure 15: The use of public wifi.....	28

List of tables

Table 1: Occupation among participants.....	15
Table 2: Displays the most common words used by respondents to describe a virus.	15
Table 3: Those who knew the risks to viruses.	16
Table 4: How individuals have been educated.	18
Table 5: Comparison of wrong and right answers from quiz question 1.	24
Table 6: Comparison of wrong and right answers from quiz question 3.	24
Table 7: Comparison of wrong and right answers from quiz question 4	26
Table 8: Comparison of wrong and right answers from quiz question 8.	26
Table 9: Respondents who said yes to being educated on public Wi-Fi dangers.....	27
Table 10: Comparison results of this study and two previous ones.....	29
Table 11: Comparison of results on social media use.....	30

Abbreviations

Wi-Fi: Wireless Fidelity

URL: Uniform Resource Locator

HTTP: Hypertext Transfer Protocol

HTTPS: Hypertext Transfer Protocol Secure

ONS: Office of National Statistics

CSO: Computer Security Office

ITRC: Identity Theft Resource Centre

Contents

1. Introduction	1
1.1 Background	1
1.2 Motivation	2
1.2 Aims, Objectives, and Hypothesis.....	2
1.3 Project Structure	3
2. Literature review	3
2.1 Knowledge of viruses	4
2.2 Open Wi-Fi networks	5
2.3 Social media	6
2.4 Obsession of social media.....	8
3. Methodology.....	9
3.1 Participants.....	9
3.2 Design	9
3.3 Materials	10
3.4 Ethics	12
3.5 Procedure	13
3.6 Analysis	13
4. Results	14
4.1 Demographic data.....	14
4.2 Knowledge of viruses	15
4.3 Experience of viruses.....	19
4.4 Impact on individuals' life.	21
4.5 Awareness to scams and malicious links	22
4.6 Knowledge and use of public wifi.....	27
4.7 Social media obsession	29
5. Discussion.....	30
5.1 Summary of findings –	30
5.2 Explanation and interpretation of the findings	31
6. Future research and limitations.....	37
7. Conclusion.....	39
8. References	40
9. Appendix	46
9.1 Appendix 1: Ethics form approval	46
9.2. Appendix 2: Research Information Sheet	52

9.3. Appendix 3: Online Survey	54
9.4. Appendix 4: Online quiz.....	59
9.5 Appendix 5: Quiz coding	63
9.6 Appendix 6: Python coding	69
9.7 Appendix 7: Demogrpahic answers	82
9.8 Appendix 8: Answers from question 2 on quiz.....	84
9.9 Appendix 9: Answers from question 6 on quiz.....	86
9.10 Appendix 10: Answers from question 1 on quiz.....	88
9.11 Appendix 11: Answers from question 3 on quiz.....	90
9.12 Appendix 12: Answers from question 4 in quiz.....	92

1. Introduction

The overall purpose of the research conducted within this project is to identify whether end-users are educated on viruses and public Wi-Fi. Exploring if their social media obsession outweighs their concerns and abilities to acknowledge malicious links, attachments, and scams. This chapter provides background and rationale for such topics. Along with aims, objectives and overall structure.

1.1 Background

The utilization of technology is rapidly growing, traditionally malware regarding computers has been around since their evolution in 1938. Malware is defined as ‘software which is used to attempt to breach a computer system’s security policy concerning confidentiality, integrity, and availability’ (Sharp, 2007). Nachenberg, (1997) further defines viruses as a ‘self-replicating program created to destroy a computer system and steal confidential data by attaching itself to executable files’.

The most common virus is web scripting, giving attackers the power to inject client-side scripting into a webpage, opening a bypass for stolen information (Das, 2019). They are attached to clickable media such as attachments or links. The browser hijacker virus is very similar, however, it modifies settings within the browsers in the absence of the users’ permission and redirects them to a webpage they had not clicked. Both are extremely dangerous and can be injected through social media and public wifi with inadequate warnings.

Wi-Fi has been around since 1991 and was first used in a public setting in 1997 and since the mania began. Public Wifi (open wifi networks) can be found in nearly every public setting; cafes, on transport and entertainment settings. By 2010 there were more than one million public networks (VPN, 2019). Although these interconnected networks have allowed effective and usually free communication for the public, it comes at an unsecured price. With security at a costly price for organisations, they expect end-users to take their precautions. Therefore, speculates whether individuals are aware of the unsecured network they continuously connect to or if consequences outweigh the convenience?

Social media was first created in 1997, with a site ‘Six Degrees’ allowing the creation of profiles and connecting with users. It gave creators and developers a glimpse of success that

the niche technology Wi-Fi has created. Sites such as MySpace and LinkedIn gained prominence in the early 2000s (Hendricks, 2013). With Facebook debuting in 2004, Twitter in 2006 and Instagram in 2010, they have a tallied total of over 3 billion users. It is in question whether these users have precautions when using the digital media platforms and have consciousness of the scams and malware injected throughout.

1.2 Motivation

This research project was motivated by how common it is for malicious attacks, and the horror stories of what it can lead to. There appeared to be a gap in existing research, with the question is there a lack of education on the risks of such an attack? Is that why many friends and family members were unsure if they are victims of such infiltrations?

Malicious software can inflict a range of emotions on a victim, including stress, depression, anxiety and worry as a hacker can obtain personal data. Researching the topic of viruses and social media this paper hopes to introduce the true knowledge users have on malware and public Wi-Fi dangers.

1.2 Aims, Objectives, and Hypothesis

The aims and objectives of this project were devised to gain a thorough understanding of the topic to enable a strict and accurate analysis based on the collected data. The hypothesis gives insight into what the potential outcome may be.

Aims

To critically examine the impact and knowledge of viruses on end-user devices and lives, whilst looking at those passed through public Wi-Fi and social media. To investigate if users can distinguish between malicious and non-malicious links. A thorough examination will determine if a social media obsession will offset the risk of becoming a malware victim.

Objectives

1. To gain a critical understanding of the level of knowledge among users on said variables and if education has been given.
2. To gain a critical understanding of what users believe makes a site/attachment or link malicious.
3. To evaluate the impact of viruses on individuals' lives.
4. To analyse the length of time users spend on social media and whether this impacts their knowledge.

5. To identify the future knowledge of said variables.

Hypothesis

The hypothesis for this research project is that the majority of individuals will not be aware or have much knowledge of what a virus is or how attacks make users the victim of cyberattacks. Nor will they be conscious of the dangers of using public wifi to go on different platforms such as social media due to their obsession. They will also be uninformed of what features make links, emails or attachments to appear 'malicious'.

1.3 Project Structure

For the above to be investigated thoroughly, the use of chapters will be used and will be used under the following.

Literature review

To thoroughly understand the chosen topic, the said chapter will outline any key findings among previous studies and literature of existing work. To help obtain objective one.

Methodology

This chapter outlines the justification of the chosen research method. Including the distribution and measures taken to collate data. Achieving objectives 2 and 3.

Results

Collated data will be presented in this section with the use of graphs and tables. Objectives 3 and 4 will be achieved.

Discussion

This chapter entails the discussion of chapter 3 against previous research and whether the hypothesis was accepted.

Conclusion

The final chapter will include a conclusion based on the previous chapters, with recommendations for future research. Allowing objective 5 to be achieved.

2. Literature review

2.1 Knowledge of viruses

The level of knowledge on viruses is deemed insufficient in corresponding with the number of malicious malware. A survey conducted by ONS, (2018) discovered that over 17 million UK individuals were attacked by viruses in 2018, resulting in over £130 billion being stolen. This figure speculates whether users are aware of the dangers. The ONS calculates the different crimes within the UK, including computer viruses. Measured by investigating the number of reports, a limitation as some viruses are silent and show no signs and therefore are not reported. However, they use bulletin reports and methods of quantitative research to determine if there's a pattern among the viruses. Their findings took them to the conclusion that there should be more information and proactive steps given to individuals about viruses or the effect they can have as stated in the hypothesis.

Reinforced by Ahmad, (2019) he demands that there is inadequate information of viruses and 'basic IT security principles will help home users eliminate the threat of computer viruses'. As does Wood et al, (1987:9) who appeals to emotion and demands 'there is a big gap in current awareness and computer security is a corporate need'. Both of these authors formulated the problem and had interpretive research orientation due to the experience individuals have and where human action would be significant.

This insufficient understanding has been acknowledged to play on users' vulnerability, believing with the purchase of a device they are purchasing security and anti-virus protection. Beardo and Whitehouse, (1993: 92) clarifies that consumers misinterpret characteristics of the device such as 'system security' or 'secure operating system'. After having an existential amount of experience, they discovered that the public takes these terms as security for malware however explain, 'systems cannot guarantee nor prevent the threat of viruses'. Focusing their study on the psychological framework it took a different approach and investigated why users have little knowledge or underestimate the threat of malware. This approach contributes significantly to this project drawing attention to the levels of knowledge users have, this will be further investigated with the users' occupation to determine if it may impact their familiarity. Due to these findings, the hypothesis is reasonable, as other scholars have found similar patterns.

Fruhlinger, (2019) highlighted his personal opinion that applications being affected by malware are transferred from device to device just like a virus in a human, referring to it as a 'logic bomb'. He explained the bomb has properties that ensure the virus' payload executes at

specific times or under certain conditions, explaining why participants in the ONS study were oblivious of a virus being present. The researcher established cause-effect relationships by highlighting ‘viruses are sneaky, they infect a device if the particular code is run for instance JavaScript within a web browser exploits security vulnerabilities to infect’ (browser scripting). Opening links passed through social media can simply activate the virus. He proposes that victims tend to be younger generation due to the amount of time they spend online, although they may be more mindful than older generations, this will be explored. His study had a positive outcome as the relevance of data analysis to the research question. Fruhlinger, (2019) concluded his report highlighting public wifi dangers, mentioning as an improvement to his study he would have considered this aspect.

2.2 Open Wi-Fi networks

Norton Security, (2017) collected data through 15,532 surveys to understand what individuals knew about public Wi-Fi including the dangers. Their findings revealed that a staggering 60% of participants believe their personal information is not at risk when using a public Wi-Fi service, nor can 53% distinguish whether the Wi-Fi is secure or not. These figures are intriguing to find out why or how individuals are unconscious of the dangers public Wi-Fi has, will be investigated throughout this research. The study had thorough and effective components within the design section which resulted in a 50 50 gender split and a positive outcome matching their hypothesis. The conclusion was validly based on the data and analysis received. With the shocking, however, expected results they drew on the need for more information given to the general public on the risks, a trend which has been seen throughout all mentioned studies.

Likewise to Norton Security, (2017), the CSO, (2019) conducted a study to combat criminal cyberattacks. Producing a report given detailed information about the dangers of public Wi-Fi for the general public. Without using scientific language to ensure the reader understands. It continuously explains that any information sent out using unsecured Wi-Fi such as e-banking, credit card information and login credentials are sent straight to the hacker who has intercepted the connection. They believe if a user has no other choice but to use public Wi-Fi they should know the 5 key steps to be proactive; different from Norton Security where they did not give security and proactive fronts on unsecured networks when they had the resources and power to do so. The CSO, (2019) mentioned ‘despite numerous warnings, headlines and efforts to educate many people still don’t understand dangers’ this is disagreed by Norton Security, (2017) who stated that there haven’t been any of these methods carried out to warn

individuals. Agreed by Bushell, (2018) highlighting the lack of information given to the public on the dangers, stimulating to find out the truth. This relates to this project as its to be investigated whether users have had information given to them on the dangers and if they have awareness on the topic.

A quantitative study conducted by ITRC, (2012) investigated consumer's knowledge of public Wi-Fi and their use of it. Aspects to be researched in this study. They discovered that users' main use of public Wi-Fi was for work purposes and the use of emails. Their results concluded that consumers use public Wi-Fi because it is free, however, this is an assumption. This research project will explore why it is used. They found that users had little or no knowledge of public Wi-Fi, which shows accuracy for the hypothesis. A similar study was conducted by Schlesinger, (2016) who found comparable results however, they revealed that younger generations use public wifi most. A problem noted with both studies was the lack of trends or patterns created with the occupation to determine why some users had more knowledge than others, which will be carried out in this research. However, the latter touched on social media use on the internet, and the potential of cyberattacks on digital media platforms.

2.3 Social media

Banerjee, (2019) is one of the few researchers who investigated this initially declaring his main finding that on social media there is always infiltration, known as 'social media hackers'. Specifying infiltration occurs through fake accounts to produce spam and the sending of malicious links and attachments. Throughout his research, he discovered that within 2012 over 12 million users had suffered identify theft and data leakages. With users in jeopardy of losing personal data, he made aware of the potential emotional damage, however, he did not investigate what type of emotions the victim would feel when ransomware blocks a computer system. This will be explored during this project, as it is important to know if it is just an assumption. Particularly looking at the fraudulent activity and attacks from social media loaded on public Wi-Fi. Digital media platforms have already been established as the main use of Wi-Fi in public areas by Tim, (2013), Katz et al, (2017) and Phillips and Young, (2009) which gives relevance to the hypothesis. This creates an explicit link between the two sections, whilst also understanding if the impact of these emotions would refrain individuals from using social media on unsecured Wi-Fi.

It is believed that 'security is a key integral issue for users' (Seyedi et al, 2011). As it grows so does the illegal infiltration of systems and deliberate duplication of users' data. The security required for social media is at an all-time high, due to the imperfection of laws and regulations. Seyedi et al, (2011) used content analysis and theoretical perspectives centred on legal laws (Gov UK, 2019). Through their research they significantly discovered that digital platforms tend to be 'cost-driven rather than value-driven in terms of security', clarifying that to improve security it should be customer's needs-oriented. A very unique point to this study was the motivation of legal acts and regulations being breached. It extenuated that users would refrain from clicking a link if it did not look legit. The severity of scam is increasing daily, with more having realistic features. This has much relevance to this project as it is to be investigated whether what users believe makes a site or links look non-malicious, investigating if users can distinguish wrong from right, what's absent within this study.

Menzheres, (2018) sums up that as technology advances so do fraudsters, giving the need for security issues to be apparent. Concurred by Chahar et al, (2013) in his report it is manifest that for successful social media the levels of security must be elevated. However, Menzheres, (2018) clarifies that not all illegal incidents which occur are due to attackers it can be as simple as the users' human error and the lack of spotting scam.

The most typical attacks on social media tend to be through links or attachments, Milletary, (2005) discovered that Facebook tends to have a higher scam rate than other platforms. The study had a strong relationship between research and theories, whilst stimulating the concern of what users portray a scam to look like. Ayers, (2013), Rosenblum, (2007) and Wu et. al (2006) have carried out different studies all producing homogeneous results that suggest the majority of computer users do not know the features on a site or link which signals legitimacy. This will be examined through this research.

Osterman, (2016) provided statistics that social media has caused one in five to be the centre of malware attack. The study explained that 'cybercrooks have utilized the social media platforms to spread malware' and described them as the perfect 'breeding ground for hackers to exploit users'. This study also took into consideration the use of email as a potential playground for hackers to attack, in which they found 80% of malware has entered networks through emails. This stood out from other studies as it incorporated emails with the spreading of viruses. Their study used questionnaires to obtain the data, which proved a downfall as

they could not support the results with any participant's comments, a frequent limitation found.

2.4 Obsession of social media

The toxic trait of social media has made individuals spend 153 minutes a day on social media (Clement, 2019). Reinforced by Feeley, (2019) who discovered that consumers spend over 50 days a year on their smartphones due to obsession. He focuses his study on social media use and establishes a trend between the ages. With 16-25-year-olds spending the most time on their phones, due to their interest in technology and the amount of extra time they have compared to other age groups. His results found that a quarter of people believe technology and digital platforms are the keys to happiness, whilst half admitted that when their device battery falls under 10 per cent they become anxious, this alone determines the high level of obsession individuals have with social media. However, he does not investigate if the public knows the dangers social media brings.

In contrast, Lovink, (2001) states that if users had awareness of the dangers social media have with malware they would not refrain from using them or changing their routine. The Norton Security, (2017) reinforced this by emphasising that social media is an inflicting enormous platform that is growing with new users every day, meaning it's the perfect place for hackers. Highlighting if users were educated on the dangers their obsession would override, however, he did not investigate how long they spent on social media or if they would use public Wi-Fi for it. Unsure of the reasoning behind users' obsession outweighing the risk of viruses, it has been inspiring to find out within this project why individuals would put their data at risk of being stolen, or whether they didn't even know this could happen? Drawing on the relevance the hypothesis has on what is already known. A limitation of this study was the small sample size used due to the thorough investigation and there was no mention of the gender ratio which can negatively affect validity.

Scholars (Aytes and Connolly, 2003, Davinson and Sillence, 2010 and Hayden, 2009) have studied these risks, using focus groups and interviews to engage with different users and get full in-depth reasoning relating to the perception of risk. They separately discovered that individuals have the motive of 'it won't happen to me' and 'it's very unlikely to happen'. More research on this approach in line with public Wi-Fi would allow a clearer insight into what individuals think, particularly through online experiments where users would be 100% anonymous.

After doing thorough research a sector that required more research is the dangers of open Wi-Fi networks and social media which is the hypothesis for this project. Also was the lack of social media obsession linkage with viruses, and whether users could acknowledge a scam email or malicious/non-malicious link. This will be investigated for this project to understand what users pick out as ‘real’ features or ‘fake’.

3. Methodology

3.1 Participants

Seventy-seven participants were recruited using an opportunity sampling method, where potential participants were enquired via email and media platforms such as Facebook. Participants were required to be over 18. Thirty-five males, forty-one females and one ‘other’ filled out the online survey and online quiz. Open-ended questions were used to gauge the knowledge and experience of each individual and allowed giving detailed answers.

3.2 Design

Two major validity concerns were biasness and influence, to ensure these did not affect this project there were many factors considered such as having a range of participants through age, gender, and occupation to ensure a fair investigation takes place. The methodological approach used was surveys, this quantitative method is very sophisticated and effective to obtain and analyse data from participating responders. Using surveys gives the ability to use both open-ended and closed-ended questions, allowing freedom to explain personal experiences.

Surveys are recommended for topics that are known to have a phenomenon where individuals may have heard of the problems being investigated, appropriate for this project. They allow patterns to be shown and categorising of answers, ensuring predictor variables of awareness and social media to be investigated with public Wi-Fi scam and users’ knowledge as outcome variables.

Another method was used to gain a deeper understanding of users’ consciousness to real-life scam and malicious links were through the creation of an online user interface which undertook the name of a ‘quiz’. A similar method was used as researcher Smith, (2017) received an in-depth investigation with positive results that relates to this hypothesis. The interface gave participants 8 different questions, within each question was a malicious link or scam and a non-malicious one. With space for individuals to explain their reasons. However,

Smith, (2017) gave multiple-choice questions to his scenarios with extremely obvious wrong answers which may have swayed the participants to the correct answer and can influence bias. Allowing the participants to give their explanations will ensure efficient measuring of alertness and knowledge.

3.3 Materials

The survey concluded of 28 short questions, designed on ‘www.surveygizmo.com’ and found at <https://www.surveygizmo.eu/s3/90197407/Computer-Viruses> whereby the first four questions were demographic about the participant; last 4 digits of their phone number, gender, age and occupation which could explain why some users may have more knowledge than others (Refer to appendix 9.3).

The online interface quiz was created using the computing languages and algorithms of HTML, MySQL, and PHP. Sorting the coding under the three files named create_database.php, test.html and process_data.php (Refer to appendix 9.5). The creation of the database was initially done first, using the create database function to ensure there was an established place for the data to be stored. After creating using the mysql functions ensured that the correct localhost name, database name and password was inserted to guarantee a successful connection. Once the overall creation of the database had been completed the particular table ‘answers’ was created (where individuals responses would automatically be stored). The name of the field was given followed by the data type and length and any special requirements of that field, for example, the last4tel field was had the primary key property as all participants should have unique telephone numbers. See figure 1.

```

1 <?php
2 $conn = mysqli_connect("sql204.epizy.com", "epiz_25149807", "pr6N7yfbqMv");
3 // $sql = "CREATE DATABASE quizzes";
4 //mysqli_query($conn, $sql) or die(mysqli_error($conn));
5
6 $sql = "USE epiz_25149807_quiz";
7 mysqli_query($conn, $sql) or die(mysqli_error($conn));
8
9 $sql = "CREATE TABLE answers(
10     last4Tel VARCHAR(4) PRIMARY KEY NOT NULL,
11     links1 VARCHAR(10),
12     why1 VARCHAR(255),
13     links2 VARCHAR(10),
14     why2 VARCHAR(255),
15     links3 VARCHAR(10),
16     why3 VARCHAR(255),
17     links4 VARCHAR(10),
18     why4 VARCHAR(255),
19     links5 VARCHAR(10),
20     why5 VARCHAR(255),
21     image1 VARCHAR(10),
22     why6 VARCHAR(255),
23     image2 VARCHAR(10),
24     why7 VARCHAR(255),
25     image3 VARCHAR(10),
26     why8 VARCHAR(255)
27 );";
28

```

Figure 1: Creating the database

To assure the database was created successfully the phpMyAdmin was used to clarify this and displayed any data entered. The test.html file was next created, (Refer to appendix 9.5) this allowed the test questions to be displayed to the end-user. It was created using the form function to ensure the data would be displayed in a professional layout, (Refer to appendix 9.4). The form method of 'post' followed by the file name 'process_data.php' to ensure when the submit button was pressed it would connect to this file. For the participants to select their question the 'radio' input type was used as shown in figure 2.

```
<html>
<center>
  <head>
    <title> Malicious Links and Emails Quiz</title>
    <h1> Malicious Links and Emails Quiz </h1>
  </head>
  <body>
    <div>
      <form method="post" action="process_data.php">
        <p> Enter the last four digits of your phone number: </p> <input type="text" id="last4Tel" name="
last4Tel"/>
        <br/><br/>
        <h1> Links </h1>
        <p> Which of the following links looks malicious? </p>
        <br/><br/>
        <input type="radio" name="links1" value="url1"/> <a href="https://www.mimecast.com/content/
malicious-email-attachments/" onclick="return false;">www.mimecast.com/content/
malicious-email-attachments</a><br/>
        <input type="radio" name="links1" value="url2"/> <a href="" onclick="return false;">bit.ly/email/</a>
        <br/><br/>
        <p> Why do you think so? </p><br/><input type="text" id="why1" name="why1"/><br/>
        <br/><br/>
        <input type="radio" name="links2" value="url3"/> <a href="https://www.santander.co.uk/" onclick="
return false;">https://www.santander.co.uk/</a><br/>
        <input type="radio" name="links2" value="url4"/> <a href="http://www.santander.co.uk/" onclick="
return false;">http://www.santander.co.uk/</a><br/>
      </form>
    </div>
  </body>
</html>
```

Figure 2: creating the quiz

To distinguish between the two options, each URL or image was given its name and assigned number which increased throughout, for example, url1, url2, url3 this was the same for any images. Each pair of URLs and images (url1 and url2 or image1 and image2) had one reason area for participants to explain why they believed their answer was right. Once all questions were entered and the test.html was displayed correctly the styling was created (Refer to appendix 9.5). The creation of the process_data.php file was then done, this allowed the data entered by participants in test.html to be extracted and placed into the database using the '\$_POST' function and created a connection to the database and placed the data within the answers table. Once this was done successfully participants were presented with the message from figure 3.

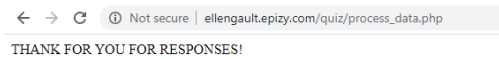


Figure 3: Quiz completed message

Once the files were linked successfully they were uploaded to the web hosting platform ‘InfinityFree’ to put the quiz on the web for participants to access any time of the day anywhere. It was then accessed through <http://ellengault.epizy.com/quiz/test.html> and participants were provided with this link.

Using this method enabled semantic analysis to be carried out, giving the participants 6 non-malicious or malicious links and 3 fraudulent emails and 3 real ones to determine if they could differentiate between them and their reasons. The computing language Python and Anaconda was used to analysis the participant's results.

Again the first question asked was the last 4 digits of the participant’s phone number as this allows a contrast to be made against their survey answers, specifically, to determine whether the user was truthful with these answers.

3.4 Ethics

For both methods, the research information outlined what the investigation was for, how they could withdraw and what will happen with their data. As well as the true aim of the study and how it would benefit them. It also upheld the right to withdraw by given contact details of the researcher to ensure they could withdraw at any point using the email address and their phone number (Refer to the appendix 9.1).

To ensure anonymity was upheld, all participants were not asked their name, address or any personal details which would give away their identity, to keep uniquely identify each set of the last 4 digits of their phone number was used.

Each participant was debriefed before taking part, to ensure they weren't deceived in any way. Additionally, there was no experiment to bias as the data was self-report and did not take place in an interview setting. This study had very low risk, therefore, they were protected from harm both physically and mentally

3.5 Procedure

Once a potential participant was recruited, they had the option to take part, in the case of taking part they clicked the link for the survey first where the research information sheet was first shown, (appendix 2). This ensured their age met the correct criteria and gave them all information about the research, contact details and information on how to withdraw from the study. The same procedure was done for the quiz.

The survey was to be completed first, followed by the quiz as this would allow participants to express their awareness, whilst the quiz determined if their awareness was to a high standard. The survey followed through the different sections of personal details, the devices they use followed by their perception of computer viruses and personal experiences of viruses. This opened up a set of questions for those whose device had a virus before to get a clear understanding of what happened, the signs of the virus and how it was eliminated, all to test the hypothesis. Once the questions about the viruses were completed it moved onto scam, which asked the users what makes a site look legit.

This section was further tested in the quiz. Following on from this, public Wi-Fi and social media were investigated, once the questions were all completed, they were thanked and contact information was provided for any inquiries they had or if they wanted to withdraw.

The quiz was then required to be completed by the same respondent, they were asked to choose the link or scam which they believed to be malicious. With their reasoning, this allowed analysis to occur between their quiz response and survey answers.

3.6 Analysis

Data analysis was conducted using the computing language Python and semantic analysis for this quantitative analysis. Using different algorithms to find the frequency, occurrence, prominence, complexity of words it took factors such as terminology used, level of correctness and length of the answer to calculate the level of knowledge among participants. Using the scale of 1-5 to rank the level of knowledge; 5 representing advanced knowledge, 4 high level, 3 medium, 2 low level and 1 no knowledge. The lower the ranking the less frequently or relevance it had to that particular question; therefore less knowledge. The

library 'numpy' would be imported along with 'nltk', 'collections' and 're'. To ensure words such as 'at', 'in' and 'it's' didn't return as common words these were replaced with space. Textblob was used to extract these common words, find the frequency and word count, (Refer to appendix 9.6).

4. Results

There were various questions asked through both methods to investigate this area of research. This project aimed to investigate whether individuals have awareness of computer viruses and if they were aware of their spread through social media and open Wi-Fi networks. This research looks at several different approaches as previously stated, therefore analysing the survey through graphs and the quiz through python and tabular data forms for comparison and strict analysis.

4.1 Demographic data

Gender

There was a gender split of 54% (41) females, 45% (35) males and 1% (1) who reported 'other', although ideally, a 50 / 50 split would be more appropriate this was not possible due to uneven gender distribution. (Refer to appendix 9.7).

Age

The majority of the respondents fell into the age category of 18-24 which has 24 participants (31%), closely followed by 45-44 with 18 (23%), 35-44 with 12 (15%), 25-34 with 11 (14%), 55-64 with 8 (12%) and 65+ with 4 (5%). Allowing there to be a clear investigation into whether age plays a role in the different levels of knowledge on the different variables. (Refer to appendix 9.7).

Occupation

The participants had a wide range of occupational status which can be seen in table 1, they have been grouped according to their occupation mentioned. The 'other' section includes those who fall under trades, manufactures, beauty, and volunteers. With only 7 participants falling under 'computing' their occupations are as followed with their age; Head of IT (55-64), Systems Manager (64+), Systems Administrator (35-44), IT Tech (25-34), Computer Science student (18-24), Administrative Officer (45-55) and Computer Science student (18-24). (Refer to appendix 9.7).

Table 1: Occupation among participants

Occupation	Number of participants
<i>Retail</i>	7
<i>Computing</i>	7
<i>Student</i>	20
<i>Education</i>	15
<i>Law</i>	3
<i>Business</i>	5
<i>Medical</i>	5
<i>Other</i>	13
<i>N/A</i>	1

Throughout the frequency, prominence, and rank will be shown. Frequency determines how often the word appeared throughout the answers whilst prominence is how important the words/answer was in regards to the question. The complexity rank was calculated through the length of their answer, the terminology used and the level of correctness, 5 representing advanced knowledge; 1 respecting no knowledge. To ensure analysing of data was to its greatest extent univariate was used each time.

4.2 Knowledge of viruses

Table 2 shows the most common words amongst the answers when participants were asked to define a virus, there was a range of terminology used to describe a computer virus. Those who used the word ‘malicious’ had much more knowledge than the 18 individuals who used the word ‘something’. The respondents who used the word ‘malicious’ were the same individuals whose occupation was ‘computing’. Whilst only 15 respondents being certain they know what a computer virus is, by stating ‘yes’. This question was crucial within the study to get an initial understanding of the knowledge the participants had.

Table 2: Displays the most common words used by respondents to describe a virus.

Word	Occurrences	Frequency	Rank
<i>Software</i>	19	4.1%	3
<i>Something</i>	18	3.9%	2
<i>Yes</i>	15	3.2%	N/A
<i>Virus</i>	15	3.2%	N/A
<i>Program</i>	9	1.9%	4
<i>Malicious</i>	9	1.9%	5
<i>Information</i>	8	1.7%	3
<i>Bugs</i>	6	1.3%	2
<i>Breaks</i>	6	1.3%	3
<i>Damage</i>	5	1.1%	3

From the analysis, when respondents were asked if they know the risks of computer viruses, only 31 respondents said ‘yes’. Those who answered with yes then listed what they believe as a risk, each answer was summarised into 2/3 words and placed into appropriate categories. This was explored against those who had previous experience with a computer virus; all 25 respondents who had also knew the potential risks. Each listing their outcome as a potential risk. It was somewhat surprising to see that 18-24 was the age gap that did not understand the risks of a virus; whereas the older generations did. It is apparent that those who had an occupation within computing used had more knowledge initially used terminology such as ‘corruption’. Shown in table 3.

Table 3: Those who knew the risks to viruses.

<i>Occupation</i>	<i>Age</i>	<i>Risk</i>	<i>Complexity Rank</i>
<i>Sales assistant</i>	<i>18-24</i>	<i>Theft</i>	<i>2</i>
<i>Early childhood</i>	<i>18-24</i>	<i>Device break</i>	<i>2</i>
<i>Education</i>	<i>25-34</i>	<i>Device break</i>	<i>2</i>
<i>Primary education</i>	<i>18-24</i>	<i>Theft</i>	<i>3</i>
<i>Computer science student</i>	<i>18-24</i>	<i>Corruption</i>	<i>5</i>
<i>Law</i>	<i>35-44</i>	<i>Theft</i>	<i>2</i>
<i>Business</i>	<i>25-34</i>	<i>Data loss</i>	<i>2</i>
<i>Volunteer</i>	<i>25-34</i>	<i>Data loss</i>	<i>2</i>
<i>Primary education</i>	<i>25-34</i>	<i>Corruption</i>	<i>2</i>
<i>Builder</i>	<i>45-54</i>	<i>Theft</i>	<i>1</i>
<i>NHS worker</i>	<i>25-34</i>	<i>Wipes system</i>	<i>2</i>
<i>N/A</i>	<i>25-34</i>	<i>Theft</i>	<i>1</i>
<i>Teacher</i>	<i>45-54</i>	<i>Theft</i>	<i>1</i>
<i>Lecturer</i>	<i>45-54</i>	<i>Hacking</i>	<i>1</i>
<i>Administrative officer</i>	<i>45-54</i>	<i>Corruption</i>	<i>5</i>
<i>Banker</i>	<i>45-54</i>	<i>Device break</i>	<i>3</i>
<i>Computer science</i>	<i>18-24</i>	<i>Data breach</i>	<i>4</i>
<i>Media and tv</i>	<i>55-64</i>	<i>Financial loss</i>	<i>1</i>
<i>Retail</i>	<i>45-54</i>	<i>Theft</i>	<i>1</i>

<i>Pharmacy</i>	<i>55-64</i>	<i>Theft</i>	<i>1</i>
<i>Retail</i>	<i>45-54</i>	<i>Theft</i>	<i>1</i>
<i>Grandparent</i>	<i>64+</i>	<i>Destroy device</i>	<i>1</i>
<i>Chef</i>	<i>35-44</i>	<i>Theft</i>	<i>1</i>
<i>Psychology</i>	<i>55-64</i>	<i>Financial loss</i>	<i>2</i>
<i>Economics</i>	<i>18-24</i>	<i>Ruins life</i>	<i>2</i>
<i>Head of IT</i>	<i>55-64</i>	<i>Corruption</i>	<i>5</i>
<i>Systems manager</i>	<i>64+</i>	<i>Corruption</i>	<i>5</i>
<i>System administrator</i>	<i>35-44</i>	<i>Theft</i>	<i>5</i>
<i>IT technician</i>	<i>25-34</i>	<i>Data loss</i>	<i>5</i>

Figure 1 shows that over 25 of the overall respondents would know what to do if they had a virus; this is the same number as those who reported having experienced a virus (figure 3). Suggesting that these users would only know due to experience; with further investigation into this only 71 participants responded to this question. Those who did not respond had little knowledge in the previous questions, which speculated not knowing how to terminate the virus.

Exploring this further to determine if all 25 respondents answered ‘yes’ to every being educated on computer viruses only 13 individuals had. Indicating there is an extremely low level of knowledge. Their education is shown in table 4.

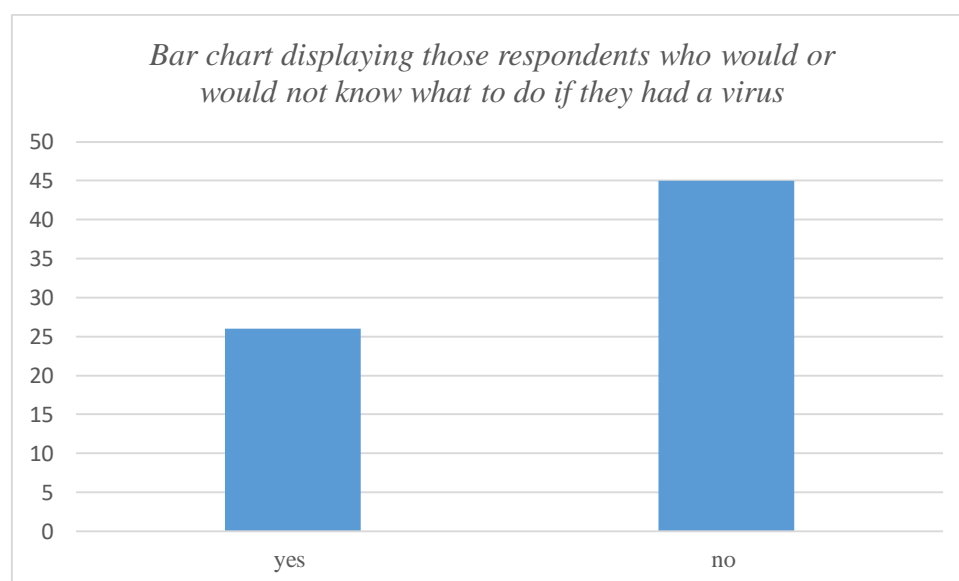


Figure 4: Future attack encounter.

Table 4: How individuals have been educated.

Occupation	Educated	Where / How
Computer science student	Yes	Highschool and dad who is a software developer taught me
Post office manager	Yes	Post office monthly training as we use computers
Manager	Yes	During my training as a manager in the shop but we have online help chat to use if we think we have one
Admin officer	Yes	Through my education to be an admin
Student	Yes	Through high school I did ICT
Sales assistant	Yes	Uncle owns a computer shop – he has warned me
Student	Yes	Learned in high school
Head of IT	Yes	Through training for the job
Systems manager	Yes	Through private course
Systems administrator	Yes	Uni course – software developer
IT Tech	Yes	Weekly courses and still go on monthly ones through school to keep up to date

Figure 2, shows that only 30% of individuals believed being connected to public wifi is an entry for malicious malware to be injected into a device; whilst 44% believed viruses could be passed through social media. Although these numbers aren't extremely low, there was an expectation that they would've have reached at least 50%, as discovered from previous research these are one of the most common ways viruses attack. 87% of participants believed that illegal websites were the main reason a device can become attacked. 76% of participants believed emails were also an entry point. It is suggested that these are so high due to those who previously experienced viruses getting onto their device through illegal websites and email.

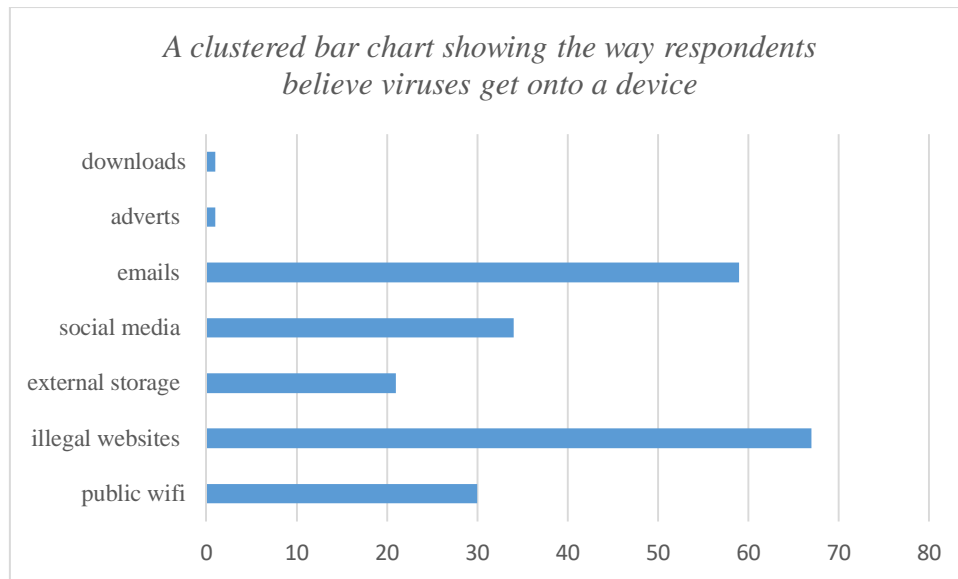


Figure 5: The platforms and apps users believe viruses get onto the device.

4.3 Experience of viruses

Figure 3 shows that nearly half of the respondents had been the victim of a virus attack and a staggering 20% of respondents are not sure if they had had a virus. While only a few opined otherwise. This implies that malware attacks are extremely common and with 20% of respondents being ‘not sure’ reinforces the lack of awareness.

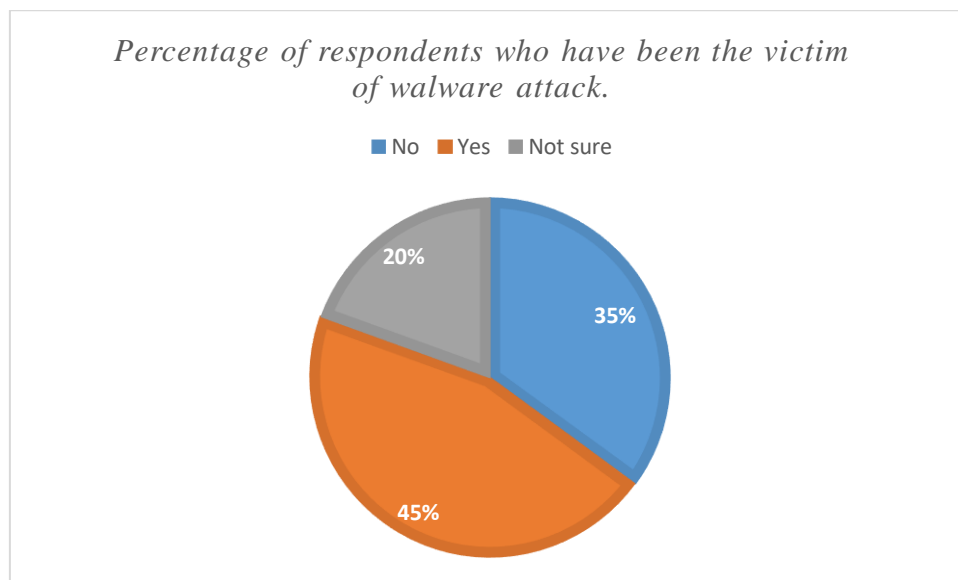


Figure 6: Virus infections encounter

For those respondents who reported ‘yes’ in figure 3, it was investigated if they knew what caused them to be at the centre of a malware attack. 13 out of the 33 respondents were ‘not sure’ how the virus injected their device. Shown in figure 4. The main cause reported was illegal movie sites, then followed by downloading files and finally through social media.

These respondents in the mentioned categories fall under the ‘web scripting’ virus and ‘browser hijacker’, the latter virus is typically used in movie sites where the end-user is found on a new web page which they did not click or enter, causing malicious code to be injected on their device.

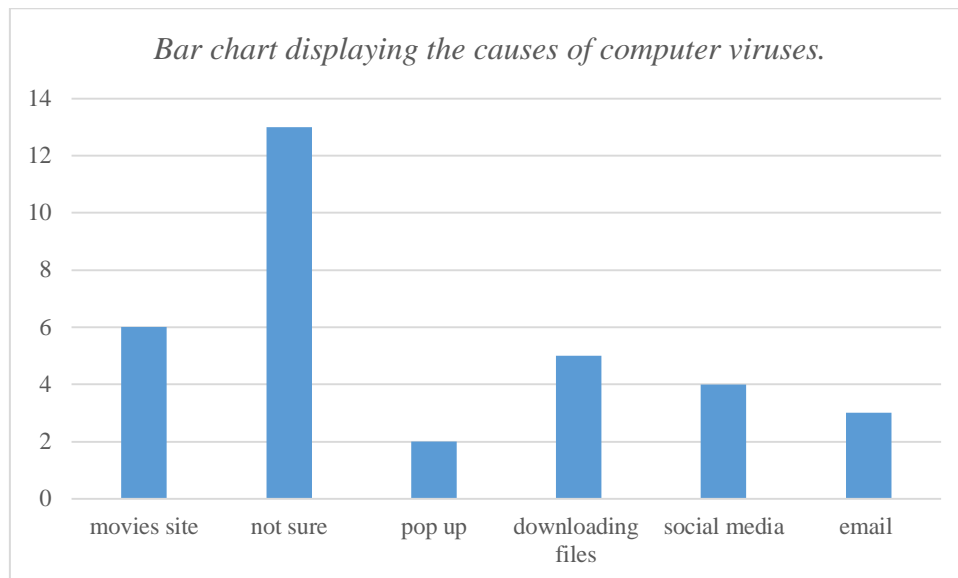


Figure 7: Virus attack causes

It was apparent to see from figure 5 that the majority of individuals who answered ‘yes’ to having experienced a virus they also reported to have installed anti-virus software on their device; with 29% of respondents who said ‘no’ it raises the question of whether they are attentive that this software exists or whether there were other factors which refrained them from purchasing.

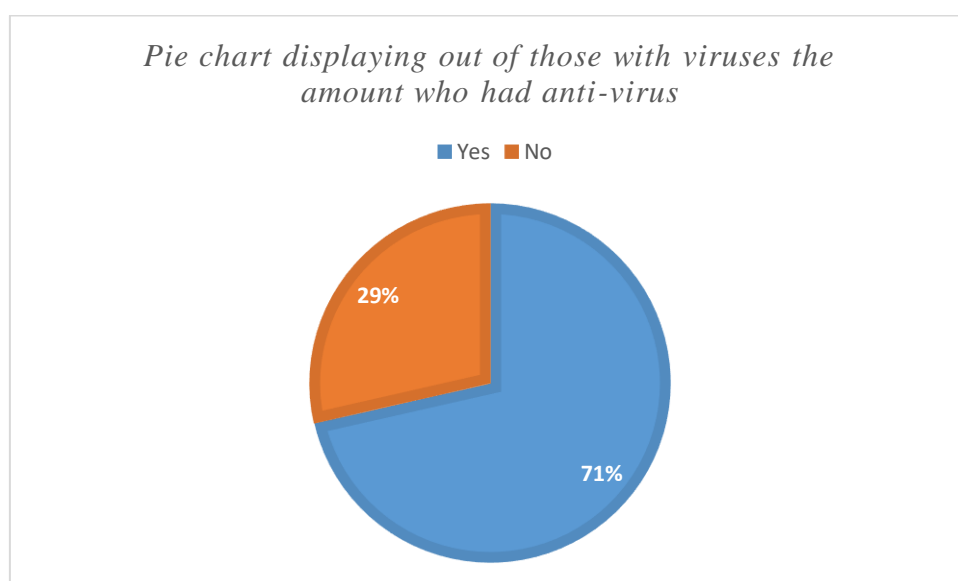


Figure 8: Anti-virus software counter

4.4 Impact on individuals' life.

Figure 6 shows only that 6 respondents' anti-virus software had detected and terminated the virus without the need for any professional help, this was surprisingly low when 25 of individuals reported they had anti-virus software. It was expected that this number would be much higher. It was appalling to see that 7 individuals required professional help and resulted in a bill of over £100, whilst 13 others reported paying under £100. Only 1 respondent paid less than £50 the others were all more.

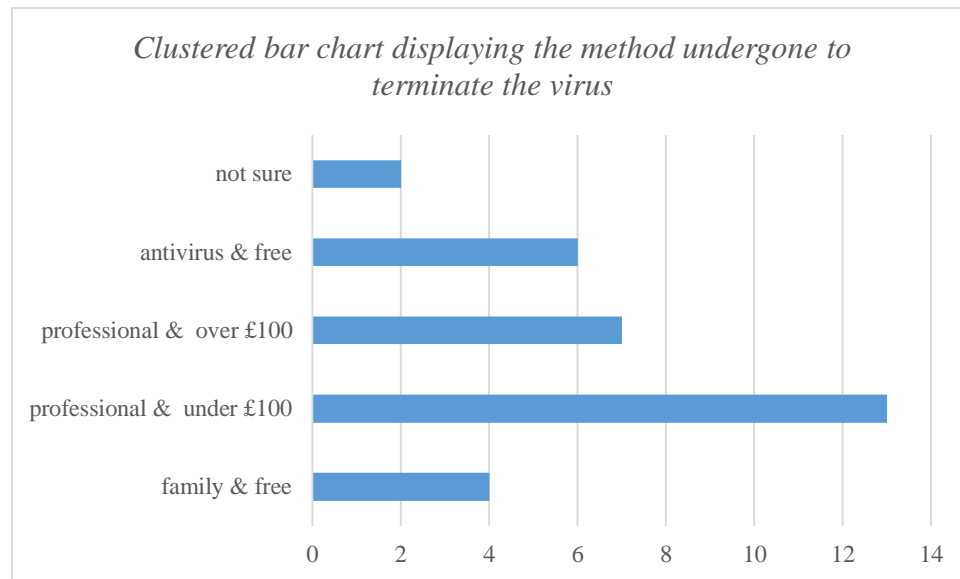


Figure 9: The termination methods undergone

From the analysis of response, it presented how users felt during an attack or if they were to be in attack, 74% of respondents agreed they would feel worried, whilst 48% reported they would encounter anxiety towards the situation and confusion was followed at 35%. The 18% of participants who reported 'other' their responses were; annoyed, angry, concerned, frustrated, not bothered, unsure how to pay, not too fussed and unsure of what to do. It was then examined whether users felt this because they were unsure of what to do; with 45 of respondents stating they would not know what to do. Shown in figure 7.

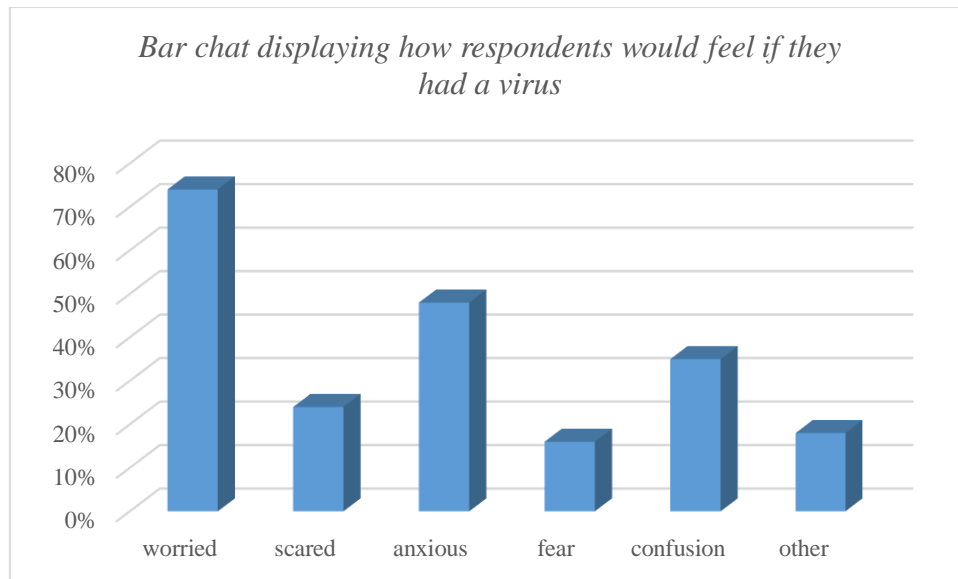


Figure 10: Emotions on virus attacks

4.5 Awareness to scams and malicious links

It was examined what users portrayed a site or URL to look legitimate, 26 of respondents knew that a padlock indicated some type of safety on a site. Shown in figure 8. This was a very shocking result as it was expected that the majority (66+ respondents) would know this. With only 12 participants recognising 'https' as a security measure; this suggests the level of knowledge is extremely low. Those individuals who said logo and popular sites suggest that they do not know any features and only use sites that have the company's logo. This can be easily duplicated by hackers and implies they are quite vulnerable and uneducated.

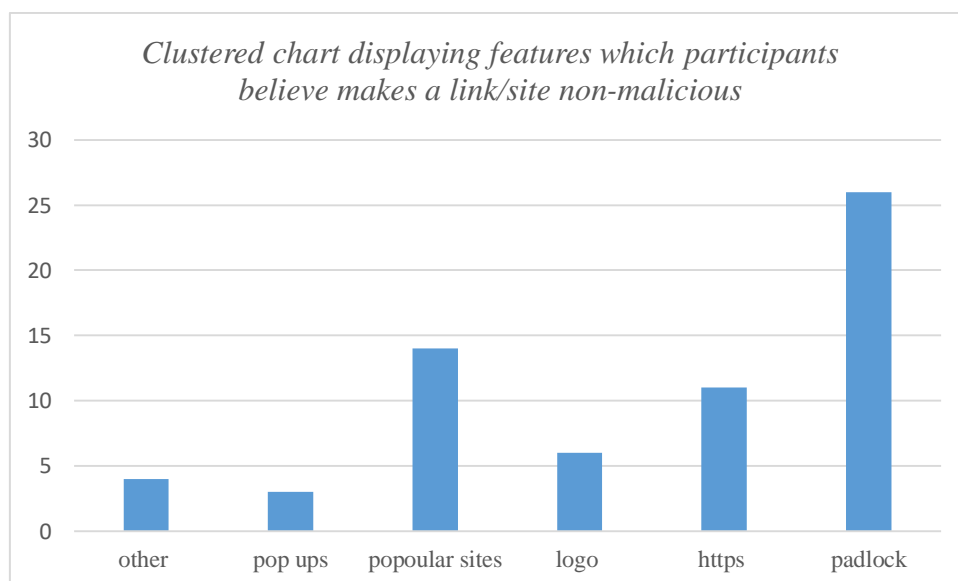


Figure 11: Security features

Figure 9 presents the participant's answers to question 2 in the quiz. With 44 respondents getting this correct, it is clear from that their reasons included the missing 's'. This gave reassurance as the 's' was mentioned 31 times. However, the 'end' mentioned 18 times and the 'two' mentioned 15 was concerning as the correct answer did not include these; then studying the answers they believed there should be two dashes at the end. It was shocking that 42% of individuals did not notice the 's' missing or did not refer to this as a security measure, and believed for a link to appear non-malicious there should be two dashes at the end. (Refer to appendix 9.8).

```
multi = C:/Users/ELIEN/Documents/0155/analysis/RESULTS /
[('s', 31), ('the', 26), ('no', 22), ('at', 21), ('there', 19), ('end', 18),
('be', 18), ('two', 15), ('should', 14), ('https', 12)]
[('question2.txt',)]
```

Figure 12 Common responses from quiz question 2

Figure 10 displays the most common words amongst all answers from question 6 in the quiz; 61% of respondents got correct and 39% got it wrong. It is clear that 'Microsoft' was the most commonly used word amongst the answers, indicating that those who answered wrong believed that Microsoft is not a correct mail server. Users believed it should be 'Gmail' with it featuring 20 times. 11 of those respondents who answered incorrectly justified their reason with 'Microsoft doesn't sound right' and 'customer service isn't real name'. Those incorrect answers were extremely poor and concerning as it is hard to comprehend some users would believe Microsoft would use 'microsoft@gmail.com' to contact their client. (Refer to appendix 9.9).

```
defaultdict(<class 'int'>, {'microsoft': 30, 'doesn't': 11, 'sound': 2, 'right': 11, 'customer': 19, 'service': 16, 'is': 4, 'not': 7, 'an': 1,
'appropriate': 1, 'email': 27, 'it': 2, 'wouldnt': 12, 'be': 13, 'gmail': 20, 'real': 11, 'for': 3, 'a': 31, 'big': 2, 'company': 8, 'weird': 1,
'would': 4, 'never': 1, 'have': 8, 'account': 12, 'services': 3, 'the': 9, 'say': 9, 'full': 1, 'address': 9, 'think': 2, 'own': 2, 'just': 1,
'looks': 2, 'fake': 1, 'isnt': 23, 'proper': 6, 'domain': 3, 'as': 1, 'like': 3, 'normal': 1, 'because': 2, 'they': 5, 'do': 2, 'look': 2,
'someone': 2, 'you': 4, 'can': 4, 'contact': 4, 'but': 2, 'use': 5, 'of': 1, 'normally': 1, 'gmailcom': 1, 'part': 1, 'customerservice': 3,
'name': 7, 'might': 1, 'should': 11, 'ends': 1, 'such': 1, 'dont': 1, 'so': 1, 'from': 1, 'someones': 1, 'personal': 1, 'their': 1, 'strange':
2})
```

Figure 13: Common answers amongst question 6 in the quiz

It was reassuring that the majority of participants got the first question of the quiz correct, table 5 shows that those who got the answer correct 28 of them acknowledged the shortness of the URL whilst the other 8 picked up on the missing 'www'. However, it was shocking to see no one referred to the top-level domain as being malicious, a usual top-level domain is known to be either .com or .co.uk. The domain featured in the quiz was '.ly' which is extremely uncommon, also making it shocking that the 30% who got it wrong would consider this to be normal. Amongst the wrong answers, it was very worrying that those who see 'malicious' in the URL automatically believed it was malicious. This also raises the question

of whether it was a guess by the participants as they had a one in two chance to get correct. Suggesting the participants may have guessed their reason based on the URL they picked. (Refer to appendix 9.10 for the full list of answers).

Table 5: Comparison of wrong and right answers from quiz question 1.

	Reason	Times appeared	Frequency %	Prominence
Right 70%	<i>Too short</i>	28	24.3	76.6
	<i>No www</i>	8	6.4	45.5
	<i>Not long enough</i>	6	4.5	29.2
	<i>Looks suspicious</i>	3	1.8	5.73
	<i>Not normal url</i>	1	0.9	5.75
	<i>Bitly never seen</i>	3	1.8	9.5
	<i>Not sure</i>	3	1.8	1.8
Wrong 30%	<i>Says malicious</i>	25	18.4	50.1
	<i>Minecast are frauds</i>	1	0.9	25.3
	<i>Obvious a spam</i>	1	0.9	22

The key findings that emerged from table 6 demonstrated the participant's knowledge to be very low with 80% of participants getting question 3 from the quiz incorrect. This result highlights that when users do not see 'www' or the full company name they are swayed to thinking it is malicious. This cast a new light on the comparison from table 3 as the malicious URL in that question did not have www either and why users were quick to believe this one was malicious and the other one was not. The participants who got the answer correctly picked up on the security feature 's', however only 7 individuals did and looking further into it it was those with the computing occupation. Those who had the reason of 'not sure' confirmed that users may be guessing on a one in two chance. (Refer to appendix 9.11 for the full list of answers).

Table 6: Comparison of wrong and right answers from quiz question 3.

	Reason	Times appeared	Frequency %	Rank
Right	<i>No https</i>	7	22.6	3
	<i>No / at the end</i>	2	6.5	4

19.4%				
	<i>Not official name</i>	1	3.2	4
	<i>Not sure</i>	5	7.1	1
<i>Wrong</i>	<i>No www</i>	51	38.9	1
80.6%				
	<i>No Hilfiger</i>	34	26	2
	<i>Not full name</i>	4	3.1	3

There was a significant decrease in the amount of knowledge demonstrated from table 2 to table 6, it was reassuring that within question 4 (table 7) of the quiz there was a 22% increase in the respondents who got the answer correct. However, it is apparent from the table that again participants believed with no ‘www’ the URL would be malicious. It was not as shocking to find that over half of the participants got this answer wrong; considering the clear trend which has been discovered through the previous questions. This trend has established those participants with jobs that do not require the use of computing have lower knowledge or awareness of malicious links. The file path ‘Index.html’ appeared again with a disappointing 84.1% frequency among the wrong answers; however, there was a theme in these responses. Those who mentioned the file path also mentioned it looked ‘strange’ or they ‘had never seen before’, which also provides insight that they have seen a question mark in the middle of a URL and a dash at the end which both symbolises suspiciousness. (Refer to appendix 9.12).

	<i>Reason</i>	<i>Times appeared</i>	<i>Frequency %</i>	<i>Rank</i>
<i>Right</i>	<i>Question mark</i>	27	29	5
	<i>/ at the end</i>	2	6.5	4
41.5%				
	<i>Not real url</i>	2	2.2	2
<i>Wrong</i>	<i>No www</i>	51	38.9	1
58.5%				
	<i>Index.html</i>	55	84.1	2
	<i>Index is strange / weird</i>	20	30.4	1

<i>Not sure what index means / never seen</i>	20	30.4	1
-----------------------------------------------	----	------	---

Table 7: Comparison of wrong and right answers from quiz question 4

Table 8 provides an overview of the final question of the quiz; 63.6% wrong and only 36.4% got it correct. Those who mentioned the email address 3 of them had ‘not sure if this is right’, ‘not sure but’ and ‘looks fake, not sure how’ which increased the speculation that this was a guess due to their lack of confidence. Those who used the reason as attn were the same individuals who said no to public WI-FI. Whilst the missing padlock was only noticed by one individual; Male, 25-34 whose occupation is a trainer. This was surprising as it was not recognised by anyone else, suggesting this one individual always looks for this feature before entering his details. Those individuals who responded within the category of ‘Apple doesn’t use invoices’ also stated they ‘had never seen this before’ which made them believe it was a scam. This was very surprising as they never observed the malicious signs of ‘attn’ instead of the customer’s name. The ‘weird font’ category had a high frequency of 22% amongst the wrong answers; this suggested that these individuals may have been guessing what to write or speculates they are android users. However, although they got the answer wrong it was reassuring to see they picked out on a ‘time limit’ and ‘negative message’ as features that could represent a scam. Although, in this case, it was not a scam and it was just offering a full refund up to 15 days. (Refer to appendix 9.13).

Table 8: Comparison of wrong and right answers from quiz question 8.

	Reason	Times appeared	Frequency %	Rank
<i>Right</i> 36.4%	<i>Email address</i>	8	9	4
	<i>Says attn</i>	7	6.5	3
	<i>No padlock</i>	1	2.2	3
<i>Wrong</i> 63.6%	<i>Apple don’t use invoices</i>	15	18	1
	<i>Never seen before</i>	12	5.8	1
	<i>Weird font</i>	20	22	2
	<i>Time limit</i>	5	5.5	3
	<i>Spelling mistakes</i>	6	7	1

<i>Sending negative message</i>	2	2.2	2
<i>Not sure</i>	3	3.6	1

4.6 Knowledge and use of public wifi

When respondents were asked if they had been educated on public wifi dangers; it was apparent only 7 users said yes. Their reasons and occupations can be seen in table 9. The pattern of occupation again was discovered here as users within the computing industry had been aware of the dangers. With only one been educated through a family member.

Table 9: Respondents who said yes to being educated on public Wi-Fi dangers.

Occupation	Educated	Where / How
<i>Computer science student</i>	<i>Yes</i>	<i>My dad warned me</i>
<i>Admin officer</i>	<i>Yes</i>	<i>Discussion with colleagues and have regular training on the subject</i>
<i>Head of IT</i>	<i>Yes</i>	<i>Through reading e-magazines on technology dangers to society – I have an extreme interest in this sector</i>
<i>Systems manager</i>	<i>Yes</i>	<i>My job is based on hackers so this is an area I research regularly</i>
<i>Systems administrator</i>	<i>Yes</i>	<i>I have been educated through online websites which updates the public and those within the field to keep up to date with the way hackers use open networks to intercept</i>
<i>IT Tech</i>	<i>Yes</i>	<i>I was educated through my online courses and training for my job; it is also an aspect I'm interested in so I always keep up to date through technology websites to see what new methods hacker intercept a network</i>

From figure 11, the majority of participants all take part in the use and connecting to public wifi; this was to no surprise. With 8% of respondents stating they do not connect to public wifi; it was the respondents shown in table 9. It was to no surprise that these individuals all had been educated through their job's training, through education before their jobs, private courses and through family members who had an interest within the sector.

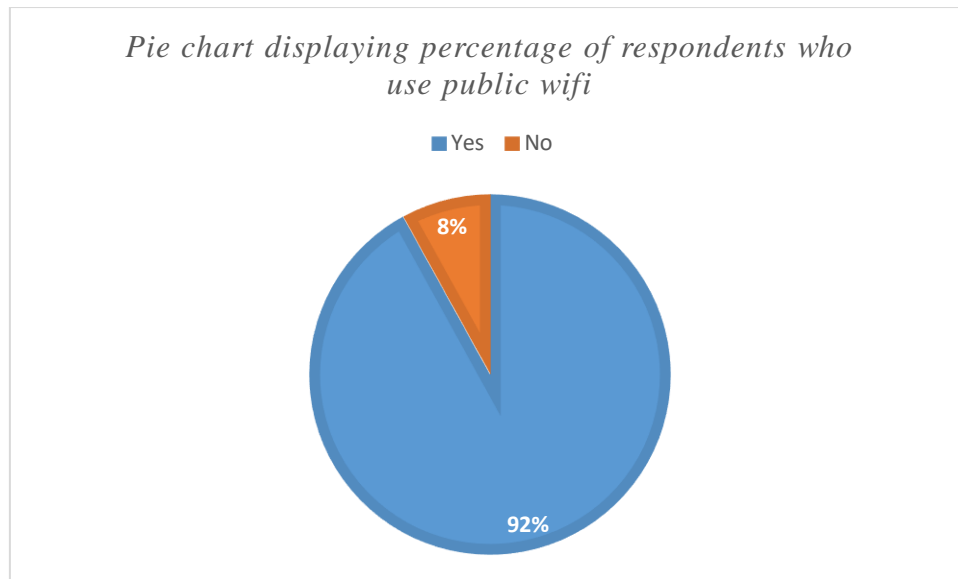


Figure 14: Those who use public wifi

The analysis of results showed that when public wifi was used for social media at 85% and communication such as WhatsApp and email came at 83%, shown in figure 12. It was very alarming that 30% of individuals would use public wifi for e-banking, and 33% would use it for online shopping. Entering login credentials for e-banking or bank card details on a public wifi network is extremely dangerous it subsequently gives the hacker all the data they are trying to hack.

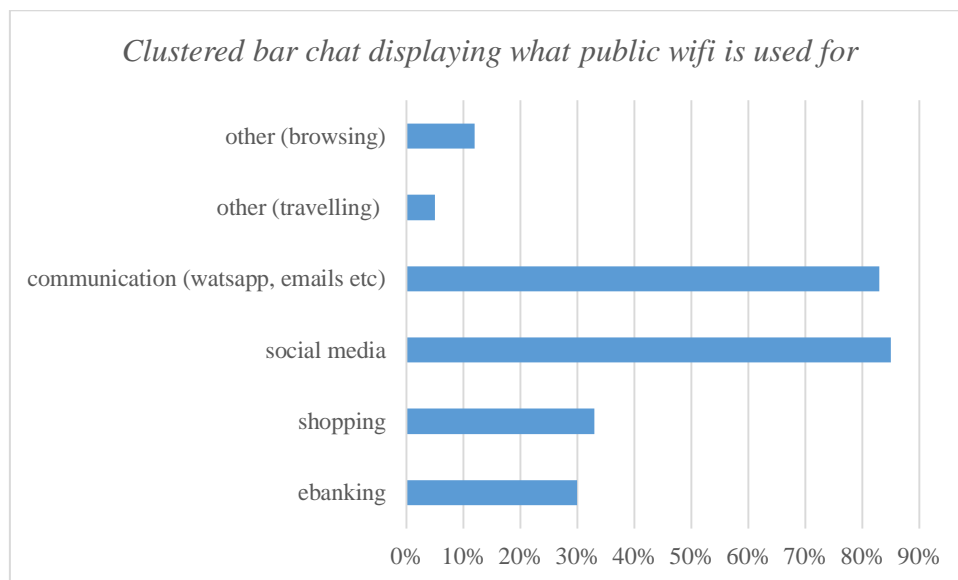


Figure 15: The use of public wifi

The results of this study were compared to two other, seen in table 10 to determine whether the results within this study are accurate. It shows the different variables investigated throughout this study, it is clear throughout all studies that the majority of users reported to

using public wifi. It was shocking to see that study 1 respondents did not report using public wifi for social media, however, emails were reported by 56% of participants. Whereas social media platforms were mostly used for public wifi in this study and the second most used in study 2. There was a pattern established among all 3 studies; that participants reported to public wifi because it is free, it gave results that users had limited or very limited knowledge on the risks public wifi has whilst over half of the participants of all studies would continue to use public wifi regardless of the dangers.

Table 10: Comparison results of this study and two previous ones.

Variable	This Study	Study 1 <i>ITRC, (2012)</i>	Study 2 <i>Schlesinger, (2016)</i>
<i>% of respondents who use public Wi-Fi</i>	92%	79%	87%
<i>Most common place to use public Wi-Fi</i>	<i>Hotel (90%) Entertainment centre (87%)</i>	<i>Coffee shop/ Restaurant</i>	<i>Airport Cafe</i>
<i>The most common use of public Wi-Fi</i>	<i>Social media (85%) WhatsApp etc (83%)</i>	<i>Emails (56%)</i>	<i>Email (58%) Social media (56%)</i>
<i>Most common risk noted by respondents</i>	<i>Majority unsure; Or theft of data</i>	<i>Identity theft</i>	<i>Theft of data</i>
<i>% of respondents who will continue to use public Wi-Fi</i>	79%	84%	95%
<i>The most common reason for using public Wi-Fi</i>	<i>‘Free and no mobile data’</i>	<i>‘It’s free’</i>	<i>‘Just want to get online’</i>
<i>Overall aware of dangers</i>	<i>Very limited</i>	<i>Limited</i>	<i>Very limited</i>
<i>% of those who think public Wi-Fi is safe</i>	70%	56%	60%

4.7 Social media obsession

The results shown in table 11 based on social media use comparison method to determine whether the results were accurate; it is positive to see another study carried out by Feeley, (2019) found very similar results to this study. With users spending an average of 3 hours of their day only on social media; with the younger generation dominating this. Both studies included the level of awareness among all participants on scams and links passed through social media; from this study, it is clear to see that users had ‘very limited’ knowledge of the scams. Calculated through the percentage of individuals who got the answer right (36.3%)

and wrong (63.4%) shown in table 6. It was discovered within both studies that there was a lack of education given to participants; with study 1 participants blatantly explaining they ‘can’t tell a difference’.

Table 11: Comparison of results on social media use.

<i>Variable</i>	<i>This Study</i>	<i>Study 1 Feeley, (2019)</i>
<i>Most time spent on social media</i>	<i>2-3 hours daily (58%)</i>	<i>3 hours daily</i>
<i>Based on the platform</i>	<i>All</i>	<i>Facebook</i>
<i>Most common age</i>	<i>18-24 25-34</i>	<i>16-25</i>
<i>Level of awareness of scams/malicious links</i>	<i>Very limited</i>	<i>Very limited</i>
<i>Reason to little awareness</i>	<i>Not educated</i>	<i>Can’t tell a difference ; ‘They look legit’.</i>

5. Discussion

5.1 Summary of findings –

The present study aimed to explore whether users knew the dangers of computer viruses and public Wi-Fi. Exploring their use of social media and if a potential obsession would cause them to become ignorant of the risks and encourage continuous insecure use. The ‘malicious links’ quiz allowed the user to demonstrate their knowledge. The factors of age, gender and occupation were examined to enlighten why respondents had more awareness than others. The results of the investigation verified that there was a significant increase in knowledge and awareness of all variables if the user reported an occupation or education within a field of computing. However, when demonstrating their knowledge within the quiz their knowledge significantly dropped.

The proposed hypothesis that the majority of individuals will not have much knowledge of viruses in general, or how public wifi is used as a playground for hackers was accepted, as was users being oblivious to the risks due to social media obsession. As was the lack of knowledge of scam and links passed through social media and emails. The hypothetical idea mentioned in chapter one was proven positive and accurate based on the results.

5.2 Explanation and interpretation of the findings

A study conducted by Gross and Rosson, (2007) highlighted that end-users are known as the weakest link in computer security; unless they are part of an official organisation who sets strict policies to ensure they are conscious of potential dangers. This study used interviews to distinguish the level of knowledge; dissimilar from this study however using interviews allows a more profound understanding of the user's perception with the qualitative data. Conversely, due to the cost and prevention of bias, this would not have been suitable for this study. The use of an online survey and quiz enabled there to be strict confidentiality, due to no face to face contact with participants nor was their name given. To determine whether users had been educated on viruses or public Wi-Fi, it was yielded using a yes/no question with clarification space for users to expand their answer. Moreover, this enabled the quick separation of who had been educated or not despite the contribution it could have to contradiction as participants may report no but demonstrate knowledge through the quiz.

Virus awareness and knowledge

To measure the level of knowledge and awareness participants had, the scale of 1-5 was used through the different variables, as discussed in chapters 3 and 4. This measurement of knowledge was created by Helenius, (1994) during his study of measuring end-user security knowledge. It proved a success and many scholars (McGarry, (2005), Bryant et. al, (2008) and Love and Arachlilage, (2014)) have all used a scale very similar and have measured respondent's knowledge and awareness accurately with success through terminology used.

Previous research has found that the general public has very basic or no knowledge of computer viruses, (Wood et al, 1987:9). Saudi, (2007) study substantiated that the term 'computer virus' has been recognised or heard by nearly every individual, particularly those who use a device or similar technology frequently. This was reinforced by this research study as shown in table 1, the most common words used by respondents to define a virus included; 'software', 'something', 'the program'. Although these answers were not completely correct it was notable that the users had a rough idea of what a virus was, the use of 'something' indicates that respondents had awareness of a virus however did not have the knowledge to define it with the correct terminology. Only 9 individuals used the correct terminology 'malicious' and 'program' given them a high rank of 4 or 5. As coincided by Helenius, (1994) confirming the hypothesis initially proposed a lack of knowledge.

This was reinforced when 20% of respondents were 'not sure' if they have been at the centre of a malware attack, scholars Wash, (2010) questioned why during his study and reported it was due to the lack of training given to home computer users. This view is shared by Fruhlinger, (2019) who also gave insight that these users could be the victim of a 'browser hijacker' where there are little or no signs and users may find themselves on a random webpage and with no knowledge did not contemplate why.

Beardo and Whitehouse, (1993: 92) revealed that users underestimate the security on such platforms and apps; they would not regard them as being dangerous because they are extremely popular. This supported the study when users were asked what caused the virus; 13 out of 33 were 'unsure' again and believed movie sites to be the biggest cause of viruses. Although this is somewhat true; the minority only acknowledged downloading files, emails, and social media. Whereas movie sites are to an extent illegal so stands out as being a virus playground.

Further analysis of the user's knowledge was demonstrated only 6 individuals' antivirus detected and terminated the virus effectively. The speculation of why users with antivirus software can still find themselves in the middle of an attack was answered by Rao and Nayak, (2014). They observed different factors that caused users' anti-virus to miss viruses the first being again, the silent virus which has many advanced properties it can be undetected during scans. They exposed that users' are not aware their anti-virus needs updating or repurchased. Norton Security requires a manual scan completed by the user; not automatically done. Users also may believe they have an anti-virus installed when they purchased the device, as agreed by Beardo and Whitehouse, (1993: 92), implying those participants who answered yes to having anti-virus software may have been oblivious of these different factors to ensure their software is functioning effectively. Thus the findings of this study support the literature as there is a strong link with the lack of knowledge users have to protect their device.

This investigation discovered that participants did not establish a link between malware and public Wi-Fi or with social media, as only the minority noted these variables as a potential malware environment. Shelke and Badiye, (2013) contribute largely to this area and highlighted that public Wi-Fi and social media are determined as 'safe spaces' by the general public with an explanation that they have been in the dark of the 'uses and abuses'. This is agreed upon by Chin. et al, (2012). In line with these studies, it stimulates the motive that if public Wi-Fi and social media were not given as options the participants may not have

considered them. Those participants also had a ranking of 1-2 throughout. This relates to the hypothetical idea that participants lack in knowledge of all variables.

Previous research has shown that the impact of malware attacks can have ameliorating effects on individual's lives (ONS, (2018) and Lévesque, (2018)). Within this study, it is reasonable to say that individuals would feel some type of emotion if they were to be at the centre of an attack. Dashora, (2011) confirmed this during their findings and said that there is an opportunity for viruses to have financial and emotional damage on end-users. Linking this to the results of this study, it showed significant impact users would go through. The different emotions of anxiety, confusion, fear and worried were all mentioned by participants; only 2 participants had 'not bothered' attitudes. Others declared they would be concerned and unsure of how to pay for the termination. Subsequently, if users were aware of preventive methods for viruses the impact on their emotions would not be as intense, with participants mentioning they would be 'unsure of what to do'. Moreover, those respondents who have already been through a malware attack and termination would not have such a significant impact on their lives, as they have experience. Therefore, contradicting the hypothesis slightly as there was no remark for those who have been at the forefront of a cyberattack.

Public Wi-Fi awareness

As noted in the literature review, there is a wide body of research that argues on the level of knowledge on public Wi-Fi. Norton Security, (2017) discovered that individuals had never been allowed to learn about the dangers of public Wi-Fi, this study supported this with only 7 knowing. This view is shared by ITRC, (2012) and Schlesinger, (2016) both demonstrated that the majority of respondents use public Wi-Fi; likewise to this study with only 8% reporting they don't. All studies discovered the most common places to use public Wi-Fi was; hotels, cafes, airports, and entertainment centres. Ullah, (2012) revealed that if an individual is said to be in a place longer than 20 minutes, they will connect to Wi-Fi. This is strongly evidenced in the current study, as respondents admitted to connecting the public Wi-Fi because 'it's free'. All found similar findings that 79%-95% of respondents would continue to use open Wi-Fi despite the dangers. This lack of knowledge was supported by sufficient evidence of a majority of respondents believing public Wi-Fi is safe. Thus the findings supported the hypothetical idea that users do not know this variable. However, it was revealed that all ages used public Wi-Fi within this study rather than the expected younger generation; as proposed in Schlesinger, (2016) study.

Occupation

Numerous investigations into users' perception and knowledge of said variables did not consider internal factors, for example, age, gender, and occupation to explore whether these had an impact. These give a full explanation as to why individuals have different levels of knowledge (Breakwell et al, 2006). In this study, there was a significant difference in those individuals who had a job or occupation outside the sector of computing and technology. Including hairdressers, students, builders, law, business all had rankings of less than 3 which indicates poor knowledge. Those within the sector had rankings of 4 and 5; using terminology and their complexity of answers. This was also reinforced when they answered 'yes' to having been educated on public Wi-Fi and viruses; through their job, private courses and personal research. Occupations ranged from 'systems manager', 'head of IT' and 'computer science student'. Passey, (2017) found similar findings. Indicating insufficient knowledge among other participants was accurate and due to unfamiliarity. This is supported by Wang et. al, (2019) revealing those working in dissimilar fields are at a disadvantage, particularly older adults (50+) as they had never the 'choice' nor the opportunity during their education period.

Age

Carpenter & Buday, (2007), Fox, (2004) and Jones, (2009) believe that older users lagged behind those younger users in awareness of viruses, scam, and public wifi dangers. However, this study found disparate results. When excluding those with technical backgrounds, 18-24 and 25-34 had the lowest knowledge rankings. Szor, (2005) proposed that it is typically the older generation who are more actively concerned about the topic. Brown, (2007) supported this study, revealing that older users were more than those younger users. This was significant through the virus section of this study. During the investigation, it exposed the older age groups to have more attacks with computer malware; implying that they have more concern and awareness through their personal experiences. Fletcher, (2019) supports this by highlighting that older generations are more exploited to malware; due to their prior lack of understanding after each attack, they are more conscious and take additional precautions. It was revealed in the current study they spent around 3.5 average hours on social media daily; with Age UK (2014) bringing light to social media scam. Moreover, this was not seen during the quiz responses.

Gender

With a very close gender split, it was discovered that there were no trends regarding male to female answers, whereas Kirkup, (2007) discovered a trend that males had more knowledge than girls due to the stereotypical association of males dominating the computing field. This statement can be supported by this study as those who had an occupation within computing were all males. However, there was no relationship established between gender and knowledge of any variable despite this. Koenig, (2008) revealed similar findings where no established link was made between gender and knowledge on this chosen topic.

Awareness of malicious links and social media

Previous research conducted by Glass et. al, (2016) discovered that hackers tend to exploit human flaws in spotting maliciousness than system flaws. Data was collated through focus groups and interviews to understand how participants engage in malicious URLs and attachments. Although the data collection was disparate from this study they found similar results. Using focus groups and interviews would allow a more in-depth analysis to be carried out. it allows capturing the natural language used by respondents, whereas the use of surveys allowed individuals to have time to think and used google for the correct terminology leading to unreliable results. However, the use of interviews and focus groups could allow more suited questions to be asked depending on the level of knowledge demonstrated in previous answers. However, the factors of time and cost would refrain from happening. It would require participants to meet at a setting to undertake the group, limiting sample size, a broad range of ages, occupations, and level of knowledge as the majority of participants would-be students. Thus, the online quiz would allow upholding anonymity and refrain from participants feeling intimidated if they answered incorrectly.

Furnell, (2017) investigated the security features within links that the general public had, it found that there was a higher number of individuals who recognised the 's' at the end of https than any other feature. The findings from this study rejected this, as it was established that only the minority of individuals oppose this feature; those who recognised were those within the computing occupation. Conversely, the majority of individuals reported the 'padlock' to be the main security feature, however, this is only when purchasing and is not part of the URL. Thus, demonstrations that they did not know about the 's', demonstrating the low level of knowledge and confirms the proposed hypothetical idea. Huang et. al, (2013) supports this view and proposed that users find it difficult to comprehend that the two dashes after https are non-malicious whereby two dashes at the end of an URL should cause concern and be seen as

malicious. This research ran parallel with the results from this study as it discovered that majority of participants believed a URL to be non-malicious as it contained two dashes at the end.

Robertson, (2010) carried out a comparative study which assessed the malicious links users click passed through social media, supporting this study as users could not comprehend if it was a scam or not. It was discovered that participants spent on average 2-3 hours on social media daily and admitting to little education to digital media scams, with Feeley, (2019) finding similar results. Bányai, (2017) and James, (2017) highlighted that if a user spends 3 hours a day on these platforms it is considered an obsession. Individuals interact with different features online, those which are malicious become normalized to the end-user as they see them frequently. This study supported this and was apparent when individuals see a question mark in random positions in a URL it's non-malicious when it's anything but. Sahoo et. al, (2007) found comparable results and discovered a question mark can be the most significant way to spot a malicious link and those who cannot have a lack of scam knowledge. It was apparent those who used social media less had a higher ranking (3 or 4) through the quiz than those who spent more than 3 hours on social media (rankings of 1 or 2). The normalisation appeared again when the majority of individuals believed a link to be malicious if it did not contain 'www', or '.com' as individuals who reported this had all scored low or no knowledge rankings. Sophos, (2014) found comparable results and noticed if users do not see these features individuals automatically have a negative attitude perception. Confirming the hypothetical idea that obsession has caused ignorance and Menzheres, (2018) voiced that attacks can occur because of human error and lack of spotting scam. Alongside the 'it won't happen to me' attitude, apparent from some participants in this study and Ayres, (2013).

As noted in chapter two Osterman, (2016) revealed how dangerous email platforms are for injecting malware into all devices, with Chen, (2011) noting during his findings that users tend to be more conscious of scams through email than other platforms. Through a critical evaluation of this study, it found dissimilar results, where individuals had higher rankings in question 1 with the shorter URL and had much lower rankings during the scam email at question 8. Individuals who choose the wrong answer believed that 'Microsoft' was not a real mail server, whereas they believed Microsoft's email address would have the server of Gmail. Anafo and Ngula, (2020) contributed to our understanding within the realm of email scam and exposed that end-users believe an email must have a well know mail server such as

‘yahoo’, ‘sky’ or ‘outlook’ to be deemed non-malicious. The small number of participants who were aware of this and got the question right were those with advanced or high knowledge for previous questions. Findings from this present study seemingly confirm previous research and the proposed hypothesis.

In conclusion, this study does support the finding that there is little knowledge or awareness when participants were asked about their knowledge of the variables and exposed to malicious links and scam. Many areas of research when investigating the different variables do not focus on factors that could affect an individuals’ knowledge. This study aids development in this area and focused on the age, gender and occupation of participants to determine if there was an impact on the demographic aspect. Moreover, there is more research required to reconcile the conflicting results.

6. Future research and limitations

For future research to ensure there is a deeper investigation all factors should be considered to significantly improve the study and make it more valid and reduce confounding variables. Incorporating a larger sample size would increase accuracy and reliability. Particularly having an even amount among all age ranges will grasp a sufficient contrast of knowledge among the pool of participants. Coinciding along with the ethical restrictions that respondents must be over 18. On the contrary, in the future extending the study to under 18s would allow a better-quality contrast of knowledge, those under 18 may have more knowledge on the variables due to newer curriculums in schools and other educational settings. This study consisted of having participants from each group, however, 18-25 held the most participants. Having a sufficient number across all groups would prevent generalisation to occur, (Smith, 2018).

This study did not consider the respondents’ level of education during analysis, it would have been sufficient to assess the respondent’s education level to determine if it impacted the results. Schilirò, (2010) investigated the level of education within his study, finding that those who attended schools 40 years ago had little or no education and significantly lacked knowledge in the explored variables. This was a limitation within this study and concerning future research, it would ensure a rich clarity of respondents’ background was analysed against their knowledge. Additionally, there has been a vast amount of research mentioned in the literature review on the lack of knowledge, it would be adequate to further investigate

those who had in-depth knowledge. Furthermore, having a different set of questions for each level would greatly improve this research area.

Implications from this study highlighted the lack of literature on malicious scams and links passed through social media and the obsession of digital media platforms. Any current research focuses on the assumption of individuals being aware of the most common security features such as 's' in https; however, this study proved differently. Future research could focus on the impact social media has on malware attacks. There are findings to suggest that social media obsession can cause users to become ignorant of the malicious signs, however no research on why individuals still choose the obsession over the potential consequences? Within this society and world, these platforms are seen as the way forward, particularly from staying in touch with family and friends across the world. However, not many people are aware of the negative dangers social media can bring about. Future research should investigate and explore the development of social media and what initially caused the obsession whilst given information to users to educate them on the cyberattacks through these platforms. Focusing on how to spot fake accounts along with malicious links, (Egele, et. al 2013).

The significant low-level knowledge on public Wi-Fi indicates there is a need to be cautious when individuals are using the networks. It had been suggested that companies require the end-user on their network to provide their own security measures (Sagers et. al, 2015). Jansen and Grance, (2011) carried out a study with findings that ran consistency with this study, however, made the assumption users are 'held in the dark' about this. They did not investigate whether users on open networks were conscious they had to make their practical security measurements. In order for generalisability can be obtained it is important that in future research this assumption is valid with evidence. In line with this current study and previous studies did not explore the measurements which users did take; if any, such as the type of firewall and anti-virus used on their device. This evaluation is lacking within this research area and can provide explanations for attacks.

Furthermore, it should be considered that some participants may endure significant learning difficulties, this is a potential issue when filling out the survey or completing the quiz. Some individuals find it difficult to read, write and describe their reasoning and even the question being asked. This could lead them to have a disadvantage of the current study or moreover, be excluded. This reduces the validity as not every group can complete the survey or the quiz.

In reference to future research, a better means of data collection could be used to ensure all minority groups could take part, for example, a focus group.

7. Conclusion

The overall purpose of this study was to investigate user's knowledge and awareness of computer viruses and whether they were aware of the potential risks of connecting public Wi-Fi particularly for social media use. Whilst exploring if their obsession made them ignorant of malicious features of scam and links. The factors of age, gender and occupation were assessed to examine if they have a role in the different levels of knowledge among participants. Past research did not look at all these factors therefore this piece of research is extremely relevant and contributes to this area. The findings of the research show that only the occupation impacted the participant's knowledge and ability to acknowledge scams; however, those with more computing occupations still failed to detect major malicious features. As well as users spending excessive amounts of time on social media caused them to become oblivious to suspicious features as they have become normalised. The overall study produced positive and accurate results regarding the hypothesis whilst demonstrating with other academics mentioned in chapter 2 there is a need of a law on educating the general public on malware, scams and public Wi-Fi dangers.

8. References

18% of Companies Get Malware Infections Through Social Media Platforms. (n.d.).

Ahlgrimm, J.M., Andrews, N.K., Binder, M.G., Eckardt, J.P., Ey, C.S., Feeley, A.P., Festa, L.M., Guden, J.C., Ingram, A.D., LaCoss, S.D. and Lopez, S., 2008. News Media and Strategic Communications Industry, Industry Study, Spring 2008. NATIONAL DEFENSE UNIV WASHINGTON DC.

Aïmeur, E. and Schönfeld, D., 2011, July. The ultimate invasion of privacy: Identity theft. In 2011 Ninth Annual International Conference on Privacy, Security and Trust (pp. 24-31). IEEE.

Anafo, C. and Ngula, R.S., 2020. On the grammar of scam: transitivity, manipulation and deception in scam emails. *WORD*, 66(1), pp.16-39.

Arachchilage, N.A.G. and Love, S., 2014. Security awareness of computer users: A phishing threat avoidance perspective. *Computers in Human Behavior*, 38, pp.304-312.

Aycock, J. (2006). *Computer Viruses and Malware*. Advances in Information Security. Springer.

Ayers, M. and McCaughey, M. (2014). *Cyberactivism*. Hoboken: Routledge.

Banerjee, S. (2019). PRIVACY, SECURITY & HEALTH RISKS OF SOCIAL MEDIA & HOW TO PREVENT THOSE. [online] RS Web Solutions. Available at: <https://www.rswebsols.com/tutorials/internet/privacy-security-risks-social-media> [Accessed 9 Dec. 2020].

Bányai, F., Zsila, Á., Király, O., Maraz, A., Elekes, Z., Griffiths, M.D., Andreassen, C.S. and Demetrovics, Z., 2017. Problematic social media use: Results from a large-scale nationally representative adolescent sample. *PLoS One*, 12(1).

Beardon, C. and Whitehouse, D. (1993). *Computers and Society*. Intellect Books.

Bishop, M., 1991. An overview of computer viruses in a research environment.

Bushell, K.E., 2018. EXAMINING THE RISKS OF INSECURE PUBLIC WI-FI NETWORKS.

Carpenter, B. D., & Buday, S. (2007). Computer use among older adults in a naturally occurring retirement community. *Computers in Human Behavior*, 23(6), 3012–3024.

Chin, E., Felt, A.P., Sekar, V. and Wagner, D., 2012, July. Measuring user confidence in smartphone security and privacy. In *Proceedings of the eighth symposium on usable privacy and security* (pp. 1-16).

Cimpanu, C. (2016). [Blog] Available at: <https://news.softpedia.com/news/one-in-five-companies-get-malware-infections-via-social-media-502603.shtml> [Accessed 13 Nov. 2020].

Dashora, K., 2011. Cyber crime in the society: Problems and preventions. *Journal of Alternative Perspectives in the social sciences*, 3(1), pp.240-259.

Drew, H., 2019. *Complete History of Social Media: Then and Now*.

Egele, M., Stringhini, G., Kruegel, C. and Vigna, G., 2013, February. Compa: Detecting compromised accounts on social networks. In NDSS.

Facebook scam ads - How can you spot them and what should you do?. (2019). [online] Available at: <https://www.telegraph.co.uk/technology/0/facebook-scams-can-spot-should-do/> [Accessed 10 Dec. 2019].

Facecrooks. (2013). 85% of Consumers Use Social Media Networks While Connected to Public WiFi. [online] Available at: <https://facecrooks.com/Internet-Safety-Privacy/85-Consumers-Use-Social-Media-Networks-While-Connected-Public-WiFi.html> [Accessed 13 Dec. 2019].

Fletcher, E., 2017. <https://www.ftc.gov/news-events/blogs/data-spotlight/2019/03/older-adults-hardest-hit-tech-support-scams>. [Blog] Consumer Protection, Available at: <<https://www.ftc.gov/news-events/blogs/data-spotlight/2019/03/older-adults-hardest-hit-tech-support-scams>> [Accessed 5 April 2020].

Fox, S. (2004). Older Americans and the internet. Retrieved March 15, 2009, from <http://www.pewinternet.org/Reports/2004/Older-Americans-and-the-Internet.aspx>

Fruhlinger, J. (2019). What is a computer virus? How they spread and 5 signs you've been infected. [online] CSO. Available at: <https://www.csoonline.com/article/3406446/what-is-a-computer-virus-how-they-spread-and-5-signs-youve-been-infected.html> [Accessed 3 Nov. 2020].

Furnell, S.M., Bryant, P. and Phippen, A.D., 2007. Assessing the security perceptions of personal Internet users. *Computers & Security*, 26(5), pp.410-417.

Garcia, A. and Wood, C. (1987). *Computer Security*. New York u.a.: Wiley.

Goodin, D. (2017). Serious flaw in WPA2 protocol lets attackers intercept passwords and much more. [online] Available at: <https://arstechnica.com/information-technology/2017/10/severe-flaw-in-wpa2-protocol-leaves-wi-fi-traffic-open-to-eavesdropping/> [Accessed 14 Nov. 2019].

GOV UK, 2019. UK To Introduce World First Online Safety Laws. GOV UK.

Gross, J.B. and Rosson, M.B., 2007, March. Looking for trouble: understanding end-user security management. In *Proceedings of the 2007 Symposium on Computer Human interaction For the Management of information Technology* (pp. 10-es).

Grossman, J. (2007). *XSS exploits*. Burlington, Mass.: Syngress, p.27.

Hu, H., Myers, S., Colizza, V. and Vespignani, A., 2009. WiFi networks and malware epidemiology. *Proceedings of the National Academy of Sciences*, 106(5), pp.1318-1323.

Huang, D., Xu, K. and Pei, J., 2014. Malicious URL detection by dynamically mining patterns without pre-defined elements. *World Wide Web*, 17(6), pp.1375-1394.

Identity Theft Resource Center, 2012. Public Wifi Usage Survey. [online] Available at: <https://www.idtheftcenter.org/images/surveys_studies/PublicWiFiUsageSurvey.pdf> [Accessed 14 September 2019].

- Jansen, W.A. and Grance, T., 2011. Guidelines on security and privacy in public cloud computing.
- James, T.L., Lowry, P.B., Wallace, L. and Warkentin, M., 2017. The effect of belongingness on obsessive-compulsive disorder in the use of online social networks. *Journal of Management Information Systems*, 34(2), pp.560-596.
- Jones, S. (2009). Generations online in 2009. Retrieved March 15, 2009, from <http://www.pewinternet.org/Reports/2009/Generations-Online-in-2009.aspx?r=1>
- Kamthane, A. and Kamal, R. (2012). Computer programming and IT. New Delhi: Dorling Kindersley.
- Koenig, A.M., 2018. Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. *Frontiers in psychology*, 9, p.1086.
- Kramer, S. and Bradfield, J.C., 2010. A general definition of malware. *Journal in computer virology*, 6(2), pp.105-114.
- Legezo, D., 2016. Research on unsecured Wi-Fi networks across the world. SecureList. Retrieved March, 25, p.2019.
- Lévesque, F.L., Chiasson, S., Somayaji, A. and Fernandez, J.M., 2018. Technological and human factors of malware attacks: A computer security clinical trial approach. *ACM Transactions on Privacy and Security (TOPS)*, 21(4), pp.1-30.
- Li, N. and Kirkup, G., 2007. Gender and cultural differences in Internet use: A study of China and the UK. *Computers & Education*, 48(2), pp.301-317.
- Lovink, G., 2011. Networks without a cause: A critique of social media (p. 24). Cambridge: Polity.
- Maimon, D., Becker, M., Patil, S. and Katz, J., 2017. Self-protective behaviors over public WiFi networks. In *The {LASER} Workshop: Learning from Authoritative Security Experiment Results ({LASER} 2017)* (pp. 69-76).
- Mathur, K. and Hiranwal, S., 2013. A survey on techniques in detection and analyzing malware executables. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(4).
- McGarry, K., 2005. A survey of interestingness measures for knowledge discovery. *The knowledge engineering review*, 20(1), pp.39-61.
- Menzheres, A. (2020). Recent E-commerce Security Issues and Best Practices (2018). [Blog] Ecommerce.
- Milletary, J. and Center, C.C., 2005. Technical trends in phishing attacks. Retrieved December, 1(2007), pp.3-3.
- Moffitt, T. (2018). Social Media Malware is Deviant, Destructive. [online] Webroot. Available at: <https://www.webroot.com/blog/2018/06/25/social-media-malware-deviant-destructive/> [Accessed 26 Nov. 2019].

- Nachenberg, C., 1997. Computer virus-antivirus coevolution. *Communications of the ACM*, 40(1), pp.46-51.
- Niranjanamurthy, M., Kavyashree, N., Jagannath, S. and Chahar, D., 2013. Analysis of e-commerce and m-commerce: advantages, limitations and security issues. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(6), pp.2360-2370.
- NORTON WI-FI RISK REPORT Report of Online Survey Results in 15 Global Markets. (2017). [ebook] Norton Security. Available at: <https://www.symantec.com/content/dam/symantec/docs/reports/2017-norton-wifi-risk-report-global-results-summary-en.pdf> [Accessed 6 Oct. 2019].
- O'Leary, D.E., Kuokka, D. and Plant, R., 1997. Artificial intelligence and virtual organizations. *Communications of the ACM*, 40(1), pp.52-59.
- Oyelere, S.S. and Oyelere, L.S., 2015. Users' perception of the effects of viruses on computer systems—An empirical research. *African journal of computing & ICT*, 8(1), pp.121-130.
- P. Bryant, S. Furnell, and A. Phippen, (2008) *Advances in Networks, Computing and Communications 4*. University of Plymouth, ch. Improving Protection and Security Awareness Amongst Home Users.
- Passey, D., 2017. Computer science (CS) in the compulsory education curriculum: Implications for future research. *Education and Information Technologies*, 22(2), pp.421-443.
- Phillips, D. and Young, P., 2009. *Online public relations: A practical guide to developing an online strategy in the world of social media*. Kogan Page Publishers.
- Rao, U.H. and Nayak, U., 2014. Malicious software and anti-virus software. In *The InfoSec Handbook* (pp. 141-161). Apress, Berkeley, CA.
- Robertson, M., Pan, Y. and Yuan, B., 2010, November. A social approach to security: Using social networks to help detect malicious web content. In *2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering* (pp. 436-441). IEEE.
- Sagers, G., Hosack, B., Rowley, R.J., Twitchell, D. and Nagaraj, R., 2015, January. Where's the Security in WiFi? An Argument for Industry Awareness. In *2015 48th Hawaii International Conference on System Sciences* (pp. 5453-5461). IEEE.
- Sahoo, D., Liu, C. and Hoi, S.C., 2017. Malicious URL detection using machine learning: A survey. *arXiv preprint arXiv:1701.07179*.
- Schilirò, D., 2010. Investing in knowledge: knowledge, human capital and institutions for the long run growth. *MJ Arentsen, W. van Rossum, AE Steenge, Edward Elgar, Cheltenham*, pp.33-50.
- Schlesinger, J., 2018. Most people unaware of the risks of using public Wi-Fi.
- Seigneur, J.M., 2015, December. Wi-Trust: improving Wi-Fi hotspots trustworthiness with computational trust management. In *2015 ITU Kaleidoscope: Trust in the Information Society (K-2015)* (pp. 1-6). IEEE.

Shelke, P. and Badiye, A., 2013. Social networking: its uses and abuses. *Research Journal of Forensic Sciences*, 1(1), pp.2-7.

Skoudis, E. and Zeltser, L., 2004. *Malware: Fighting malicious code*. Prentice Hall Professional.

Smith, B., 2018. Generalizability in qualitative research: Misunderstandings, opportunities and recommendations for the sport and exercise sciences. *Qualitative research in sport, exercise and health*, 10(1), pp.137-149.

Statistics on Crime in England and Wales. (2020). Assessment of compliance with the Code of Practice for Official Statistics. [online] Office of National Statistics. Available at: https://www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-assessmentreport268statisticsoncrimeinenglandandwale_tcm97-43508-1.pdf [Accessed 3 Oct. 2019].

Szor, P., 2005. *The Art of Computer Virus Research and Defense: ART COMP VIRUS RES DEFENSE _p1*. Pearson Education.

Ullah, I., 2012. A study and analysis of Public WiFi.

Urmee Khan, D. (2012). 12 million people suffered a computer virus attack in the last six months. [online] *Telegraph.co.uk*. Available at: <https://www.telegraph.co.uk/technology/news/5317908/12-million-people-suffered-a-computer-virus-attack-in-the-last-six-months.html> [Accessed 10 Oct. 2019].

Vanhoef, M. and Piessens, F., 2017, October. Key reinstallation attacks: Forcing nonce reuse in WPA2. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1313-1328).

Vishwanath, A., Herath, T., Chen, R., Wang, J. and Rao, H.R., 2011. Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems*, 51(3), pp.576-586.

VPN, G., 2019. The History Of Public Wi-Fi And Why It Has Become A Problem - GOOSE VPN Service. [online] GOOSE VPN service. Available at: <<https://goosevpn.com/blog/the-history-of-public-wi-fi-and-why-it-has-become-a-problem>> [Accessed 3 November 2020].

Wang, P., Dawson, M. and Williams, K.L., 2019. Improving cyber defense education through national standard alignment: case studies. In *National Security: Breakthroughs in Research and Practice* (pp. 78-91). IGI Global.

Wang, P., González, M.C., Hidalgo, C.A. and Barabási, A.L., 2009. Understanding the spreading patterns of mobile phone viruses. *Science*, 324(5930), pp.1071-1076.

Wash, R., 2010, July. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security* (pp. 1-16).

Wiethölter, S., Emmelmann, M., Andersson, R. and Wolisz, A., 2012, June. Performance evaluation of selection schemes for offloading traffic to IEEE 802.11 hotspots. In *2012 IEEE International Conference on Communications (ICC)* (pp. 5423-5428). IEEE.

Yazdanifard, R., Edres, N.A.H. and Seyedi, A.P., 2011. Security and privacy issues as a potential risk for further ecommerce development. In International Conference on Information Communication and Management-IPCSIT (Vol. 16).

Alghamdi, B., Watson, J. and Xu, Y., 2016, October. Toward detecting malicious links in online social networks through user behavior. In 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW) (pp. 5-8). IEEE.

Zhauniarovich, Y., Khalil, I., Yu, T. and Dacier, M., 2018. A survey on malicious domains detection through DNS data analysis. ACM Computing Surveys (CSUR), 51(4), pp.1-36.

Sänger, J., Hänsch, N., Glass, B., Benenson, Z., Landwirth, R. and Sasse, M.A., 2016, May. Look before you leap: improving the users' ability to detect fraud in electronic marketplaces. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 3870-3882).

9. Appendix

9.1 Appendix 1: Ethics form approval

Section 1 Your Details

Researcher

Ellen Gault (17001172)

Title of Proposed Project

Analysis and an investigation into the impact viruses have to mobile phones and personal computers users particularly focusing on those passed through social media platforms and open WiFi networks. Looking in-depth at whether these users have an awareness of what a virus is and their knowledge of them.

Programme Title and Level of Study

INFORMATION TECHNOLOGY AND MEDIA & COMMUNICATION

Research Dates

Start Date:

2019-11-01

End Date:

2020-04-21

Department

First department

MATHEMATICS, COMPUTER SCIENCE AND ENGINEERING

First Supervisor Name

Dr Neil Buckley

If your research is part of two departments, please enter the details of the second department here:

Second department

MEDIA AND COMMUNICATIONS

Second supervisor name

Anthony-Ridge Newman

Professional Guidelines Referenced

If your research does not reference any professional guidelines, please enter "No Guidelines Referenced"

No Guidelines Referenced

Section 2. Who will be taking part?

Will other people be taking part in your research?

Human participants

Section 3.1 General

The full title of the research project

Analysis and an investigation into the impact viruses have to mobile phones and personal computers users particularly focusing on those passed through social media platforms and open wifi networks. Looking in-depth at whether these users have awareness and knowledge of what a virus is and they're knowledge of them.

Aims and Objectives

The aims of this research project is to examine the impact and effects of viruses passed through both social media and open wifi networks, considering whether the users are really aware of what a virus is and besides, are they even conscious they have one? The project will go through the way this malware can get onto their devices and the dangers of open wifi networks for vulnerable computer users, looking at the dangers links and messages passed through social media may have.

The objectives and outcomes for this research proposal are to ensure users are mindful of what links they open on social media, their security if they have any with open wifi networks in particular, using e-commerce. The result should make users more aware that clicking links or using wifi networks may cause 'silent' harm to their personal devices, give them a deeper perception of how to act on these viruses.

Please give a brief outline of the research study

Ensure that you include details of the design (qualitative/quantitative, etc) as well as the methods and procedures (questionnaire, interviews, experimental trial, observation etc)

There are now millions and millions of malicious software, subsiding them into different categories. Including Logic Bomb, Trojan Horse, Back Door, Virus, Worm, Rabbit, Spyware, Adware and Hybrids as discussed by Aycock, (2006 : 3). The most common malware today is viruses, which is a very board term as there are multiple types within this particular malware, this research project will focus on the Web Scripting and Browsers Hijacker viruses also known as Cyber Attacks.

A crime survey conducted by ONS, released information that over 17 million UK computer and mobile device users were attacked by viruses in 2018 alone, resulting in over £130 billion being stolen. Given the reasons for conducting a research project based on this topic as it is evident that it is a growing concern in today's world and if users were more acknowledge on the malware it could reduce these numbers dramatically. Ahmad, (accessed 24/10/2019), reinforces this in an article called 'Computer Viruses as a Threat to Home Users', he highlights that there should be an increasing awareness of computer viruses to users and 'basic IT

security principles will help home users eliminate the threat of computer viruses’.

To obtain this I will use quantitative research, in the use of questionnaires or surveys as this will allow me to gain a full understanding of the user's knowledge of the particular topic, I will also use semantic approaches to get a full analysis of the results.

Some types of research must be referred (by the Faculty Ethics Research Sub-Committee) to the University Research Ethics Sub-Committee. Therefore, please state here if your research involves or may involve deception, the use of covert methods, matters involving national security, illegal activity or might endanger the University’s reputation. Please also highlight the key aspects which cause it to fall into one or more of these categories.

This is not applicable for my research study, as it is based on anti-virus malware, which the participant will be of knowledge.

Where will the study take place and in what setting? If in a workplace, or if the participants are from a workplace (e.g. a school), identify what your connections are with that workplace

The study will be conducted online, using a survey or questionnaire software. Allowing participants to complete this online rather than in a workplace or setting means they can do it in confidence as I will not personally collect them or meet them face to face, resulting in more honest and open answers. I will distribute these surveys through emails, social media and online messaging services.

Give a brief description of your target sample (e.g. age, occupation, gender)

The target sample will be for all genders, and those over the age of 18 as the study doesn't apply to children, It will be based on students as this is the popular target audience mobile phones and typically have their personal computers, I will ask for their occupation in terms of study or jobs, meaning I can see if they have any prior knowledge in this area.

Is the participation individual or as part of a group?

Individual participant

Define the special arrangements which will be made to deal with issues of informed consent (e.g. is parental/guardian agreement to be obtained, and if so in what form?) and also of the participants’ freedom to withdraw from the research at any time

As the audience is based on 18 years plus, this will mean there will be no need for any parental/guardian agreements required. To ensure participants are aware of what the study is there will be a declaration at the beginning before they start giving details of what the information is for, how it will be used and will highlight that confidentiality will be a key aspect. I will add my uni email for contact if they decide they do not want their information used if they decide during the questionnaire they don't want to complete it they can simply close it down and will not feel the need to submit.

How will participants be selected, approached and recruited? Identify clearly and analyse fully any issues of power relations that might arise, and say what steps you will take to alleviate them. This applies particularly if the location of the research is a place of the researcher's employment, or if they have other strong links with the participants.

I will email students from the university and put the survey on social media for participants to fill out, there will be no power relations that might arise as it is based on students and will be done in their comfort.

Is written consent to be obtained?

Yes

How will the participants' right to withdraw be ensured?

To ensure I withdraw the right participant there will be a field to ask the last 4 digits of their telephone number as this will uniquely identify each survey, but is a way in which they cannot be identified as a person due to no names being asked to keep it confidential. To ensure the participant knows their work is not being used and has been withdrawn appropriately by deleting the work they have submitted and reanalyse the rest of the work accordingly.

Section 3.2 Risk & Ethical Procedures

What potential risks are there of physical harm to participants? Please specify, and explain any steps you will take to address them.

Please note that "No risk" is not an acceptable answer - you must at least note a "Minimal risk" and explain this.

The names of the participants will not be used as this ensures confidentiality is not breached, the only minimal risk which may occur is mild fatigue because they are completing a survey which may be heavily based on questions about their experiences.

What potential psychological risks are there to participants? In particular, how might participate in this research cause discomfort or distress to participants? Please specify, and explain any steps you will take to address these issues.

Please note that "No risk" is not an acceptable answer - you must at least note a "Minimal risk" and explain this.

The only risk which may occur from this research project could be distress by the participation as it will open their awareness of viruses on laptops by asking certain questions if they have little or no knowledge they may panic as their device could have one without knowing.

Are there any risks to you as the researcher (and / or your co-researchers, if you have any) in this project? If so, outline the steps you will take to minimise them.

Please note that "No risk" is not an acceptable answer - you must at least note a "Minimal risk" and explain this.

Stress and fatigue may occur as the researcher, as it is very important to get the right meaning a lot of trial and error will be done when creating the survey/questionnaire. To minimise this

risk I will make sure the designing and conducting will be done continuously through a certain time to ensure there is no stress.

How might participants benefit from taking part in this research?

It will be to their benefit as it will raise their awareness of anti-virus software and malware which may cause harms, particularly asking if they have heard of silent viruses which do not show any signs or give messages on computers. This has the possibility for them to go back and do their research on the topic and find out if when using open WiFi networks it is safe or not, for example, the padlock when completing payments.

Does any aspect of your research require that participants be naïve (i.e. they are not given full or exact information about the aims of the research)? Please explain why and give details of the debriefing procedures you would use when the need for the naïveté is over

No this will not occur, I will ensure the declaration at the beginning will give the participant full insight of why they are completing it, including the aims and objectives I would like it to achieve from it.

Section 3.1. Data Security, Confidentiality, Anonymity and Destruction

Where and how do you intend to store any data collected from this research? Give details of the steps you will take to ensure the security of any data you collect.

Note that data protection regulations stipulate that data must be stored securely and not be accessible or interpretable by individuals outside of the project. Hence, data should be stored in a password-protected file on a password-protected device such as a desktop or laptop, and not on easily movable devices such as USB keys or CD ROMs

The data collected will be based online, meaning there will be a password to log onto the account and only I will have access to that password, to ensure I follow the Data Protection Act the data submitted by each individual will not be discussed with anyone else nor will it be stored on a USB which can be easily lost.

What steps will you take to safeguard the anonymity and confidentiality of personal records?

I will ensure there will be no names or addresses received during the completion by participants, once it is over I will destroy and delete all irrelevant information following the Data Protection Act.

Will this research require the use of any of the following?

Video Recordings

No

Audio Recordings

No

Photos

No

Observation of participants

No

Please provide a more detailed explanation of how you will ensure confidentiality and anonymity

Deletion of personal data

I will destroy all personal data recorded by the end of my degree programme

You should NOT make this point dependent on a successful outcome of your studies.

The date you will delete the data:

15th June

9.2. Appendix 2: Research Information Sheet



LIVERPOOL HOPE UNIVERSITY

RESEARCH INFORMATION SHEET

Outline of the research (a couple of sentences in non-specialist language)

Who is the researcher?

Name: Ellen Gault

Institution: Liverpool Hope University

Researcher's University email address: 17001172@hope.ac.uk

What will my participation in the research involve?

Your participation will involve answering questions laid out online, answering honestly will give me the chance to fully analyse the answers appropriately.

Will there be any benefits to me for taking part?

It will raise your awareness of virus malware, by asking questions if you have any knowledge on them.

Will there be any risks to me in taking part?

There are no serious risks to your taking part, however, you may feel mild fatigue when carrying it out, therefore, take breaks where necessary.

What happens if I decide that I don't want to take part during the actual research study, or decide that the information given should not be used?

Please email me on the email above, and I will remove your results, once removed I will email you back ensuring you know it has been destroyed or deleted.

How will you ensure that my contribution is anonymous?

I will not ask you for your name or address to ensure I cannot identify you as a person, I will ask the last 4 digits of your phone number in case you do decide to not take part after submitting it will allow me to use these 4 digits to identify the right results are being deleted.

Please note that your confidentiality and anonymity cannot be assured if, during the research, it comes to light that you are involved in illegal or harmful behaviours which I may need to disclose to the appropriate authorities.

9.3. Appendix 3: Online Survey

Name of researcher: Ellen Gault

Researcher contact email address: 17001172@hope.ac.uk

What was the purpose of the project?

The purpose of this study is to systemically analyse the impact computer viruses have on devices and the awareness users have to them, whilst thoroughly looking at the link between malware passed through social media.

What will happen to the information I have provided?

The information provided will be used in a dissertation study at Liverpool Hope University.

How will this benefit me?

It will benefit yourself, by increasing your awareness on the particular topic, making you more conscious of links, attachments and social media accounts.

Was I deceived in any way? If so, why was I required to be naive?

You were not deceived in any way; the true aims of the study were stated.

If I change my mind and wish to withdraw the information I have provided, how do I do this?

If this is the case please remember your 4 digits of your mobile number and email 17001172@hope.ac.uk asking to withdraw stating these 4 digits.

1. Please write the last 4 digits of your phone number. *

2. What age are you? *

- ☐ 18-24
- ☐ 25-34
- ☐ 35-44
- ☐ 45-54
- ☐ 55-64
- ☐ Other - Write In (Required) *

3. What is your occupation/study? *

4. What is your gender? *

- ☐ Male
- ☐ Female
- ☐ Rather not say
- ☐ Other - Write In (Required)

5. Which of the following do you use? Please select all that apply. *

- ☐ Mobile Phone
- ☐ Desktop Computer
- ☐ Laptop
- ☐ Tablet (iPad etc)
- ☐ I don't own any
- ☐ Other - Write In (Required) *

6. Do you know what a virus is? Please give a brief description. *

7. Have you ever experienced a computer virus? If no / not sure please move to question 12. *

- ☐ Yes
- ☐ No
- ☐ Not sure

8. How did you get the virus and how did you know about it?

9. Did you have anti-virus protection?

- ☐ Yes
- ☐ No

10. How did you get rid of the virus? Please specify if you required professional help and if comfortable how much it cost.

11. How did the virus impact your device and did it require any loss of money?

12. How do you believe viruses get onto your device? Select where appropriate. *

- ☐ Public WiFi
- ☐ Illegal websites
- ☐ External storage
- ☐ Social Media
- ☐ Emails
- ☐ Other - Write In (Required) *

13. How would you feel if you had a computer virus? *

- ☐ Worried
- ☐ Scared
- ☐ Anxious
- ☐ Fear
- ☐ Confusion
- ☐ Other - Write In (Required) *

14. Would you know what to do if you thought you had a virus on your device? If so, what? *

15. Do you know the dangers/risks a virus has on your device/life, if so, please list them? If not please write N/A. *

16. Have you ever been educated or made aware of computer viruses; i.e. what they are, their dangers, how to get rid of them and signs of them. If yes, please state where.

17. Do you use public Wi-Fi? If no please move onto question 27. *

- ☐ Yes
- ☐ No

18. Where do you use public wifi?

- ☐ Hotel
- ☐ Cafes
- ☐ Public Transport
- ☐ Entertainment Centres
- ☐ Airport
- ☐ Other - Write In (Required)

19. What do you use public wifi for? Please select where necessary.

- ☐ Banking
- ☐ Shopping
- ☐ Social Media
- ☐ Other - Write In (Required) *

20. Do you think public wifi is safe or do you believe there are risks when using it?

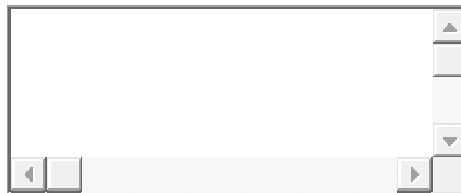
21. If you were aware of public wifi dangers would you still use it and why?



22. Have you ever been educated on the risks of public wifi? If yes, please state where.



23. What do you believe makes a site/link safe?



24. What social media platforms do you use? Please select as appropriate. *

- ☐ Facebook
- ☐ LinkedIn
- ☐ Twitter
- ☐ Instagram
- ☐ Other - Write In (Required) *

25. How long do you roughly spend on social media each day? *

- ☐ 0 hours
- ☐ 1-2 hours
- ☐ 2-3 hours
- ☐ 3-4 hours
- ☐ 5+ hours
- ☐ Other - Write In (Required)

9.4. Appendix 4: Online quiz

Malicious Links and Emails Quiz

Name of researcher: Ellen Gault

Researcher contact email address: 17001172@hope.ac.uk

What was the purpose of the project?

The purpose of this study is to systemically analyse the impact computer viruses have on devices and the awareness users have to them, whilst thoroughly looking at the link between malware passed through social media.

What will happen to the information I have provided?

The information provided will be used in a dissertation study at Liverpool Hope University.

How will this benefit me?

It will benefit yourself, by increasing your awareness on the particular topic, making you more conscious of links, attachments and social media accounts.

Was I deceived in any way? If so, why was I required to be naive?

You were not deceived in any way; the true aims of the study were stated.

If I change my mind and wish to withdraw the information I have provided, how do I do this?

If this is the case please remember your 4 digits of your mobile number and email 17001172@hope.ac.uk asking to withdraw stating these 4 digits.

Enter the last four digits of your phone number:

Links

Which of the following links looks malicious?

- ☐ www.mimecast.com/content/malicious-email-attachments
- ☐ bit.ly/email/

Why do you think so?

- ☐ <https://www.santander.co.uk/>
- ☐ <http://www.santander.co.uk//>

Why do you think so?

- ☐ <https://uk.tommy.com/>
- ☐ <http://www.tommyhilfiger.com>

Why do you think so?

- ☐ www.currys.co.uk/gbuk/index.html
- ☐ www.currys?pcworld.html

Why do you think so?

- ☐ www.techwalla.xml
- ☐ www.techwalla.html/pdf

Why do you think so?

Emails

Please select the following images which look like a scam.

- ☐

From : **Microsoft Customer Service Team** <microsoft@gmail.com>

To :

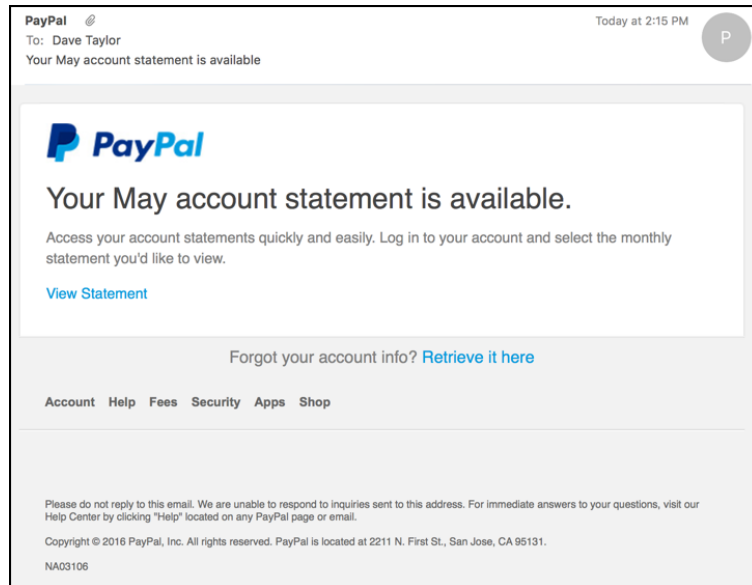
Subject :
- ☐

From : **Microsoft Customer Service Team** <customerservice@microsoft.com>

To :

Subject :

Why do you think so?



ID: #R56L6D9R406MH

Your account has been limited.

Dear customer.

We've limited your access and the reason is the last login attempt, we've limited your account for security reasons.

To fix this problem you have to login and update your personal information by following this link.

[Secure your account](#)

Copyrights Reserved 1999 - 2017

[Help & Contact](#) - [Fees](#) - [Security](#) - [Features](#)

Why do you think so?

Nationwide online@e-authorize.co.uk via home.pl
to



Attn Nationwide Client,

As part of our ongoing commitment to provide the Best Possible service and protection to all our customers, we require every member to validate their Internet Banking Profile using our safe SSL servers.


You are required to comply with immediate effect.

[Validate your Online Access](#)

Note You are advised to adhere strictly to this feature to avoid future service denial.

It is all about keeping you safe Online.



 Invoice

APPLE ID


BILL TO

INVOICE DATE
30 Dec 2019

SEQUENCE NO.
1-3235319425

ORDER ID
M/23JBHWYS

DOCUMENT NO.
180311157015

iCloud	PRICE
 iCloud: 50 GB Storage Plan Monthly Renews 28 Jan 2020	£0.79

Inclusive of VAT at 20%

Subtotal £0.66

VAT charged at 20% £0.13

TOTAL £0.79

If you have any questions about your bill, [contact support](#). This email confirms payment for the iCloud storage plan listed above. You will be charged each plan period until you cancel by [downgrading](#) to the free storage plan from your iOS device, Mac or PC.

Apple for a full refund within 15 days of a monthly subscription upgrade or within 45 days of a yearly refund are available where required by law.

Why do you think so?

Submit your answers

9.5 Appendix 5: Quiz coding

create_database.php

```
<?php
$conn = mysqli_connect("sql204.epizy.com", "epiz_25149807", "pr6N7yfbqMv");
// $sql = "CREATE DATABASE quizzes";
// mysqli_query($conn, $sql) or die(mysqli_error($conn));

$sql = "USE epiz_25149807_quiz";
mysqli_query($conn, $sql) or die(mysqli_error($conn));

$sql = "CREATE TABLE answers(
    last4Tel VARCHAR(4) PRIMARY KEY NOT NULL,
    links1 VARCHAR(10),
    why1 VARCHAR(255),
    links2 VARCHAR(10),
    why2 VARCHAR(255),
    links3 VARCHAR(10),
    why3 VARCHAR(255),
    links4 VARCHAR(10),
    why4 VARCHAR(255),
    links5 VARCHAR(10),
    why5 VARCHAR(255),
    image1 VARCHAR(10),
    why6 VARCHAR(255),
    image2 VARCHAR(10),
    why7 VARCHAR(255),
    image3 VARCHAR(10),
    why8 VARCHAR(255)
);";

mysqli_query($conn, $sql) or die(mysqli_error($conn));

mysqli_close($conn);
?>
```

process_data.php

```
<?php
$last4Tel = $_POST["last4Tel"];
$links1 = $_POST["links1"];
$why1 = $_POST["why1"];
```

```

$links2 = $_POST["links2"];
$why1 = $_POST["why2"];
$links3 = $_POST["links3"];
$why1 = $_POST["why3"];
$links4 = $_POST["links4"];
$why1 = $_POST["why4"];
$links5 = $_POST["links5"];
$why1 = $_POST["why5"];

$image1 = $_POST["image1"];
$why1 = $_POST["why6"];
$image2 = $_POST["image2"];
$why1 = $_POST["why7"];
$image3 = $_POST["image3"];
$why1 = $_POST["why8"];

$conn = mysqli_connect("localhost", "root", "root", "quiz");
$sql = "INSERT INTO answers('last4tel', 'links1', 'why1') VALUES('$last4tel',
'$links1', '$why1');"
if (mysqli_query($conn, $sql)) {
    echo "THANK FOR YOU FOR RESPONSES!";
}

mysqli_close($conn);
?>

```

test.html

```

<html>
<center>
    <head>
        <title> Malicious Links and Emails Quiz</title>
        <h1> Malicious Links and Emails Quiz </h1>
    </head>

    <body>
        <div>
            <form method="post" action="process_data.php">
                <p> Enter the last four digits of your phone number: </p> <input
type="text" id="last4Tel" name="last4Tel"/>
                <br/><br/>
                <h1> Links </h1>
                <p> Which of the following links looks malicious? </p>
                <br/><br/>

```



```

        <input type="radio" name="links1" value="url1"/> <a
href="https://www.mimecast.com/content/malicious-email-attachments/" onclick="return
false;">www.mimecast.com/content/malicious-email-attachments</a><br/>
        <input type="radio" name="links1" value="url2"/> <a href=""
onclick="return false;">bit.ly/email/</a><br/><br/>
        <p> Why do you think so? </p><br/><input type="text" id="why1"
name="why1"/><br/>
        <br/><br/>

```

```

        <input type="radio" name="links2" value="url3"/> <a
href="https://www.santander.co.uk/" onclick="return
false;">https://www.santander.co.uk/</a><br/>
        <input type="radio" name="links2" value="url4"/> <a
href="http://www.santander.co.uk/" onclick="return
false;">http://www.santander.co.uk/</a><br/>
        <p> Why do you think so? </p> <input type="text" id="why2"
name="why2"/><br/>
        <br/><br/>

```

```

        <input type="radio" name="links3" value="url5"/> <a
href="https://uk.tommy.com/" onclick="return false;">https://uk.tommy.com/</a><br/>
        <input type="radio" name="links3" value="url6"/> <a
href="http://www.tommyhilfiger.com" onclick="return
false;">http://www.tommyhilfiger.com</a><br/>
        <p> Why do you think so? </p> <input type="text" id="why3"
name="why3"/><br/>
        <br/><br/>

```

```

        <input type="radio" name="links4" value="url7"/> <a
href="www.currys.co.uk/gbuk/index.html" onclick="return
false;">www.currys.co.uk/gbuk/index.html</a><br/>
        <input type="radio" name="links4" value="url8"/> <a
href="www.currys?pcworld.html" onclick="return
false;">www.currys?pcworld.html</a><br/>
        <p> Why do you think so? </p><input type="text" id="why4"
name="why4"/><br/>
        <br/><br/>

```

```

        <input type="radio" name="links5" value="url9"/> <a
href="www.techwalla.xml" onclick="return false;">www.techwalla.xml</a><br/>
        <input type="radio" name="links5" value="url10"/> <a
href="www.techwalla.html/pdf" onclick="return false;">www.techwalla.html/pdf</a><br/>

```

```

        <p> Why do you think so? </p> <input type="text" id="why5"
name="why5"/><br/>
        <br/><br/>

        <h1> Emails </h1>

        <p> Please select the following images which look like a scam. </p>
        <br></br>

        <input type="radio" name="image1" value="image1"/>  <br/>
        <input type="radio" name="image1" value="image2"/> <br/>
        <p> Why do you think so? </p> <input type="text" id="why6"
name="why6"/><br/>
        <br/><br/>

        <input type="radio" name="image2" value="image3"/> <br/>
        <input type="radio" name="image2" value="image4"/> <br/>
        <p>Why do you think so? </p> <input type="text" id="why7"
name="why7"/><br/>
        <br/><br/>

        <input type="radio" name="image3" value="image5"/> <br/>
        <input type="radio" name="image3" value="image6"/> <br/>
        <p>Why do you think so?</p> <input type="text" id="why8"
name="why8"/><br/>
        <br/><br/>

        <input type="submit" value="Submit your answers"/>

    </form>
</div>

</center>

<style>
    input[type=text], select {
    width: 50%;
    padding: 5px 10px;

```

```

margin: 8px 0;
display: inline-block;
border: 1px solid #ccc;
border-radius: 4px;
box-sizing: border-box;
align-content: center;
}

input[type=submit] {
width: 100%;
background-color: #4CAF50;
color: white;
padding: 14px 20px;
margin: 8px 0;
border: none;
border-radius: 4px;
cursor: pointer;
}

input[type=submit]:hover {
background-color: #45a049;
}

div {
border-radius: 5px;
background-color: #f2f2f2;
padding: 20px;
}

body {
background-color: lightblue;
}

h1 {
color: black;
font-size: 15px;
font-family: lucida bright;
}

p {
font-family: lucida bright;
font-size: 15px;
}

```

```
input[type="radio"] {  
    font-family: lucida bright;  
    font-size: 15px;  
}
```

```
title {  
    color: black;  
    font-size: 15px;  
    font-family: lucida bright;  
}
```

```
</style>  
    </body>  
</html>
```

9.6 Appendix 6: Python coding

The 'text=' paragraph under each question displays the answers from all participants.

Question 1

```
from textblob import TextBlob
```

```
text = "too small its too short very short It's small  It's shorter Doesn't look like a real URL  
Doesn't have www. At the start  It says malicious in it  Not www. not sure  Has the word  
malicious in it  too short  Because it's short  very small  Doesn't have www. At the start  the  
link is not long enough  It says malicious in the link  No www or company name  It's vague  
too short  It says malicious in it  too small  It says malicious on the link too small its not real  
looking too small  very short to be a proper url  says malicious which is obvious its malicious  
too short  Says malicious  Because of the bit  too short  too short  too small Says malicious in  
it  Think it's too short its too short its too short says malicious in it  says malicious in it  says  
malicious in it  says malicious in it  shortened url says malicious in it  it has malicious in it  
its too short says malicious in it  Says malicious in it  bit.ly looks suspect shortened link.  
Companies will use their company name in website address Everyone knows mimecast are  
frauds too small says malicious  too short  Says malicious in it  the link is not long enough  It  
says malicious in it  says malicious in it  bit.ly looks suspect It says 'malicious' on the  
link It says 'malicious' on the link the link is not long enough  not sure  too small  its  
too short too short  it has malicious in it  too short  bit.ly looks suspect too short  its too short  
its not real looking too small  too short  Doesn't have www. At the start  Doesn't have www.  
At the start  too small says malicious in it  the link is not long enough"
```

```
text = text.lower()  
text = text.replace(".", "")  
text = text.replace("'", "")  
text = text.replace(",", "")  
text = text.replace("â€œ", "")  
text = text.replace("â€™", "")
```

```
text = text.replace(" at ", " ")  
text = text.replace(" in ", " ")  
text = text.replace(" its ", " ")
```

```
blob = TextBlob(text)
```

```
print(blob.words)  
print(blob.noun_phrases)  
print(blob.word_counts)
```

```
import collections
```

```

import re

words=re.findall(r'\w+', open('question1.txt', encoding="utf8").read().lower())
most_common = collections.Counter(words).most_common(10)
print(most_common)

import nltk

ngramlist=[]
raw=('question1.txt')

x=1
ngramlimit=3
tokens=nltk.word_tokenize(raw)

while x <= ngramlimit:
    ngramlist.extend(nltk.ngrams(tokens, x))
    x+=1

print(ngramlist)
import collections

```

```
import re
```

Question 2

```
from textblob import TextBlob
```

text = "too small its too short very short It's small It's shorter Doesn't look like a real URL Doesn't have www. At the start It says malicious in it Not www. not sure Has the word malicious in it too short Because it's short very small Doesn't have www. At the start the link is not long enough It says malicious in the link No www or company name It's vague too short It says malicious in it too small It says 'malicious' on the link too small its not real looking too small very short to be a proper url says malicious which is obvious its malicious too short Says malicious Because of the bit too short too short too small Says malicious in it Think it's too short its too short its too short says malicious in it says malicious in it says malicious in it says malicious in it shortened url says malicious in it it has malicious in it its too short says malicious in it Says malicious in it bit.ly looks suspect shortened link. Companies will use their company name in website address Everyone knows mimecast are frauds too small says malicious too short Says malicious in it the link is not long enough It says malicious in it says malicious in it bit.ly looks suspect It says 'malicious' on the link It says 'malicious' on the link the link is not long enough not sure too small its too short too short it has malicious in it too short bit.ly looks suspect too short its too short its not real looking too small too short Doesn't have www. At the start Doesn't have www. At the start too small says malicious in it the link is not long enough"

```

text = text.lower()
text = text.replace(".", "")
text = text.replace("'", "")
text = text.replace(",", "")
text = text.replace("â€œ", "")
text = text.replace("â€\x9d", "")

text = text.replace(" at ", " ")
text = text.replace(" in ", " ")
text = text.replace(" its ", " ")

blob = TextBlob(text)

print(blob.words)
print(blob.noun_phrases)
print(blob.word_counts)

import collections

import re

words=re.findall(r'\w+', open('question2.txt', encoding="utf8").read().lower())
most_common = collections.Counter(words).most_common(10)
print(most_common)


import nltk
ngramlist=[]
raw=('question2.txt')

x=1
ngramlimit=3
tokens=nltk.word_tokenize(raw)

while x <= ngramlimit:
    ngramlist.extend(nltk.ngrams(tokens, x))
    x+=1

print(ngramlist)

```

Question 3

```

from textblob import TextBlob

```

text="there is no www because there is no www no www No www It's only http Looks so the least real No 'https' at start and no / at end Doesn't say the whole brand name Not https should be http www. tommyhilfiger .com here is no https No www or Hilfiger no www It hasn't go the full name of the company Has http and www. In the same link because they haven't put www. The UK coming before the brand name No www heading It doesn't have UK and it's a world wide website no www No www or Hilfiger theres no www or hilfiger Same reason than in the previous one no www or hilfiger no www no www or hilfiger there is no www. or hilfiger no www or hilfiger No www or Hilfiger Not the right one no www or hilfiger no www or hilfiger no www or hilfiger No www or Hilfiger Should have the full name no www or hilfiger because there is no www no www or hilfiger no www or hilfiger no www or hilfiger no www or hilfiger prefixed uk at the start doesnt say www or hilfiger should be www or hilfiger no www or hilfiger no tommy or hilfiger No www or Hilfiger not https Again this is https so it is a secure site. Quick google check also shows it is the official site name No such thing there is no www there is no www. or hilfiger no www or hilfiger No www or Hilfiger because they haven't put www. No www or Hilfiger no www or hilfiger not https Same reason than in the previous one no s again because they haven't put www. here is no https theres no www or hilfiger no www or hilfiger no www should be www or hilfiger no www or hilfiger not https no www or hilfiger because there is no www no www no www No 'https' at start and no / at end No 'https' at start and no / at end no www or hilfiger no tommy or hilfiger because they haven't put www"

```
text = text.lower()
text = text.replace(".", "")
text = text.replace("'", "")
text = text.replace(",", "")
text = text.replace("â€œ", "")
text = text.replace("â€\x9d", "")
```

```
text = text.replace(" at ", " ")
text = text.replace(" in ", " ")
text = text.replace(" its ", " ")
```

```
blob = TextBlob(text)
```

```
print(blob.words)
print(blob.noun_phrases)
print(blob.word_counts)
```

```
import collections
```

```
import re
```

```
words=re.findall(r'\w+', open('question3.txt', encoding="utf8").read().lower())
most_common = collections.Counter(words).most_common(10)
print(most_common)
```



```

import nltk
ngramlist=[]
raw=('question3.txt')

x=1
ngramlimit=3
tokens=nltk.word_tokenize(raw)

while x <= ngramlimit:
    ngramlist.extend(nltk.ngrams(tokens, x))
    x+=1

print(ngramlist)

```

Question4

```

from textblob import TextBlob

```

```

text="there is a index.html there is index? theres ? Question mark There's a ? Never seen
gbuk before Question mark in just and no .com or .co.uk The question mark Not a secure site
it just doesnt look right Index looks weird gbuk looks As it has a question mark in it the
question mark Has a question mark because there is a question mark The ? in the middle of
the website link The question mark The question mark index.html is weird Index . Html is
weird index.html is strange Because of the æœ?œ• In the main part of the link .index is
weird not sure what index means theres no .com or .co.uk index.html is strange and never
seen before .index is weird .index is strange Had a question mark in index doesnt look or
sound right for a web address there is index? full currys name is not in it I don't know what
index means Shouldn't have ? In it there is index? index.html never seen before .index is
not normal no .com or uk index.html should be .com .html should be .com ? inbetween
currys & pc world index isnt right for a url index looks weird dont think it should be there
question mark is weird index.html is strange should be .com or .uk The index is weird I've
never seen that not sure what it means ? not / The url for this site isnt a full URL address so
more likely to be malicious Why's it asking a question? there is a index.html index.html is
strange and never seen before .index is weird I don't know what index means because there
is a question mark Index . Html is weird index.html should be .com ? not / Because of the ?
Because of the ? because there is a question mark it just doesnt look right index.html is not
right there is index? index.html is weird index looks weird dont think it should be there
.index is weird ? not / .index is weird index.html never seen before not sure what index
means gbuk looks Question mark in just and no .com or .co.uk Question mark in just and no
.com or .co.uk .index is weird index.html is strange should be .com or .uk because there is a
question mark "

```

```

text = text.lower()
text = text.replace(".", "")
text = text.replace(" ", "")

```

```

text = text.replace(", ", " ")
text = text.replace("œ", " ")
text = text.replace("€\x9d", " ")

text = text.replace(" at ", " ")
text = text.replace(" in ", " ")
text = text.replace(" its ", " ")

blob = TextBlob(text)

print(blob.words)
print(blob.noun_phrases)
print(blob.word_counts)

import collections

import re

words=re.findall(r'\w+', open('question4.txt', encoding="utf8").read().lower())
most_common = collections.Counter(words).most_common(10)
print(most_common)

import nltk
ngramlist=[]
raw=('question4.txt')

x=1
ngramlimit=3
tokens=nltk.word_tokenize(raw)

while x <= ngramlimit:
    ngramlist.extend(nltk.ngrams(tokens, x))
    x+=1

print(ngramlist)

```

Question 5

```
from textblob import TextBlob
```

text="there is a .xml not sure what the xml means but think pdf is ok no dot .xml this isn't right I have no idea hahaha PDF part at the end Weird ending Not a PDF Not a full email address Doesn't sound right .xml is weird .xml never seen before It has pdf in it the xml part Because of xml the link is shorter Haven't heard of xml before I think xml might be a programme extension and pdf is a document PDFs are usually a safe format .xml doesnt look right / pdf looks strange I've never seen it .xml looks very weird Because I know what pdf

is but no xml never seen .xml normally com or co uk no idea they both look wrong .mls should be .com ? i have no idea what .xml means .xml isnt right and looks strange .xml is weird Think it's a dodgy one have no idea what xml is but pdf is right .xml havent heard of its meant to be .com .xml i dont think is the end of a correct url Xml isn't real it should be com or co uk Should have / in it .xml is meant to be .com .xml .xml is strange it should be .com not .xml meant to be co . uk .xml means malicious it should be com or co uk at the end .xml should be com or co uk should be .com or .co.uk should be .com .xml looks suspicious never seen that for a website I dont know what .xml is. PDF is secure there is a .xml i have no idea what .xml means .xml isnt right and looks strange Xml isn't real it should be com or co uk the link is shorter / pdf looks strange I've never seen it meant to be co . uk I dont know what .xml is. Because I know what pdf is but no xml xml is weird the link is shorter Doesn't sound right .xml looks very weird .xml is meant to be .com .xml doesnt look right .xml should be com or co uk .xml isnt right and looks strange I dont know what .xml is. .xml isnt right and looks strange .xml no idea they both look wrong .xml never seen before Weird ending Weird ending never seen .xml normally com or co uk should be .com the link is shorter "

```
text = text.lower()
text = text.replace(".", "")
text = text.replace("'", "")
text = text.replace(",", "")
text = text.replace("œ", "")
text = text.replace("€\x9d", "")
```

```
text = text.replace(" at ", " ")
text = text.replace(" in ", " ")
text = text.replace(" its ", " ")
```

```
blob = TextBlob(text)
```

```
print(blob.words)
print(blob.noun_phrases)
print(blob.word_counts)
```

```
import collections
```

```
import re
```

```
words=re.findall(r'\w+', open('question5.txt', encoding="utf8").read().lower())
most_common = collections.Counter(words).most_common(10)
print(most_common)
```

```
import nltk
ngramlist=[]
raw=('question5.txt')
```

```

x=1
ngramlimit=3
tokens=nltk.word_tokenize(raw)

while x <= ngramlimit:
    ngramlist.extend(nltk.ngrams(tokens, x))
    x+=1

print(ngramlist)

```

Question 6

```
from textblob import TextBlob
```

```

text="@microsoft doesnt sound right customer service is not an appropriate email it wouldnt
be @ gmail @ gmail wouldnt be real for a big company It's a weird email Microsoft would
never have a gmail account Doesn't customer services In the email The email doesn't say
customer service Not a full email address Would think Microsoft would have own Just looks
fake @ microsoft isnt a proper email domain As it doesn't say it's customer service it's
@gmail Looks like a normal email because they do not look like someone you can contact
but you can contact customer service Microsoft wouldn't have a gmail account Use of a
gmail address Microsoft would use a Microsoft email account microsoft isnt a right @
Microsoft isn't real email @microsoft isnt real A company normally doesn't use the
'@gmail.com' part wouldnt be gmail customerservice isnt real name customerservice
might not be a real name microsoft wouldnt be the name should be a proper email address
@ Microsoft isn't a email address microsoft isnt right @ microsoft isnt a email account
microsoft isnt a right @ @ Microsoft isn't a real email company Should have customer
service in the email wouldnt be gmail they wouldnt have gmail @ microsoft isnt a email
account doesnt say customer service microsoft isnt a right @ microsoft isnt a right email
address email address ends in gmail should say customer service should say customer service
they wouldnt have a gmail account for such a big company should say customer service
Should say customer service in the name gmail account for microsoft, dont think so! Gmail
account is from someones personal account not company account. Microsoft have their own
domain It's not @microsoft @microsoft doesnt sound right microsoft wouldnt be the name
should be a proper email address @ @ Microsoft isn't a real email company customer
service email isnt right Microsoft isn't real email microsoft isnt a right @ gmail account is
strange a company doesn't use gmail a company doesn't use gmail customer service email
isnt right customer service isnt real name @microsoft isnt real wouldnt be gmail microsoft
isnt a right @ should say customer service should be a proper email address @ gmail
account is strange should be a proper email address @ they wouldnt have gmail
customerservice isnt real name @ microsoft isnt a proper email domain Doesn't customer
services In the email Doesn't customer services In the email wouldnt be gmail should say
customer service because they do not look like someone you can contact but you can contact
customer service "

```

```
text = text.lower()
```

```

text = text.replace(".", "")
text = text.replace("'", "")
text = text.replace(",", "")
text = text.replace("â€œ", "")
text = text.replace("â€\x9d", "")

text = text.replace(" at ", " ")
text = text.replace(" in ", " ")
text = text.replace(" its ", " ")

blob = TextBlob(text)

print(blob.words)
print(blob.noun_phrases)
print(blob.word_counts)

import collections

import re

words=re.findall(r'\w+', open('question6.txt', encoding="utf8").read().lower())
most_common = collections.Counter(words).most_common(10)
print(most_common)

import nltk
ngramlist=[]
raw=('question6.txt')

x=1
ngramlimit=3
tokens=nltk.word_tokenize(raw)

while x <= ngramlimit:
    ngramlist.extend(nltk.ngrams(tokens, x))
    x+=1

print(ngramlist)

```

Question 7

```
from textblob import TextBlob
```

text="not sure spelling mistake i dont have paypal i dont know Don't think they provide statements It's trying to make you put in ur details Secure account could be a dodgy link Asks for personal information Doesn't give a clear reason They want personal information"

mmmmk Asking details again Dear customer is weird it says customer It gave you a link to update login The PayPal logo isn't right looks like they are wanting me to sign up for something that i may not want to do Asking to enter personal details Second one requires clicking on a link whilst first asks you to log into the site normally It contains less information about PayPal also the other email says don't reply looks weird Doesn't look like a serious enough email for it to tell the customer their account is limited i have never had a statement from paypal before They say "access quickly" to motivate people to click on the link it doesnt say persons name ive never used paypal so unsure if these are real or not statements is normal from bank but not sure about paypal attempt should be attempt because its a statement Statements are strange for PayPal paypal statement looks fake ive never had paypal statements statements arent right with paypal Statements are weird for PayPal Shouldn't be able to retrieve your account details without private passwords etc because its a statement ive never had paypal statements because its a statement and theyre only for banks statements are strange i think ive never had paypal statements because its a statement Phishing attempt to enter log in details statements are strange statements are not normal for paypal theres spelling mistakes - attempt and grammar because its a statement Statement for PayPal might be malicious or suspicious as only banks do this and it's not a bank looks like phishing Spelling mistakes in email. Bad punctuation not sure attempt should be attempt its a statement Statements are weird for PayPal overall looks like a scam Doesn't look like a serious enough email for it to tell the customer their account is limited ive never had paypal statements looks like phishing the quick access is very intriguing the quick access is very daunting looks like a scam Asking details again i have never had a statement from paypal before because its a statement looks weird statements are not normal for paypal its a statement looks like phishing its a statement ive never had paypal statements ive never used paypal so unsure if these are real or not it says customer Asks for personal information Asks for personal information it doesnt say persons name because its a statement looks like they are wanting me to sign up for something that i may not want to do "

```
text = text.lower()
text = text.replace(".", "")
text = text.replace("'", "")
text = text.replace(",", "")
text = text.replace("â€œ", "")
text = text.replace("â€\x9d", "")
```

```
text = text.replace(" at ", " ")
text = text.replace(" in ", " ")
text = text.replace(" its ", " ")
```

```
blob = TextBlob(text)
```

```
print(blob.words)
print(blob.noun_phrases)
print(blob.word_counts)
```

```
import collections
```

```

import re

words=re.findall(r'\w+', open('question7.txt', encoding="utf8").read().lower())
most_common = collections.Counter(words).most_common(10)
print(most_common)

import nltk
ngramlist=[]
raw=('question7.txt')

x=1
ngramlimit=3
tokens=nltk.word_tokenize(raw)

while x <= ngramlimit:
    ngramlist.extend(nltk.ngrams(tokens, x))
    x+=1

print(ngramlist)

```

Question 8

```
from textblob import TextBlob
```

text="looks fake not sure how tho the border the font doesnt look real Spelling mistakes It's a weird font Email addresses of nationwide looks fake Email looks dodgy unsure They want personal information Looks fake Looks not sure never seen this before apple dont send invoices There's a time limited its about money Log in button instead of link they are sending a very negative message that someone will easily fall for because they may get scared and will quickly input their information It gives a time frame for you to complete it in Spelling mistakes There are spelling mistakes and 24 hours to validate an account isn't very long apple dont send invoices Looks not sure never seen this before not sure It says to you to do it before 24 hours doesnt say persons name the attn is weird the email address font and border is not sure never seen this before apple dont send invoices Email address looks weird Because of the email font is not sure never seen this before and border looks funny the border looks not sure never seen this before the font is not sure never seen this before looking It looks very not sure never seen this before Hasn't got a padlock anywhere to prove it's safe apple dont send invoices apple dont send invoices attn is weird looks not sure never seen this before and professional email address isnt legit doesnt say customer says attn email address looks suspicious the whole email looks not sure never seen this before not sure if im right but the email looks funny the attn is strange should be the customers name, and the second image gives time limit it says attn which doesnt look right and the email is weird Email address isn't real Email contains a typo Spelling mistakes It says via home looks fake not sure how tho font and border is not sure never seen this before apple dont send invoices It looks very not sure never seen this before its strange looking Looks not sure never seen

this before email address isnt legit Email has spelling mistakes time limit time limit to complete its strange with the font Looks fake not sure apple dont send invoices apple dont send invoices not sure if im right but the email looks funny apple dont send invoices Email has spelling mistakes apple dont send invoices apple dont send invoices the attn is weird apple dont send invoices Email looks weird Email looks dodgy doesnt say persons name it says attn which doesnt look right and the email is weird they are sending a very negative message that someone will easily fall for because they may get scared and will quickly input their information "

```
text = text.lower()
text = text.replace(".", "")
text = text.replace("'", "")
text = text.replace(",", "")
text = text.replace("its", "")
text = text.replace(" a ", "")
text = text.replace(" it ", "")
text = text.replace("â€œ", "")
text = text.replace("â€\x9d", "")
```

```
text = text.replace(" at ", " ")
text = text.replace(" in ", " ")
text = text.replace(" its ", " ")
```

```
blob = TextBlob(text)
```

```
print(blob.words)
print(blob.noun_phrases)
print(blob.word_counts)
```

```
import collections
```

```
import re
```

```
words=re.findall(r'\w+', open('question8.txt', encoding="utf8").read().lower())
most_common = collections.Counter(words).most_common(10)
print(most_common)
```

```
import nltk
ngramlist=[]
raw=('question8.txt')
```

```
x=1
ngramlimit=3
tokens=nltk.word_tokenize(raw)
```

```
while x <= ngramlimit:
```



```
    ngramlist.extend(nltk.ngrams(tokens, x))  
    x+=1  
  
print(ngramlist)
```

9.7 Appendix 7: Demographic answers

Q1	Q2	Q3	Q4
phone number	gender	age	occupation
1384	M	18-24	Sales assistant
5558	M	18-24	Computer science
5559	F	18-24	Primary ed
3358	F	18-24	Primary ed
7525	F	18-24	Early childhood
7618	F	25-34	Education
1998	F	35-44	Primary ed
2926	F	18-24	Student teacher
8603	F	25-34	Primary teacher
5267	F	18-24	Student
6016	M	35-44	Law
361	M	25-34	Business
7120	M	45-54	Engineering manager
9364	F	18-24	Student
7377	F	25-34	Teaching
9148	F	25-34	Volunteer
2072	F	18-24	Fine art
379	F	18-24	Forensic psychology
1443	F	45-54	Primary teaching
5832	F	45-54	Hairdresser
6450	F	25-34	Primary ed
4627	M	45-54	Builder
6597	M	35-44	Managerial
6666	F	35-44	Relationship counsellor
3496	F	35-44	Childmilder
6337	F	45-54	Housewife
9014	F	45-54	Post office manager
8902	F	45-54	Hairdresser
9878	M	25-34	Ambulance care assistant
7662	M	25-34	NA
9874	M	35-44	Lecturer
3248	M	35-44	Head of learning
4900	M	45-54	Teacher
6145	M	25-34	Trainer
5824	F	45-54	Lecturer
3674	F	18-24	Teacher
5151	F	18-24	Nursery
7012	M	45-54	Manager
202	F	18-24	Teacher
6642	F	35-44	Health care worker
7887	F	18-24	Teacher

8931	M	45-54	Administrative officer
2040	M	45-54	Airport worker
3628	F	45-54	Banker
2928	F	18-24	Shop assistant
6058	M	18-24	Airport worker /student
82	F	18-24	Customer assistant
6015	M	18-24	Computer science student
7379	M	55-64	Media and tv
303	M	45-54	Retail
7803	F	55-64	Pharmacy
734	M	45-54	Staff nurse
6163	N/A	18-24	Student
195	F	45-54	Retail
1201	F	18-24	Student
7621	M	55-64	Sales assistant
3261	M	55-64	Lecturer
6177	M	45-54	Practice manager
8884	F	45-54	Hospitality management
1633	M	64+	Granda
595	F	18-24	Hotel and events management bachelor student
2952	M	55-64	B&q
3359	F	18-24	Student law
4912	F	55-64	Tourism and business
5098	M	35-44	Chef
3838	M	55-64	Psychology
5985	F	18-24	Study economics
3384	F	64+	Geology
7749	M	55-64	Head of it
588	M	64+	Systems manager
9780	M	35-44	Systems administrator
4750	M	55-64	It technician
6594	M	35-44	Grounds maintenance worker
7761	F	25-34	Vet
4976	M	35-44	Consultant
5922	M	64+	Dog walker
6639	F	18-24	Student

9.8 Appendix 8: Answers from question 2 on quiz.

phone number	q2		reason 2
1384	url2	wrong	there is no s
5558	url3	wrong	im not sure
5559	url4	wrong	no s
3358	url4	wrong	No s
7525	url4	wrong	It is only http
7618	url4	wrong	2 // at end
			Doesn't have https at start, the s at the end means
1998	url4	wrong	safe
2926	url3	wrong	Doesn't have the s
8603	url3	wrong	Not https should be http
5267	url4	wrong	theres two /
6016	url4	wrong	Not sure
361	url3	wrong	the second url looks quite long
7120	url3	wrong	As it has one less /
9364	url4	wrong	the number of dashes
7377	url4	wrong	Has two / at the end
9148	url4	wrong	there are 2 dashes at the end of this one
2072	url4	wrong	Has two // at the end
379	url4	wrong	Extra back slash
1443	url3	wrong	It doesn't have two forward slashes
5832	url4	wrong	two //
6450	url3	wrong	there's an s which shouldn't be there
4627	url3	wrong	there shouldnt be a dash at the end
6597	url4	wrong	Https doesn't have an "s"
6666	url4	wrong	theres no s
3496	url3	wrong	theres an s which doesnt look right
6337	url3	wrong	dont know the difference between the two
9014	url4	wrong	no s
8902	url3	wrong	too // after https
9878	url3	wrong	There's no two / at the end and there should be
7662	url4	wrong	Wrong address
			no s which im pretty sure means secure or
9874	url4	wrong	something
3248	url3	wrong	think after https there should only be one dash
4900	url3	wrong	there shouldnt be 2 / before www
6145	url4	wrong	No s
5824	url4	wrong	Should have an s after http
3674	url4	wrong	theres no s
5151	url4	wrong	theres no s
7012	url3	wrong	there should be 2 // at the end

202	url3	wrong	im not sure
6642	url4	wrong	theres no s
7887	url3	wrong	santander is spelt wrong
8931	url4	wrong	No secure connection
2040	url3	wrong	should be two // at the end
3628	url3	wrong	should be two dashes at the end
2928	url4	wrong	no s
6058	url3	wrong	there should be two // at the end
82	url3	wrong	Should be two // at the end
6015	url4	wrong	should only be one /
7379	url4	wrong	always use https not http. As you know it is secure
303	url3	wrong	No such thing
7803	url2	wrong	there is no s
734	url4	wrong	no s
6163	url3	wrong	too // after https
195	url4	wrong	No s
1201	url4	wrong	there are 2 dashes
7621	url3	wrong	there's an s which shouldn't be there
3261	url4	wrong	theres no s
6177	url4	wrong	should only be one /
8884	url4	wrong	Https doesn't have an "œ"•
1633	url4	wrong	no s
595	url4	wrong	there are 2 dashes
2952	url4	wrong	theres two /
3359	url3	wrong	there shouldnt be a dash at the end
4912	url4	wrong	theres no s
5098	url4	wrong	two //
3838	url3	wrong	should be two dashes at the end
5985	url3	wrong	too // after https
3384	url4	wrong	should only be one /
7749	url3	wrong	too // after https
588	url4	wrong	theres no s
9780	url3	wrong	theres an s which doesnt look right
4750	url3	wrong	the second url looks quite long
6594	url4	wrong	Doesn't have https at start, the s at the end means safe
7761	url4	wrong	Doesn't have https at start, the s at the end means safe
4976	url4	wrong	theres no s
5922	url3	wrong	there should be two // at the end
6639	url4	wrong	there are 2 dashes at the end of this one

9.9 Appendix 9: Answers from question 6 on quiz.

phone number	q6	reason 6
1384	image2 wrong	@microsoft doesnt sound right
5558	image2 wrong	customer service is not an appropriate email
5559	image1 wrong	it wouldnt be @ gmail
3358	image1 wrong	@ gmail wouldn't be real for a big company
7525	image1 wrong	It's a weird email
7618	image1 wrong	Microsoft would never have a gmail account
1998	image1 wrong	Doesn't customer services In the email
2926	image1 wrong	The email doesn't say customer service
8603	image1 wrong	Not a full email address
5267	image1 wrong	Would think Microsoft would have own
6016	image2 wrong	Just looks fake
361	image2 wrong	@ microsoft isnt a proper email domain
7120	image1 wrong	As it doesn't say it's customer service
9364	image1 wrong	it's @gmail
7377	image1 wrong	Looks like a normal email
9148	image1 wrong	contact customer service
2072	image1 wrong	Microsoft wouldn't have a gmail account
379	image1 wrong	Use of a gmail address
1443	image1 wrong	Microsoft would use a Microsoft email account
5832	image2 wrong	microsoft isnt a right @
6450	image2 wrong	Microsoft isn't real email
4627	image2 wrong	@microsoft isnt real
6597	image1 wrong	A company normally doesn't use the "ægmail.com" part
6666	image1 wrong	wouldnt be gmail
3496	image2 wrong	customerservice isnt real name
6337	image2 wrong	customerservice might not be a real name
9014	image1 wrong	microsoft wouldnt be the name
8902	image2 wrong	should be a proper email address @
9878	image2 wrong	Microsoft isn't a email address
7662	image1 wrong	
9874	image2 wrong	microsoft isnt right
3248	image2 wrong	@ microsoft isnt a email account
4900	image2 wrong	microsoft isnt a right @
6145	image2 wrong	@ Microsoft isn't a real email company
5824	image1 wrong	Should have customer service in the email
3674	image1 wrong	wouldnt be gmail
5151	image1 wrong	they wouldnt have gmail
7012	image2 wrong	@ microsoft isnt a email account
202	image1 wrong	doesnt say customer service
6642	image2 wrong	microsoft isnt a right @
7887	image2 wrong	microsoft isnt a right email address
8931	image1 wrong	email address ends in gmail

2040	image1	wrong	should say customer service
3628	image1	wrong	should say customer service
2928	image1	wrong	they wouldnt have a gmail account for such a big company
6058	image1	wrong	should say customer service
82	image1	wrong	Should say customer service in the name
6015	image1	wrong	gmail account for microsoft, dont think so!
7379	image1	wrong	Gmail account is from someones personal account
303	image1	wrong	It's not @microsoft
7803	image2	wrong	@microsoft doesnt sound right
734	image1	wrong	microsoft wouldnt be the name
6163	image2	wrong	should be a proper email address @
195	image2	wrong	@ Microsoft isn't a real email company
1201	image1	wrong	customer service email isnt right
7621	image2	wrong	Microsoft isn't real email
3261	image2	wrong	microsoft isnt a right @
6177	image1	wrong	gmail account is strange
8884	image1	wrong	a company doesn't use gmail
1633	image1	wrong	a company doesn't use gmail
595	image1	wrong	customer service email isnt right
2952	image2	wrong	customer service isnt real name
3359	image2	wrong	@microsoft isnt real
4912	image1	wrong	wouldnt be gmail
5098	image2	wrong	microsoft isnt a right @
3838	image1	wrong	should say customer service
5985	image2	wrong	should be a proper email address @
3384	image1	wrong	gmail account is strange
7749	image2	wrong	should be a proper email address @
588	image1	wrong	they wouldnt have gmail
9780	image2	wrong	customerservice isnt real name
4750	image2	wrong	@ microsoft isnt a proper email domain
6594	image1	wrong	Doesn't customer services In the email
7761	image1	wrong	Doesn't customer services In the email
4976	image1	wrong	wouldnt be gmail
5922	image1	wrong	should say customer service
6639	image1	wrong	do not look like someone you can contact

9.10 Appendix 10: Answers from question 1 on quiz.

phone number	q1	reason 1
1384	url2	too small
5558	url2	its too short
5559	url2	very short
3358	url2	It's small
7525	url2	It's shorter
7618	url2	Doesn't look like a real URL
1998	url2	Doesn't have www. At the start
2926	url1	It says malicious in it
8603	url2	Not www.
5267	url2	not sure
6016	url1	Has the word malicious in it
361	url2	too short
7120	url2	Because it's short
9364	url2	very small
7377	url2	Doesn't have www. At the start
9148	url2	the link is not long enough
2072	url1	It says malicious in the link
379	url2	No www or company name
1443	url2	It's vague
5832	url2	too short
6450	url1	It says malicious in it
4627	url2	too small
6597	url1	It says 'malicious'• on the link
6666	url2	too small
3496	url2	its not real looking too small
6337	url2	very short to be a proper url
9014	url1	says malicious which is obvious its malicious
8902	url2	too short
9878	url1	Says malicious
7662	url2	Because of the bit
9874	url2	too short
3248	url2	too short
4900	url2	too small
6145	url1	Says malicious in it
5824	url2	Think it's too short
3674	url2	its too short
5151	url2	its too short
7012	url1	says malicious in it
202	url1	says malicious in it
6642	url1	says malicious in it
7887	url1	says malicious in it
8931	url2	shortened url

2040	url1	says malicious in it
3628	url1	it has malicious in it
2928	url2	its too short
6058	url1	says malicious in it
82	url1	Says malicious in it
6015	url2	bit.ly looks suspect
7379	url2	Companies will use their company name in website address
303	url1	Everyone knows mimecast are frauds
7803	url2	too small
734	url1	says malicious
6163	url2	too short
195	url1	Says malicious in it
1201	url2	the link is not long enough
7621	url1	It says malicious in it
3261	url1	says malicious in it
6177	url2	bit.ly looks suspect
8884	url1	It says 'malicious' on the link
1633	url1	It says 'malicious' on the link
595	url2	the link is not long enough
2952	url2	not sure
3359	url2	too small
4912	url2	its too short
5098	url2	too short
3838	url1	it has malicious in it
5985	url2	too short
3384	url2	bit.ly looks suspect
7749	url2	too short
588	url2	its too short
9780	url2	its not real looking too small
4750	url2	too short
6594	url2	Doesn't have www. At the start
7761	url2	Doesn't have www. At the start
4976	url2	too small
5922	url1	says malicious in it
6639	url2	the link is not long enough

9.11 Appendix 11: Answers from question 3 on quiz

phone number	q3	reason 3
1384	url1 wrong	there is no www
5558	url5 wrong	because there is no www
5559	url5 wrong	no www
3358	url5 wrong	No www
7525	url6 wrong	It's only http
7618	url5 wrong	Looks so the least real
1998	url6 wrong	No 'https' at start and no '/' at end
2926	url5 wrong	Doesn't say the whole brand name
8603	url5 wrong	Not https should be http www. tommyhilfiger .com
5267	url6 wrong	here is no https
6016	url5 wrong	No www or Hilfiger
361	url5 wrong	no www
7120	url5 wrong	It hasn't go the full name of the company
9364	right	
7377	url6 wrong	Has http and www. In the same link
9148	url5 wrong	because they haven't put www.
2072	url5 wrong	The UK coming before the brand name
379	url5 wrong	No www heading
1443	url6 wrong	It doesn't have UK and it's a world wide website
5832	url5 wrong	no www
6450	url5 wrong	No www or Hilfiger
4627	url5 wrong	theres no www or hilfiger
6597	url6 wrong	Same reason than in the previous one
6666	url5 wrong	no www or hilfiger
3496	url5 wrong	no www
6337	url5 wrong	no www or hilfiger
9014	url5 wrong	there is no www. or hilfiger
8902	url5 wrong	no www or hilfiger
9878	url5 wrong	No www or Hilfiger
7662	url5 wrong	Not the right one
9874	url5 wrong	no www or hilfiger
3248	url5 wrong	no www or hilfiger
4900	url5 wrong	no www or hilfiger
6145	url5 wrong	No www or Hilfiger
5824	url5 wrong	Should have the full name
3674	url5 wrong	no www or hilfiger
5151	url5 wrong	because there is no www
7012	url5 wrong	no www or hilfiger
202	url5 wrong	no www or hilfiger
6642	url5 wrong	no www or hilfiger
7887	url5 wrong	no www or hilfiger
8931	url5 wrong	prefixed uk at the start

2040	url5	wrong	doesnt say www or hilfiger
3628	url5	wrong	should be www or hilfiger
2928	url5	wrong	no www or hilfiger
6058	url5	wrong	no tommy or hilfiger
82	url5	wrong	No www or Hilfiger
6015	url6	wrong	not https
7379	url6	wrong	Again this is https so it is a secure site. Quick google check also shows it is the official site name
303	url5	wrong	No such thing
7803	url1	wrong	there is no www
734	url5	wrong	there is no www. or hilfiger
6163	url5	wrong	no www or hilfiger
195	url5	wrong	No www or Hilfiger
1201	url5	wrong	because they haven't put www.
7621	url5	wrong	No www or Hilfiger
3261	url5	wrong	no www or hilfiger
6177	url6	wrong	not https
8884	url6	wrong	Same reason than in the previous one
1633	url6	wrong	no s again
595	url5	wrong	because they haven't put www.
2952	url6	wrong	here is no https
3359	url5	wrong	theres no www or hilfiger
4912	url5	wrong	no www or hilfiger
5098	url5	wrong	no www
3838	url5	wrong	should be www or hilfiger
5985	url5	wrong	no www or hilfiger
3384	url6	wrong	not https
7749	url5	wrong	no www or hilfiger
588	url5	wrong	because there is no www
9780	url5	wrong	no www
4750	url5	wrong	no www
6594	url6	wrong	No 'https' at start and no / at end
7761	url6	wrong	No 'https' at start and no / at end
4976	url5	wrong	no www or hilfiger
5922	url5	wrong	no tommy or hilfiger
6639	url5	wrong	because they haven't put www.

9.12 Appendix 12: Answers from question 4 in quiz.

phone number	q4	reason 4
1384	url1 wrong	there is a index.html
5558	url7 wrong	there is index?
5559	url8 wrong	theres ?
3358	url8 wrong	Question mark
7525	url8 wrong	There's a ?
7618	url8 wrong	Never seen gbuk before
1998	url8 wrong	Question mark in just and no .com or .co.uk
2926	url8 wrong	The question mark
8603	url8 wrong	Not a secure site
5267	url9 wrong	it just doesnt look right
6016	url7 wrong	Index looks weird
361	url7 wrong	gbuk looks
7120	url8 wrong	As it has a question mark in it
9364	url8 wrong	the question mark
7377	url8 wrong	Has a question mark
9148	url8 wrong	because there is a question mark
2072	url8 wrong	The ? in the middle of the website link
379	url8 wrong	The question mark
1443	url8 wrong	The question mark
5832	url7 wrong	index.html is weird
6450	url7 wrong	Index . Html is weird
4627	url7 wrong	index.html is strange
6597	url8 wrong	Because of the " " In the main part of the link
6666	url7 wrong	.index is weird
3496	url7 wrong	not sure what index means
6337	url8 wrong	theres no .com or .co.uk
9014	url7 wrong	index.html is strange and never seen before
8902	url7 wrong	.index is weird
9878	url7 wrong	.index is strange
7662	url8 wrong	Had a question mark in
9874	url7 wrong	index doesnt look or sound right for a web address
3248	url7 wrong	there is index?
4900	url7 wrong	full currys name is not in it
6145	url7 wrong	I don't know what index means
5824	url8 wrong	Shouldn't have ? In it
3674	url7 wrong	there is index?
5151	url7 wrong	index.html never seen before
7012	url7 wrong	.index is not normal
202	url7 wrong	no .com or uk
6642	url7 wrong	index.html should be .com
7887	url7 wrong	.html shoud be .com
8931	url8 wrong	? inbetween currys & pc world

2040	url7	wrong	index isnt right for a url
3628	url7	wrong	index looks weird dont think it should be there
2928	url8	wrong	question mark is weird
6058	url7	wrong	index.html is strange should be .com or .uk
82	url7	wrong	The index is weird I've never seen that not sure what it means
6015	url8	wrong	? not /
7379	url8	wrong	The url for this site isnt a full URL address so more likely to be malicious
303	url8	wrong	Why's it asking a question?
7803	url1	wrong	there is a index.html
734	url7	wrong	index.html is strange and never seen before
6163	url7	wrong	.index is weird
195	url7	wrong	I don't know what index means
1201	url8	wrong	because there is a question mark
7621	url7	wrong	Index . Html is weird
3261	url7	wrong	index.html should be .com
6177	url8	wrong	? not /
8884	url8	wrong	Because of the ?
1633	url8	wrong	Because of the ?
595	url8	wrong	because there is a question mark
2952	url9	wrong	it just doesnt look right
3359	url7	wrong	index.html is not right
4912	url7	wrong	there is index?
5098	url7	wrong	index.html is weird
3838	url7	wrong	index looks weird dont think it should be there
5985	url7	wrong	.index is weird
3384	url8	wrong	? not /
7749	url7	wrong	.index is weird
588	url7	wrong	index.html never seen before
9780	url7	wrong	not sure what index means
4750	url7	wrong	gbuk looks
6594	url8	wrong	Question mark in just and no .com or .co.uk
7761	url8	wrong	Question mark in just and no .com or .co.uk
4976	url7	wrong	.index is weird
5922	url7	wrong	index.html is strange should be .com or .uk
6639	url8	wrong	because there is a question mark