Østfold University College

MASTER´S THESIS

Study of Social Network Analysis to Support the Old Town Fredrikstad – Smart Tourism

Pooja Shrestha

02-05-2020

Master of Applied Computer Science
Faculty of Computer Sciences

Abstract

User generated content plays an important role in the tourism sector, many researchers are found to be studying the generalizability features of social network analysis. This research studied the existing data on the specific tourist destination on TripAdvisor. We have chosen both qualitative and quantitative analysis method for the research methodology. Big data analytics has been popular among researchers using many data resources and various predictive analysis tools like natural language processing, web textual mining tools to make the data useful and predict about the reality. By focusing on the Old Town of Fredrikstad, we have tried to study the role of web 2.0 and travel 2.0 in tourism marketing of the Old Town. The semi structured interview and predictive analysis methodology has helped to gain insights for this research. The result from the interview and the predictive analysis from the reviews from TripAdvisor showed that according to the visitors that visited the Old Town were quite pleased with the destination. There were many suggestions for the improvements. From the interview result, we could find that most of the local people were happy with the Old Town as it is whereas the businesspersons were willing to focus towards the successful sustainable business. This research will help to make clear interpretation of visitor`s experience, their information sharing behaviour and interest on online travel communities.

Keywords: Social network analysis, sentiment analysis, predictive analysis, smart tourism, data visualization, social media marketing, historic site, online review, big data.

Acknowledgments

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis describes the study of social networking analysis to support the old town of Fredrikstad smart tourism. The main purpose of this study is to discover the possible ways to find the visitor's preferences, opportunities, and challenges found through the analysis process by conducting a network analysis.

Fredrikstad is located in Viken with the total number of 81,139 inhabitants (Bureau, 2018; Tufte, 2001). It is Norway's second-largest city and the seventh-largest municipality (Wikipedia, 2018a). Being more specific towards the old town of Fredrikstad, it has its unique urbanization with well-preserved historical fortifications sites in northern Europe with only 450 inhabitants. It is the oldest part of Fredrikstad inside a bastion fort at the mouth of the Glomma river, which was established on 12th September 1567 in the eastern part of Glomma. King Frederik II signed the city letter and the name after the king became the first Norwegian city founded after the Middle ages. With its unique urbanization and well-preserved historical fortifications of 1660 eras make the old town northernmost well-kept fortress city in northern Europe. It is considered as Norway's first renaissance city where it is known that the Moss convention was signed by Crown Prince Carl Johan on the behalf of the Swedish King (Wikipedia, 2018a, 2018b)

In the old town, there are several small shops selling designed art, crafts, furniture, food and much more (Kommune, 2018). There is a small flea market happening every Saturday in the cosy square of the town. The museum, old town model railway shop, art galleries, restaurants and cafes around the old town has helped to bind the 16th century to the present time. There are many events, concerts happening throughout the year. There are 100 campsites, 24 motel rooms and 18 cabins in the activity town also called Fredrikstad Motel and Camping for guests. There are activities like golf, miniature golf, football golf, tennis and cafe in the same area (Wikipedia, 2018a, 2018b). However, due to multiple changes in the town, several businesses have left the town which has made a significant impact on the city atmosphere. The old town now requires enhancing the tourism strategy making beneficial use of digital resources. It can support businesses and stakeholders of the old town with social network analysis. The result from the analysis and the tourist information can help them in the future for making services more advanced and reliable. This can motivate the sustainable development of the old town. There are diverse kinds of elements that are known to play a significant role in tourism activities like tourists, service-oriented organizations, tourism resources, destinations, and public institutions. So, to understand the core of tourism and its relations to the elements, we need to be able to examine those important elements involving tourism. Implementation of information and communications technology (ICT) and new smart technologies are bound to be helping the tourism industry.

Smart tourism is a new buzzword which defines the increasing reliance of tourism destinations, tourism industries and tourist in emerging forms of ICT allowing big data that can be transformed into valued (Gretzel, Sigala, Xiang, & Koo, 2015). The word "smart" involves technological, economic and social development promoting technologies relying on "*sensors, big data, open data, new ways to connectivity and exchange of information like the Internet of things, RFID and NFC*" as well as ability to infer and reason (Gretzel et al., 2015). Big data is known to be

offering new opportunities in tourism studies with high spatial and temporal data including a larger group of tourists (Salas-Olmedo, Moya-Gómez, García-Palomares, & Gutiérrez, 2018). In recent years, big data is found to be popular with the increasing quantity of data. It covers a wide range of areas like internet searches, bank card transactions, mobile phone activity documents, social networks, water and electricity consumption, meteorological data, images, etc. Big data can give a huge amount of information that can be studied to analyse the tourist's behaviours. The digital "footprint" left behind by tourists during their visit can be helpful for further analysis. The emergence of modern technologies has offered data analysis and real-time forecasting capabilities to the entire travel cycle which has opened immense opportunities for the travel industry. Big data can become a very handy tool in the ongoing battle for a competitive environment and personalization.

Social networking sites provide a powerful source for data collections since such sites store demographic information about people, their relationships with other users in the same site and their ranking, reviews or even rating information depending on the nature of the sites (Alfantoukh & Durresi, 2014). Garton, Haythornthwaite, and Wellman (1997) defined social networking as a group of people, organization or social entities that are connected through social relationships, co-working and information exchange. The use of web 2.0 widely through its boundless networking possibilities had extended social networking to a greater future. Social networking analysis is described as an analysis of characteristics of the pattern of distribution that includes relational ties and drawing of inferences about the network as a whole or those belonging to individuals or groups (Hanneman & Riddle, 2005).

The effective network visualization is based upon the deep knowledge and care about the substance, quality and integrity of the content(Tufte, 2001). Informative visualization is defined as pragmatic and artistic means of expressions (Lei, Ni, Zhu, & Zhang, 2018). Data visualization is described as any effort for helping people understand the significance of data with the help of visual context. It is the technique to create images, diagrams, graphs, animations to express messages and ideas. There are different types of visualization like 1D or linear, 2D or Planar, 3D or volumetric, temporal, multidimensional, tree or hierarchical, network. There are two type of data visualization which are a) exploration: explore the data to find the story and b) explanation: tells story to audience about the data.

The pictorial representation of information and data can be done through charts, graphs and maps. These visualizations provide access to understand latest trends, outliers and patterns through data. Processing information by using charts or graphs can help to visualize large and complex data and networks. This can show important areas which require "*attention and improvement, clarify the beneficial factors to influence customer behaviour, helps companies to understand which products to place where and predict sales volumes*" (SAS, 2019).

1.1 Background, Problem Statement & Motivation

The emergence of Web2.0 and Travel 2.0 has planted the concept of social networking firmly into the tourism industry (Y. Huang, Basu, & Hsu, 2010). An increasing number of social networking sites and a higher level of user interaction are required to be studied at a deeper level. Smart technologies have touched all of us in some way or the other. Smart technologies are often unnoticed and taken for granted by society. In the last few years, tourist destinations have also started to implement smart technologies within its operation to enhance the visitor experience. In the competitive edge of today's business, it has become important to understand travellers and their views. This can support the travel industry to predict real customers, their travel intentions and tailor the best offers to them.

Larger number of networks can be found everywhere through platforms like social networks, connection among people for example: kinship relation, friendship, trade between

organizations or countries; citation and co-authorship networks. The data obtained from web science; telephone calls; flow charts in computer science; organic molecule in chemistry for example: DNA protein its interaction network; connection between words in text or dictionaries; transportation network for example airlines, maps etc (Mrvar & Batagelj, 2016). Big data can bring expected impact on sales and marketing strategy by analysing the view of new markets, their expansions or propose the latest trends by delivering relevant and right information. The use of previous travel data sources can enable better prediction of the future and take proper actions. The enriched analysis can increase in sale opportunities with the combination of traveller related information like social media content, markets and propose a personalized offer to actual shoppers giving benefit to both business and customers.

The terminology "Smart City" was coined in the nineties of the twentieth century by the Californian Institute for SMART communities (Albino, Berardi, & Dangelico, 2015). Basically, this expression involves six different characteristics it focuses on smart economy, smart people, smart governance, smart mobility, smart environments and smart living (Giffinger, Fertner, Kramar, & Meijers, 2007). A Smart City is defined as "*a city well performing in a forward-looking way in these six characteristics, built on the 'smart' combination of endowments and activities of self-decisive, independent and aware citizens* (Giffinger et al., 2007). Going back, "smart" has become a rather common term in marketing which is enhanced by technology and technology has become essential to the tourism industry over the last few years, not just regarding ways of communication, but also in terms of booking procedures and information searches, etc. Smartness is usually closely related to enhancing the usability of certain products by implementing technologies.

The tourism sector has been contributing to the economic growth globally with its direct contribution of travel and tourism to global GDP with 2,893 billion US Dollar in 2019 and overall GDP worldwide contribution was over 9,258 billion US Dollar(S.Lock, 2020a).

From the report published by Lauren Uppink Calderwood (2019), Norway is known to be Northern Europe`s largest and most competitive travel and tourism economy but declined from 18 to 20th rank globally. Due to best quality of air transport infrastructure(19th), Norway has been ranked 5th globally. It is also ranked 4th in high airport density and 28th in the number of operating carriers. It was ranked 2nd for environmental sustainability worldwide.

According to the survey report of Innovation Norway (Norway, 2018), the tourists including both international and national people had total expenditure of 103.7 million NOK spent in overnight stays in Norway in 2018. 128.8 billion NOK expenditure was directly connected to the total amount of expenditure spent by tourist during their holiday in Norway. The international tourist overnight stays related to holiday was 9.7 million whereas 1.6 million was for business related travel. Most of the Norwegians are found to be staying overnight stays for holiday related. So, the total of 78.1 million overnight stays were holiday related and 14.3 million was business related stays. The southern part of Norway is found to be having 6.3 million overnight stays and have 11 billion NOK expenditure. The 49% of the stay was non rented type of accommodations like private cabins, staying with friends/family, tents etc, 38% of rented accommodation was used like hostels, rented cabin, camp sites, Airbnb, Hurtigruten and only 13% of stay was in hotel, motels. The eastern part of Norway excluding Oslo and Akershus was 25 million overnight stays and 23.6 billion NOK expenditure. Oslo and Akershus had 11.1 million overnight stays and 14 billion NOK expenditure in 2018 (Norway, 2018).

It is found that every 10th dollar spent, and quarter billion jobs are linked to the travel and tourism industry. According to the statistics from World Tourism Organization (Organization, 2018). In 2017 it is estimated that 1.32 billion international tourists had travelled around the world. This denotes a 7% increase in trips across the national border by 83 million (Organization, 2018). The statistic shows the robustness of the tourism industry is increasing despite problems like economic conditions, political unrest, terrorism, and natural disasters. People are found to be

more interested in visiting new places, tasting new food and traditions. In the year 2017, showed that tourists have concentrated travelling to see the fjords of Norway and northern part of Norway(visitnorway, 2017) .

Some of the benefits from tourism are greater prosperity, economic growth, and employment but the economic benefit is not found to be shared equally throughout the world. Similarly, tourism in Norway is traditionally found to be seasonal from May to August which is the summer months (visitnorway, 2017). The main challenge faced by Norway is a limited number of destinations with a larger number of tourists only visiting a few weeks in a year. The largest increase in overnight stays were from South Korea by 37% and USA by 35%.

Norway is found to be a highly competitive and safest country for traveling in the world(Norway, 2017). The lower value of Norwegian krone has been beneficial for tourists to get more value for their money. The tourists visiting Norway are found to be active ones rather than traditional ones as most of them are here to enjoy and experience fjords and mountains, visit historical sites, try local food and drinks and know the local lifestyle of people and their culture.

According to a report by Innovation Norway (Norway, 2017) there was a 16 % growth in online product search in 2017 whereas 8.68 million tourism-related searches were made for exclusively Norway (visitnorway, 2017). The most searches were done for Germany, the USA, and the United Kingdom. The main keywords specially used for Norway were "*fjords, glaciers and the northern lights*" following the various types of accommodations, cities and outdoor activities (visitnorway, 2017).

According to an interview conducted by innovation Norway (Norway, 2017) in the summer season of 2018 throughout the country, the research found the highest number of tourists came from the following countries: Germany, Sweden, USA, other outside Europe, Denmark, United Kingdom, other Europe, Netherland, China/Hongkong/Macau. The Germans are found to be in top list spending long holidays with 1.7 million overnight stays and 2.3 billion NOK expenditures(Norway, 2018).

Another researcher Xie (2020) has explored the economic determinants of seasonal tourist concentration to fulfil the gap in tourism literature. The research explains that Asia has been one of the most growing tourism markets for Norwegian tourism where China and Japan are the important source of markets. The result from the research found that there are three factors affecting tourism in Norway as climate and institutional reason like June, July being the exceptional season with school and institutional holidays whereas economic factors like income, (Consumer Price index)CPI and exchange rate also affect tourism seasonality.

Another researcher Heslinga, Hartman, and Wielenga (2019) has described that beautiful nature and scenic landscapes like fjords and glaciers contribute in popularity as a holiday destination in Norway. Various places like "*Tromsø, Trolltunga, Preikestolen, Lofoten, Geirangerfjørd and Sognefjørd*" are highly experiencing tourism growth. The "allemansretten" or right to roam around which is outdoor recreation right has also been one of the advantages to tourists visiting Norway.

The oldest and natural way for intercultural communication is through travelling(Prameswari, Surjandari, & Laoh, 2017). It provides stories about the places, know about inhabitants and their culture which has developed the richer and diverse genre known as travelogue online. With evolving technology and communication developments has brought explosion of tourism industry and increased accessibility of traveling foreign to wider domain. This has led to increase in popular interest by emerging online tourism forums. These forums has become the primary tools for tourists to find travel information. With increasing number of travellers has created large input of online data sharing their experiences, images, ratings etc. There are several travel forum websites like TripAdvisor, Yelp, Citysearch, virtual tour etc for travel related discussion as well as a place for travellers to view and write reviews. These travel forums are popular worldwide as google search yields 542 million results. TripAdvisor is found

first in the list with keyword as "Travel forums". The new formats of writing and sharing travelling information has attracted marketing experts. Leisure and entertainment while travelling bring richer source of big data describing tourist`s preferences, their behaviour, satisfaction or dissatisfactions and suggestions for future.

Similarly, we have used TripAdvisor page to view the network of people interacting in official website of Fredrikstad tourism to find out leading network of people, their topics of interest so that the office can improve their marketing strategy and use the social media platform for better performance. As many researcher are concerned about the electronic word-of-mouth which can directly affect the image of destination and tourist`s travel decision (Atay & D'Silva, 2019). The study of these travel forums can be adapted to tourism management strategies, help the future prediction for beneficial investment on the long run, have better control in competitive hospitality market, increase interaction and attract more tourists to old city of Fredrikstad.

There are several benefits of predictive analysis for this project. Networked organizations can share unified purposes, clear coordination structures, consensus for alike interests, strengthen in numbers, better representation, interactive communications in both horizontal and vertical dimensions, increase in dynamism and creativity due to clear communications. This can help to influence larger social network and access diverse forms of capital, information and knowledge content about social networks. Social network focuses on the structure of relationships among set of actors. Mapping and measuring both formal and informal relationships can identify what can "*facilitate or impedes the information and knowledge flow that binds interacting units, visualize, who knows whom and who shares what information and knowledge with whom through what media*" (Serrat, 2015). It is to recognise a community by mapping the relationships that connects them as network, try to draw out main individuals, various groups within the network (important components) or association between the individuals. Practicality of the approach provides an objective to find replicable representation of community applying SNA tool.

SNA shows the structural gaps and social capital in the organization, focusing manager`s attention on critical informal networks (Aydin, 2018). Evaluating strategically important networks inside the organization and focus in invisible groups, make it visible in interaction and work with effective collaboration. Targeting responses by the process of mapping the important networks, links can allow to tailor to specific individuals, events that may be associated to tourism attraction for example core network and peripheral networks. Multiple uses like visualizing the dataset through many analytical approaches can help to resolve local issues and promote tourism.

1.2 Research objectives and questions

The main aim of the thesis is to study social networking analysis and the issues related to tourism marketing of the old town of Fredrikstad. In order to accomplish the goal, we have developed two main research questions and three sub research questions in order to find the results according to our research plan. The questions are listed as following:

RQ1: How can an analysis of the social network support the old town of Fredrikstad with the availability of big data service?

In RQ1, we have a focus on the implementation of social network analysis related to smart tourism using big data from social networking sites.

RQ2: How will social media analysis help the management of Visit Fredrikstad and Hvaler in their daily operations?

In RQ2, we are interested in finding the best ways to manage tourism marketing with visit Fredrikstad and Hvaler with long term use of social media analysis tools.

1.3 Overview of methodology

After the systematic literature review, we will be performing a mixed method. For conducting the quantitative methodology, we have selected social network analysis. The analysis of available social media tourism data pertaining specifically to the old town of Fredrikstad. The tourist authorities use four social media platforms: Facebook, Twitter, YouTube and flicker for their tourism promotion. They also have its own official website providing various information about the old town and current events. In addition, interviews will be conducted with the stakeholders to get first-hand feedback. Mixed method research is known as the research where the researcher collects and analyses data, integrates the findings and draws inferences using both qualitative and quantitative design (Rittichainuwat & Rattanaphinanchai, 2015). Using this method can provide better insights into the data analysis.

After the systematic literature review, we will be performing a mixed method. For conducting the quantitative methodology, we have selected predictive analysis under which we have chosen social network analysis and sentimental analysis. The analysis of available social media tourism data pertaining specifically to the old town of Fredrikstad. The tourist authorities have used four social media platforms: Facebook, Twitter, YouTube and Flicker for their tourism promotion. They also have its own official website providing various information about the old town and current events. In addition, qualitative methodology has been used with semi structured interviews will be conducted with the participant in Old Town to get first-hand feedback. Mixed method research is known as the research where the researcher collects and analyses data, integrates the findings and draws inferences using both qualitative and quantitative design (Rittichainuwat & Rattanaphinanchai, 2015). Using this method can provide better insights into the data analysis.

1.4 Relevance, significance and audience

The result of this study will be beneficial to the tourist authorities. This study can elevate the way of social media marketing for tourism and attract more visitors. This will be profitable to find important networks and ways to invest in long term tourism marketing. The research has studied significant factors that are suggested to attract tourist towards old town of Fredrikstad. Professional social media usage can lift the information about the old town. The new generations search for popular and recommended travel destinations so, usage of social networking marketing strategy has become a necessity today.

This study can also be beneficial to the municipality. As the municipality also plays an important role in the investment and development of its city, it can view the important side of social media marketing.

The local business can use the study outcome for their strategic marketing use. They can view the importance of social media and its usage for business marketing.

1.5 Assumptions, limitations and delimitations

Assumptions: Assumption for this research is that the data imported from TripAdvisor consist of a real person. The data posted by users online are assumed to be true. The data collected from the interviewee is assumed to reflect the general population.

Limitations: There are a few limitations we have found for this thesis project. We are analysing the social network activities, not the people and their behaviour. There are limited data sets with no knowledge of real users due to privacy purpose and could potentially be a fake ID used by the same person. There are two methodologies both qualitative and quantitative methodologies chosen for the analysis process. Recently, there has been changes implemented in the API of social media imposes a limitation to its data being scraped at once.

Delimitations: The delimitation of this research is to focus on old town of Fredrikstad rather than a larger tourist destination. The problem identified in this beautiful historical old town is to save tourism which is found to be interesting to us. There many other social media platforms which could be chosen for the data analysis, but we have restricted it to TripAdvisor. There are many ways of analysing data, but we have chosen predictive analytics particularly analysis of the social network followed by a sentiment analysis and an interview. Choosing both quantitative and qualitative research method for this research can help to elevate the problem identified and provide a better end solution.

1.6 Definition of terms

The following terms have been introduced throughout this research. To avoid miscommunication, this section defined the terminologies introduced throughout this research.

- **Big Data:** Big data is a term that describes the large volumes of data "large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information(Gandomi & Haider, 2015).
- **ICT:** Information and Communications Technology (ICT) will coordinate all activities and services, leading to connected, better informed and engaged tourists. ICT makes cities more accessible and enjoyable for both residents and visitors through interactive service interconnecting all local organizations to provide real-time services and use data centrally for better coordination (Buhalis & Amaranggana, 2013).
- **Social Network:** It is a social structure made of individuals or organizations known as "nodes" which are connected by many specific types of interdependency like friends, people with a common interest, dislikes, knowledge, beliefs. A connected group of individual agents who make production and consumption decisions based on the actions (signals) of other agents on the social network; not necessarily (e.g., family, friends, and colleagues), not necessarily regular, but may hold hubs, weak and strong connections, and close and distant connections. Furthermore, agents may exhibit significant heterogeneity with respect to their connections in social network (Potts, Cunningham, Hartley, & Ormerod, 2008).
- **Social networking sites**: as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site (Boyd & Ellison, 2007).
- **Social Network Analysis:** is the process of investigating social structures through the use of networks and graph theory. It characterizes networked structures in terms of nodes (individual actors, people, or things within the network) and the ties, edges, or links (relationships or interactions) that connect them (Otte & Rousseau, 2002).

- **Sentimental Analysis:** is a natural language processing tool useful for monitoring web 2.0 applications for identification and categorization of opinions expressed in a piece of texts which is generally positive, negative or neutral (A. Valdivia, Luzón, & Herrera, 2017).
- **Smart City:** involves six different characteristics where the focus is on: smart economy, smart people, smart governance, smart mobility, smart environment and smart living (Giffinger et al., 2007). A Smart City is defined as *"a city well performing in a forward-looking way in these six characteristics, built on the 'smart' combination of endowments and activities of self-decisive, independent and aware citizens".*
- **Smart destination**: is considered an innovated tourist destination. In more detail it is defined as "*an innovative tourist destination, built on infrastructure of state of art technology guaranteeing the sustainable development of tourist areas, accessible to everyone, which facilitates the visitor's interaction with and integration into his or her surrounding, increases the quality of the experience at destinations and improves resident's quality of life*" (Lopez de Avila, 2015).

1.7 Outline of the Thesis

The thesis is divided into five main chapters: introduction; literature review; methodology; results, and conclusion, implications, recommendations, and summary. Here, the main chapters are briefly explained.

Chapter 1: Introduction – The introduction chapter includes main aim of this thesis. It describes the importance of SNA and its impact on tourism. In succession, the social network analysis approach and the research questions are proposed for further study.

Chapter 2: Literature Review – This chapter establishes the research criteria to study the broader area and theory for the topic. We have focused on main research element, common terms, existing research methods and current measures taken by SNA community related to this research.

Chapter 3: Methodology – This chapter presents the overall process and research design used in the thesis. The methodology chosen for data collection and manipulation is also found in this chapter.

Chapter 4: Results – This chapter includes representation of data after its interpretation. The analysis of data and its result are described in detail.

Chapter 5: Conclusion, Implications, Recommendations, and Summary – In this final chapter, based on the findings from above chapter, conclusion of the research is drawn.

1.8 Chapter Summary

This chapter established the foundation for this research. It further describes background of the real problem, research objectives and research question. The methodology chosen in the area of predictive analysis is described and its importance in e-tourisms. This is followed by relevance, significance and the targeted audience for the research. This chapter also includes assumption, limitations and delimitations of the research. It concludes by the definition of common terms used in the research.

Chapter 2

Literature Review

This study focuses on investigating the impact predictive analysis has on smart tourism. The literature study will comprise of identification of relevant researches to social media, social network analysis, sentimental analysis that has already been conducted.

2.1 Review scope

A systematic literature review based on the guideline of Kitchenham has been applied (Brereton, Kitchenham, Budgen, Turner, & Khalil, 2007). The systematic literature review is "*systematic, explicit and reproducible methodology for identifying, evaluation and synthesizing the existing body of completed work produced by researchers, scholars and practioners*" (Okoli & Schabram, 2010). A literature review requires to be systematic in its methodological approach, explicit to explain the procedures by which the research was conducted, comprehensive in its scope of including all the relevant materials and finally reproducible to others who would like to follow the same approach in the research (Okoli & Schabram, 2010).

The main purpose of the literature review is to include the theoretical background of similar researches conducted, learning the methodology followed by the research on similar topics and even answering the questions in existing analyses on the topic (Okoli & Schabram, 2010).

The main goal of a systematic literature review is to study existing works of literature on smart tourism and its relation towards social network analysis. For this research, we have chosen a systematic literature review (SLR) to achieve a comprehensive overview of the topic. The SLR is defined as literature surveys with defined research questions, search process, data extraction, and data representation, whether or not the researchers referred to their study as systematic literature review (Kitchenham et al., 2009). The process was followed using the reference manual Brereton et.al. Brereton et al. (2007).

2.2 Search terms

The search term helps to identify all the important keywords that have been used in the RQ. The synonyms and other used words are also included in the search terms. The two Boolean operators found in the search category as "AND" and "OR" were also applied to search for keywords that could be used together. The quoted keywords also helped to find specific texts during the search. The final keywords used in this research are listed below.

"Smart city" OR "smart tourism destinations" OR
"Smart tourist attraction" AND "Social network analysis
" AND "Sentimental analysis"
OR "online review"

2.3 Search Procedure

The search procedure contains four steps for the selection of the best literature articles. The first step is the identification of the key concept and terms in the selected database. We have selected five database sources as **ACM Digital library, IEEE, Science Direct, Springer, Semantic scholar and Taylor & Francis Online**. Under the advanced search option, we executed the search. The second step was to select relevant papers according to the title and keywords, disqualifying the irrelevant papers. The third step was to include the papers that were describing key concepts terms in their abstract and exclude the others as further refinement. Finally, the papers were read in detail with full text for the final selection. The overall process is described in the figure below:
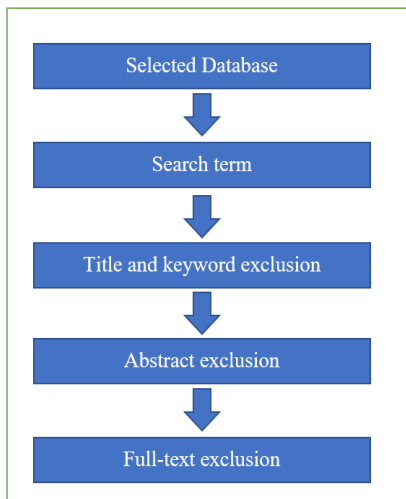
Figure 2. 1: Literature Search Procedure

2.4 Inclusion and Exclusion Criteria

After the implementation of the initial searching procedure, the selected papers were selected through both inclusion and exclusion criteria. The criteria are explained in detail as below:

Inclusion criteria
- Papers that explicitly include smart cities, smart tourism
- Papers that describe smart tourism and social network analysis
- Accessible papers
- Papers that are from 2010 to 2020

Exclusion criteria
- Papers that do not contain smart city and smart tourism
- Papers that do not describe smart tourism and social network analysis
- Paper that was not accessible

2.5 Data sources and analysis process

The data extraction process was carried out by following the guideline provided by researchers Brereton et al. (2007). The data has been documented in a form to record the information found after the primary study and data described to answer the RQs. The information displayed in the table includes the following details:

- The title of the paper.
- The source of paper.
- Criteria met with title and keyword filter: The selected paper must satisfy the first filter and is indicated by Yes/No field.
- Criteria met with abstract filter: The selected paper must then satisfy the second filter and is indicated by Yes/No field.
- Criteria met with fill text filter: The selected paper must then further satisfy the third filter.
- Accessibility of the paper: Must be accessible.
- Notes: Verifying the exclusion or inclusion of papers in detail description.
- The selected papers that successfully went through all three filtrations with a Yes field is then moved to a new sheet to extract the necessary data that addresses the RQs. The definitions, characteristics, and social network analysis.

2.6 Results of the structured literature review (SLR)

From the search string, a total of 10494 papers were retrieved. 10186 papers were excluded after the title and keywords were checked. Once the abstract was confirmed the search procedure reduced the total to 85 and after the full text was gone through the final applicable papers were 66.

Library	Initial results without filtration	Title & Keyword selected	Abstract selected	Full text selected
IEEE	433	44	10	5
ScienceDirect	1768	39	31	29
Springer	220	64	11	4
ACM	741	50	15	8
Taylor & Francis Online	1624	61	10	4
Semantic Scholar	5708	50	9	6
Websites				10
Total	10494	308	85	66

Table 2. 1: Paper filtering phase

2.7 Results and Findings

The articles found were split into the aforementioned RQs to ensure literature were found for both questions.

Research questions	Publications
RQ1	6,35,40,41,42,43,44,45,46,47,48,49,50,51,52,53,67,68,69,70,71,72,73,74,75,76
RQ2	7,13,54,55,56,57,58,59,60,61,62,63,64,65,66,77,78,79,80,81,82,83,84,85,86,87, 88, 89,90,91,92,93,94,95,96,97,98,99,100

Table 2. 2: Studies in detail according to research question

The table below displays the results of the final literature review where the articles have been divided into various categories pertaining to the topic.

Research terms	Publications	Total articles
Role of web tourism	**6,35,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56**	19
Social networking sites	**7,57,58,59,60,61,62, 63,64,65,66,67,68,69**	14
E-tourism	_7,45,46,69,_ **70,71,72,73,74,75,76,77,78,79**	10
Social network analysis	**13,**_44,45_**,80, 81,82,83,84, 85,86,87,88**	10
Sentimental Analysis	**89, 90, 91,92,93,94,95,96,97, 98,99,100,101**	13
Total		66

Table 2. 3: Studies in detail according to research categories

2.8 Related work

In this section the related work is addressed. It has been divided into these four areas: a) role of Web 3.0 in e-tourism, b) Social networking sites, c) e-tourism, and d) social network analysis.

2.8.1 Role of Web 3.0 in e-tourism (Digital approach in tourism)

The world wide web(www) is known as the most prominent part of the internet which is defined as a techno-social system that interacts humans based on technological networks (Aghaei, Nematbakhsh, & Farsani, 2012). The techno-social system is a system that enhances human cognition, communication, and cooperation (Patel, 2013). There has been much progress and further advancement in technologies in the past three decades.

Web 1.0 is a web of information connections known as world wide web from era 1996-2004 (Kujur, 2015). The founder was Tim Berners Lee. It was one directional with read only system. It had roughly 45 million of users in 1996(Kujur). Netscape navigator was a browser developed for web 1.0 concept. Travel 1.0 was conducting offline operations where authors could only change the contents.

Web 2.0 was introduced by Tim O`reilly Dale Dougherty (O'Reilly, 2012). Its era was from 2004 to 2016 known as the social web. It was both read and write web, where participation was allowed to interact as bi-directional. The web had 1 billion of users by 2006. Blogs and social media page like Facebook and YouTube are some examples. The internet has brought an enormous revolution which changed the world economy and social life of people(Roberta Milano 2011). It further has changed the way we communicate with others, work and even conduct business.

Web 2.0 even has larger impact to general activity of human where tourism could not remain unnoticed. Tourism had important transformation as it also changed into Travel 2.0 as "*touristic version of web 2.0*"(Roberta Milano 2011). The boundary between information producer and customer has been removed in this era. Shifting from Travel 1.0 to 2.0 has enabled exclusive online operation experience enabling travellers to book, shop and leave reviews, comments and build travel itineraries via www. It is found that social media has left tourism marketers to effectively compete for consumer`s attention(Xiang & Gretzel, 2010). Web 2.0 does not only include content consumers but also allow users to generate content and collaborate with other users so that users are involved as active role and can create a virtual community. It includes websites like "*blogs-WordPress, blogger, media content-Prezi, YouTube, Flickr, wikis-Wikipedia, wikispace, collaboration-Dropbox, google docs and social networks like twitter, Facebook, google+*"(Del Chiappa & Baggio, 2015). Users are found to be checking for opinions shared by other users in forum, blogs Nd social networks before planning to travel, purchase a product or use services (A. Valdivia et al., 2017).

Web 3.0 is a semantic web introduced by Tim Berners Lee. Its era has started from 2016. It was an executable web with a multi-user virtual environment which has trillions of users. Web 3.0 can improve "*data management, support accessibility of mobile internet, stimulate creativity and innovation, encourage globalization phenomena, enhance customer satisfaction and assist collaboration of social web*" (Aghaei et al., 2012). This research verified the impact of two online social networking sites, Facebook and twitter.

Many tourism organizations are found to be adapting their approach to present themselves online due to the importance of interactive tools and features of Travel 2.0 focusing in social media platform (Schegg, Liebrich, Scaglione, & Ahmad, 2008; Yu, 2010). Most of the researches has been done on the social and psychological effects which play vital role in the source of information and various issues related to transportation and accommodation (Chung & Buhalis, 2008; Inversini & Buhalis, 2009; Kasavana, Nusair, & Teodosic, 2010).

The researcher Kasavana et al. (2010) studied the "evolution, scope and types of social networks" and its potential implications for hospitality businesses. It used in depth literature examining intended and unintended significances in social media. The findings of the study were social networking can be helpful in assisting tourism business by strengthening loyalty and satisfaction of guests. But exposure to social media were found to have disadvantages like biased criticism and baseless speculations. Some companies were found using sponsor websites to try incentivising interactivity by "*peer to peer and staff to administrator participation*" which risked in having negative relations and stressed working conditions.

Another research Parra-López, Bulchand-Gidumal, Gutiérrez-Taño, and Díaz-Armas (2011) proposed theoretical model explaining the factors defining the intention to use social media for organizing or having vacation trips. The understanding of antecedents of tourists is found to be valuable for organizations and destination policy makers. The theoretical model and its hypothesis were tested based on structural equations with PLS techniques. The study had 404 internets users' sample who travelled in last 12 months. The result found that use of social media directly affected the perceived benefits of that use in fictional, psychological, hedonic and social ways however cost did not affect predisposition to such technologies. It also showed that the incentives like altruism, availability, trust in other`s contribution facilitate and promote social media technology while organizing and taking trips.

The research Schmallegger and Carson (2008) has examined how web 2.0 web applications and the increasing number of travel blogs influence the five key function of internet contribution as "*promotion, product distribution, communication, management and research*". For the study they used published literatures and real-life examples in destination marketing organizations and tourism enterprises using blogs as a part of their marketing strategy. The result found that there are numerous advantages of web-based content like blogs. They were easily

updated, flexible in structure, encourage interaction among user and author, allow users to participate in sharing information's.

However, there were few challenges found in effectively using blogs for marketing business in travel and tourism (Gretzel et al., 2015). Some of the challenges were business require substantial amount of time for incorporating blogs, creating and maintaining destinations require regular attentions, credibility and capacity to competing market can be compromised if it is outdated. Building creative and interesting contents, responding to negative comments and incorrect information can be challenging too(Xiang, Magnini, & Fesenmaier, 2015). Unclear limitation to how corporate can response and insufficient researches to which blog can be connected to other resources like booking engines, corporate websites is unknown side. The customer can perceive information directly from the blogs. On the other hand, it can be undesirable interference in consumer to consumer information exchange process. The research also suggested the most useful application of travel blog is found to be monitorization of customer`s attitudes.

The researcher Chung and Buhalis (2008) examines the connection between perceived benefits and participation in an online travel community through social networking sites to find what makes the users participate in social networks. There were three factors that was revealed from the result as "*information acquisition, social psychological and hedonic*". They were found to be influencing user`s participation and the attitude toward online travelling community. With the use of multiple regression analysis showed information acquisition as the most important factors.

Web 4.0 is the web of integration as the fourth generation of the web to this date. The basic idea was to define data structure and link them to more effective discovery, automation, integration and reuse in other applications to obtain additional information. Web 4.0 is still known as an underground idea and no exact definition of how it would be. But it is found to be moving toward using artificial intelligence as intelligent web making machines clever enough to execute website with superior quality and performance with powerful commanding interfaces. Changes in web evolution have a significant impact on E-tourism, on its development and business model.

The researcher Reichstein and Härting (2018) has studied the potential benefits of digitization in tourism by developing conceptual model. With the help of empirical data collection method and interview with 10 German experts. The main goal of the research was to recommend tourist companies facing challenges and changing market environment through digitization. The result found four main factors influencing the potentials of changing customers' needs which are digital services, digital marketing, data mining and online travel communities. The result showed that personalized approach for a customer with the use of digital communication channel has become essential requirements in the future. In order to meet the changing requirements of customers, we need to have bilateral communication among customer and companies during the entire customer journey. The research found that tourists planning their trip ahead can impact their travel experience in a positive way. Presenting own travel experiences to their customers personally as well as digitally can enhance trust so, use of digital communication channels has become essential requirement to win and retain customer in the long run.

Tourism has been one of the most vital components in online commerce area (Roberta Milano 2011). It has profoundly changed the business structure of the tourism industry. Online tourism has been anticipated since the new market dynamics and customer behaviour has changed. With interactive features of web 2.0, tourism marketing has attracted wide age range of internet users.

In another research by Matloka and Buhalis (2010) they explore the user personalized content (UPC) which is a new web 2.0 form of "*customized information access and streaming based on content aggregators and widgets*". They have investigated in the range and value of widgets

and assesses the potential of UPC for marketing destinations. The results from the research shows that UPC has bigger opportunity to deliver customised information according to people needs and preference in travelling. Advancement of destination marketing by implementation of UPC can be useful for efficient management of online information's and personalization of products and services.

2.8.2 Social networking sites

In October 2019 the most popular social networking site was Facebook, with 2,414 million active users(J.Clement, 2019b). The second popular platform was YouTube with 2000 million users, third is WhatsApp with 1400 million users and Facebook messenger with 1300 million users. Facebook, YouTube and Reddit has been popular for more than ten years whereas the other social media sites on the market are newer ones (J.Clement, 2019b).

The social media platforms have significantly found to be changing their features along with their competitors. For example, twitter allowed user to upload pictures and videos as the user are found to be more engaged with the content. The digital world population in January 2020 for active internet users was 4.54 billion whereas 3.8 billion are for social media usage which shows that information has a huge opportunity to reach new people using these platforms (J.Clement, 2020).

The most popular Facebook page in Norway focusing in Nature was Aurora Borealis Observatory- visit Senja with more that 4 million fans (H.Tankovska, 2020). Second most popular page was Team Drift monkey about drifting car page with 3.7 million fans. The Facebook penetration rate of Norway was high with 90% used by people from 18 to 29 year old in first quarter of 2019 (H.Tankovska, 2020).

Social networking sites have known to take a convenient place for smart tourism due to its cost-effectiveness (Park, Lee, Yoo, & Nam, 2016). It has become a powerful marketing tool that can instantly reach millions of people in Europe. It was found that in 2018, 95% of women and 86% men from age 18 to 29 were Facebook users in Norway(H.Tankovska, 2020).This has also become a hub where local governments, tourism agencies, industry partners and many other businesses can come together and build stronger business relationships in this sector. Social network analysis can further provide information about investment opportunities, tourism industry research results, PR opportunities, and travel trade possibilities. A city like old town of Fredrikstad can also take this opportunity to increase their benefits.

Furthermore, J.Clement (2019a) has stated the total time spent using internet worldwide through mobile in average every day in first quarter of 2020 by age range from 16-24 is 4 hours 1 minute which is the highest amount, age 25-34 is 3 hour 45 minutes, age 35-44 is 3 hour 5 minutes, age 45-54 is 2 hour 22 minutes and age 55-64 is 1 hour 42 minutes. Similarly, internet used worldwide in 2020 by laptop is 16-24 3 hour 3 minutes, 25-34 is 3 hour 37 minutes which is the highest amount, 35-44 is 3 hour 21 minutes, 45-54 is 3 hour 23 minutes and 55-64 is 3 hour 11 minutes. From the data above we can find that the internet has become nearly universal, the data shows that every age range are using internet for minimum 1 hour 42 minutes. The younger age range are using internet more in mobile whereas age range from 25-34 have been highly using laptop for internet usage.

Tourist is found to be selecting destinations for specific purposes to reduce uncertainty while exploring the area and wanting to perceive the greatest reward for its effort. An analysis of tourist's mobility showed that they are willing to choose attraction or skip it based on the specific areas of city centres, where they can find the main tourist attractions (historical buildings, parks, museum, theatres, concert halls, etc.), places with leisure and shopping conveniences, accommodations near the walking distance near major attraction in the city (Salas-Olmedo et al.,

2018). So, with the help of social media analysis, we will be able to find the requirements of tourists and the possibility of further activities.

The research (Oliveira, Araujo, & Tam, 2020) studied about the reason to share travel experiences by people in social media. The tourism practise and travelling to tourist destination decisions are now affected by the views of trusted friends which is channelled through social media. There other people known as "lurkers" who does not share their travel experience with others. The study has drawed social influence theory based on identification, internalization and compliance, user`s personality to explore the issue. Based on 381 response, results revealed that perceived enjoyment was the most important motive to share their experience on online social media and travel websites. The security and privacy issues were the top reason for not sharing information.

Park et al. (2016) investigated the utilization of Facebook for tourist development by the local Korean government. The result showed that the local government had been actively using Facebook for tourism promotion to the public. Facebook was found to be correlated as a smart tourism ecosystem and activates the local tourism promotion.

Y. Li, Hu, Huang, and Duan (2017)have described the benefits of smart tourism. They stated that smart tourism can change tourist information search behaviours. As the touring information is easily accessible through the internet, tourists can find previous reviews on tour behaviour, clicking action on the internet, spending records and many other data. Tourists can experience three-dimensional virtual views of the destination. This can attract tourist by giving them lots of information about the destination, discount coupons and option to make reservations through intelligent terminals.

The research Lewis-Pryde and Evans (2016) explores currently available literatures that are related to social networking adoption within tourism industry and study case analysis of three multinational tourism industry British Airways, Thomas Cook and Marriott Hotels. They have evaluated four social media sites as Facebook, Twitter, YouTube and LinkedIn. The result found that the lowest level of engagement with social media was Thomas Cook, Then Marriott hotel and then British Airways. There was much information found on the latest trends, where one shares relevant posts, consistency etc.

The book by Safko (2010) describes the key element for successful social media marketing is impact analysis. It describes the importance to measure company`s social media account impact on the potential customers, repetitive customers and employees. Incorrect use of social media can be waste of time and resources if the outcome is ineffective. By the examination of outcome company can find if the strategy is profitable or losing income.

The research by Minazzi and Lagrosen (2013)aims to explore the use of social media in hospitality industry. The purpose of the study is to compare the behaviour of 12 European hotel brands from year 2012 to 2018. Using the result of the previous research of the researchers have studied the main dimensions which influence the use of social networks and the ways used by hotel brands to interact with customer by social media. The research both uses quantitative and qualitative research methods. The research explored the previous study adapting to current situation and analysed Facebook page of top 12 European hotel brands.

The research by Bigne, Ruiz, and Curras-Perez (2019) explored how online reviews, valence (positive or negative), content style(general or specific) and destination familiarity interact, influence digital destination image (DDI) and intention to visit particular tourist destination (TD). They have run 2x2x2 experimental design with 1055 TripAdvisor users' sample. The finding from the investigations suggested that positive online reviews, specific online reviews and familiarity of a destination enhance DDI and intention to visit TD. On the other hand, impact of valence of the review on DDI and intention to visit TD decrease when destination is familiar and finally the impact of the valence of the review on DDI and intention to visit TD decreases when the review is specific.

The research by Seo and Park (2018) analysed the effect and social media marketing activities based on brand equity and customer response in the airline industry. The study conducted survey among 302 passengers who used social media managed by airlines. The data was analysed by using structural equation modelling which showed trendiness was the top priority for social media marketing contents. The social media marketing activities showed significant effect by both brand awareness and brand image of the airline. It was found that brand awareness directly affected commitment and brand image affected online word of mouth and commitment.

There are several benefits of social media advertisements like tailor messages to the best possible audience, reach out to many new customers instantly, get feedbacks and comments in real time data, scale the ads as much as required(Seo & Park, 2018). The smart technologies bring both opportunities and challenges in the tourism industry. It can also help to empower supply-side in a destinations, tourism businesses and demand-side as travellers for identifying, customization and purchasing or producing tourism products(Choe & Fesenmaier, 2017). With further marketing can help travellers achieve diverse information, come across more possibilities for creating memorable experiences. As the way for travelling to a destination and experiencing at the destinations have been constantly evolving depending the technologies serving as a new generation of tools.

2.8.3 E-Tourism

The widespread recognition of ICT, internet of things and cloud computing in the tourism industry has provided instruments and platform to disseminate the information and knowledge enhancing innovation and destination competitiveness among stakeholders. The researcher Del Chiappa and Baggio (2015) defined "*smart tourism destination*" as a collective network of stakeholders and their technological representations. The research has studied three Italian destinations Elba, Gallura, and Livigno through the epidemic diffusion model and other network analytic methods. The result concluded that a strong structural cohesion between real and virtual components of destination exist so, knowledge-based destination management studies should be considered as a component of the system. Smart digital business ecosystem for smart tourism destination can use the diffusion of information and knowledge to be an important basis for innovation and consensus development.

Salas-Olmedo et al. (2018) has analysed the digital footprint of urban tourist in Madrid City using various data sources like geolocated photographs with big data sources like Panoramio, Four square and Twitter. Tourism requires the involvement of diverse kinds of activities like tourism flow, movement of tourist throughout the destinations and routes, internal business relations among alliances and franchises, stakeholder's relations in the destination, informal relation among actors, managing of relations with client and tourists (Casanueva, Gallego, & García-Sánchez, 2016). Many researchers studied the importance of using network analysis in relation to destinations, marketing, geography and sustainable tourism as described by Casanueva et al. (2016).

E- tourism can also make significant changes in tourist pattern behaviour through the flexibility of touring arrangements. It can become easier to change the arrangement or get the latest information about the destinations. Sharing the tour experience through photographs, travel notes, blogs or lining photos to google map and sharing information with friends can be done. Smart tourism can change the aspect of the tourism market. Touring information services can be viewed through traces and other marketing information. Interactive information exchange between stakeholders and tourists through various media, channels can help in the publicity and marketing of local products.

ICT and smart technologies have empowered tourism firms to become smarter (Sigala & Marinidis, 2012). In terms of becoming smarter by increasing performance and competitiveness. ICT and its implementation have several benefits like automating, obtaining information and transform traditional business functions and processes like marketing, procurement, supply chain management, human resource management, customer service, and its management.

Booyens and Rogerson (2017) has examined the nature of networking and learning by tourism firms in relation to accessing knowledge for innovation. The article was about the nature of tourism learning and networking, geographical network linkage and systematic characteristics of relationship among tourism firms, government agencies, higher educational institutes and others in Western Cape tourism system. It used 182 tourism firms, tourism systems and contextual interviews. The result from the research was the network linkages between local and regional actors for supporting tourism innovation are weak leading to the underdevelopment of local and regional innovation network or systems in Western Cape. The paper suggests enhancing competitiveness of tourism firms, support for stimulating networks and strengthening systemic relationships for growth of tourism (Booyens & Rogerson, 2017).

Wei, Song, and Rutherford (2014) has described that companies with social media webpage and private company`s blog can help in better performance by better revenue, profits and reputation that company without it. However, implementation of just social media pages or Facebook strategy does not ensure success in marketing business. Many companies are found to have problem to determine the social media marketing in their profitable strategy link (Ng, 2016).

Though SNSs (social networking sites) has many advantages, it has a few disadvantages too. Improper use of these platform can bring negative impact on the customer`s attitude towards the official page and reputation. Though it can provide opportunity to share, interact and create contents the company needs to understand the customer`s motivation and requirement otherwise the marketing can fail. It is important to study consumer`s motivation, decision making, preferences and behaviour in the context of SNSs to attract new members and encourage to repeat visits to the page. The motive to participate and engagement in social network to visit the company`s page should be researched.

In the research by Ladhari, Rioux, Souiden, and Chiadmi (2019) they found evidence that users were found to be concerned about discounted items, read recommended views, participate in contests, learn about exclusive promotion to FB fan page members, find new products sold in store and get discounts. The monetary profits like promotions, contests, coupons and content profit like tips and information on new products can foster loyalty to the page. Both information and monetary profits has enhanced their interest and bring positive attitude towards the page.

The research by de Vries, Peluso, Romani, Leeflang, and Marcati (2017) also found that videos are more popular comparing to text and images. The result showed that use of influencers and known endorsers in videos had increased in number of likes, share and comments. Users were found to be reacting more when the post appeal to their emotions than informational post. The effect of individual and cultural values while people use social media, their motivation in using corporate pages, their attitude towards their page and purchase intention are important information to the company(Pookulangara & Koesler, 2011).

The main purpose of the study by Hernández-Méndez and Muñoz-Leiva (2015) was to investigate the effectiveness of online advertising on different sites under the influence of web 2.0 philosophy in tourism industry(eToursim 2.0). Firstly, analysis is conducted based on type of banner (image /text) used in different eToursim 2.0 tools have greater influence on the attention of potential tourists of a hotel and secondly types of banners (static/animated) draw attention. Based different classification of tourist in gender, age and experience level in these tools a mixed experiment was conducted by eye tracking method and self-administered questionnaire. The results revealed that tourists tend to take longer periods and higher prior fixations to notice text

than image. Tourist tend fixate on static banners. It suggests implementation on effective advertisement on eToursim tools in future.

Another researcher Koleva-Tsankova (2018) has studied factors that influence destination marketing in the tourism industry of Norway and its relation to each other. They have used both qualitative and quantitative method as in-depth interviews, literature review and questionnaire in the tourism industry. Factors like *"accessibility, affordability, image, quality of service, attraction and experience"* are found to affecting tourism industry as the result from research.

2.8.4 Social network analysis

In today's era social network is the fastest growing application over the web. The main feature of interaction medium among the people throughout the world has made social media more popular. Social network is known to be complex graphical visualisation of data by collection of nodes connected through edges (Mislove, Marcon, Gummadi, Druschel, & Bhattacharjee, 2007). Nodes are the individual people inside the network and edge is the relationship between these actors. Revolution of internet as a platform to share information through various mediums has changed drastically in few decades. According to the book by Scott and Carrington (2011) social network and the techniques to analyse these networks existed decades ago.

Another researcher Benckendorff and Zehrer (2013) studied network analysis of the most influential scholars and works that contributes to tourism in the field of academic study. The finding from this research was to identify the most influential scholars, journals and seminal work on tourism research and the relationship between influential authors and works, networks and tribes in tourism research.

Social networking sites are user interaction-based platforms; therefore, playing a significant role in sharing information and communicating among tourists. In research by Luo and Zhong (2015) they used social network analysis to study the communication characteristics of travel-related reviews in social networking sites using ego and whole network methods. The research found that it provides a dynamic perspective and communication to be a dynamic dissemination process.

Much of the research conducted on social networking sites (SNS) is mostly focused on user motives and behaviours. The research has been found to be divided into two categories. First, many studies considered SNS as travel-related information search engines. On the other hand many researches (Bruns & Burgess, 2011; Y. Huang et al., 2010; Xiang & Gretzel, 2010; Yoo, Lee, Gretzel, & Fesenmaier, 2009) found SNS was not the main source for travel-related information but official websites by the tourist authorities, travel agencies or other third party websites were found to be trustworthy.

Other studies in tourism have identified the use of SNS in terms of user characteristics and motivation to share (Ráthonyi, 2013). The Lo, McKercher, Lo, Cheung, and Law (2011) found that many of the tourist sharing photos were of young people, with a high level of education, good income, familiar with travel and had the willingness to be involved in sharing their experience while traveling at various destinations(Lo et al., 2011). A study (Lo et al., 2011) conducted among Hongkong residents revealed that 89% of the travellers take photographs but only 41% posted them online, most of them posted through SNSs, instant messages, online photo album or personal blogs.

Furthermore, Y. Huang et al. (2010) explored the motivation and barriers to travel information sharing in SNSs among undergraduate and graduate students in the United States (Y. Huang et al., 2010). The result showed that three functional motivation was: the obtaining of travel information, information dissemination and personal documentation whereas on the other hand, privacy concerns like online threats, losing privacy control through advertisements, other

potential risks, and time-related issues were known as key issues for not using SNSs based on travel information.

The research by X. Li and Law (2020) investigates on comprehensive network analysis for studying current state of big data research in tourism by investigation of multidisciplinary contributions relevant to big data. The comprehensive network analytical method consists of co-citations, clustering and trend analysis which has been used from publications from 2008 to 2017. The two unique datasets were used from web of science. The first dataset involved in big data research of tourism and hospitality whereas second dataset involved in computer science comparing to tourism. The result from the research suggested that social media applications and user generated content are gaining momentum on the other hand theory-based research on big data in tourism is limited. Tourism and other similar domain were found to have same challenges like "*privacy, data quality and appropriate data use*"(X. Li & Law, 2020).

2.8.5 Sentimental analysis

The research A. A. L. Valdivia, M. Victoria; A Herrera, Francisco (2017) conducted on three well known monuments of Spain, Alhambra, Mesquita, Cordoba and Sagrada Familia. They studied sentimental analysis using TripAdvisor webpages considering ratings of the page as one and two in negative category, three as neutral and four and five positive categories. They have used four sentimental analysis method(SAMs) as SentiStrength, Bing, Syuzhet and CoreNLP and extracted the polarity of the opinions. The result found that there was high level of misclassification of the ratings because they observed users write negative comments on positive user rating. This study showed that there is requirement of analysing opinions beyond the ratings. The research suggested three steps as a)handle the negative comments with SAMs through learning models, b) get good clusterization by consensus degree among SAMs and c) discover relationships among common topics to characterize the cause behind negative comments.

Emotions has been studied from earlier times through fields like "*psychology, philosophy, sociology, biology*" etc (B. Liu, 2010a) . However, there are no fix sets for basic emotion found in researches. According to the book by Parrott (2001) has described six types of primary emotions which are "*love, joy, surprise, anger, sadness and fear*"(Parrott, 2001). These emotions can be divided into many secondary and tertiary level. The strength of opinions lies in the intensity, but the concept is largely intersected. An objective sentence describes "*some factual information about the world whereas subjective sentence express personal feelings, desire, suspicions, allegations and beliefs*" (B. Liu, 2010a).

Another research Cortez, Levachkine, and de la Cruz (2014) has evaluated tourist information by using sentimental analysis. They have studied the comments to classify the hotels of the port of Mazatlán Sinaloa, Mexico. They have used TripAdvisor for the collection of comments. The study used ontologies for the analysis of comments. The goal of the research was to use computer analysis for the comments found in TripAdvisor and perform automated classification of the hotels for possible tourists. The major challenge found in the research was the accuracy of the result as the comments were found confusing. The main reason was the Spanish language, which was found to have ambiguity, many synonyms.

The research by Salehan and Kim (2016) used 35000 online reviews for 20 different products from Amazon.com website using crawling software developed by the researchers themselves. The final sample was 2616 after many filtrations. The sentiment analysis was carried on by using SentiStrength software. The result from the study found higher level of positive sentiment in the title had more readerships, even neutral polarity in the text was perceived as helpful. The length and longevity of the review did not affect readership and helpfulness. This study recommends for development of scalable automated system to sort and classify online consumer reviews which can be beneficial to vendors and consumers too.

Sentimental analysis method is divided into two approaches as knowledge-based approach and machine learning based approach(Schouten & Frasincar, 2015). The difference between two approaches is that machine learning models require to train certain amount of data set into training dataset to perform better whereas the knowledge-based models have power in the quality of used knowledge repository. Various researchers have found machine learning algorithms highly successful in sentimental analysis(Q. Liu, Zhang, Zeng, Huang, & Wu, 2018; Maas et al., 2011; Tang et al., 2014). Similarly, many researches have also used knowledge-based model using ontology, classifying texts and calculating sentiments (Cambria, 2016; Federici & Dragoni, 2016; Zhao & Li, 2009).

Understanding of emotions has been always linked with personal development, growth and key tile for human intelligence (Cambria, 2016). With the advancement of Artificial intelligence, emotion processing has been playing important part for polarity detection. Capturing general public`s sentiment in the topics related to social events, politics, marketing campaigns and products has raised interest in both scientific and business world. This has led towards the increase in affective computing and sentiment analysis researches. Leveraging "*human computer interaction, information retrieval, and multimodal signal processing of people`s sentiment*" can be easily retrieved from the rising amount of social online data (Cambria, 2016).

The research Q. Liu et al. (2018) studied aspect based sentiment classification. The research found that prevalent content attention mechanism used for sentiment classification is not useful for multi aspect sentences and complex sentence structure. The researcher have proposed new content attention-based classification model as sentence level content capable to capture important information and context level attention mechanism simultaneously make order of words, calculate their correlation etc. This resulted to outperform using the model and described that proposed mechanism played the key role.

Another research Maas et al. (2011) found unsupervised vector based approaches fail to capture sentiment information in many important words in NLP task. They proposed to use both unsupervised and supervised techniques to learn words vectors capturing semantic terms based on star rating of online documents. They evaluated the model based on small widely used sentiment and subjectivity and found that it outperformed previously used methods for sentiment classification.

Another research Chang, Ku, and Chen (2019) conducted has studied social media analytics by extraction and visualization of Hilton hotel ratings and reviews using TripAdvisor and data from Google trend. They have used aspect-based sentimental analysis method. It consist of five steps as: "*data collection and processing, sentiment sensitive tree construction, convolution tree kernel classification, aspect category detection and data visualization*"(Chang et al., 2019). The result from the study found that sole source for the data, lexicon and methods are not adequate enough for discovering value of rating and review. The proposed method have outperformed the baseline models which can be beneficial to improve hotel services and increase marketing opportunities.

2.9 Chapter Summary

This chapter presents the systematic literature review, where we went through the process of manual search for specific literature related to the thesis topic. Based on Kitchenham guideline we have six steps as review scope, search terms, search procedure, inclusion/exclusion criteria, data sources and analysis process and finally results of the structured literature review. The selected articles were then described in related work title under the basic keywords selected for this research.

Chapter 3

Methodology

A research methodology plays an important part in a research project because of its two main functions(Kumar, 2019). The first function is for the identification or developing the process and logistic arrangements required for the research for example outlining the whole process detail of the research journey, explain study methods chosen for the research and why, who are the respondents, their selection purpose, how to gain the required information and its analysis. The second function is to ensure the importance of quality in the process to emphasize the validity of the data, its objectivity and accuracy ensuring that chosen methodology is adequate in the research.

It is commonly found that authors generally use two approaches both qualitative and quantitative. As both methodologies has their own advantages and disadvantages, they can be used for better outcomes. The quantitative approach can be used for analysing the structural features in the network or actors. This method is found to be criticized for lacking conceptual consideration, difficulty in its utilization and understanding, presenting actor`s structural features from their individual features. It is also found that quantitative methodology lacks in studying deeply about contents of relationship, their history, specificity etc which is possible by using qualitative methodology. Qualitative methodology on the other side, is found to be lacking relationship visualization and problems to make easy and explicit synthesis. So, it is necessary to use both approaches for better visualization and understanding of the issues for analysis.

We have chosen mixed methodology for this research to gain a better understanding of the issues found with the tourism marketing strategy by the authorities of Visit Fredrikstad and Hvaler. The old town is a well-known tourist destination in Fredrikstad. It is popular for its fortification shape historical area with many summer attractions specially

3.1 Research objectives and questions

This research attempts to answer the following questions:

RQ1: How can an analysis of the social network support the old town of Fredrikstad with the availability of Big Data service(social network)?

In RQ1, we have a focus on the implementation of social network analysis related to smart tourism using big data from social networking sites.

RQ2: How will social media analysis help the management of Visit Fredrikstad and Hvaler in their daily operations?

In RQ2, we are interested in finding the best ways to manage tourism marketing with visit Fredrikstad and Hvaler with long term use of social media analysis tools.

3.2 Mixed methods

Over past two decades there has been increase in the number of studies either being quantitative or qualitative which has led to researchers miss out on important parts of their studies(Domínguez & Hollstein, 2014). Many researchers have found best results by combining both methodologies. There are many researchers who have conducted mixed methods(Domínguez & Hollstein, 2014). Mixed methods bring the strength of both methodology strategies and compensate for weaknesses of both methods.

The mixed method has three features which need to be fulfilled to conduct such researches (Domínguez & Hollstein, 2014). First, the study should use both qualitative and quantitative data, secondly, both qualitative and quantitative strategies for data analysis is used and finally in any one stage of the research process there is requirement for integration either in data or in data analysis or in the results of the research. It is also known as hybrid, combined and multi-method.

In a network research, quantitative methodology includes mathematical descriptions, analysis of interaction and network structures. The measurement of value, numbers, like for example density, centrality measures, triad census. More sophisticated analysis has formal models, statistical procedures include block model analysis, exponential random graph modelling or regression analysis (Scott & Carrington, 2011).

Qualitative analysis includes all those methods in empirical social research which aim to understand the meaning and its framework to reference (Hollstein, 2011). Qualitative data is normally known for text and they are meant to provide the context of action or meaning to the system. There are other ways to collect data for this type of research like open ended method for data collection. It can be interviews, unstructured observation method and interpretive methods of data analysis. Some of the benefits of mixed methodology are enhancing both qualitative and explanatory power of data, contribute to deeper and broader understanding of social phenomenon. While combining different perspective of social phenomena, mix method supports "*development of measurement and improvement of implementation, the validation and confirmation of results and contribute to more comprehensive image*"(Greene, Caracelli, & Graham, 1989). Using mix methodology contributes in investigating social networks specially in three areas as "*a) detail description of networks, network practises and interpretations, b)network effects and c)network dynamics.*" (Greene et al., 1989).

3.3 Predictive Analytics

Predictive analytics consists of text, web, and social media analytics. Where the process begins by scraping the data from a social media platform or site. In this case TripAdvisor has been utilized. Sentimental analysis is all about information retrieval and text mining. Text mining includes knowledge discovery of textual data and text analytics. Sentiment analysis is all about information retrieval and categorizing data based on comments. Sentiment analysis is also a category of natural language process (NLP).

Data mining can be used for various purpose in tourism industry. Some of the successful application of data mining are 1) predicting sales in different services like airplane tickets, accommodations in hotels, rental cars in order to maximize revenues, 2) forecast demand at different locations to better allocate limited organizational resources, 3)forecast the profitable customers and provide them with personalized services to the repeat the visit and 4) retain valuable employees by identifying and acting on the root causes for attrition.

Data miner allows us to scrape thousands of queries at one time. Https encryption is used to protect the privacy and security of the reviewers but still a lot of private details can be retrieved.

The figure 3.2 below shows that we have taken the reviews of the fortified town located in Fredrikstad from the TripAdvisor website. The reviews were grouped into similar ontology in order to make the correct categorization. Then we have analysed all the reviews separated by three categories as positive, negative and neutral.
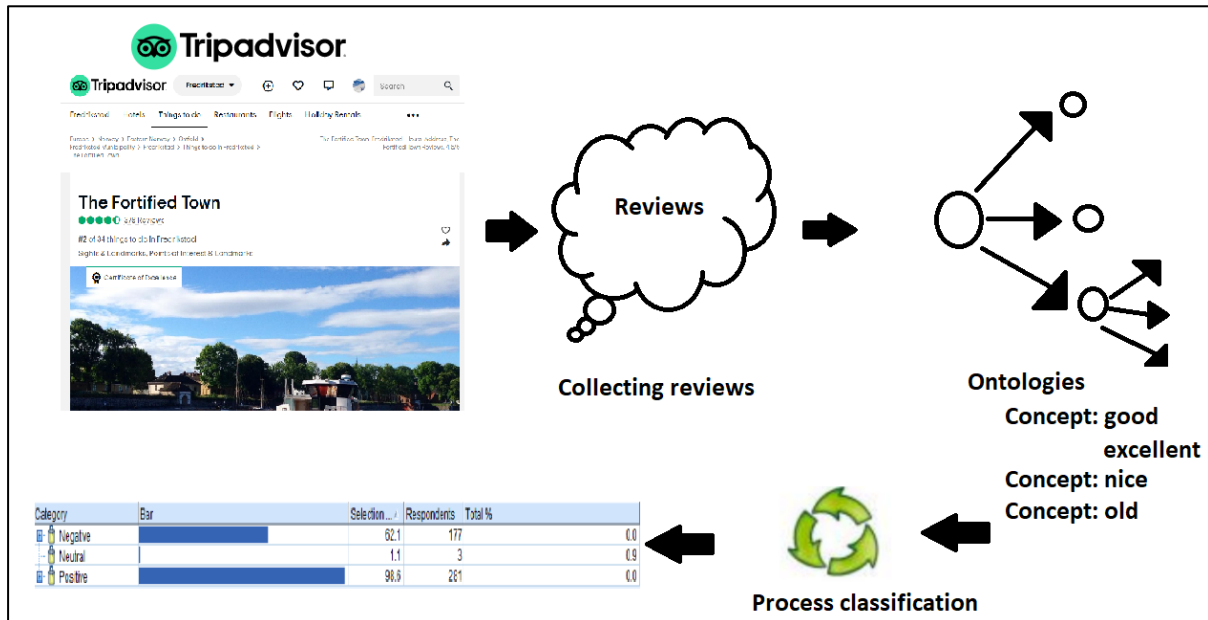
Figure 3. 1: Methodology used for classification of comments

The motivation for applying data mining are to gain valuable insight from large amounts of data that otherwise could not be seen. The process would otherwise be quite tedious both reading and sorting through volumes of data. Being able to automate some of the process is quite important these days. The figure below is the workflow for text analysis in MATLAB software . There are five steps included for text analysis process. The first step is to read the data from file, images etc. Second step is to clean the raw data by removing common words, stemming, splitting text into words, optical character recognition, text to speech, sentiment analysis, word counting and word2vec. The third step convert data into numeric form to make it understandable to computer. The fourth step is to build predictive model by tying analytics with machine learning and deep learning models. Finally, the fifth step is to visualize by using variety of plots to bring insights from the data visualization.
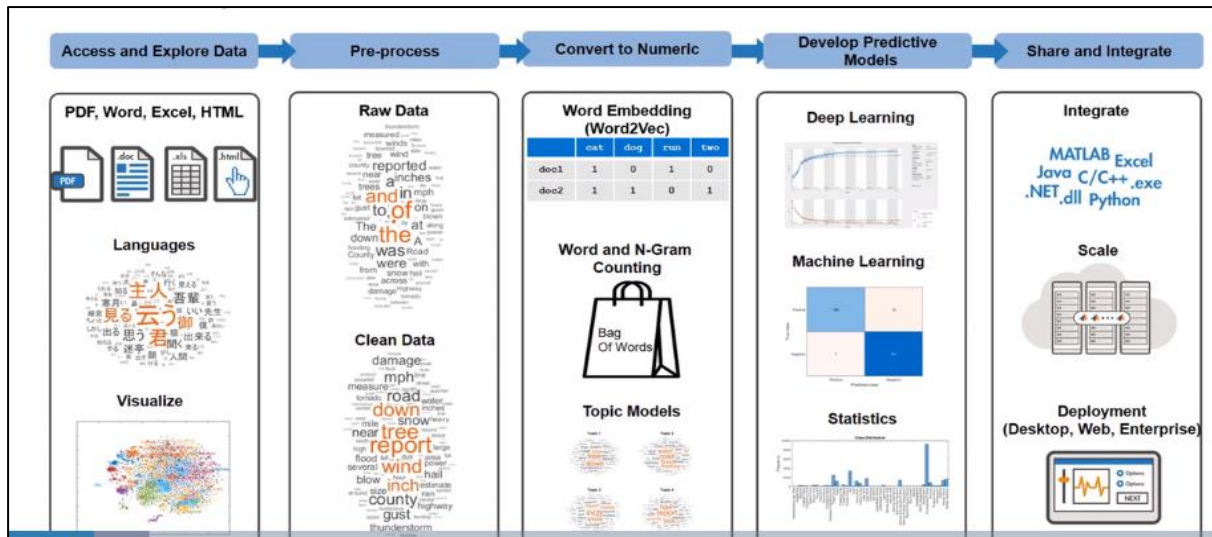
Figure 3. 2: Text Analytics Workflow

3.3.1 Sentiment Analysis

Sentiment analysis/ opinion mining/text mining/text analysis is defined as "*a series of methods, techniques and tools about detecting and extracting subjective information like opinion, sentiments, attitudes from language*"(Mäntylä, Graziotin, & Kuutila, 2018). Traditionally, sentimental analysis was about opinion polarity categorised into "positive, negative and neutral opinion about specific subject"(Mäntylä et al., 2018). The main objectives of sentimental analysis are about a product, service whose review has been written public on internet. Categorising many synonym words into one are often used for accuracy of emotional reviews. Sentiment analysis helps for data analysis to find public opinions, conduct market research, monitor brand and product reputation and user stand customer`s experience.
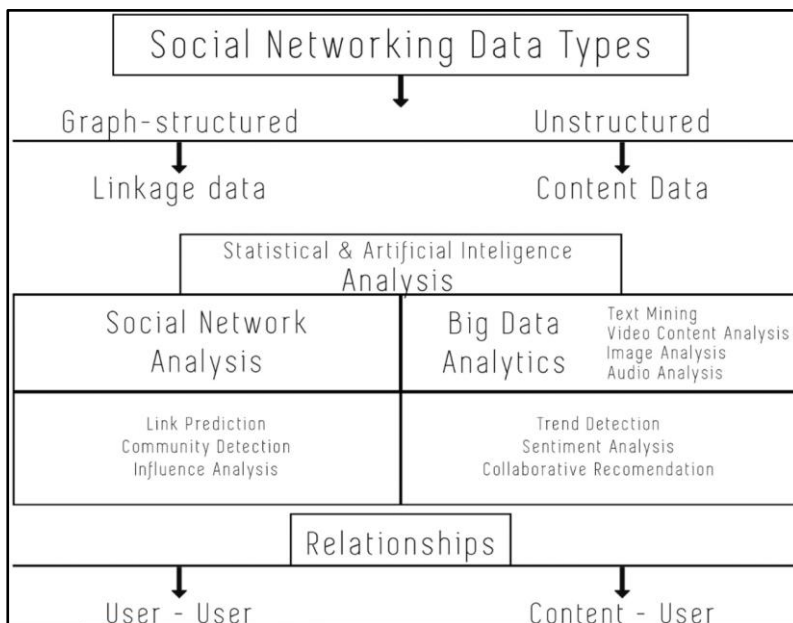

Figure 3. 3 : Types of data and analysis(Sapountzi & Psannis, 2018)

As in the figure above, the data are further found to be structured and unstructured respectively depending on their predefined manner. The biggest challenge in a large dataset is to find what the data contains and how to analyses them. The importance of social network is that the tremendous amount of contents and linkage data can be used for analysis. The data contains nodes (people, organization, products) and edges connecting each other to interact is measured through social network analysis. The graphical analysis can enhance the information from their relationships. On the other hand, the unstructured data are the user generated content or the online data contents including "*text, images, videos, tweets, product reviews and many other multimedia data*" for studying content-based analysis (Sapountzi & Psannis, 2018). The techniques for these data are structuring data, summarizing the type of data and applying various analysis approaches. Text mining extracts the patterns through textual data through information retrieval, text summarization and natural language processing.

Text mining and SNA co-exist in data analysis because they are not exclusive approaches mutually. Information from different part of network are found to be closely connected to its structure. Performing both sources of information can help to achieve better in analysis. Previously, sentiment analysis methodologies identified texts as independent source whereas social network consider data are networked. So, both features should not be overlooked. Social connection among user are considered to be equally valuable in tourism industry as patterns of content shared among users reveal about information they are sharing.

Public application provider interfaces (API) are the standard platform to retrieve social networking data from the cloud that can be used for various purposes like plugin, commercial scraping tools etc. Combination of API and crawlers are also used as crawlers is used to exact information that are not automated in API service.

There has been massive increase in the researches focusing in sentimental analysis and opinion mining in recent years(Mäntylä et al., 2018). The researcher Mäntylä et al. (2018) found nearly 7000 papers published after 2004 in sentimental analysis reports making it the fasted growing research area. Sentimental analysis has its various purposes like the prediction analysis in politics, terrorism, financial markets, business etc. Overlapping Sentimental and natural language processing has contributed in irony detection and multilanguage support. In terms of emotional reviews has been further advanced from polarity detection to complex variation of emotions and differentiating them into negative emotions like anger, grief etc.

Sentiment analysis also known as text analysis which is a form of qualitative analysis is the extraction of useful information from texts for example open ended responses to obtain key concepts. The text is grouped into many categories for finding appropriate information. The complexity of data analysis depends on the length of the texts.

TripAdvisor is known for larger text source storing huge number of reviews for tourist business worldwide. Sentimental analysis can be useful to extract the insights from the data found in the page. Classification of sentiments is the best task featured by sentimental analysis which aims to find sentiments within the document, sentence or an aspect. The analysis task is divided into three steps as 1) Polarity detection (labelling text in sentence as positive, negative or neutral), 2) aspect selection or extraction (obtaining the features for structuring texts) and classification (Applying machine learning or lexicon approach for classification of text)(A. Valdivia et al., 2017).

The Sentimental analysis method is trained to detect sentiment polarity detection which can automatically detect sentiments in documents, sentences or words. There are different methods for addressing categories for the text found in blogs, reviews, tweets etc. However, textual information found worldwide are specially categorized into two types as facts and opinions. Facts lie under objective expression about entities, events and their properties whereas opinion are usually subjective expression describing human`s sentiments, appraisal or feelings towards entities, events and their properties. For this research, we have chosen opinion and its expression conveying people`s positive, negative and neutral sentiments.

Many researches are focusing on textual information processing towards data mining and retrieve factual information. Some of the examples of the process are "*information retrieval, web search, text classification, text clustering and many other text mining and natural language processing tasks*"(B. Liu, 2010b). As the web has changed the way to share user`s view and opinions drastically, it is easily accessible to everyone in anytime. The data from product reviews, internet forums, discussion groups, blogs is collectively known as "*user generated content*"(B. Liu, 2010b). Due to diverse source with large amount of opinionated texts hidden in long forum posts and blogs can be formidable task to find relevant information. We can find tremendous value in practical applications though natural language processing task.

3.3.2 Analysis of Social Networks

Social network analysis has become popular in scientific researches from 1990`s to the starting of 21 centuries. SNA was used in different areas where we could study relationship between actors like individuals, companies, events etc., which is considered important (Czernek-Marszałek, 2019). Researches in social network in tourism sector has been found more important because of its economical perspective depending on entities like creating tourist products and tourism supply.

Network can be found universally, only few people know about them or understand them (Foundation, 2019). Networks are created when things are connected to each other. Communication between people, businesses trade between companies, machines connecting to other machines can create networks. Social network analysis is a powerful technique to organize these connected networks around the world. Analysis can reveal many insights into the ways, things connect to each other. Even though society consists of multiple networks many people do not know about these networks, how to find one, calculate them or visualize insights of its finding. Network analysis has been made simpler with the many tools available and the processing power of the computer.

There are various type of measures(metrics) used for social network analysis. These measures are commonly used by researchers and other users for commercial purpose. Some of the important measures used in graph theory are explained below:

- **Direct/ Indirect graph type**: There are two notable kind of network diagrams (Smith, 2018). Direct network diagram shows the connection between entities one way or two way through arrows while undirected network diagram only shows the connection between entities. We have selected directed graph for the analysis.
- **Vertices**: Networks are known as sets of vertices connected with edges. The vertices are the total amount of people within the network. In other field science, math, sociology, the term nodes are known as nodes, sites, points or actors. There are three types of vertices from immediate connections: in degree, outdegree and degree in or out. There are two type of centrality from entire graph as betweenness and closeness(Smith, 2018).
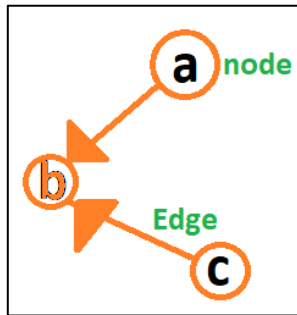
Figure 3. 4 : Simple example of network

- **Edge**: It is the connection between two vertices representing their relationship. In the network diagram by mapping out connected systems we can interpret the structure of network by clustering vertices, how dense the vertices are connected or its layout(Smith, 2018).
- **Unique edges**: The unique edges calculates the number of connections between A and B when there are multiple connections is counted as only once(Smith, 2018).
- **Edges with duplicates**: It is the total number of multiple connection between A and B vertices(Smith, 2018).
- **Total edges**: It is the total number of multiple connections between A and B vertices are all counted(Smith, 2018).
- **Self-loops**: It is an edge which starts and ends in the same vertex(Smith, 2018).
- **Reciprocated vertex pair ratio:** When vertices A and B link to each other, the connection is known as "reciprocated"(Smith, 2018). The percentage of vertices which have reciprocal relationship is identified here.
- **Reciprocated edge ratio:** If edge A and B is joined by another edge from B to A, the connection is called as "reciprocated"(Smith, 2018). The percentage of edges having reciprocal relationship is calculated here.
- **Connected components:** When a group of vertices are all connected is called as "component"(Smith, 2018). It is the number of separate sets of connected vertices. Single vertex connected components is the count of vertices with zero connection or isolated.
- **Maximum vertices in a connected component:** It is a connected component composed by number of vertices. It counts the vertices with largest connected component (Smith, 2018).
- **Maximum edges in a connected component:** It is a connected component composed by number of edges. It counts the total edges with largest connected component(Smith, 2018).
- **Maximum Geodesic distance(diameter):** A geodesic is the path or chain made of edges linking two vertices possibly through intermediate vertices. The minimum number of connection required for linking vertices A and B is the "shortest path" whereas the longest "shortest path" is the maximum geodesic distance(Smith, 2018). The average length of the paths is average geodesic distance.
- **Graph density:** Density of the graph is measured by the total number of edges among the group of vertices identifying if everyone is connected to everyone. High graph density shows that most of the people are connected to each other whereas low graph density are not connected to each other(Smith, 2018).
- **Modularity:** It is the measurement of the fitness of the group(set of vertices) created in a clustered network. The clustered group algorithm intends to find strongly connected groups. Modularity also measure the number of edges which leave a group connecting to vertices in a different group(Smith, 2018).

NodeXL is a SNA tool which can place network analysis into the context of the familiar Excel spreadsheet. It is an application supported by Social Media Research Foundation (SMRF) with the tagline called "open tools, open data, open scholarship for social media" known as .NetMap formally. NodeXL is an additional add into widely familiar framework of Excel spreadsheet with its own menus and features. It helps in simplified collection of large complex network data, store them, analysis, visualization and publish network data. It also generates reports to share insights in connected structures. According to the Social media research foundation, 4th September 2019(Meier, 2019), Facebook has changed its API so many research app has been stopped to access about public post from Facebook. Due to this change NodeXL Facebook data importers has also been affected.

We are using NodeXL as a tool for social network analysis. Node xl is a free and open source networking analysis and visualization software. It is used in Microsoft excel. It has both basic and premium version. The premium version includes access to social media data importer, advanced network metrics and automation.

There are two type of NodeXL (Marc, 2017). NodeXL Basic is free of cost and permit users to create networks, view basic level of visualization and analysis tool. Whereas NodeXL pro needs to be subscribed and it gives access to wider range of data that includes SNA packages in GraphML and other file formats. NodeXL is developed to help end users with easy features, automated frequently used process which are commonly used in network data analysis. Simplified network analysis with automated process can help in useful visualization and reports without depending on human engagement. The users are able to publish and share NodeXL datasets through NodeXL Graph Gallery, other network database, reports and visualization as contribution to the NodeXL group forum.

3.4 Interviews

Semi-structured interview is known to be very popular method for data collection due to its versatility and flexibility (Kallio, Pietilä, Johnson, & Kangasniemi, 2016). We have chosen semi-structured interview for the qualitative research methodology. Semi structured interview is defined as a data collection method where simple conversation improvising questions based on the response of participants and allowing space for verbal expressions of the interviewee.

The main advantage of the semi structured interview is that it is found successful to synchronous communication in certain time and place, interviewer can know the social cues of interviewee like voice, intonation, body language, can find more in-depth information due to verbal answers for example provides opportunity to express experiences of participants freely, no time delay by waiting for answer responses (Opdenakker, 2006). Few disadvantages of interview´s can be both time and cost for example travel cost, Interview place is faraway. We have separated interview process into four steps for this research based on the research from Kallio et al. (2016) . They are described as following:

a) Identification of prerequisites to use semi structured interviews related to the purpose of the research: the main purpose of the research is to find the importance of social media in tourism and analysis to support the old town of Fredrikstad with the availability of Big Data service.
b) Retrieving and using existing literature review articles information from chapter two: The detail study of literatures and articles found till date has been helpful to choose best methodology for the research. The adequate amount of comprehensive studying and understanding of subject was helpful from previous researches.
c) Development of the interview questions: covering the relevance of the study and identifying the requirement of data are included in the questions, we have developed

practical open-ended questions related to the research and created follow-up question furthermore when necessary.

d) Collection of data from the interview: we then retrieved appropriate data from the interviews taken and used it for further analysis.

3.5 Data Sampling

The data sampling for this research was chosen for wider perspective of social network. For the network analysis and sentimental analysis, we have chosen data from social media specifically TripAdvisor. There several platforms in social media for the collection of the data but we have chosen TripAdvisor to narrow the goals of the research. The page identified the old town as "The fortified town". TripAdvisor is known as platform for tourist to get information about the travel destinations, so it has described about the old town. The page has altogether 377 reviews related to the old town with tourism perspective. We have used the data for further analysis many analysis tools like NodeXL, SPSS, TextMiner, MATLAB, Azure machine learning.

Semi structured interview was chosen to find the result which would not only elucidate individual participant situation but wider population into the account. We have chosen random sampling for finding the interviewee participants. Random sampling provides every member of the population so that they have equal chance to participate as it can refer the findings representing their group. The study took place one point in time and considered cross sectional. We developed two sets of questions one for visitors and one for stakeholders (businesses). The questions are generated based upon the goals of the research. The consent form can be seen in appendix A and the questionnaire can be found in appendix B.

3.5 Chapter summary

In this chapter the research objective and questions were presented, followed by the methodology. In this case web scraping, a sentiment analysis, social network theory and the proposed interview questions were conducted.

Chapter 4

Results

The main purpose of this study was to examine the perspective of visitors towards the old town of Fredrikstad. It has further examined whether the demographic variables like age, gender, usage of social media etc., affect the marketing strategy in tourism. TripAdvisor was founded by Stephen Kaufer, Langley Steinert and others in February 2000, which is an American based travel company based in Needham, Massachusetts(Tripadvisor, 2017). Firstly, it was a site that listed information from guidebooks, newspaper and magazines. Later it was purchased by InterActiveCorp in 2004 and collaborated with business travel group Expedia. After this, the site was turned into user generated content. It has become the most popular travel destination and accommodation website with 390 million unique users visiting the website monthly with 465 million reviews and opinion for more than 7 million accommodations, restaurants and attraction in 49 markets worldwide (A. Valdivia et al., 2017). The TripAdvisor branded sites and forums offers online travel guide by free user generated reviews for travel related contents . With large user generated content, it has been extremely popular among tourist and tourist industry.

Statistica reported that (J.Clement, 2019b) 19% of consumers trust online reviews if they were authentic whereas 6% did not trust the review at all. TripAdvisor reported that 8 out of 10 users which is 86% agreed that TripAdvisor helps in planning better trips . Whereas 63% of the users wanted one single platform for reading reviews, comparing prices and book their destination. The important things that TripAdvisor users use while reading review are recency (up to date) by 42% and quantity by 30%. The readers want the reviews from recent date and many reviews can help to summarize the overall opinions. Further, 83% of the global TripAdvisor users search two words as "usually" and "always" for reference in the TripAdvisor reviews before deciding to book hotels (Tripadvisor, 2017). Whereas 62% worldwide users are influenced by TripAdvisor to research about their trip online. Business are found to be interested in gaining the piece of market share, engaging with TripAdvisor can result in improvement of the business, better customer satisfactions, revenue increase and continuous growth like more reviews equals more customers resulting more revenues.

By total engagement and taking care of the listing on the site can help business to use the benefits of TripAdvisor. Business details like contact details, announcements, special offers, listing images, managing reviews and analysis can help business to potential customers.

According to statistica, there are 859 million user reviews and opinions given worldwide in 2019 (S.Lock, 2020b). It covers over 8 million listings for restaurants, hotels, vacation rentals and attractions (S.Lock, 2019). TripAdvisor has been generating revenue by advertising business model relying by click based advertisement and transaction. TripAdvisor is an U.S based online travel company generating 1.56 billion U.S dollars in 2019(S.Lock, 2020b). TripAdvisor has several travel brands such as "City maps, Cruise Critic, Booking Buddy, Family Vacation Critic and Travel Pod"(S.Lock, 2020b).

4.1 Social network results

The initial screen shows the layout of the comments made on TripAdvisor pertaining to the Fortified Town located in Fredrikstad. As one can see these are all individual comments and there

is no communication between the users which is quite normal on TripAdvisor. The site is primarily used to leave a comment about a particular site or stay.
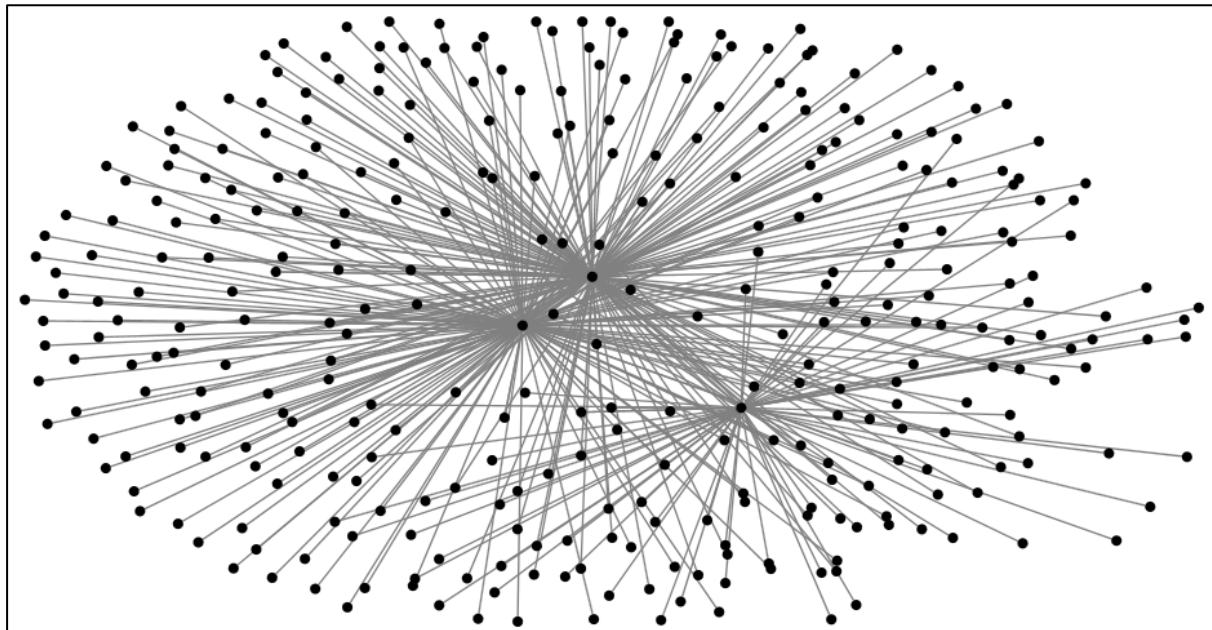

Figure 4.1 :TripAdvisor – Fortified Town

From the network one can identify who made the post and each user is assigned each own user ID. In this dataset we have gathered data on the number of reviews each reviewer has made, followed by the number of likes, the title of their post, date of review, full review, the star rating they left, the date of experience, city and country of the reviewer.

The reviewer assigned their own star rating from 1 – 5 (where 1 is the lowest and 5 highest). The figure shows that majority assigned the old town a 5 (60%), followed by 4 (33%), 3 (5%), 2 (1%) and 1(1%) stars. In addition, majority of the people making comments were local visitors, followed by visitors from other parts of Norway and last international visitors. Dark blue = 5, light blue = 4, dark green = 3, light green = 2, and red = 1.
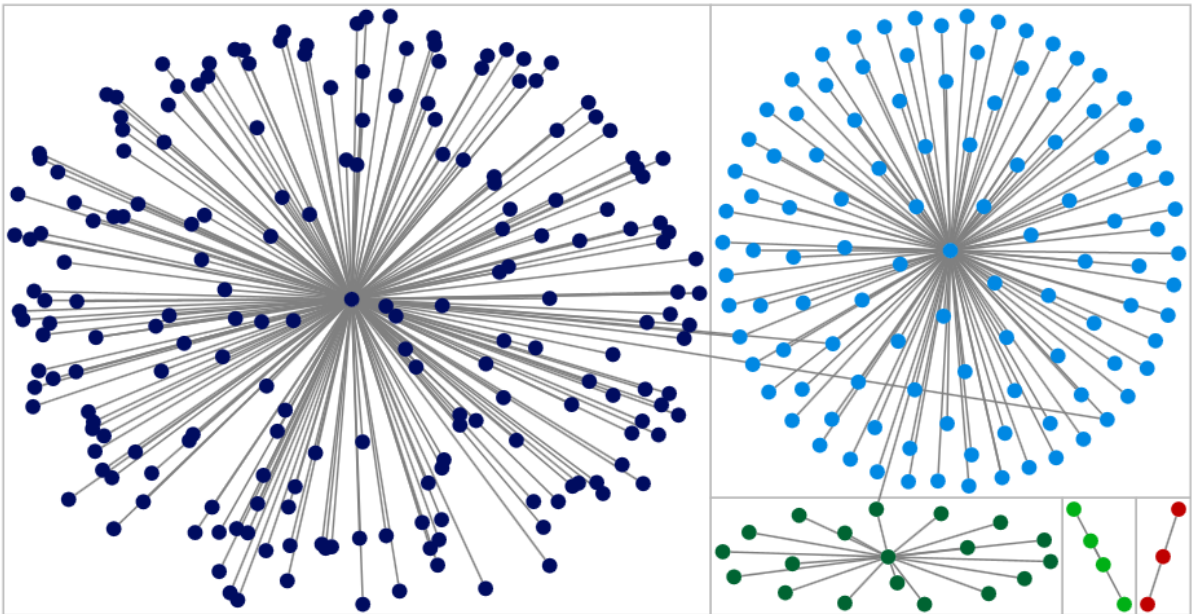
Figure 4. 2 : Reviewer Star Rating

From the data one can clearly see that the majority of the visitors on TripAdvisor are from Norway where dark blue represents local residents, light blue other parts of Norway and green international visitors. These are folks that have particularly searched for the old town of Fredrikstad and left reviews based on their recent visit. This type of information can help the tourist authority to better plan their marketing efforts.
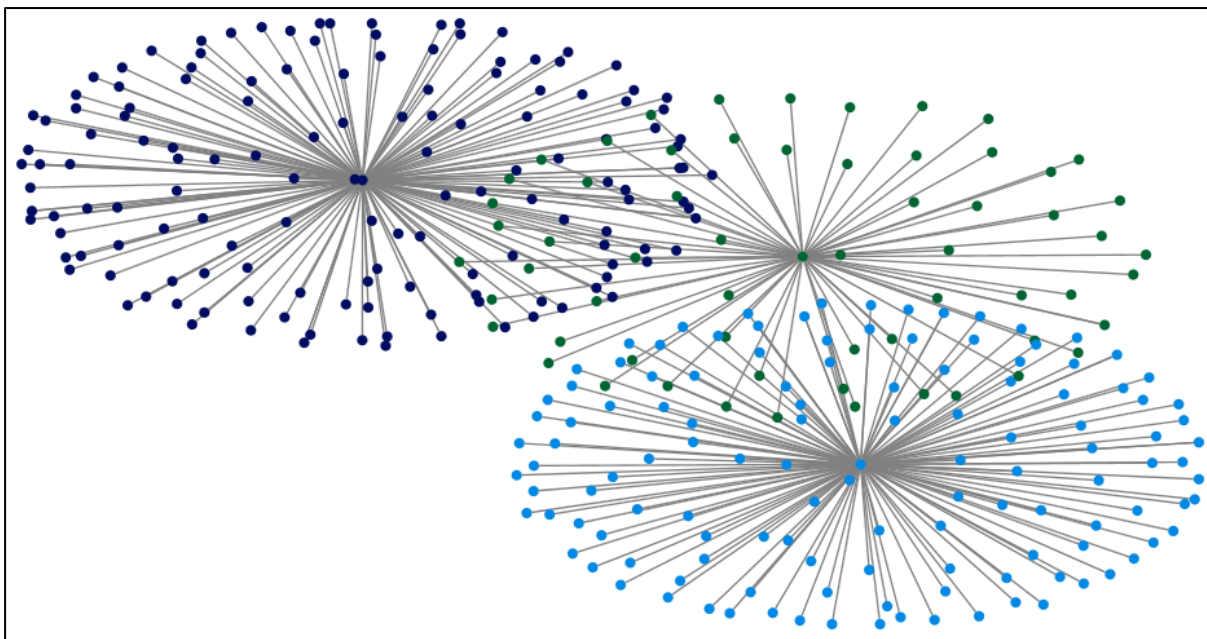

Figure 4. 3: Location of Reviewer

Further, the data has been visualized showcasing some of the comments made. The comments have been coloured coded based on the star rating assigned by the reviewer. Dark blue colour identified the users with 5 out of 5 rating, light blue with 4 out of 5 rating, green with 3 or 2 rating

and lastly black with 1 out of 5 rating. The feedback from the people interviewed matched the rating displayed in figure 4.4. People were satisfied with the place as is.
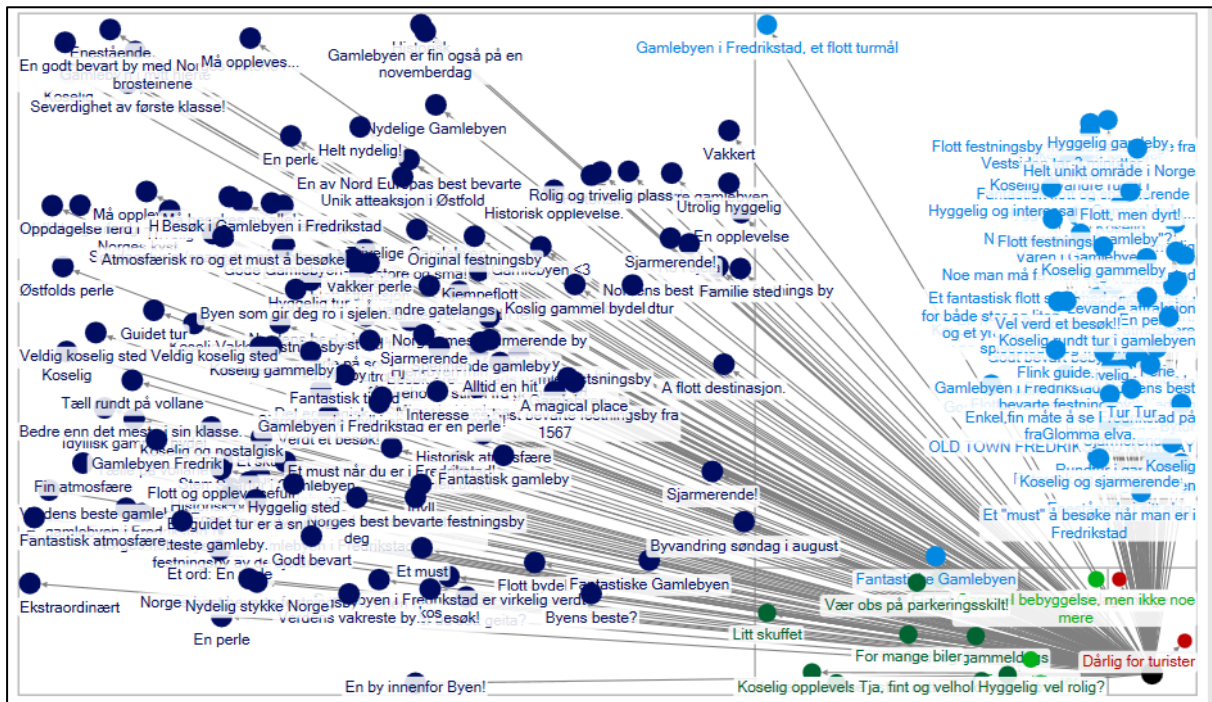
Figure 4. 4 :Comments based on ratings

The diagram below shows one of the many positive comments about the fortified town, connected to different variations of comments.
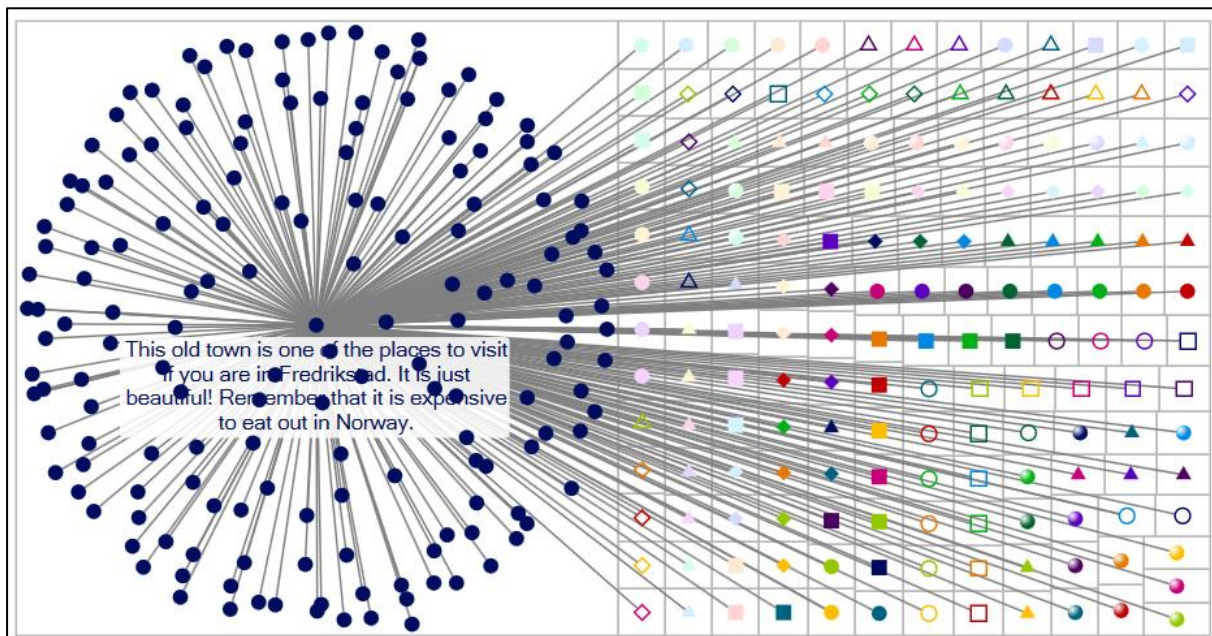

Figure 4. 5: Highlighting the main theme

The graph created were directed, indicating the direction of the conversation. There is a total of 331 vertices (connecting lines/network) TripAdvisor users which is the number of rows on the vertices worksheet. There are reported to be 22 duplicated edges which is the number of repeated vertex pairs on the edge worksheet. A duplication may occur when person A replies to person B. The total edges (people) is 341 which is the number of comments collected in English and Norwegian. There are no self-loops meaning none of the reviewers commented to themselves. In a network analysis diagram, it would visually show up as a circular edge that comes out of a vertex and returns back to the same vertex. Further, in this case, there is one connected component meaning once can get from one vertex to all other vertices. Also, there are no isolated vertices in this case. The maximum geodesic distance is reported to be two which is the length of the shortest path between two reviewers. The average geodesic distance is 1,987. The graph density operates with values between 0 and 1 indicating how interconnected the vertices are in the network. Our value is reported to be 0,006 indicating the vertices are less dense and that most people are not connected. Lastly, when the modularity is low as in this case the groups are well defined.

Graph Metric	Value
Graph Type	Directed
Vertices	331
Unique Edges	319
Edges with Duplicates	22
Total Edges	341
Self-Loops	0
Reciprocated Vertex Pair Ratio	0
Reciprocated Edge Ratio	0
Connected Components	1
Single-Vertex Connected Components	0
Maximum Vertices in a Connected Component	331
Maximum Edges in a Connected Component	341
Maximum Geodesic Distance (Diameter)	2
Average Geodesic Distance	1,987934
Graph Density	0,006042296
Modularity	-0,468262

Table 4. 1: Graph Metric

The figure below showcases where the comments are coming from around the globe. Even though the Old town is a small tourist destination it has been popular among many people. Platform like TripAdvisor can be useful source for tourist information office to focus on tourists.

Tourist from all around the world can get useful information instantly and even their reviews can be important source for future improvement.
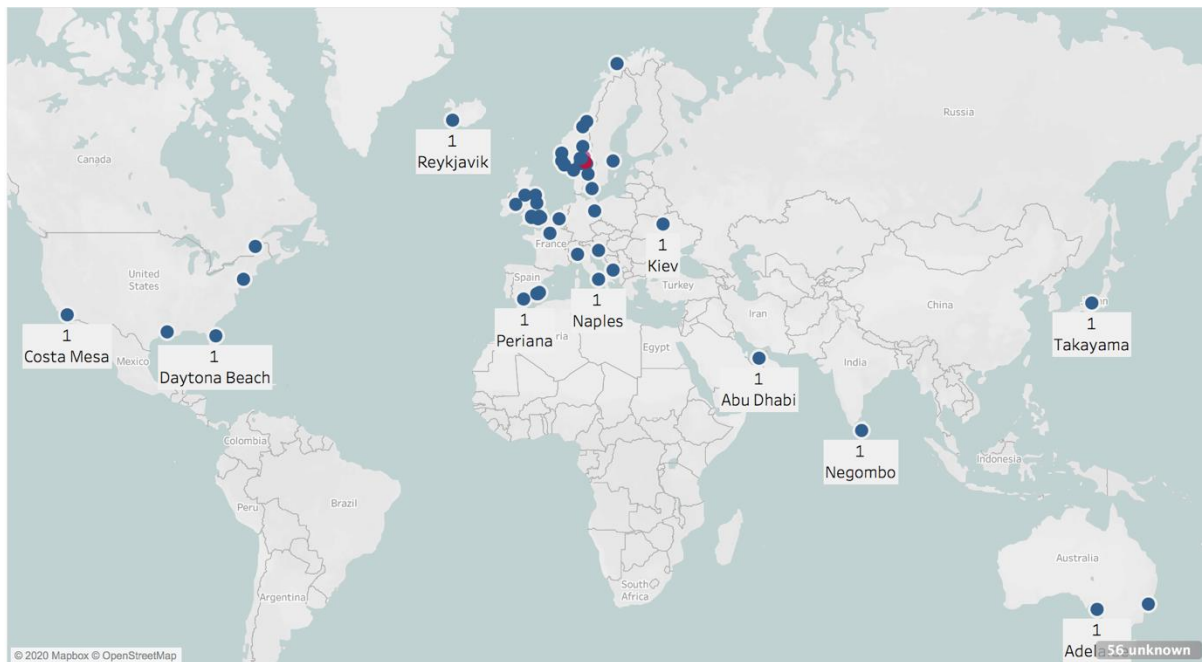

Figure 4. 6: Location of Review Comments.

4.1.1 Discussion of social network analysis

The social network analysis was done using the TripAdvisor data. Focusing on the reviews posted specifying "the fortified town" was selected for choosing the review. The network showed that data was directed as it was posted by visitors. The reviews are posted to express the experience about the place they have visited. The ratings from the page showed that the reviews were highly positive with nearly 60% is 5-star rating. The analysis showed that the reviews were mostly from the local people living in Viken, then the visitors from Norway and small number from International tourists. This analysis can be useful to find the type of people who tends to usually visit the town. This help to plan the future focusing on the type of people who are likely to visit, make events special related to these people or the place etc. Based on the ratings from comments the visitors were satisfied with the Old town. They liked the well preserved old ancient building and monuments. The network analysis can be useful to reach out more possible tourists. This can bring better opportunity for the tourist office find what type of network is useful to advertise, learn about the groups and their topics of interests.

4.2 Sentiment Analysis

We firstly created the question as "What is the general view about the old town of Fredrikstad by the tourists?" for searching the data input. For the data gathering process we decided to limit the scope by filtering the data written specific about the Old town. We have chosen Norwegian and English reviews which were the highest number of reviewers who had given the reviews for the Old town.

We have extracted data from TripAdvisor particularly pertaining to the fortified town located in Fredrikstad . Allowing us to scrape data one can see the number of review posts, the

number of review likes made by the person making the review. The title of the post, comment, data of post, data of visiting the site, star rating, ID, and where the reviewer is visiting from.

The figure 3.1 below describes that the reviewer was largely from Norway with 283 reviews, there were 12 reviews from UK and 12 from USA in the TripAdvisor review.
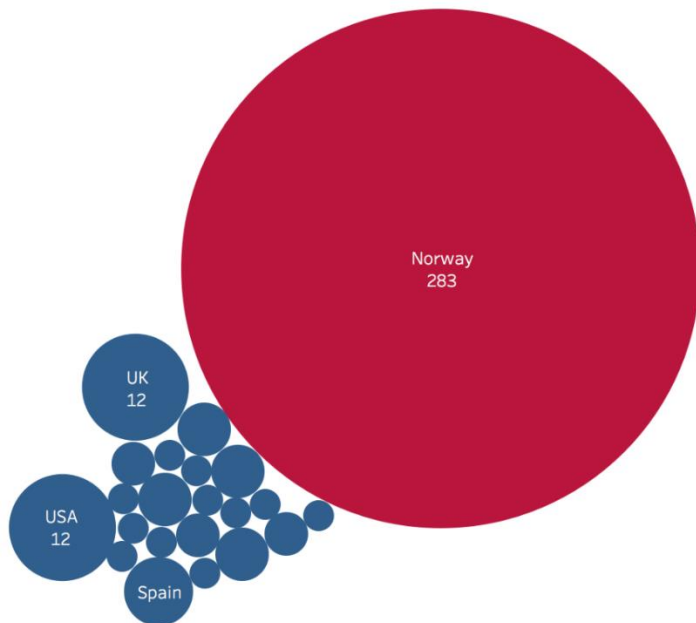

Figure 4. 7: Location of reviewers

According to TripAdvisor, we found 341 customer reviews altogether for Norwegian and English comments. The page also shows that 60% of the review was excellent from latest January 2020 to June 2015. Only 7% of the remark was rated average and below. The page has its heading named as "The Fortified town". The page shares informative information and 243 good images of the Old town of Fredrikstad. It has also given the direct link to official page of Fredrikstad and Hvaler as "https://www.visitoestfold.com/en/fredrikstad-and-hvaler/The-Fortified-Town/" .
 The reviews about the place has very good ratings about the Old Town. The page has described the old town as Northern Europe`s best preserved Fortified town that has a thriving history, local shops, art galleries and cafés welcoming visitors.

From the comments made on trip advisor, a sentimental analysis was performed. First, we scraped the data. Next we divided the comments into three categories: positive, neutral, and negative. With the subcategories under the main three categories to make the reviews easier for grouping.

The next step consisted of cleaning the data to make the data readable by applying spelling correction, grammar correction, removing punctuation marks, change the comments into lowercase letters, removing the numbers etc. For cleaning the data, we then translated the Norwegian comments into the English language one can either apply a library or perform the process manually. The data transcript was then formatted as simpler data for the analysis in the excel sheet. Then the text was tokenized by splitting the text into smaller pieces. The text is then changed into document-term matrix so that computer machine can read them changing the text values as word counts into true or false. The words are then categorized into the Positive, Negative and Neutral categories as they are separated where they belong. The terms are separated by the synonyms, words into similar group by using several dictionaries. One can notice that the words used get colour coded based on theme and tone of voice.

Figure 4. 8: *Sample of responses being classified under different categories*

Various libraries were applied to the text so that the comments can be categorized in a meaningful way. As an example, the words easy to use can be written down in different ways, so synonyms are applied. One can also see that the dictionaries applied are in English.

Figure 4. 9: Synonyms used in the dictionary

The most common words mentioned by the reviewers were old, town, good, excellent shops, cafes, place, small, free, comfortable etc. Meaning that the reviews were discussing the old town, their shopping experience, cafes, and that it was free to take the ferry as well as enter the area. One can also see the number of times the word was used while commenting.

old (217)
town (172)
good (149)
excellent (128)
shops (108)
cafes (92)
place (86)
small (83)
free (71)
comfortable (68)
restaurants (68)
visit (66)
fredrikstad (63)
buildings (60)
walk (59)
no (48)
ferry (46)
market (46)
beautiful (45)
city (43)
history (41)
enjoyable (40)
summer (37)
adorable (36)
galleries (35)
worth (35)
would recommend (32)
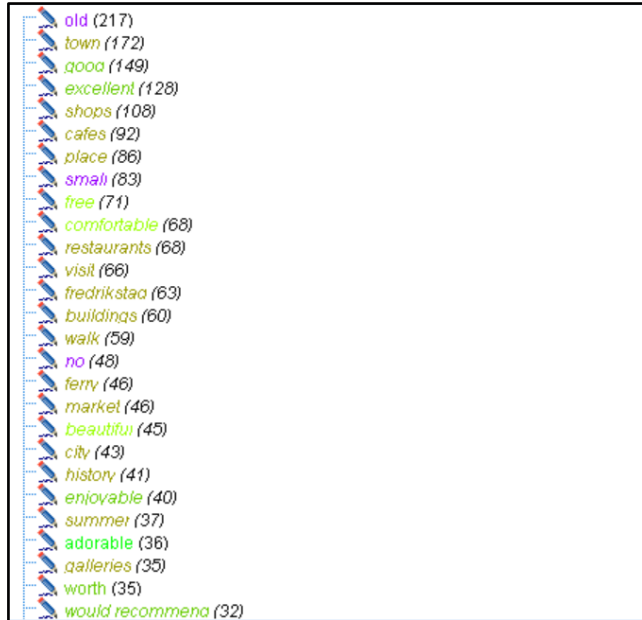
Figure 4. 10 : Top word extractions

Further one can group the positive words used. In this case the top five words used were good, excellent, enjoyable, worth, would recommend. The top positive words were also categorized based on feelings, in this case words such as comfortable, beautiful, relaxing, quiet and sunny were used.
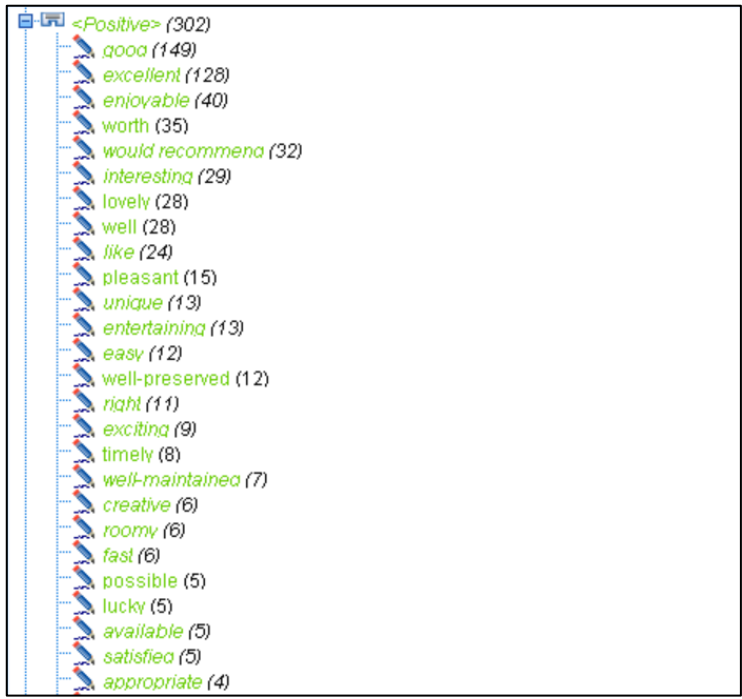
Figure 4. 11: Positive words

Figure 4. 12: Positive words based on feelings

There were also reviewers less satisfied with the place and used words such as would be good, bad, problem, stop, and depressing. Those reviewers expressing negative feelings mentioned words such as not relaxing, not attractive, undrinkable, not comfortable and boring.
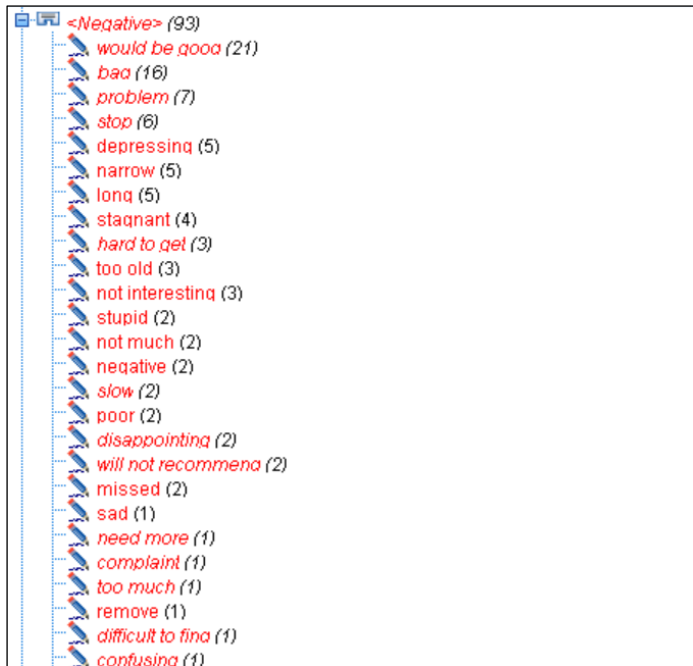
Figure 4. 13: Negative words
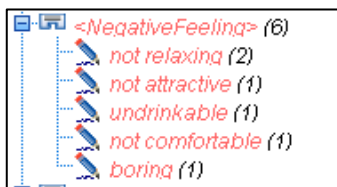
Figure 4. 14: Negative feelings

We were further able to display the positive and negative comments in a bar graph showing the various sub-categories and the percentage agreement. If we look at the negative comments, the general dissatisfaction scored the highest however one would need to look at the particular comments to see what they rated it the way they did. The next category was dissatisfaction towards the place, service and finally price. Most likely the visitors found the place to be expensive but there were other issues that they were more concerned about. Looking at the positive comments the design of the city received the highest score, followed by general satisfaction and service.

Category	Bar	Selection %	Respondents	Total %
Positive		98.6	281	0.0
Pos: Service: Accessibility		27.0	77	22.9
Pos: General Satisfaction		28.8	82	24.0
Pos: Service: General		28.1	80	23.5
Pos: Usability in the city		4.9	14	4.1
Pos: City: Public Image-Reputation		2.5	7	2.1
Pos: Design-Features of city		52.6	150	44.0
Pos: Price		27.4	78	22.9
Pos: Service: Knowledge		1.8	5	1.5
Pos: Service: Attitude		14.7	42	12.3
Pos: General Service		2.8	8	2.3
Pos: Description of city in size		4.9	14	4.1
Pos: Information of the city		2.8	8	2.3
Pos: No Plan to Change-Would Reco		11.2	32	9.4
Pos: Variety		3.2	9	2.6
Negative		62.1	177	0.0
Neg: Price		2.8	8	2.6
Neg: Dissatisfaction towards place		8.4	24	8.5
Neg: Service Dissatisfaction		7.7	22	7.6
Neg: General Dissatisfaction		55.8	159	47.8
Neutral		1.1	3	0.9

Figure 4. 15: Total categories divided under Positive, negative and neutral labels

The reviews found in the page was separated into various categories. In the example below is one the review which has four labels according to the text used in the review. The review category can be different for each review.

🔍 Id ▭	📄 Response	🔒 Categories	Category Count
688086016	Nice historical place. Nice place and stroll and look at the old houses, small shops, cafes, museums, ramparts etc. The only thing you can draw is the ridiculous parking solution, which is a space quite far away and hard to find when you get there. Very limited parking in the old town itself. Not good enough. However, the old town is moped. It's a city boat going there. Recommended.	Negative/Neg: Service Dissatisfaction Negative/Neg: Dissatisfaction towards place Positive/Pos: No Plan to Change-Would Recommend Positive/Pos: Service: Accessibility/In Store	4

Figure 4. 16: The different types of Categories in a sentence

The positive category with subcategory as general satisfaction has 28 words which was included in formula. If the review contains the words selected in subcategory then the review will be selected for positive category. We can find word "good" was counted 12 and excellent was counted 21-time overall reviews.

🔒 Pos: General Satisfaction	28	33
ƒx [well-run]		0
ƒx [satisfied + .]		0
ƒx [worth a try]		0
ƒx [meets needs]		0
ƒx [no dislike]		0
ƒx [reliable + .]		0
ƒx [honest]		0
ƒx [expectations exceeded]		0
ƒx [general satisfaction & !(<Budget>) & !(* service *)]		0
ƒx [(* program* * | * plan* *) & <Positive>]		0
ƒx [no problem + .]		0
ƒx [(<Products> | <Product>) & (like	good	excellent	no p		0
ƒx [good value for money]		0
ƒx [* value * & <Positive>]		0
ƒx [<Customer> & <Positive>]		0
ƒx [* process * & <Positive>]		0
ƒx [(* work * | * job *) & <Positive>]		0
ƒx [* result* * & <Positive>]		0
ƒx [* plan* * & <Positive>]		0
ƒx [always improving]		0
ƒx [appropriate + .]		0
ƒx [(* company * | * organization * | <Organization>) & <Positi		0
ƒx [time-saving]		0
ƒx [good + .]		12
ƒx [* program* * & <Positive>]		0
ƒx [well-made]		0
ƒx [* improvement * & (<Positive> | significant | large)]		0
ƒx [excellent + .]		21

Figure 4. 17: Formula used for category in positive general satisfaction

Neg: General Dissatisfaction	24	3
fx [not honest]		0
fx [needs improvement + .]		0
fx [<Organization> & <Negative> & !(slow)]		0
fx [better + .]		1
fx [disappointing + .]		1
fx [nothing + .]		1
fx [* advertisement * & (<Negative> | less | better | no)]		0
fx [too many problems + .]		0
fx [(* brandname * | * brand * | * name *) & <Negative>]		0
fx [intention to change]		0
fx [frustrating + .]		0
fx [dislike + .]		0
fx [not satisfied + .]		0
fx [nothing specific + .]		0
fx [* *mail * & (<Negative> | less) & !(too long)]		0
fx [doesn't meet expectation + .]		0
fx [(* spam * | * pop-up * | * junk *) & !(<Positive>)]		0
fx [easier + .]		0
fx [no intention to use]		0
fx [(* public image* | * reputation *) & (<Negative> | better | no)]		0
fx [irritating + .]		0
fx [* visibility * & (<Negative> | better | no)]		0
fx [will not recommend & !(<CustomerSupport>)]		0
fx [waste of time]		0

Figure 4. 18: Formula used for category in negative general dissatisfaction

Neutral	9	3
fx [no suggestion + .]		0
fx [no comment + .]		0
fx [don't remember + .]		0
fx [no information + .]		0
fx [no difference + .]		0
fx [don't know + .]		3
fx [no need + .]		0
fx [not important + .]		0
fx [as stated before]		0

Figure 4. 19: Formula used for category in neutral

To look at the accuracy of the sentiment analysis, we also performed Azure Machine Learning on the same dataset. It is the predictive analytics software API with the MPQA subjectivity lexicon which is a generic dictionary including 5097 negative and 2533 positive words were applied to categorize the dataset. Each word is assigned to strong or weak polarity. It is usually best to the short sentences.
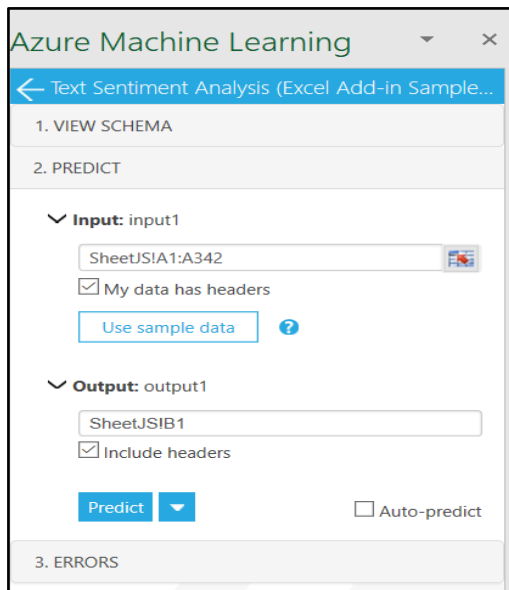
Figure 4. 20: Microsoft excel add in

1	tweet_text	Sentiment	Score
2	great place for big and small with lots of history here you will find good food small niche shops and a nice playground for the children	positive	0.989173174
3	fredrikstad is a beautiful municipality old town is a nice place to be feed the ducks play with the kids look at the animals	positive	0.961019754
4	over to the old town which takes 2 minutes here you can stroll around cobblestones among old venerable buildings sit down in an equally cosy corner cafe and take 5	positive	0.978734016
5	according to the advertisement the old town is the best-preserved fortress town in northern europe and the city is well worth a visit well-preserved park-like fortresses	positive	0.985602558
6	a cosy and charming place where you find peace of mind interesting history diversity and market on weekends	positive	0.854689956
7	very cosy and charming district which appears to be well kept cake and coffee at mormors cafe are recommended the goats were a highlight for the little man easy to get there both by car and city ferry	positive	0.811714947
8	was on a pension trip to the old town in fredrikstad had a good and knowledgeable guide who showed us around this can definitely be recommended very interesting despite the rainy weather	positive	0.928016901
9	a paradise on earth highly nice place to walk around official eateries and cafe official shops come back	positive	0.879082024
10	fantastic area and great plus that local artists get a lot of space cosy to stroll around here	positive	0.815879047
11	well-kept and vibrant old town a very interesting story and we had a guide who made the experience top notch	positive	0.891565084

Table 4. 2: The sentimental analysis in excel using add-in

The overall result from the sentimental analysis is provided below. One can clearly see that the average score was overwhelmingly positive for the positive comments. Followed by less neutral comments and even fewer negative comments.

Row labels	Average of Score	Count of review text
negative	0.217866709	44
neutral	0.521916659	21
positive	0.894487097	276
Grand total	0.78423702	341

Table 4. 3: Result from excel sentimental analysis

4.2.1 Discussion of sentimental analysis

The result from sentimental analysis showed that the reviews were on the positive side and with less negative comments. We have analysed each review following sentimental analysis methodology. The reviews were categorised into three category which was then grouped into subcategories related to the review. The most top 5 frequent positive words used to describe the Old town in the review were: good(149), excellent(128), enjoyable(40), worth(35) and would recommend (32) times. The top 5 frequent negative words used in the review were: would be good(21), bad(16), problem(7),stop(6) and depressing(5). The word frequency shows user`s expression and attitude while sharing their experiences on the review. The result sheds the light in finding the experience shared by tourists in the Old Town digitally.

The visitors have described that they like the Old town due to its ancient style of buildings making them go back in time. The word "Old" has been used 217 times to either specifically talk about the old town or describe the key feature of the town like "well preserved, well maintained ,old style, cobbled streets, beautiful green surrounding, hidden gem, living city, fortress, vibrant,". 34 visitors have used word "interesting" to refer historical sites, its story, buildings of the Old town. Even in 20th century, the town has preserved its historical era including the building, streets and the landmarks. The visitors have expressed that though the town was small they would recommend strolling around the town by walking. 79 visitors have shared about the free ferry that was accessible to everyone to reach the old town entrance which was described as "*convenient, cosy, popular, 3-minute ride, free* ". They have written that they were happy to ride the ferry, which was interesting, exciting and of great experience.

Though most of the reviews has positive reviews, they also included recommendations and few bad experiences. 19 visitors talked about the parking, they described parking with "ridiculous, confusing, expensive, problem, hard to get, limited, poor parking signs, shame, difficult parking conditions". Some of them have expressed the accessibility to move around the city was difficult due to small streets with cobbled stone which can also be difficult for wheelchair users. 72 visitors have mentioned about the narrow streets. One visitor described that there were no signs to the way showing towards the Old town, even the tourist office sign was nowhere to be found in the entrance. Signs about the special feature located in the town was also found to be missing. Two visitors who visited the town in wintertime found dead, sad and boring.

130 users have also notified about the shops inside the old town was unique and from olden days. The speciality of the shops was different from the one that are found today. They have described the shop as "cool, fun, unique, hidden little treasure, small" in the review. The 44 visitors have also described about galleries, museum they visited in the Old town which can be useful for art and painting lovers. The largest model of railway miniatures was also recommended by 21 visitors. 75 visitors expressed that cafes and restaurants with words like "cake and coffee recommended, nice, small dining place, full of life, good food, lovely food". The flea market happening every week was also found to be enjoyed by the seven visitors.

50 visitors have described that summer was the best time to visit the place with words like "beautiful weather, gem of summer, perfect, a nice summer sunny day, summer walk". 12 visitors have also written about Christmas markets in Christmas time. The moon festival has been described by 12 visitors which happens once a year in summertime. 34 visitors have recommended tourist to visit the Old town once.

The word of bags in the figure below was generated from MATLAB by using the TripAdvisor reviews. The screenshots of the coding process is provided in Appendix B. The raw data was shown in the words of bag from the data without cleaning them. The other word of bag

was generated after cleaning the data. The words highlighted in red are the most frequent used words. The size of the words identifies the important words which was used in the reviews many times. This word cloud helps us to identify what the review mostly talks about, the important topics that users were interested in talking. It is the quickest way to find important words described in the huge data. It gives the meaningful information to sort out the raw data.
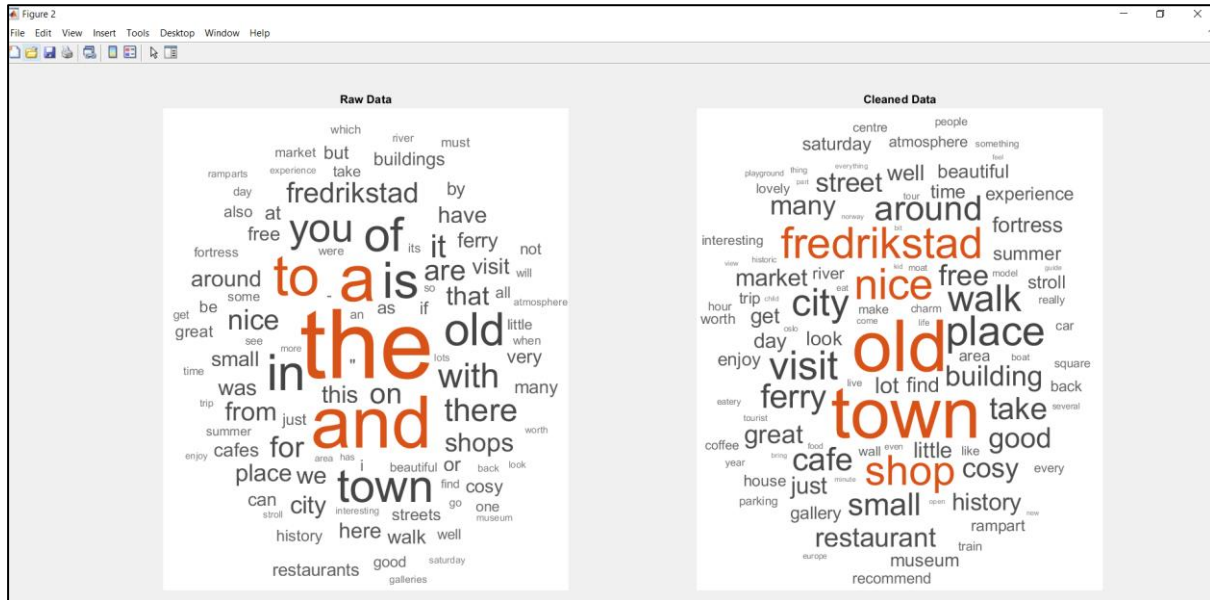
Figure 4. 21: Word Cloud from TripAdvisor data

4.3 Interview results

A total of 25 individuals were interviewed. The participants were general visitors and business owners/workers. The interview took place in the old town of Fredrikstad in one setting. Both groups were asked the same first four questions, the next group of questions were split depending if the participant were a visitor and business owner/worker. The questionnaire is listed in appendix B. The responses were written down and later transcribed that day. The questions were both closed ended and open ended allowing the participants to speak what was on their mind.

The questions itself focus on the issues related to the old town of Fredrikstad. It is important to understand the perspective of the people living, using and working in the old town. The interview tends to find the importance of individuals in tourism marketing and their intentions, if they were influenced by the social media ratings, comments etc.

Question 1. What is your gender?

ANSWER CHOICES	RESPONSES	
Female	64%	16
Male	36%	9
TOTAL		25

Table 4. 4: The total number of people interviewed based on their gender in question 1

The table above describes the total number of participants taking part in this study. A total of 16 participants were female whereas 9 participants were male.

The next table showcase the age category of the participants that volunteered to be interviewed. The age category was divided based upon the primary generations: a) Maturist pre – 1945; b) Babyboomers 1945-1960; c) Generation X 1961-1980; d) Generation Y 1981-1995; and e) Generation Z 1995 - present. The generations are known to have similar characters such as communication, shopping, technology and motivational preferences. Each of the generations have similar values, beliefs and develop the same set of expectations. By separating people into categories can help to predict bigger amount of population. In this case all age categories were represented.

Question 2. Which age category describes you?

ANSWER CHOICES	RESPONSES	
75+	12%	3
60 - 74	12%	3
40-59	16%	4
25-39	28%	7
18-24	32%	8
TOTAL		25

Table 4. 5: The age category of the interviewee in question 2

The figure below displays the distribution of local residents residing in Østfold and visitors from other places. In this case 23 were local residents and 2 were categorised as visitors.
Question 3. Are you a local resident or visitor?
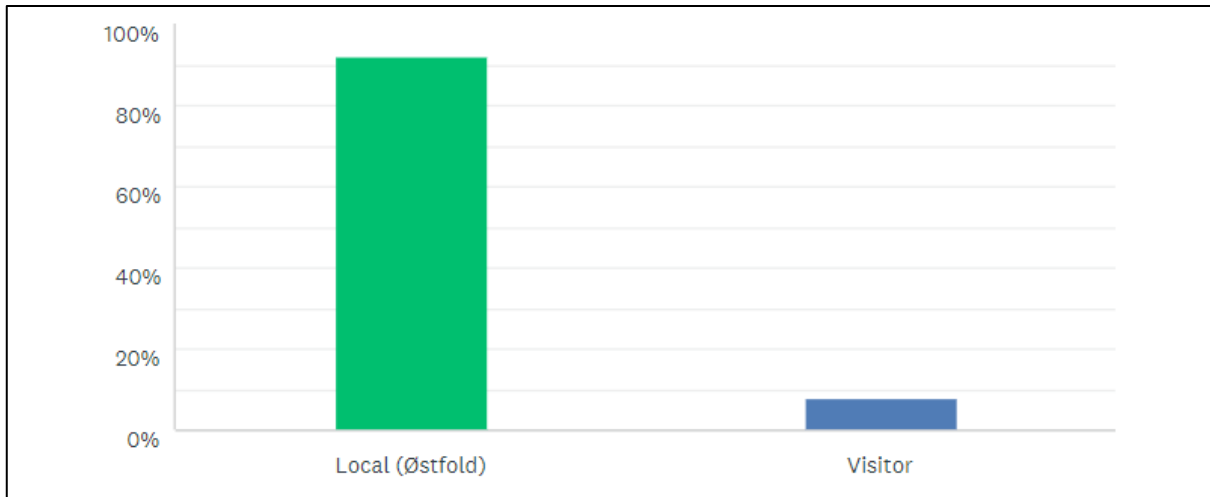
Figure 4. 22: Total number of visitor or local resident in the interview in question 3

The next question recorded the purpose of the participants trip to the old town of Fredrikstad. From the data 28% if the participants were going for a walk that day, 28% were passing through on their way to work or heading towards downtown Fredrikstad, 4% were going to shop or eat, 24% were working in the old town, and 16% of the participants lived in the old town. As one can notice from the data, majority of the people are not there to spend money. If you live on the east side, it is very convenient to cut through the old town to get on a ferry to take

you to the other side of the river or further up the river. Increasing the number of people going there to shop or eat is extremely important if wanting to make this a lively part of town.

Question 4: What is the purpose of your visit today?
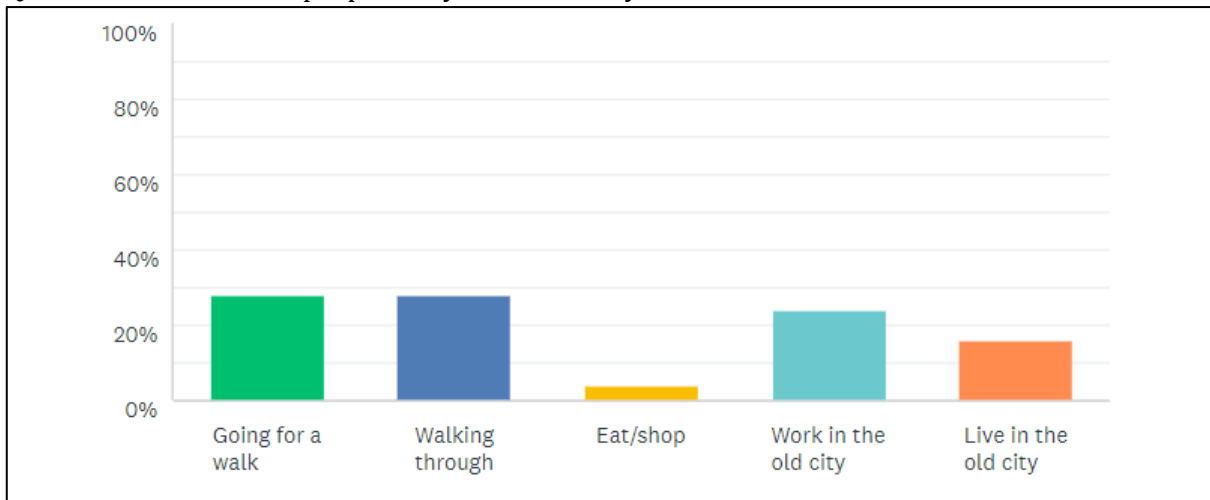
Figure 4. 23: The purpose of visiting old town in question 4

Question pertaining to visitors

The next set of questions pertained to the visitors. The visitors were asked about their online behaviour such as which online medium they relied on the most if wanting to find out information about a particular place. From the results, 11% relied on social media platforms such as Facebook, Twitter or Instagram. Whereas 5% relied on the company website itself, 5% on the tourist authorities website, 26% on booking sites such as Expedia, Travelocity, Hotels.com, Trivago etc., 26% on TripAdvisor, 53% googled the destination or business they were looking for, and 26% did not rely on the internet at all.

Question 5: Which online site do you rely on the most when looking for tourist information ?
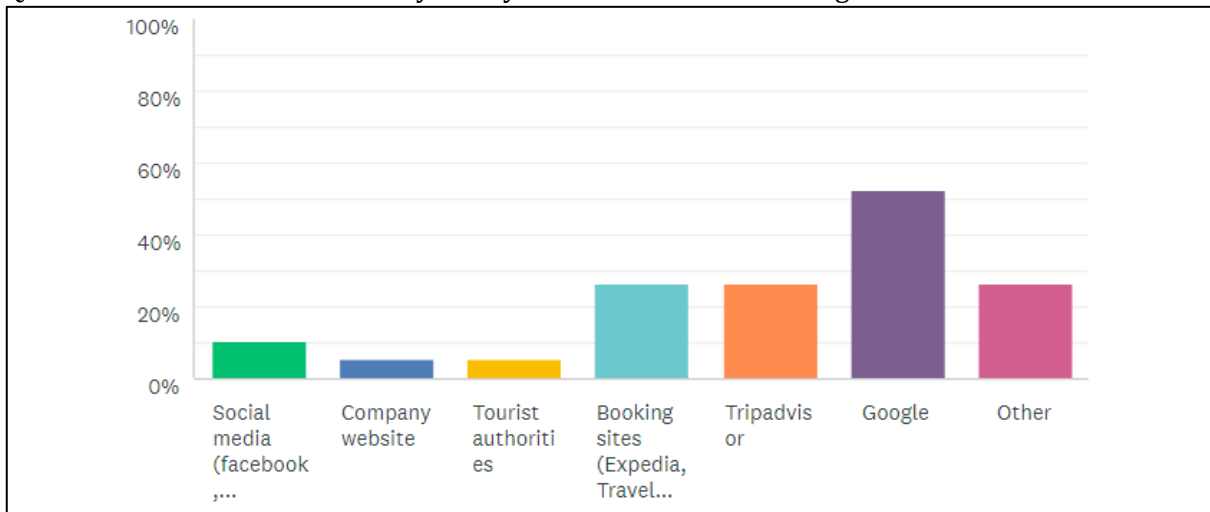
Figure 4. 24: Participants relying in online sites for tourist information in question 5

For question 6 – 8 the participants were asked to rate their behaviour on a scale from 1 – 5 where one is the lowest and five is the highest, if the rating online is <u>POOR</u>, what their

likelihood of going to visit the site is like and if the rating is <u>GOOD</u>, what the likelihood of them visiting the site would be and lastly define how often they rely on these types of ratings?

Question	Average score
6: If the rating is poor, what is the likelihood of you going to visit the site?	2.16
7: If the rating is good , what is the likelihood of you going to visit the site?	3.79
8: How often do you rely on these types of ratings?	3.16

Table 4. 6: Table of the average scores for the closed-ended questions

From the results, one can clearly see that the participants would be less likely to go if the online rating was poor and more likely to go if the rating was good. However, the participants were more neutral in terms of relying on these types of ratings. This is quite surprising as the majority of the participants were on the younger side. The individual scores are presented in the figure below to better show the dispersion of the scores.
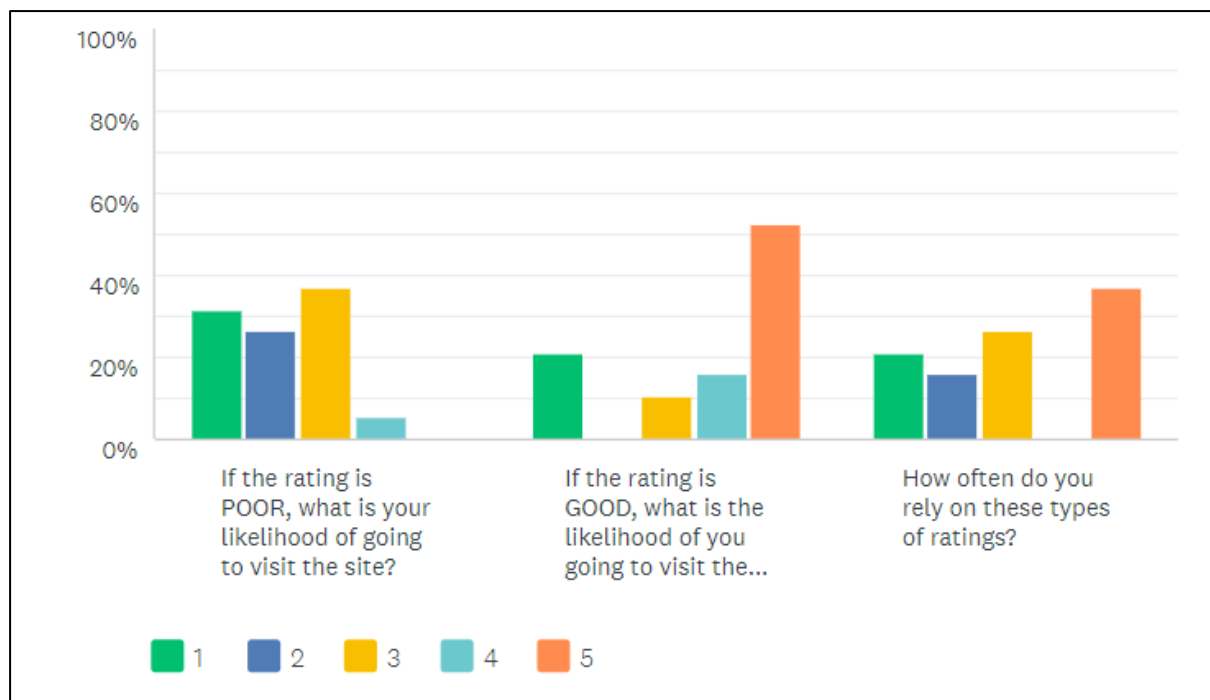
Figure 4. 25: The influence of poor and good ratings among participants in question 6,7 and 8

When it comes to the open-ended question (question 9), we received the following feedback:

Question 9: In regard to the old city of Fredrikstad, what is your general impression? Any changes you would like? All participants gave an answer to this question and were quite pleased with the old town. Majority of the participants wanted the city to remain as is and considered the current offering to be satisfactory. They liked the free ferry and liked to go for a stroll around town. Many visited the place quite often.

Questions pertaining to businesses

The next set of questions pertained to the business owners/workers in the old town of Fredrikstad. The results reported an average score of 4.33 on a scale ranging from 1 – 5 meaning the business owners/workers were online at least once a week. All of the business had its own Facebook page and Instagram account.

Question	Average score
5: How frequently do your make a post?	4.33
6: How frequently do your reply to a comment?	4.33

Table 4. 7: Table of the average scores for the closed-ended questions.

However, from the individual results one can see that either the participants were extremely active on social media or chose not to market their business there at all. There were no in-between in this case. This can be understandable as not everyone has grown up with computers and social media. It can be hard to run a business and, on the side, spend all of the time online.
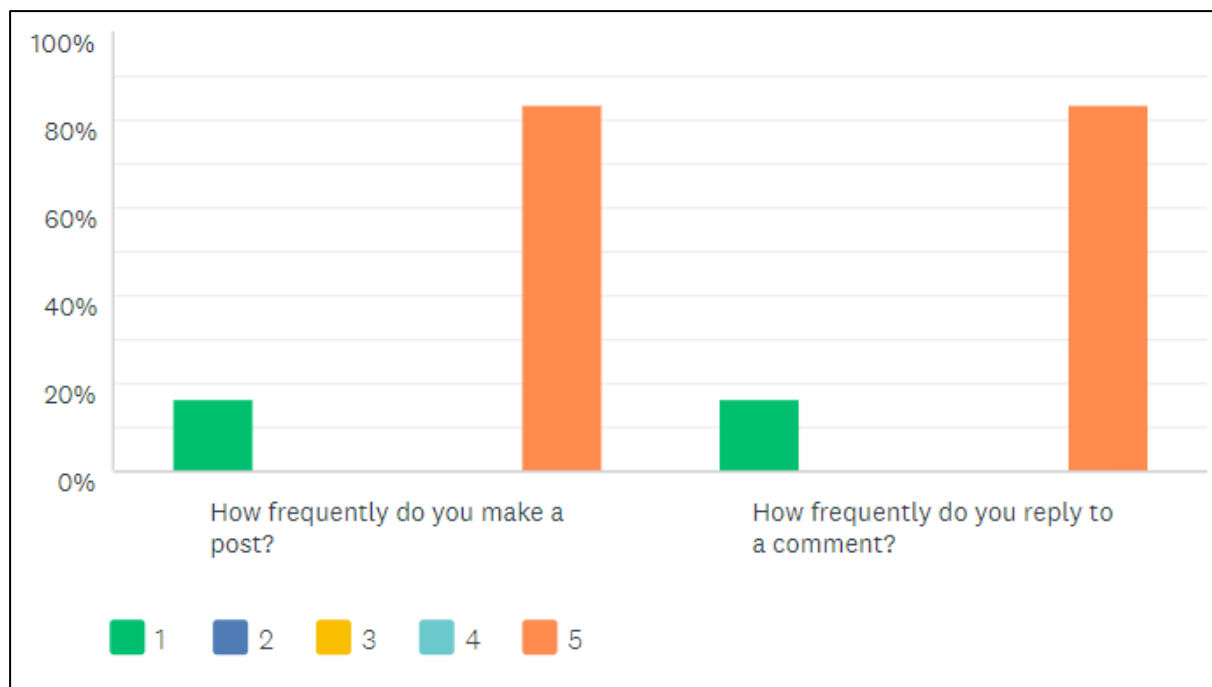
Figure 4. 26: Business involving in social media in question 5 and 6

When it comes to the open-ended questions (question 7 and 8), we received the following feedback.

Question 7: Do you see any positive changes in your business due to the use of social media? The feedback received ranged from I do not know, sometimes, and yes. One store owner noted that when she had a store in Sarpsborg city, people came based on her postings made online. However, in the old town this was not really the case. Here either people stopped by based on curiosity or they were locals knowing that the store existed.

Question 8: In regard to the old city of Fredrikstad, what is your general impression? Any changes you would like? All participants gave an answer to this question and were quite happy to provide feedback. The business owners/workers provided more in-depth feedback regarding their own experience working in the city. The expressed both challenged and opportunities and a concern for the future number of customers as people want convenience when going shopping/eating.

4.3.1 Discussion of interview

To our big surprise, people were pleased with the old town as it is, despites of all the negative publicity in the papers over the last few years. Businesses has been leaving, the old town due to high rent, less customers, and the lack of convenience. All the 25 interviewees were pleased to take part in this study. The participants were given a consent form and were they were introduced to the research and their role of participation in the research.

According to question 5 responses, we could find nearly all the participants used online media to get information about a tourist site. The visitors used online ratings for visiting new places, shopping. They were found to be highly relying on ratings, comments so that they can have a good time or make good planning beforehand. Five of the participants had commented on various sites for different purposes. The main idea behind these ratings and comments was to give advice to new people in advance. Poor rating of a site did not influence visitors whereas good rating was found to be highly appreciated by all the visitor participants.

All the six business owners or workers accepted that using social media for marketing purpose has helped to some extent in their business. They were using social media like Facebook and Instagram for posting and commenting. They are highly interactive with the customers, but two business participants found that tourist were randomly visiting the shop based on social media postings, but rather passing by the place. Local visitors come to visit the shop based on their offering. The six shops that were visited during the interview were found to be unique unlike the modern shops today and had many vintage items.

All the participants had responded to the final question regards to the general impression of the old town in Fredrikstad and the question was same to both the set of interviewees. The visitors were found to have everyday general suggestions whereas business owners had broader more specific perspective leaning towards business sustainability in the old town. All the participants including visitors and business owners found the free ferry entering the old town to be a bonus, as it made it easier to reach the old town. However, access to the ferry was not enough as the participants identified that they wanted to have more parking places for cars inside the town or nearby. The old town has small paved roads made of cobblestones and the total area of pedestrian trail covers 4.9 kilometres(Nilsen, 2018). So, parking has been found as a big issue in the area. The participants suggested that they would like to have free parking in the area. The businessperson found that parking inside the small town made the area smaller and congested. So, there is the requirement for a serious solution to everyday parking issues of locals and businesspersons. The opinions were of course mixed within each group.

We interviewed the shops in the old town which were found to have retro style gift shops. The cafes and restaurants were found to be simple and have vintage styles. Any business can thrive by including more outdoor seating areas and lively presence as two business owners also suggested to have more outdoor eatery places specially during summertime.

Three local participants shared their problems for not having local grocery shops for everyday use. They had to go to the other side of the city to purchase food and other daily items. A store of this nature will also draw more business to the other stores in the area. One of the participants even suggested if the town could have weekly vegetable and flower markets to help boost business. A market of such sort would not be in competition to the existing businesses. This would only offer more variety to buy things but also involve more people outdoor.

Three business participants even found that the few stores like liquor store, pharmacy etc inside the town were closed down which effected their business. They suggested that opening shops with important daily requirements can attract more people to visit and enhance tourism. Two of the participants suggested that they would love to visit a place if they had its unique features for example restaurants with its own special menus or shops with unique things to offer new tourists.

All the business participants accepted that they had less business during the winter season. They suggested that if there would be more additional activities like festivals, concerts during winter managed by the municipalities it could enhance their business. One of the participants found that there is a lack of organization representing the stakeholders, this led to weaker business. The participant wanted to have one representative for all the business to share the problems they are facing today, overcome them and plan for future activities.

Five participants identified that there is the necessity of several activities or speciality shops in the area to attract and gather to people from all the age groups. Three of the participants from the generation Z category found that they would like to have more affordable cafes and restaurants. Activities which could include all the generations could be attractive for both tourist or local visitors to visit the old town again and again. Two visitor participants suggested that there is the need of more signs to locate the tourist office at both ends of the fortress as well as downtown Fredrikstad.

A word cloud was created to sum up the most important key words from the conversation. For the majority of the businesses, parking was a big issue. Majority wanted parking to make it more convenient for customers to shop and eat. However, there were also people who rather wanted the old town free of cars. An option would be to have the parking lot outside of the old town free of charge and set up a small train or electric tram to transport shoppers inside the fortress. People were interested to keep the buildings as is. Visitors were pleased with the current offering the town had to offer.

Figure 4. 27: Word cloud of interview data

4.4 Chapter summary

In this chapter the results from the analysis of the social network and the interviews were presented, followed by the discussion. Various tools were applied such as NodeXL, SPS, Azure Machine Learning, TextMining, MATLAB to analyse the data from different angles to draw meaningful information out of the data available.

Chapter 5

Conclusions, Implications, Recommendations, and Summary

This chapter presents the conclusion, implication, recommendation and overall summary of the thesis.

5.1 Conclusion

This thesis aimed to study social network analysis for supporting the old town of Fredrikstad. The structured literature review helped to gain information from the previous researches conducted for this topic. The study revealed that there has been several researchers conducting similar studies to this research, but they have used various tools and models for the analysis. For this research, we have chosen both qualitative and quantitative analysis methodology. We have chosen semi-structured interview for qualitative analysis and predictive analysis for quantitative methodology. Both methodologies have played an important role for this research.

      The first research question of this study is: "*How can an analysis of the social network support the old town of Fredrikstad with the availability of big data service?*" Focusing on the solution of this question, we searched for the popular social media platforms which is explicitly dedicated to tourism marketing. We chose TripAdvisor and selected the page that was specifically about the "fortified town of Fredrikstad". We chose to conduct social network analysis and sentimental analysis for studying the quantitative data collected from the TripAdvisor page. The result from the analysis was very useful to find how social media can be used for marketing purpose by the Old town of Fredrikstad. The data analysed from the page showed what tourist experienced after visiting the Old town. Similarly, for qualitative analysis, we chose semi-structured interview which was conducted to find the opinion from people found in the Old town. The result from the interview helped to find the perspective of the local people, visitors and businesspersons from the Old town. The results found from the research has showed many possibilities for the enhancement of tourism in the Old town by using social media marketing strategies. The tourism organization of the Old town has been involving in social media platforms like Facebook, Instagram, Twitter and YouTube, it was found to be lacking in potential implication of many social media strategies for tourism marketing. From the visualization of social network analysis, it showed that there was not much interactions among the people, tourism office had posted about the Old town but strengthening interests, loyalty and satisfaction of tourist to visit the old town again was found to be missing. The marketing through social media was found to be challenging in the old town. We found that there was a gap between stakeholders and the tourism office. This is quite common as social media pages are linked to an individual. The stakeholder had their own social media pages, wrote posts and interacted with customer themselves. The events happening in the Old town was required to be seen by the visitors in the event calendar page themselves. The visitor`s problems, suggestions posted online about the Old town were not found to be acknowledged. The popular topics that were talked about in the social media regarding Fredrikstad or the Old town did not have any support from the authorization. As we have studied in the literature researches about the ways of sustainable tourism marketing, we

could find that the authorization need to work on making the social media pages more informative, welcoming, attractive and address the popular events happening in the Old town including people from all age ranges. We even found that the businesspersons in the Old town suggested that they required one representative which could unite them as one so that they can plan future events, share problems, give suggestions etc. The other important problems could also be addressed by this one representative which can be the tourist authority. We found that the Old town should focus on the community participations, use integrated marketing tools, follow tourism destination marketing trends and tourism laws for better outcome.

The second research question of this study is: "*How will social media analysis help the management of Visit Fredrikstad and Hvaler in their daily operations?*" The answer to this question is that from the predictive analysis we could find that majority of the tourist reviews from the social media platform were found to be positive. The authorities can also use similar type of tools for the analysis of the data found in social platforms they like. The visualization of data can help the authorization to find the real time facts from the tourists reviews. The analysis of the important networks, trending topics, useful information etc was easily found through the analysis. The interview results showed that most of the visitors were local so, these kind of analysis can be done from time to time by the authorization to find experiences from people who like to visit the Old town, the suggestion and complaints can be useful for better improvement and the visitors can also feel that they are important and their problems are addressed. The businesspersons can be united to have one representative and plan for sustainable business. Even including them for tourism marketing by the authorities can be a better option for win win results. In today`s era every business is required to upgrade and join digital platforms so, the authorities can also use the methodologies similar to this research for enhancing marketing of the Old town towards e-tourism. The continuous development of social media platforms has been found to have greater impact in all tourism activities. There has been many changes in tourist behaviour from searching methods, discovering, reading and using new travel information today. So, tourist authorities can bring social media as a new opportunity for efficient utilization, sharing social media resources of the Old town, taking timely interaction with visitors, make precise marketing to establish positive image and reputation of Old town, improve visitor`s loyalty and increase in revisit rate.

5.2 Implications

The result indicates that the authorities needs to focus on both efficiency and effectiveness of customers sharing reviews and become more proficient to influence potential tourist seeking information through social media platforms. Encouraging past tourists to generate contents according to their experiences, tourists sharing images and videos should be rewarded like grant discounts for next holiday, create challenges or competition among tourists can also gain attention from others(Dedeoğlu, Taheri, Okumus, & Gannon, 2020).

This research can help the authority of the Old town of Fredrikstad and the tourists. Using the predictive analysis tools similar to this research can help the authorities to get the big data information on the current situation on social media about the Old town. This can be helpful for authorities to plan for future focusing on resolving the problems addressed. There are lots of tools and methods for deeper data analysis used which can be useful for the authority to implement in their daily usage. The information acknowledged by the authority from tourists can also help the tourist in future though solving the requirements from the tourist has not been the goals of this study.

The useful information that may be posted by authorities for the tourists in social media can be helpful to all the visitors in the Old town. This study contributes in finding a better way for analysis of the data found on social media. This research not only analysed the data but also included the real participants to find their perception of the Old town.

The business persons can also be benefitted by the data analysis of the visitors review so that they can focus on the real requirement of tourists, selling the products they like, what type of tourists have the possibility to visit the Old town and also prepare for events happening in future planning with the tourism authority.

Privacy of the data can be the obvious concern to tourists, which can be both useful and vulnerable at the same time for consumers. Privacy in tourism is found to be a ´special case as the relationship with providers and their app or service is short lived which is limiting the trust building. The need for information can easily persuade tourist to forego privacy which can leave behind the digital traces to manifold in various uses. Mindless capture, storage, retrieval, information management of the digital data requires to be handled carefully to ensure safety and security of tourist.

Extreme technology dependence for implementation of smart tourism can be difficult due to widening of digital divide. Due to older age, illiteracy or the people who are not using smart technologies, does not use social media cannot benefit from smart tourism. Other problems like information overload, lack of serendipity essential to meaningful tourism experience, increasing desire to escape technologies in vacation can be another dimension for resistance toward smart tourism concept.

Needing of artificial intelligence and issues in human efficiency to automate the understandability of big data in smart tourism cannot be undenied(Gretzel et al., 2015). Despite of stronger dependency in ICT, tourism is found to be struggling with innovation deficiency. Human resource issues, happy collaboration between various actors, smart self-regulating ecosystem are known as utopian view of smart tourism.

Despite of the concerns mentioned above, smart tourism is found to be having incredibly promising future resulting in safety, convenient, exciting and sustainable destination for citizens and tourists. Personalized and relevant tourism marketing can bring positive experiences and greater opportunity for businesspersons to provide new services, business models and markets to find flexibility in structures , perspective on value creation(Gretzel et al., 2015). Data science can further create new opportunities supplied by smart tourism and advance towards new value creation taking advantages of simple technological platforms. As the sustainability is the main concern of tourism marketing, true implementation of advanced technologies and artificial intelligence is a must for better analysis of social media platforms.

5.3 Recommendations

Smart tourism has been described to increase reliance of tourism destination, its industry and tourists emerging from ICT allowing huge amount of data for the transformation to valuable propositions(Gretzel et al., 2015). This study shows the implementation of smart tourism, shedding lights of current smart tourism trends and using technological and managerial aspects for the improvement. As the word "smart" describes technological, economic and social development aspects fuelled by technologies depending on "*sensors, big data, open data, new methods for connectivity and exchanging information like Internet of Things, RFID and NFC also ability to infer and reasoning*"(Gretzel et al., 2015). So, in this study we have been more focused on analysis of big data. The analysis tools used here can be useful for the further analysis of another social media platform. We have used TripAdvisor whereas authority can choose various platforms and many tools for making the analysis applicable. The main goal of the study was to support e-tourism of the Old town so, we used predictive analysis which is just one branch under the umbrella for big data analysis. This study would recommend using similar approaches like social network analysis and sentimental analysis to other data found in other social media too. The use of analysis tools can be dependent on how to use the data and why so, using this method of this study can guide the authority to find the right purpose of the analysis. As a tourist

destination, for attracting tourist nationally and internationally, authorities need to take care of the interest of the local community for obtaining economic development, improve quality of life of the citizens living there and also meet the satisfaction of the visitors(Ramos, Andraz, & Cardoso, 2020).

The quality of experiences shared by tourist about the destination which directly reflects on their satisfaction and creation of beautiful memories. So, using smart ICT technologies can improve and add value to the experience of tourists, residents living there. Finding the online searches made by tourist before and during their visit in Old town can be fruitful. Also, by the dissemination of its authenticity, values and culture of the destination online, sharing experiences of the tourists, residents and the official destination management organization in official webpage can be helpful.

Visit Fredrikstad Hvaler or any other town or business can analyse its own pages using similar techniques applied here. Be aware that there is some API restriction in place, username, data limitation due to the Cambridge Analytica Data scandal. Each social media site or platform must be analysed separately. Other multiple data resources can be used to collect reviews however, there is no clear evidence directly found in the difference of reviews exist in those resources(Chang et al., 2019). As we can find increase in marketing though mobile devices which can be the opportunity for new spectrum for marketing campaigns of the Old town (N. Huang, 2016). There has also been concerns about fake and paid online reviews which are said to be manipulative so there is deeper investigations required in reviewer`s profile, their reviews and writing behaviours (Filieri, Alguezaui, & McLeay, 2015). This thesis has focused on social media and user generated contents, but we can recommend future researches can also be done in tourist behaviours, tourism marketing, crisis management, tourism culture and many more.

Despite of useful findings from the research, there were few limitations found in the present study. First this research contributes theoretically with literature reviews by showcasing how social media can affect tourists . Furthermore, this research presents the findings indicating how big data can be examined in detail and guide the foundation for future study of social media networks.

5.4 Summary

Finding balance within sustainable tourism and using e-smart tourism to attract tourists to the Old town of Fredrikstad has been the foremost aim of this study. As many other destinations have been adapting to the smart concept. Applying innovative technologies to achieve satisfactory high level services and becoming attractive, competitive at the same time is to become a smart tourist destination(Avelar, 2019). This study presents very interesting findings for the tourism authorities to further analyse after implementation of smart tourism using ICT. It is important for the involvement of all the important actors in the Old town of Fredrikstad to enhance the overall competitiveness of the tourist destination(Roopchund, 2020). It is found that there is requirement for greater efficiency in using the resources by deployment of right social media platform and right analysis procedure to enhance better service quality of tourism authority. The Old town requires to engage in sustainable tourism approach and become smart tourist destination.

The paper presents the big data analysis and semi structured interview. The results concludes that the application of big data analysis tools can allow the Old town authorities to define smart destination services, understand requirement of tourist in today`s trend, manage destination stakeholders requirement effectively and lastly, become more competitive and sustainable toward e-tourism.

References

Aghaei, S., Nematbakhsh, M. A., & Farsani, H. K. (2012). Evolution of the world wide web: From WEB 1.0 TO WEB 4.0. *International Journal of Web & Semantic Technology, 3*(1), 1.

Albino, V., Berardi, U., & Dangelico, R. M. (2015). Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology, 22*(1), 3-21.

Alfantoukh, L., & Durresi, A. (2014). *Techniques for Collecting data in Social Networks*. Paper presented at the Proceedings of the 2014 17th International Conference on Network-Based Information Systems.

Atay, A., & D'Silva, M. U. (2019). *Mediated Intercultural Communication in a Digital Age*: Routledge.

Avelar, S. (2019). *Island Tourism: Keep It Smart and Sustainable—An Exploratory Approach*, Cham.

Aydin, N. (2018). Social Network Analysis: Literature Review. 34. doi:10.5824/1309-1581.2018.4.005.x

Benckendorff, P., & Zehrer, A. (2013). A network analysis of tourism research. *Annals of Tourism Research, 43*, 121-149.

Bigne, E., Ruiz, C., & Curras-Perez, R. (2019). Destination appeal through digitalized comments. *Journal of Business Research, 101*, 447-453. doi:https://doi.org/10.1016/j.jbusres.2019.01.020

Booyens, I., & Rogerson, C. M. (2017). Networking and learning for tourism innovation: evidence from the Western Cape. *Tourism Geographies, 19*(3), 340-361. doi:10.1080/14616688.2016.1183142

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication, 13*(1), 210-230.

Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software, 80*(4), 571-583. doi:https://doi.org/10.1016/j.jss.2006.07.009

Bruns, A., & Burgess, J. E. (2011). *The use of Twitter hashtags in the formation of ad hoc publics.* Paper presented at the Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011.

Buhalis, D., & Amaranggana, A. (2013). Smart tourism destinations. In *Information and communication technologies in tourism 2014* (pp. 553-564): Springer.

Bureau, S. C. (2018). Population and population changes. Retrieved from https://www.ssb.no/befolkning/statistikker/folkemengde

Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems, 31*(2), 102-107.

Casanueva, C., Gallego, Á., & García-Sánchez, M.-R. (2016). Social network analysis in tourism. *Current Issues in Tourism, 19*(12), 1190-1209. doi:10.1080/13683500.2014.990422

Chang, Y.-C., Ku, C.-H., & Chen, C.-H. (2019). Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. *International Journal of Information Management, 48*, 263-279. doi:https://doi.org/10.1016/j.ijinfomgt.2017.11.001

Choe, Y., & Fesenmaier, D. R. (2017). The Quantified Traveler: Implications for Smart Tourism Development. In Z. Xiang & D. R. Fesenmaier (Eds.), *Analytics in Smart Tourism Design: Concepts and Methods* (pp. 65-77). Cham: Springer International Publishing.

Chung, J. Y., & Buhalis, D. (2008). Information needs in online social networks. *Information Technology & Tourism, 10*(4), 267-281.

Cortez, P., Levachkine, S., & de la Cruz, C. (2014). Tourist Information Evaluation Using a Social Network. *Journal of Advances in Computer Networks, 2*(3).

Czernek-Marszałek, K. (2019). Applying mixed methods in social network research–The case of cooperation in a Polish tourist destination. *Journal of Destination Marketing & Management, 11*, 40-52.

de Vries, L., Peluso, A. M., Romani, S., Leeflang, P. S. H., & Marcati, A. (2017). Explaining consumer brand-related activities on social media: An investigation of the different roles of self-expression and socializing motivations. *Computers in Human Behavior, 75*, 272-282. doi:https://doi.org/10.1016/j.chb.2017.05.016

Dedeoğlu, B. B., Taheri, B., Okumus, F., & Gannon, M. (2020). Understanding the importance that consumers attach to social media sharing (ISMS): Scale development and validation. *Tourism Management, 76*, 103954. doi:https://doi.org/10.1016/j.tourman.2019.103954

Del Chiappa, G., & Baggio, R. (2015). Knowledge transfer in smart tourism destinations: Analyzing the effects of a network structure. *Journal of Destination Marketing & Management, 4*(3), 145-150. doi:https://doi.org/10.1016/j.jdmm.2015.02.001

Domínguez, S., & Hollstein, B. (2014). *Mixed methods social networks research: Design and applications* (Vol. 36): Cambridge University Press.

Federici, M., & Dragoni, M. (2016). *A knowledge-based approach for aspect-based opinion mining.* Paper presented at the Semantic Web Evaluation Challenge.

Filieri, R., Alguezaui, S., & McLeay, F. (2015). Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management, 51*, 174-185.

Foundation, S. M. R. (2019). Social Nework analysis tool NodeXL. Retrieved from https://www.smrfoundation.org/

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137-144. doi:https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Garton, L., Haythornthwaite, C., & Wellman, B. (1997). Studying online social networks. *Journal of computer-mediated communication, 3*(1), JCMC313.

Giffinger, R., Fertner, C., Kramar, H., & Meijers, E. (2007). City-ranking of European medium-sized cities. *Cent. Reg. Sci. Vienna UT*, 1-12.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational evaluation and policy analysis, 11*(3), 255-274.

Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: foundations and developments. *Electronic Markets, 25*(3), 179-188. doi:10.1007/s12525-015-0196-8

H.Tankovska. (2020, 28/01/2020). Most popular Facebook pages in Norway by number of Fan in 2019. Retrieved from https://www.statista.com/statistics/621805/most-popular-facebook-pages-in-norway/

Hanneman, R. A., & Riddle, M. (2005). Introduction to social network methods. In: University of California Riverside.

Hernández-Méndez, J., & Muñoz-Leiva, F. (2015). What type of online advertising is most effective for eTourism 2.0? An eye tracking study based on the characteristics of tourists. *Computers in Human Behavior, 50*, 618-625. doi:https://doi.org/10.1016/j.chb.2015.03.017

Heslinga, J. H., Hartman, S., & Wielenga, B. (2019). Irresponsible responsible tourism; observations from nature areas in Norway. *Journal of Tourism Futures*.

Hollstein, B. (2011). Qualitative approaches. *The Sage handbook of social network analysis*, 404-416.

Huang, N. (2016). Four important mobile marketing strategies for hotels. Retrieved from https://www.pegs.com/blog/four-important-mobile-marketing-strategies-for-hotels/

Huang, Y., Basu, C., & Hsu, M. K. (2010). Exploring Motivations of Travel Knowledge Sharing on Social Network Sites: An Empirical Investigation of U.S. College Students. *Journal of Hospitality Marketing & Management, 19*(7), 717-734. doi:10.1080/19368623.2010.508002

Inversini, A., & Buhalis, D. (2009). Information convergence in the long tail: The case of tourism destination information. In *Information and communication technologies in tourism 2009* (pp. 381-392): Springer.

J.Clement. (2019a, 31/07/2019). Average daily internet usage worldwide 2019 by age and device. Retrieved from https://www.statista.com/statistics/416850/average-duration-of-internet-use-age-device/

J.Clement. (2019b, 21/11/2019). Global Social networks ranked by numbers of user 2019. Retrieved from https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

J.Clement. (2020, 03/02/2020). Worldwide digital population as of January 2020. *Global digital population as of january 2020.* Retrieved from https://www.statista.com/statistics/617136/digital-population-worldwide/

Kallio, H., Pietilä, A.-M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing, 72*(12), 2954-2965. doi:10.1111/jan.13031

Kasavana, M. L., Nusair, K., & Teodosic, K. (2010). Online social networking: redefining the human web. *Journal of Hospitality and Tourism Technology, 1(1)*, 68-82.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering–a systematic literature review. *Information and software technology, 51*(1), 7-15.

Koleva-Tsankova, M. B. (2018). *Destination marketing within tourism industry in Norway.* University of Stavanger, Norway,

Kommune, F. a. H. (2018). Activity city in old town Fredrikstad. Retrieved from https://www.visitoestfold.com/no/fredrikstad-og-hvaler/aktiviteter-og-attraksjoner/Barn-og-familie/?TLp=1338474&Aktivitetsbyen-Gamle-Fredrikstad-tennis-fotballgolf-minigolf-og-golf

Kujur, P., & Chhetri, B. (2015). *Evolution of World Wide Web : Journey from Web 1.0 to web 4.0.* Paper presented at the International Journal of Computer Science and Technology.

Kumar, R. (2019). *Research methodology: A step-by-step guide for beginners*: Sage Publications Limited.

Ladhari, R., Rioux, M. C., Souiden, N., & Chiadmi, N.-E. (2019). Consumers' motives for visiting a food retailer's Facebook page. *Journal of Retailing and Consumer Services, 50*, 379-385. doi:https://doi.org/10.1016/j.jretconser.2018.07.013

Lauren Uppink Calderwood, M. S. (2019). *The travel and tourism competitiveness report 2019*. Retrieved from http://www3.weforum.org/docs/WEF_TTCR_2019.pdf

Lei, T., Ni, N., Zhu, Q., & Zhang, S. (2018). *Aesthetic Experimental Study on Information Visualization Design Under the Background of Big Data*, Cham.

Lewis-Pryde, J., & Evans, R. D. (2016). *A Social Networking Strategy for Improving Knowledge Management and Communication in the Travel Industry*. Paper presented at the Proceedings of the The 3rd Multidisciplinary International Social Networks Conference on SocialInformatics 2016, Data Science 2016, Union, NJ, USA. https://doi.org/10.1145/2955129.2955143

Li, X., & Law, R. (2020). Network analysis of big data research in tourism. *Tourism Management Perspectives, 33*, 100608. doi:https://doi.org/10.1016/j.tmp.2019.100608

Li, Y., Hu, C., Huang, C., & Duan, L. (2017). The concept of smart tourism in the context of tourism information services. *Tourism Management, 58*, 293-300. doi:https://doi.org/10.1016/j.tourman.2016.03.014

Liu, B. (2010a). Sentiment analysis and subjectivity. *Handbook of natural language processing, 2*(2010), 627-666.

Liu, B. (2010b, 01/01). *Sentiment analysis and subjectivity*.

Liu, Q., Zhang, H., Zeng, Y., Huang, Z., & Wu, Z. (2018). *Content attention model for aspect based sentiment analysis.* Paper presented at the Proceedings of the 2018 World Wide Web Conference.

Lo, I. S., McKercher, B., Lo, A., Cheung, C., & Law, R. (2011). Tourism and online photography. *Tourism Management, 32*(4), 725-731.

Lopez de Avila, A. (2015). *Smart destinations: XXI century tourism.* Paper presented at the ENTER2015 conference on information and communication technologies in tourism, Lugano, Switzerland.

Luo, Q., & Zhong, D. (2015). Using social network analysis to explain communication characteristics of travel-related electronic word-of-mouth on social networking sites. *Tourism Management, 46*, 274-282.

Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). *Learning word vectors for sentiment analysis*. Paper presented at the Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1, Portland, Oregon.

Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review, 27*, 16-32. doi:https://doi.org/10.1016/j.cosrev.2017.10.002

Marc, H. I. a. S. (2017). NodeXL. In *The International Encyclopedia of Communication Research Methods* (pp. 1-3).

Matloka, J., & Buhalis, D. (2010). Destination marketing through user personalised content (UPC). *Information and Communication Technologies in Tourism 2010*, 519-530.

Meier, H. (2019). NodeXL Pro Facebook Data Importer is no longer fuctional. *Insights into Network with NodeXL Pro.* Retrieved from https://www.smrfoundation.org/2019/09/05/nodexl-pro-facebook-data-importers-no-longer-functional/

Minazzi, R., & Lagrosen, S. (2013). *Investigating Social Media Marketing in the Hospitality Industry: Facebook and European Hotels*, Cham.

Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2007). *Measurement and analysis of online social networks*. Paper presented at the Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, San Diego, California, USA.

Mrvar, A., & Batagelj, V. (2016). Analysis and visualization of large networks with program package Pajek. *Complex Adaptive Systems Modeling, 4*(1), 6. doi:10.1186/s40294-016-0017-8

Ng, M. (2016). Factors influencing the consumer adoption of Facebook: A two-country study of youth markets. *Computers in Human Behavior, 54*, 491-500. doi:https://doi.org/10.1016/j.chb.2015.08.024

Nilsen, J. T. (2018). Fredrikstad. Retrieved from http://www.gonorway.no/norway/counties/ostfold/fredrikstad/7635f8bd10f1b30/index.html

Norway, I. (2017). *Key figures for Norwegian Tourism*. Retrieved from https://assets.simpleviewcms.com/simpleview/image/upload/v1/clients/norway/Key_Figures_2017_pages_9b3f82d5-43f4-4fe9-968c-7a85a36704b2.pdf

Norway, I. (2018). *Key figures for Norwegian travel and tourism 2018*. Retrieved from https://assets.simpleviewcms.com/simpleview/image/upload/v1/clients/norway/Key_figures_for_norwegian_tourism_2018_f9ac4f82-7b02-4fee-a67b-dcf98c4bd403.pdf

O'Reilly, T. (2012). What is web 2.0. *Design patterns and business models for the next generation of software. Accessed, 13*.

Okoli, C., & Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research.

Oliveira, T., Araujo, B., & Tam, C. (2020). Why do people share their travel experiences on social media? *Tourism Management, 78*, 104041. doi:https://doi.org/10.1016/j.tourman.2019.104041

Opdenakker, R. (2006). Advantages and Disadvantages of Four Interview Techniques in Qualitative Research. *Forum qualitative sozialforschung/forum: Qualitative social research, 7*(4). doi:10.17169/fqs-7.4.175

(2018). [Television series episode]. In Organization, W. T. (Executive producer), *International Tourism trends*

Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science, 28*(6), 441-453. doi:10.1177/016555150202800601

Park, J. H., Lee, C., Yoo, C., & Nam, Y. (2016). An analysis of the utilization of Facebook by local Korean governments for tourism development and the network of smart tourism ecosystem. *International Journal of Information Management, 36*(6, Part B), 1320-1327. doi:https://doi.org/10.1016/j.ijinfomgt.2016.05.027

Parra-López, E., Bulchand-Gidumal, J., Gutiérrez-Taño, D., & Díaz-Armas, R. (2011). Intentions to use social media in organizing and taking vacation trips. *Computers in Human Behavior, 27*(2), 640-654. doi:https://doi.org/10.1016/j.chb.2010.05.022

Parrott, W. G. (2001). *Emotions in social psychology: Essential readings*: Psychology Press.

Patel, K. (2013). Incremental journey for World Wide Web: introduced with Web 1.0 to recent Web 5.0–a survey paper. *International Journal of Advanced Research in Computer Science and Software Engineering, 3*(10).

Pookulangara, S., & Koesler, K. (2011). Cultural influence on consumers' usage of social networks and its' impact on online purchase intentions. *Journal of Retailing and Consumer Services, 18*(4), 348-354. doi:https://doi.org/10.1016/j.jretconser.2011.03.003

Potts, J., Cunningham, S., Hartley, J., & Ormerod, P. (2008). Social network markets: a new definition of the creative industries. *Journal of Cultural Economics, 32*(3), 167-185. doi:10.1007/s10824-008-9066-y

Prameswari, P., Surjandari, I., & Laoh, E. (2017, 25-26 Oct. 2017). *Opinion mining from online reviews in Bali tourist area.* Paper presented at the 2017 3rd International Conference on Science in Information Technology (ICSITech).

Ramos, C. M. Q., Andraz, G., & Cardoso, I. (2020). The Role of ICT in Involving the Tourist and in Sustainability of Tourism Destinations. In V. Ratten (Ed.), *Technological Progress, Inequality and Entrepreneurship: From Consumer Division to Human Centricity* (pp. 29-45). Cham: Springer International Publishing.

Ráthonyi, G. (2013). Influence of social media on tourism – especially among students of the University of Debrecen. *Applied Studies In Agribusiness And Commerce, 7*, 105-112. doi:10.19041/Apstract/2013/1/18

Reichstein, C., & Härting, R.-C. (2018). Potentials of changing customer needs in a digital world – a conceptual model and recommendations for action in tourism. *Procedia Computer Science, 126*, 1484-1494. doi:https://doi.org/10.1016/j.procs.2018.08.120

Rittichainuwat, B., & Rattanaphinanchai, S. (2015). Applying a mixed method of quantitative and qualitative design in explaining the travel motivation of film tourists in visiting a film-shooting destination. *Tourism Management, 46*, 136-147.

Roberta Milano , R. B., Robert Piatteli. (2011). *The effects of online social media on tourism websites*.

Roopchund, R. (2020). *Mauritius as a Smart Tourism Destination: Technology for Enhancing Tourism Experience*, Singapore.

S.Lock. (2019). TripAdvisor - Statistic & Facts. Retrieved from https://www.statista.com/topics/3443/tripadvisor/

S.Lock. (2020a). Economic contribution of travel and tourism to GDP worldwide 2006-2019. Retrieved from https://www.statista.com/statistics/233223/travel-and-tourism--total-economic-contribution-worldwide/

S.Lock. (2020b, 24-02-2020). Number of user reviews and opinion on Tripadvisor worldwide 2014-2019. Retrieved from https://www.statista.com/statistics/684862/tripadvisor-number-of-reviews/

Safko, L. (2010). *The social media bible: tactics, tools, and strategies for business success*: John Wiley & Sons.

Salas-Olmedo, M. H., Moya-Gómez, B., García-Palomares, J. C., & Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management, 66*, 13-25. doi:https://doi.org/10.1016/j.tourman.2017.11.001

Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems, 81*, 30-40. doi:https://doi.org/10.1016/j.dss.2015.10.006

Sapountzi, A., & Psannis, K. E. (2018). Social networking data analysis tools & challenges. *Future Generation Computer Systems, 86*, 893-913. doi:https://doi.org/10.1016/j.future.2016.10.019

SAS. (2019). Data Visualization. Retrieved from https://www.sas.com/en_au/insights/big-data/data-visualization.html

Schegg, R., Liebrich, A., Scaglione, M., & Ahmad, S. F. S. (2008). An exploratory field study of Web 2.0 in tourism. In *Information and communication technologies in tourism 2008* (pp. 152-163): Springer.

Schmallegger, D., & Carson, D. (2008). Blogs in tourism: Changing approaches to information exchange. *Journal of Vacation Marketing, 14*(2), 99-110. doi:10.1177/1356766707087519

Schouten, K., & Frasincar, F. (2015). Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering, 28*(3), 813-830.

Scott, J. P., & Carrington, P. J. (2011). *The SAGE Handbook of Social Network Analysis*: Sage Publications Ltd.

Seo, E.-J., & Park, J.-W. (2018). A study on the effects of social media marketing activities on brand equity and customer response in the airline industry. *Journal of Air Transport Management, 66*, 36-41. doi:https://doi.org/10.1016/j.jairtraman.2017.09.014

Serrat, O. (2015). *A Guide to Social Network Analysis*.

Sigala, M., & Marinidis, D. (2012). Web map services in tourism: a framework exploring the organisational transformations and implications on business operations and models. *International Journal of Business Information Systems, 9*(4), 415-434.

Smith, M. (2018). What do all of the overall graph metrics mean. Retrieved from https://www.smrfoundation.org/faq/what-do-all-of-the-overall-graph-metrics-mean/

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., & Qin, B. (2014). *Learning sentiment-specific word embedding for twitter sentiment classification.* Paper presented at the Proceedings of the

52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).

Tripadvisor. (2017). Tripadvisor network effect and the benefits of total engagement. Retrieved from https://www.tripadvisor.com/TripAdvisorInsights/w828

Tufte, E. R. (2001). *The visual display of quantitative information* (Vol. 2): Graphics press Cheshire, CT.

Valdivia, A., Luzón, M. V., & Herrera, F. (2017). Sentiment Analysis in TripAdvisor. *IEEE Intelligent Systems, 32*(4), 72-77. doi:10.1109/MIS.2017.3121555

Valdivia, A. A. L., M. Victoria; A Herrera, Francisco. (2017). Sentiment Analysis in TripAdvisor. *IEEE Intelligent Systems, 32*, 72 - 77. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8012330&tag=1

visitnorway. (2017). Tourism Norway. *Facts and figures on Norwegian toursim 2017.* Retrieved from https://www.visitnorway.com/media/tourism-in-norway/

Wei, Y., Song, P., & Rutherford, B. (2014). Corporate blogs, social media links and firm performance: A study of Fortune 500 companies. *Journal of Digital & Social Media Marketing, 2*(2), 159-175. Retrieved from https://www.ingentaconnect.com/content/hsp/jdsmm/2014/00000002/00000002/art00008

Wikipedia. (2018a). Fredrikstad. Retrieved from https://no.wikipedia.org/wiki/Fredrikstad

Wikipedia. (2018b). Old Town (Fredrikstad). Retrieved from https://no.wikipedia.org/wiki/Gamlebyen_(Fredrikstad)

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management, 31*(2), 179-188.

Xiang, Z., Magnini, V. P., & Fesenmaier, D. R. (2015). Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet. *Journal of Retailing and Consumer Services, 22*, 244-249.

Xie, J. (2020). The economic determinants of tourism seasonality: A case study of the Norwegian tourism industry. *Cogent Business & Management, 7*(1), 1732111. doi:10.1080/23311975.2020.1732111

Yoo, K.-H., Lee, Y., Gretzel, U., & Fesenmaier, D. R. (2009). Trust in travel-related consumer generated media. *Information and communication technologies in tourism 2009*, 49-59.

Yu, C. J. a. R. (2010). Travel industry uses Facebook and Twitter to reach customers. Retrieved from https://usatoday30.usatoday.com/tech/news/2010-09-07-travelsocialmedia07_CV_N.htm

Zhao, L., & Li, C. (2009). *Ontology based opinion mining for movie reviews.* Paper presented at the International Conference on Knowledge Science, Engineering and Management.

Appendix A

Consent form

My name is Pooja Shrestha and I am a student in the Applied Computer Science program at Østfold University College. I am currently working on my master thesis and therefor asking for your help to take part in the interview.

Please know that we value your privacy. All responses to this interview will be kept strictly confidential and the interview data will be reported in aggregate form. Your information is secure and will be used exclusively for research purposes.

- **Purpose:** The purpose of this project is to study social networking analysis and the issues related to tourism marketing of the Old town of Fredrikstad.
- **Procedures:** After you agree to participate in this interview, we will ask you a set of questions, it will take no more than 10 minutes.
- **Benefits:** Benefits include the ability to better understand the visitor's perspectives and learn what interest people online.
- **Risk:** The risk associated with taking part in this interview is minimal. If you feel any discomfort, you may voluntarily withdraw at any time.
- **Privacy:** Every effort will be made by the researchers to keep all information collected in this study private. Only the researchers will have access to the information and will be kept in a secure location. Once record discarded.
- **Subject rights:** You understand that taking part in this research is completely voluntary. You may refuse to answer any questions or withdraw your consent to take part in any part of this interview.
- **Questions:** If you have any questions please contact Pooja Shrestha pooja.shrestha@hiof.no or my supervisor Dr. Cathrine Linnes cathl@hiof.no

Sincerely
Pooja Shrestha
Østfold University College

Appendix B

Interview Questions

The following interview questions will be asked a set of visitors and business owners/workers.

1. What is your gender?
   a. Male
   b. Female

2. Which age category describes you?
   a. 75 and older   Maturist pre – 1945
   b. 60 – 74   Babyboomers 1945-1960
   c. 40 – 59   Generation X 1961-1980
   d. 25 – 39   Generation Y 1981-1995
   e. 18 – 24   Generation Z 1995

3. Are you a local resident or visitor?
   a. Local (Østfold)
   b. Visitor _____

4. What is the purpose of your visit today?
   a. Going for a walk
   b. Walking through
   c. Eat / Shop
   d. Work in the old city
   e. Live in the old city
   f. Other_____

Questions for Visitors

5. Which online site do you rely on the most when looking for tourist information (eat, shop, attraction, relaxation)?
   a. Facebook_____
   b. Twitter_____
   c. Company website_____
   d. Online newspaper_____
   e. Tourist authorities_____
   f. Booking sites (Expedia, Travelocity, Hotels.com)_____
   g. Other_____

6. If the rating online is <u>POOR</u>, what is your likelihood of going to visit the site?

   Scale 1 – 5 (1 is lowest and 5 is highest)

7. If the rating is <u>GOOD</u>, what is the likelihood of you going to visit the site?

   Scale 1 – 5 (1 is lowest and 5 is highest)

8. How often do you rely on these types of ratings?
   Scale 1 – 5 (1 is lowest and 5 is highest)

9. In regard to the Old City of Fredrikstad, what is your general impression? Any changes you would like?

Questions for Business Owners/Workers

5. How frequently do you make a post?

   Scale 1 -5 (1 is lowest and 5 is highest)
   Never, Yearly, Monthly, Weekly, Daily

6. How frequently do your reply to a comment?

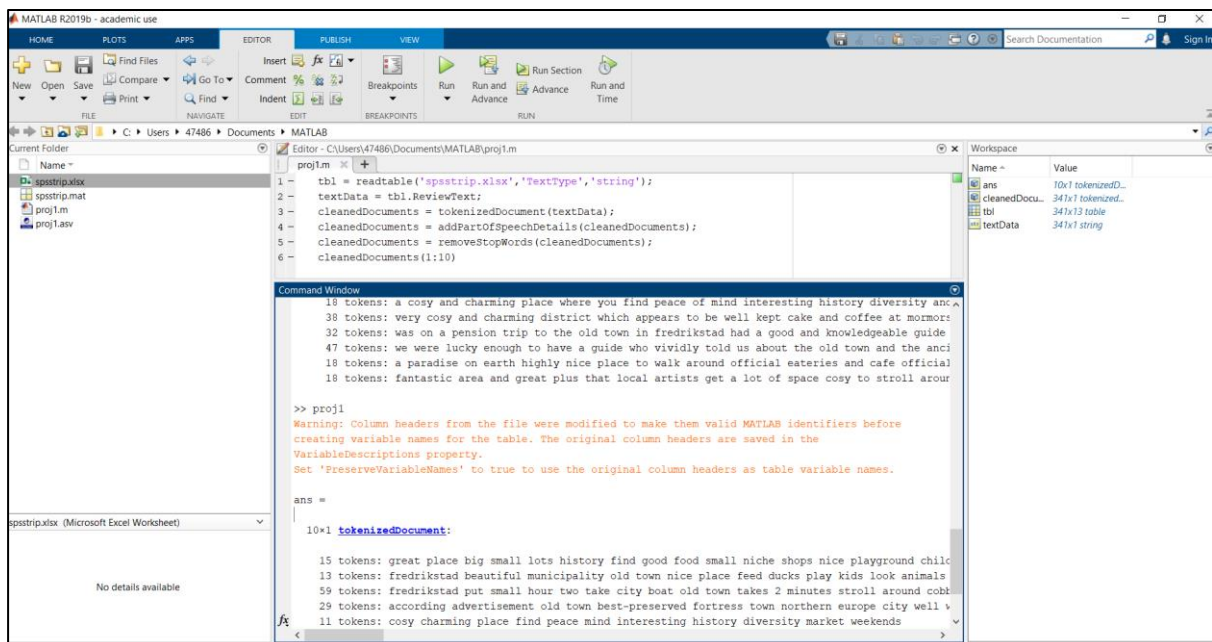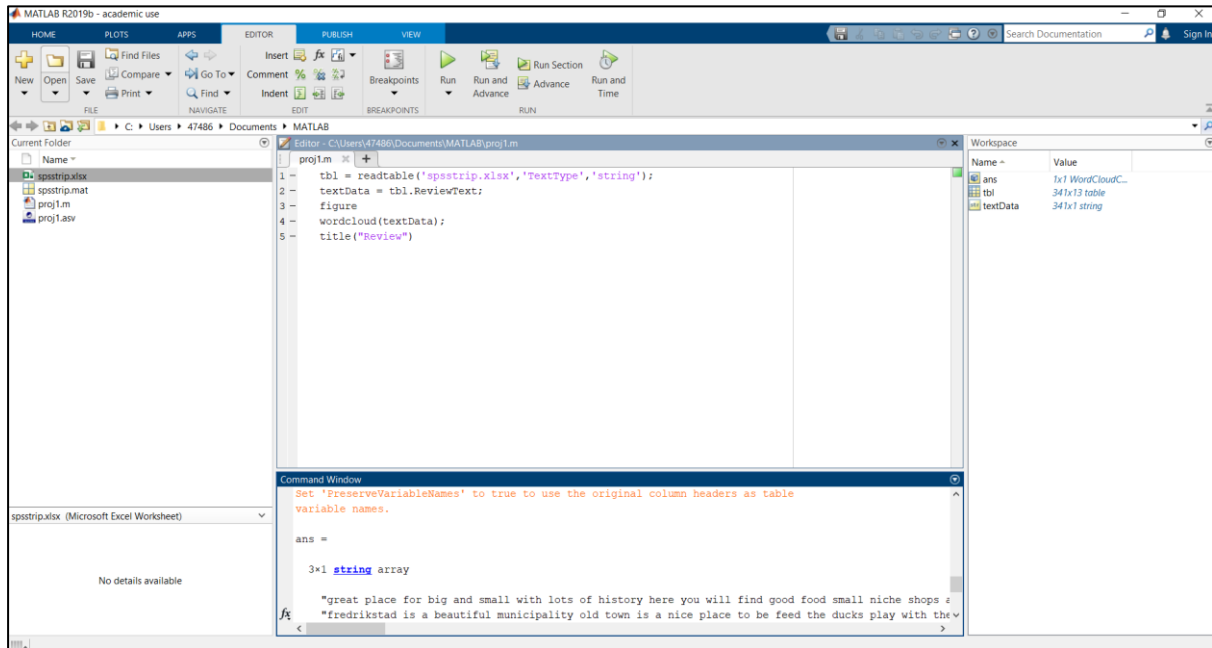   Scale 1 – 5 (1 is lowest and 5 is highest)
   Never, Yearly, Monthly, Weekly, Daily

7. Do you see any positive changes in your business due to the use of social media?

8. In regard to the Old City of Fredrikstad, what is your general impression? Any changes you would like?

Appendix C

The screen code of MATLAB:

```matlab
1   tbl = readtable('spsstrip.xlsx','TextType','string');
2   textData = tbl.ReviewText;
3   cleanedDocuments = tokenizedDocument(textData);
4   cleanedDocuments = addPartOfSpeechDetails(cleanedDocuments);
5   cleanedDocuments = removeStopWords(cleanedDocuments);
6   cleanedDocuments = normalizeWords(cleanedDocuments,'Style','lemma');
7   cleanedDocuments = erasePunctuation(cleanedDocuments);
8   cleanedDocuments = removeShortWords(cleanedDocuments,2);
9   cleanedDocuments = removeLongWords(cleanedDocuments,15);
10  cleanedDocuments(1:10)
```



```matlab
1   tbl = readtable('spsstrip.xlsx','TextType','string');
2   textData = tbl.ReviewText;
3   cleanedDocuments = tokenizedDocument(textData);
4   cleanedDocuments = addPartOfSpeechDetails(cleanedDocuments);
5   cleanedDocuments = removeStopWords(cleanedDocuments);
6   cleanedDocuments = normalizeWords(cleanedDocuments,'Style','lemma');
7   cleanedDocuments = erasePunctuation(cleanedDocuments);
8   cleanedDocuments = removeShortWords(cleanedDocuments,2);
9   cleanedDocuments = removeLongWords(cleanedDocuments,15);
10  cleanedBag = bagOfWords(cleanedDocuments)
11  cleanedBag = removeInfrequentWords(cleanedBag,2)
12
```
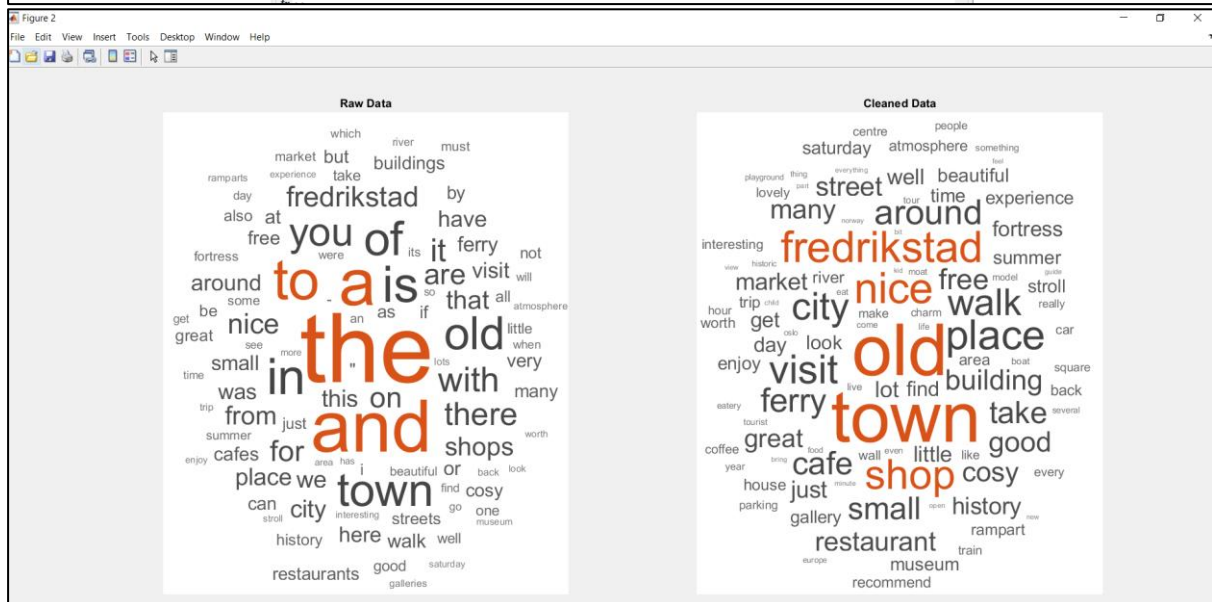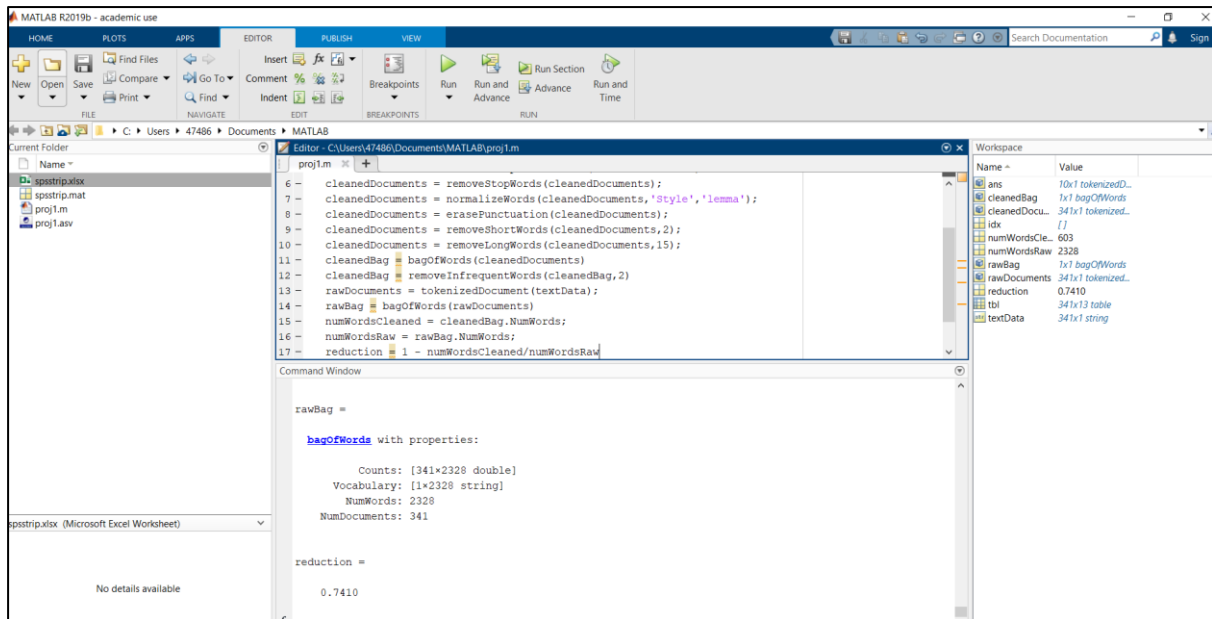
Command Window

```
    bagOfWords with properties:

            Counts: [341×1775 double]
        Vocabulary: [1×1775 string]
          NumWords: 1775
      NumDocuments: 341


cleanedBag =

    bagOfWords with properties:

            Counts: [341×603 double]
        Vocabulary: [1×603 string]
          NumWords: 603
      NumDocuments: 341
```

Raw Data   Cleaned Data

```matlab
tbl = readtable('spsstrip.xlsx','TextType','string');
textData = tbl.ReviewText;

cleanedDocuments = tokenizedDocument(textData);
cleanedDocuments =
addPartOfSpeechDetails(cleanedDocuments);
cleanedDocuments = removeStopWords(cleanedDocuments);
cleanedDocuments =
normalizeWords(cleanedDocuments,'Style','lemma');
cleanedDocuments = erasePunctuation(cleanedDocuments);
cleanedDocuments = removeShortWords(cleanedDocuments,2);
cleanedDocuments = removeLongWords(cleanedDocuments,15);
```

```matlab
cleanedBag = bagOfWords(cleanedDocuments)
cleanedBag = removeInfrequentWords(cleanedBag,2)
rawDocuments = tokenizedDocument(textData);
rawBag = bagOfWords(rawDocuments)
numWordsCleaned = cleanedBag.NumWords;
numWordsRaw = rawBag.NumWords;
reduction = 1 - numWordsCleaned/numWordsRaw
figure
subplot(1,2,1)
wordcloud(rawBag);
title("Raw Data")
subplot(1,2,2)
wordcloud(cleanedBag);
title("Cleaned Data")

tbl = readtable('spsstrip.xlsx','TextType','string');
textData = tbl.ReviewText;

cleanedDocuments = tokenizedDocument(textData);
cleanedDocuments =
addPartOfSpeechDetails(cleanedDocuments);
cleanedDocuments = removeStopWords(cleanedDocuments);
cleanedDocuments =
normalizeWords(cleanedDocuments,'Style','lemma');
cleanedDocuments = erasePunctuation(cleanedDocuments);
cleanedDocuments = removeShortWords(cleanedDocuments,2);
cleanedDocuments = removeLongWords(cleanedDocuments,15);
cleanedBag = bagOfWords(cleanedDocuments)
cleanedBag = removeInfrequentWords(cleanedBag,2)
rawDocuments = tokenizedDocument(textData);
rawBag = bagOfWords(rawDocuments)
numWordsCleaned = cleanedBag.NumWords;
numWordsRaw = rawBag.NumWords;
reduction = 1 - numWordsCleaned/numWordsRaw
figure
subplot(1,2,1)
wordcloud(rawBag);
title("Raw Data")
subplot(1,2,2)
wordcloud(cleanedBag);
title("Cleaned Data")
function documents = preprocessReview(textData)

% Tokenize the text.
```

```matlab
documents = tokenizedDocument(textData);

% Remove a list of stop words then lemmatize the words.
To improve
% lemmatization, first use addPartOfSpeechDetails.
documents = addPartOfSpeechDetails(documents);
documents = removeStopWords(documents);
documents = normalizeWords(documents,'Style','lemma');

% Erase punctuation.
documents = erasePunctuation(documents);

% Remove words with 2 or fewer characters, and words with
15 or more
% characters.
documents = removeShortWords(documents,2);
documents = removeLongWords(documents,15);

end
```