NTNU – Trondheim
Norwegian University of
Science and Technology

CORONAVIRUS PANDEMIC INFLUENCE ON INTERNET USAGE AND PERFORMANCE IN NORWAY

ABRAHAM KWAME RICHMAN

Submission date: June 2021
Supervisor: Poul E. Heegaard, IIK
Co-supervisor: Poul E. Heegaard, IIK

NTNU – Norwegian University of Science and Technology
Department of Information Security and Communication Technology

Title: CORONAVIRUS PANDEMIC INFLUENCE ON INTERNET USAGE AND PER

Student: ABRAHAM KWAME RICHMAN

Problem description:

The rapid increase in the spread of the coronavirus since its arrival in Norway on the 26th of February, 2020 forced the Norwegian government to take preventive measures to address this pandemic issue. The nation was put under total lockdown which in turn forced a paradigm shift in the way people interact with each other and their immediate surroundings. The lockdown, basically, ensured all physical interactions were ceased and an alternative be taken to replace these physical interactions. This had a great impact on everything and the Internet is no exception.

So, with the lockdown in full effect, the people turned to the Internet and begun to explore various platforms that could allow multiple people interact with each other and at the same time, work, learn and entertain themselves. Videoconferencing platforms like Zoom, Microsoft Teams and Google duo that allowed people organize seminars, have meetings and lectures had an extreme increase in its number of users during this coronavirus pandemic than it has been recorded [1]. However, for entertainment, social media, streaming and gaming services like Facebook, Netflix, Twitch TV and YouTube also saw a spike in its daily customer usage [2].

With almost all work and activities moved online, how substantially the internet was used and performed during this time is yet to be fully researched. This paper sets out to address the impact the coronavirus pandemic has had on internet usage and performance since its arrival in Norway. We analyze the data provided by NKOM's Nettfart app (`nettfart.no`) which provide internet speed test results from mobile devices across the whole country of Norway and beyond for before and during the lockdown period.

The following tasks will be undertaken;

– Filter out data in order to eliminate minor errors and surplus data.

– Extract performance characteristics data.

– Study statistics of data and investigate traffic flow.

– Compare data from before and during the pandemic.

– Take a survey to know people's experience with Internet performance before and during the pandemic.

Date approved: 2021-02-18

Supervisor: Poul E. Heegaard, IIK

Abstract

Norway's response to the Coronavirus pandemic was to initiate a lockdown. With Lockdown measures in place, the population was advised to work from home and use the internet for office work, meetings, lectures, conferences, studying and leisure activities. This study aims to determine how much change the Internet experienced during the Coronavirus pandemic in Norway. It investigates the performance of the Internet exploring performance metrics speed and latency during the pandemic in comparison to the previous years. In this context, performance is simply described as the a network's ability to do what it was designed to do.

To test the hypothesis that the Internet experienced certain changes during the pandemic, a mixed method research was used to analyse mobile and web browser speed test results collected by Nettfart from the year 2012 and responses from an online survey distributed to the population who lived in Norway before and during the pandemic. Speed test results gave information about upload and download speed and latency of users' network. For the survey, respondents were asked to share their user experience with internet quality and rate the Internet performance before and during the pandemic and finally, why users performed speed tests. The results showed that the origin of traffic changed from business or office areas to more residential areas as hypothesized. Analysing data from the year 2012 to 2021, the study results revealed that average upload and download speed got better as the years went by whiles latency performance did not change significantly. However, there was a great increase in the number of speed tests following the lockdown in March 2020 and the study reveals it was as a result of network quality change.

These results indicates that during the pandemic year, users experienced Internet quality issues in one way or the other. These network quality issues were not as a result of internet speed or latency changes as the pandemic did not have any significant influence on these metrics.

Preface

Before you lies the master thesis report "Coronavirus Pandemic Influence On Internet Usage and Performance In Norway", focusing on the performance of the Internet during the pandemic in relation to speed and latency of users network. Aside my personal interest in this study, it has been written to fulfill the requirements to complete a two-year Master program in Communication Technology at NTNU's Department of Information Security and Communication Technology.

I would like to thank my supervisor, Poul Heegard for his magnificent guidance and support throughout the process and always being available and willing to answer my questions. I also wish to express gratitude to all the respondents of the survey, without whose cooperation I would not have been able to perform the analysis.

I hope you have a good read.

Abraham Kwame Richman
Trondheim, 16th June,2021

Contents

List of Figures

List of Tables

List of Algorithms

Listings

Chapter 1
INTRODUCTION

1.1 Background and Motivation

As a response to the rising number of coronavirus infections across the globe, most governments imposed a national lockdown which has been accompanied by a sharp increase in the use of the Internet. According to a research by Feldmann et al [9], there was 15-20% increase in Internet traffic volume as a result of the lockdown.

As the pandemic prevents people from having physical interactions, the population of Norway has been strongly advised to self-isolate and stay at home in order to slow down the spread of the coronavirus. Hence, thousands of Norwegians have turned to the Internet for communication, work, learning, leisure and pleasure, leading to a high demand for Internet services and presumably causing a number of changes. These changes include, Traffic load increase, application mix change (more video conferences and streaming), change in geographical origin of traffic data (most likely originating from residential areas rather than school and business areas) and traffic load variation change (day-to-day variations). These changes are a potential cause of congestion in various networks and poor Internet service quality (or poor speed). The Internet as we know it, will have to adjust to meet all the requirements to make the new digital life possible.

Even though there was a high demand for internet services across the globe, the Internet managed to adjust itself to meet the requirements to make the digital shift possible all because of its resilient and redundant nature from its design [10]. This, however, grants us the opportunity to study and investigate how the Internet was used and also how it performed during these times, explicitly in Norway.

1.2 Aim and Objectives

This master thesis aims to analyse and investigate to what extent the coronavirus pandemic has affected the performance of the Internet in Norway based on speed test results across the nation. With the tasks stated in the problem description, they aim to help achieve the following research objectives.

- Perform a qualitative research using the survey.
- Filter out data based on the desired variables from the data set.
- Analyze and quantify Internet performance metrics for before and during pandemic.
- Compare results for before and during the pandemic.
- Establish statistical conclusions based on statistical Inference

1.3 Research Questions

With the data set provided by NKOM's nettfart as well as the conducted survey, they should be sufficient to help answer the following research questions.

1. **To what extent has the users experience changed in user quality compared to before?**
2. **Did the average Internet speed before and during the pandemic change?**
3. **Does the internet speed change over the day?**

1.4 Limitations

Considering the available data set from Nettfart, there are surely some limitations that needs to be recognized. First and most importantly, the data set provided is limited to information about speed, geographical information, network and operators and therefore provides no information about the traffic mix or user behaviour. Therefore we can not study the Usage aspect of this thesis and limits us to performance only.

Secondly, the data set available was collected from mobile devices only which were on either cellular network (4G, 3G, 2G) or WLAN. As a result, there are several factors the customer and the provider (cellular or WLAN) have no control over but may influence the speed test results. So, when performing the test with NETTFART, factors like the browser being used, whether measurements is via wireless router, user device, user operating system, geographical location, distance to base station (for cellular networks), weather conditions and even the thickness of the wall may influence test results [11]. According to Ookla[1] [12], the physical distance to test server can influence test results. To overcome this limitation for a better overall analysis, considering benefactors that has been measured several times is a better option. However, BEREC[2] provides guidelines for accurate and reliable measurement results when you consider a user initiated measurement with possible environmental factors that could lead to measurement errors [13].

Thirdly, as mentioned in the specialisation project report [6], the data set is speculated to be bias since possible times users would like to access the app was when users finds themselves in a new location out of curiosity or when most especially the network is having challenges or is slow. With a survey answered by 70 people randomly across Norway, we can affirm that speculation.

Also, according to nettfart.no, whenever there is a speed greater than 100Mbits/s, the uncertainties in measurements results are higher. They attributed it to the fact that the current version of Nettfart.no does not contain contain a subscription with a higher download capacity than 100Mbits/s. This is relevant because the data set provided for this master thesis has several recorded speed greater than 100Mbits/s. For cellular networks that have speed greater than 100Mbits/s, they will be filtered out of the data set

Finally, With a data set that spans a over a year, we cannot ignore the possibility providers of the networks may have improved the network in a way, either hardware or software. However, there is also a possibility that these improvements may have caused under-performance.

1.5 Structure

This section briefly describes the structure of the remaining of this report. In chapter 2, we give an overview of our methodology on a high level and a summary of several related work concerning the internet and the Coronavirus pandemic; Chapter 3 discusses the detailed methods used for this research explaining the how and what was done in order to help assess the authenticity and rationality of this research; Chapter 4 reports the results of the research pertinent to the research questions and objectives listed in sections 1.3 and 1.2 respectively; Chapter 5 presents the meaning of the results and what they may signify; Chapter 6 concludes the thesis providing a clear understanding of the central argument.

[1] https://www.ookla.com/
[2] Body of European Regulators for Electronic Communications https://berec.europa.eu/

Chapter 2
LITERATURE AND RELATED WORK

2.1 Introduction

In order to know the impact of the pandemic on Internet performance, understanding what network performance entails is crucial therefore creates the need to study the literature. Although the focus of this thesis is on speed testing during the pandemic, other performance metrics and measurement methodologies will be discussed as well. Also, the impact of Coronavirus on the internet in general has recently become a very popular research topic researchers are looking into. There are several literature about how the internet was affected by the coronavirus pandemic. This chapter studies the literature of this thesis and presents some related works.

2.2 Network Performance

Network Performance, according to ITU-T, can be defined[1] as the ability of a network to provide the functions that allows users to communicate [14], videlicet, resources needed to deliver communication services [15]. To describe the performance of a network at any time or within a certain period, we can also consider the quality of service (QoS) and the quality of experience (QoE). QoS[2] focuses on performance from the system perspective whiles QoE[3] focuses performance from the user perspective [16]. So, in that sense, the data set for this thesis seems to provide information from the system perspective (QoS) whiles the survey helps provide information from the user perspective (QoE).

[1]Note that there are several possible definitions of Network performance.

[2]In ITU-T Rec. E.800 (1994) [14], QoS is defined as “The collective effect of service performance which determine the degree of satisfaction of a user of the service.”

[3]In Qualinet White paper [16], QoE is defined as "the degree of delight or annoyance of the user of an application or service."

2.3 Network Performance Metrics

To measure network performance, there are important metrics that can be used. Some common metrics include speed (throughput), latency and packet loss. For this thesis, the metric latency and speed (throughput) will be of focus and are described below. Moreover, these metrics qualify as network measurement metrics as well which will be discussed in the next section.

2.3.1 Packet Loss

On any network, packets (small units of data) are generated and are transmitted from client to destination and are expected to be received but that is not the case all the time. Some packets never arrive at its destination at all and this is termed as packet loss. An ideal network should have little to no packet loss. High packet loss leads to problems for end users of the network. For instance, when packet loss occurs during a video transmission, the quality of video is degraded and causes a lot of buffering.

There is usually a timeout mechanism that initiates a retransmission anytime there is a packet loss. For example, if a packet does not arrive at destination in 3 seconds (timeout), it is considered lost and it is transmitted again. However, even if the packet arrives at the destination but after timeout it is still considered lost [3].

Packet loss may occur due to several reasons and can be prevented or fixed by handling these causes. Some causes of packet loss include, network congestion, software bugs, security threads, network hardware problems, deficient infrastructure and Overtaxed devices [17].

2.3.2 Latency

When a packet is generated, the time it takes for the packet to be correctly received at the destination is defined as latency and is measured in milliseconds (ms) [18]. To explain further, if a network has 10ms latency, it means it will take 10ms for a packet to be received at the destination. Latency can be measured in either one-way or round-trip time (RTT) using the ping service [19]. One-way measurement refers to measuring the time it takes for a packet to be sent either from source to destination only or the other way around whiles RTT is the total time for the combination of both directions. Latency can also be described as the measure of delay.

An ideal network is one with low latency (towards zero) as it is impossible to have no latency at all. This is because, technically, the transmission of packets from source to destination and/or back is usually not a straight path. They are transmitted over a network link and most likely go through several routers, switches and hosts. These

Table 2.1: Latency values across several network types

Network	**Range of Latency(ms)**
Fibre Networks	several hundreds
Cable/VDSL Networks	few tens
Satellite Networks	500ms (GEO), 125ms(MEO), 20ms(LEO)
WiFi Networks	few tens

segments in a network is bound to introduce certain delays. A Tutorial on Network Latency and its measurements [19] categorizes and describes these delays as;

- Processing delay; This delay is caused by the time needed by a router, switch or host to process the packet received. Upon receiving packets, routers for instance, have the obligation to decide the next hop and forwarding of the packets. They do so by reading the destination address and look up the the next hop on the routing table to determine where the packet is supposed to be forwarded to. All this take time and hence causes processing delay.

- Queuing delay; This type of delay is caused when the network is congested or processing is slow or the network link for packet transmission is slow forcing the packets to wait in a queue.

- Transmission delay; Before data is transmitted on the network link, all the bits of the packet need to be received first. The time it takes to receive and put the entire bits of packet onto the network link causes transmission delay.

- Propagation delay; The time it takes for the packet to travel over the network link. This type of delay has to do with physical distance between routers, switches and/or host. This means the greater the distance, the higher the propagation delay.

Whether measurement is for a one-way or round-trip delay, these delays are all possible and therefore the sum of these delays results in the total delay/latency in the network.

The value of latency varies across communication networks. Table 2.1 shows a range of latency for some common networks based on research by the EMEA Satellite Operators Association (ESOA) [20].

2.3.3 Speed

Internet Speed is basically the measure of how much data is sent/uploaded from our devices to Internet or received/downloaded from the Internet to our devices within a given period of time using throughput in bits per second (or sometimes in bytes per second) [21]. This means a network which has throughput of 200 bps can send 200 bits of data every second. Also, an ideal network is one with a high throughput which gives an indication that more data could be transmitted per second from our device to the internet or vice versa.

In order to measure true network performance in terms of throughput, one must know the bandwidth of the network. In other words, we measure throughput against bandwidth as these two work hand in hand. Bandwidth tells the potential capacity of a network while throughput tells the actual capacity of the network used [22]. With respect to the Internet, measuring speed is done in two ways; download and upload speed. Download speed represents how much data is being received from the server to client and the upload speed represents how much data is being sent from client to server.

According to ookla's Speedtest, in Norway, Telenor has the highest speed score[4] of about 68.04 followed by Telia and Ice (49.57 and 31.03 respectively) [23]. This gives an idea of how much throughput is available on Norway mobile networks.

2.4 Network Measurement

One of the main motives behind network measurement is to be aware of network performance by measuring the metrics explained in section 2.2. Other reasons may be to troubleshoot or identify network problems, probe into the nature of an existing traffic, network capacity or other network characteristics [24].

A measure of the metrics explained in section 2.2 serves as an indicator of how well a network is performing by presenting results for example, how much delay a network has (Latency), how much packets have been lost (Packet Loss) and/or how fast the network transmits data (throughput). Aside these metrics mentioned, connectivity, bandwidth and jitter are also metrics that can be used to measure performance. There are methods used in network measurements and these methodologies are categorized into active and passive and hybrid measurement [3].

The BEREC, however, provides guidance on a measurement methodology for a harmonised performance (QoS) which provide recommendations for IAS speed

[4]Speed Score is calculated using the values of download and upload speed. Available: https://www.speedtest.net/awards/methodology

measurements, delay and delay variation (Latency) measurements and packet loss measurements whiles ensuring network neutrality at the same time [13].

2.4.1 Passive Measurement

As the word passive implies, this methodology does not play any active role in the traffic to be measured but just collect them (traffic). Passive measurement uses only existing traffic and instigates no new traffic on the network for its measurements therefore having no influence on the performance and measurement analysis [24], [25]. However, the accuracy of measurements is highly influenced by how well the measurement devices perform instead [26].

An example of a passive measurement is illustrated in Figure 2.1 where data is collected from routers into a database using a link splitter or hub. Wireshark is one of the most commonly used passive measurement tool nowadays which allows the user to see the behaviour of an existing traffic on any network.

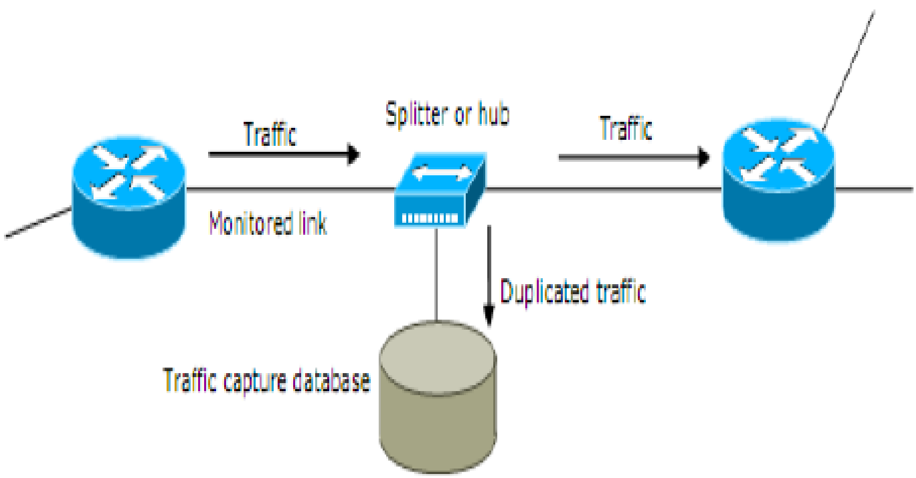

Figure 2.1: An Illustration of a passive measurement [3]

For a passive measurement, the high amount of data that could be collected and the equipment needed to collect these data is mainly the challenge of this methodology. Venkat et al. [3] gives a detailed illustration of how much data and storage capacity could be a challenge for passive measurements. Even though high amount of data collection is a major challenge it can be corrected with procedures like filtering, classification and sampling of packets [25]. Aside these challenges, passive measurements usually have low overheads, provide accurate measurements and are very scalable [27].

2.4.2 Active Measurement

Active measurement on the other hand, deliberately instigates new traffic onto the network. This methodology does not use existing traffic but rather creates its own traffic by generating special probe packets and sending them over the network in order to measure the performance metrics (throughput, latency or bandwidth for example) of the flow of the packets [3]. These probe packets are special because they are purposely designed for the measurement job [25]

The example in Figure 2.2 illustrates an active measurement where an Agent (web client) on a computer sends a web page request to a web server across a network and measures the response time. Traceroute and ping commands are the most common ways to perform active measurements due to its availability on all computers. According to [28], the recognized best practice for Speed testing is active measurement.

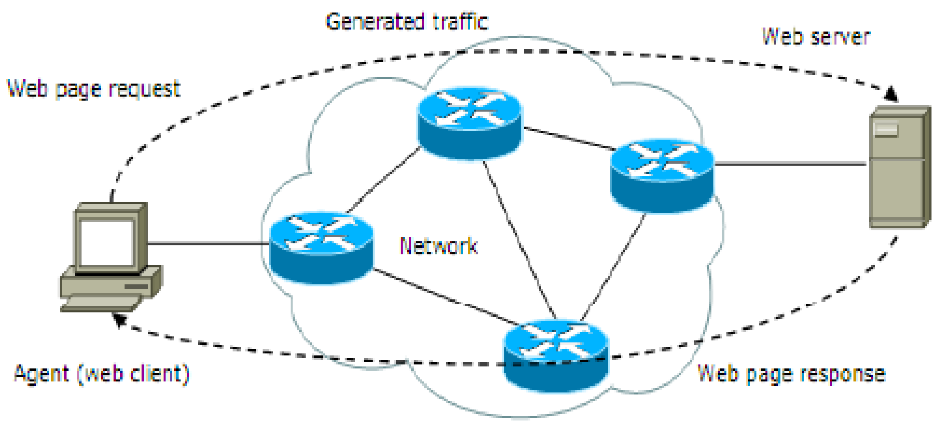

Figure 2.2: An Illustration of an active measurement [3]

By instigating these special probe packets, active measurement is challenged by the fact that it may cause some disturbance in the network [27]. This is because these packets add to the existing traffic and share network resources with the existing traffic. The measurement process might go ahead to debase the performance of the network in general [26]. For this reason, before execution, active measurement must be strategized properly. On the bright side, privacy issues and storage space is not a problem as compared to passive measurement [3].

2.4.3 Hybrid Measurement

Since active and passive measurements produce measurement results for different aspects of a network, a combination of the two might reveal the actual state of a network and that is what hybrid measurement is about [3].

2.5 Speed Test

As speed is an important way to measure how best your internet is performing within a given period, knowing the speed is vital. Hence, carrying out a speed test can help reveal this mystery. Running an Internet speed test has become quite common around the globe now with an increasing number of speed test platforms.

2.5.1 What is Speed Test?

A speed test can be described as the measure of how fast a network's data is downloaded and/or uploaded per unit time. Speed tests reveal the download speed (Mb/s), upload speed (Mb/s) and ping or latency (ms) of a network. There are several speed test platforms which are accessible on several devices. However, in Norway, the most common ones are Ookla's speedtest[5] and NKOM's NETTFART[6].

2.5.2 Requirements and Reasons For Speed Test

From the user perspective, to perform a speed test, all you need is a working network connected to your device and a supported web browser or mobile app (supported on Android and iOS devices) that will access the speed test platform. On the other hand, multiple parallel Transmission Control Protocol (TCP) connections is used for measuring throughput [28]. Figure 2.3 illustrates the components involved in performing a speed test.

The requirement for a speed test is very simple but as to the reason why speed test are usually performed could be very subjective. Some reasons behind a speed test is as follows;

- To check whether ISP is providing exactly what has been agreed on.
- To know how your network is performing at a particular time.
- To help in planning and upgrading networks.

[5] https://www.speedtest.net
[6] https://www.nettfart.no

2.5.3 How Speed Tests Work

In Section 2.4, we discussed the categories of network measurement methodologies and we can say that speed testing uses the active measurement methodology because in order to measure performance of the network, they instigate new traffic onto the network.

When a speed test is started, there are several steps that occur in order to reveal the results for the speed test. Even though most speed test platforms might use different approach for obtaining measurements, the goal is the same. It is evident in their presentation of their measurement results. They all show similar parameters; download speed, upload speed and ping or latency.

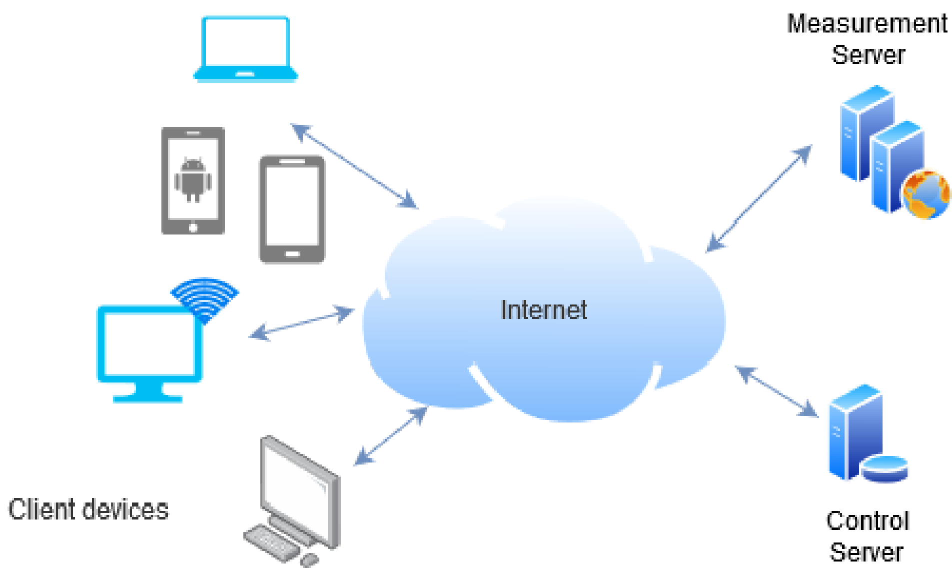

Figure 2.3: Main Components of Nettfart measurement platform [4]

First, the client (Speed test platform) determines location and finds the closest test server to the network being tested to establish a connection. Ookla's Speedtest for instance, looks up your IP address in a GeoIp database to obtain your estimated location and uses spherical geometry to determine up to five servers nearby once you press "Begin Test" before it finally selects the closest one [29].

After finding and establishing a connection with the closest test server, the measurement process begins and the first parameter that is determined is the latency (ping). It is determined by the client sending multiple requests to the server and measuring the time it takes to receive all responses from the server in milliseconds (ms). The result is shown depends on the client. For example, NETTFART shows the median of all the measured times as the result for latency measurement [4] and Ookla's Speedtest shows the lowest value as the result [12].

The next parameter to measure is the Download speed. At this stage, the client will now download data from the server whiles measuring the throughput. According to Speed of me [7], though the download test starts with downloading the smallest data size (128KB) whiles watching the download time, the download speed result is based on the sample size that took more than eight seconds to download [30].

Finally, measuring the upload speed is next. Usually, the approach used in measuring the download speed is the same as that of the upload speed. This time it is done in the opposite direction where the client uploads data to the server and measures the throughput.

Concerning accurate measurement results, the BEREC, however, provides guidance on a measurement methodology for a harmonised performance (QoS) which provide recommendations for IAS speed measurements, delay and delay variation (Latency) measurements and packet loss measurements whiles ensuring network neutrality at the same time [13].

2.5.4 Limitations of Existing Speed Tests

There are a number of limitations when it comes to speed testing. Based on a research by Nick Feamster and Jason Livingood [28] about the current challenges of existing speed tests, they reveal that several challenges arise from Wide-Area networks, Test infrastructures, Test Designs and user pespectives.

2.6 Research on Coronavirus and the Internet

Most of these papers emphasised internet traffic increased and not on performance. Papers and articles like [9], [31], [10] and [2] confirmed the speculation of traffic increase as mentioned in [6]. Moreover, [32] and [10] show how Internet proved to be robust and resilient amidst all the traffic increase. Now that this effect is evident, this paper seeks to show a quantitative research where we look into the relationship between the available data variables and systematically study whether there is a domino effect between these data variables from the perspective of performance.

One of the most relevant papers is the one by Massimo et al. [33], where aside acknowledging increase in Internet traffic, they study the coronavirus impact

[7] https://speedof.me/index.html

on Internet latency. They believed that this study could help provide estimated performance of the internet. Their main focus was on Italy but also provided some results concerning Sweden, Spain, France and Germany. They came to a conclusion that Latency was high in the evenings, the time of the day where traffic was mostly entertainment related and less latency during other times of the day where traffic was solely distant learning, e-teaching and work related. This gives an insight to which times of the day this research could investigate to measure performance.

Chapter 3
RESEARCH METHODOLOGY

3.1 Introduction

This chapter describes and explains the research methodology used to conduct this study as well as the procedures related to the selected research methodology when investigating the impact of the coronavirus pandemic on Internet performance in Norway using speed test results from NETTFART and also predict certain occurrences.

To justify the research methodology chosen for this study, we discuss the types of research methodology and provide a brief review of the literature of these research methodologies and how it is implemented in the field of Speed testing and measuring internet performance. However, this chosen methodology is geared towards achieving the research objectives and answering research questions.

Quantitative and Qualitative research are the commonly used research approach in modern research and there has been debates about which is the best research approach. Well, a suitable methodology/approach for this thesis is the Mixed Methods research approach which is basically, an adaptation of both quantitative and qualitative research approaches. This is because only a qualitative or quantitative research may not be enough to address the research questions and also a combination of both qualitative and quantitative data together will help provide a better understanding of this study.

This chapter presents an overview of the Mixed Methods research and proceeds to discuss a more detailed review of the qualitative and quantitative methodologies independently.

3.2 Mixed Methods Research

The description of Mixed Methodss Research has been given by many researchers and most of them agree on the fact that this type of research usually combines a quantitative and qualitative research approach ([7], [34], [35], [36], [37]). For instance, Ivankova, Cresswell and Stick [7] and [34] respectively describes Mixed Methodss as:

> *"an approach to inquiry that combines or associates both qualitative and quantitative forms. It involves philosophical assumptions, the use of qualitative and quantitative approaches, and the mixing of both approaches in a study"*

> *"a procedure for collecting, analyzing, and "mixing" or integrating both quantitative and qualitative data at some stage of the research process within a single study for the purpose of gaining a better understanding of the research problem"*

An editorial by Tashakkori and Creswell [38] highlighted the importance of distinguishing between Mixed Methods as "a collection and analysis of two types of data (qualitative and quantitative) and as the integration of two approaches to research (quantitative and qualitative)" and this is evident in this study. This study collects a quantitative data from NETTFART and surveys and a qualitative data from Surveys as well.

The Strategies involved in Mixed Methods research are categorized into three by John Creswell namely Sequential, Concurrent and Transformative Mixed Methodss [7]. Table 3.1 gives a brief distinction of categories.

Table 3.1: John Creswell's Strategies in Mixed Methods Strategies [7]

Sequential	**Concurrent**	**Transformative**
Data is collected sequentially (begin with quantitative followed by qualitative or vice versa)	Both quantitative and qualitative data is collected at the same time with qualitative and quantitative addressing process and outcomes respectively	Creates and uses a framework for quantitative and qualitative data collection (Framework could involve sequential or concurrent Mixed Methods)

Table 3.2 gives a summary of the types of mixed methods design. According to John Creswell and Vicki P. Clark [5], there are four major types of Mixed Methods

designs; Embedded design, Triangulation design, Explanatory design and Exploratory design.

Table 3.2: The Four Major Types of Mixed Methods Designs [5]

Mixed Method Design	**Description**
Triangulation design	Obtains a different but complementary data on the same study topic to help provide more clarity on the research problem. This design can be used to either directly compare and contrast quantitative statistical results with qualitative findings or validate or expand quantitative results with qualitative data.
Embedded design	A quantitative data set may provide a supportive, secondary role in a study based primarily on a qualitative data or vice versa. This design can include qualitative or quantitative data to answer research questions within a largely quantitative or qualitative study.
Explanatory design	This design involves a qualitative data helping to spell out or build upon initial quantitative results. This is a two-phased design.
Exploratory design	This design starts qualitatively and the result can help give more information about the quantitative method. This can also be a two-phased design

The mixed methods research is supposed to be a superior research methodology since it involves both quantitative and qualitative approaches but it has somehow inherited some of their strengths and weaknesses. Therefore, the mixed methods research undoubtedly have benefits to a researcher and limitations as well. Sami Almalki presents a research on the Benefits and challenges in integrating quantitative and qualitative data in Mixed Methods Research [35] and Johnson et al. [37], also, reviewed several definitions of the mixed methods research and several issues emerged. The table 3.3 highlights briefly some benefits and limitations of the mixed methods research based on these researchers and a few others.

Looking at the benefits of the mixed methods research, strategies and possible designs, it will be a suitable methodology for this study. In relation to the strategies of data collection, a sequential strategy is implemented where the quantitative data was collected before the qualitative data. Although this strategy was implemented, the outcome of this study would stay the same if data was collected vice versa or concurrently as data collection instruments are independent of each other.

Table 3.3: Brief Highlight of Some Benefits and Limitations of the Mixed Methods Research

Benefits	**Limitations**
■ The designs in the mixed methods research encourage gathering information from different sources whiles utilising different methods making it very efficient [35].	■ Difficulty in agreeing at which stage of the research process to perform mixing or integration [37]. Coupled with this, there is no clarification on the effective strategies for the integration process at different stages of the research process. [37].
■ To address research questions, the mixed methods research may offer a range of methods to make this possible [39].	■ Also, the skill set, effort and expertise needed to utilize the mixed methods research correctly and effectively may be a challenge [35] [39].
■ The opportunity of Skills enhancement, experience and broadening one's horizon of methodologies is presented when one undergoes mixed methods research [39].	■ Deciding which mixed methods design (Embedded, Triangular, Explanatory or Exploratory) is suitable for a particular kind of study [35].
■ Implementation is straightforward especially when structure is two-phased and researchers collect one type of data at a time whiles conducting the two methods [5].	■ Since this research integrates other research approaches it is tagged as time consuming [35].

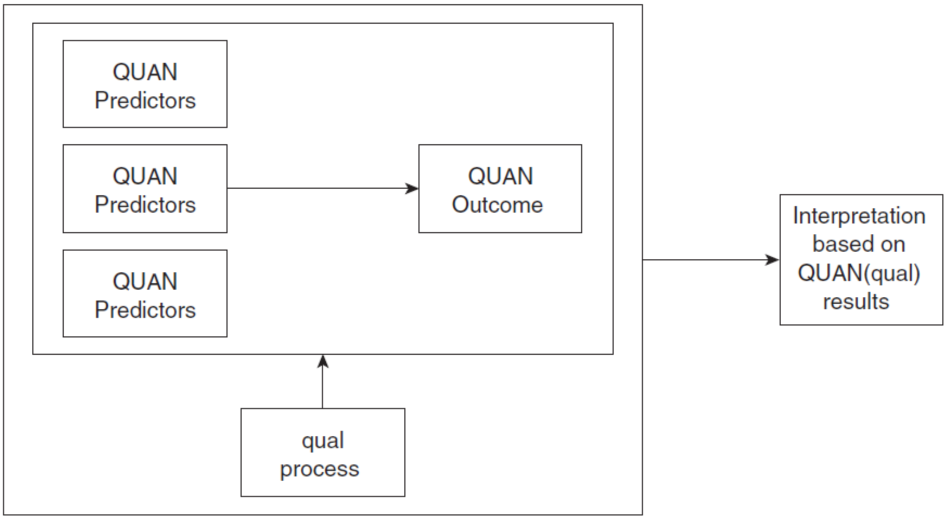

Figure 3.1: Embedded Correlational model [5]

In addition, looking at the data set available, it is made up of a largely quantitative data (from NETTFART and Survey) and a secondary or supportive qualitative data (from Survey only). The design approach used for this study was the Embedded design (see table 3.2). We use the Embedded Correlational model [5] which is illustrated in figure 3.1. In this study, with this type of embedded design, we wanted to explain and expand on the quantitative outcomes from a survey which included few open-ended qualitative questions within one survey instrument. The qualitative data was meant to provide interesting insights that we can use to elaborate the quantitative findings.

3.3 Quantitative Research and The Approach

As this thesis is a scientific one, using the quantitative approach is important. The word Quantitative relates to quantity which involves numbers and so does the quantitative approach. A good definition and description of this approach is given in the Research Guides from University of Southern California [40] which says

> *"Quantitative methods emphasize objective measurements and the statistical, mathematical, or numerical analysis of data collected through polls, questionnaires, and surveys, or by manipulating pre-existing statistical data using computational techniques. Quantitative research focuses on gathering numerical data and generalizing it across groups of people or to explain a particular phenomenon"*

As described above, in order to perform a quantitative research, data can be collected using two strategies which is in line with Creswell's Quantitative Strategies [7]. So, the approach which collects data through polls, questionnaires and surveys according to Creswell's Quantitative Strategies [7] is known as the survey research and is meant to study a sample of a population (in this case Norway) and provide a numeric or quantitative description of opinions, trends or attitude of that population. On the other hand, by manipulating pre-existing statistical data using computational techniques is the experimental research which seeks to discover what influences lead to a certain outcome by introducing one group to a specific treatment and restraining from the other group and then learn what either groups perform on an outcome [7]. These were the strategies adopted for the quantitative part hence making it relevant to this study.

3.4 Qualitative Research and The Approach

In contrast, qualitative research involves non-numerical data. It rather seeks to understand the perspective of people or a group and generate theories. The word Qualitative refers to the quality of something rather than its quantity (Oxford Dictionary) and so John Creswell [7] describes a qualitative research as:

> *"a means for investigating and understanding the meaning individuals or groups give to a problem by emerging questions and procedures, data typically collected in the participant's setting, data analysis inductively building from particulars to general themes, and the researcher making interpretations of the meaning of the data".*

The major qualitative methodologies are Narratives, Phenomenologies, Ethnographies, Grounded theories and case studies. This study contains a real or hypothetical situations, a suitable qualitative methodology was a case study where we studied the impact of an ongoing pandemic on internet performance using a survey.

3.5 Statistical Methods

According to Dharmaraja Selvamuthu and Dipayan Das [41], Statistical methods can be defined as the "mathematical formulas, models, and techniques that are used in statistical inference of raw data". The raw data for this thesis has variables[1] which is mostly numbers making it largely quantitative with supporting qualitative data therefore qualified for the use of statistical methods.

[1] A variable is a property that takes on different values and logically group attributes [42]

Planning, designing, collecting data, analysing, outlining relevant interpretation and reporting of the research findings are some examples of the statistical methods [43]. In carrying out this study, the statistical methods involved are, inferential statistics, descriptive statistics and conducting a survey.

3.5.1 Descriptive Statistics

To summarise and describe the data and the relationship between its variables in a sample we use a descriptive statistics [43]. Descriptive analysis is essential for, and provides the basis for inferential statistics and therefore it is usually recommended to be undertaken before inferential statistics [44]. For this study, the descriptive statistics involved are the measure of frequency and central tendency.

Measure of Frequency is one of the most common ways to describe variables. Frequency shows how often a variable may occur. Measure of frequency could be Absolute or Relative. Absolute frequency shows the number of times a particular value occurs in the data whiles relative frequency is just absolute frequency relative to the total number of values for that variable [44]. For this thesis, we measure the relative frequency of the number of speed tests performed by county, annually, monthly, weekly and even hourly. Visually, measure of frequency is usually illustrated using graphs and tables. We used bar graphs and line graphs for easy interpretation of the data (4).

Measure of Central tendency usually gives a thorough overview of the whole data set. According to [45], it is described as *"the descriptive summary of a data set through a single value that reflects the center of the data distribution"*. Mean (average), median () and mode (most frequent) are the three primary measures of tendency. For this study, we find the mean or average as well as the median of the internet speed. Since the average alone may not be the true centre of the data since it is highly influenced by very small or large values in the data set (if present), we find the median as well to ensure it is not much of a difference.

3.5.2 Inferential Statistics

After conducting descriptive statistics, inferential statistics can be done for a more meaningful result. Inferential Statistics is defined by Professor Amin [46] as the statistics that *"predicts or estimates characteristics of a population from a knowledge of the characteristics of only a sample of the population"*. Inferential Statistics can be used for making judgements on probability and also to make inferences from a set of data to general conditions [47] but this thesis uses the latter.

For instance, with this study, the population is the whole Norwegian population but since it is impossible to reach every single one, we use a survey that reaches only a few hundreds and based on their answers we make inferences.

3.5.3 Statistical Survey

Statistical survey is an inquisition (usually structured) about the attributes of a given population by collecting data from a sample of that population and evaluating their attributes through the systematic use of statistical methodology [48]. The data collected could be qualitative or quantitative. The efficiency, external and internal validity, geographical spread of samples and flexibility characteristics of surveys [49] is why we conducted a survey for this study.

Surveys may take the form of cross-sectional, longitudinal and Explanatory or Correlational [49]. This study, however, seeks to explore causal relationships between several variables. We wanted to know whether users experienced Internet quality issues and its relationship between speed testing hence an explanatory or correlational survey was used.

We used a questionnaire (in appendix) to collect data from a sample population although there are other methods of collecting data like face-to-face interviews and telephone interviews. We discuss more about the design and construction of the questionnaire used for this study in the upcoming sections.

3.6 Statistical tools

To make data collection, statistical analysis, interpretation and making meaning out of raw data, there are statistical tools that can make these possible. A research may incorporate one or more tools in order to achieve its goals and objectives.

For statistical tools, there are several alternatives that exist for this study considering the data set. With these tools, the intention was to create methods for filtering of the data (e.g. select columns) and statistical methods. Feasible statistical tools include the following:

- Google earth to point out and display some specific locations based on latitude and longitude information provided in the data set.
- GNU plot or any plotting software for plotting the data
- Linux terminal tools: script languages to access the huge data
- Python and/or the R language for creating algorithms
- RStudio to provide an environment to filter out data.

- Mathamatica, matlab and Excel might be useful for plotting when working with a smaller sets of data.
- Microsoft forms to help create questionnaires for survey

3.6.1 Selection criterion

There are several options of tools to choose from when performing statistics. These tools are to aid and simplify statistical analysis during statistical research. Although these tools might have similar properties, each of them might have unique features as well. Since this thesis is a statistical one, the following criteria was taken into consideration. The statistical tool must be

- designed and suitable for statistical learning. The statistical tool must able to perform statistical methods such as calculating data average and median and also represent them graphically (line and bar graphs). In addition, mapping geographical locations should be a present feature
- intuitive, comprehensive and easy to use.
- able to handle large data. The data set for this study contains millions of data entry so a statistical tool that can handle such large data is suitable.

With this criteria, the R language and RStudio best fits.

3.6.2 R Language

R is a programming language for statistical data analysis written by Ross Ihaka and Robert Gentleman with the first version released in 1993 [50] and its source code archive management and modification is now the responsibility of the R Core Team since 1997 [51]. R is often referred to as an implementation of the S [2] programming language since it was built on S language [52].

R has a lot of features that suits better for statistical learning. It is open source and is compatible with all computers running any operating system [53]. It supports importing data set files from various platforms. For instance, R was used to import the data set for this thesis from a CSV file. R is able to perform statistical and numerical analysis of large data sets, data modelling and presents a much better data visualization [54]. One exceptional feature is the fact that R provides several avenues for getting help due to its active and massive user community [53].

[2] https://web.archive.org/web/20181014111802/http://ect.bell-labs.com/sl/S/

Table 3.4: Packages used and their description

Packages	Description
sf [3]	It provides support for simple features like encoding spatial vector data, reading and writing data, geometrical operations and data transformations [55].
dplyr [4]	A tool for manupulating data frames like stringing together a sequence of actions [56].
lubridate [5]	It consist of functions to work with date-times and time-spans. So with a data set, it is possible to parse, extract, update and manipulate all time based objects [57].
readr [6]	It provides a way to read cross-sectional data with multiple statistical variables (like 'csv', 'tsv', and 'fwf') [58].
ggplot2 [7]	It converts a data frame and plot specification into a complete graphic and is based on "The Grammar of Graphics" [59].
RCurl [8] (A wrapper for "libcurl") [9]	Simply provides all functions relating to HTTP [60].
DT (Datatables)[10]	It provides features like filtering, pagination, and sorting when the data frames are rendered as HTML tables using the JavaScript library.
RColorBrewer [11]	It provides color schemes.
jsonlite [12]	converts JSON data from or to R objects and also generates, parses, stream and validate JSON data [61].
leaflet [13]	It helps to create and customize interactive maps.

In order to perform functionalities that are not present in the base R and also, facilitate easier programming, R makes use of add on packages and libraries of which there are thousands of them in the R repository [53]. Only a hand full of these packages and libraries were used in this thesis (Table 3.4).

R is simply a system for statistical computation and graphics [51] which is why it was used mainly for this master thesis. It helped in analaysing, plotting and graphing of the data in this thesis.

[4] https://github.com/r-spatial/sf/
[5] https://dplyr.tidyverse.org
[6] https://github.com/google/cctz
[7] https://github.com/tidyverse/readr
[8] github.com/tidyverse/ggplot2
[9] https://CRAN.R-project.org/package=RCurl
[10] https://github.com/rstudio/DT
[11] http://colorbrewer2.org
[12] https://arxiv.org/abs/1403.2805
[13] https://rstudio.github.io/leaflet/

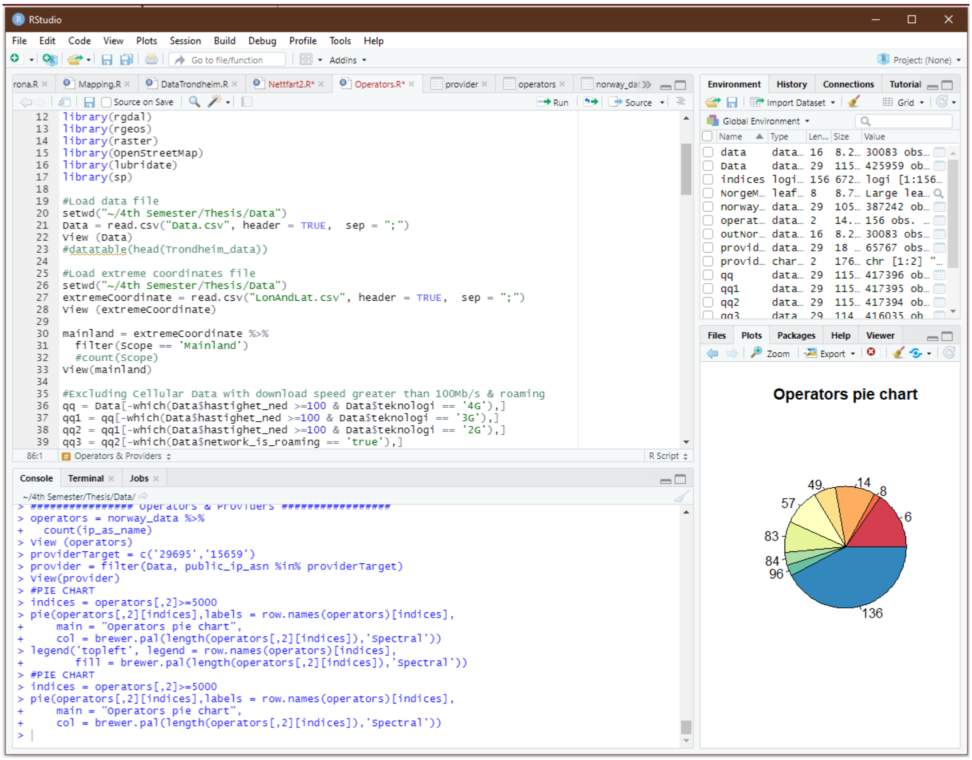

Figure 3.2: RStudio interface on a desktop

3.6.3 RStudio

RStudio is an integrated development environment (IDE) for R. From Figure 3.2 showing RStudio interface on a desktop, there is a script editor (top left), a console (bottom left), Environment/History (top right) and Files/Plots/Packages/Help (bottom right). RStudio basically does code execution, plotting and graphing. Since its founding in 2009, Rstudio has been used by millions of people around the world to make sense of data [62].

3.7 Questionnaire and Survey

To obtain an in-depth knowledge about a problem, a questionnaire is constructed for a sample population by presenting a series of questions individuals that could be later used (most likely) in a survey. The data obtained is the answers of the questions. The data obtained could be in a textual format (qualitative) and/or numerical form (quantitative). As discussed in section 3.2 where we used an embedded correlational design, we constructed a questionnaire that produced largely quantitative results and qualitative results as well to support the quantitative findings.

Questionnaires can be conducted by an interviewer face-to-face, on phone or on the web. We used mostly on the web style of conducting the questionnaires.

3.7.1 Design and Construction of The Questionnaire

Since we decided to adopt the embedded correlational design for mixed methods, we need to obtain a qualitative data to explain and expand our quantitative findings. Following Bryn Farnsworth [63] six steps to a good questionnaire design, we constructed a feasible questionnaire.

First and foremost, we identified the purpose of this questionnaire which should be in line with the research aims and objectives. The purpose of this questionnaire was to discover why and when People mostly perform speed tests, if users actually experienced quality issues and understand the user perspective of internet performance before and during the pandemics. This is in relation to the first research question.

Secondly, we defined the target respondents and how to reach these respondents. Since this study is Norway bound, the target respondents were those who live in Norway only. Participants included students, lecturers and workers spread across Norway.

Thirdly, we went ahead to develop and write the questions. Each question was aimed at testing one or more research questions. Afterwards, we chose to ask both open and closed type of questions. Open questions mostly to give qualitative results and closed questions for quantitative results [63]. We ensured that questions were very clear and concise. Also, we did not ask too many questions to make the questionnaire lengthy. Next, we planned the overall layout and how questions will flow in the questionnaire.

Finally, before sending it off to the masses we ran a pilot. We handed the questionnaire to five people in the sample population before pushing it to the masses. It helped identify issues in understanding and readability of the questionnaire.

3.7.2 Assessing the Validity and Reliability of the Questionnaire

Obtaining pertinent knowledge during a research is what questionnaires hope to achieve and they do so in a reliable and valid way. Validity and Reliability describes how accurate and consistent (respectively) a measuring method (in this case a questionnaire) should be [64]. A questionnaire is usually acceptable when it has these features. Table 3.5 gives a summary

Table 3.5: Summary of Validity and Reliability according Fiona Middleton [8].

	Reliability	Validity
What does it tell us?	if the research is conducted again under the same conditions, the results will be the same	the results represents the true measurement that is being sought-after
How is it assessed?	By checking the consistency of results across several parameters like time, observers and parts of the test itself	By checking how well the results correspond to established theories and other measures of the same concept.
How do they relate?	A reliable questionnaire is not valid all the time. Although the results might be repeatable, but they will not be necessarily correct	A valid questionnaire is mostly reliable. if a questionnaire produces accurate results, they should be repeatable.

To ensure validity and reliability of results, we used a strong research design where we chose appropriate methods and conducted the survey carefully and consistently.

3.7.3 Data Analysis and Interpretation of Survey

After collecting valid and reliable data, we then went on to perform the survey by further analyzing and making inferences. We Followed John Creswell's [7] series of steps for a complete discussion of the data analysis procedures.

First, we reported information about the number of members of the sample who did and did not return the survey by describing the respondents and non-respondents in a table with numbers and percentages.

Secondly, we discussed the procedures used to check response bias. Whether responses from non-respondents would have a major impact on the current responses. Using the respondent/non-respondent analysis, we contacted few non-respondents by emails, texts and phone to see whether their responses will be any different.

Thirdly, we indicated the averages and range of scores during analysis and provided a descriptive analysis of the data we received from the responses.

Fourthly, we identified the suitable statistical computer program for testing the research questions and hypotheses. To be able to make inferences from a sample and project it to a population, we compare the variables relating to the hypotheses and research questions.

Finally, the data analysis ended with presentation of results and interpreting the results from a statistical point of view. Conclusions were drawn from the results for the research questions and hypotheses. We interpreted the results by first, reporting how they answered or contradicted the research questions or hypothesis, then explained why the results happened as it did and finally discussed the implications and further studies on the topic.

3.8 Mixed Method Analysis

Since data analysis in mixed methods usually is in line with the research strategy that has been chosen, making it clear in the design is important. John Creswell highlights several mixed method analysis procedures [7]. For a concurrent and sequential strategy, the most popular data analysis approach highlighted were Data transformation (quantifying qualitative data), exploring outliers (mostly from quantitative data), developing instruments, examining multiple levels and creating matrices.

Julie Combs made an elementary research into sequential mixed analysis used to analyse doctoral student test score data and survey responses [65]. She described it as a Quantitative dominant Mixed Analysis. This is quite similar to this study in a way since we have speed test results and survey responses as well. This analysis involved six stages and could be adopted for this study but was not. This is because the limitations of this approach gave the impression that the analysis was not thorough enough and could have been explored further.

We adopted Onwuegbuzie and Teddlie's [66] concept of mixed methods analysis which consist of a seven-step process listed below;

1. **Data Reduction**
2. **Data Display**
3. **Data Transformation**
4. **Data Correlation**
5. **Data Consolidation**
6. **Data Comparison**
7. **Data Integration**

Chapter 4

DATA ANALYSIS

4.1 Introduction

This study has two groups of data set. One from Nettfart containing speed test results and results of a survey conducted during the period of this study. In this chapter, we present the analysis procedures and results of this study. We do so by analysing and presenting the results from both groups independently. However, before we discuss these analysis and results of this study, we take a look at our preliminary results from the pre-project as well [6].

4.2 NETTFART Data

The data set was obtained using Nkom's Nettfart. Nettfart is an online speed test tool which is accessible as a mobile app (Figure 4.1) or as a public website (Figure 4.2) where one can measure Internet speed or perform speed tests for broadband lines and is owned by The National Communications Authority (Nasjonal kommunikasjonsmyndigheit - Nkom) [1] of Norway [67].

4.2.1 Data Set and Description

This section describes the data sets used in the analysis for this thesis. The data set is basically speed test results from mobile devices and web browser across the whole country of Norway and beyond. The speed test results shows the upload and download speed (in Mbps), ping, jitter (in ms) and the name of the network the device is connected to as shown in figures 4.1 and 4.2.

For this thesis, the data set collected are the speed test results from mobile devices and web browsers. For the data set from mobile devices, there are over 420,000 entries and the data spans a period of 3 years (from May 2017 to July, 2020). On the other hand, the data set from web browsers are categorised annually and contains

[1] nkom.no

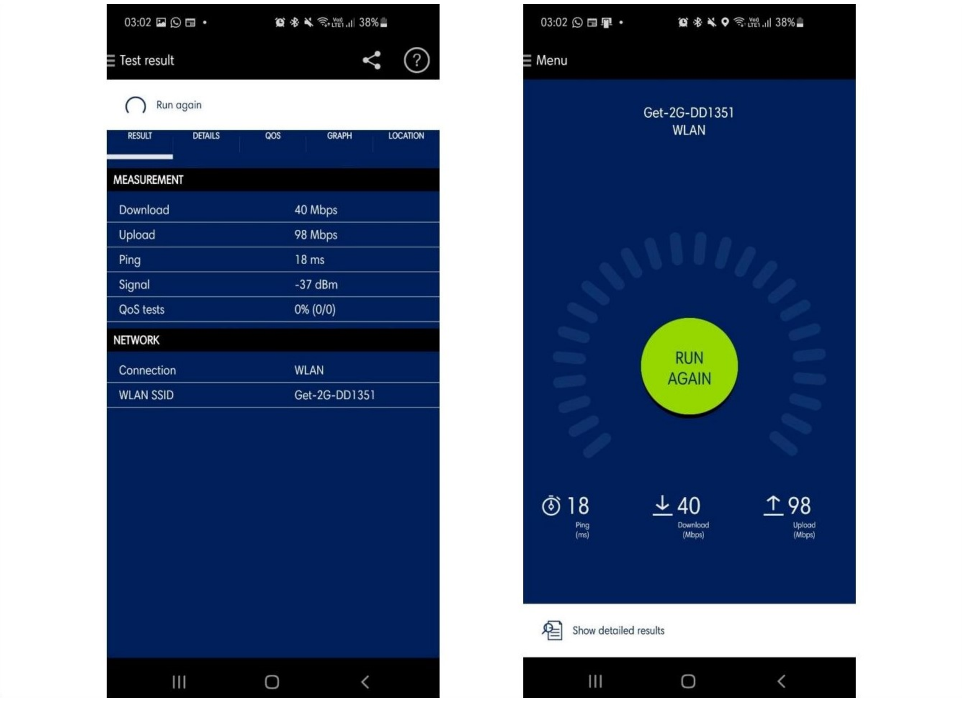

Figure 4.1: Nettfart on mobile app

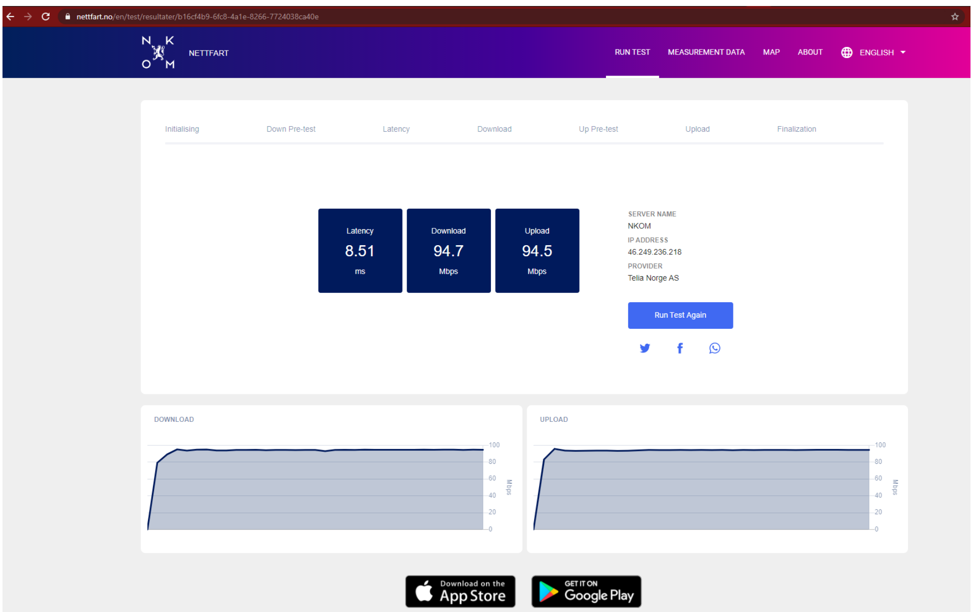

Figure 4.2: Nettfart on the web

over a million entries (between 1.1 and 1.5 million) each year spanning a period of about 8 years (from January 2012 to January 2021). The data set had variables that gave information on the following:

- download speed (kbs and Mbs)
- upload speed (kbs and Mbs)
- ping
- timestamp (Date and Time)
- locations (in longitude and latitude, counties, cities)
- client type (Mobile, web browser)
- network information (technology (4G/WLAN), Public IP AS name, Network Type and Provider, Country GeoIP and ASN, Public IP RDNS, network is roaming)
- device info (OS version, platform, device name and model)
- Signal Strength,
- cell locations (Area code and Location ID)
- mobile network info (operator name, sim operator and sim operator name)
- number of threads requested
- number of threads
- target measurement server
- ping median

With the help of a mapping algorithm using R [2] and R studio [3] (discussed in Chapter 3), we discovered that some entries were beyond Norway and thus with the help of filtering algorithms, we excluded them since the focus of this thesis is within Norway only. Also, we discovered that there were over 150 providers and the most common of them were, Telenor, Telia, Altibox, Ice, Global Connect AS, Eidsiva and NextGentel. These providers constitute over 80% of the entries. Figure 4.3 illustrates the share of operators in 2020 for example.

[2] https://www.R-project.org
[3] http://www.rstudio.com/

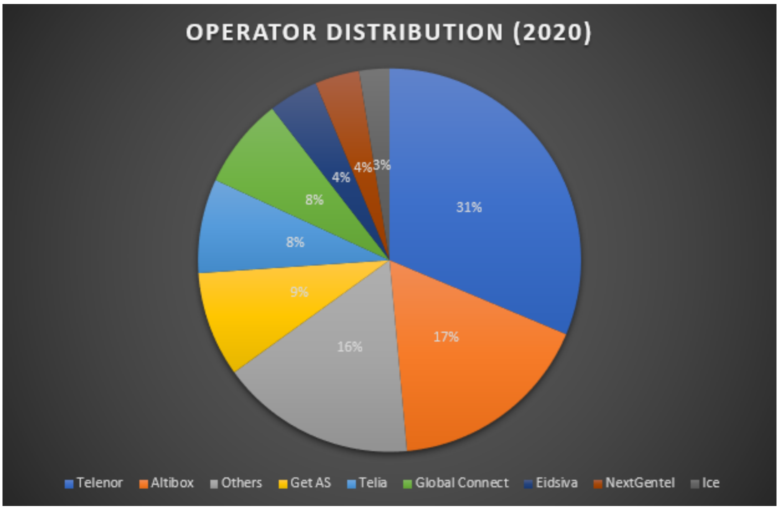

Figure 4.3: Pie chart Showing Operators Share

4.2.2 Preliminary Results

For the pre-project, the data set available was Nettfart speed test results from mobile devices only. With this data set available, we analyzed the geographical aspects of the data sets to know whether the location specific traffic patterns were affected. First, we filtered out the dates 1st October,2019 through to 11th March,2020 to represent the period before the pandemic and 12th March,2020 through to 6th July,2020 represented the period during the pandemic. We then mapped out Trondheim city and projected the speed test performed within the city before and during the pandemic and is illustrated in figures 4.4 and 4.5.

The figure 4.4 shows a slight increase in speed tests performed in residential areas and figure 4.5 shows a decrease in business areas. For instance, areas like Byasen and its surroundings (mainly residential) saw a slight increase and areas like sentrum and its surroundings (mainly business) saw a slight decrease.

In addition, we performed descriptive statistics on this sample and found the mean/average, standard deviation (σ) and the variance (σ^2) of the download and upload speed. This analysis was done to find out if the average speed before and after the pandemic changed and if it was dependent on the type of technology used (4G or WLAN). The result is presented in tables 4.1 and 4.2 and it shows a decrease in the average/mean, standard deviation (σ) and the variance (σ^2) of the upload speed in both 4G and WLAN . On the other hand, we see an increase in the average/mean, standard deviation (σ) and the variance (σ^2) of the download speed in both 4G and WLAN.

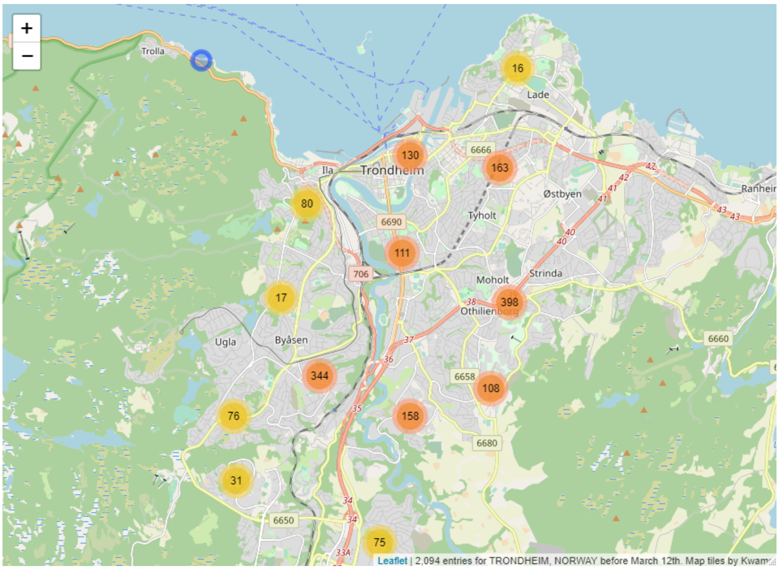

Figure 4.4: Trondheim Map Showing Exact Mobile Speed test Locations Before Lockdown [6]

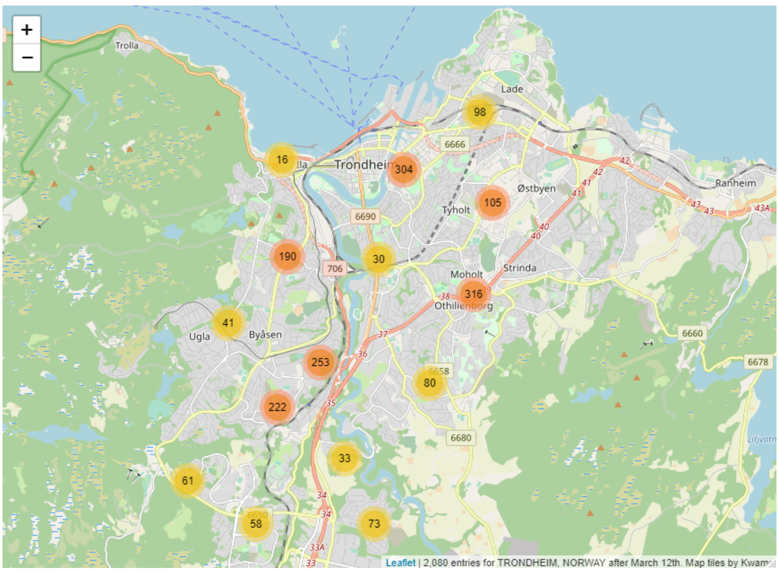

Figure 4.5: Trondheim Map Showing Exact Mobile Speed test Locations During Lockdown [6]

Table 4.1: Statistics for Upload and Download Speed for Trondheim Before Lockdown [6]

Tech	**n**	**mean (up)**	σ **(up)**	σ^2 **(up)**	**mean (dwn)**	σ **(dwn)**	σ^2 **(dwn)**
4G	192	20.92	17.51	306.73	70.58	75.17	5651.26
WLAN	1895	71.58	89.52	8013.67	110.26	101.82	10367.6

Table 4.2: Statistics for Upload and Download Speed for Trondheim During Lockdown [6]

Tech	**n**	**mean (up)**	σ **(up)**	σ^2 **(up)**	**mean (dwn)**	σ **(dwn)**	σ^2 **(dwn)**
4G	148	19.79	16.54	273.44	89.28	95.08	9039.88
WLAN	1914	69.94	90.88	8258.61	112.27	108.26	11719.74

4.2.3 Analysis and Results from Web Browser

Mapping

Here, we had available the data set from web browser in addition to the data from the pre-project (mobile data only). First, we performed analysis on the web browser data as we did during the pre-project. Filtering out dates from 1st January,2020 through to 11th March,2020 to represent the period before corona and 12th March,2020 through to 31st April,2020 represented during corona period. Again, mapping out Trondheim city produced the same result with residential and business areas but this time with a higher sample size (Figures 4.6 and 4.6).

Number of Speed Tests

Afterwards, we proceeded to making descriptive statistics. Since data spanned over an 8 year period (from 2012 to 2021), We analysed the web browser data year by year excluding the year 2021 since it only had data from a few days in that year. First, to know how often the population performed speed tests, we found the measure of frequency. We measured frequency from different perspective. We measured based on counties, monthly, weekly, daily and even hourly whiles comparing them annually.

From the County perspective, obviously, counties with a higher population had the highest measure of frequency. Over the years, there was a small change (either a small increase or decrease) in the measure of frequency in nine of the counties with no particular order. Oslo and Viken, the most populated counties in Norway, rather showed a steady increase annually with a very significant increase in the year 2020. In more detail, the measure of frequency in Oslo increased annually by about 7.2%

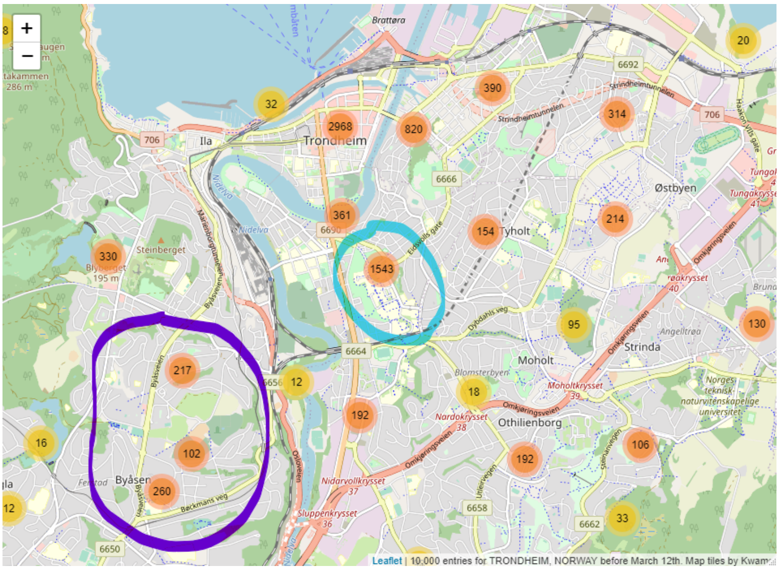

Figure 4.6: Trondheim Map Showing Exact Browser Speed test Locations Before Lockdown

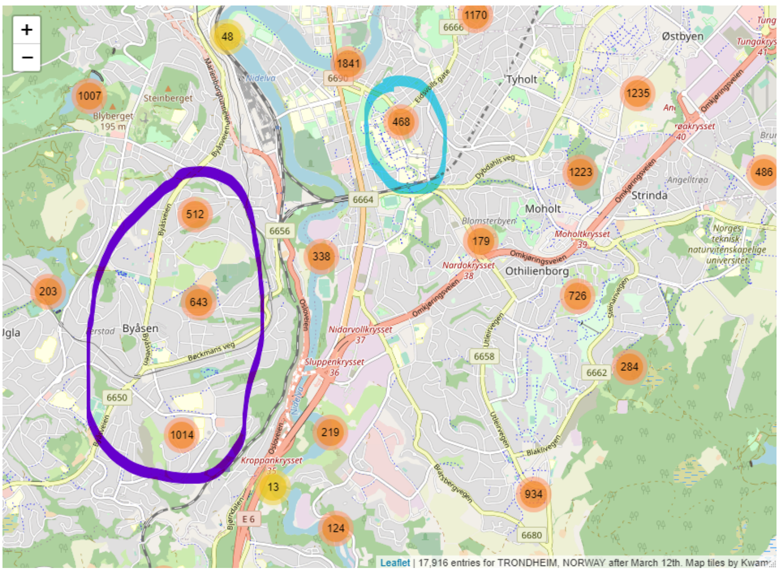

Figure 4.7: Trondheim Map Showing Exact Browser Speed test Locations During Lockdown

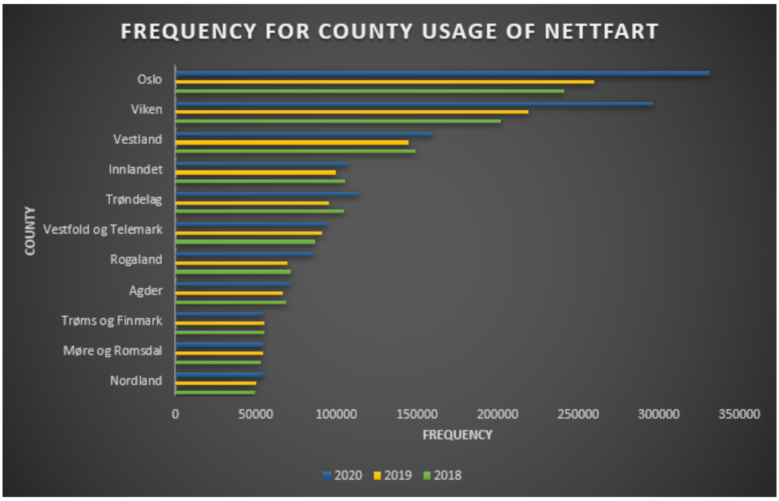

Figure 4.8: Number of Speed tests done in the Counties

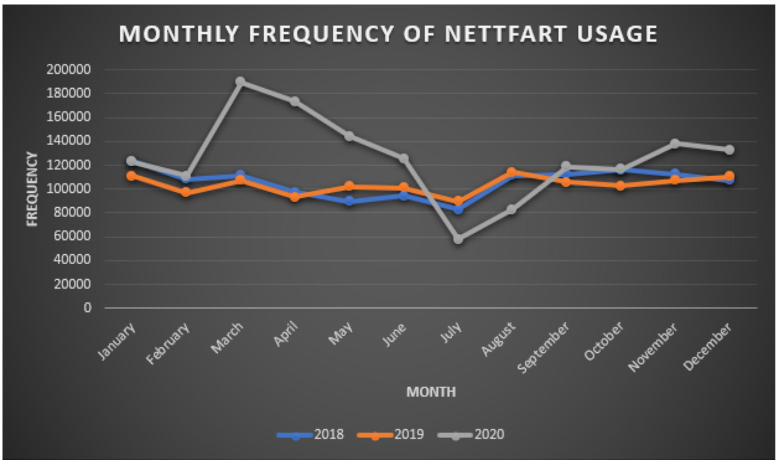

Figure 4.9: Number of speed tests done in a month

but saw about 21.7% increase in the year 2020. In Viken, the measure of frequency increased annually by about 7.8% and about 26% increase in the year 2020. Figure 4.8 shows the graphical representation of results.

Considering the monthly measure of frequency, there was no significant changes in the previous years. Every month saw slight changes up until March which produced a very great spike of about 12.5% increase. Also, in the month of July each year, it recorded the least measure of frequency. A graphical representation is shown in figure 4.9.

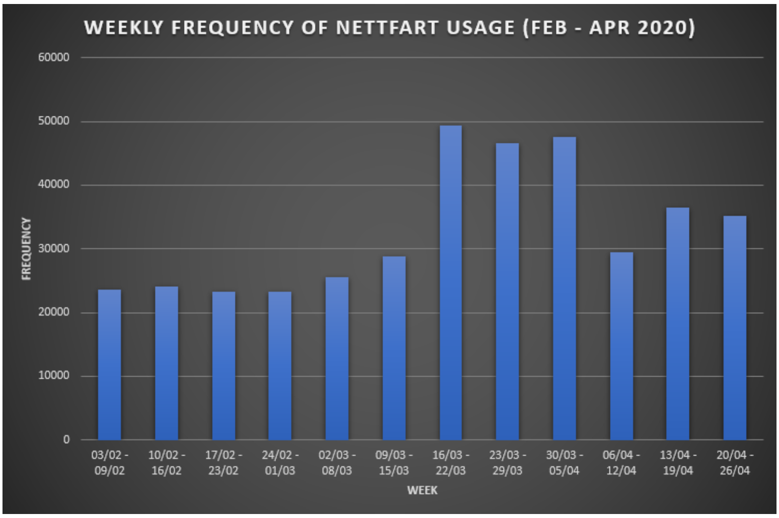

Figure 4.10: Number of speed tests done from week 6 to week 17

In view of the spike in the month of March, we went ahead and measured the number of speed tests from week 6 to week 17 of 2020 (from 03/02 to 26/04). We then saw little to no changes week after week until another spike in week 12 (16/03 - 22/03) and was similarly high for the next two weeks until a drop in week 15 (06/04 - 12/04). After the drop in frequency in week 15, changes in the following weeks were visible but not as much as in week 12. Figure 4.10 gives a visual of the result.

Further, we measured the frequency from the perspective of time of the day. The results showed the year 2020 having a high frequency between 10am and 10pm throughout the year whiles the subsequent years always saw a gradual increase from 7am until its highest at 10pm followed by a drop after 10pm till 7am. Graphical illustration of results in figure 4.11.

Change in Average Speed

The next step in our analysis was to find the average speed for before and during the pandemic. With speed categorized as upload and download speed, we measured the averages just like we did in the pre-project. We measured the averages based on counties, months, weeks, days and even hours whiles comparing them annually like we did for the measure of frequency.

The average download and upload speed experienced a steady increase. In other words, when we considered the averages by counties, we saw each counties' average speed increase steadily as the years went by no matter the population size. This steady increase is reflected in the monthly, weekly and even hourly analysis as the

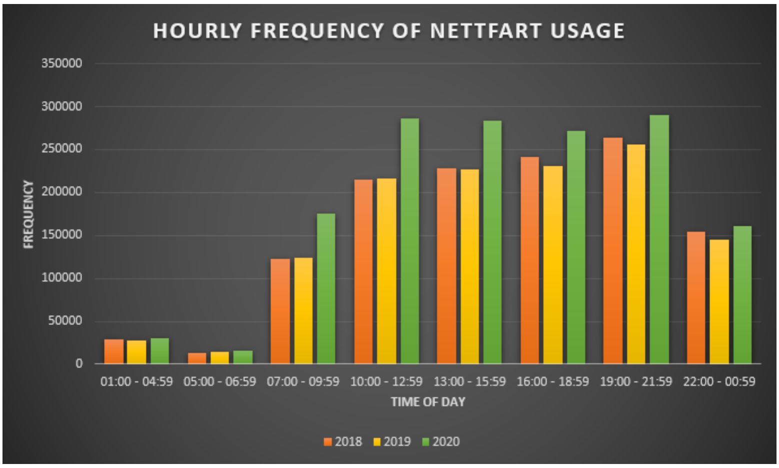

Figure 4.11: Number of speed tests done during the day

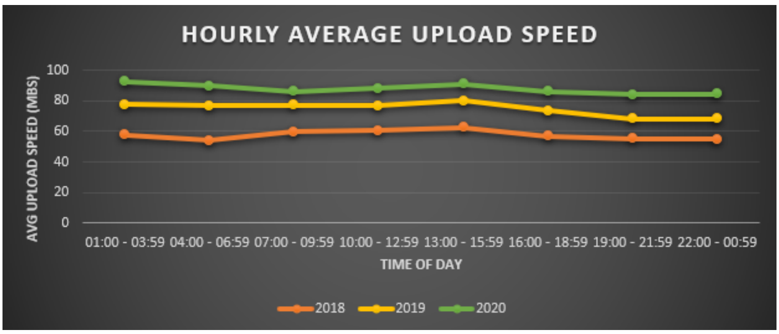

Figure 4.12: Change in Average Upload Speed per hour during the day

averages increased as the years also increased (See Appendix B for graphs). There was little to no significant changes in the average speed value for the monthly, weekly and hourly analysis. For instance, difference in average speed between a particular month and its previous or next month is always minimal. Same can be seen with the time analysis (Figure 4.12). Figure 4.12 also shows the nature of the graphical representation of the results when we considered the average upload and download speed based on counties, months, weeks, days and hours when we make an annual comparison.

Change in Average Latency

In addition to speed, we measured the average latency as well. Here, there was no pattern of an annual steady increase. Instead, the year 2020 always had the highest average latency compared to previous years. The change in value of the

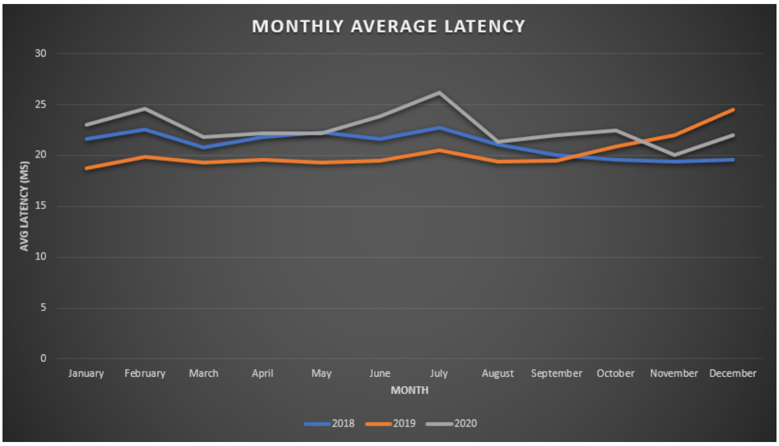

Figure 4.13: Change in Average Latency over the years

average latency over the years is minimal (usually between 1ms and 3ms) as well. Graphical representation in Figure 4.13 shows how latency changed during the year. In Appendix B, there are results for weekly and hourly analysis of latency which have the same nature.

4.3 Survey Data

After our quantitative analysis, we then went on to obtain some more quantitative data qualitative data by taking a survey using a questionnaire. The targeted number of participants were 100 people but there were about 74 participants in total for this survey. Five people participated in the pilot test and the remaining 69 participated in the main survey. This section describes our questionaire and presents the results from the survey based on the questions asked in the questionnaire.

4.3.1 Questionnaire and Description

The questionnaire was created with Microsoft forms and links were shared to the masses. It was mainly an online based questionnaire. The questionnaire had about 69 responses. All respondents were in Norway before and during the pandemic. Also, the questionnaire was open from March to April 2021 and was responded by students and lecturers (in and outside NTNU) and other workers in Norway.

The questionnaire had 12 questions in total but less for each respondent when answering. This is because the questionnaire was designed to have follow up questions and implemented skip-rules where participants skip questions that do not apply to

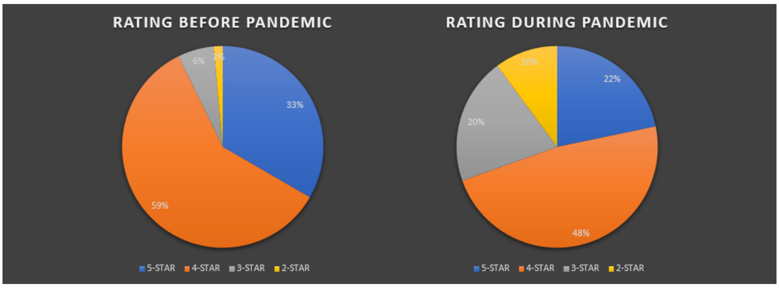

Figure 4.14: Internet Rating Before and During Pandemic

them [63]. There were about three open-ended questions and other closed questions as well.

The questionnaire is found in Appendix C and appears in the same order it was presented to the participants.

4.3.2 Results from Survey

We asked random people living in Norway before and during the pandemic to answer the questionnaire about the pandemic's influence on the Internet performance. Some results were expected (poor Internet quality experience) and others were less predictable (experiencing poor Internet quality due to construction activities).

Most importantly, there are several things we learned from the results. The key findings are as follows;

1. Participants felt the internet under-performed during the pandemic as compared to before the pandemic. The average performance rating out of five before the pandemic was about 4.25 with majority (about 59%) rating it as 4. On the other hand, average Internet performance rating was about 3.81 with a greater number rating below 4.

2. Also, participants responded that they actually faced Internet quality issues during the pandemic. However, they associated the quality issues with several reasons. Because the question was an open one, some of the responses were not straight forward. Out of the 69 participants, we had about 60% confirming Internet quality issues, 37% of participants said they experienced their normal Internet quality and the rest gave inexplicable answers. One participant said their Internet even got better and that was an answer we never expected.

3. Finally, the survey revealed the most likely reasons or situation people would perform speed tests. About 81% of the participants who responded to this

question said they are most likely to perform a speed test when the network is bad. In addition, performing speed test randomly and when one is in a new location was the next likely reasons. One participant also responded, they perform speed test when they have made changes to the network setup.

To prove our responses valid and reliable we checked for response bias. We went ahead and contacted few non-respondents by emails, texts and especially phone to see what responses they would have given (respondent/non-respondent analysis). Their responses was quite similar to the others. Only one of them had a different perception of the reason for a speed test. The respondent said they only perform speed tests when they have multiple devices that heavily use the web at the same time.

Chapter 5
DISCUSSION

5.1 Introduction

This chapter discusses the study results and its relationship with the research questions and hypothesis. We interpret the results and show how it answers or contradicts our research questions and hypothesis. The research questions for this study is as follows;

1. **To what extent has the users experience changed in user quality compared to before?**
2. **Did the average Internet speed before and during the pandemic change?**
3. **Does the internet speed change over the day?**

5.2 The Survey, User Experience and Internet Quality

The user experience and how users feel about the performance of the Internet is a very important subject. We set out to find out if the user experience changed during the pandemic. From the survey results in chapter 4, there is an indication that user experience in quality changed.

The survey had majority of the participants affirming this change. Coupled with this, the ratings of the internet performance for before and during the pandemic affirmed this change. Also, the third key finding in the survey linked speed testing to bad or poor network. Now with speed testing associated with bad or poor network, lets make a more quantitative discussion.

Considering the measure of frequency which reveals how often people performed speed tests, we saw a significant increase in the year 2020. Figure 4.9 in chapter 4 shows a great increase in the measure of frequency for the month of March, when the national lockdown was imposed and the masses moved several activities online.

There was huge increase in the number of speed testing in that month followed by subsequent months. Over the years speed testing has never seen this sharp increase. This is an indication of a change in Internet quality in the year 2020 especially in the month of March.

The extent by which the pandemic affected user experience can be drawn from the measure of frequency during the day (Figure 4.11). In previous years, frequency of speed tests steadily increased over the day but in the year 2020, frequency of speed tests was high all day. This gives an indication that internet quality changed and users experienced it almost through out the day (10am to 10pm) unlike previous years when it was usually between 7pm and 10pm.

5.3 Geographical change in Origin of Speed Tests

As mentioned in chapter 1, section 1.1, with most school and work activities moved online, the origin of traffic was expected to change geographically due to majority working from home and not their offices and schools. Observing figures 4.6 and 4.7 which points out the exact locations where speed tests were done. We observe an actual change in the number of speed tests done in certain areas before and after the pandemic. The Light blue circle is NTNU's Gløshaugen campus and its immediate surroundings. This is a place where there are expected to be a great number of students, lecturers and workers present in that area almost all the time. Before and during the pandemic saw a significant change. In contrast, the violet circle points to Byåsen a more residential area in Trondheim. Here, we witness an opposite change. The figures show more speed tests were performed during the pandemic in residential areas and a few in business/work areas.

5.4 Average Internet Speed Before and During Pandemic

In this study, we observed a steady increase in average upload and download speed every year. As average speed increases annually, it indicates that upload and download speed gets better every coming year. We observed the same pattern in the year 2020 as well which was not much expected. Any speed analysis that we performed ended up following the same pattern when we made annual comparisons. This means average speed usually increased with time. So the average speed during the pandemic is higher than before the pandemic and this is not because of the pandemic's impact but rather as result of a trend over the years (See Figure 4.12 or Appendix B).

However, the average speed and measure of frequency for the month of July always dropped. This could be attributed to July being the month people usually go on vacation leaving the population size to its minimum around that period.

5.5 Average Internet Speed Over The Day

Over the day, we usually differentiate working hours from home hours. Working hours last between 7am up until 4pm and the rest of the day considered home hours. As section 5.4 highlighted the annual average increase, we observed the same trend here.

Observing the download and upload speed over the day, there is a change in the average speed value every hour but the difference is not that much. The difference in average download speed during the day is between 7Mbs and 15Mbs and that of upload speed is between 8Mbs and 12Mbs. Considering the fact that there was a huge increase in the measure of frequency over the day, there were no significant changes even during the supposed peak times. This gives an indication that the average speed over the day is not impacted by the number of speed tests performed within those hours (See Figure 4.12 or Appendix B).

Chapter 6
CONCLUSIONS AND FUTURE WORK

There was no significant impact on the average upload and download speed and latency. However, the pandemic impacted the number of times the population performed speed tests as there was an obvious increase in speed testing around the country and per the survey conducted we may attribute it to a change in user quality.

There was an indication that the user experience was affected by the pandemic as interpreted from our survey and frequency of speed testing in Norway. Its effect was experienced almost throughout the day (working and home times).The survey primarily gave an indication that there was internet quality issues during the pandemic even though the causes differ from respondents. However, the increased number in speed tests had no relationship with the change in average download and upload speed and latency.

The study discovered that the average internet speed increased annually. Aside the annual increase, there was little change in average speed during the day, even during the times when number of speed tests was very high but from a user perspective it might be enough change to perform a speed test.

In the future, one can study about the annual increase in average speed. The causes and factors that influence this trend could be explored. Also, this study can be repeated in some years to come when the annual data is complete. One can study the year 2021 and explore how the internet performed in that year in comparison to the previous years. Any other year with a whole data set could be studied.

Finally, If traffic mix and user behaviour for the data sets for this study is made available, one can go ahead and study the usage of the internet and the pandemic's impact on it. Also, if the pandemic should end, one can study the internet performance and usage after the pandemic.

References

[1] Novet Jordan. Zoom has added more videoconferencing users this year than in all of 2019 thanks to coronavirus, February 2020.

[2] Ella Koeze and Nathaniel Popper. The virus changed the way we internet, April 2020.

[3] Venkat Mohan, Y. R. Janardhan Reddy, and K. Kalpana. Active and passive network measurements: A survey. *International Journal of Computer Science and Information Technologies (IJCSIT)*, 2(4):1372–1385, June 2011.

[4] NKOM Nettfart. About nettfart. [Online]. Available:https://nettfart.no/en/om#toc-12.

[5] John Creswell and Vicki Plano Clark. Designing and conducting mixed methods research, 2007.

[6] Abraham Kwame Richman. Coronavirus pandemic influence on internet usage and performance. Project report in TTM4502, Department of Information Security and Communication Technology, NTNU-Norwegian University of Science and Technology, Dec 2020.

[7] Creswell W. John. Research design: Qualitative, quantitative, and mixed methods approaches, 2009.

[8] Fiona Middleton. Reliability vs validity: what's the difference? [Online]. Available:https://www.scribbr.com/methodology/reliability-vs-validity/, June 2021.

[9] Feldmann Anja, Gasser Oliver, Lichtblau Franziska, Pujol Enric, Poese Ingmar, Dietzel Christoph, Wagner Daniel, Wichtlhuber Matthias, Tapiador Juan, Vallina-Rodriguez Narseo, Hohlfeld Oliver, and Smaragdakis. Georgios. The lockdown effect: Implications of the covid-19 pandemic on internet traffic. [Online]. Available:https://doi.org/10.1145/3419394.3423658, October 2020.

[10] Craig Timberg. The washington post: Your internet is working. thank these cold war-era pioneers who designed it to handle almost anything. [Online]. Available:https://www.washingtonpost.com/technology/2020/04/06/your-internet-is-working-thank-these-cold-war-era-pioneers-who-designed-it-handle-almost-anyt April 2020.

[11] Martin Sauter. From gsm to lte-advanced pro and 5g: an introduction to mobile networks and mobile broadband, 2017.

[12] Carin Overturf. How does speedtest measure my network speeds? [Online]. Available:https://help.speedtest.net/hc/en-us/articles/360038679354-How-does-Speedtest-measure-my-network-speeds-, 2020.

[13] BEREC. Net neutrality regulatory assessment methodology. [Online]. Available:https://berec.europa.eu/eng/document_register/subject_matter/berec/regulatory_best_practices/methodologies/7295-berec-net-neutrality-regulatory-assessment-methodology, October 2017.

[14] ITU-T. Terms and definitions related to quality of service and network performance including dependability. *ITU-T Recommendation E.800*, 1(1):003–004, August 1994.

[15] Peder Emsta, Poul Heegard, Bjarne E. Helvik, and Laurent Paquereau. Dependability and performance in information and communication systems. *Department of Telematics, NTNU*, 11(1):019–040, July 2016.

[16] Brunnström Kjell, Beker Sergio Ariel, de Moor Katrien, Dooms Ann, Egger Sebastian, Garcia Marie-Neige, Hossfeld Tobias, Jumisko-Pyykkö Satu, Keimel Christian, Larabi Mohamed-Chaker, and et al. Qualinet white paper on definitions of quality of experience. *Qualinet White papaer*, 1(2):006–007, March 2013.

[17] What is packet loss? packet loss defined, explained, and explored. [Online]. Available:https://www.forcepoint.com/cyber-edu/packet-loss.

[18] Xiaolin Jiang, Hossein Shokri-Ghadikolaei, Gabor Fodor, Eytan Modiano, Zhibo Pang, Michele Zorzi, and Carlo Fischione. Low-latency networking: Where latency lurks and how to tame it, 08 2018.

[19] Minseok Kwon. A tutorial on network latency and its measurements.

[20] ESOA. Latency in communications networks. [Online]. Available:https://www.esoa.net/Resources/1527-ESOA-Latency-Update-Proof4.pdf.

[21] Tony Fortunato. Internet performance testing: Speed vs. quality. *Network Computing*, 1(1):1, May 2017.

[22] Goyal Anshika. Performance of a network. [Online]. Available:https://www.geeksforgeeks.org/performance-of-a-network/, December 2020.

[23] Fastest mobile network, norway: Q2 - q3 2020. [Online]. Available:https://www.speedtest.net/awards/norway/2020/?award_type=carrier&time_period=q2-q3.

[24] Merry Mou. Evaluating a tcp model-based network performance measurement method. *MIT*, 1(1):013–021, May 2017.

[25] Kim Ervasti. A survey on network measurement: Concepts, techniques, and tools, December 2016.

[26] H. Wang, L. Song, H. Chen, and L. Yan. Study on network measurement technologies and performance evaluation methods. *Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC)*, 1(1):0377–0380, July 2013.

[27] Maheen Hasib and John Schormans. Limitations of passive and active measurement methods in packet networks. [Online]. Available:https://www.researchgate.net/publication/228913017_Limitations_of_passive_active_measurement_methods_in_packet_networks, jan 2003.

[28] N. Feamster and J. Livingood. Internet speed measurement: Current challenges and future recommendations. *ArXiv*, abs/1905.02334, 2019.

[29] Carin Overturf. How does speedtest select a server. [Online]. Available:https://help.speedtest.net/hc/en-us/articles/360038679834-How-does-Speedtest-select-a-server-, 2020.

[30] How it works. [Online]. Available:https://speedof.me/howitworks.html.

[31] Clement J. Coronavirus: impact on online usage in the u.s. - statistics & facts. [Online]. Available:https://www.statista.com/topics/6241/coronavirus-impact-on-online-usage-in-the-us/#dossierSummary___chapter1, January 2021.

[32] Thomas Favale, Francesca Soro, Martino Trevisan, Idilio Drago, and Marco Mellia. Campus traffic and e-learning during covid-19 pandemic. [Online]. Available:https://www.sciencedirect.com/science/article/pii/S1389128620306046, 2020.

[33] Massimo Candela, Valerio Luconi, and Alessio Vecchio. Impact of the covid-19 pandemic on the internet latency: A large-scale study. [Online]. Available:https://www.sciencedirect.com/science/article/pii/S1389128620311622, 2020.

[34] Ivankova Nataliya, John Cresswell, and Sheldon Stick. Using mixed-methods sequential explanatory design: From theory to practice. *Sage Journal*, 18(1):003–020, February 2006.

[35] Sami Almalki. Integrating quantitative and qualitative data in mixed methods research—challenges and benefits. *Journal of Education and Learning*, 5(3):001–009, July 2016.

[36] Martin Gunnell. Research methodologies: A comparison of quantitative, qualitative and mixed methods. [Online]. Available:https://www.linkedin.com/pulse/research-methodologies-comparison-quantitative-mixed-methods-gunnell/, dec 2016.

[37] R. Johnson, Anthony Onwuegbuzie, and Lisa Turner. Toward a definition of mixed methods research. journal of mixed methods research, 1, 112-133. *Journal of Mixed Methods Research*, 1:112 –133, Apr 2007.

[38] Abbas Tashakkori and John W. Creswell. Editorial: The new era of mixed methods. *Journal of Mixed Methods Research*, 1(3):001–006, January 2007.

[39] Julia Brannen. Mixed methods research: A discussion paper.

[40] Research guides. [Online]. Available:https://libguides.usc.edu/c.php?g=235034&p=1561756.

[41] Dharmaraja Selvamuthu and Dipayan Das. Introduction to statistical methods, design of experiments and statistical quality control, 2018.

[42] Simer Preet Kaur. Variables in research, December 2013.

[43] Zulfiqar Ali and S Bala Bhaskar. Basic statistical tools in research and data analysis. *Indian Journal of Anaesthesia*, 60(9):662–669, October 2016.

[44] Parampreet Kaur, Jill Stoltzfus, and Vikas Yellapu. Descriptive statistics. *International Journal of Academic Medicine*, 4(1):60–63, 2018.

[45] Central tendency. [Online]. Available:https://corporatefinanceinstitute.com/resources/knowledge/other/central-tendency/.

[46] Professor Tarek Tawfik Amin. Inferential statistics, March 2019.

[47] Prof William M.K. Trochim. Research methods knowledge base. [Online]. Available:https://conjointly.com/kb/inferential-statistics/.

[48] OECD. Glossary of statistical terms. [Online]. Available:https://stats.oecd.org/glossary/detail.asp?ID=2620.

[49] Nigel Mathers, Nick Fox, and Amanda Hunn. Surveys and questionnaires. [Online]. Available:https://www.rds-yh.nihr.ac.uk/wp-content/uploads/2013/05/12_Surveys_and_Questionnaires_Revision_2009.pdf, 2009.

[50] Ross Ihaka and Robert Gentleman. R: A language for data analysis and graphics. [Online]. Available:https://doi.org/10.2307/1390807, September 1996.

[51] Kurt Hornik. R FAQ. [Online]. Available:https://CRAN.R-project.org/doc/FAQ/R-FAQ.html, 2020.

[52] John Chambers. S, r, and data science. *The R Journal*, 11(1), July 2020.

[53] Roger Peng. R programming for data science, December 2019.

[54] Data-Driven Science. Python vs r for data science: And the winner is.. [Online]. Available:https://medium.com/@datadrivenscience/python-vs-r-for-data-science-and-the-winner-is-3ebb1a968197, January 2018.

[55] Edzer Pebesma. Simple Features for R: Standardized Support for Spatial Vector Data. [Online]. Available:https://doi.org/10.32614/RJ-2018-009, 2018.

[56] Mailund Thomas. R data science quick reference. [Online]. Available:https://doi.org/10.1007/978-1-4842-4894-2_7, August 2019.

[57] Garrett Grolemund and Hadley Wickham. Dates and times made easy with lubridate. [Online]. Available:https://www.jstatsoft.org/v40/i03/, 2011.

[58] Wickham Hadley. Read rectangular text data. [Online]. Available:https://readr.tidyverse.org,https://github.com/tidyverse/readr, October 2020.

[59] Hadley Wickham. ggplot2: Elegant graphics for data analysis. [Online]. Available:https://ggplot2.tidyverse.org, 2016.

[60] Wickham Hadley. General network (http/ftp/...) client interface for r. *The R Journal*, 1(9), March 2021.

[61] Jeroen Ooms. The jsonlite package: A practical and consistent mapping between json data and r objects. [Online]. Available:https://arxiv.org/abs/1403.2805, 2014.

[62] RStudio Team. Rstudio: Integrated development environment for r. [Online]. Available:http://www.rstudio.com/, 2020.

[63] Bryn Farnsworth. How to design a questionnaire. [Online]. Available:https://imotions.com/blog/design-a-questionnaire/, September 2019.

[64] Hamed Taherdoost. Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. *International Journal of Academic Research in Management (IJARM)*, 5(3):28 – 36, 2016.

[65] Julie P. Combs. Data analysis in mixed research: A primer. *International Journal of Education*, 3(1):006–009, May 2011.

[66] Anthony J. Onwuegbuzie and Charles Teddlie. A framework for analyzing data in mixed methods research. *Handbook of mixed methods in social and behavioral research*, 2:397–430, 2013.

[67] NKOM Nettfart. Test the speed of your network. [Online]. Available:https://nettfart.no/en/test.

Appendix A

Algorithms

```
# load packages
library(sf)
library(readr)
library(ggplot2)
library(rjson)
library(jsonlite)
library(leaflet)
library(RCurl)
library(DT)
library(dplyr)
library(RColorBrewer)
library(rgdal)
library(rgeos)
library(raster)
library(OpenStreetMap)
library(lubridate)
library(sp)
```

Listing A.1: Loading Packages from repository

```
#Load mobile
setwd("~/4th Semester/Thesis/Data")
Data = read.csv("Data.csv", header = TRUE,  sep = ";")

#Load 2020
setwd("~/4th Semester/Thesis/Data/NETT FART")
year20 = read.csv("nettfart-2020.csv", header = TRUE,  sep = ";",
                  quote = "\"")
year20$Time1 = year20$Time
year2020 = separate(year20, col = Time1, into = c("Date", "Clock"),
                    sep = " ")%>%
  mutate(Date = as.Date(Date), Download = Download/1000,
         upload = upload/1000 ) %>%
  dplyr::select(Download, upload, Ping, Time, Date, Clock, ISP,
                City, County,
                Country, Latitude, Longitude, Platform) %>%
  filter(Country == 'NO')

```

```
#Load 2019
setwd("~/4th Semester/Thesis/Data/NETT FART")
year19 = read.csv("nettfart-2019.csv", header = TRUE,  sep = ";",
                  quote = "\"")
year19$Time1 = year19$Time
year2019 = separate(year19, col = Time1, into = c("Date", "Clock"),
                    sep = " ")%>%
  mutate(Date = as.Date(Date), Download = Download/1000,
         upload = upload/1000) %>%
  dplyr::select(Download, upload, Ping, Time, Date, Clock, ISP,
                City, County,
                Country, Latitude, Longitude, Platform) %>%
  filter(Country == 'NO')

#Load 2018
setwd("~/4th Semester/Thesis/Data/NETT FART")
year18 = read.csv("nettfart-2018.csv", header = TRUE,  sep = ";",
                  quote = "\"")
year18$Time1 = year18$Time
year2018 = separate(year18, col = Time1, into = c("Date", "Clock"),
                    sep = " ")%>%
  mutate(Date = as.Date(Date), Download = Download/1000,
         upload = upload/1000) %>%
  dplyr::select(Download, upload, Ping, Time, Date, Clock, ISP,
                City, County,
                Country, Latitude, Longitude, Platform) %>%
  filter(Country == 'NO')
```

Listing A.2: Loading the CSV data file from the computer

```
########## TRONDHEIM ################
trondheim = year2020 %>%
  filter(City == 'Trondheim')

######### BEFORE CORONA ###########
trondb4Corona = trondheim %>%
  filter(Time <= as.Date("2020-03-11") &
           Time >= as.Date("2020-01-01"))

trondb4Corona <- trondb4Corona %>%
  flatten(recursive = TRUE) # remove the nested data frame

#Using the datatable function from the DT package
#to see the first 6 rows of data
datatable(head(trondb4Corona))

trondMap <- leaflet(year2018) %>%
  addTiles( attribution='10,000 entries for TRONDHEIM,
            NORWAY before March 12th. Map tiles by Kwame') %>%
  addCircleMarkers(~Longitude, ~Latitude,
                   clusterOptions = markerClusterOptions())
trondMap

```

```
######### DURING CORONA #############
trondinCorona = trondheim %>%
  filter(Time <= as.Date("2020-05-31") &
           Time >= as.Date("2020-03-12"))

trondinCorona <- trondinCorona %>%
  flatten(recursive = TRUE) # remove the nested data frame

#Using the datatable function from the DT package
#to see the first 6 rows of data
datatable(head(trondinCorona))

trondMap <- leaflet(trondinCorona) %>%
  addTiles( attribution='17,916 entries for TRONDHEIM,
           NORWAY after March 12th. Map tiles by Kwame') %>%
  addCircleMarkers(~Longitude, ~Latitude,
                   clusterOptions = markerClusterOptions())
trondMap
```

Listing A.3: Mapping algorithm for Figures with maps

```
############ PLOTTING BY COUNTY'S AVERAGE DOWNLOAD SPEED ################
#COUNTIES COUNT
counties = year2020 %>%
  filter(Country == 'NO')%>%
  count(County)
view(counties)

grouped = year2020 %>% group_by(County, Download)

extr_cols = data.frame(County=grouped$County,Download = grouped$Download)

mean_by_county = aggregate(Download ~ County, extr_cols, mean)

ordered_freq = mean_by_county %>% arrange(desc(mean_by_county$Download))

head(ordered_freq)

bar_graph = ggplot(data=ordered_freq, aes(x=reorder(County,Download), y=Download),
                   fill = County)+
  geom_bar(stat='identity', col = "burlywood", fill = "lightgreen",
           width=0.7,position=position_dodge(width=.8)) + coord_flip()+
  geom_text(aes(label = round(Download)), hjust = -0.3, size = 3)

bar_graph + labs(x = 'County', y = 'Average Download Speed (Mbs)',
             title = "Internet Performance based on County (2018)",
             subtitle = "Average Download speed by County from NETTFART",
```

```
               caption = "Graph by Kwame, 2021") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9))


############### PLOTTING BY AVERAGE MONTHLY DOWNLOAD SPEED ####################
download_by_month = table(randaberg$Month_abbr)

neworder = c('Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','Sep','Oct','Nov','Dec')
# # setcolorder(sales_by_month_freq, neworder)

month_data = data.frame(download_by_month)
names(month_data)[1]<-'Month'

bar_graph1 = ggplot(month_data, aes(x=factor(Month, level = neworder), y=Freq),
                     colour = Month)+
  geom_bar(stat='identity', col = "brown", fill = "brown4") +
  geom_text(aes(label = Freq), vjust = -0.5, size = 3)

# add y-label and add Title
bar_graph1 + labs(x = 'Month', y = 'Frequency',
                title = "Monthly Test of Internet Speed Using NETTFART (Randaberg, 2018)",
                subtitle = "Frequency by Month",
                caption = "Graph by Kwame") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9))


#Alternatively
grouped_month = trondheim %>% group_by(Month_abbr, Download)

extra = data.frame(Month_abbr=grouped_month$Month_abbr,Download = grouped_month$Download)

mean_by_month = aggregate(Download ~ Month_abbr, extra, mean)

neworder = c('Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','Sep','Oct','Nov','Dec')

month_data<-data.frame(mean_by_month)
names(month_data)[1]<-'Month'

line_graph1 = month_data %>%
  ggplot(aes(x=factor(Month, level = neworder), y=Download)) +
```

```r
  geom_point(size = 3, alpha = 0.5, col = "burlywood1") +
  geom_line(col = "burlywood", group = 1)

line_graph1 +
  #facet_wrap(~) +
  labs(title = "Internet Performance by Months from NETTFART (Tronheim, 2018)",
       subtitle = "Average Monthly Download Speed results by Month",
       x = "Months",
       y = "Average Download speed (Mbs)",
       caption = "Graph by Kwame") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9))


############### PLOTTING BY AVERAGE DAILY DOWNLOAD SPEED ####################
download_by_day_of_month = table(year$Day)

day_freq = data.frame(download_by_day_of_month)

names(day_freq)[1] = 'Day_of_Month'

bar_graph2 = ggplot(data=day_freq, aes(x=Day_of_Month, y=Freq)) +
  geom_bar(stat='identity',width=0.7,position=position_dodge(width=.8),
           col = "brown", fill = "brown4")+
  theme(text = element_text(size = 15.3))

bar_graph2 + labs(x = 'Day of Month', y = 'Frequency', fill = NULL,
                  title = "Daily Test of Internet Using NETTFART (Randaberg, 2018)",
                  subtitle = "Frequency by Day of Month",
                  caption = "Graph by Kwame, 2021") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9),
        axis.text.x = element_text(angle = 45))

#Alternatively
grouped_day = trondheim %>% group_by(Day, Download)

extra_day = data.frame(Day=grouped_day$Day,Download = grouped_day$Download)

mean_by_day = aggregate(Download ~ Day, extra_day, mean)

day_freq = data.frame(mean_by_day)

names(day_freq)[1] = 'Day_of_Month'
```

```

bar_graph3 = day_freq %>%
  ggplot(aes(x=Day_of_Month, y=Download)) +
  geom_bar(stat='identity',width=0.7,position=position_dodge(width=.8),
           col = "burlywood", fill = "burlywood1")

bar_graph3 + labs(x = 'Day of Month', y = 'Average Download Speed (Mbs)',
                  title = "Daily Internet Performance from NETTFART (Trondheim, 2018)",
                  subtitle = "Average Download Speed by Day of Month",
                  caption = "Graph by Kwame, 2021") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9),
        axis.text.x = element_text(angle = 45))+
  scale_x_continuous(breaks = seq(1, 31, 1))


############### PLOTTING BY HOURLY DOWNLOAD SPEED ####################
download_by_hour_freq = table(randaberg$Hour)

hour_freq = data.frame(download_by_hour_freq)

names(hour_freq)[1] = 'Hour'

bar_graph4 = ggplot(data=hour_freq, aes(x=Hour, y=Freq)) +
  geom_bar(stat='identity',width=0.7,position=position_dodge(width=.8),
           col = "brown", fill = "brown4")

bar_graph4 + labs(x = 'Hour', y = 'Frequency',
                    title = "Hourly Speed Test of Internet Using NETTFART (Randaberg, 2018)",
                    subtitle = "Frequency by Hour (24-Hour)",
                    caption = "Graph by Kwame, 2021") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9),
        axis.text.x = element_text(angle = 45))

#ALTERNATIVELY
grouped_hour = oslo %>% group_by(Hour, Download)

extra_hour = data.frame(Hour=grouped_hour$Hour,Download = grouped_hour$Download)

mean_by_hour = aggregate(Download ~ Hour, extra_hour, mean)

hour_freq = data.frame(mean_by_hour)

```

```r
names(hour_freq)[1] = 'Hour'

bar_graph5 = ggplot(data=hour_freq, aes(x=Hour, y=Download))+
       geom_bar(stat='identity',width=0.7,position=position_dodge(width=.8),
                col = "burlywood", fill = "burlywood1")

bar_graph5 + labs(x = 'Hour', y = 'Avearge Download speed (Mbs) ',
                       title = "Daily Internet Performance from NETTFART (Oslo, 2018)",
                       subtitle = "Hourly Average Download Speed",
                       caption = "Graph by Kwame, 2021") +
       theme_bw() +
       theme(plot.title = element_text(hjust = 0.5),
             plot.subtitle = element_text(hjust = 0.5),
             plot.caption = element_text(hjust = 0.9),
             axis.text.x = element_text(angle = 45))+
       scale_x_continuous(breaks = seq(00, 23, 1))


############### PLOTTING BY WEEKLY DOWNLOAD SPEED ####################
troms$day_of_week = format(troms$Date, format = "%A")

week_days = table(troms$day_of_week)

day_freq = data.frame(week_days)

names(day_freq)[1] = 'Day'
# day_freq

# Need to Use so Columns of Graph are correctly ordered: look at aes(x=)
days_ordered = c('Monday','Tuesday','Wednesday','Thursday','Friday','Saturday','Sunday')

bar_graph6 = ggplot(data=day_freq, aes(x=factor(Day, level = days_ordered), y=Freq), fill = Day)+
  geom_bar(stat ='identity',width=0.7,position=position_dodge(width=.8),
           col = "brown", fill = "brown4") +
  geom_text(aes(label = Freq), vjust = -0.5)

bar_graph6 + labs(x = 'Day of Week', y = 'Frequency',
                  title = "Weekly Speed Test of Internet Using NETTFART (Troms & Finnmark, 2018)",
                  subtitle = "Weekly Frequency",
                  caption = "Graph by Kwame, 2021") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9))
```

```
#ALTERNATIVELY
trondheim$day_of_week = format(trondheim$Date, format = "%A")

grouped_week = trondheim %>% group_by(day_of_week, Download)

extra_week = data.frame(day_of_week=grouped_week$day_of_week,
                        Download = grouped_week$Download)

mean_by_week = aggregate(Download ~ day_of_week, extra_week, mean)

week_freq = data.frame(mean_by_week)

names(week_freq)[1] = 'Day'
# week_freq

# Need to Use so Columns of Graph are correctly ordered: look at aes(x=)
days_ordered = c('Monday','Tuesday','Wednesday','Thursday','Friday','Saturday','Sunday')

bar_graph7 = ggplot(data=week_freq, aes(x=factor(Day, level = days_ordered), y=Download))+
  geom_bar(stat='identity',width=0.7,position=position_dodge(width=.8),
           col = "burlywood", fill = "burlywood1")  +
  geom_text(aes(label = round(Download)), vjust = -0.5)

bar_graph7 + labs(x = 'Day of Week', y = 'Avearge Download Speed (Mbs)',
                  title = "Weekly Internet Performance from NETTFART (Trondheim, 2018)",
                  subtitle = "Weekly Average Download Speed",
                  caption = "Graph by Kwame, 2021") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9))
```

Listing A.4: Download Speed variable Analysis

```
############ PLOTTING BY COUNTY'S AVERAGE SPEED ################
#COUNTIES COUNT
counties = year2020 %>%
  filter(Country == 'NO')%>%
  count(County)
view(counties)

grouped = year2018 %>% group_by(County, upload)

extr_cols = data.frame(County=grouped$County,upload = grouped$upload)

mean_by_county = aggregate(upload ~ County, extr_cols, mean)

ordered_freq = mean_by_county %>% arrange(desc(mean_by_county$upload))
```

```

head(ordered_freq)

bar_graph = ggplot(data=ordered_freq, aes(x=reorder(County,upload), y=upload),
                   fill = County)+
  geom_bar(stat='identity', col = "burlywood", fill = "lightgreen",
           width=0.7,position=position_dodge(width=.8)) + coord_flip() +
  geom_text(aes(label = round(upload)), hjust = -0.3, size = 3)

bar_graph + labs(x = 'County', y = 'Average upload Speed (Mbs)',
                 title = "Internet Performance based on County (2018)",
                 subtitle = "Average upload speed by County from NETTFART",
                 caption = "Graph by Kwame, 2021") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9))


############### PLOTTING BY AVERAGE MONTHLY UPLOAD SPEED ####################
upload_by_month = table(year2020$Month_abbr)

neworder = c('Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','Sep','Oct','Nov','Dec')
# # setcolorder(sales_by_month_freq, neworder)

month_data = data.frame(upload_by_month)
names(month_data)[1]<-'Month'

bar_graph1 = ggplot(month_data, aes(x=factor(Month, level = neworder), y=Freq),
                    colour = Month)+
  geom_bar(stat='identity', col = "brown", fill = "brown4") +
  geom_text(aes(label = Freq), vjust = -0.5, size = 3)

# add y-label and add Title
bar_graph1 + labs(x = 'Month', y = 'Frequency',
                  title = "Monthly Test of Internet Speed Using NETTFART (2018)",
                  subtitle = "Frequency by Month",
                  caption = "Graph by Kwame,2021") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9))


#Alternatively
```

```r
grouped_month = year2018 %>% group_by(Month_abbr, upload)

extra = data.frame(Month_abbr=grouped_month$Month_abbr,upload = grouped_month$upload)

mean_by_month = aggregate(upload ~ Month_abbr, extra, mean)

neworder = c('Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','Sep','Oct','Nov','Dec')

month_data<-data.frame(mean_by_month)
names(month_data)[1]<-'Month'

line_graph1 = month_data %>%
  ggplot(aes(x=factor(Month, level = neworder), y=upload)) +
  geom_point(size = 3, alpha = 0.5, col = "burlywood1") +
  geom_line(col = "burlywood", group = 1)

line_graph1 +
  #facet_wrap(~) +
  labs(title = "Internet Performance by Months from NETTFART (2018)",
       subtitle = "Average Monthly upload Speed results by Month",
       x = "Months",
       y = "Average upload speed (Mbs)",
       caption = "Graph by Kwame") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9))


############### PLOTTING BY AVERAGE DAILY UPLOAD SPEED ####################
upload_by_day_of_month = table(year2018$Day)

day_freq = data.frame(upload_by_day_of_month)

names(day_freq)[1] = 'Day_of_Month'

bar_graph2 = ggplot(data=day_freq, aes(x=Day_of_Month, y=Freq)) +
  geom_bar(stat='identity',width=0.7,position=position_dodge(width=.8),
           col = "brown", fill = "brown4")+
  theme(text = element_text(size = 15.3))

bar_graph2 + labs(x = 'Day of Month', y = 'Frequency', fill = NULL,
                  title = "Daily Test of Internet Using NETTFART (Randaberg, 2018)",
                  subtitle = "Frequency by Day of Month",
                  caption = "Graph by Kwame, 2021") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
```

```
        plot.caption = element_text(hjust = 0.9),
        axis.text.x = element_text(angle = 45))

#Alternatively
grouped_day = trondheim %>% group_by(Day, upload)

extra_day = data.frame(Day=grouped_day$Day,upload = grouped_day$upload)

mean_by_day = aggregate(upload ~ Day, extra_day, mean)

day_freq = data.frame(mean_by_day)

names(day_freq)[1] = 'Day_of_Month'

bar_graph3 = day_freq %>%
  ggplot(aes(x=Day_of_Month, y=upload)) +
  geom_bar(stat='identity',width=0.7,position=position_dodge(width=.8),
           col = "burlywood", fill = "burlywood1")

bar_graph3 + labs(x = 'Day of Month', y = 'Average upload Speed (Mbs)',
                  title = "Daily Internet Performance from NETTFART (Trondheim, 2018)",
                  subtitle = "Average upload Speed by Day of Month",
                  caption = "Graph by Kwame, 2021") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9),
        axis.text.x = element_text(angle = 45))+
  scale_x_continuous(breaks = seq(1, 31, 1))


############### PLOTTING BY HOURLY UPLOAD SPEED ####################
upload_by_hour_freq = table(year2020$Hour)

hour_freq = data.frame(upload_by_hour_freq)

names(hour_freq)[1] = 'Hour'

bar_graph4 = ggplot(data=hour_freq, aes(x=Hour, y=Freq)) +
  geom_bar(stat='identity',width=0.7,position=position_dodge(width=.8),
           col = "brown", fill = "brown4")

bar_graph4 + labs(x = 'Hour', y = 'Frequency',
                  title = "Hourly Speed Test of Internet Using NETTFART (Randaberg, 2018)",
                  subtitle = "Frequency by Hour (24-Hour)",
                  caption = "Graph by Kwame, 2021") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9),
```

```
        axis.text.x = element_text(angle = 45))

#ALTERNATIVELY
grouped_hour = year2018 %>% group_by(Hour, upload)

extra_hour = data.frame(Hour=grouped_hour$Hour,upload = grouped_hour$upload)

mean_by_hour = aggregate(upload ~ Hour, extra_hour, mean)

hour_freq = data.frame(mean_by_hour)

names(hour_freq)[1] = 'Hour'

bar_graph5 = ggplot(data=hour_freq, aes(x=Hour, y=upload))+
  geom_bar(stat='identity',width=0.7,position=position_dodge(width=.8),
           col = "burlywood", fill = "burlywood1")

bar_graph5 + labs(x = 'Hour', y = 'Avearge upload speed (Mbs) ',
                  title = "Daily Internet Performance from NETTFART (Oslo, 2018)",
                  subtitle = "Hourly Average upload Speed",
                  caption = "Graph by Kwame, 2021") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9),
        axis.text.x = element_text(angle = 45))+
  scale_x_continuous(breaks = seq(00, 23, 1))


############### PLOTTING BY WEEKLY UPLOAD SPEED ####################
aprweek4$day_of_week = format(aprweek4$Date, format = "%A")

week_days = table(aprweek4$day_of_week)

day_freq = data.frame(week_days)

names(day_freq)[1] = 'Day'
# day_freq

# Need to Use so Columns of Graph are correctly ordered: look at aes(x=)
days_ordered = c('Monday','Tuesday','Wednesday','Thursday','Friday','Saturday','Sunday')

bar_graph6 = ggplot(data=day_freq, aes(x=factor(Day, level = days_ordered), y=Freq), fill = Day)+
  geom_bar(stat ='identity',width=0.7,position=position_dodge(width=.8),
           col = "brown", fill = "brown4") +
  geom_text(aes(label = Freq), vjust = -0.5)
```

```
bar_graph6 + labs(x = 'Day of Week', y = 'Frequency',
                  title = "Weekly Speed Test of Internet Using NETTFART (2018)",
                  subtitle = "Weekly Frequency",
                  caption = "Graph by Kwame, 2021") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9))

#ALTERNATIVELY
aprweek4$day_of_week = format(aprweek4$Date, format = "%A")

grouped_week = aprweek4 %>% group_by(day_of_week, upload)

extra_week = data.frame(day_of_week=grouped_week$day_of_week,
                        upload = grouped_week$upload)

mean_by_week = aggregate(upload ~ day_of_week, extra_week, mean)

week_freq = data.frame(mean_by_week)

names(week_freq)[1] = 'Day'
# week_freq

# Need to Use so Columns of Graph are correctly ordered: look at aes(x=)
days_ordered = c('Monday','Tuesday','Wednesday','Thursday','Friday','Saturday','Sunday')

bar_graph7 = ggplot(data=week_freq, aes(x=factor(Day, level = days_ordered), y=upload))+
  geom_bar(stat='identity',width=0.7,position=position_dodge(width=.8),
           col = "burlywood", fill = "burlywood1")  +
  geom_text(aes(label = round(upload)), vjust = -0.5)

bar_graph7 + labs(x = 'Day of Week', y = 'Avearge upload Speed (Mbs)',
                  title = "Weekly Internet Performance from NETTFART (2018)",
                  subtitle = "Weekly Average upload Speed",
                  caption = "Graph by Kwame, 2021") +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5),
        plot.caption = element_text(hjust = 0.9))
```

Listing A.5: Upload speed Analysis

Appendix B
Graphs

As stated in Subsection 4.2.3, the following graphs present the results to show that analysing the change in download and upload speed from different perspectives like counties, monthly and hourly produced results similar in nature as the results presented in figure 4.12

Also, when results on latency was presented in subsection 4.2.3, we mentioned that weekly and hourly analysis led to the results of the same nature.

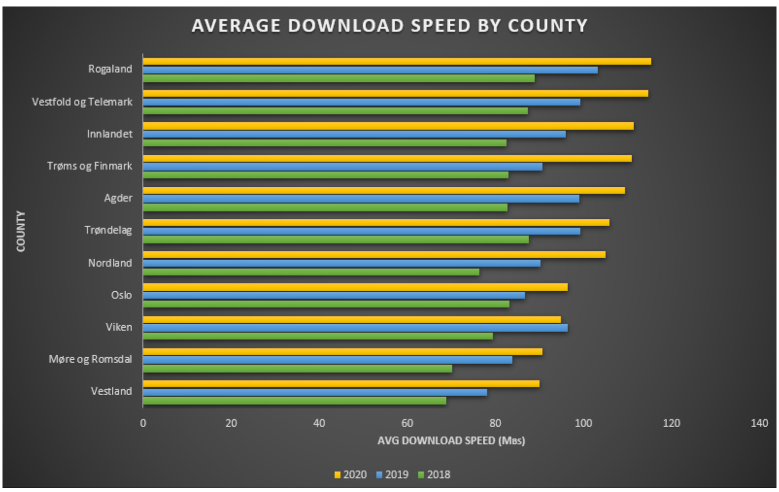

Figure B.1: Change in Average Download speed in the counties

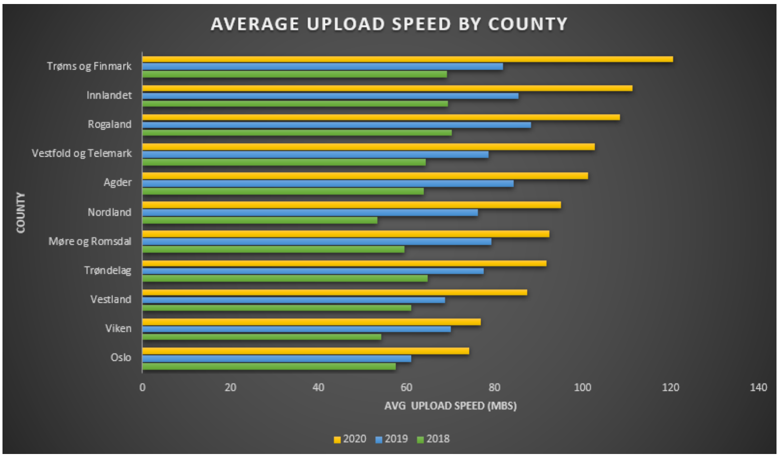

Figure B.2: Change in Average Upload speed in the counties

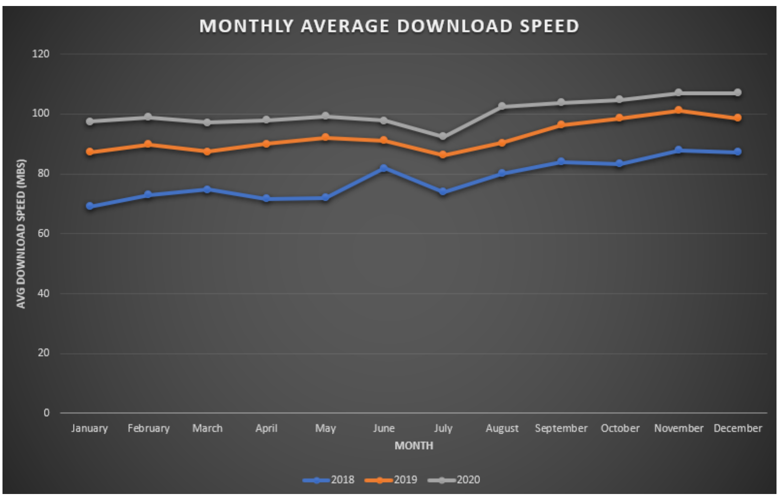

Figure B.3: Change in Average Download speed every month each year

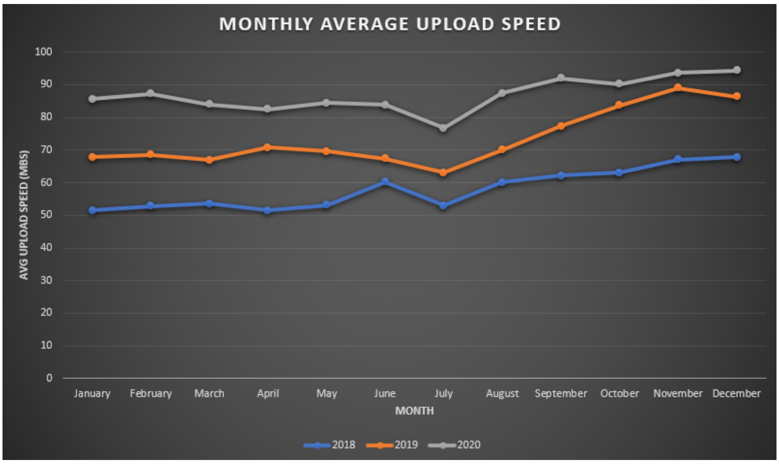

Figure B.4: Change in Average Upload speed every month each year

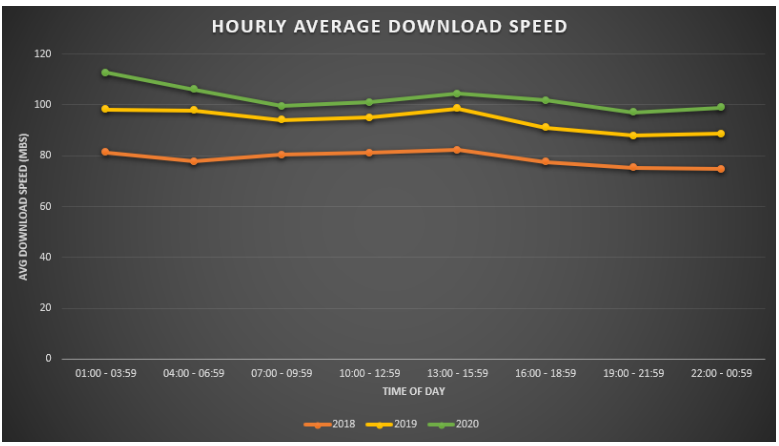

Figure B.5: Change in Average Download speed per hour during the day

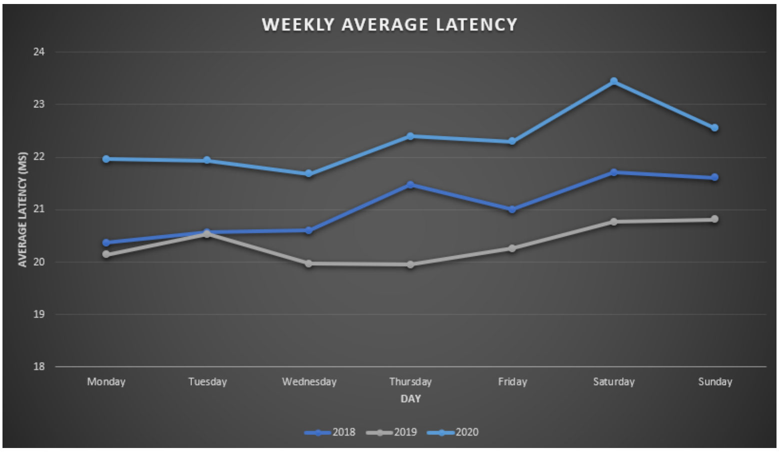

Figure B.6: Change in Average Latency over the week

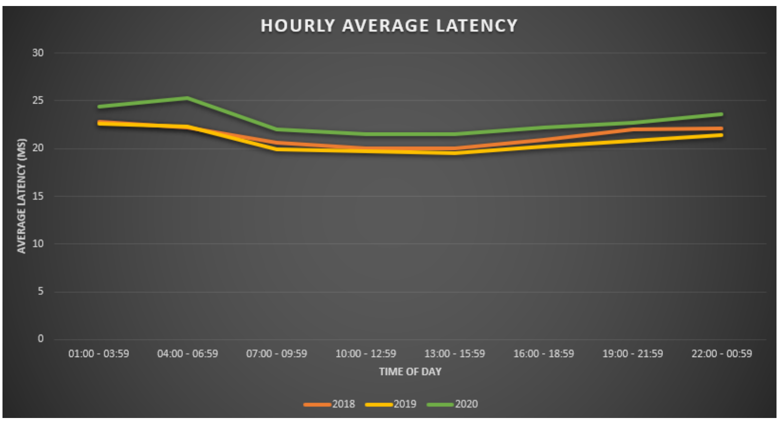

Figure B.7: Change in Average Latency per hour during the day

Appendix C Questionnaire

Internet Speed test and Coronavirus Pandemic

This Survey is to find out IF, WHEN and WHY People may use NETTFART platform (nettfart.no) and/or other speedtest platforms in general to test internet speed at any given time. Also, to know customer experience with internet quality before and during the pandemic. The responses will be used for research purposes for the Master thesis (TTM4905) by ABARAHAM KWAME RICHMAN (Department of Information Security and Communication Technology, NTNU) and are completely anonymous.

1.Since the coronavirus pandemic forced us to work and study from home, have you experienced any Internet quality issues at home, school and office during that period? If yes kindly elaborate.

2.If you have experienced poor Internet quality, was it confirmed by a speed test?

☐ Yes

☐ No

3.What was your reaction? E.g., did you try again shortly? Leave the service? etc.

4.Have you performed an Internet speed test using any of these platforms? (multiple answers is allowed)

☐ NKOM's Nettfart.no

☐ Ookla's Speedtest.net

☐ Netflix Fast.com

☐ Never

☐ Other (Kindly specify below)

5.Please specify if you chose "Other".

6.How do you access your Internet speed test?

☐ Mobile app

☐ Browser

☐ Both

☐ Never

☐ Other (Kindly specify below)

7.Please specify if you chose "Other".

8.Why/ under which circumstance(s) do you usually do a speed test? (multiple answers is allowed)

☐ When internet is bad

☐ Random

☐ When in a new location

☐ Other (Kindly specify below)

9.Please specify if you chose "Other".

10.If you have NEVER performed a speed test, WHY NOT and if you were to test your Internet speed, under which circumstance(s) would you likely do that?

11.How would you rate your overall Internet performance before the pandemic?

☐ 1

☐ 2

☐ 3

☐ 4

☐ 5

12.How would you rate your overall Internet performance for the last 12 months (during the pandemic)?

☐ 1

☐ 2

☐ 3

☐ 4

☐ 5