Universitetet i Stavanger

FACULTY OF SOCIAL SCIENCES,
NORWEGIAN SCHOOL OF HOTEL MANAGEMENT

MASTER'S THESIS

STUDY PROGRAM: Master's degree in service leadership in International Business	THESIS IS WRITTEN IN THE FOLLOWING SPECIALIZATION/SUBJECT: Tourism IS THE ASSIGNMENT CONFIDENTIAL? No

TITLE: Impact of Online Reviews for Selecting Norway as a Tourist Destination Choice – Drawn in by Theory of Planned Behavior

AUTHOR		SUPERVISOR:
Student number: 268607	Name: Sahiba Inayatullah	Prof. Olga Gjerald

Acknowledgment

I would like to express my heartfelt gratitude to all those who have played a significant role in the completion of my Master's thesis. First and foremost, I would like to extend my deepest appreciation to my supervisor, Professor Olga. Her invaluable guidance, insightful feedback, and constant encouragement have been instrumental in shaping this thesis. Her expertise, dedication, and unwavering support have been truly inspiring.

I would also like to acknowledge the immense contribution of Professor Torvald, my former mentor on this project and my course instructor for MHR101: Research Methods. His extensive knowledge, challenging insights, and mentorship have greatly influenced my research work. I am grateful for his guidance and the valuable lessons I have learned under his tutelage.

Special thanks go to my loving husband, Omer for his unwavering support, patience, and guidance throughout this demanding journey. My parents, for their relentless belief in my abilities that have been a constant source of motivation. And my deepest gratitude to my (now) 7 months old son, Mustafa, for being my guiding light and been the driving force behind my determination to succeed. His presence has truly impacted my life’s perspective in most positive and profound way, and I am forever grateful my loved ones’ presence in my life.

I would like to extend my gratitude to my acquaintances, Sumaira and Suleiman. I am deeply appreciative of their unwavering availability, patience, and genuine dedication in assisting me in navigating the complexities of this study. Thank you for your commitment to my academic development and for sharing your expertise selflessly.

Finally, I would like to acknowledge the contributions of all the individuals who have been a part of this research, whether directly or indirectly. Your participation and assistance have been invaluable.

Words cannot adequately express my gratitude to each and every one of you for your unwavering support and guidance throughout this thesis. I am truly honoured to have had such incredible individuals by my side, and I will forever be grateful for the opportunity to work with each of you.

Abstract

Purpose: In today's digital age, the integration of digitalization into our daily work and business processes has become increasingly important. Information is often considered as the new gold and innovative companies are producing revenues from the vast amounts of data available online. Particularly, the surge in e-commerce activity shows the use of online platforms. With people and businesses conducting transactions online, customers frequently share their opinions and experiences through online reviews, which have become pivotal in shaping consumer decisions across various industries, including tourism. In this study, I attempted to leverage the wealth of online information to explore its influence on travel decision-making, specifically regarding trips to Norway.

Methods: Using a statistical and computational approach, I aimed to analyze the impact of online reviews on travel intentions, drawing upon the Theory of Planned Behavior (TPB). To achieve this, I devised a novel hybrid methodology that integrated data from Google-Form surveys and reviews sourced from TripAdvisor.

Around 8000 online reviews from 17 tourist destinations across Norway from TripAdvisor were used for data analysis purpose. Prominent themes were identified and analyzed through Python 3. Futhermore, sentiment intensity analysis were conducted to process the online reviews and classified them into the components of the TPB framework. This allowed us to systematically evaluate the factors influencing travel decisions.

For cross-comparison of findings, a Google-form was filled by 59 respondents comprising of mainly two cohorts; International Tourists who traveled to Norway and International Students in Norway. Subsequently, regression analysis was applied to both, the Google-form Survey responses and TripAdvisor reviews to assess the relationship between electronic Word of Mouth (eWOM) and the intention to travel to Norway.

Results: The regression analysis revealed a significant correlation between eWOM and the intention to travel to Norway, highlighting the importance of online reviews in shaping travel behavior. Regression models of both cohorts showed significant results by representing a good percentage (60% and 80%) of variance in Attitude, Subjective norm, Perceived Control and eWOM. I was also able to analyze the ~8000 reviews from TripAdvisor and classified them in the framework of Theory of Planned Behavior and applied regression analysis on them.

Originality / Value: The outcome of this study is a novel method of a hybrid approach for studying the relationship between Attitude, Subjective Norm, Perceived Control and eWOM with the Intention to travel using regression analysis. This work is a step forward towards utilizing the already available data for understanding the behavior of travelers. This information can be useful for Institutions promoting tourism in their respective countries, as well as for businesses in tourism industry for upscaling.

Keywords: Tourist Destination; Online Reviews; eWOM; Theory of Planned Behavior; Content Analysis; Python, User-generated Content, Decision Making Process.

List of Abbreviations

CSV	Comma-Separated Values
eWOM	Electronic Word of Mouth
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
NLTK	Natural Language Toolkit
PCA	Principle Component Analysis
TPB	Theory of Planned Behavior
URL	Uniform Resource Locator

Table of Contents

List of Tables

Table of Figures

Chapter 1: Introduction:

1.1 Background of the Study

Electronic Word of Mouth (eWOM) is a term that recently is been used for the exchange of people's opinions, experiences, and suggestions about companies, brands, services, or travel destinations through digital platforms such as social media, video blogging, live videos, online reviews, and in text blogs. In the current time of a digitally connected globe, eWOM has not only become an unavoidable tool for shaping end-user behavior and influencing their travel and destination choice decisions, but also impacting countries' reputation for tourism. This research thesis explores the concept of eWOM, its uses, and its implications for businesses, institutions, related government departments and consumers Brown et al. (2007). In this modern times of digitalization, traditional word-of-mouth has evolved into electronic word-of-mouth, that covers all forms of human communication and interaction on digital platforms including Snapchat, Instagram and tiktok etc (Cheung & Lee, 2012). eWOM includes online reviews, social media posts, blogs, video making platforms, new papers and websites, online forums, Facebook groups, pages and other user-generated content that can influence consumer perceptions about different travel destinations, buying decisions, attitudes towards countries, and behaviors towards products and companies.

1.2 Types of eWOM

Although with such speed of digitization, every day there are new ways of spreading word-of-mouth, but I would try to gather and cover some of the famous forms of eWOM. The top and most classic is Product Reviews where consumers share their opinions and experiences with items they purchased or services they used. It is mainly done through a user-friendly interface provided by the app or the company themselves in written format on online platforms such as Temu, Amazon, Finn, Yelp, and TripAdvisor etc. The second way or type of eWOM is Social Media Posts. In this type users share their thoughts, recommendations, suggestions or complaints about items and companies on social networking sites like Snapchat, Twitter, Facebook, reddit, medium, Instagram, and LinkedIn. Third most famous type is through Blogs and Forums where influencers, celebrities, bloggers, stars, and consumers engage in discussions, share product recommendations, and give good or bad feedback on dedicated platforms and video making websites like tiktok and YouTube. Fourth form is Online Ratings in which users rate purchased items or services using star ratings, thumbs up/down, or vote up or vote down, or selection of an item using radio button or by creating polls, or any other rating systems available on various online platforms.

1.3 Uses of eWOM

Although use of eWOM nowadays can just be limitation of the mind as it’s influence is not limited to forming opinions related to buying products and services, but also shaping attitude and holistic view of end-user about contemporary world overall, i.e: opinion-forming related to and ranging from electoral processes of countries to raising awareness about social issues. Discussing a few directly related ones here. First, it influences Purchase Decisions; Online buyers mostly rely on reviews and recommendations from other users to evaluate products or services before using a service or purchasing a product or for traveling to a destination (Filieri & McLeay, 2014). Some of the main websites are purely based on online reviews like TripAdvisor etc. Positive reviews can increase brand credibility and reach more people, hence leading to increased sales and customer trust. The second most common use is Brand Reputation Management; In this era most of the companies closely monitor and invest resources in managing eWOM to maintain and increase their company's reputation. In some cases, a negative review, a fake campaign to malign the name or viral social media posts can damage a company's repute and credibility. Therefore, companies engage in online image management strategies to deal with the customer complaints immediately and process the potential damage. The third most common use of eWOM is Marketing and Advertising (Hajli et al., 2017). eWOM is a cost-effective marketing tool for businesses. The money that was used to buy and rent billboards is now invested in the social media posts that cost less money and have greater reach. In these cases, paid positive reviews and recommendations from customers act as endorsements, influencing potential buyers' perceptions and driving sales. Companies hire celebrities to post on their social media with their brand and it reaches millions of people in a split second. Brands use influencer marketing and user-generated posts and videos to increase their brand messaging and reach a broader audience through video-making apps and other websites. eWOM is widely used for Market Research and Customer Insights**.** eWOM provides valuable insights into users' preferences, priorities, interests, opinions, and trends. Companies and third-party software analyze online conversations, sentiment analysis, and social listening tools to gather business intelligence, identify new trends, and understand costumer behavior in a more intelligent and educated way. This kind of intelligence-based knowledge then leads the product development, marketing strategies, and customer service initiatives. eWOM also encourages the creation of online communities where like-minded individuals share common interests, experiences, and passions in the form of facebook groups, pages they like and many other community tools. Brands leverage these communities to engage with customers, process and understand the feedback, and design product strategy (Hennig-Thurau et al., 2004). By building strong relationships with consumers, businesses can enhance brand loyalty and advocacy.

In conclusion, electronic word-of-mouth (eWOM) has been adopted as a powerful tool in the digital era, which influences consumer behavior in purchasing products, selecting a vacation destination, brand reputation, and marketing strategies. Current companies must recognize and acknowledge the importance of eWOM and actively engage with consumers on digital platforms to leverage its benefits effectively. By understanding the uses and implications of eWOM, businesses can harness its potential to drive sales, build brand loyalty, and gain a competitive edge in the marketplace.

1.4 Relevance of the Study

In the modern world of digitalization, eWOM has emerged as an important factor influencing user behavior in several fields, especially in tourism. Countries now have specialized ministries and departments to manage their reputation online by hiring social media and YouTube influencers. eWOM includes the sharing of thoughts, opinions, experiences, and recommendations about travel destinations, countries, accommodations, attractions, and services through different websites such as social media websites, review websites like Trip Advisor, video and photo-making apps like Instagram and Snapchat, and travel forums etc. In this thesis I will be validating this relevance of electronic word-of-mouth on people's behavior and then their intention to select Norway as a travel destination. eWOM plays an important role in changing travelers' perceptions and expectations of destination countries. Positive eWOM contributes to the creation of a favorable destination image, attracting more tourists and boosting tourism revenue, hence improving its economy as well (Doosti et al., 2016). A lot of modern travelers rely on eWOM to gather information and plan their trips in an efficient way. Content and reviews generated by the real users provides valuable insights into attractions, and activities which eventually leads to educated decision making (Litvin et al., 2008). eWOM uses the power of social influence, as travelers need validation and recommendations from their social networks. Engaging with eWOM allows travelers to share their experiences, seek advice, and connect with like-minded individuals, fostering a sense of community (Wang, 2014). eWOM plays a multidimensional role in the tourism industry, affecting traveler behavior, perception about the people in destination countries, and market dynamics.

1.5 Scope of the Study

Although, there can be multifaceted study and analysis done for eWOM in tourism, but scope of this current thesis will include online review analysis, testing of Theory of Planned behavior model, validation through online surveys, and comparison between Google-form Surveys and online reviews extracted from tripadvisor.com

Reviews obtained from both sources are subjected to content analysis to identify common sentiments and patterns as per Theory of planned Behavior’s three main attributes. This qualitative analysis helps in understanding the Attitude towards selecting Norway as travel destination, Subjective Norms, Perceived Control Behavior and Intention to travel to Norway. Utilizing natural language processing (NLP) techniques, sentiment analysis of reviews will be performed to categorize reviews into positive, negative, or neutral sentiments but I will also find the intensity of positivity or otherwise. This quantitative approach enables the quantification of review sentiment across categories of Theory of Planned Behavior. A comparative analysis will be conducted to see the findings from TripAdvisor reviews with the insights derived from Google Form surveys. This comparison allows for the identification of discrepancies, similarities, and areas of improvement in consumer feedback from different sources.

1.5.1 Objectives of the Study:

The purpose of the study is to validate the Theory of planned behavior model by eWOM effects in deciding Norway as travel destination. One of the main objectives of this work is to attempt a novel method of designing a hybrid approach to validate TPB. I used

Google Forms as a source of using new data, and reviews from TripAdvisor as a motivation to use existing data for the analysis and application of a model. Here I will attempt to explain the Theory of Planned Behavior:

Theory of Planned Behavior (TPB):

The TPB is a prominent psychological framework used to understand and predict human behavior (Ajzen, 1991) It explains that behavioral intentions are influenced by three key factors: Attitude towards the behavior, In tourism, attitudes refer to travelers' positive or negative evaluations of a destination or service. The second is Subjective norms which means perceived social pressure to perform the behavior. Subjective norms in tourism involve the influence of social and peer opinions on traveler choice of destination and the third is Perceived behavioral control which tell us about the perceived ease or difficulty of performing the behavior, traveling in our study. This aspect relates to the fact that how easy or difficult is it to travel to a destination for example Norway. These factors collectively shape a travelers' intention to travel. The TPB can be applied to understand the factors influencing consumers' intentions to engage in eWOM activities, such as posting reviews or recommendations online and then taking the decision to travel or not. Attitudes towards reading opinions and reviews online, perceived social norms regarding eWOM behavior, and perceived control over expressing opinions digitally are unavoidable factors in tunning travelers' intentions to engage in traveling.

The purpose of this thesis is to explore the interplay between the Theory of Planned Behavior and Electronic Word-of-mouth using online reviews from Tripadvisor.com and Google-form Surveys. Specifically, the study aims to investigate how attitudes, subjective norms, and perceived behavioral control influence consumers' intentions to travel to Norway using existing data resources. Here I have done text analysis on large datasets, by using the power of sentiment analysis tools and intensity measurement libraries to categorize the reviews into TPB and then done detailed regression analysis to find the variables affecting the most. I have also done feed-forward and feed-backward regression analysis to validate and present the best possible model trained using Google form surveys (new data resource) and reviews from Tripadvisor (existing data resource).

1.5.2 Research Question and Hypothesis

For this study, I am using a similar hypothesis as Jalilvand and Samiei (2012) suggested in their model for testing intention to travel, in the context of Norway as a tourist destination. Following hypotheses will be tested against the constructs:

Hypothesis 1: Online reviews (eWOM) have a positive impact on attitudes towards visiting Norway.

Hypothesis 1a: Attitudes toward visiting Norway has a positive impact on intention to travel.

Hypothesis 2: Online reviews (eWOM) have a positive impact on subjective norms.

Hypothesis 2a: Subjective norm has a positive impact on intention to travel.

Hypothesis 3: Online reviews (eWOM) have a positive impact on perceived behavioral control.

Hypothesis 3a: Perceived behavioral control has a positive impact on the intention to travel.

Hypothesis 4: Online reviews (eWOM) have a positive impact on intention to travel.

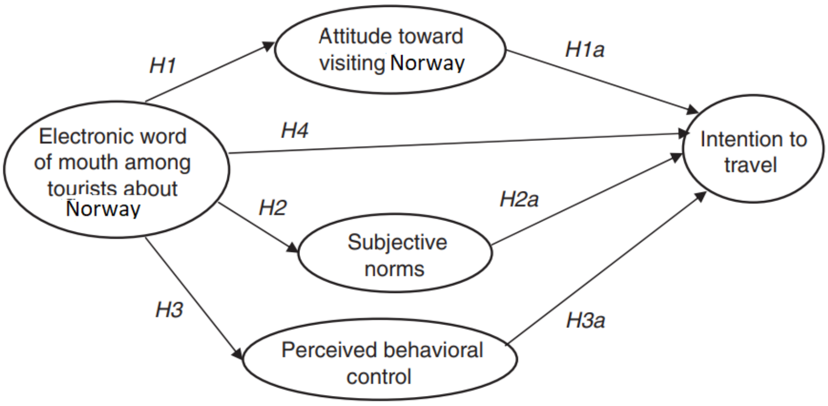

Figure 1: Research model

The figure shows a model which was originally produced by Jalilvand and Samiei (2012), as visual representation of their hypotheses used for the study in the context of Isfahan as their destination choice. I have added it here with modification in the context of Norway for our study. Adapted from "The impact of electronic word-of-mouth on a tourism destination choice: Testing the theory of planned behavior (TPB)," by M.R. Jalilvand and N. Samiei, 2012, Internet research, Vol. 22 Issue 5, Pages 591-612. Copyright reserved with the original authors as quoted above. I will be testing this model on data from both Google Forms and Reveiws from TripAdvisor.

1.6 Outline of the Subsequent Chapters

I have talked about the eWOM and theory of planned behavior along with the purpose and scope of study in the introduction chapter, the second chapter comprises of literature review and related work, third chapter describes the methodology, and the fourth chapter discusses results. The last chapter is about the conclusion and discussion.

Chapter 2 Literature review:

The Theory of Planned Behavior (TPB), developed by Icek Ajzen in the late 1980s, is a widely recognized psychological framework used to understand and predict human behavior (Ajzen, 1991). This theory posits that an individual's behavioral intention is influenced by three primary factors: Attitude towards the behavior, Subjective Norms, and Perceived Behavioral Control (Cheung & Thadani, 2012). In terms of electronic word-of-mouth (eWOM), the TPB may provide valuable insights into how travelers engage with and respond to online reviews and recommendations given by vloggers and tourists on different websites.

In our study, Attitude towards travel play an important role in the TPB framework as it is one of the main components of the model which reflects on an individual's evaluation of the behavior and its associated outcomes (Ismagilova et al., 2020). The travelers' attitude towards online reviews influences their likelihood of engaging with and trusting the information provided about the people, cultures and norms about the destination. Positive attitude towards eWOM may lead individuals to perceive online reviews as trusted and valuable sources of information. Therefore, it can hugely influence their intentions and behaviors towards traveling to a destination.

The second main component of the TPB is the Subjective normative. The Subjective norm explains the perceived pressure that you have on you from your peers or expectations from relatives and close people regarding the behavior in question (Nilashi et al., 2022). In regards of eWOM and intention to travel, the subjective norms may be catered to through influence from around and individual, such as the suggestions and recommendations of close friends, family, or online influencers from across the platforms. Positive subjective norms regarding online reviews may reinforce people's trust in electronic word-of-mouth, leading to increased engagement with and reliance on online review platforms like Trip Advisor (Fernandes et al., 2022).

Then Perceived behavioral control means an individual's belief in their ability to perform the behavior successfully (Ajzen, 2020). In terms of eWOM, perceived behavioral control has factors such as perceived ease of use and confidence in online review platforms such as Tripadvisor as I have considered for our study.

Moreover, if I consider the study and model of the TPB framework with the study of eWOM, it can provide valuable insights into users' online review sentiments, their behaviors, and decision-making processes of traveling to a specific destination influenced by those sentiments and behaviors. Governments and departments can develop more effective marketing strategies for managing online reputation, enhancing the credibility of the destination, and influencing traveler perceptions and behaviors(Litvin et al., 2008) by investigating the underlying factors affecting individuals' attitudes, subjective norms, and perceived behavioral control toward eWOM. The model of the theory of planned behavior provides a theoretical base for investigating the impact of enhancing positive eWOM behaviors and mitigating the impact of negative online reviews on traveler's perceptions (Renzi & Klobas, 2008).

A lot of work has been done to study Theory of Planned Behavior to test and validate different behaviors like buying, executing an act, traveling etc. The influence of social media and eWOM on tourism destination choice has also been a field of focus. Towards understanding members' general participation in and active contribution to an online travel community.

In 2004, Wang et al. employed qualitative interviews and content analysis, and the study investigated travelers' participation in online travel communities. This work performed in-depth exploration of online community dynamics, while the subjectivity of qualitative analysis is a drawback (Wang & Fesenmaier, 2004). One of the studies used a quantitative survey approach. This study examined the relationships between customer perceived value, satisfaction, loyalty, and switching costs. They performed a rigorous statistical analysis, while their work has some common method biases in self-reported data (Yang & Peterson, 2004). In 2004, Hennig et al tried to capture the answer to question; What Motivates Consumers to Articulate Themselves on the Internet? Their work utilized qualitative interviews and content analysis to investigate consumer motivations for engaging in electronic word-of-mouth (eWOM) on the internet. The strengths are the in-depth exploration of consumer behavior, while study have some potential biases in self-reported motivations (Hennig-Thurau et al., 2004). Litvin, Goldsmith, and Pan (2008) investigated electronic word-of-mouth in hospitality and tourism management, providing a foundational understanding of eWOM dynamics. The paper has also briefly discussed the issue of potential abuse of online reviews. However, the study's did not discuss its ability to capture longitudinal trends (Litvin et al., 2008). Another interesting study I found is by Huang et. Al. where through a quantitative survey, their research explored a cognitive-affective model of destination image. They have used of a comprehensive model, while the complexity of measuring cognitive and affective components cannot be ignored (San Martín & Del Bosque, 2008). Sparks et al studied the impact of online reviews on hotel booking intensions where they used quantitative surveys, where this study investigated the influence of online reviews on hotel booking intentions and trust perception in 2011. This work also included the large-scale data collection like our work, while weakness is that they included social desirability biases in self-reported intentions (Sparks & Browning, 2011). There is another study which explored persuasive design principles in destination recommendation systems by applying persuasion theories to Tripadvisor. The study's strength lies in its innovative approach to enhancing user experience, but limitations exists in the complexity of implementing persuasive design elements in existing platforms (Kim & Fesenmaier, 2017). Another study in 2016 investigated the impact of social media reviews on consumer decisions using an extended TPB. This study has performed a good job in integration of social media marketing theories with established behavioral models. The generalizability of findings to diverse social media platforms is limited (Rauschnabel et al., 2016). A study extended the Technology Acceptance Model and TPB to understand users' acceptance of voice-based interactive services in 2019. This article innovative approach addresses emerging technologies such as voice and a system in hospital called voice smart care (Jian et al., 2022). In 2019, A study employed a quantitative survey approach to examine the impact of social media and electronic word-of-mouth (eWOM) on tourists' destination choices. Strengths include the use of robust research methodology and real-world data collection. However, potential limitations may arise from self-reporting biases inherent in survey research (Tham et al., 2013). Many works attempted to validate the TPB by using surveys. One such work is done by employing quantitative surveys, the work examined the impact of online reviews on hotel booking intentions, considering language and tone. A positive point of this study is that they included the incorporation of linguistic analysis and they presented a detailed review of eWOM and TPB, while some sample are added as representativeness and do not have the real user (Huete-Alcocer, 2017). Through quantitative surveys, this research explored the influence of electronic word-of-mouth (eWOM) on travel intentions using the Theory of Planned Behavior (TPB) model. Strengths include the use of a theoretical framework, while limitations may arise from potential biases in self-reported intentions (Leung et al., 2015). This study used consideration set theory to model the impact of online reviews on consumer decision-making. They have used hotel reviews that included both positive and negative reviews. They used real-world data, while this work include biases in declaring that negative reviews do not cause harm (Vermeulen & Seegers, 2009).

Li, Liu, and Du in 2021 used the Theory of Planned Behavior (TPB) to understand travelers’ intentions to visit memorial tourism destinations. Although the TPB offers a robust framework for analysis, but the study face challenges in capturing the nuanced motivations behind tourists' decisions to engage with sensitive heritage sites (Dimitrovski et al., 2017). To examine the effects of online reviews on hotel booking intentions, incorporating perceived trust into the TPB, a study was conducted. By integrating trust dynamics into behavioral models, the research provided valuable insights into the role of trust in eWOM, yet potential biases in data influenced the results (Chakraborty, 2019). A recent study explored electronic word-of-mouth (eWOM) in wellness tourism, explaining the emerging trends in traveler behavior. This article helps to understand electronic Word-of-mouth dynamics in a tourism segment. This study showed a significant effect of the destination scene on the wellness of tourist satisfaction and his intention to travel based on online reviews. The shortcoming of the study is that it was conducted during the COVID-19 pandemic and sample size was very small (Goyal & Taneja, 2023). In 2021, one of the studies presents a comprehensive analysis of past, present, and future trends in electronic word-of-mouth (EWOM), offering valuable insights into the evolution of online consumer behavior. It clearly explains the longitudinal progression of the field. However, the study's focus on synthesizing existing that limited empirical contributions and it can use a recommender system to further improve the studies (Verma & Yadav, 2021). This other article investigated the influence of user-generated content on traveler behavior, focusing on the effects of eWOM on hotel online bookings. The article attempted to contribute in understanding eWOM effects statistically and presented a regression model (Ye et al., 2011). Most of these studies used Surveys and some also worked in existing data. I believe there is a lack of a study where we can see how many of the answers and insights are comparable from the analysis of both new sources of data and already existing data sources from the internet.

Chapter 3 Methodology:

3.1 Data Collection and Treatment

3.1.1 Location of Research

This research was conducted as part of a master thesis at the University of Stavanger, Norway. As mentioned earlier, the primary data has been collected through online surveys. Whereas the secondary and main source of data was collected through the website Tripadvisor.com, there I selected several different destinations within Norway to include diversity in this research study. Here are the following destinations that I have added:

Table 1: List of destinations

S No	Destination	City
1	Nordisk bibelmuseum	Oslo
2	Tromso Fjords	Tromso
34	Mount Floyen and the Funicular	Bergen
5	Bakklandet	Trondheim
6	Lofoten Krigsminne Museum Svolvaer Vagan	Lofoten Islands
7	Gruve3 Longyearbyen Spitsbergen	Svalbard
8	Brekkefossen Flam	Aurland Municipality
9	Slaatta Skisenter Geilo Snowsports	Geilo Hol
10	The Norwegian National Opera Ballet	Oslo
11	The Bergen Railway	Bergen
12	Munch	Oslo
13	Karl Johans gate	Oslo
14	Astrup Fearnley Museet	Oslo

15	Vigeland Museum	Oslo
16	Viking Ship Museum	Oslo
17	Nasjonalmuseet National Museum	Oslo
18	Holmenkollen Ski Museum	Oslo
19	Akershus Castle and Fortress Akershus Slott og Festning	Oslo

The above table shows a list of destinations. The first column shows the serial number, the second column shows the name of the destination from where reviews were extracted from the TripAdvisor website while the third column shows the city of the destination.

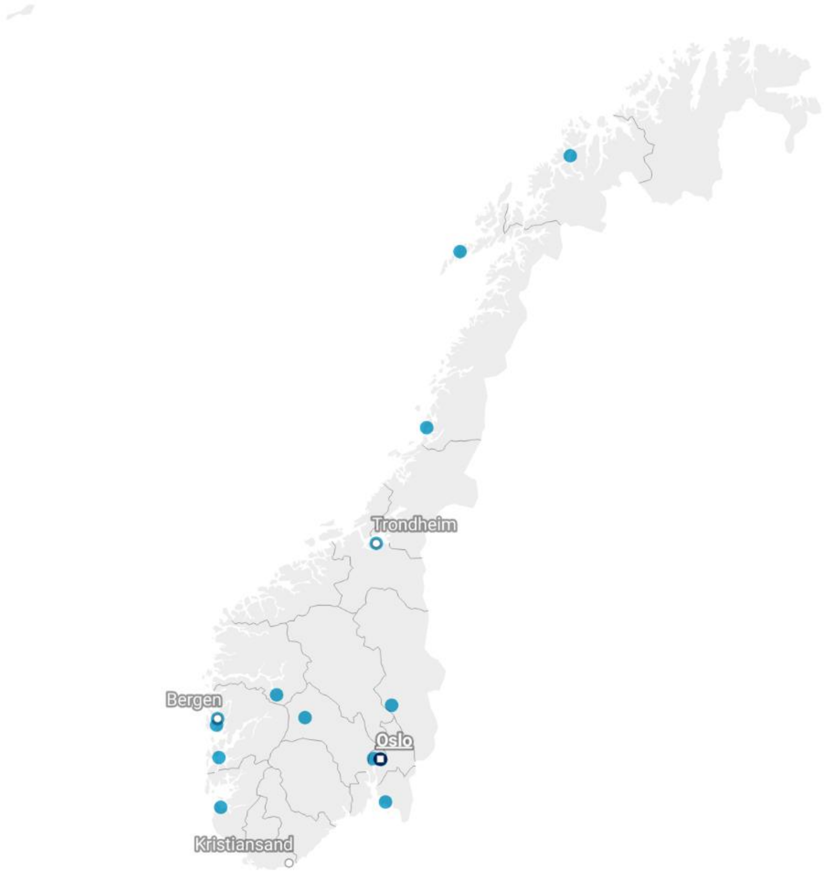

Figure 2: Distribution of destinations across the Norway

This figure shows a map of Norway with destinations marked in blue. Details of destinations are also given the table.

3.1.2 Populace:

For one part of this study, I have used reviews from Trip Advisor, which encompasses a diverse range of contributors spanning various geographical locations worldwide. This emphasizes the breadth and diversity of reviewers on the platform. This description highlights the global nature of Trip Advisor's user base, indicating that the reviews utilized in the study are not limited to a specific demographic or geographic region. By incorporating reviews from individuals worldwide, the analysis benefits from a rich and varied pool of perspectives, enhancing the comprehensiveness and relevance of the research findings. I adopted this approach to acknowledge the global reach and influence of Trip Advisor as a platform for user-generated content, underscoring its significance as a valuable source of information for academic inquiry and decision-making processes and validating the model of the Theory of Planned Behavior.

Here I present a table to show demographics of the study by showing top 10 countries whose residents' reviews are included:

Table 2: List of top 10 countries of tourists

Position	Country	Frequency
1	United Kingdom	1341
2	United States	1297
3	Australia	587
4	Norway	538
5	Canada	213
6	Romania	174
8	Germany	169
9	India	127
10	Italy	107

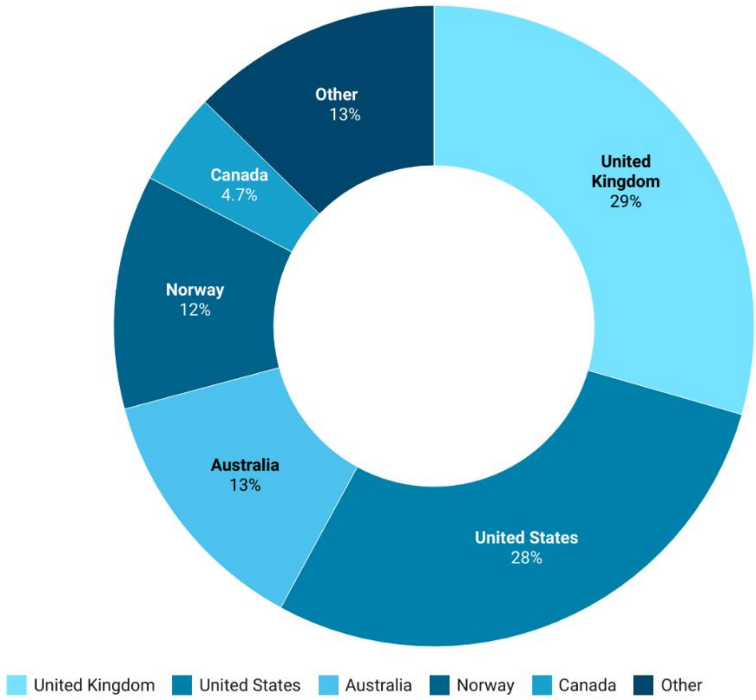

Figure 3:Demographic Distribution of Reviewers of top 10 countries

This Figure shows the distribution of frequencies of travellers from top 10 countries. There were 29% of top 10 countries from United Kingdom, 28% were from US, Australia and Norway were 13% and 12% respectively. Other countries in top 10 were Italy, India, Romania, Germany and Sweden.

For the second part of this study, where I collected primary data through Google forms, a random sample of 59 respondents were taken. Out of which cohort 1 comprised of 25 International Students in Norway (eWOM was tested, along with TPB), while cohort 2 consisted of rest of 34 respondents, who were International Tourists (Only TPB was tested). For the simplification purposes, information on their demographics was not collected. However, it was ensured that respondents are either visiting Norway as international students or had visited Norway atleast once in their life (International Tourists).

3.1.3 Method of Sampling:

Studies usually focus on surveys to test TPB but here I adopted a hybrid approach to mainly test TripAdvisor reviews but also validate models using Google-form Surveys (random sampling) and perform a comparative analysis. The study utilizes two main data collection methods:

- **TripAdvisor Web Scraping**: Reviews from TripAdvisor, a popular travel platform, are extracted using web scraping techniques by Python. These reviews provide valuable insights into customer experiences, opinions, and ratings related to travel destinations, and attractions. I have been able to extract 8000 reviews from 17 destinations and approximately 1600 web pages. I also noted gender, names, reviews, the country from where the user belongs, date of travel, and the company with which they traveled to the destination of their choice.

- **Google Form Surveys**: Online surveys are conducted using Google Forms to gather structured feedback from participants through random sampling. The survey questions were designed to capture diverse aspects of TPB which incorporates attitudes, subjective norms, control behavior and intention to travel toward the subject of study. Those questions were combinedly adapted from Sparks and Pan (2009) study and modified in context of Norway and in the context of past tense , while the original paper had questions in the future tense. I also included questions related to eWOM, which were adapted from Bambauer-Sachse and Mangold (2011) study and modified it in tourism context. Multiple items were used to measure each variable construct (see Appendix). Seven-point scale was used to measure elements of TPB (Attitude, Subjective norms, and Perceived control behavior), whereas five-point scale was used to measure eWOM and Behavioral Intention, as per original questionnaire format. Internal consistency was measured by Cronbach's alpha for both cohorts in the Google form survey responses. In total, I was able to gather 59 samples and applied regression analysis to test the hypothesis.

3.2 Data Analysis Technique

As I was able to extract around ^8000 reviews, on which I have applied regression analysis to test the Theory of Planned Behavior and Hypotheses that I wanted to evaluate. There were some steps that needed to be taken to get the data into a shape suitable for TPB. These preprocessing steps were necessary to know where a review lies. I first did some text mining where I classified the review into 'Attitude towards a behavior', 'Subjective norms', 'Perceived control' and 'intention to perform that behavior'. For this purpose, I was supposed to do sentiment analysis. As I was supposed to mimic exactly the same strategy for Google Form surveys, I decided to find the intensity of the review as well. I used an intensity score of 1 as strongly disagree to 7 as strongly agree in our Google form surveys on same questions as are in previously used questionnaire in this study. Then I used the intensity of these reviews to assign an intensity score ranging from 1 to 7, for which I used a library of NLTK text intensity finder in Python (Loper & Bird, 2002). I have written our extraction and analysis codes in Python language. Codes can be provided upon request.

3.2.1 Sentiment Intensity Analyzer:

There is a tool for text analysis using natural language processing methods. One famous tool that is implanted in Python is called NLTK. The NLTK Sentiment Intensity Analyzer is a component of the Natural Language Toolkit (NLTK), which is a well-known platform for building Python programs to work with language and text data. The Sentiment Intensity Analyzer is a specialized method designed for sentiment analysis tasks, aiming to quantify and analyze the sentiment expressed in text data which I used here to mimic the Google-form Surveys answers.

Components of NLTK Sentiment Intensity Analyzer:

Lexicon-Based Approach: The Sentiment Intensity Analyzer uses a lexicon-based approach, where sentiment scores are assigned to each word based on their orientation in a predefined dictionary.

Intensity Scoring: It assigns numerical values to each word indicating the strength of its sentiment, allowing for the computation of overall sentiment intensity scores for sentences. I used this technique in our work to make it comparable with the Google-form Survey answers and to perform more analysis with numbers.

Sentence-Level Analysis: The Analyzer performs sentence-level sentiment analysis by aggregating the sentiment scores of individual words to derive the sentiment of the entire sentence.

Normalization: It is also important to make sure that accurate sentiment analysis across different texts is done. So, normalization techniques are applied to handle variations in word usage, contexts of the sentences and words, and sentence structure.

Here I discuss the examples of sentiment intensity analyzers. A review will be considered a positive review is it has this text: "The stunning fjords of Norway are breathtaking." This sentence has a higher Sentiment Score showing that intensity is high. Another example of negative sentiment is "The weather in Norway was dismal and ruined our trip." This is a negative review with moderate intensity. Words like terrible, and horrible are more intense in terms of negativity. A neutral review would look like this: "Norway offers a diverse range of attractions for travelers.". Here I have discussed simple examples of sentiments and their intensity but reviews in real life are usually complex. However, analyzing sentiment in reviews can provide insights into travelers' experiences, preferences, and perceptions about different destinations in and outside Norway. It can help assess the overall satisfaction levels of tourists visiting various attractions in Norway. It also allows tourism authorities and businesses to identify areas for improvement and address concerns raised by travelers and design marketing strategies accordingly.

In summary, here in our work the NLTK Sentiment Intensity Analyzer served as a valuable tool for extracting insights from reviews about travel destinations in Norway, enabling governments and institutions to make data-driven decisions to enhance the overall tourism experience for the tourists in Norway.

3.2.2 Rating Score:

The sentiment intensity analyzer outputs some values that consist of four keys; neg: that shows a negative sentiment score and lies between 0 and 1. The second key is ‘neu’ which shows the neutral sentiment score and the range of values is between 0 and 1. The third value is ‘pos’ and it shows the positive sentiment score between 0 and 1. The last one gives an overall overview of the whole text which is named as compound. The compound value gives us the overall sentiment score between -1 and 1. I used the value compound in our analysis where I divided this range of -1 to 1 into 1 to 7 as per our Google-form Surveys values.

3.3 Explanation of Codes:

To do a customized analysis, I have written some codes. The first code was designed to extract reviews from TripAdvisor and this code can be reused just by giving the URL of any trip advisor page as input. I then did sentiment intensity analysis and then divided the reviews into categories like Attitude toward behavior, subjective norms, and perceived control. At the end I designed a code to do regression analysis of Google-form Surveys and reviews as well. Here I will explain the codes in detail.

3.3.1: Reviews scrapping from TripAdvisor.com:

I wrote Python code that extracts reviews from TripAdvisor URLs for the seventeen travel destinations across Norway. I prepared a URLs list that contains a list of TripAdvisor URLs for different pages of reviews for the different attractions provided above in the thesis. I Iterate through each URL in the list. Sends an HTTP GET request to the URL to fetch the webpage content. I also do an error check if the response status code is 403 (Forbidden). If yes, skips processing for that URL. Otherwise, I process the page. I provided a unique identifier for the attraction from the URL to give a unique id to every review that I store. I have parsed the HTML content of the webpage using BeautifulSoup (Richardson, 2007) library.

I search HTML elements containing ratings, names, review headers, reviews, dates, countries, and contributions. I extracted data from these elements and stored them in lists. I did data handling by organizing the extracted data into rows, including columns for name, rating, review header, review text, company, date, country, and attraction and appended each row to a DataFrame for storing. I printed the DataFrame containing the extracted data to make sure that I have what is needed. To print a summary on screen I printed the total number of processed URLs and successful extractions. I saved the extracted data to a CSV file named "data_reviews.csv". The code is designed to systematically scrape reviews from multiple pages of a TripAdvisor attraction and store them for further analysis or storage. This general code be used by anyone just by giving the list of URLs of TripAdvisor and it will extract and store the information.

3.3.2 Sentiment analysis code:

I provided Python code that implements sentiment analysis, a computational technique used to determine the emotional tone of a review. I used the Natural Language Toolkit (NLTK) and the VADER lexicon, a rule-based sentiment analysis tool, to assess the sentiment of textual reviews. The code first downloads the VADER lexicon, necessary for sentiment analysis. It defines functions to calculate sentiment scores and ratings based on the polarity of the text. The sentiment score represents the overall sentiment intensity, while the rating categorizes the sentiment into discrete levels, ranging from 1 as highly negative to 7 as highly positive. The code reads each review from a CSV file, conducts sentiment analysis on each review using the defined functions, and appends the sentiment scores and ratings to the DataFrame. Finally, it exports the analyzed data to a new CSV file for further analysis or visualization. I demonstrated the process of loading, preprocessing, and analyzing textual data using Pandas, a Python data manipulation library. Overall, the code can be used as a practical implementation of sentiment analysis techniques for analyzing textual data in any review from any website.

3.3.3 Classification of reviews:

The code uses libraries like Pandas, and Counter for counting review ratings to have an overall idea of where the sentiment lies. Data is loaded from a CSV file named "sentiment_data_reviews.csv" that contains the reviews that I extracted and performed sentiment intensity analysis into a Pandas DataFrame. The code creates masks for words like "enjoy", "enjoying", "enjoyed", and "enjoyable" to identify positive sentiments, and masks for words like "bad", "good", "worst", "worsen", and "better" to identify negative sentiments. These masks are combined using logical operators to filter the DataFrame and extract relevant reviews. The filtered results are saved into separate CSV files ("sentiment_data_reviews_enjoy.csv" and "sentiment_data_reviews_bad_good.csv"). This is to mimic the questions in Google-form Survey forms. In the same way, I have classified all reviews by searching words and creating masks to divide them into eWOM, attitude towards visiting Norway, subjective norms, perceived behavior and intention to travel. Finally, the sentiment ratings are counted and printed for each category ('Enjoy/Unenjoy' and 'Good/Bad'). I also created Boolean Masks for Each Word: Boolean masks are created for words associated with certain sentiments such as 'foolish', 'fun', 'silly', 'idiotic', and 'crazy'. These masks identify whether each review in the Data Frame contains any of these words, ignoring case sensitivity and handling missing values. The masks created in the first step are combined using logical operators, specifically the OR operator, to create a single composite mask. This composite mask identifies reviews containing any of the specified words, indicating belonging to a category.

The analysis results are then printed to show the ratio of all sentiment reviews to the total number of reviews in the subset. The Counter function is used to count the occurrences of sentiment ratings in the filtered results. More explanation of this can be found in the result section under the heading of descriptive investigation and review categorization.

3.3.4 Regression analysis:

I wrote Python code that conducts a regression analysis using the Pandas, Scikit-learn, Numpy, and Statsmodels libraries. In the first part, the code imports necessary libraries such as Pandas, Scikit-learn, and Numpy. It then reads an Excel file that contains the answers of the Google-form Surveys into a pandas DataFrame named 'df'. I utilized Principal Component Analysis (PCA) from scikit-learn to reduce the dimensionality of the data. It applies PCA separately to different subsets of the DataFrame 'df' and it is because I have used many questions to cater the ‘attitude towards traveling to Norway’ while I needed this as one variable in our regression analysis. Before the PCA I also transformed data using Min-Max scaling to bring all values on the same ranges of levels. The dependent variable was taken as ‘I’ stating the intention to travel in our analysis. While eWOM, Attitude, Subjective norms and Perceived control are considered as independent variables. A new DataFrame named 'df_final' is created, containing columns 'A', 'S', 'P', ‘E’ and 'I' which store the transformed PCA components and the calculated dependent variable 'I'. Using Statsmodels, the code specifies the independent variables ('A', 'S', 'P', ‘E’) and adds an intercept to them. It defines the dependent variable 'y' as 'I'. Then, it fits an Ordinary Least Squares (OLS) regression model to predict 'I' using the independent variables. Finally, the code prints the summary statistics of the fitted regression model, including coefficients, standard errors, t-values, and p-values for each predictor variable.

This code essentially performs regression analysis to predict the dependent variable 'I' using principal components 'A', 'S', ‘E’ and 'P'. The PCA and MinMaxScaler are used to preprocess the data, and the OLS regression model is employed for prediction and analysis.

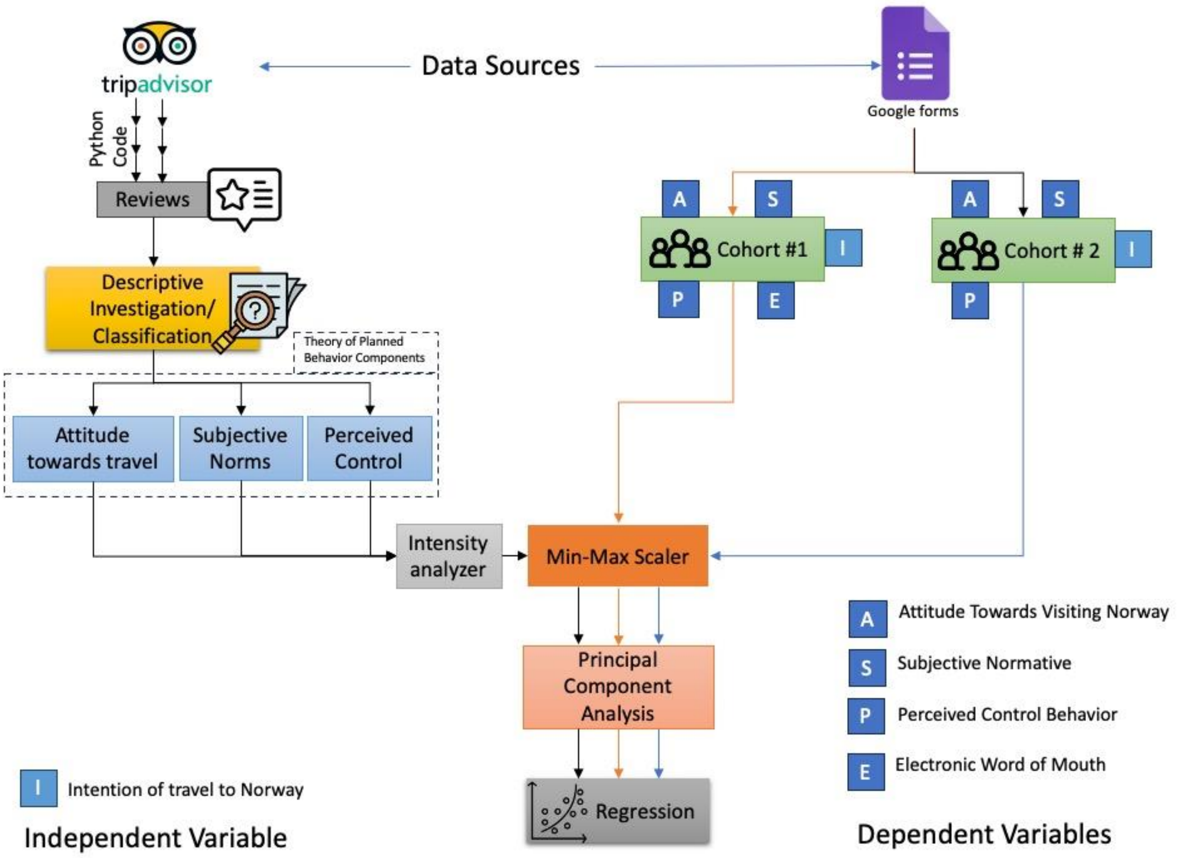

Figure 4: Overview of Complete Methodology

This figure shows the complete layout of this thesis. It can be seen that there are two data sources used in the work; TripAdvisor and Google forms. Reviews are extracted from the tripadvisor.com using a Python code. Reviews are then subjected to descriptive investigation and classification into components of Theory of Planned Behaviour. Sentiment Intensity is then calculated using an Intensity Analyser. A min-max scaler is applied and then PCA is applied to reduce the number of responses for each component to 1 variable. At the last regression analysis is perform. For google surveys responses, they were directly subjected to min max scaler, PCA and then regression analysis was performed for each cohort.

Chapter 4: Results

To diversify our analysis, I have divided our responses based on two cohorts. The first cohort was of international students where I have taken eWOM, Attitude towards travel, Subjective norms and Perceived control as our independent variables. The first cohort consisted of 25 respondents. In the second cohort, I had visitors and those who are on visa status other than international students. These included skilled workers, tourist visitors, and some permanent residents etc. The second cohort of people had 34 respondents. In the following section I have described the results of both cohorts.

4.1: Regression Analysis:

4.1.1 Regression with data of first cohort – Google-form Survey:

I performed regression analysis on 24 respondents where I asked questions regarding four independent variables that consisted of eWOM, Attitude towards travel, Subjective norms and Perceived control. There were six questions regarding electronic word-of-mouth labeled as ‘E’, six questions cater ‘Attitude towards travel to Norway’ denoted as ‘A’, four questions were asked about ‘Subjective normative influence’ denoted as ‘S’, three questions about ‘Perceived Control’ denoted as ‘P’, and two questions to find out the ‘Intention towards traveling to Norway’. Intention towards traveling to Norway was considered as our dependent variable while other three were dependent variables denoted here as ‘I’.

The data was collected through a Google-form Survey, with questions targeting eWOM, attitudes towards traveling, subjective norms, perceived control, and intention to travel. There were multiple questions targeting hypotheses and I labeled the responses as E1, E2, E3 for eWOM, A1, A2 and so on for attitude towards travel related questions, S1, S2 for subjective norms and P1, P2 for perceived control. The responses were subjected to regression analysis, with the intention to travel as the dependent variable. In the preprocessing step of the data before regression, min-max scaling was applied to normalize the values, and principal component analysis (PCA) was used to reduce the dimensions of the variables. By reducing the question responses, I get one value for each of our hypotheses along with the dependent variable.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      I   R-squared:                       0.812
Model:                            OLS   Adj. R-squared:                  0.772
Method:                 Least Squares   F-statistic:                     20.52
Date:                Mon, 06 May 2024   Prob (F-statistic):           1.11e-06
Time:                        18:44:44   Log-Likelihood:                -20.714
No. Observations:                  24   AIC:                             51.43
Df Residuals:                      19   BIC:                             57.32
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       5.551e-17      0.132   4.22e-16      1.000      -0.275       0.275
A              0.4543      0.078      5.851      0.000       0.292       0.617
S              0.0675      0.105      0.643      0.528      -0.152       0.287
P              0.1391      0.111      1.255      0.225      -0.093       0.371
E              0.0628      0.082      0.766      0.453      -0.109       0.235
==============================================================================
Omnibus:                        0.534   Durbin-Watson:                   2.002
Prob(Omnibus):                  0.766   Jarque-Bera (JB):                0.623
Skew:                          -0.284   Prob(JB):                        0.732
Kurtosis:                       2.452   Cond. No.                         2.88
==============================================================================
```

Figure 5: Regression summary for first cohort

The regression analysis yielded significant insights into the factors influencing intention to travel. R-squared value was 0.81 which means the model explains 81.2% of the variance in intention to travel, indicating a strong predictive power. The adjusted R-squared value of 0.77 suggests that approximately 77% of the variance is explained by the independent variables, adjusting for the number of predictors. F-statistic with a value of 20.52 and a corresponding p-value of 1.11e-06, the F-statistic indicates that the regression model is statistically significant.

Coefficient of Attitudes (A): The coefficient of 0.4543 ($p < 0.001$) suggests a significant positive relationship between attitudes towards traveling and intention to travel.

Subjective Norms (S): The coefficient of 0.0675 ($p = 0.528$) indicates a non-significant relationship between subjective norms and intention to travel.

Perceived Control (P): With a coefficient of 0.1391 ($p = 0.225$), perceived control shows a non-significant effect on the intention to travel.

Electronic Word-of-mouth (E): The coefficient of 0.0628 ($p = 0.453$) suggests a non-significant association between eWOM and intention to travel.

The Cronbach alpha for first cohort 1 is as follows: A is 0.92, S is 0.91, P is 0.76, I is 0.76 and Cronbach alpha for E is 0.83, which is quite significant and above threshold limit of 0.7 (Nunnally, 1981).

The findings highlight the significant role of attitudes towards traveling in shaping individuals' intention to travel. However, subjective norms, perceived control, and eWOM do not show significant effects on the intention to travel in this study. These results underscore the importance of considering individual attitudes and perceptions in understanding travel behavior.

4.1.2: Regression with data of second cohort – Google-form Survey:

I then performed regression analysis on 34 respondents where I asked questions regarding three independent variables that consisted of Attitude towards travel, Subjective norms and Perceived control. There were six questions regarding 'Attitude towards travel to Norway' denoted as 'A', four questions were asked about 'Subjective normative influence' denoted as 'S', three questions about 'Perceived Control' denoted as 'P', and two questions to find out the 'Intention towards traveling to Norway'. Intention towards traveling to Norway was considered as our dependent variable while other three were dependent variables denoted here as 'I'.

```
==============================================================================
Dep. Variable:                      I   R-squared:                       0.602
Model:                            OLS   Adj. R-squared:                  0.562
Method:                 Least Squares   F-statistic:                     15.11
Date:                Mon, 29 Apr 2024   Prob (F-statistic):           3.56e-06
Time:                        22:40:16   Log-Likelihood:                -39.707
No. Observations:                  34   AIC:                             87.41
Df Residuals:                      30   BIC:                             93.52
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       2.776e-17      0.142   1.95e-16      1.000      -0.290       0.290
A             -0.0389      0.076     -0.514      0.611      -0.193       0.116
S              0.3865      0.100      3.872      0.001       0.183       0.590
P              0.4017      0.124      3.243      0.003       0.149       0.655
==============================================================================
Omnibus:                        3.005   Durbin-Watson:                   2.577
Prob(Omnibus):                  0.223   Jarque-Bera (JB):                1.727
Skew:                          -0.381   Prob(JB):                        0.422
Kurtosis:                       3.798   Cond. No.                         2.01
==============================================================================
```

Figure 6: Regression summary for second cohort

The regression analysis yielded significant insights into the factors influencing intention to travel. R-squared value was 0.60 which means the model explains 60% of the variance in intention to travel, indicating a strong predictive power. The adjusted R-squared value of 0.56 suggests that approximately 56% of the variance is explained by the independent variables, adjusting for the number of predictors. F-statistic with a value of 15.11 and a corresponding p-value of 3.56e-06, the F-statistic indicates that the regression model is statistically significant.

Coefficient of Attitudes (A): The coefficient of -0.0389 ($p = 0.611$) suggests a non-significant negative relationship between attitudes towards traveling and intention to travel.

Subjective Norms (S): The coefficient of 0.3865 ($p < 0.005$) indicates a significant relationship between subjective norms and intention to travel.

Perceived Control (P): With a coefficient of 0.40 ($p < 0.005$), perceived control shows a positive significant effect on intention to travel.

The Cronbach alpha for first cohort 2 is as follows: A is 0.86, S is 0.32, P is 0.78, I is 0.43, which shows values of A and P are significant and above the threshold limit of 0.7 (Nunnally, 1981). One justification for low Cronbach alpha for S and I could be low-inter item correlation.

4.2 Feedforward regression analysis

To check every angle of analysis, I also conducted feed forward analysis on google survey responses, where I performed the analysis on each individual dependent variable. I first run the regression analysis with only one variable Attitude towards travel, then Subjective Normative and so on but the results of all combined had the highest representation of variance of variable.

4.3 Descriptive Investigation for the categorization of reviews:

I also performed a descriptive analysis of our reviews to categorize them into three categories of TPB namely 'Attitude towards behavior', 'Subjective Norms' and 'Perceived Control'. The analysis revealed a diverse distribution of reviews across various dimensions related to all these categories.

The Python code implements this noval methodology to filter and extract specific reviews using predefined keywords related to the sentiment of "enjoyment" or its variants. As can be seen in the appendix, sentiment of enjoyment is a question covered in Attitude towards travel section. By looking for the related reviews, I tried to cater those questions of respective category. The keyword search was performed using the str.contains() method. Each mask searches for variations of the keyword "enjoy" in the 'review' column of the DataFrame, ignoring case sensitivity. Variations such as "enjoying," "enjoyed," and "enjoyable" are included to ensure comprehensive coverage. Combining masks of individual masks are then combined into a single composite mask using logical operators (specifically, the bitwise OR operator |). This combination ensures that any review containing at least one of the specified keywords will be selected. Such as a review “We really enjoy traveling around the European countries, this summer we wanted to experience the natural beauty of Fjords in Norway” is categorized in ‘Attitude towards a behaviour’.

The combined mask is applied to the original data of reviews using boolean indexing, resulting in a new data of selected reviews containing only the rows that match the combined conditions. Additionally, the sentiment ratings of the extracted reviews are counted and displayed using the Counter function, providing insight into the distribution of sentiment ratings among the selected reviews. For example, out of 100 reviews in the ‘Attitude’ category 47 reviews were given an intensity of 7 on a scale of 1 to 7 using the intensity analyzer code, 20 reviews were 6 score, 18 reviews were 1 score, 9 reviews were assigned a score of 2, 1 review got 4, 3 reviews were scored 5 and 2 reviews were scored 3. I do get results in the same way for all the categories.

Among the aspects of attitude, the categories of "enjoy," "good_bad," "fun_foolish," "pleasure," "favorable," and "like" garnered varying levels of attention, with the highest count observed in the "favorable" category, totaling 1126 reviews. In terms of subjective norms, categories such as "popular among family and friends," "important people in my life recommended," and "I have heard in past 12 months" received considerable engagement, with 831, 565, and 764 reviews, respectively. In the theory of planned behavior, the Perceived control including "enough time to travel," "enough money to travel," and "if I wanted to travel," also attracted attention, although to varying extents, with review counts of 965, 101, and 681, respectively. Furthermore, eWOM (electronic word-of-mouth) generated interest, with 421 reviews, while the intention to travel to Norway, represented by the category "intention," saw 187 reviews. These findings offer valuable insights into the perceptions, sentiments, and intentions of travelers regarding various facets of traveling to Norway. This is a noval method that I tried in our study where I attempted to categorize the reviews and applied regression to them as an average. I searched for words and and then tried to filter using the masking of these words as describe earlier in the text.

4.4 Regression Analysis Results on Reviews:

After extracting the reviews, and extensive analysis I was able to get the intensity values for our reviews. As stated in the methods above, I prepared our data for the regression analysis. Here, I attempted to take same A1, A2, A3 up to A6 for Attitude towards traveling to Norway, P1, P2, P3 for perceived control, S1, S2, S3 for subjective norms and E1 for eWOM. The reason for taking multiple instances of A, P, and S was to mimic the results from Google forms. The responses in the form of reviews with its sentiment intensity set from 1 to 7 were then subjected to regression analysis, with the intention to travel as the dependent variable as I have done in Google-form Survey. The model explained only 5% of the variance in dependent variables. In the preprocessing step of the data before applying regression, min-max scaling was applied to normalize the values and get them all in the same range, and principal component analysis (PCA) was used to reduce the dimensions of the variables to 1 PCA. By reducing the question responses, I get one value for each of our hypothesis along with the dependent variable in a separate data frame. As we got different numbers of reviews in different categories, we applied regression to the 100 reviews from each component from the TPB.

4.5 Reviews Vs Google Form Survey:

One of the objectives of our study was to understand the connection between the reviews and Google form surveys. These two data sources refer to two types of data; new data taken using a survey and existing data from TripAdvisor. I could see that a survey is filled completely by only one person while in terms of reviews, a review that is classified in Attitude category is given by some person while review that is classified in Perceived Control is written by some other person. For that reason, there might not be a very good connection between the responses and hence the model does not cover a lot of variances of the dependent variables. However, the results shown from a survey part usually highlight an overall pattern of the regression line and the same can be achieved using the reviews.

Chapter 5: Discussion:

There has been a lot of work done on sentiment analysis. Research has also been performed on electronic word-of-mouth and online reviews. Multiple fields also attempted to fit the data on Theory of planned behavior. Here in our work, I attempted a noval method of trying out TPB with reviews from the trip advisor. Our findings offer valuable insights into the perceptions, sentiments, and intentions of travelers regarding traveling to Norway. This is a noval method that I tried in our study where I attempted to categorize the reviews according to TPB and applied regression to them as an average. I searched for words and and then tried to filter using the masking of these words. The study's findings have implications for various stakeholders, including businesses, marketers, and policymakers. By understanding and leveraging insights from online reviews, governments and tourism departments can enhance customer satisfaction, tailor their offerings to meet consumer preferences, and improve overall service quality. I also successfully validated the TPB using our data from the Google-form Surveys and regression results shows significant dependence of our dependent variables. It acknowledges the hypotheses that I made. This work is also a valuable initiative to conduct hybrid studies as I used a hybrid model of online reviews and Google-form Surveys forms. Although our r-squared value is not extremely high in case of Google-form Surveys but it is very significant in the field of behavioral sciences. In some fields like behavioral sciences, R-squared values of 0.3 or 0.4 may be considered very good, while in others, such as physics or engineering, values over 90% may be expected due to precise measurements. (Maximus, 2020)

It is essential to acknowledge the limitations of the study. I believe work and specifically regression results of the TripAdvisor can be improved if I spend more time on finding the right words and improve the categorization of reviews. In the current studies, such potential biases in online reviews and sample representativeness exists due to word selection and masking. Additionally, the study's scope is delimited to a geographic region such as Norway to ensure focused and meaningful analysis. . It can also be noted that a survey is filled completely by only one person while a review that is classified in 'Attitude' category is given by one person while a review that is classified in Perceived Control is written by some other person. For that reason, there might not be a significant connection between the responses and hence the model does not cover a lot of variance of the dependent variables.

Through a comprehensive examination of online reviews from TripAdvisor and Google Form surveys, this study aims to provide valuable insights into traveler behavior and preferences affected by Attitude towards traveling, subjective norms, and perceived control, thereby contributing to the broader understanding of tourist satisfaction in the selecting Norway as travel destination. The work can further be improved to include complete destinations of Norway from the TripAdvisor and making it a more bigger data problem. In current work, I used word search for the classification of reviews into categories but more complex models can be used to understand the contextual meanings and interpretation for the classification and hence will improve TripAdvisor review model. I can also include more sources and move towards more hybrid and complex structure. I also believe that it is right time to take advantage of the data mining, artificial intelligence and machine learning methods in tourism and understanding human behavior.

Chapter 6: Reference

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes, 50*(2), 179-211.

Ajzen, I. (2020). The theory of planned behavior: Frequently asked questions. *Human behavior and emerging technologies, 2*(4), 314-324.

Bambauer-Sachse, S., & Mangold, S. (2011). Brand equity dilution through negative online word-of-mouth communication. *Journal of Retailing and Consumer Services, 18*(1), 38-45.

Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of interactive marketing, 21*(3), 2-20.

Chakraborty, U. (2019). Perceived credibility of online hotel reviews and its impact on hotel booking intentions. *International Journal of Contemporary Hospitality Management, 31*(9), 3465-3483.

Cheung, C. M., & Lee, M. K. (2012). What drives consumers to spread electronic word of mouth in online consumer-opinion platforms. *Decision support systems, 53*(1), 218-225.

Cheung, C. M., & Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision support systems, 54*(1), 461-470.

Dimitrovski, D., Senić, V., Marić, D., & Marinković, V. (2017). Commemorative events at destination memorials–a dark (heritage) tourism context. *International Journal of Heritage Studies, 23*(8), 695-708.

Doosti, S., Jalilvand, M. R., Asadi, A., Khazaei Pool, J., & Mehrani Adl, P. (2016). Analyzing the influence of electronic word of mouth on visit intention: the mediating role of tourists' attitude and city image. *International Journal of Tourism Cities, 2*(2), 137-148.

Fernandes, S., Panda, R., Venkatesh, V., Swar, B. N., & Shi, Y. (2022). Measuring the impact of online reviews on consumer purchase decisions–A scale development study. *Journal of Retailing and Consumer Services, 68*, 103066.

Filieri, R., & McLeay, F. (2014). E-WOM and accommodation: An analysis of the factors that influence travelers' adoption of information from online reviews. *Journal of travel research, 53*(1), 44-57.

Goyal, C., & Taneja, U. (2023). Electronic word of mouth for the choice of wellness tourism destination image and the moderating role of COVID-19 pandemic. *Journal of Tourism Futures*.

Hajli, N., Sims, J., Zadeh, A. H., & Richard, M.-O. (2017). A social commerce investigation of the role of trust in a social networking site on purchase intentions. *Journal of Business Research*, *71*, 133-141.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *Journal of interactive marketing*, *18*(1), 38-52.

Huete-Alcocer, N. (2017). A literature review of word of mouth and electronic word of mouth: Implications for consumer behavior. *Frontiers in psychology*, *8*, 271126.

Ismagilova, E., Slade, E. L., Rana, N. P., & Dwivedi, Y. K. (2020). The effect of electronic word of mouth communications on intention to buy: A meta-analysis. *Information Systems Frontiers*, *22*, 1203-1226.

Jalilvand, M. R., & Samiei, N. (2012). The impact of electronic word of mouth on a tourism destination choice: Testing the theory of planned behavior (TPB). *Internet research*, *22*(5), 591-612.

Jian, W.-S., Wang, J.-Y., Rahmanti, A. R., Chien, S.-C., Hsu, C.-K., Chien, C.-H., Li, Y.-C., Chen, C.-Y., Chin, Y.-P., & Huang, C.-L. (2022). Voice-based control system for smart hospital wards: a pilot study of patient acceptance. *BMC health services research*, *22*(1), 287.

Kim, J., & Fesenmaier, D. R. (2017). Sharing tourism experiences: The posttrip experience. *Journal of travel research*, *56*(1), 28-40.

Leung, X. Y., Bai, B., & Stahura, K. A. (2015). The marketing effectiveness of social media in the hotel industry: A comparison of Facebook and Twitter. *Journal of Hospitality & Tourism Research*, *39*(2), 147-169.

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism management*, *29*(3), 458-468.

Loper, E., & Bird, S. (2002). Nltk: The natural language toolkit. *arXiv preprint cs/0205028*.

Maximus. (2020). *What is a Good R-Squared Value?* reddit. Retrieved 29 April from https://www.reddit.com/r/statistics/comments/i2wnzu/d_what_is_a_good_rsquared_value/

Nilashi, M., Abumalloh, R. A., Alrizq, M., Alghamdi, A., Samad, S., Almulihi, A., Althobaiti, M. M., Ismail, M. Y., & Mohd, S. (2022). What is the impact of eWOM in social network sites on travel decision-making during the COVID-19 outbreak? A two-stage methodology. *Telematics and Informatics*, *69*, 101795.

Rauschnabel, P. A., Krey, N., Babin, B. J., & Ivens, B. S. (2016). Brand management in higher education: the university brand personality scale. *Journal of Business Research*, *69*(8), 3077-3086.

Renzi, S., & Klobas, J. (2008). *Using the Theory of Planned Behavior with Qualitative Research (Version Working Paper Number 12): Centre for Research on Social Dynamics (DONDENA), Università Commerciale Luigi Bocconi.*

[Record #42 is using a reference type undefined in this output style.]

San Martín, H., & Del Bosque, I. A. R. (2008). Exploring the cognitive–affective nature of destination image and the role of psychological factors in its formation. *Tourism management*, *29*(2), 263-277.

Sparks, B., & Pan, G. W. (2009). Chinese outbound tourists: Understanding their attitudes, constraints and use of information sources. *Tourism management*, *30*(4), 483-494.

Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism management*, *32*(6), 1310-1323.

Tham, A., Croy, G., & Mair, J. (2013). Social media in destination choice: Distinctive electronic word-of-mouth dimensions. *Journal of Travel & Tourism Marketing*, *30*(1-2), 144-155.

Verma, S., & Yadav, N. (2021). Past, present, and future of electronic word of mouth (EWOM). *Journal of interactive marketing*, *53*(1), 111-128.

Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism management*, *30*(1), 123-127.

Wang, P. (2014). Understanding the influence of electronic word-of-mouth on outbound tourists' visit intention. Digital Services and Information Intelligence: 13th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2014, Sanya, China, November 28-30, 2014. Proceedings 13,

Wang, Y., & Fesenmaier, D. R. (2004). Towards understanding members' general participation in and active contribution to an online travel community. *Tourism management*, *25*(6), 709-722.

Yang, Z., & Peterson, R. T. (2004). Customer perceived value, satisfaction, and loyalty: The role of switching costs. *Psychology & marketing*, *21*(10), 799-822.

Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human behavior*, *27*(2), 634-639.

Appendix

Questionnaire:

Questionnaire for the first and second cohort is provided below. Please note, in the second cohort the first section of electronic word of mouth questions was not tested.

Electronic word of mouth

I often read other tourists' online travel reviews to know what destinations make good impressions on others.
To make sure I choose the right destination, I often read other tourists' online travel reviews.
I often consult other tourists' online travel reviews to help choose an attractive destination.
I frequently gather information from tourists' online travel reviews before I travel to a certain destination.
If I don't read tourists' online travel reviews when I travel to a destination, I worry about my decision.
When I travel to a destination, tourists' online travel reviews make me confident in travelling to the destination.

Attitude

Unenjoyable	_.___.___.___.___.___.___._	Enjoyable
Bad __	_.___.___.___.___.___.___.___	Good
Foolish	.___.___.___.___.___.___.___	Fun
Unpleasant___	.___.___.___.___.___.___.___	Pleasant
Unfavourable___	.___.___.___.___.___.___.___	Favourable
Dislike___	.___.___.___.___.___.___.___	Like

Subjective norm

I took a holiday in Norway within these past 12 months because it is popular among my friends or family.
Important people in my life would probably think it would have been a good idea to take a holiday in Norway within past 12 months.
Friends or family had recommended that I take a holiday to Norway within past 12 months.
I had visit Norway within these past 12 months because I heard a lot about this destination from friends or family.

Perceived behavioural control

I felt I had enough time to take a holiday to Norway within past 12 months.
I felt I had enough money to take a holiday to Norway within past 12 months.
I felt there is nothing that prevents me from taking a holiday to Norway within these past 12 months if I wanted to.

Intention to travel

I intended to take a holiday to Norway within these past 12 months.
How likely were you to take a holiday to Norway within past 12 months?

Results of all participants (Country of origin):

Following are the number of reviews from each country. As these results are completely produced using Python code, please expect some duplicates such as UK, and United Kingdom may be reported as separate countries. We have fixed these while reporting in demographic table.

'UK': 1316, nan: 912, 'Australia': 587, 'Norway': 538, 'CA': 271, 'HI': 225, 'Canada': 213, 'Romania': 174, 'Germany': 169, 'FL': 160, 'NY': 150, 'India': 127, 'The Netherlands': 114, 'TX': 114, 'Italy': 107, 'SP': 97, 'Sweden': 88, 'Spain': 83, 'Finland': 77, 'MA': 73, 'Ireland': 69, 'GA': 67, 'OH': 64, 'WA': 64, 'IL': 64, 'Singapore': 63, 'SC': 62, 'VA': 60, 'CO': 55, 'NJ': 54, 'MN': 53, 'Argentina': 51, 'PA': 50, 'NC': 49, 'VT': 48, 'Belgium': 48, 'France': 47, 'Hungary': 45, 'DC': 43, 'Switzerland': 39, 'Denmark': 39, 'Poland': 39, 'AZ': 38, 'United Kingdom': 37, 'MD': 35, 'United States': 35, 'United Arab Emirates': 32, 'Greece': 32, 'South Africa': 31, 'Portugal': 31, 'New Zealand': 31, 'WI': 30, 'Malaysia': 29, 'China': 29, 'Thailand': 28, 'MI': 27, 'Czech Republic': 27, 'Israel': 26, 'OR': 23, 'TN': 21, 'Türkiye': 20, 'Mexico': 18, 'MO': 17, 'CT': 17, 'ME': 15, 'RS': 15, 'UT': 15, 'Indonesia': 15, 'RJ': 14, 'NV': 14, 'LA': 14, 'Austria': 13, 'Croatia': 12, 'Russia': 12, 'Serbia': 12, 'IA': 11, 'Qatar': 11, 'IN': 11, 'Slovakia': 11, 'Cyprus': 10, 'KY': 9, 'OK': 9, 'AL': 9, 'Luxembourg': 9, 'Iceland': 8, 'Japan': 8, 'NH': 8, 'MT': 8, 'Estonia': 8, 'Lithuania': 7, 'RI': 7, 'Slovenia': 7, 'SD': 7, 'AK': 7, 'Latvia': 7, 'Sri Lanka': 7, 'NM': 7, 'NE': 7, 'USA': 7, 'Philippines': 7, 'Virginia': 6, 'Bulgaria': 6, 'Egypt': 6, 'AR': 6, 'Panama': 5, 'England': 5, 'Vietnam': 5, 'Lebanon': 5, 'PR': 5, 'Texas': 4, 'North Carolina': 4, 'Saudi Arabia': 4, 'KS': 4, 'Pakistan': 4, 'Trinidad': 4, 'California': 4, 'MS': 4, 'Dominican Republic': 4, 'Ontario': 3, 'Georgia': 3, 'Ecuador': 3, 'Guatemala': 3, 'Costa Rica': 3, 'ID': 3, 'Bahrain': 3, 'DF': 3, 'Malta': 3, 'Missouri': 2, 'Bosnia and Herzegovina': 2, 'Ukraine': 2, 'DE': 2, 'Deutschland': 2, 'Nigeria': 2, 'Florida': 2, 'Nepal': 2, 'Kazakhstan': 2, 'Ohio': 2, 'ny': 2, 'nh': 2, 'Dc': 2, 'New York': 2, 'Myanmar': 2, 'Taiwan': 2, 'Pennsylvania': 2, 'holland': 2, 'Brazil': 2, 'Fl': 2, 'ND': 2, 'Ca': 2, 'Grand Cayman': 2, 'WV': 2, 'Republic of North Macedonia': 2, 'Kenya': 2, 'wis': 2, 'Cambodia': 2, 'Netherlands': 1, 'Kent': 1, 'WY': 1, 'BA': 1, 'Tanzania': 1, 'Morocco': 1, 'District Of Columbia': 1, 'Belize': 1, 'Usa': 1, 'Maharashtra': 1, 'Mozambique': 1, 'Botswana': 1, 'Maryland': 1, 'Oh': 1, 'España': 1, 'Suriname': 1, 'Laos': 1, 'North Korea': 1, 'Uruguay': 1, 'PB': 1, 'uk': 1, 'Bangladesh': 1, 'Hawaii': 1, 'Minnesota': 1, 'c': 1, 'Mali': 1, 'Albania': 1, 'Rhode Island': 1, 'Arizona': 1, 'N.S.W.': 1, 'Tennessee': 1, 'Jordan': 1, 'Central Mexico and Gulf Coast': 1, 'Armenia': 1, 'PI': 1, 'Korea': 1, 'Uganda': 1, 'West Yorks': 1, 'Fiji': 1, 'Colorado': 1, 'Tunisia': 1, 'New Hampshire': 1, 'Connecticut': 1, 'Maine': 1, 'South Korea': 1, 'il': 1, 'va': 1, 'Kuwait': 1, 'Oman': 1, 'ES': 1

Corresponding Author:
Sahiba Inayatullah can be contacted at: sahiba.inayat@hotmail.com