Using Social Media Data Sources to Effectively Market Tourism in Sri Lanka.

Ashan Bogollagama

Harrisburg University of Science and Technology

Abstract

Being in sync with technological advancements is critical for all industries and businesses to succeed. A well-established business can go under very quickly if it is not aware of how technology is influencing the way in which its customers are interacting. Sri Lanka has an economy that is heavily dependent on its tourism industry. However, it is very traditional on its stance in promoting and capturing insights on its travelers. The purpose of this paper is to comprehend the effect social media has on the travel industry and to understand how relevant it is in the decision making process amongst many travelers.

Keywords:

Social Media, Databases, Data Mining, Natural Language Processing, Text Analytics, User Generated Content

Table of Contents

Relationship to CPT

"NO CPT."

Introduction

Tourism is a major commodity in the Sri Lankan industry similar to many south Asian countries such as Thailand, Malaysia, Indonesia, Maldives and Taiwan. Tourism can greatly effect a country's economic status as it promotes new businesses and brings in outside money. For this reason, it is important to be competitive against other countries in this market.

According to Rocket Fuel's white paper (Jeremy Kressmann, 2015) digital ad spending by the US travel industry alone spent over $4.85 billion in 2015. Dave Rumsey (CIO, Tourism Australia) and Estela Marine-Roig (Author, Tourism Analytics with Massive User Generated Content) both agree that cities in Australia and Spain have heavily invested in data analytics platforms to better understand its tourism market and to continue its stance as a smart destination. Switzerland hosted an international conference in Lugano from February 3-6 of 2015 to address information communication technologies in tourism.

With the amount of global research and financing going into the tourism industry, new ways of gathering, analyzing and reporting tourism information is exhausted. For this reason, it is vital for country's whose economy heavily depends on the tourism industry to be up to date with information communication technologies and analytical methodologies to compete on a global scale.

Sri Lanka, formerly known as Ceylon while under British rule, was once a very popular and prosperous tourist destination. After enduring a civil war that started in 1983 and lasted over 20 years (Fatma Ahmad, 2013), the tourism industry was greatly affected. The 5 star hotels that were

once built for outsiders were mainly utilized by locals. However, after the war ended, Sri Lanka tourism has tripled from 2008 with 438,475 tourists to 1,527,153 in 2014 (Sri Lanka Tourist Development Authority – Annual Statistical report 2014, Appendix A). This new wave of tourists visiting the country means new practical methods of data collection and analysis needs to be implemented to efficiently and precisely understand the tourists and what they are seeking on their visits.

The Sri Lankan Tourist Board currently uses immigration data to record the tourists coming into the country. All this information can be found online in the Sri Lanka Tourism Development Authority website. Monthly statistical reports are published stating number of tourists visiting the country and where they originate from. The type of analysis done is mainly descriptive and is not progressive in determining new markets, adhering to current issues and determining future trends.

The purpose of this paper is to comprehend the importance of social media in tourism and the advantage of analyzing real-time user generated content from social media sites to depict patterns and trends in user behavior in Sri Lanka's tourism.

Problem Statement and Justification

Is Sri Lanka investing enough on developing its information and communication infrastructure to effectively analyze and promote tourism? Is the tourist board operating with enough resources to utilize the vast amount of influential information from third party sources (social media)?

As mentioned in the introduction, Sri Lanka's tourism industry plays a vital role in the country's economy. The contribution of travel & tourism to gross domestic product was 7.5% (US$3,3881 million) and the contribution to employment was 503,000 jobs in 2009. This is equivalent to 6.2% of total employment or 1 in every 16 jobs (Ruwan Ranasinghe, Ravindra Deyshappriya, 2010). These numbers are further illustrated in Figures 1 and 2 (World travel and tourism Council Economic impact report – 2009) with comparison to other neighboring countries.

As depicted in Figure 1 and 2, tourism is not only a major commodity for Sri Lanka but for a lot of its neighboring countries as well; some engaging in a tourist market that determines over 50% of the national GDP. With stakes this high within neighboring countries, the demand to attract tourism is undoubtedly increasing and the only way to succeed is by being one step ahead of the competition.

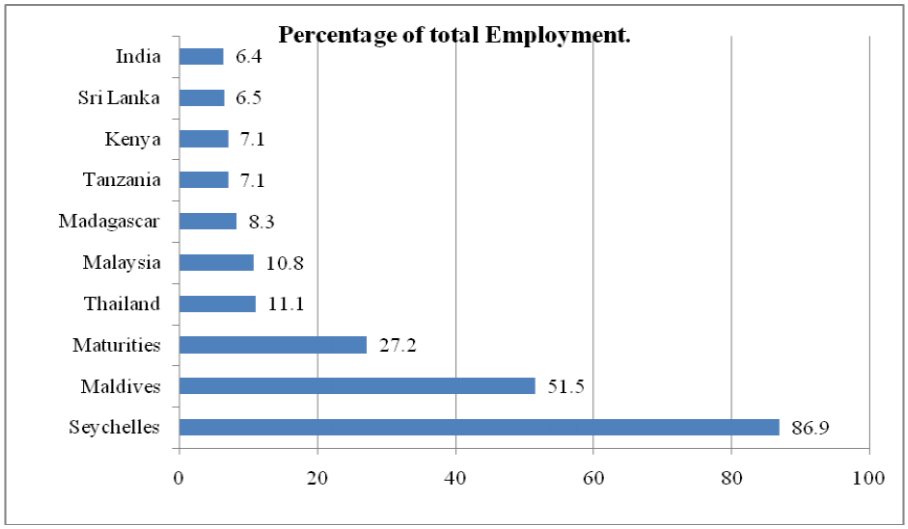


Figure 1- Percentage of Total Employment in the Tourism Industry; World travel and tourism Council Economic impact report – 2009

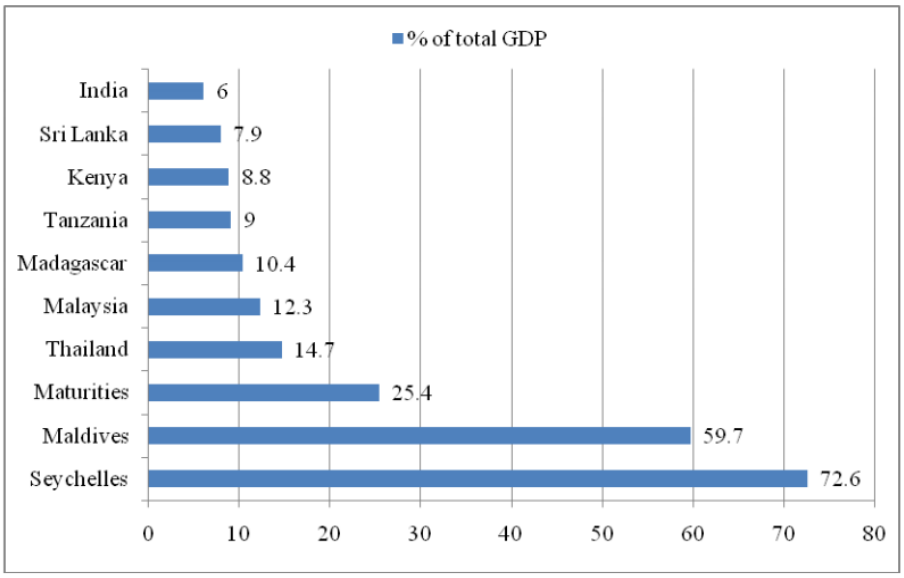


Figure 2 - Percentage of Total GDP in the Tourism Industry; World travel and tourism Council Economic impact report – 2009

According to Sri Lanka's Tourism Development Authority, the country set up a Tourism Bureau as far back as 1937, decades before tourism was recognized as a global industry. Under the British Empire the main objective of the bureau was to accommodate passengers sailing from the West to

the East through the port of Colombo by providing facilities and services. Sri Lanka's primary geographic location attracted many cruise ships, and passengers on board enjoyed sightseeing within the country during their short visits. Apart from the duration of the world wars, tourism within the country rapidly increased up until the 1950's, when aircrafts became a prominent medium for civilian transport. Sri Lanka failing to understand the importance of updating international airports to accommodate the demand for air travel lost its statue as the main gateway between the West and the East. Sri Lanka who hadn't updated its airports until 1965 lost many travelers to Singapore, who were quick to assess and implement the potential of developing modern airports (Sri Lanka Tourist Development Authority).

The purpose of the timeline of events above, is that Sri Lanka already has a background that depicts the consequences of not being up to par with global technology changes. An industry like tourism is not determined internally but by what is trending on a global level. For these reasons alone, the competition of neighboring countries for tourism (some having higher stakes), and the lessons learned in history, it is vital for Sri Lanka to update its information and communication technologies to effectively and efficiently utilize its resources to get mass insight into the travel industry.

Dissecting information created by travelers into a more granular level, an area of study that is vital to understanding consumer behavior, or in this case travelers' behavior is the study of user generated content (UGC) from social media sites. The problem faced by the government of Sri Lanka is adapting to technological changes that will aid in gaining insight into travelers which is vital for effective marketing strategies.

For this reason, the two main objectives of this paper it to analyze the involvement of social media in tourism and methods for analysis of user generated content from social media sites. The second study is more of an analytical study and methodology as to how specific analytical tools such as data mining can be implemented into capturing valuable insights into tourist behavior from large datasets. The first study would be a questionnaire based analysis and would have to pertain certain deliverables; measurable levels of relevancy to accurately understand how relevant social media truly is in peoples travel choices.

Literature Review

Traditionally many travelers used to consult with travel agents on options for planning out trips abroad or to unfamiliar regions. However, this process is made a lot more efficient with easily accessible online user interfaces provided by hotels and airlines, making it easier for their clients to access and book simple transactions. With increased interaction online came about online travel reviews (OTR) and travel blogs, enabling the traveler to engage in first hand travel experiences through narration, photos, videos and other communication methods from previous travelers. Social media involvement is defined as a person's level of interest, emotional attachment or arousal with social media (Suzanne Amaro, Paulo Duarte, 2015).

In recent years, we have witnessed tremendous growth of search engines, new distribution channels, virtual tourism communities and numerous social media platforms that are essential for travelers to make smarter decisions (Estela Marine-Riog, 2015). For this reason, information communication technology (ICT) has never been so tightly knit into the fabric of the travel experience and the management of the tourism product than today (Buhalis, 2014). PhocusWright, a well-known research firm in the travel industry, concluded that 80% of travelers read at least six reviews before booking an accommodation (Suzanne Amaro, Paulo Duarte, 2015). The same study concluded that these reviews usually trump pricing. Therefore, it is important to understand a traveler's involvement with social media. A Forrester Research study concluded that 75% of internet users use social media, but less than half actively participate (Osborn, 2009).

Is this a common trend amongst travelers? Understanding this question is vital as travelers with higher levels of involvement represent themselves as key marketing outlets. Another Forrester

study confirms this and have termed these types of travelers as conversationalists; those who are actively engaged in social media conversations and have the ability to persuade decisions of others and drive sales (Harteveldt et al. 2010)

Information acquisition is the main reason travelers participate in online travel communities (Chung and Buhalis, 2008). As the relationship between ICT and tourism grow stronger online, more user-generated content (UGC) is produced from OTRs to travel blogs. UGC offers huge possibilities for e-commerce, business intelligence, marketing and social studies. UGC data in tourism is considered a good source of information for National Tourism Organizations (NTOs), Destination Marketing Organizations (DMOs), stakeholders and future travelers because it consists of freely expressed opinions (Estela Marine-Riog and Salvador Anton Cave, 2015)

UGC data has been around since the internet and has been growing exponentially. Initially with the small amount of UGC generated, this data was collected by hand. As the amount of data started growing a research study shows that this method of data collection is very time consuming and limited the sample size (Lu and Stepchenkova, 2014**)**, which can affect the results if it doesn't accurately represent the population.

According to Statista, an online statistics portal that collects information from over 18,000 sources, the travel industry has doubled within a span of 10 years. For the first time in 2012 the number of international tourist arrivals exceeded over 1 billion. This stat alone can give you an idea to the amount of UGC generated online. Social media can be classified into blogs, review sites, media sharing, question and answer sites, social book marking, social-networking and social news. It is

claimed at present the amount of data generated from these sites have increased from terabytes to petabytes (Ganomi & Haider, 2015). An illustration of different media sites and its contents is represented in Figure 3.

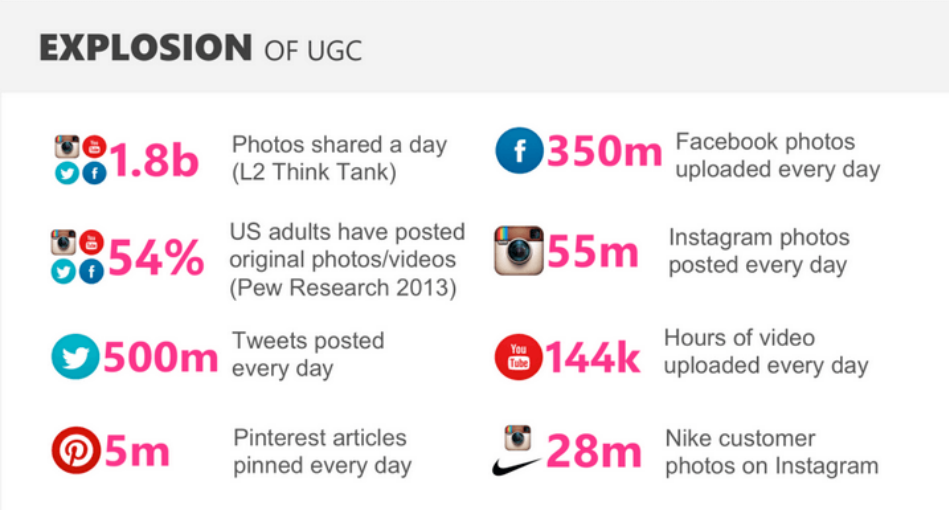


Figure 3 - Six Ways to Inspire Influential Content from Your Facebook Fans, Dan Virgillito

To put this into perspective one terabyte has enough space to store 16 million Facebook photographs. One petabyte equals 1024 terabytes. It is estimated that Facebook in 2010 stored over 260 billion photos, using a storage space of over 20 petabytes (Beaver, Kumar, Li, Sobel and Vaigel, 2010). The data stored and analyzed by big companies is already moving from petabyte to exabyte (Hu, Wen, Chua, and Li, 2014). The graph in Figure 4 illustrates the UGC increase from 2005 to 2008 and Figure 5 exemplifies the exponential growth in data.

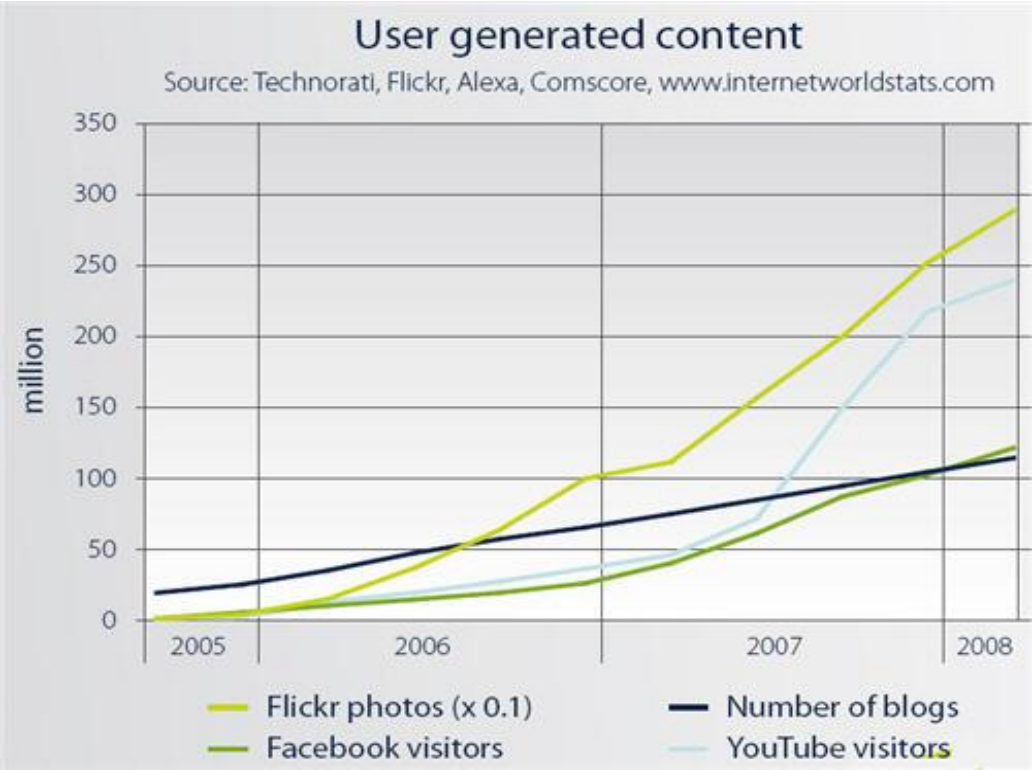


Figure 4 - User Generated Data, Technorati, Lickr, Alexa, Comscore

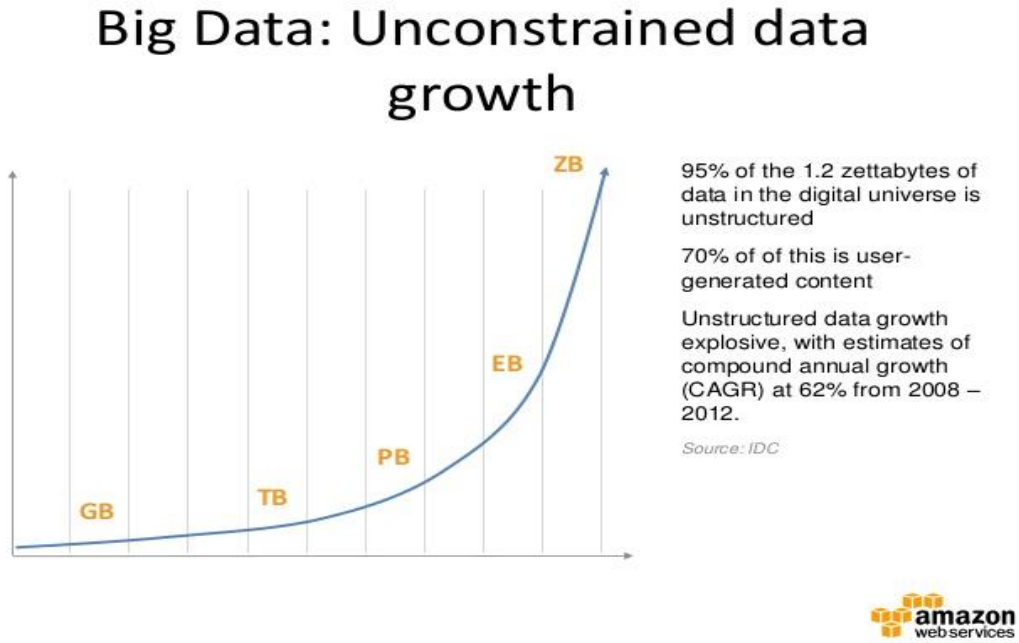


Figure 5 - Solving Big Data Problems, Rajnish Malik

Analyzing the two graphs it can be established that there is a high velocity, variety and volume of data to filter through in social media sites alone. With 95% being unstructured data the only alternative for analysis is using big data technologies. Traditional data-management systems and analysis is proving insufficient in handling the amount of data coming in at its current pace where as big data technologies are creating real-time intelligence.

Proposed Solution Approach

Study Design:

The main objectives we will seek during the course of this research is to gain insight into the development of a methodology in analyzing UGC from social media datasets to recognize behavioral patterns within tourists; Along with a proposed measure to analyze the importance of social media involvement within travelers. Therefore, I will be simultaneously conducting two different studies.

The first study, involving the analysis of UGC from data sources would require using big data and analytical tools. The three basic steps in web extraction and analysis are (Abburu and Babu, 2013***)***:

1. Finding URL's of webpages (web crawling).
2. Extracting information from webpages (information extraction)
3. Data analysis/reporting.

The information extraction phase can be further sub-divided into cleaning and debugging phases to get rid of noise (Estela Marine-Riog and Salvador Anton Cave, 2015).

Study 1 Methodology:

For analyzing UGC, I will implement a similar study to Marine-Roig and Clave's approach to analyzing UGC data for Catalonia. Certain attributes would have to be altered however considering Catalonia and Sri Lanka provide different experiences with regards to tourism. This study will be broken down into six main tasks.

Task 1 (Website Selection)

The first step in this analysis is the selection of websites. As discussed earlier, there are a vast number of websites pertaining UGC such as travel reviews, photos, blogs etc... A study would have to be conducted determining contents from which sites would best help us analysis tourism behavior in Sri Lanka.

Task 2 (Data Download)

Next it is important to make sure the URLs selected contains the data required and is not hyperlinked to another website. Once this is confirmed the data would have to be downloaded using the most efficient software and filters for minimizing download space and time.

Task 3 (Data Cleaning)

The URL data downloaded are usually in HTML format and would have to be cleaned to get rid of unnecessary noise such as code on where to place figures and contents on a webpage.

Task 4 (Debugging)

The data has to be checked for grammatical errors, misspelling and other common errors.

Task 5 (Content Analysis)

Once the following procedures are completed, a content analysis can be done to determine the underlying message of the UGC. Considering we are dealing with 1000's of reviews, reading through them all would be impractical. This is where big data tools come convenient. Text mining, natural language processing (NLP), machine learning and other tools will be used to effectively process the data and return real time feedback. With the aid of a programming platform such as python, text mining and NLP can be utilized to identify common key words used in reviews of Sri Lanka. The words obtained could be a mix of emotions, names of places, food, people, adjectives etc.… This will aid in detecting trends within travelers, popular places to visit during a certain time of the year, general emotions towards their time spent in Sri Lanka and popular people associated with Sri Lanka. These are all key marketing insights if obtained immediately. Sentiment analysis can be used to identify between positive and negative reviews. Negative reviews can be useful in understanding areas of improvement similar to constructive criticism. Machine learning techniques can be used to optimize this process for future reference.

Task 6 (Conclusion)

The findings of this study will be generated in the form of a report or visual representation using business intelligence tools such as Tableau. It will give valuable insights into tourism within Sri Lanka and will suggest new marketing outlets with up to date information communication systems.

The second study conducting the relevance of social media within travelers will be less complex. This study involves active participation from third party sources. For this reason, time management is critical as adequate time has to be allocated for outside participation.

When it comes to understanding social media involvement, it is understood that individuals interact with social media in three different ways (Shao, 2009):

1. Consuming – Individuals who only read or view but never participate.
2. Participating – User to user interaction and user to content interaction.
3. Producing – Individuals only create content.

Consuming and participating are the most common in social media. People engage in user generated media for similar entertainment reason as traditional media such as relaxation, time fulfillment, emotional release and escaping from problems (Shao, 2009). For this reason, enjoyment levels are also a determining factor in recognizing involvement levels with social media. To factor in all these attributes, I will be conducting a questionnaire based analysis, reviewing four dimensions proposed by Amaro and Duarte for measuring a traveler's social media involvement:

1. Social Media Consumption
2. Creation of Social Media Content
3. Perceived Playfulness
4. Level of Interest

Task A (Questionnaire Design)

The key is to developing an engaging questionnaire that can be easily filled out to obtain insights into the involvement of social media. Designing an eye-catching, effective, quick questionnaire is vital to getting as many participants as possible. Considering this analysis is to determine online user interaction with social media, an online questionnaire can be used as all participants that we are trying to study would have access to the internet.

Task B (Questionnaire Distribution Medium)

The link to the questionnaire will be distributed by various means such as Facebook, LinkedIn, email, and by word of mouth. The goal is get as many people as possible to fill out the survey.

Task C (Time Allocation)

Considering this is a time sensitive research, a time frame will be allocated in which the study participants can fill out the questionnaire. Plenty of time should be set aside to retrieve as much feedback as possible.

Task D (Information Retrieval)

Once the deadline is met, the questionnaires will be retrieved along with the participants' information. Participants information cross referenced with their questionnaires can help gain further insights outside of our scope of study.

Task E (Information Analysis)

An analysis such as confirmatory factor analysis (CFA) would be conducted to determine the validity of the questionnaires retrieved. Statistical tools will be used to evaluate and compare variables that determine the involvement of social media within travelers.

Task F (Output Medium)

Dashboards and other visualizations would be used to display findings of this study.

Task G (Conclusion)

The conclusion of this study will give us quantitative ratings on the four dimensions discussed earlier to determine the involvement of social media within travelers and more specifically travelers visiting Sri Lanka.

Analysis of Results and Discussion:

Study 1 – Analyzing UGC

Website Selection:

Having lived and traveled to Sri Lanka numerous times, and interviewing representatives of local businesses such as Cinnamon Hotels, Mahaweli Reach Hotel, Jetwing Hotels, Cantoulope, Sunset Hotel and Mount Lavinia Hotel, it was determined that Trip Advisor is the most sought out international travel site to get insights into booking accommodation and transport, and seeking insights into restaurants along with things to do within Sri Lanka.

According to Forbes, Trip Advisor is claimed to be one of the largest travel site's in the world, with more than 60 million members and over 170 million reviews and opinions of hotels, restaurants, attractions and other travel-related businesses. Considering the sites reputation and credibility on UGC, Trip Advisor will be used as the main website to obtain the UGC needed for this particular research.

Data Download, Cleaning, Debugging:

Once the website was chosen, the initial plan was to use a language such as Python to connect to the Trip Advisor API and download the necessary information such as reviews. However, after thorough investigation this process was costlier and more time consuming than anticipated. Upon further research it was discovered that many existing website scrappers were able to download all the reviews by location, hotel, restaurants and things to do. Due to time constraints and efficiency this was the sought out method to extracting the data. A Trip Advisor scrapper developed by

Botguruz was used to scrape all the reviews by location and were exported onto Microsoft Excel for further analysis. Any review that had the key word “Sri Lanka” in it from August 2006 till present (June 2016) was downloaded. A total of 20,539 reviews were obtained. Each review or record consisted of a date, hotel url, rating, review and a username.

Content Analysis/Discussion:

Three different analytical studies were conducted on Excel using the data gathered.

The first analysis was a comparison study of the growth and patterns of reviews for each year. The best way to visually represent a comparison of the measurable attribute is through means of a graph. The data first had to be cleaned and organized before being able to represent it. By using excels many functions such as “COUNTIF”, the reviews were separated by month and year as depicted in Table 1 below.

Table 1 : Number of Reviews by Month

	JAN	**FEB**	**MAR**	**APR**	**MAY**	**JUN**	**JULY**	**AUG**	**SEP**	**OCT**	**NOV**	**DEC**
2007	1	0	0	0	0	0	0	0	0	0	0	1
2008	0	0	0	0	0	1	0	0	0	5	0	1
2009	0	1	1	4	0	0	2	3	3	0	2	3
2010	5	2	9	7	4	0	5	3	1	7	5	6
2011	21	19	10	6	1	8	10	15	10	9	12	
2012	41	48	51	57	32	29	40	66	47	52	67	61
2013	132	148	135	129	116	77	111	146	140	126	122	147
2014	305	289	257	267	229	134	228	313	281	230	242	276
2015	498	478	499	521	383	308	472	654	698	524	569	920
2016	1462	1324	1690	1747	1284	508						

These values were then plotted using a line graph by year for comparison purposes as seen in Figure 6. As portrayed there is significant growth in reviews for each year, however a comparable cyclic pattern is sustained across the different years. The number of reviews tend to dip in June (6th month), rise from August till September (Summer) and peak once more in December (Christmas). Regardless of the number of reviews, this pattern remains consistent with each year. Assuming these reviews were written during the travelers stay in Sri Lanka, the graph below can help understand peak travel seasons within the country. It will be interesting to compare this trend with immigration data on the number of tourists visiting the country and see if there is any correlation. This relation will help determine a third party data source in depicting tourism trends within Sri Lanka.

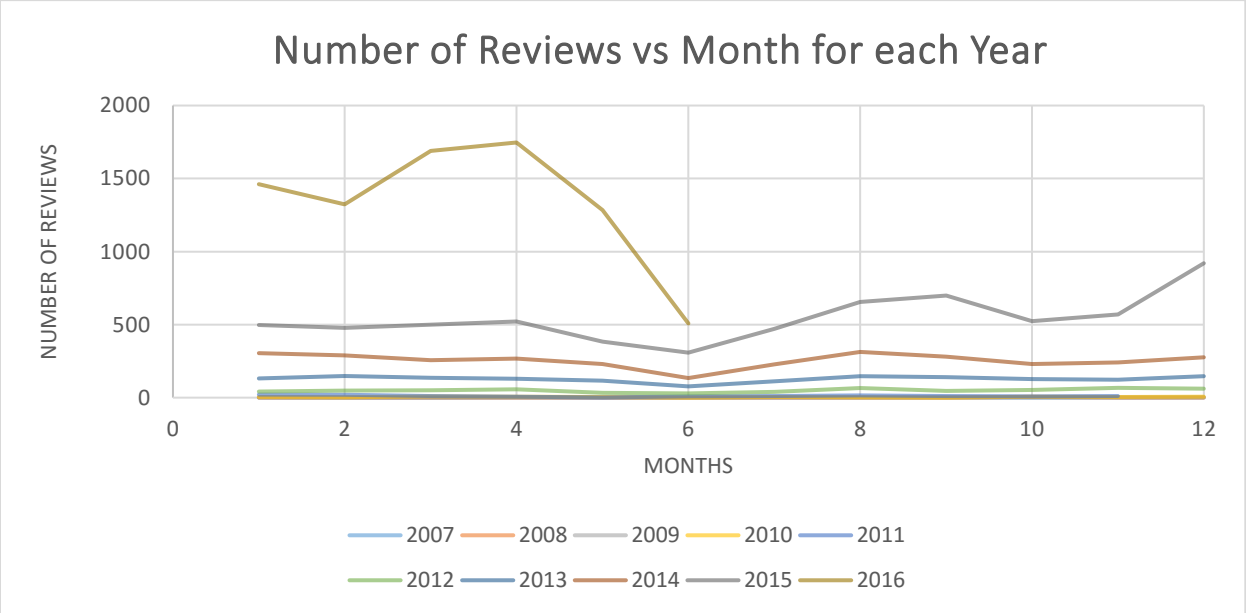


Figure 6 - Number of Reviews Vs Month

The second analysis determines the overall growth in social media involvement by analyzing the number of reviews that contained the word “Sri Lanka” over the years. Figure 7 illustrates this behavior by showing an exponential growth between the years 2006 and 2016. Trip Advisor was founded in 2000 and has taken time to catch on amongst travelers, this will most likely explain the low number of reviews between 2006 and 2011. A trend analysis was conducted to understand the pattern in growth and determine the number of reviews in upcoming years. A polynomial trend line with the equation $y = 0.0036x3 - 0.4322x2 + 14.537x - 103.03$ closely resembled the pattern obtained having a R^2 value of 0.8154.

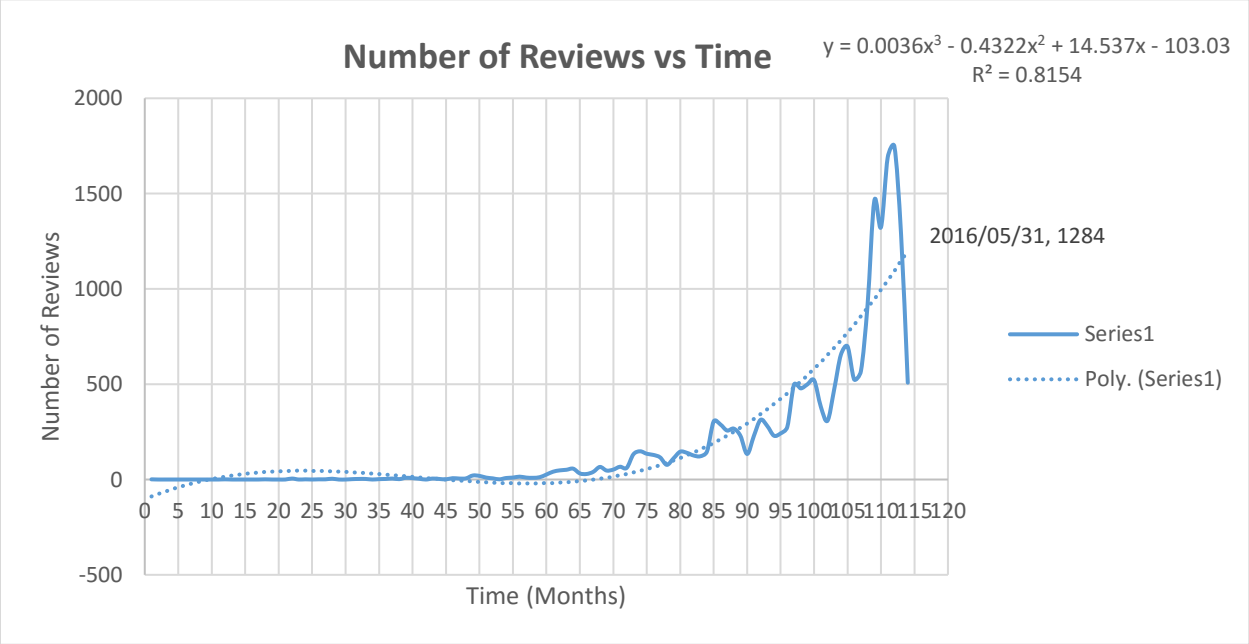


Figure 7: Number of Reviews vs Time

It is hard to validate the trend obtained in Figure 7 due to minimal results obtained in the first couple years. The graph shows a constant line without any variation till approximately January of 2012, as there weren’t enough reviews obtained. For this reason, as seen in Figure 8, a similar analysis was conducted starting with the values from the beginning of January 2012 till present.

With this data an exponential trend line with the equation $y = 34.24e0.0658x$ closely resembles the growth of reviews. At 60 months or January of 2017, It could be predicted that the number of reviews to be close to 1775 according to the equation line obtained (y = $34.24e0.0658 * 60$)

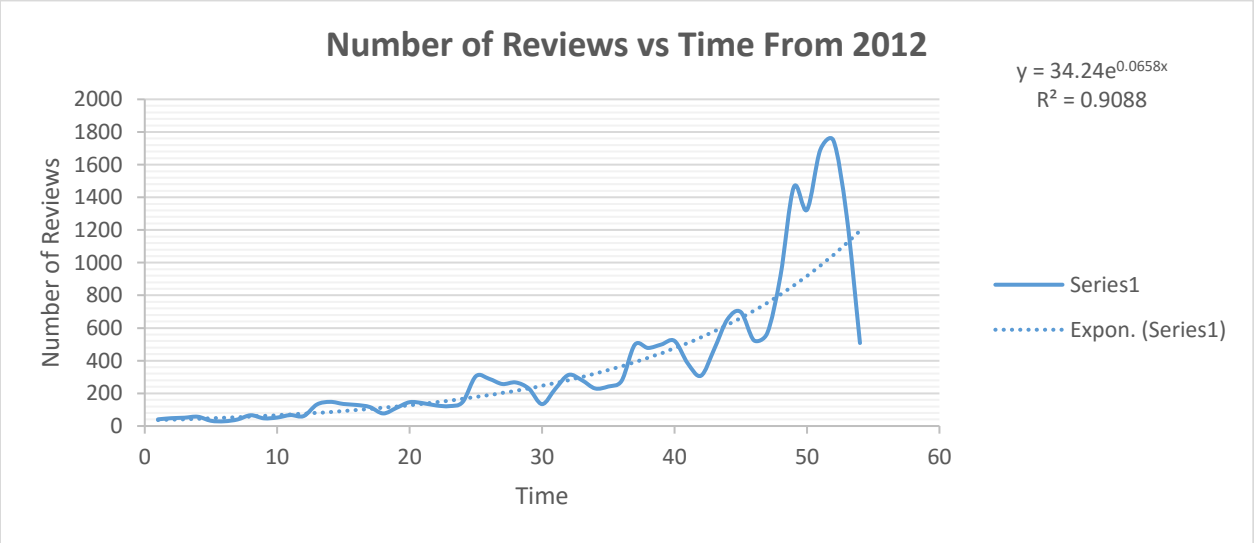


Figure 8: Number of Reviews vs Time (From 2012)

Figure 8 gives a generalized estimation and visual to how fast social media interaction is growing amongst people traveling to Sri Lanka throughout the years. It is important to understand that until further analysis is undertaken, these results do not reflect the number of tourists entering the country. However, if a correlation is found with immigration data on the number of tourists visiting the country, the number of reviews can be used as a key performance indicator (KPI) to predict future trends in travel.

The third analysis conducted was to study the change in customer ratings over the years. Using excel, a descriptive statistics analysis was done using a 95% confidence level on all reviews

starting from 2012. As seen in the Table 2 below, the mean, standard error, median, mode and other variables for the ratings were determined.

Table 2 - Decriptive Analysis

Year	Mean	Standard Error	Median	Mode	Standard Deviation	Sample Variance	Sum	Count
2012	4.097719	0.043314171	4	5	1.073282845	1.151936065	2516	614
2013	4.27636595	0.02572727	5	5	1.0206952	1.0418187	6731	1574
2014	4.31222009	0.017857825	5	5	0.998442498	0.996887422	13480	3126
2015	4.33402675	0.01187226	5	5	0.97395934	0.9485968	29168	6730
2016	4.340685977	0.010147476	5	5	0.92174615	0.849615966	35815	8251

An ANOVA test with a 95% confidence level was run on the following results to determine if there was any significant correlation between the means. The test resulted in a high P value (greater than 0.05) concluding; easily noticed by glancing at the table, there is no significant difference in the customer ratings as the years changed. It might be worth while taking a look at the ratings on a more granular level, like city.

Business Intelligence tools are able to give a great deal of insights into data at first glance. For this reason, some of the data obtained was uploaded onto Tableau to obtain some visuals on the popularity of reviews by city. However, the location field obtaining the city name was populated with other text and characters that would influence the analysis. For example, the location field pertaining a review for a hotel in Colombo, used the following attribute for its location "*http://www.tripadvisor.com/Restaurant_Review-g187870-d942395-Reviews-Al_Colombo-Venice_Veneto.*html*"*. To obtain the word Colombo on its own, Tableau has a useful function

known as splitting, where the data can be split before and after given characters. In the example above, the characters that separated the word Colombo were the "_" before Colombo and "- "after. It was the same format that followed all the other cities, and therefore the same function was copied across the field.

Figure 9 below illustrates the most popular cities mentioned with the aid of a bubble map. As you will notice immediately, Kandy seems to have the biggest bubble as it is mentioned the most in the reviews analyzed. A more detailed analysis can be sought out using the bar chart in Figure 10.

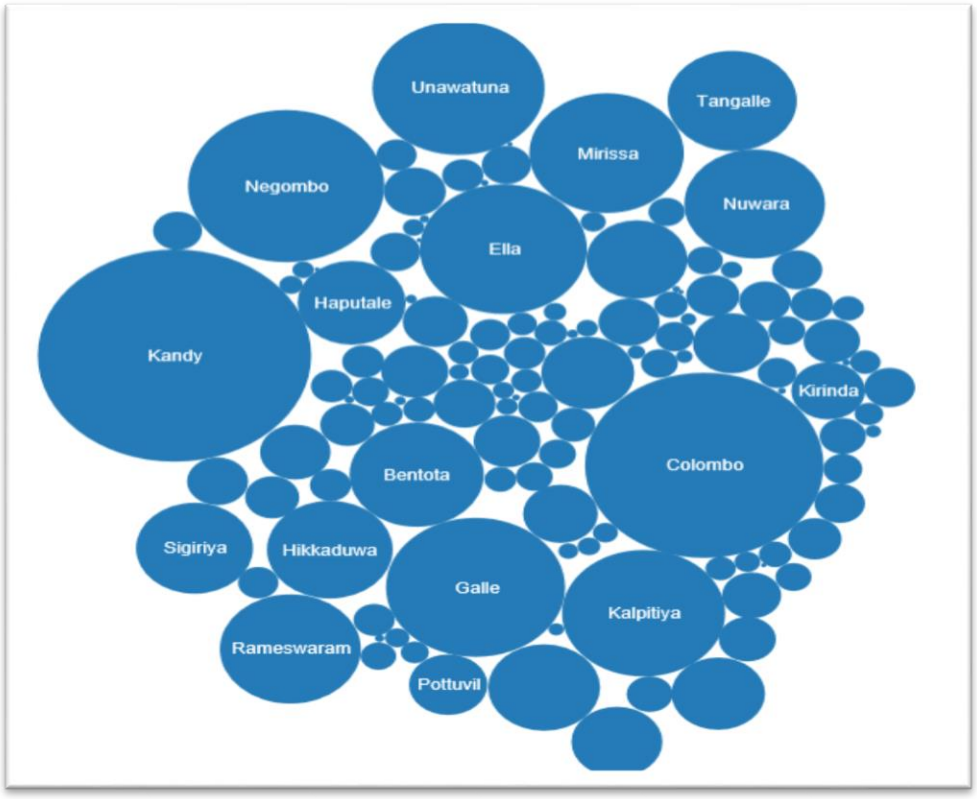


Figure 9: Bubble Map of Most Popular Destinations in Sri Lanka

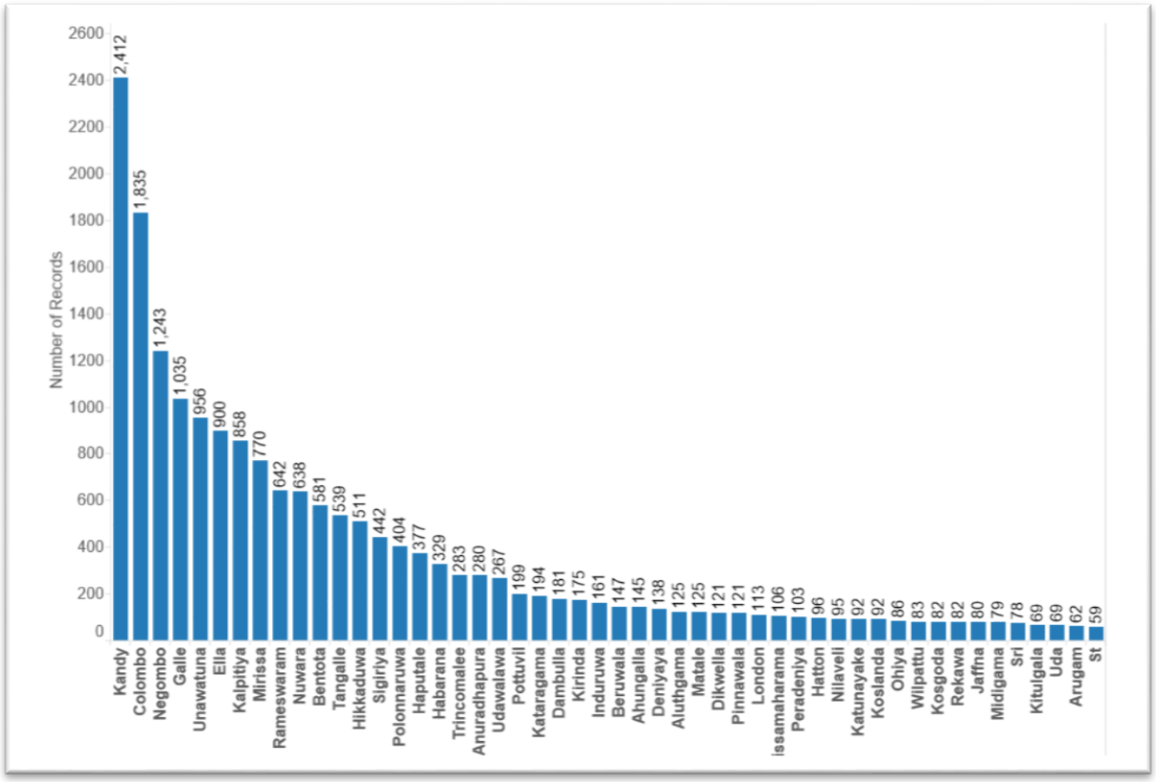


Figure 10 - Bar Chart of Most Popular Destinations in Sri Lanka

Tableau is well known for its interactive style of reporting. It is hard to see this on a picture, however if access to Tableau is obtained, hovering over the visuals will enable further insight into the data, along with capabilities of filtering to a granular level.

To get better insight into how the reviews are spread throughout the country, an analysis was undertaken to place the number of reviews on a map of Sri Lanka. This task right away encountered some obstacles as the Tableau database is not familiar with most of the cities in Sri Lanka. For this reason, an online software known as Batch Geocoding was used, which broke down all the cities entered into longitudinal and latitudinal coordinates. These coordinates enabled the plotting of the cities obtained on a map in Tableau.

Figure 11 below illustrates a geo spatial illustration of the number of reviews by city. The bigger the circle and darker in color, the more populated with reviews. As you will notice right away in the figure, Kandy is a popular destination where a lot of reviews come from. This visual interesting considering we are able to place the city with respect to the country. Sri Lanka, for many years has been well known as a beach/costal destination, however one the most populated cities by tourists is inland. Further analysis into this will help a country understand what tourists are seeking on their travels.

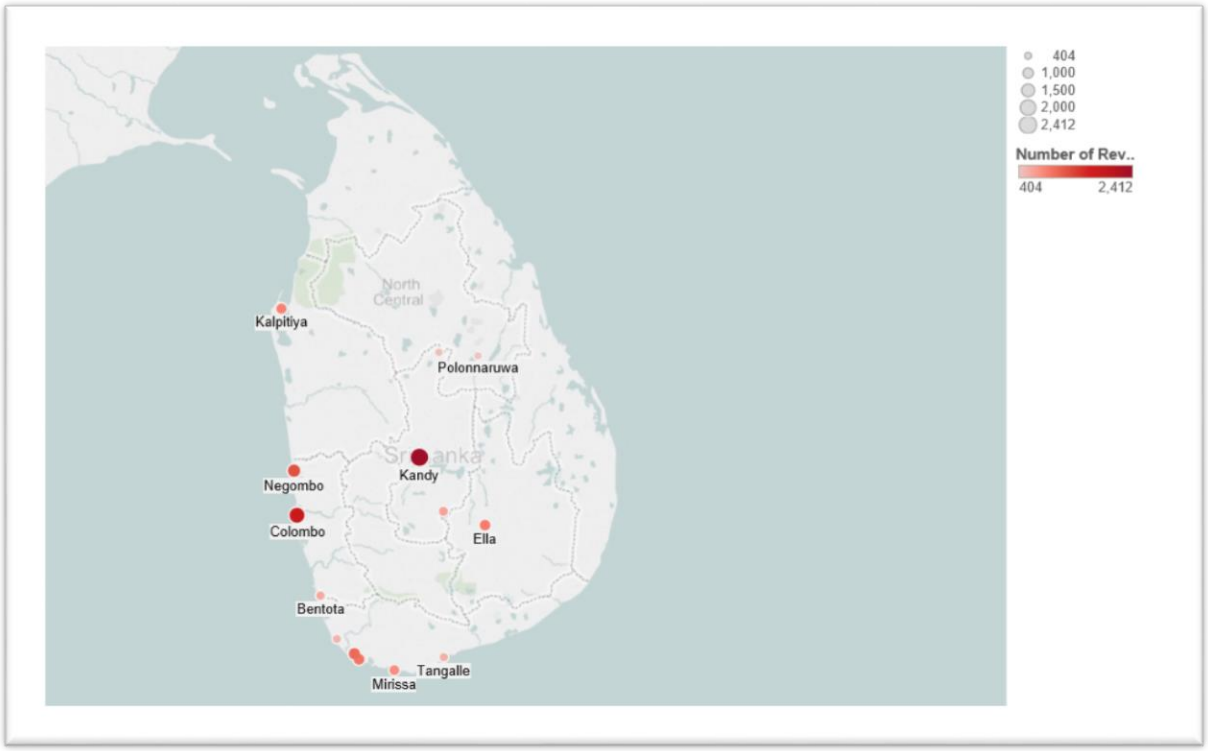


Figure 11 - Geographic Illustration of Number of Reviews

Study II – Social Media Involvement

Questionnaire Design:

The second study was to understand social media interest amongst travelers by engaging in a proposed measure conducted by Amaro and Duarte. With the aid of Google Forms, an anonymous questionnaire was designed to determine four different scenarios with regards to social media involvement:

- Social Media Consumption (Extent to which individuals use social media for travel related information.)
- Creation of Social Media Content (Participation involvement levels)
- Perceived playfulness (Extent to which social media travel websites are perceived as entertaining.)
- Level of Interest (Overall interest in related social media)

The first couple questions in the questionnaire were personal questions to determine the travelers background and ideally hoping to pull some correlations with respect to the answers on the survey. The questions included:

- Gender
- Nationality
- Age
- Highest Education Level
- Annual Income Bracket
- Visited Sri Lanka

To understand social media consumption a total of 5 questions were generated using a five-point Likert scale with responses from 1 – "Never" to 5 – "Always". Questions included were:

Before Travel

- I read hotel reviews from other travelers
- I search for travel information on social media websites
- I look at activity/attraction reviews of other travelers
- I read other travelers experience and tips

While Traveling

- I search for travel information on social media websites (For example where to eat or things to do)

Similar questions using a Likert scale with responses from 1 – 5 were generated to apprehend the creation of social media content within travelers. Questions included:

While Traveling

- I check in to the location I am at/update my location on social media (Facebook, Twitter, Foursquare)

After Traveling

- I write hotel reviews on social media
- I post photos on social media websites
- I write reviews of activities/attractions on social media websites
- I put videos on social media websites
- I write reviews of the place and/or monuments I visited on social media websites

To evaluate perceived playfulness or understand how social media can be used as a form of entertainment, a total of 5 questions were produced using a five-point Likert scale with responses from 1 – "Strongly Disagree" to 5 – "Strongly Agree". Questions comprised:

- Using social media for travel purposes is enjoyable
- Using social media websites for travel purpose is fun
- Using social media websites for travel purposes stimulates my curiosity
- I consider the use of social media for travel purpose a big hassle
- I don't have time to interact with social media for travel purposes

Overall interest in social media was measured using a five point differential scale. The scales ranged from:

- Unimportant – Important
- Unexciting – Exciting
- Doesn't matter to me – Matters to me
- Boring – Interesting
- Useful – Useless

And finally the questionnaire determined the top three choices of social media outlets for travel purposes.

A total of 28 questions were formulated into the survey, Considering the simplicity of the questions and the use of multiple choice on most, it was approximated that each question should take no

longer than five seconds to answer. This theory was tested on a sample group of 10 participants, the results are displayed in Table 3 below.

Table 3 - Questionnaire Time Results

Participant	Time in (Mins)
1	1:50
2	1:55
3	2:10
4	2:05
5	2:15
6	1:45
7	2:05
8	2:12
9	1:57
10	1:58

The mean time to take the survey was concluded to be 1.85 seconds, 2 minutes when rounded up. (The complete survey can be found in Appendix B).

Questionnaire Distribution Medium:

A couple different options are presented when trying to share a survey through Google forms. The two most common are via email or a URL link. For this research a shortened URL link was copied and distributed through different means via Facebook. The survey was distributed amongst friends,

and social media groups that are dedicated to communities that enjoy to travel. Some of these groups included:

- Travel Blogs Promotion
- NOMADS – a place where travelers…
- Backpacking
- Travelette Hosts

Time Allocation:

According to the work plan a total of 30 days were assigned to get as many participants as possible to take the survey. However, this was extended an additional 5 days as getting permission from administrators to post the survey on social groups took a little longer than expected.

Information Retrieval and Analysis:

Once the time allocated surpassed, the results of the questionnaire were downloaded onto excel spreadsheets. A total of 1151 responses were obtained. With the aid of Tableau some visuals were created to understand the population of survey takers. Figure 12 illustrates a dashboard containing some of the demographics. At first glance you will notice that travelers aged 24-30, female, attaining a bachelor's degree and having an annual income between $20,000-$50,000 make up a vast sample of the survey population. This is further represented in Figure 13.

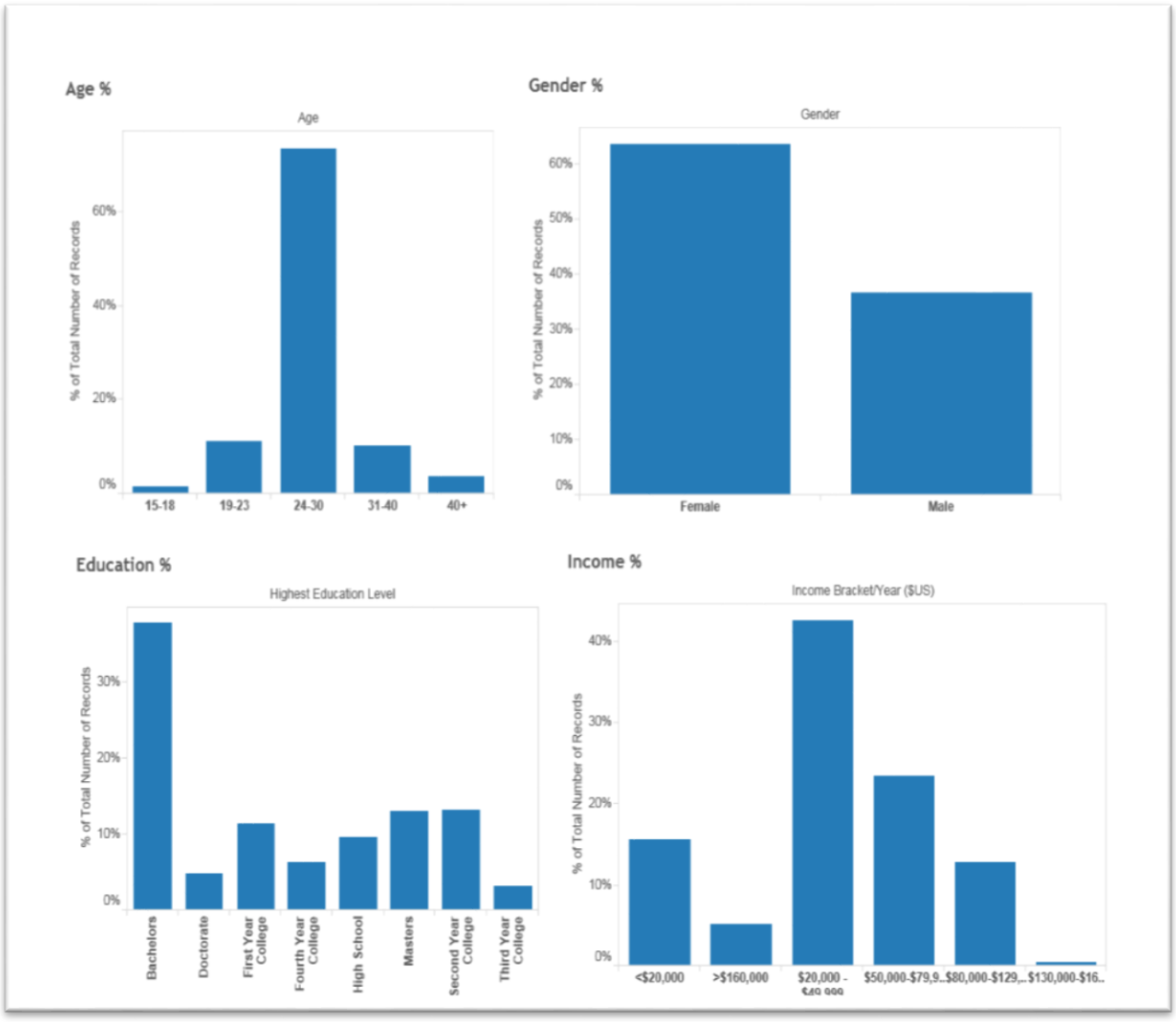


Figure 12: Demographics of Participants

Figure 13: Demographic Bubble Chart

An error encountered when creating these visuals were having an open ended answer for nationality in the questionnaire. Different people had different interpretations to what nationality meant to them. It was commonly replaced with race and ethnicity. A more appropriate question would have been, “country of origin” or a drop down menu with set options. However, the data was cleaned up on Excel with the aid of the “COUNTIF” function and uploaded onto Tableau to

get a clean visual of survey takers, displayed on a geographic map as seen in Figure 14. It is no surprise that the majority of the participants are from the United States considering this is the main point of contact, where the research took place.

Figure 14: Geographic Map of Participants

Considering this research is narrowed down to understanding social media interest amongst travelers within one specific country, it is important to draw out comparisons in the demographics of travelers who have visited Sri Lanka. The results of this analysis can be seen in Figure 15.

As you will notice right away the gender does not play a significant role amongst travelers visiting Sri Lanka. However, the age group (24-30), Education (Bachelors) and Annual Income (<20000) seem to stand out from the rest of the variables within the category.

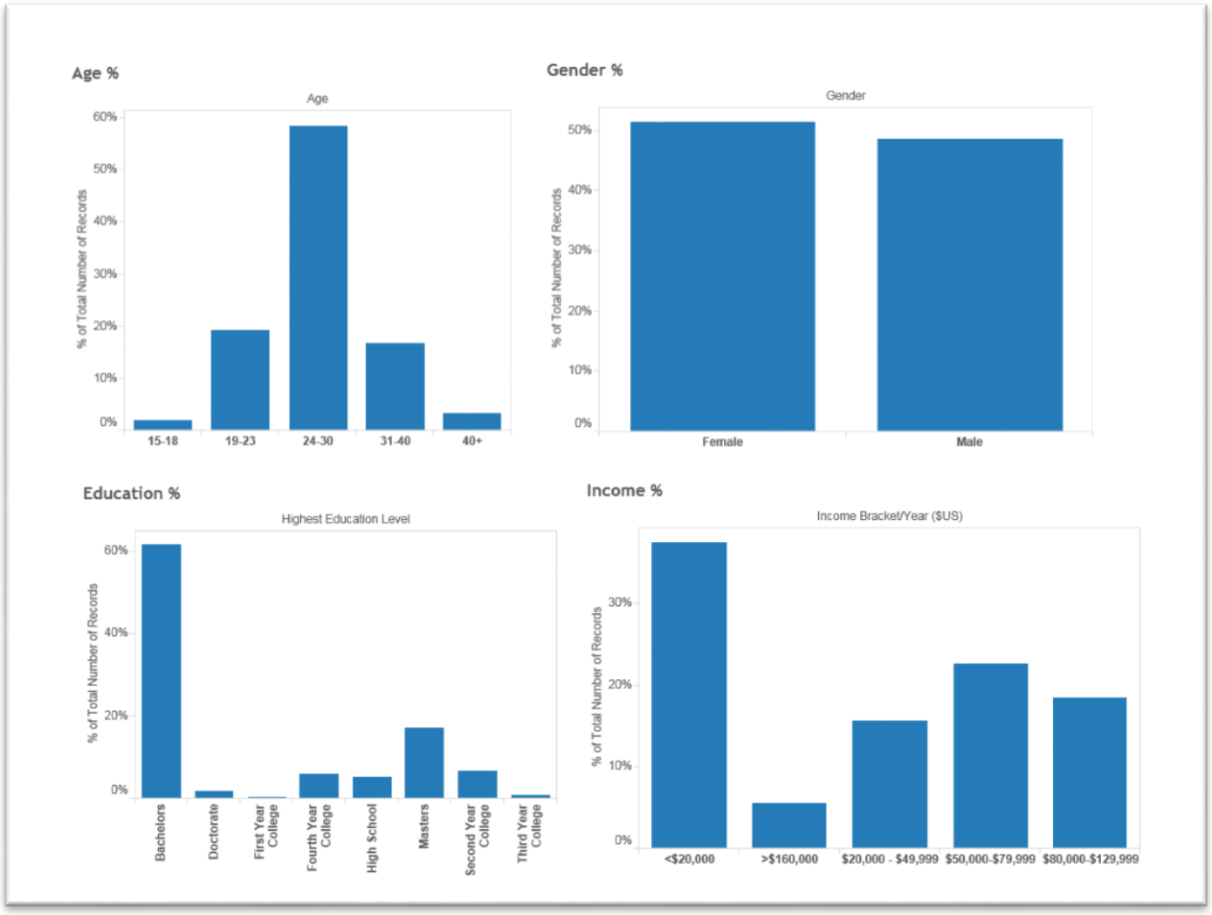


Figure 15: Demographics of Visitors to Sri Lanka

To get a better understanding as to whether the difference in demographics has any correlation to the people visiting Sri Lanka, an analysis of variance was done (ANOVA) on Gender, Age, Education and Annual Income. Considering all four variables were assessed as categorical values due to simplicity in taking the questionnaire, the first step was to convert them into numeric values. The tables below demonstrate the values used.

GENDER

Table 4: Gender Numerical Values

MALE	1
FEMALE	0

AGE

Table 5: Age Gap Numerical Values

15-18	1
19-23	2
24-30	3
31-40	4
40+	5

EDUCATION

Table 6: Education Level Numerical Values

High school	1
First Year College	2
Second Year College	3
Third Year College	4
Fourth Year College	5
Bachelors	6
Masters	7
Doctorate	8

Using excel the data was singled out to only show travelers that have visited Sri Lanka. Once this information was obtained and the categorical values were converted into numeric, an ANOVA table was obtained as seen below.

Table 7: ANOVA Table

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Gender	262	85	0.324427	0.220014
Age	262	787	3.003817	0.402284
Education	262	1277	4.874046	3.964918
Income	255	660	2.588235	1.833719

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2744.259	3	914.7531	570.4046	6.9E-219	2.613486
Within Groups	1663.028	1037	1.603691			
Total	4407.287	1040				

The null hypothesis was tested, which in this case states that the means of all variables are the same and have no significant effect on the demographics of people traveling to Sri Lanka. Studying the P-value obtained, and seeing that it is extremely close to 0 (below 1%) we have to reject the null hypothesis. With access to software such XLMiner (Excel Add-On), further analysis on a granular level can be run, such as regression analysis, to determine which variables in specific have a significant impact on the hypothesis.

Once the demographics of the reviewers were studied, the next step was to analysis the results of the questionnaire to fully understand social media interaction amongst travelers.

Figure 16 breaks down the mean result for each question, illustrated with the aid of a bar chart. At first glance you will notice that apart from the bar “I put phots on social media websites” the next top 3 bars are for “I look at activity/attraction reviews of other travelers”, “I read hotel reviews from other travelers” and “I read other travelers experiences and tips”. This goes to show that there are more travelers consuming social media on their travels than actually creating it. This claim is further supported in Figure 17.

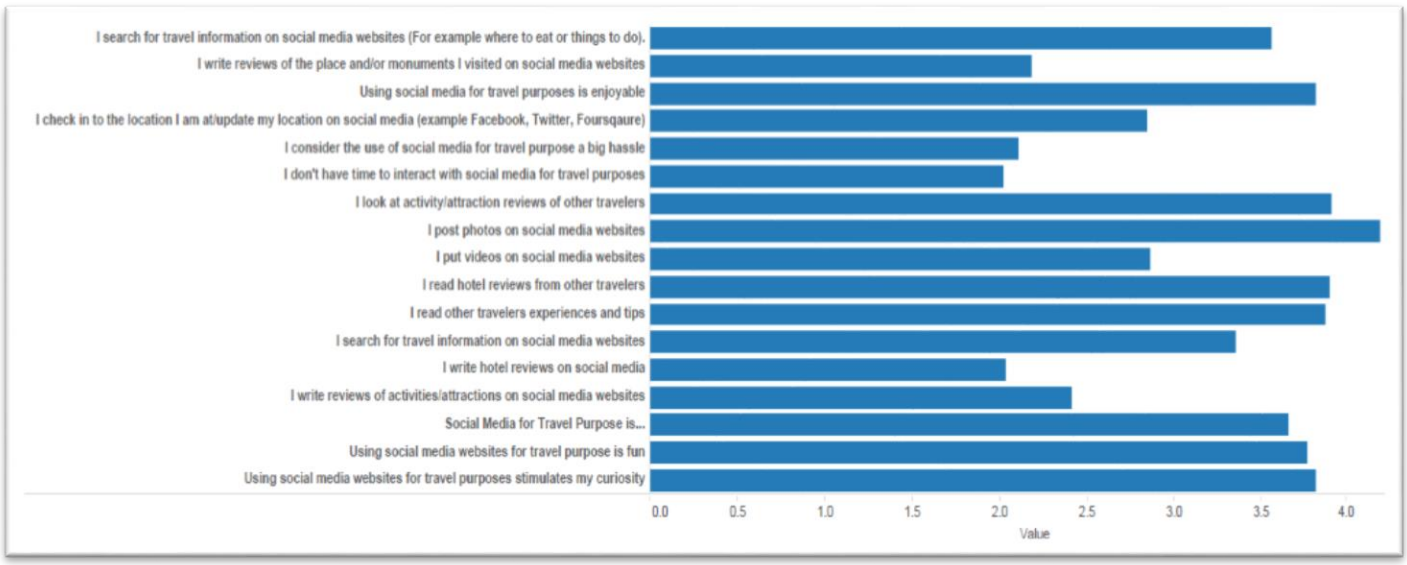


Figure 16: Questionnaire Results

Figure 17 also shows that there is a great deal of interest in Social Media amongst travelers.

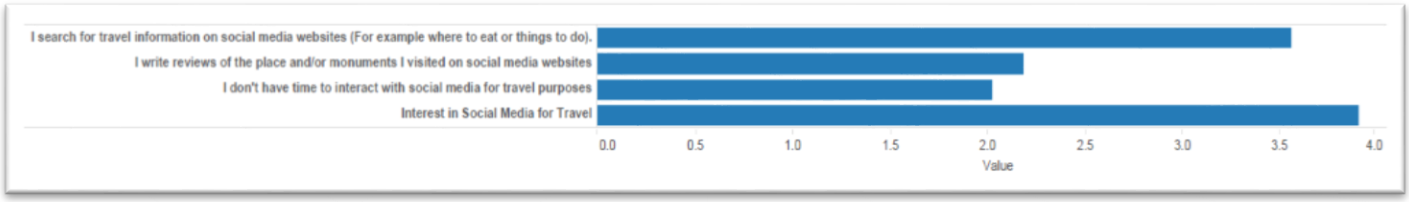


Figure 17: Grouped Questionnaire Results

Conclusion and Recommendations:

As stated earlier the purpose of this paper was to apprehend the importance of social media in decision making amongst travelers and understanding how the user generated content can be analyzed and leveraged to effectively market tourism within Sri Lanka.

The results obtained can verify some of the topics that were assessed in the literature review. Osborn's statement that, "75% of internet users use social media but less than half participate" can be clarified in the questionnaire study (Figure 17). This is vital information to marketing groups in promoting holiday destinations. If a hotel for example is aware of a conversationalist (travelers with higher levels of involvement) staying at their premise, they should emphasize on meeting the needs of this particular customer, as their interaction with social media can greatly influence many other travelers' decisions for accommodation. As stated before, 80% of travelers read 6 reviews before booking accommodation, and the same study found that reviews trump pricing. For these reasons alone, and with that discovered in this study, tourist destinations need to put a great deal of emphasize in discovering conversationalists on social media sites, to leverage their marketing campaigns.

The analysis and visualizations of the data obtained can be beneficial to many different industries in promoting tourism within Sri Lanka. The geographic map (Figure 11), pin points the most popular cities visited. This visual can be analyzed to determine a number of factors such as, what areas/terrain most travelers are seeking, how to enhance the transportation infrastructure between popular areas, when and where promotional events should be held, what areas should not be

considered for development projects along with a number of other aspects. Geo-location tracking is beneficial in understanding the movement of travelers within a country. Considering each city or location within a country does not have a border to track movement within, the right technology and platform can assist in getting accurate results.

The Sri Lankan government has already taken initiative to lay down a platform for ICT by partnering with Google in the Loon Project (The Guardian, 2015). Project Loon is Google's balloon powered high speed internet service that aims to help connect remote regions of the world at cheaper rates. Google has come to agreement with the Sri Lankan government along with Indonesia and parts of Australia to trial the technology. If everything goes to plan the whole country would have access to cost effective high speed internet. This would greatly influence the already exponentially growing number of reviews of the country on social media in the years to come. With the vast amounts of unstructured data being created through reviews of Sri Lanka on social media sites, it would be beneficial to implement a plan to capture this data for analytical purposes with the aid of up to date database management systems (NoSQL) and Business Intelligence tools for quick real time decision making.

Business Intelligence tools are great into getting insights into data at a much faster rate. For example, Sri Lanka being an island along the equator, most people would assume that it is a beach destination and is what a lot of travelers seek during their visits. However, an analyst who is not a domain expert in tourism within Sri Lanka, can glance at Figure 11 created by Tableau, and determine that a vast majority of the reviews are coming from a city inland (Kandy). Kandy is a city up in the hills located towards the center of the country, well known for its historical relevance

and monuments such as temples, architecture along with a cooler climate compared to the coastal areas. Further analysis into this quick observation can reveal many insights into what is attracting many tourists and help promote those aspects onto similar minded people. Considering many of the neighboring countries along the equator provide similar beach destination getaways, an insight as mentioned above can be what gives Sri Lanka the edge it needs for marketing. The proper implementation of Business Intelligence tools can also give rise to predictive analytics as described in Figure 8. Information as such would be beneficial to hotels, the tourist board, and other departments in understanding what to expect in the coming years and how to utilize resources to accommodate.

Effective marketing strategies can leverage the information obtained to target similar groups visiting Sri Lanka. Figure 15, breaking down the demographics of reviewers, aids in understanding the most receptive traveler to visit Sri Lanka. For example, more funds would be spent on promotional campaigns to advertise to travelers of the age 24-30 over other age groups according to the data obtained.

In recent years many marketing firms and agencies have started utilizing the power of big data by partnering with DSP's (demand side platforms) and altering the medium and method in which ads are projected. Traditionally, firms used the spray and pray model for reaching their clients. Placing ads on a billboard, magazine, paper and other mediums and hoping it would get the attention of clients and pick up traction. However, DSP's with the aid of DMP'S (Data Management Platforms) have been able to capture, store and track information on many different users over the internet at a more granular level. The technology is able to distinguish many different users according too

different KPI's (Key Performance Indicators). So if the goal was to run a marketing campaign to promote tourism within Sri Lanka, we would use the data obtained in this research and pass the information along to a DSP. The DSP would utilize this information and target advertise via the internet to people who meet this specific criterion along with the data they have obtained using their own KPIs. Examples of a DSP's KPI would include, the best time to reach a person that falls into specific groups, what medium or device would the campaign be most effectively utilized on, what sites generate the most traffic for the following criteria of individuals to display the add on. Understanding the emphasize and influence big data has on advertising and marketing is essential in competing on a global platform.

It can be concluded that social media has a great impact on the tourism industry. UGC is seen by users as more honest and reliable. Therefore, it is taking priority on decision making over previous influences such as price. Destinations not featured on top travel sites have a greater probability of losing out to competitors. The exponential growth of UGC on social media and the necessity to understand consumer behavior has given birth to new industries and technologies to bring insight. This development in the tourism industry is a bi-product of the big data evolution. As social norms merge with society and the ability to capture and dissect the vast amounts of information being created, the customer vs client playing field has evolved drastically and is still changing at a rapid pace. Many businesses and whole industries have crumbled for not keeping up to par with technology and Sri Lanka has personal experience and has faced repercussions with the aviation industry. For these reasons alone it is important in invest the time and resources into implementing a platform to understand UGC created by travelers to promote tourism growth within a country.

Appendix A:

Sri Lanka Tourism Development Authority
Statistical Report 2014

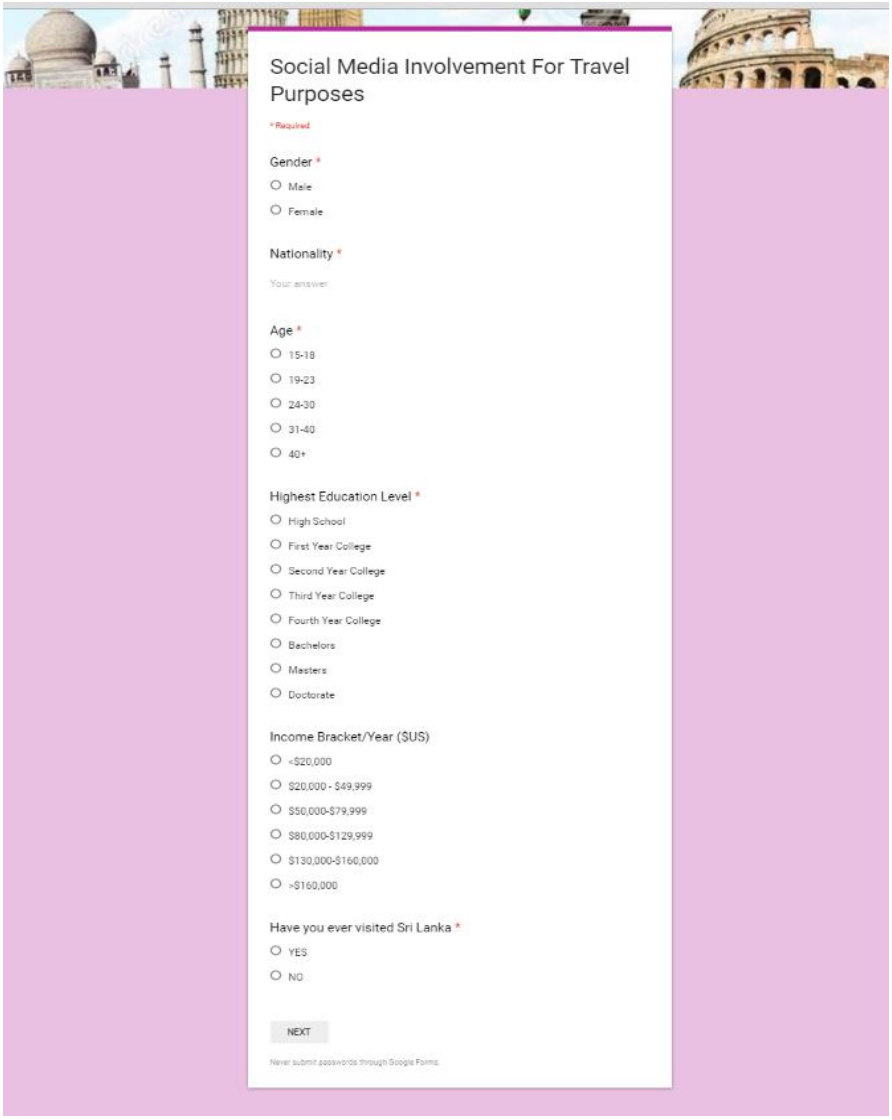
Table 1

Tourist Arrivals by Country of Nationality - 2008 to 2014

Country of Nationality	2008	2009	2010	2011	2012	2013	2014
NORTH AMERICA	23,203	25,044	40,552	53,658	56,694	70,003	72,653
Canada	9,745	10,785	21,231	26,090	28,786	30,382	32,156
U.S.A.	13,458	14,259	19,321	27,568	27,908	39,621	40,497
LATIN AMERICA & THE CARIBBEAN	3,721	666	594	1,167	1,001	3,109	3,202
WESTERN EUROPE	165,822	170,186	255,172	320,431	377,873	418,395	498,758
Austria	2,684	2,411	3,963	6,185	6,601	9,394	11,201
Belgium	2,394	2,613	5,371	10,853	13,013	10,432	11,532
Denmark	1,336	1,306	4,301	6,640	9,753	6,746	10,745
Finland	489	742	1,971	3,819	6,230	2,763	8,859
France	10,703	16,205	31,119	50,175	55,473	62,771	82,874
Germany	30,154	29,664	45,981	55,339	69,652	91,150	105,432
Italy	9,137	7,214	11,512	14,182	17,661	17,860	22,520
Netherlands	13,180	11,297	17,628	26,004	25,564	21,989	23,165
Norway	1,620	1,669	3,884	5,135	9,093	8,971	5,876
Spain	2,251	2,383	4,489	5,879	9,709	8,852	9,864
Sweden	3,745	3,578	7,128	11,715	15,165	15,423	13,287
Switzerland	5,261	6,247	9,514	13,161	21,644	17,329	18,965
U.K	80,214	81,682	105,938	106,895	111,828	135,425	153,875
Others	2,654	3,175	2,373	4,449	6,487	9,290	20,563
EASTERN EUROPE	29,352	26,177	35,630	51,922	70,437	112,737	141,984
Russia	15,831	11,390	13,312	21,291	30,156	47,265	49,876
Ukraine	765	5,177	5,582	9,067	20,530	34,067	35,674
Others	12,756	9,610	16,736	21,564	19,751	31,405	56,434
MIDDLE EAST	16,701	23,821	37,501	36,376	58,901	62,680	88,654
AFRICA	2,354	1,591	2,249	6,736	6,912	7,717	10,674
ASIA	175,944	174,382	245,753	334,274	383,748	527,656	640,045
Bangladesh	1,438	1,295	1,846	4,934	5,748	10,037	9,598
China (P.R.)	10,436	9,899	12,234	17,013	21,220	51,704	112,867
India	88,628	83,650	125,112	178,359	191,281	229,674	238,951
Indonesia	1,258	1,039	1,281	2,011	2,812	11,161	26,786
Japan	10,578	10,931	14,998	20,951	23,421	33,506	42,136
Korea (South)	4,102	3,595	4,318	5,965	6,133	11,700	13,543
Malaysia	5,021	6,878	13,101	15,915	29,181	20,914	22,120
Maldives	31,458	31,890	35,401	43,926	45,321	78,726	82,342
Nepal	897	679	602	858	984	2,019	3,296
Pakistan	7,702	7,388	9,001	15,857	14,543	24,095	24,657
Philippines	1,520	1,421	1,369	2,394	4,761	11,745	15,210
Singapore	5,702	7,976	12,514	10,666	15,453	15,020	17,643
Thailand	3,510	3,198	3,713	5,403	7,904	9,323	8,765
Taiwan (P.C.)	1,988	2,711	5,019	7,067	11,231	10,736	7,564
Others	1,706	1,832	5,244	2,955	3,755	7,296	14,567
AUSTRALASIA	21,378	26,023	37,025	51,411	50,039	72,296	71,183
Australia	19,145	23,249	33,512	43,737	42,310	62,242	59,786
New Zealand	2,148	2,623	3,301	5,175	5,837	7,023	8,854
Others	85	151	212	2,499	1,892	3,031	2,543
Total	438,475	447,890	654,476	855,975	1,005,605	1,274,593	1,527,153

15

Appendix B:



Social Media Involvement For Travel Purposes

* Required

Gender *
- Male
- Female

Nationality *

Your answer

Age *
- 15-18
- 19-23
- 24-30
- 31-40
- 40+

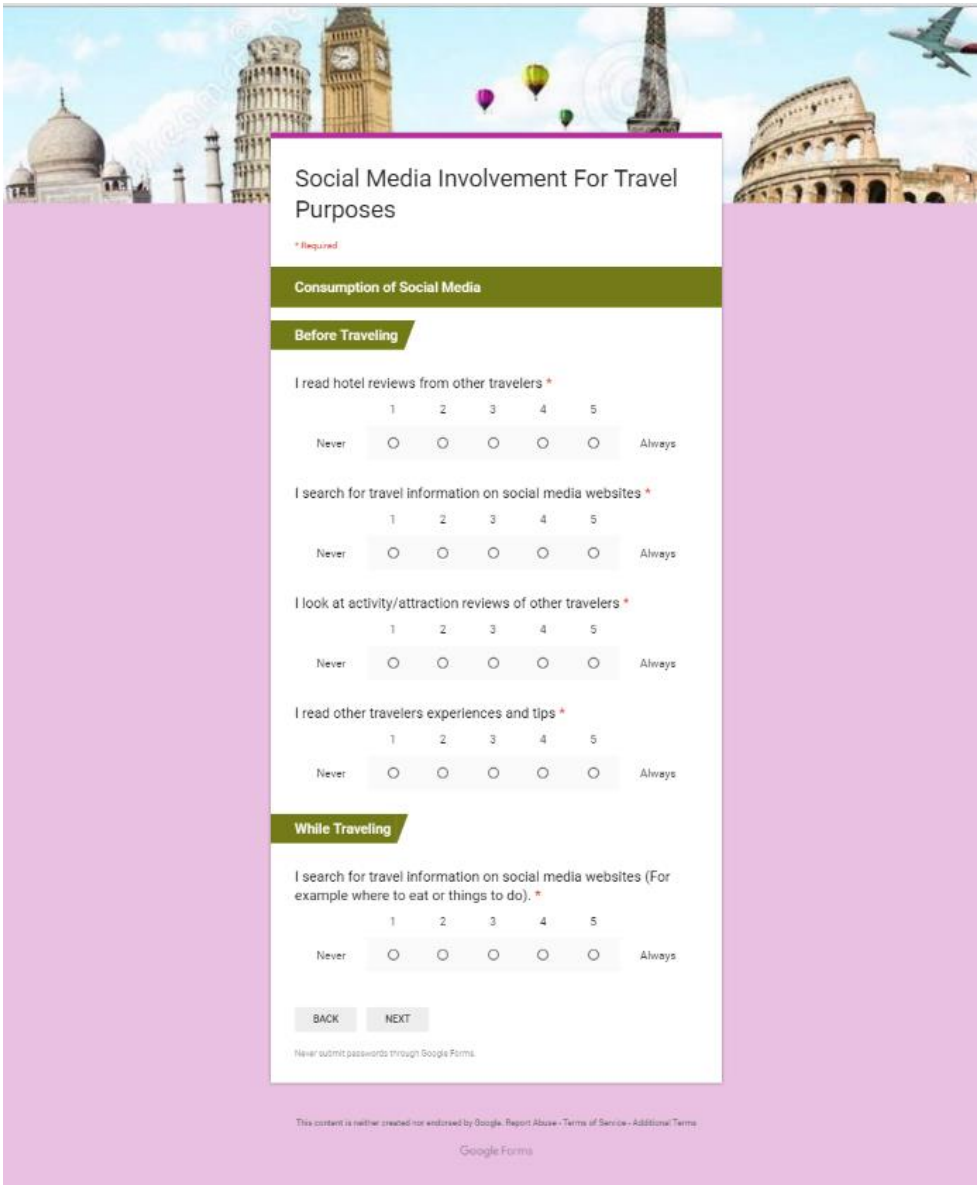
Highest Education Level *
- High School
- First Year College
- Second Year College
- Third Year College
- Fourth Year College
- Bachelors
- Masters
- Doctorate

Income Bracket/Year ($US)
- <$20,000
- $20,000 - $49,999
- $50,000-$79,999
- $80,000-$129,999
- $130,000-$160,000
- >$160,000

Have you ever visited Sri Lanka *
- YES
- NO

NEXT

Never submit passwords through Google Forms.



Social Media Involvement For Travel Purposes

* Required

Consumption of Social Media

Before Traveling

I read hotel reviews from other travelers *

	1	2	3	4	5	
Never	○	○	○	○	○	Always

I search for travel information on social media websites *

	1	2	3	4	5	
Never	○	○	○	○	○	Always

I look at activity/attraction reviews of other travelers *

	1	2	3	4	5	
Never	○	○	○	○	○	Always

I read other travelers experiences and tips *

	1	2	3	4	5	
Never	○	○	○	○	○	Always

While Traveling

I search for travel information on social media websites (For example where to eat or things to do). *

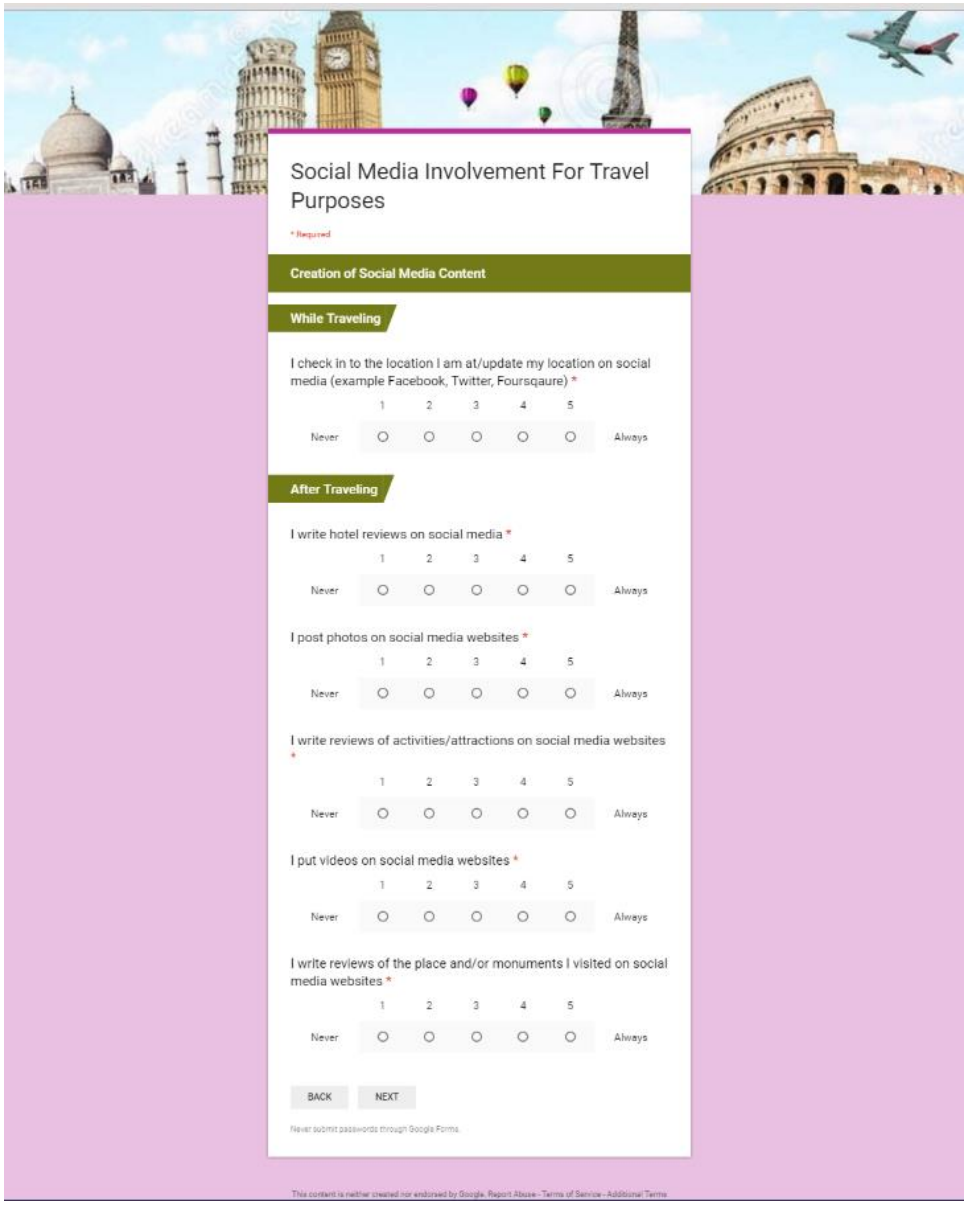
	1	2	3	4	5	
Never	○	○	○	○	○	Always

BACK NEXT

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Additional Terms

Google Forms



Social Media Involvement For Travel Purposes

* Required

Creation of Social Media Content

While Traveling

I check in to the location I am at/update my location on social media (example Facebook, Twitter, Foursqaure) *

	1	2	3	4	5	
Never	○	○	○	○	○	Always

After Traveling

I write hotel reviews on social media *

	1	2	3	4	5	
Never	○	○	○	○	○	Always

I post photos on social media websites *

	1	2	3	4	5	
Never	○	○	○	○	○	Always

I write reviews of activities/attractions on social media websites *

	1	2	3	4	5	
Never	○	○	○	○	○	Always

I put videos on social media websites *

	1	2	3	4	5	
Never	○	○	○	○	○	Always

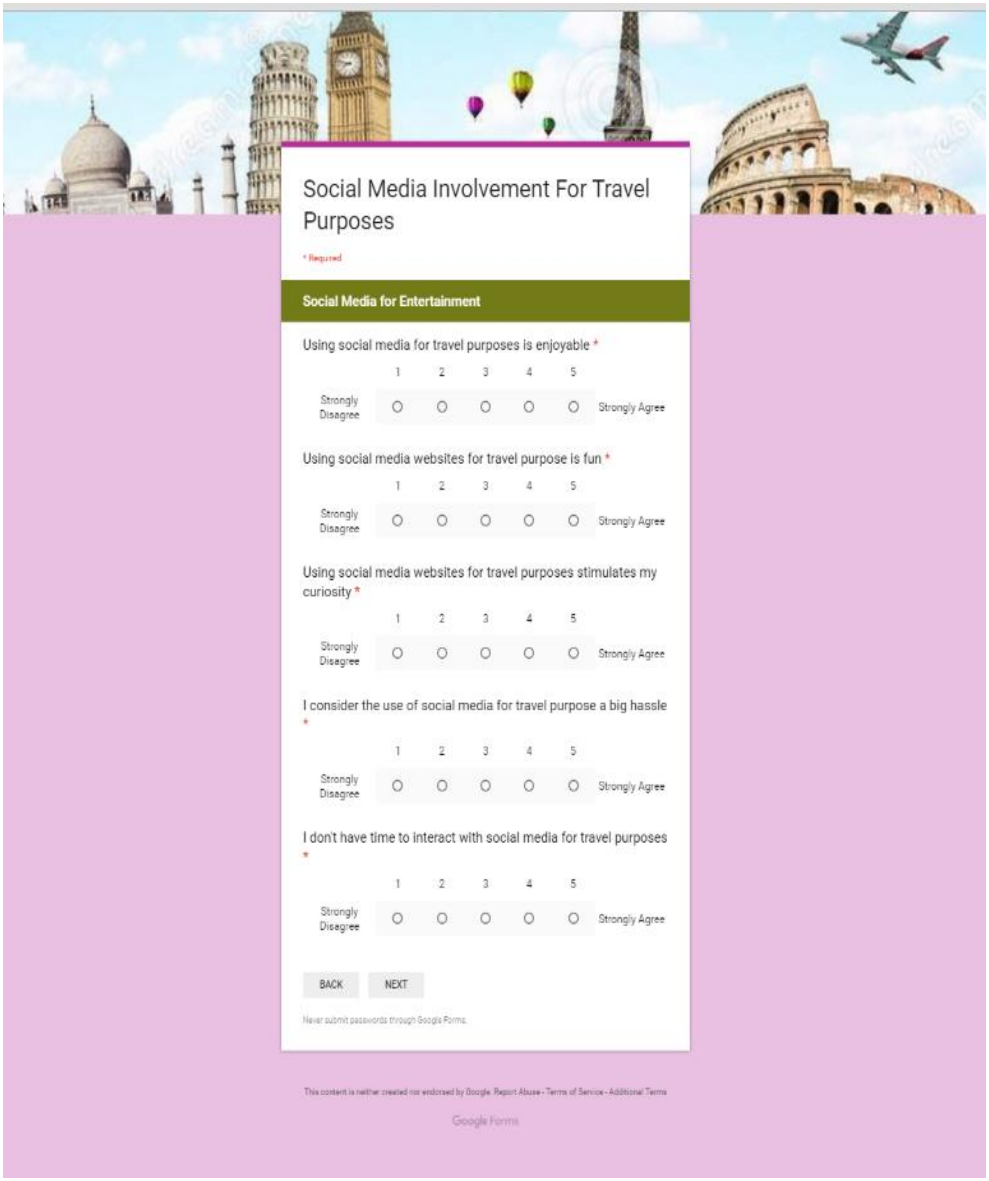
I write reviews of the place and/or monuments I visited on social media websites *

	1	2	3	4	5	
Never	○	○	○	○	○	Always

BACK NEXT

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Additional Terms



Social Media Involvement For Travel Purposes

* Required

Social Media for Entertainment

Using social media for travel purposes is enjoyable *

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

Using social media websites for travel purpose is fun *

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

Using social media websites for travel purposes stimulates my curiosity *

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

I consider the use of social media for travel purpose a big hassle *

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

I don't have time to interact with social media for travel purposes *

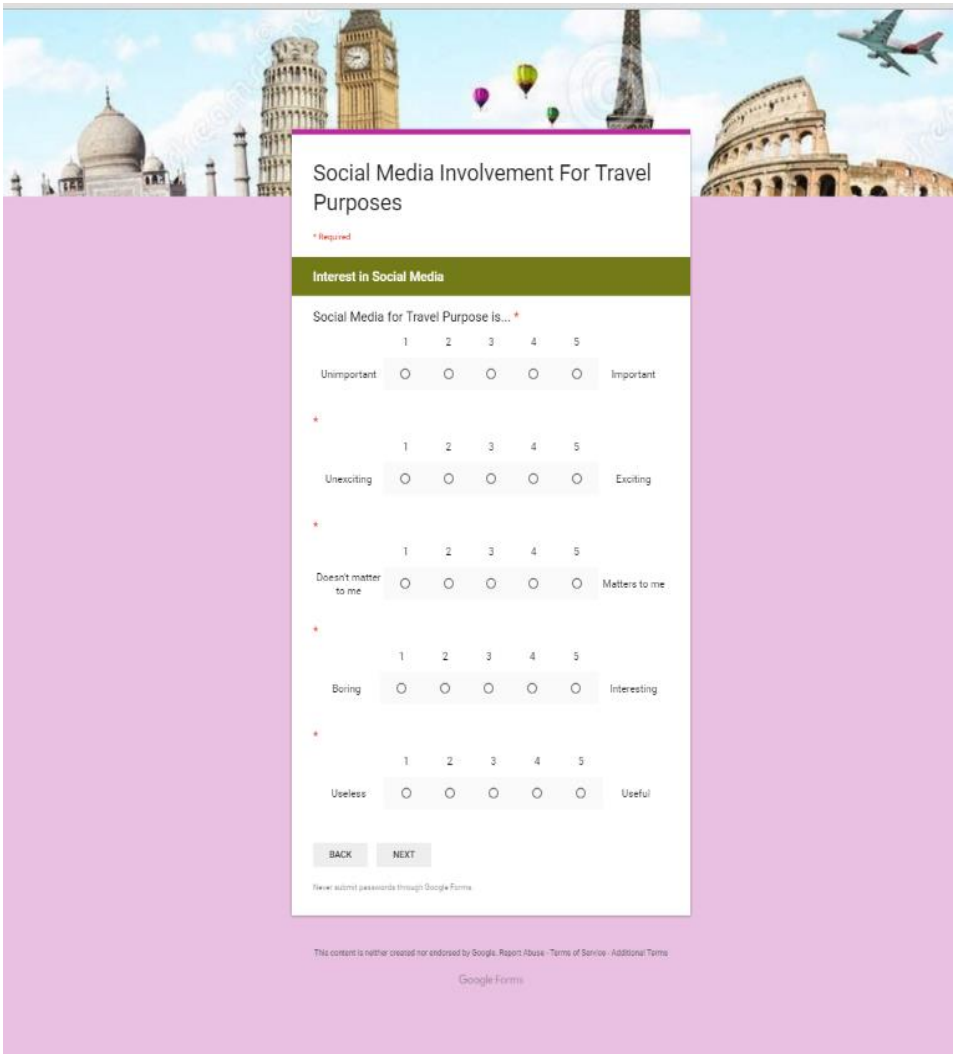
	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

BACK NEXT

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Additional Terms

Google Forms



Social Media Involvement For Travel Purposes

* Required

Interest in Social Media

Social Media for Travel Purpose is... *

	1	2	3	4	5	
Unimportant	○	○	○	○	○	Important

*

	1	2	3	4	5	
Unexciting	○	○	○	○	○	Exciting

*

	1	2	3	4	5	
Doesn't matter to me	○	○	○	○	○	Matters to me

*

	1	2	3	4	5	
Boring	○	○	○	○	○	Interesting

*

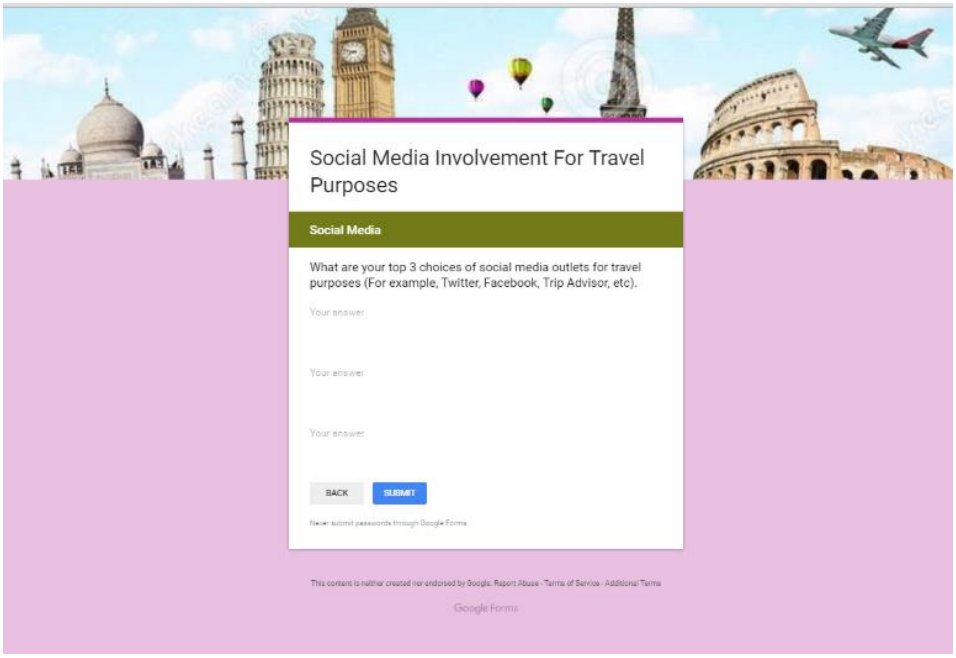
	1	2	3	4	5	
Useless	○	○	○	○	○	Useful

BACK NEXT

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Additional Terms

Google Forms


Social Media Involvement For Travel Purposes
Social Media
What are your top 3 choices of social media outlets for travel purposes (For example, Twitter, Facebook, Trip Advisor, etc).
Your answer
Your answer
Your answer
BACK
SUBMIT
Never submit passwords through Google Forms.
This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Additional Terms
Google Forms

REFERENCES

Beaver, Kumar, Li, Sobel and Vaigel (2010). Finding a needle in a haystack: Facebook's photo data storage. In Proceedings of the ninth USENIX conference on operating systems design and implementation.

Buhalis, D., and Amaranggana (2014). A Smart Tourism Destination – Information and Communication Technologies in Tourism

Estela Marine-Riog (2015). Tourism Analytics with Massive User Generated Content – A Case Study by Barcelona

Estela Marine-Riog and Salvador Anton Cave (2015). A Method for Analyzing Large Scale UGC Data for Tourism: Application to the Case of Catalonia.

Fatma Ahmad (2013). Peace Building in post Srilankan Civil War: Reconstruction and Rehabilitation Efforts

Ganomi and Haider (2015). Beyond the Hype. Big Data Concepts, Methods and Analytics. International Journal of Information Management.

Hu, Wen, Chua and Li (2014). Toward scalable systems for big data analytics: A technology tutorial.

Jeremy Kressmann (2015). The US Travel Industry 2015 – Digital Ad Spending Forecasts and Trends

Lu and Stepchenkova (2014). User generated content as a research mode in tourism and hospitality application: Topics, methods and software. Journal of Hospitality Marketing & Management.

Ruwan Ranasinghe, Ravindra Deyshappriya (2010), Analyzing the significance of tourism on Sri Lankan Economy; an econometric analysis.

Suzanne Amaro, Paulo Duarte (2015). Travel Social Media Involvement: A Proposed Measure.

Shao (2009). Understanding the appeal of user generated media.