ISCTE Business School
Instituto Universitário de Lisboa

MAIN CONTRIBUTION OF ICONIC ATTRACTIONS TOWARDS INCREASING POPULARITY OF TOURISM DESTINATIONS: AN ANALYSIS OF TWITTER POSTS AND LOCATIONS

Anel Imanbay

Dissertation submitted as partial requirement for the conferral of

Master in Marketing

Supervisor:

Prof. Paulo Miguel Rasquinho Ferreira Rita, Full Professor, ISCTE Business School, Department of Marketing, Operation and Management

Co-supervisor:

Prof. João Ricardo Paulo Marques Guerreiro, Assistant Professor, ISCTE Business School, Department of Marketing, Operation and Management

September 2017

"If a man empties his purse into his head, no man can take it away from him. An investment in knowledge always pays the best interest."

- Benjamin Franklin

DEDICATION AND ACKNOWLEDGEMENTS

Dedication: This dissertation is dedicated to my University ISCTE and the beautiful Portugal. I owe them both for the immense knowledge that I have gained regarding the field of marketing, and the amazing experience that I have had with the faculty and friends here. I was blessed to be able to discover not only the unique nature and culture of this land, but also it helped me to grow in my professional, student and personal life.

Acknowledgments: From the very early stage of writing this dissertation, I felt connected and inspired by the idea of being able to contribute to the theory of tourism and associated marketing trend, as I believe travelling brings number of positive impact in many aspects of our life. However, in spite of the growth in my professional career, I had to face many challenges with the personal discipline, time management and patience. Yet life brought me to meet and know exceptional people, without support of whom I would not be able to succeed.

I would fore mostly like to thank my supervisor Professor Paulo Miguel Rasquinho Ferreira Rita, and my co-supervisor Professor João Ricardo Paulo Marques Guerreiro, for their relentless support and encouragement throughout my study tenure. The execution of the study and its fine completion within the allotted time frame would not have been possible without their extended support and commitment. Even though I had the experience of living and studying in many countries, it was my first time to get to know such mentors, who would offer the help despite their busy and tight schedules, and always finding availability regardless if it is via phone, skype, email or in person. Their enhanced knowledge of the various aspects of Marketing, support and believe in me till the last minute was the main source of my motivation and inspiration.

I am also grateful to my fiancée Ricardo Jorge Godinho Cortez Dos Santos, for all the moral support that he has provided me with, and his ability to creatively brainstorm ideas and guide me whenever I felt lost.

Finally, I am indebted to my family in Kazakhstan for positive thoughts, trust and supporting me at every corner of the road.

ABSTRACT

The social media platforms, due to their universal and comfortable interface, have become the real enablers of a microblogging services. Moreover, with the evolution of online reviews, consumers feel comfortable to express their opinions and share their personal experiences not only about the brands, but also about the travel destinations. Henceforth, social networks such as, Twitter, became important source of information. In this study, author analyzes 4,000 Twitter posts about 2 popular and 2 less popular locations and associated derived sentiments. The study demonstrates that there is a certain difference in perception of locations with a different popularity rank. In terms of information exposure, more popular locations tend to have a higher message diffusion activity, with most of them being of neutral polarity. Additionally, results showed that negative affection is observed more for less popular locations, providing valuable insight for Destination Marketing Organizations. In addition, for both groups, role of followers' base was ineffective, demonstrating that topic of message sentiment and diffusion are key in tourism domain. Thus, from a methodological point of view, the main contribution of this research is the usage of random and unstructured data in Twitter to the measurement of the perception of the potential visitors of tourist attractions based on the sentiment analysis of posts associated to them. From theoretical point of view, using the sentiment orientation, the study relates to the user exposure and affection of the iconic attractions by the perceived difference in their popularity in accordance with external destination ranking.

Keywords: Iconic Attractions, Popular and Less Popular Tourism Destinations, Sentiment analysis of Twitter Posts; Text Mining.

JEL Classification system:

M310 - Marketing

Z320 - Tourism and Development

RESUMO

As redes sociais, devido ao seu interface universal e confortável, tornaram-se reais facilitadores de serviços de microblogging. Por conseguinte, a evolução dos reviews on-line, conferiu aos consumidores maior conforto para expressar as suas opiniões e partilhar as suas experiências pessoais, não apenas sobre as marcas, mas também sobre os seus destinos de viagem. As redes sociais, como o Twitter, tornaram-se importantes fontes de informação. Neste estudo, o autor analisa os sentimentos derivados de 4.000 publicações do Twitter acerca de 2 locais turísticos mais populares e 2 menos populares. O estudo demonstra que há uma certa diferença na percepção dos locais em função do seu grau de popularidade. Em termos de exposição, os locais mais populares tendem a ter uma maior atividade de difusão nas suas mensagens, sendo a maioria delas de polaridade neutra. Adicionalmente, os resultados mostraram que o sentimento negativo é mais partilhado em locais menos populares, fornecendo informações valiosas para Organizações de Marketing. Não obstante, para ambos os grupos, a dimenção da base de seguidores foi irrelevante, demonstrando que o tema da mensagem sentimento e difusão são fundamentais no domínio do turismo. A nível metodológico, o principal contributo desta pesquisa é a análise do sentimento de dados aleatórios e desestruturados do Twitter para a medição da percepção acerca de atracções turísticas com base na. Do ponto de vista teórico, o estudo relaciona-se com a exposição do usuário e o sentimento das atrações icônicas pela diferença percebida na sua popularidade de acordo com um ranking de destinos externo.

Palavras-chave: Atracções icónicas, destinos turísticos populares e menos populares, análise de sentimento de postagens no Twitter, Text Mining.

JEL Classification System:

M310 - Marketing

Z320 - Tourism and Development

Table of Contents

INDEX OF FIGURES

INDEX OF TABLES

LIST OF ABBREVIATIONS

UNWTO	United Nations World Trade Organization
GDP	Goss Domestic Product
DMO	Destination Marketing Organization
UGC	User Generated Content
NTO	National Tourism Organization
eWOM	Electronic Word-of-Mouth
SPSS	Statistical Package for the Social Sciences

1 INTRODUCTION

Tourism, at a global scale, is defined as a socio-economic phenomenon that has experienced continued growth over its history, being classified as one of the fastest growing economic sectors in the world. In modern times, as suggested by UNWTO (2016), tourism has been attributed with more development and increased number of tourist destinations, leading it to surpass the business volume of oil exports, food products and even automobiles. Interestingly, the tourism industry renders a major impact on the development of the home economy. It yields benefits in terms of both income creation as well as employment generation (Bennett, 2014). For a host of countries, tourism remains an important source of welfare. Hence, the ability of a national economy to benefit from a thriving tourism industry fundamentally depends upon the strength of its infrastructure and its ability to cater to the needs of the tourists (Bennett, 2014). Thus, being a significant economic driver, the tourism industry generates income for the local community and may play a major role in the poverty alleviation of underdeveloped countries as well. Apart from economic benefits, tourism tends to bring about social benefits in terms of a sense of pride and identity for communities, essentially through showcasing the distinct characteristics of the local culture, heritage and traditions (UNWTO, 2016). Finally, the tourism industry inevitably yields environmental benefits in terms of conserving the ecosystems and the natural resources of the regions, adding more value to the local tourism business (UNWTO, 2016).

According to Christian (2015), travel and tourism renders both direct and indirect impact on the economy. The direct impacts include direct contribution to the GDP and direct contribution to the employment. Similarly, domestic travel and tourism spending also increases, along with visitor exports and government individual spending. On a similar note, the indirect and induced impacts include enhanced capital investment in terms of infrastructural development and improved government spending on aspects such as tourism promotion and administrative services (Christian, 2015).

Owing to the proliferation of the internet and the widespread use of digital technologies the global tourism industry has witnessed a paradigm shift, which has not only impacted the consumer side of the spectrum, but also the business side. While consumers increasingly rely on these technologies to seek and share relevant information, travel organizations deploy them to connect to the prospective clients, engage with them and make offers to promote the tourist destinations and products, reaching out to masses. Social media has emerged as a revolutionary tool that has impacted virtually all spheres of consumer life. Gradually becoming a part of their daily routine, consumers are using social media for literally every ongoing activity, may it be business, social life, education, or travel. According to Hays, Page and Buhalis (2013), the mere advent of the internet as well as its increased accessibility have significantly altered how tourists access information, how they plan their travel trips, and how they share their travel related experiences.

Gloor et al (2016) add that social media has enabled “seeing what is happening in the world” much easier, allowing real time interaction and information sharing over multiple networks, and promoting collective awareness around social events. Similarly, the study on Twitter by Hay (2010) concluded that consumers share and acquire information regarding common travel encounters in Twitter, and henceforth the platform can be used and classified as a tourism marketing tool for the destination marketing organizations.

Therefore, the use of social media, particularly twitter posts and locations, for improving the attractiveness of iconic attractions and tourist destination, remains the primary premise for this research. This delineates upon the important theoretical, conceptual and empirical aspects of the phenomenon. The research is intended to provide meaningful theoretical and practical contributions to the body of academic literature as well as practical realms of the DMO industry globally.

It has been established through preliminary research that while prior studies have been conducted on the impact of social media on the choice of consumers in terms of selecting tourist destinations, very little research has been conducted upon its impact on the activities of DMOs and consumers’ contribution to the attractiveness of such tourist locations. Moreover, country specific studies have been conducted on analyzing the role of Twitter in terms of enhancing the attractiveness of tourist destinations. For instance, Nguyen and Wang (2012)’s study was aimed as Swedish DMOs and their use of social media to streamline online marketing activities. Similarly, Bayram and Arici (2013)’s study was aimed at 12 Balkan countries’ DMOs and their use of Twitter in particular to attract tourists. However, no study so far has been aimed at providing a global perspective on the topic in terms of enhancing the popularity of tourist destinations, by comparison popular and less popular locations sentiments. Moreover, previous literature also neglected to use sentiment analysis as a measurement of popularity of tourist destinations that could significantly contribute towards competition evaluation of DMOs.

Therefore, Given the research gap identified above, the primary objective of the research remains to investigate the impact on tweeting activity on the popularity of a given tourist destination. Travel popularity, for the purpose of this research, is essentially defined by the rank position of given location as per TripAdvisor (2017a). Hence, the main goal of this dissertation is to examine how twitter posts contribute to the popularity of the tourism destination and the difference between sentiments of the iconic locations in general using text mining tools. Thus, the main research question is identified as following:

Are the popular and less popular destinations perceived the same way by the twitter users?

The structure of the paper is as follows: The second chapter provides a detailed critical review of the literature, discussing the works of numerous other authors on similar subject. It delineates upon the role of social media and user-generated content with relevance to the tourism industry, the theoretical framework behind text mining and data mining, and previous studies conducted on the use of Twitter and TripAdvisor to promote the popularity of various tourist destinations. The third section of the chapter provides the theoretical framework, where literature suggests to expand the study to the multiple locations, which serves as a base for the construction of the conceptual model and further builds the detailed research hypotheses that govern this dissertation.

The fourth chapter discusses the research methodology underpinning the study at hand. It begins with reiterating the research problem and moves forward to discuss the data mining and processing tools, the data set process, and the detailed data description.

Obtained results are shown in fifth chapter, together with empirical findings.

The last chapter concludes the research, while highlighting the discussion of usefulness of sentiment analysis for identification of the exposure and affection difference within the popular locations. The final section also presents the major findings with the limitations of the study as well as opportunities for further future research. It concludes with valid recommendations to enhance the popularity of iconic tourist destinations through strategic and tactful use of Twitter.

2 LITERATURE REVIEW

2.1 ROLE OF SOCIAL MEDIA IN TOURISM

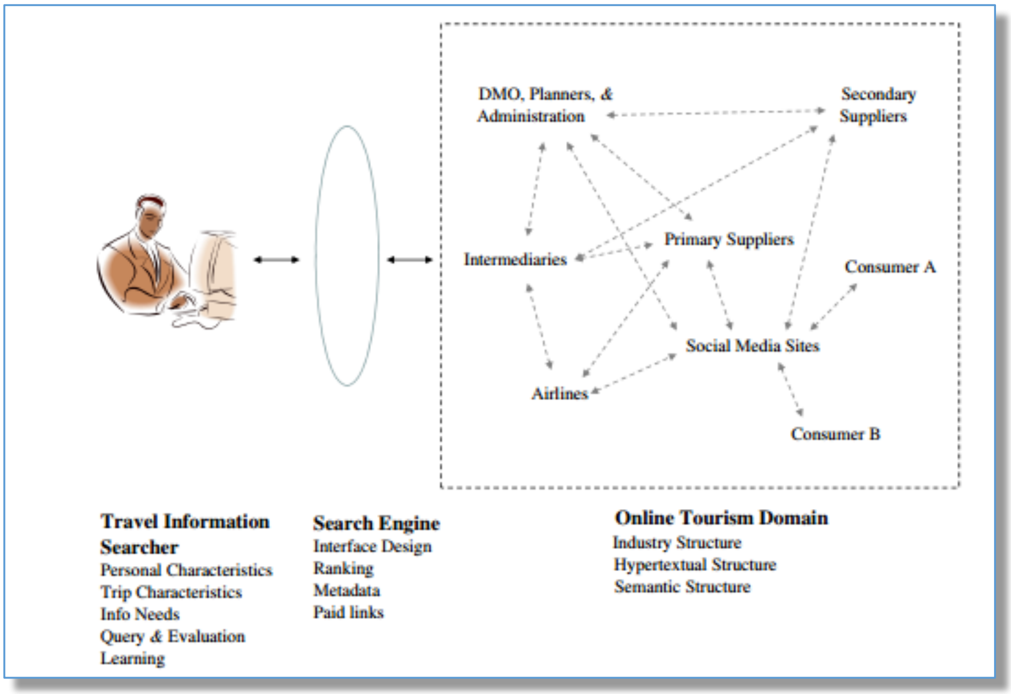
Given the rise of digital technologies and their extremely common usage in everyday life affairs, social media has evolved as a major breakthrough intervention to keep the tech-savvy global consumer connected around the clock. Speaking true of the tourism industry in particular, social media has emerged as a "mega trend" (Leung et al., 2013) that has rendered a major impact, allowing tourists to engage with social networking platforms to not only research trips and make informed decision about their next tourist destination, but also to share their travel experiences both in terms of texts and images (Moro, Rita and Vala, 2016). This Mega trend is equally popular amongst travelers as well as marketers for all sorts of travel related activities including information search, travel plans organizing, and sharing travel experiences (Leung et al., 2013). The importance of social media is further amplified given the experiential nature of the tourism industry, since the content generated by other travelers tends to have a strong influence on the information search as well as the decision making process of other travelers (Yoo and Gretzel, 2012).

Hence, according to Zeng and Gerritsen (2014), leveraging social media to market tourism destinations has proven to be an excellent strategy, particularly capitalizing on social media's tendency to facilitate information research as well as influence decision-making behaviors of tourists. Leung et al., (2013) add that social media is essentially driven by user-generated content which covers a vast array of activities including sharing photos, videos and reviews and comments. This characteristic of social media in particular tends to enhance the capability of the tourism and hospitality companies to engage with their potential clients, improve their online presence, and ultimately drive online revenues. Miguens, Baggio and Costa (2008) also discuss that the tourism sector holds a close relationship with the new innovative information and communication technologies whereby tourism managers can largely benefit from a good understanding of the quality and quantity of the information spreading mechanism. Therefore, it is critical to encourage comprehensive investigation to study the impact of social media as part of effective tourism management strategy on the overall socio-economic contribution of popular tourist destinations.

Deloitte (2013) highlights that owing to the rise of digital technologies, organizations in the tourism economy must necessarily invest in technologies to not only **segment and understand the tourists' needs** better but also to offer more relevant choices. Marketing efforts must be directed through online channels and social media to build awareness of the tourism destinations as well as to fight competition from other countries. It, however, remains imperative to note that social media plays an important role in influencing the travel planning process of the trailer's in all phases of the trip; the pre-trip information search phase, the during-trip phase, and the post-trip phase (Leung et al., 2013). Hays, Page and Buhalis (2013) add that the current economic climate calls for regional as well as local destination marketing organizations to invest in new technologies to provide multi-lingual websites, social media platforms and international public relations expertise to facilitate global tourists and travelers. Growth of social media has been acknowledged as a key trend in the international tourism economy, acting in the capacity of a mainstream driver of the tourism market. Furthermore, the report highlights an important point that social media is anticipated to present new challenges for the consistency of brands while also opening up new forms of dialogue with the consumers. Hence, social media particularly offers significant opportunities to not only build brand awareness but also to build brand communities.

Xiang and Gretzel (2010) discuss that social media plays a crucial role as **information source** for travelers. The online tourism domain offers all the relevant important information for travelers as well as establishes a touch point between the consumers and the industry, whereby social media plays a mediating role between travel planners and the travelers (Yoo and Gretzel, 2012). Xiang and Gretzel (2010) further conclude that in a typical setting, there are three important components: 1) the online traveler, who has certain trip-related needs; 2) the online tourism domain, which constitutes all the relevant international entities; 3) the search engine, which determine the representation of the tourism domain. The information offered by the third component in terms of search result rankings, etc. tend to influence the travelers' perception and decision making process. Figure 1 illustrates the social media and tourism domain within the context of using a search engine.

Figure 1: Social Media and Tourism Domain



Source: Xiang and Gretzel, 2010, pp 182

According to Blackshaw and Nazzaro (2006), social media typically allows the users to **share their experience** for easy access to other impressionable consumers, using a variety of applications in the technical sense. These applications allow the users to post, tag or blog on various social platforms to share their travel related content, with the intent of educating others (Yoo and Gretzel, 2012). Since an increasing number of travelers tends to tap into this "collective intelligence" over the web, it inevitably challenges the established marketing practices of the conventional tourism businesses as well as destinations (Yoo and Gretzel, 2012). Xiang and Gretzel (2010) also note that the consumer generated content over social media is essentially a mixture of facts and opinions, as well as impression and sentiment. Offering others with the insight on one's own travel experiences typically influences the socio-psychological aspects of consumer behavior. These authors also note that virtual tourist communities have emerged in the wake of increasing use of social media for sharing travel related information and experiences, that allow tourists to exchange their opinions and experiences on topics of common interest, such as LonelyPlanet and IGoUGo.

Moreover, evidence suggests that social media has emerged as an **effective destination marketing tool** widely being used by tourism organizations to attract tourists. Hays, Page and Buhalis (2013) suggest that since tourism is an "information sensitive industry", social media plays an immensely significant role in influencing the decision making process of the consumers through user-generated content as well as effective marketing strategies deployed by tourism organizations. These authors add that the trip planning process of the consumers is particularly facilitated through the information obtained through such content which significantly assists in making informed decisions regarding travel destinations, accommodation options, dining facilities, and favorite tourist attractions (Hays, Page and Buhalis, 2013). Given the fact that tourism experiences cannot be assessed prior to consumption due to being intangible in nature, personal recommendations and word of mouth referrals over social media are extremely influential. This is typically why online travel communities have sprung up to engage consumers and provide a platform to them for sharing their travel experiences.

Interestingly, **corporate communications** have radically changed in lieu of the rise of social media. Given the fact that marketing is essentially the process of satisfying consumers through continuous two-way communication, creation and exchange of value has significantly changed with the increasing empowerment of the internet. Hvass and Munar (2012) add that social media has a strong role to play in terms of transforming the ways in which customer relationships are established and maintained, while also drastically changing the communication patterns between the companies and their client base. This has inevitably shifted the "command-and-control" approach to a new digital, technology-based approach that rests on the principle of engaging and empowering individuals.

Moreover, the proliferation of social media usage for destination marketing has raised serious implications for corporate reputation amongst companies. According to Dojkmans, Kerkhof and Beukeboom (2015), the company's social media activities are directly related to the consumers' (as well as non-consumers') perceptions of corporate reputation. The authors discuss that the level of consumers' engagement in a given company's social media activity is positively related to their perceptions of corporate reputation, part of which can be linked to emotional contagion; that is, the ability and tendency of being influenced by the emotions of others. Interestingly, this relationship between social media engagement and corporate reputation is rather more pronounced amongst the non-consumers, who essentially constitute the largest part of the company's target markets.

Miguens, Baggio and Costa (2008) conducted a specific research aimed at analyzing the role of social media in **promoting tourism destinations**, focusing the case study on TripAdvisor, an online social travel networking platform that allows tourists to plan their trips online. While providing a platform to interact with others, it also offers reviews sharing regarding hotels and local tourist attractions, etc. The results indicate that TripAdvisor is a major tool for rating hotels, based on user-generated content. Given the increasing importance of the online travel market, social media on the whole has emerged as a crucial factor influencing the popularity of tourist destinations. Moreover, not only is social media an important determinant for promoting single operators, but also the destination on the whole.

Similarly, Lange-Faria and Elliot (2012) also add that social media has become the modus operandi of the 21st century tourism industry, witnessing unprecedented growth in usage amongst all generations. Being equally popular amongst Generation X, Generation Y and baby boomers, social media has evolved as a remarkable tourism destination marketing tool that not only allows consumers to search their desired travel related information but also allows them to socialize with other travelers of common interest to plan their next trips. Miguens, Baggio and Costa (2008) add that influenced by social networking, social media is forcing the buyers as well as the suppliers to attach more value to opinions, reviews and referrals of fellow travelers.

Interestingly, Mukherjee and Nagabhushanam (2016) further state that social media is particularly important in **influencing consumer behavior** in the pre-travel stage which is essentially why tourism organizations such as airline companies, hotels and travel agencies have embraced social media as a mainstream marketing tool, both as a means of promoting tourism products and destinations. The authors also delineate upon the branding aspect in the tourism industry and conclude through literary evidence that the element of trust in the social media brand tends to exert intense influence in developing brand loyalty. This trust in the social media brand is technically impacted by brand characteristics, company characteristics and customer characteristics. Moreover, Hudson and Thal (2013) also support the stance of social media impacting the consumer decision making process of tourists, primarily owing to increased sophistication in terms of how consumers engage with DMOs and brands of preference. Hence, tech savvy consumers are particularly benefitting from social media engagement with destination marketers in today’s rapidly changing digital environment.

Munar and Jacobsen's (2014) study aimed at expanding the construct of **information and knowledge sharing** over social media to encompass the overall sharing of the users' tourism experiences. The results indicated that the primary motivations for sharing their travel related information on social media were essentially altruistic and community-related. Social and emotional support ranked as top reasons for participating in online groups and sharing own travel related experiences with others would essentially help others chose their next travel spot. Many participants claimed that they basically contributed to online so as to inform and discourage people from using bad products, and to provide advice on practical travel related matters such as selection of tourist locations. Hence, sharing of personal travel experiences over social media appeared as "valuable expression of sociability", although having a low level of relevance as information source for holiday makers. At the same time, such information adds to the reputation of the travel site since tourists' associate social and emotional support for them.

2.1.1 User-Generated Content and Social Media

According to Akehurst (2009), due to the increasing bargaining power of the consumers in lieu of the proliferation of the internet and digital technologies, there is dire need for the tourism organizations to improve market intelligence to be able to facilitate timely decision making on the part of the consumers. The use of various platforms that support User-Generated Content (UGC), such as blogs, peer-to-peer web applications and virtual communities, has become a common phenomenon in the age of social media. Ye et al (2010) add that social media essentially empowers the internet users to engage in two-way information communication in travel and tourism, generating a huge amount of user generated content on travel destinations, services, hotels and restaurants, and so forth. At the same time, UCG has radically transformed the way consumers search, gather, share and consumer travel related data, while largely influencing the consumers' decision-making process (Lu and Stepchenkova, 2015).

Stankov et al (2010) discuss that the National Tourism Organizations (NTOs) have more recently acknowledged the role of online communities in creating awareness about various tourist destinations amongst both national and international markets. Aiming to capitalize upon the power of social media, these NTOs are particularly encouraging local residents, local businesses and tourists to narrate their destination's story over various social networking platforms. At the same time, these NTOs are also actively engaged in creating and sponsoring their own pages on social networking sites, run polls, and also create applications that users can include on their own pages. Therefore, social media has emerged as an important technological intervention that tends to significantly influence the tourism industry, particularly through the UGC that encompasses media impressions created by users.

However, Cox et al (2009) state that while social networking sites containing user-generated content play a major role in shaping up the search and travel behavior of consumers, they are still not yet considered as credible and trustworthy as the traditional sources of travel information such as "*government sponsored tourism websites*". It is an established fact that UGC tends to act as an "additional source of information" for the tourists to consider as an important part of their search process. Nevertheless, the more recent studies highlight an important role of the UGC. For instance, Ayeh, Au and Law (2013) further add that the subjective nature of UGC is one of the leading causes of credibility concerns since in an online setting, the interpretation of opinions and reviews is essentially subjective. Also, UGC is quite vulnerable to manipulation and abuse. For instance, fake traveler identities and posting dishonest comments and reviews regarding a particular topic can seriously undermine the credibility of UGC in the online social media environment. Consequently, as found by several studies, different reviews further impact sales and consumer purchase decision, given the valence of sentiments (Blal and Sturman, 2014; Tang et al., 2014; Floyd et al., 2014; Kostrya et al., 2015).

Interestingly, UGC varies greatly across different social media platforms. Smith, Fischer and Yongjian (2012) discuss that that is wide assortment of brand related UGC across different social media networks. For instance, a YouTube video is completely different from a Facebook wall post, which is again different from a Twitter tweet. The key to success for marketers then lies in better understanding of these differences which can deploy these social media for co-creation of their brands. At the same time, UGC is considered one of the strongest forms of electronic word-of-mouth (eWOM) marketing, since it portrays the level of satisfaction of the user (Lu and Stepchenoka, 2015). Particularly in the case of travel and tourism industry, UGC provides all the important information and personal experiences to shape up the travel decisions and plans of other potential travelers (Lu and Stepchenkova, 2015). Moore et al. (2012) and Garnefeld et al. (2011) also noted that level of personal experience share can impact the travelers emotions and perceptions. Moreover, Ye et al (2010) discuss that online UGC has important implications for the managers in terms of brand building as well as product/service development and quality assurance. Given the fact that in the tourism industry positive reviews regarding particular tourist destinations tend to improve the overall perception of the destination, sophisticated technologies such as the social media play a significant role in influencing potential consumers. Therefore, Stevenson and Hamill (2012) presents the importance of social media monitoring by examining the comments on Facebook about ten top ranked destinations. Bekk et al (2016) further add that the growing phenomenon of sharing online travel reviews is increasingly being considered as an interactive marketing tool for the travel and tourism industry since it carries immense economic potential in terms of using information from the customer as opposed to using information about the customer. Moreover, recent study of Duan et al. (2015) highlights that the user reviews lead to user evaluation and initiate content generation, which then if measured correctly, can improve experience of service quality and performance.

Also, since the Generation Y is more technology-driven, the tourism industry can significantly benefit from extracting important information from social media to further align its marketing strategy to better cater to the consumer requirements (Gretzel et al, 2000). According to Lu and Stepchenkova (2015), Generation Y is the largest consumer group in the US, where around eight out of ten respondents acknowledge the influential nature of UGC to impact their purchase decisions. Yoo and Gretzel (2008) further add that since user generated content tends to support higher credibility, relevance and empathy as compared to the marketer-generated content, consumers' trust in the travel company is particularly enhanced, while it also results in longer and deeper relationship between the consumer and the service provider.

Also, this traveler information shared across various social networking platforms can act as a predictor of future events. According to Gretzel and Yoo (2008), travel planning and decision making is largely influenced by the UGC relevant to various tourist destinations. Since consumers tend to rely on reviews from other travelers while choosing their next tourist spot, their shared experiences can act as a major factor that the tourism organizations can use in term of tailoring their future offers as well as predict future traveler behavior (Sparks et al, 2013).

Milano et al (2011) further add that social media has evolved as a powerful influence on the popularity of tourism websites. It is not only tends to improve the attractiveness of various tourist destinations but also helps the tourism based companies to effectively channel their resources towards the most desirable products and services. The study conducted by the authors further elaborates that UGC shared on Facebook and Twitter have been noted to impact the number of visits to a tourism website, thereby acting as an important influence on the travel decision processes of other potential tourists and travelers. Moreover, one of the major business implications of the study relates to the fact that it is particularly difficult to motivate consumers to post their travel reviews since the motivations are intrinsic to a great extent. Hence, tourism organizations may want to invest considerable time and effort into decreasing the barriers to writing and encouraging travelers to share their travel experiences over social media more openly (Gretzel et al, 2007).

Hence, the primary purpose of this research remains to delve into the dynamics of using twitter as a social media platform to enhance the attractiveness of iconic travel destinations, and assess its feasibility and success. It remains critical to encourage comprehensive investigation into the impact and influence of social media, as an integral component of tourism destination marketing strategy, on virtually all aspects of tourism industry including key stakeholders such as local communities, and the economy at large. The following section defines and discusses the comprehensive approach to deriving data for the purpose of this research and the data analysis process. Text mining and sentiment analysis are the two frameworks that will be used, and their context justification and relevance to the realm of social media are discussed as under.

2.2 TEXT MINING AND WEB MINING

Text mining is a data derivation technique that attempts to glean meaningful information from natural language text, which is otherwise unstructured and amorphous, and therefore very difficult to analyze algorithmically (Calheiros, Moro, and Rita, 2017). It is used as a reliable tool to unveil the hidden patterns in textual contents such as comments and reviews, through typical tasks including (but not limited to) text categorization, text clustering, and sentiment analysis. According to Radovanovic and Ivanovic (2008), text mining is an innovative aspect of computer science that encourages establishing strong connections with the natural language processing, as well as data mining and machine learning, while facilitating information retrieval and knowledge management. It is essentially a technique to analyze large quantities of natural language text, detecting lexical as well as linguistic usage patterns to derive useful information. Hence, it is about identifying patterns in text. Within the larger domain of text mining, there are a number of information processing tools available including data mining, web mining, computational linguistics and natural language processing, information retrieval, and statistics (Berry and Kogan, 2010).

Text mining is particularly a tough task, primarily because of the ambiguous nature of language, whereby context is needed to clarify. While the same word can mean different things (homographs) and hence support different connotations, the different word can also mean the same thing (synonyms). Similarly, language is subtle. Moreover, word extraction usually results in huge number of dimensions which makes room for numerous new fields whereby each field typically supports low information content. Finally, abbreviations, misspellings and spelling variants all add to the complexity of analyzing language (Berry and Kogan, 2010; Srivastava and Sahami, 2009).

Pang and Lee (2008) suggest that text mining is gaining immense momentum in combination with sentiment analysis, particularly on social networking platforms such as Twitter and Facebook. This combination is significantly important while predicting stock market trends, forecasting churn outs, and expecting customer influence. Sentiment analysis, as described by Radovanovic and Ivanovic (2008), deals with analyzing opinions found in documents, classifying text in terms of positivity and negativity of expressed opinion at the very basic level. Sentiments, precisely defined, are opinions and personal judgments about something specific. They are expressions of emotion or opinion or attitude, and are thought of as the prediction of people’s personal interests (Calheiros, Moro, and Rita, 2017). Hence, sentiment analysis has wide applications in various business functions and aspects of everyday life as suggested by Mukherjee (2012). It has wide applications in consumer market for product reviews, in marketing for understanding consumer trends and attitudes, and in social media for seeking the general consumer opinions regarding what is “in” and “trending”.

However, it remains imperative to note that it is a highly complex task to extract knowledge from unstructured text through text mining. While it may be feasible to do it for a few tweets or for a few reviews, it is practically not feasible to do it for thousands or millions of tweets. Hence, text mining is particularly useful tool to semi-automatically extract the most important topics latent in text (Calheiros, Moro, and Rita, 2017).

Interestingly, the possibility of applications of digital data are vast. The traces left behind by Internet users are called digital footprints and they can be used as source of data to perform a wide variety of research (Santos, Rita, and Guerreiro, 2017). With recent technological infrastructure, the capacity to collect data and submit to a second degree data mining operation, widens the spectrum of digital footprints eligible to generate data that may be used for a given analysis. Hence, the data may range from uploaded photos to posted reviews or clicked links. Moreover, this data can be collected through almost any website and used as a reliable source for marketing and research (O'Neill et al, 2006).

A classic example of application of data mining is that of the study conducted by Moro, Rite and Coelho (2017), which models TripAdvisor score for 21 hotels situated in the Strip, Las Vegas, using over 500 reviews published. Conducting a sensitivity analysis over the constructed model, the study concluded that the reviews provided on TripAdvisor play a critical role in influencing the granted scores. Moro, Rita and Vala (2016) further suggest that data mining is a useful approach to study and predict social media performance metrics of posts made on various social networking platforms. Studying Facebook posts published on official Facebook page of a cosmetic brand, the study concluded that a status post tends to capture twice the attention of the posts that carry links, photos or videos.

Girardin et al (2008a) categorized digital footprints into two different types. One, Active Footprints represent those where a user has added some content to the website such as uploaded photos or written reviews. The other form of digital footprints are designated as Passive Footprints where a user had some interaction with the website but did not add any content, such as, a user browsing products on a website. Individual product recommendations on websites are generated through this type of data.

Additionally, current mobile technology allows for users to take advantage of the functionalities of their devices and perform a variety of tasks in an easy and convenient manner. For instance, a user would take a photo on his phone and immediately share it with friends on social media. Exploring this premise Girardin et al (2008b), uncovered further uses to digital footprints by distinguishing between implicit and explicit interactions with infrastructures. Implicit interactions are considered involuntary actions such as the generation of a location log through a mobile device communication with network and explicit interactions are voluntary actions such as the publication of content.

Text mining social media, according to Gemar and Jimenez-Quintero (2015), is a highly effective strategic tool used for competitive analysis. Santos, Rita, and Guerreiro (2017: 9) add that the ultimate rationale behind text mining remains that "*each document can be represented by the frequency of its terms*." While social media is increasingly being used by companies to communicate and interact with consumers, the use of text mining tools has largely facilitated organizations to assess their own standing in relation to their competitors and, thereby, improve their competitive strategies. Bach et al (2013) also suggest that data mining applications have gained significant importance in the tourism industry. Data mining applications based on keywords analysis allow tourism based organizations to use travelers' shared experiences over social media to further streamline their strategies and also assess the popularity of various tourist destinations. Sentiment analysis can then be conducted through two approaches to classify sentiments: machine learning and lexicon-based (Santos, Rita, and Guerreiro, 2017). Furthermore, these applications of data mining cover a wide range of fields including forecasting, tourism management, personalization, machine learning techniques such as multi-agent systems and particle swarm optimization, and tourism systems such as recommendation systems, etc (Santos, Rita, and Guerreiro, 2017).

The tourism and hospitality industry in particular has benefitted from the application of data mining procedures to study various aspects of travel and tourism. Moro, Rita and Coelho (2017) discuss that data mining has been used for a wide variety of applications such as to study the influence of a hotel's marketing strategies on behavior of potential customers. In addition, Stevenson and Hamill (2012), gave an outline of the advantages of web-based social networking monitoring and investigated the scope of web crawler that was accessible at the time. They could produce data, such as number of mentions, the channel sources, key influencers and brand sentiments, that was derived from their examination of social media content related to the world's top 10 destinations. they could produce data, such as number of mentions, the channel sources, key influencers and brand sentiments. However, privacy of the Facebook content limited their research.

2.3 PREVIOUS STUDIES ON TWITTER AND TRIPADVISOR

With the advent of social media, communications have drastically changed and consequently marketing or tourism destinations and businesses has also revolutionized. According to Sotiriadis and Zyl (2013), Twitter has emerged as a widely used tool for eWOM and referral/recommendations strategy, increasingly used in tourism services with high rates of involvement. This tool has essentially provided an integrated communications marketing platforms for tourism service providers as well as tourists.

Jenders, Kasneci and Naumann (2013) conducted a study aimed at analyzing and predicting viral tweets over Twitter. Sentiment analysis was conducted to evaluate whether tweets with positive sentiments underwent a different diffusion process as compared to tweets with negative sentiments. The findings of the study suggest that the most important factors that impact the tendency of a tweet becoming viral primarily include the number of followers, the tweet length, hashtags and mentions. Boyd et al. (2010) found that information diffusion and engagement of the users serve as a main cause of the retweets. Similarly, a study on retweet dynamics of retweetability of Suh et al. (2010) found that tweets with URLs and hashtags are likely to be retweeted more, as well as strong relationship between a number of followers the high probability of the retweet was also identified. Moreover, despite that Sobel et al. (2009) and Berger et al. (2010) suggest that Twitter is also used as promotional tool, and therefore results on high numbers of shares of positive content, Pfitzner et al. (2012) argues that only high emotional divergent comments are more retweeted. Hence, several researches conclude that twitter is has enormous impact given the large cascade of information exposure (Bakshy et al., 2011; Hong, 2011; Zhou, 2011).

A similar study conducted on semantic sentiment analysis of twitter and found that Twitter has evolved as a fast and effective way of monitoring the public feelings towards any given brand or an organization (Saif, He and Alani, 2012). The authors discuss that Twitter has evolved as a "gold mine for organizations" facilitating them in monitoring their reputation and brands by deploying sentiment analysis of the tweets made by the public about them. They further argue that adding the semantic feature for classifying positive or negative sentiments in the tweets is a major facilitator in assessing the impact of tweets and also improves the accuracy of sentiment analysis. Similarly, Stepchenkova, Kirilenko and Kim (2013) studied the use of twitter for branding Florida to attract tourists and promote Florida as an attractive tourist destination. Acknowledging the fact that social networking sites have become a dominant source of word-of-mouth referral for travel experiences, it remains imperative to assess the suitability of using publically accessible data to gain tourist insights and perceptions about a particular travel destination. The key findings suggest that the use of adjectives to describe Florida as a travel destination tend to determine the impact of tweets on other tourists' perception, and the word "amazing" remained the most influential, followed by "beautiful" and "magnificent". Moreover, tweets that depicted positive sentiments generally expressed happy attitude and thereby added to the attractiveness of Florida for other travelers.

From the perspective of DMO's, social media has again emerged as a lucrative tool to enhance and promote their brands and to reach out to their potential visitors. Aiming to study the use of social media amongst 12 Balkan countries' DMOs, Bayram and Arici (2013) concluded that nine maintained official Twitter accounts, and one of the measures of Twitter user's effectiveness as used by the study, is the number of followers attracted. The results suggested that DMOs had between 260 and 21,799 followers, confirming the growing importance of social media by DMOs. Nguyen and Wang (2012) also conducted a similar study, seeking to provide insights into the application of social media in the tourism industry, specifically from the perspective of DMOs, keeping Visit Sweden as case in point. While the study relied on other social media as well, part of the study records use of Twitter as a mainstream tool. The study concluded that DMOs need to streamline their focus in terms of emphasizing the wide participation in online marketing as well as social media activities to be able to maximize their benefits. Integration of online marketing and social media activities with traditional marketing approaches have become the need of the hour for today's DMOs.

Regional studies also foster the importance of social media as an important tool in terms of enhancing the attractiveness of tourism destinations. For example, according to Sevin (2013) five prominent American destinations - San Francisco, Texas, Illinois, Idaho, and Milwaukee – effectively use Twitter for increasing their tourist attractiveness. While the primary objective of the study was to explore and understand the overall trends and usage patterns of microblogging with reference to the "social media ecology" and "place branding", it concluded that destination marketing projects predominantly used Twitter to share important destination related information; these primarily included events such as festivals, fairs, and concerts.

It further remains important to note that the attitude of DMOs towards the use of social media also remains of utmost significance. According to Hassan (2013), social media has played a significantly important role in positioning Egypt, particularly after the 25 January revolution. 180 local DMOs were involved in the study and it was found that the majority were using Twitter as a mainstream social medium. However, there were numerous loopholes identified in terms of optimal utilization of social media by DMOs for positioning a new image of Egypt and their products. For instance, while the hotel industry has realized the significance of social media in panting an important image of Egypt to the outside world, other sectors of the tourism industry, such as the travel agencies, have failed to follow the trend. Similarly, there is a general lack of knowledge regarding the use of social media technology in the region, while there is also general fear of not being able to completely control the information dissemination. Similarly, Antoniadis, Vrana and Zafiropoulos (2014) add that DMOs are increasingly using social media, particularly twitter, to share destination-related information and knowledge, adopting clearly thought out strategies and initiatives to not only increase destination related awareness, but also to achieve influence and to promote their country's image. Using the convenience of marketing as offered by Twitter, DMOs capitalize on building relationships and on expanding online branding opportunities to provide information as well as encourage the country's destination image. Hence, as more and more DMOs embrace social media to reach out to their customers, Twitter has evolved as a plausible opportunity to target global audience "at a relatively low cost", capitalizing upon both promotion and branding, and, therefore, can achieve significant insight into consumer behavior using direct observation through Twitter.

However, in another study Antoniadis, Zafiropoulos and Vrana (2015) suggest that while Twitter remains a widely used microblogging platform for sharing travel related information and experiences, it only serves as a public notice board for the DMOs as well as the followers. Needless to say, it still remains an important tool in sharing and annotating the travel stories and experiences of consumers, as well as an important tool for DMOs in promoting their countries' images.

TripAdvisor has more recently emerged as a widely used reference point for international travelers and tourists. It is essentially a content-based recommender system for producing recommendations for various tourist destinations across the globe. Being a web-based application, TripAdvisor posts autonomously generated information by users, including post reviews, comments and ratings on a given destination, a hotel or an attraction, for other tourists and travelers. The rankings assigned by TripAdvisor are essentially based on the quality, recency and quantity of reviews received by a given business from the travelers, to produce ranking for the business and act as a reliable recommender system for other tourists. Hence, the popularity of a given attraction, according to TripAdvisor, is determined through more reviews, good reviews, and recent reviews. With the improved Popularity Ranking Algorithm in 2016, TripAdvisor's results are now deemed more accurate, representative, and reliable (TripAdvisor, 2016).

Miguens, Baggio and Costa (2008) conducted a study on the hotels of Lisbon, Portugal, using data from TripAdvisor; all hotels listed on TripAdvisor were included in the study. Lisbon as a destination was analyzed to compare different ways of rating hotels in the TripAdvisor travel community, capitalizing upon the major strength of TripAdvisor in terms of using user-generated content. Hence, the study used TripAdvisor as a platform to gain deeper insights and understanding about the way tourist, hotels and residents of a given destination respond towards tourism in general.

Crotts, Mason and Davis (2009) conducted another research aimed at measuring guest satisfaction and delight for tourism and hospitality industry. The study highlights the fact that guest satisfaction is a crucial factor to add to the popularity of any given tourist destination. Hence, while the purpose of the study remained to introduce a new technique of measuring guest satisfaction, the study used TripAdvisor as a source of guest comments, as a reliable measure for popularity of tourist attractions.

Similarly, Wang, Chan and Ngai (2012) conducted a study on the reliability and applicability of demographic recommender systems for tourist attractions, taking TripAdvisor as the case in point. Acknowledging the fact that majority of existing recommender systems rely on historical rating information to rank various tourist destinations, the authors aimed at establishing a new recommender system that did not need any extra knowledge or history of ratings to make recommendations for attractions. Various machine learning methods were used to serve the purpose, while the focus remained attractions ranked top as per TripAdvisor. Let alone the findings, the study itself validates the use of TripAdvisor as a proxy measure for ranking tourist attractions.

Hence, given the credibility attached to TripAdvisor as a benchmark for ranking the popularity of tourist destinations, this study uses TripAdvisor's ranking as a measure of a given tourist location popularity. It remains imperative to highlight that the researcher was unable to find official reliable data on the number of visitors for the tourist locations chosen for the purpose of this study. Therefore, counting the reliability and diversity of its ranking system, TripAdvisor has been used as a proxy measure.

3 THEORETICAL FRAMEWORK

3.1 RESEARCH PROBLEM AND OBJECTIVES

Usage of Twitter, Facebook, email and other online social media significantly facilitate the process of a collective awareness around the events. (Sparrow et al. 2011). As a result, such collective awareness can lead to the short-term impact of the global awareness about the particular event/twitter post (Gloor et. al, 2016). In addition, as suggested by Pike (2016), individuals may form an opinion and image of destination both prior and after visitation, which emphasizes the importance of studying users' perceptions regarding locations via social media tools such as Twitter.

A number of studies performed research Twitter posts and their applications in business and marketing. However, as it was described in the earlier chapter, most of the researches have been limited to either single location or to the countries tourism twitter account. For instance, study on by Stepchenkova, Kirilenko and Kim (2013) examines the visitors' positive perceptions on particular destinations in Florida. On the other hand, Antoniadis, Zafiropoulos and Vrana (2015) stated in their research of 37 European countries tourism accounts, that they can only serve as public notice board for announcements, and generate no further discussion by their followers.

However, as to the best knowledge of the author of this study, there are only few researches that compare the difference in consumer perception towards popular and less popular locations. For instance, recently, the study on analysis of attraction features of Tourism in Shenzhen by Gu et al (2016) suggests that tourism destination images are perceived differently than what is on official marketing. Even though the research is mainly oriented on evaluation of the check-in times through data-mining, it still leads to the research opportunity for this dissertation, to study difference between popular locations, especially as was suggested by Pike (2016) that change in perception on destination image occurs very slowly.

As mentioned earlier, several studies have been performed to evaluate countries' or single city oriented tourism performance by analyzing Twitter posts (Nguyen et al., 2012; Stepchenkova et al., 2013; Antoniadis et al.,2014). Nonetheless, other than these studies, research investigating the perception of the social media users on different popular locations through sentiment analysis has been neglected in the tourism and marketing literature and therefore represents an important research gap.

Therefore, in order to conclude the broader findings, considering limitations of the previous studies, the comparisons of several different locations regardless of the sentiment rank seemed to be the most relevant approach for this research.

Hence, the main goal of this dissertation is to examine how twitter posts contribute to the popularity of the touristic iconic destination and the difference between sentiments of the iconic attractions in general using text mining tools.

Table 1 below shows the most important findings of the previous researches, grouped by topics, as well as their limitations for the consideration of further resolution in this dissertation.

Table 1: Main Contributions to the Literature on Twitter Posts, and Limitations that contribute to the purpose of this Dissertation.

Research Topics	Authors	Main Contributions	Limitations

Twitter Performance as Tourism indicators	Stepchenkova et al. (2013)	• Research examines how the particular destination, such as Florida, is viewed by visitors and residents, and various descriptors that characterize the affective state of positive surprise associated with Florida.	• The authors state that only one affective state of a positive surprise was studied and limited to only one location. As a result, they suggest that future research will need to expand the number of key words or destination attributes.
	Antoniadis et al. (2014)	• The findings propose that Twitter performance is associated with countries' actual tourism performance as it is measured by three official Tourism indicators; • Authors conclude that Twitter does not fail to provide information and possibly to promote countries' Destination Image.	• A main limitation of this study is that the content of the tweets was not considered, and research was limited to quantitative rather actual message study. Therefore, authors suggest to record Twitter content and compare possible difference regarding the content in Twitter, by different countries.
	Nguyen, V. H. and Wang, Z. (2012)	• Authors investigated the approach of practicing online marketing with social media as one of digital marketing communication options to build destination brands and engage audience to reach potential visitors. • The study evaluated the crucial role of online marketing and social media (including Twitter) in destination marketing strategies and managements.	• One of the limitations that affected the measurements of main conclusions is that the empirical data gathered from single case study VisitSweden, as well as it was broad research across many social networks, giving narrow findings on Twitter. Hence, it gave limited possibilities for generalization of the practice and benefits of online marketing and social media to apply for other DMOs.
	Stevenson and Hamill (2012)	• Authors examined city destinations using social media monitoring (comments, mentions). They presented implications for travel, tourism and hospitality businesses using different techniques.	• Due to limited availability of text mining tools, only high level Bran Sentiment was identified, without giving specific attention quantitative study and content. Thus, it was suggested to study relationship between business performance and social media mentions, i.e. how an increase in mentions of a destination translated to an increase in visitor members.

Sentiment Analysis on Tweeter	Jenders, Kasneci and Naumann (2013)	• The study addressed an essential findings on how the information is spread on Twitter, prediction model of tweet virality and its further impact.	• Research was rather generic and informative than specific to the travel topic, therefore further narrowed subjects could be studied using the same findings
	Bakshy et al. (2011)	• Quantification of influence on Twitter was presented in the study, highlighting that number of followers and retweet can play an important role.	• One of the proposal to the future studies was to examine eWOM by adopting more precise metrics of influence, and exploiting more generic and ordinary topics.
	Saif et al. (2012)	• The study concentrated on research of semantics on Twitter through Semantic Analysis. Authors measure correlations of the presented concepts with negative/positive sentiments.	
	Philander et al. (2016)	• Using integrated resort property in Las Vegas, authors study customers attitudes/perceptions through sentiment analysis of Twitter Data with cross-check to the TripAdvisor ranking.	• The study was focused on a certain category of hospitality firm in 1 city, therefore findings may not bring same results for different industries or other locations.

Source: Author's own elaboration

Thus, in this study, the author will try to close the existing gaps by conducting broad analysis of several popular destinations inclusive of twitter elements. Also, the research will examine if emotions of potential tourists differentiate as per level of popularity of destinations, given the tweet diffusion.

Ultimately, this dissertation will attempt to interpret the marketing implications by creating a support for future researches regarding how positive and negative posts on twitter may be used for branding the destinations' image and impact the tourism popularity. Thus the main research question is identified as following:

Are the popular and less popular destinations perceived the same way by the twitter users?

3.2 RESEARCH HYPOTHESES

As it was stated by Santana (2001), that in tourism, ***perception*** is reality when it comes to decision making, the main objective of the hypothesis that are set forth to define weather the popular and less popular locations are perceived different for the Twitter users. Henceforth, firstly author will try to examine level of exposure of popular and less popular locations given the tweet posts which is followed by analysis of affection of the users related to such posts.

Exposure. As it was concluded by Suh et al. (2010) in Twitter, the major tool that leads to the spread of the information and most powerful mechanism to disseminate the data is the retweeting. Similarly, Boyd et al. (2010) adds that beside the information diffusion, retweets can be also used to engage other users. In addition, retweets have been seen to be reliable in providing tractable information for diffusion model (Hong, 2011; Zhou, 2011). Therefore, considering that retweets cause the topic exposure and previous research found that higher ranked locations lead to a very large number of online conversations (Stevenson and Hamill, 2012), following first hypothesis is identified:

H1: In Twitter, tweets about more popular touristic attractions, generate a higher number of retweets than less popular ones.

As mentioned earlier, it is important to determine whether the Tweeting activity associated with a certain location has an impact on its popularity and if similar patterns are observed across groups between the same city or with comparable popularity rank. One of the issues that could impact the results is the difference between the average number of followers, across the different locations. According to a study by Jenders, Kasneci and Naumann (2013), the number of followers were among the variables that most contributed to a tweet being shared. Similarly, Bakshy et al. (2011) found that the wide exposure of the tweets is the most common for those users who have many followers. Hence, a location could be more popular due to the fact that it received higher levels of exposure. Besides testing for this effect, the second hypothesis is meant to understand if additional measures need to be taken in order to normalize the retweet activity by the number of viewers that were directly exposed to that tweet. In other words, if the role of followers has a strong effect on retweets. Leading to the following hypothesis:

H2: In Twitter, for more popular locations, the number of followers has a stronger significant impact on the number of retweets than for less popular locations.

Affection. Previous research highlighted that perception on actual experience and associated emotions can be influenced by the scope of shared experience (Moore, et al. 2012; Garnefeld et al. 2011). As per Russell's circumflex model of affect, such emotions can be explained by blend of differing degrees of two fundamental measurements, valence (pleasure measurement) and arousal (actuation measurement) (Russell, 1980). According to Kostrya et al. (2015), positive valence of online reviews, has strong effect on customer's choice and volume moderates its impact. On the other hand, Floyd et al. (2014), found that negative also had an effect on costumers' choice by dissuading potential customers from purchasing. Furthermore, Tang et al. (2014), found that neutral sentiment of online reviews has a mixed effect in sales. The authors defined neutral reviews as either mixed or indifferent with both sets of categories generating a positive and negative effect on sales respectively. This leads to conclusion that studying the affection of the Twitter users about popular and less popular location may derive an important implications. Additionally, a study by Saif et al. (2012), highlighted the potential of Twitter as a monitoring tool for any kind of organization when combined with the applications of semantic analysis. Moreover, research of Blal and Sturman (2014) demonstrated that the type of product generates different effects of eWOM on sales. The authors found that for upper scale hotels the valence of reviews drive sales while for economy the volume does. Therefore, given that there's a difference in valence of reviews for higher ranked hotels, the following hypothesis, posits that touristic locations in popular and less popular destinations differ in terms of their average sentiment score:

H3: In Twitter, tweets about more popular touristic locations, carry a higher sentiment score than less popular ones.

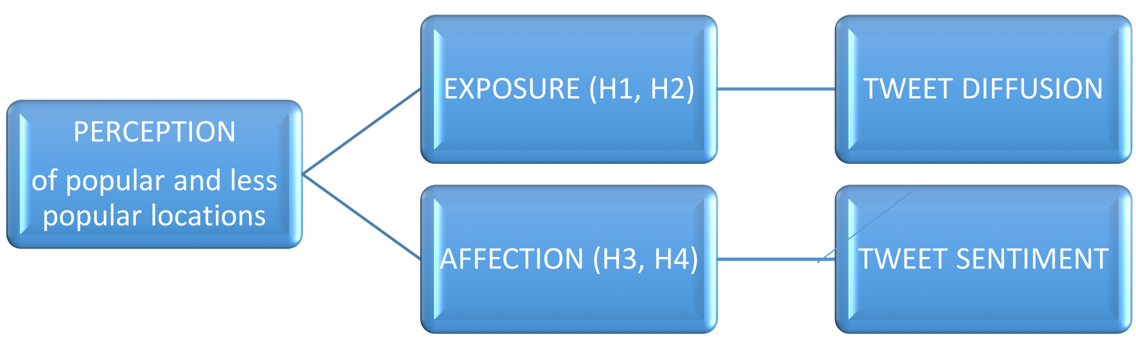
Another issue that may influence the retweet rate is the polarity of tweets. Previous literature has given evidence that social media users tend to follow a self-promoting role, therefore sharing more messages with a positive sentiment. (Sobel et al. 2009; Berger & Milkman, 2010). On the other hand, Pfitzner et al. (2012) found that both negative and positive tweets have more chance to be retweeted if they are with a high emotional diversity. Henceforth, our fourth hypothesis, is aimed at understanding which sentiments of posts are more like to be retweeted and potentially influence the popularity of a location. Furthermore, since research of Stepchenkova et al. (2013) suggested that the study was limited to the only one affective state of positive surprise and only one location, therefore in this dissertation more emphasis is given to negative posts to evaluate if they have a more immediate effect on popularity. Moreover, research of Cheng et al. (2011) also indicated that users are more likely to express negative sentiment about location, hence the final hypothesis of this research paper is as follows:

H4: In Twitter, negative posts about touristic attractions are the most retweeted.

3.3 Conceptual Model

As a result of evaluation of the previous literature, this dissertation builds from the need of the study whether the popular and less popular locations viewed the same way by potential visitors in twitter. Therefore, Figure 2 is constructed to fulfill the proposed gap in literature and the Marketing field.

Figure 2. Conceptual Model of the Current Research



Source: Author's own elaboration

The conceptual model suggests to first examine the popular and less popular locations in terms of their exposure in twitter audience by assessment of the tweet diffusion. Secondly, since research of the previous literature showed that the affection of the potential visitors could be monitored through tweet sentiments, conceptual model proposes to evaluate the sentiments within the popular and less popular destinations.

4 METHODOLOGY

4.1 DATA MINING AND PROCESSING TOOLS

Twitter

Twitter is a social networking platform that, like others of the same kind, has become an enabler of a microblogging service. Thus, the platform allows people to informally express their opinions on variety of matters by posting real time messages, which are designated as tweets. These messages are short, being restricted to 140 characters and can be written in an open format. Hence, they are prone to the incorporation emoticons as well as acronyms, which in turn, limit any filter for spelling mistakes.

The level of proliferation of a message in Twitter is dependent on a user's willingness to share it. In Twitter terms, this factor can be accounted for as the number of retweets which by definition, represents the number of times a particular message created by a single user, has been shared by other users

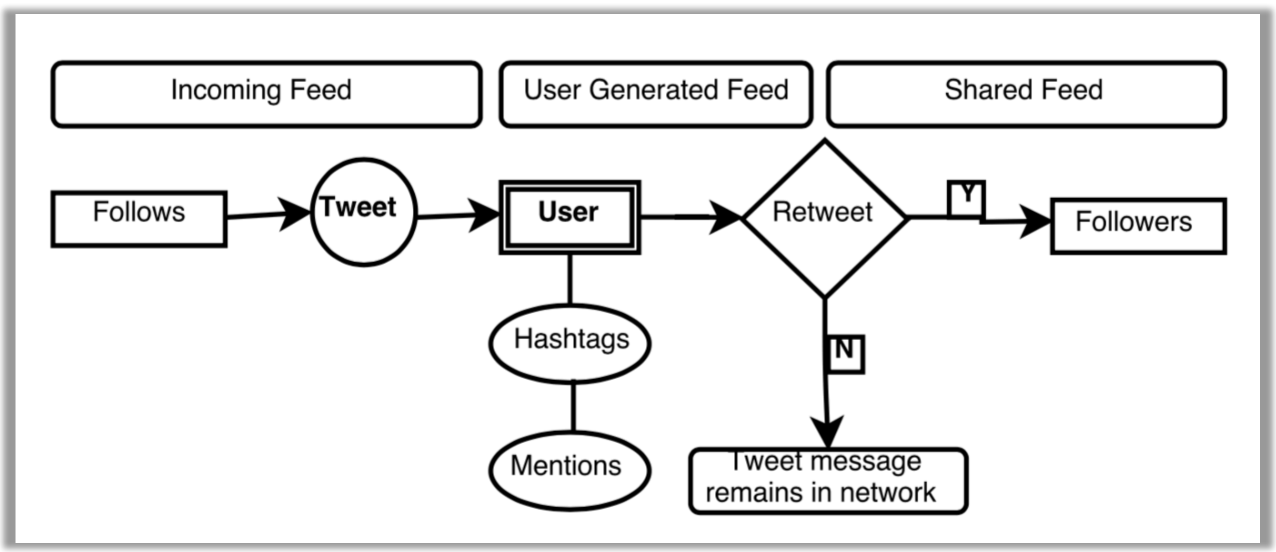
Moreover, there are other factors affecting this relationship, namely the following:

- **Content of the message:** Refers to the text, emojis, characters that together compile the tweet message.
- **Followers:** Represents the number of users, a single user has receiving his feed (contentit generates) and the number of retweets.

- **Following:** The number of users in network and the feed it receives.
- **Hashtags:** single word tokens which follow the hash symbol '#'. They are used to tag a tweet and can only be used by its author.
- **Mentions**: Referrals to use or places.

Figure 2 below showcases the dynamics of the twitter network from the perspective of a single user. The user receives twitter feed from the users he follows and the users who follow him, receive the feed that he generates. Hence, when a user opts to retweet a message, its contents are spread to his network of followers through his feed.

Figure 3: Twitter Network Process



Source: Author's own elaboration

Twitter Archiver

Data from twitter was extracted automatically using the web tool "twitter archiver". Through premium version, the data is automatically retrieved though Google sheets, where the tweets are organized in an ascending chronological order by date and segmented by user. For each tweet, different elements were collected such as the tweet text, number of followers, number of likes, number of retweets and the date since the user has been active. The twitter archiver premium retrieves data up to 10 days before the search query is set and continues to update it as long as the sheet is active.

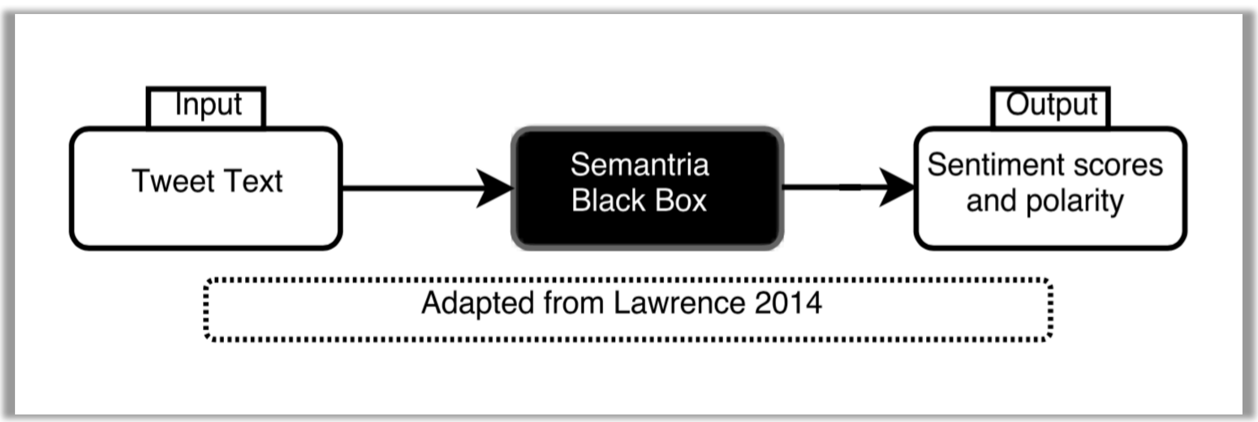
Semantria

For sentiment analysis, an automated tool was used. There is a variety of web services tools available such as AlchemyAPI, Lymbix and etc. However, Semantria was chosen primarily due to two reasons. One has to do with the free trial version setup, which compared to other tools allows the user to perform more transactions. These are run on a single line of text, in this case a tweet.

The second reason is related to the reliability, according to several studies (Gao, Hao, & Fu, 2015; Peisenieks & Skadiņš, 2014; Serrano-Guerrero, Olivas, Romero, & Herrera-Viedma, 2015); it was one of the tools that produced more robust results. Semantria is a Web service platform, developed by Lexalytics that based on machine learning techniques classifies sentiments as either neutral, positive or negative. Semantria is also capable of analyzing, negations, acronyms and emoticons.

Lawrence (2014) describes Semantria works as a blackbox, meaning it only allows the user to input data and receive its output. Figure 3 below indicates how the process works:

Figure 4: Flow Semantria Process



Source: Author's own elaboration

4.2 Data Collection and Processing

Initially the prime locations for this research as chosen as top popular destinations of the European Union, such as Spain and France. However, after initializing the twitter searches for iconic destination, it was concluded that the language of search creates a barrier in generating enough tweets. Therefore, it was decided to base the search in English speaking countries.

The data on Twitter can be accessed by Application Programming Interface (Horn, 2010). The HTTP serves as a base for the API and it sends requests to retrieve data from Twitter. There are 3 types of APIs that allow for Twitter data retrieval. The Search API allows for queries against recent Tweets and it will find only Tweets that are no more than 7 days old and correspond to the used query (Twitter, 2013e). The tool Twitter Archiver was a form of automating this step. The following sections describe the processes undertaken until the final database was constructed.

Before starting the tweet collecting process, several steps had to be carried out in order to ensure that the data retrieved was in accordance with object of the study. The collection tool, Twitter Archiver, had to be setup for input and output of data. The input was the data from twitter, which required a user account to access and the output was stored on Google sheets, which could be accessed anytime by the user via Gmail account.

Having completed the configuration process, the next step revolved around setting up the appropriate filters to narrow the data to what was desired, in this case, tweets referencing specific touristic locations.

In order to automate the data collection process, Twitter Archiver required defining a twitter search rule. In other words, selecting the criteria that its search engine would use for the automated extraction of tweets. The parameters allow the use of specific words, phrases, language or even the location where the tweet was posted. As explained above, the information posted on a tweet, may take a wide variety of forms and given the fact that Semantria, the tool chosen to run the semantic analysis was developed in English, retrieving data originally posted in English, could enhance the accuracy and robustness of the results.

Therefore, two search rule restrictions were set:

- Constant for all search queries, which was the language setting to English
- Variable which was a hashtag of the location.

During this step, we selected a variety of locations in different countries in order to understand the data disposition and which locations could be feasible for the purpose of the study.

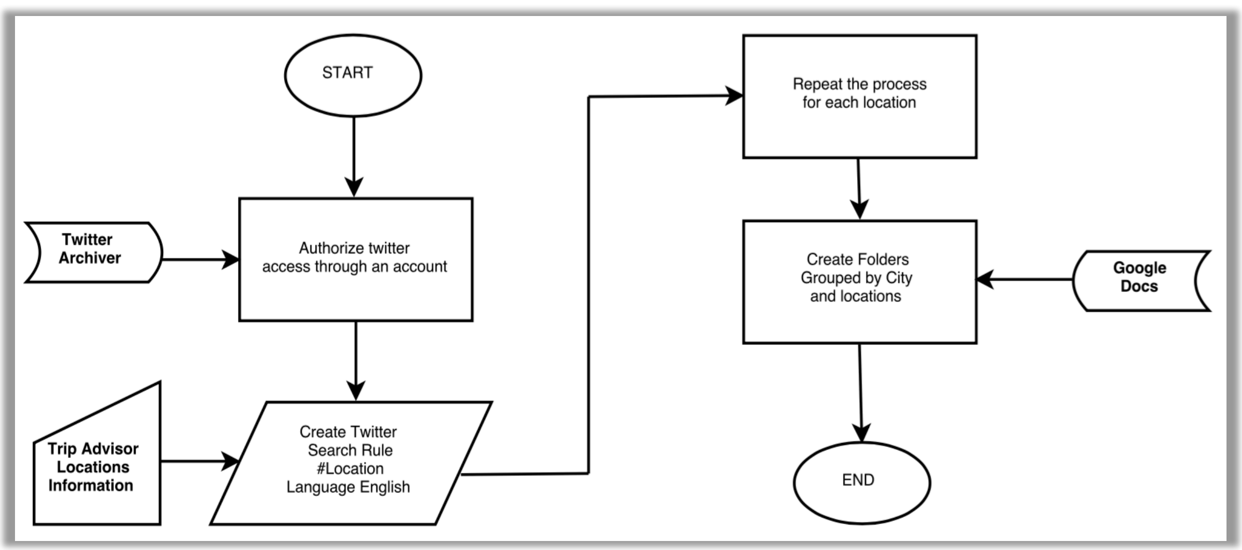
Location Selection

In order to select the locations, ranking information from TripAdvisor (2017a) was used as a benchmark for popularity. Top cities from each continent were considered and then a variety of touristic locations ranging in popularity from each city chosen to complete the search query. The data was cross-referenced with Global destinations report of the MasterCard (2016) and the cities were chosen based on their overnight visitors ranking. Given the characteristics of the study, urban cities were preferred, as they tend to have better infrastructure in regards to access and a wider variety of landmarks.

Therefore, the city selection was done as follows:

- Europe: London, Paris, Rome, Crete, Barcelona.
- America: New York, Las Vegas, Rio de Janeiro, Buenos Aires, Sao Paulo
- Asia: Bangkok, Singapore, Kuala Lumpur, Tokyo, Seoul, Hong Kong.

Figure 5: Flow Chart of Pre-Process Steps for Tweet Collection



Source: Author's own elaboration

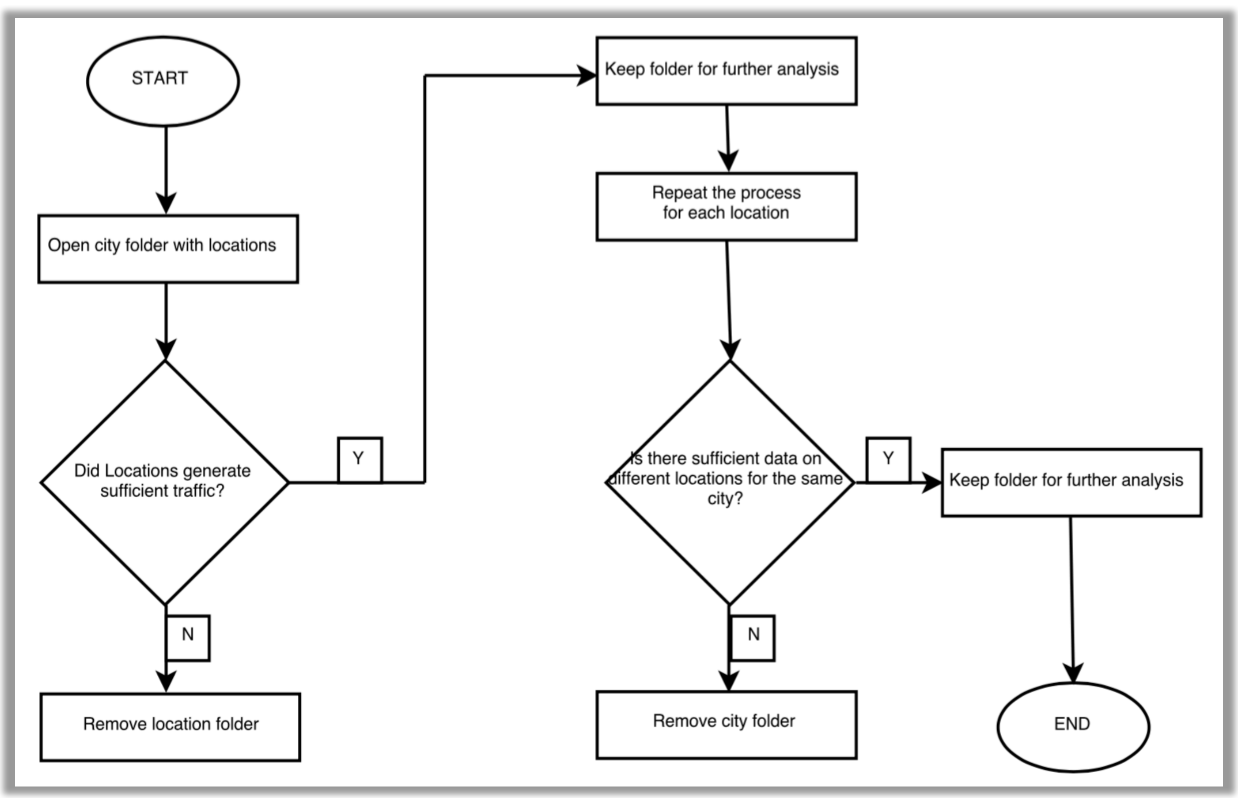
Having understood the appropriate parameters for the data collection process, it was necessary to ensure that the information collected was suitable to elaborate a data set for the study. The twitter Archiver premium retrieves data up to 10 days before the search query is set and continues to update it as long as the sheet is active. The search query was setup as a hashtag (#) of know popular locations in cities all around the world and for the reason explained above, the language parameter was set to English.

Having completed the step above, we analyzed our initial data, and came across the following limitations:

1. The English language parameter limits the number and variety of tweets.
2. There was a large discrepancy in terms of sample size between locations. Locations from non-Anglophone countries retrieve too few tweets.

Furthermore, in light of these limitations, it seemed more convenient to choose cities from an Anglophone market. Therefore, touristic locations from America and the United Kingdom were chosen, due to the size as well diversity of their tourism market resulting in a new tweet collection process being arranged with locations in London and New York. Figure 5 below illustrates the flowchart of the tweet collection process:

Figure 6: Flow Chart of tweet collection process



Source: Author's own elaboration

The collection period was established to be from the 28th of April 2017 until the 28th May 2017. The period chosen in an effort to reduce the bias arising from external events which were coincidental to the data collection period. The events in reference were the Easter holidays, or the terror attacks which occurred in London (September 11) as they could potentially influence the general sentiment expressed in regards to a location during that period of time.

With the Google sheets active and constantly updating, after the predetermined collection period, the data was exported in excel format for processing. It was then cleaned and kept organized in individual sheets segregated by the location in order to facilitate the subsequent reprocessing through Semantria and the incorporation of additional information.

Furthermore, previous studies, list the number of followers of an account, number of other accounts an account follows (following), and number of tweets, as relevant indicators of Twitter performance (Anger & Kittl, 2011; Bakshy et al., 2011; Bayram & Arici, 2013; Crump, 2011; Rossi & Magnani, 2012; Sevin, 2012).

In the case of this study, the "performance" is not specifically related to the activity of an account but rather on the popularity of location. Therefore, from the data that was extracted from twitter, the variables that were believed to be more relevant and kept for further processing, were the number of retweets, which as explained above could be used as proxy of proliferation for a post, the number of followers which could be a regulator for exposure the sample size (number of tweets retrieved) which could be an indication of how popular a location is and finally the tweet text which is the basis for sentiment. The table below represents a short summary of the data that was gathered and the preliminary analysis that was done in excel. The three locations listed were the ones that generated most tweets out of each city.

Table 2: Initial Data Set for Processing

		Count				Average			Correlation		
	Location	**Trip Advisor Ranking**	**Master card's Global Index**	**Sample Size**	**Number of Retweets**	**Followers**	**Retweets**	**Likes**	**Followers & Likes**	**Followers & Retweets**	**Retweets & Likes**

New York, USA	Times Square	#26	#5	20,086	2,820,385	14,003.58	70.31	0.21	2.41%	2.41%	- 0.70%
	Central Park	#1		6,769	1,566,222	7,145.57	231.38	0.67	46.26%	-0.92%	- 1.19%
	Statue of Liberty	#14		1,720	64,781	7,501.76	37.66	0.57	32.73%	-1.27%	- 0.66%
London, UK	Bucking ham Palace	#45	#2	16,151	6,217,809	6,582.63	384.98	0.46	31%	-4%	-1%
	Big ben	#5		3,866	36,317	5,765.66	9.39	0.37	25%	1%	0%
	London Eye	#24		2,644	30,772	8,271.95	11.64	0.46	4%	-3%	2%

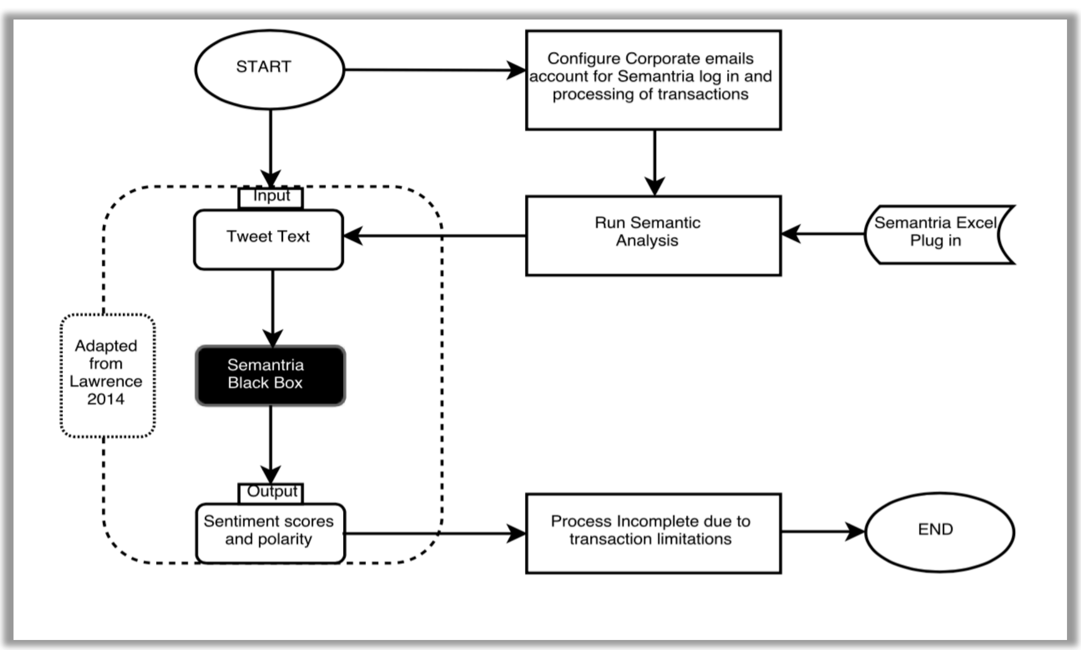
4.3 Data Set Process

4.3.1 Semantic Analysis

Once the data was preprocessed in Excel, it was then ready to be processed in Semantria via the Excel Plug in. As previously noted, the location data was organized in individual sheets so that the process could be repeated for each location. In order to simplify an analysis, 1,000 documents were gathered through Semantria respectively to each chosen location.

The diagram below indicates how the process was carried out:

Figure 7: Flowchart of Semantic Analysis Process



Source: Author's own creation

The data files for each location were imported to SPSS and a random sample of 1000 tweets was generated per location.

After, the data was again preprocessed in excel and prepared for processing in Semantria.

Table 3 below represents summary of the random sample data set.

Table 3: Random Sample of 1000 tweets per location

		TripAdvisor		Twitter	
	Location	**Ranking**	**Number of Reviews**	**Average Retweets**	**Average Followers**
New York	Central Park	1	14,290	204	12,702
	Statue of Liberty	14	4,175	40	7,933
London	Big Ben	5	2,854	9	6,326
	London Eye	24	9,340	12	10,050

Source: Author's own elaboration

After this process, the semantic scores and polarity were added to the data files of each location. Before the data was uploaded to SPSS, a final reprocessing in excel revealed two locations as outliers. Times Square and Buckingham Palace both had a considerable amount of negative tweets arising from recent incidents which would significantly bias the data set. This effect can be easily observed by the word cloud of phrases generated by Semantria. Hence these locations were removed.

Figure 8: Word cloud Times Square

Figure 9: Word Cloud Buckingham Palace

Furthermore, the locations were grouped according to their level of popularity. The TripAdisor (2017a) relative rank of city attractions was used to combine the locations into two groups: More and Less popular. The table bellow summarizes the key variables used for the study.

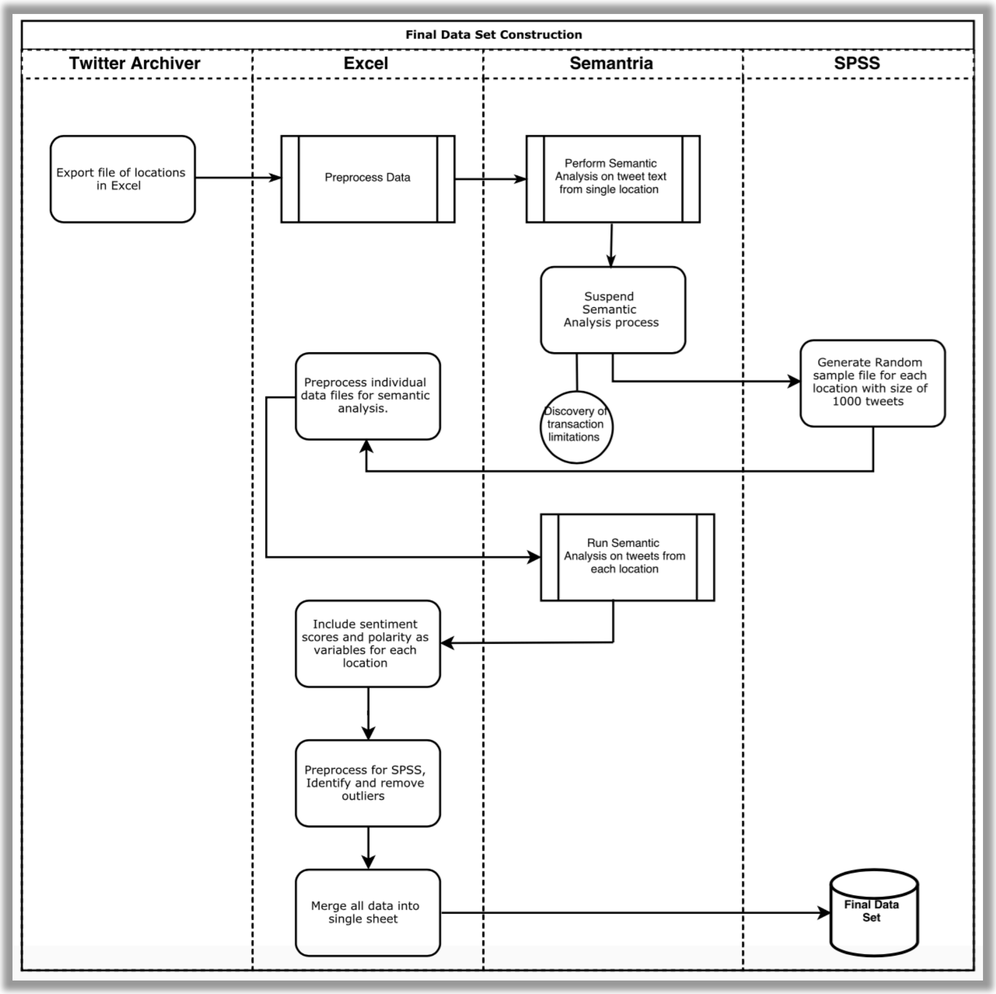
Table 4: Sample grouped by popularity ranking

	Location	TripAdvisor Ranking	Averagre retweets	Averagre Followers	Averagre Sentiment Score
More Popular	Central Park	1	106.607	9514,62	0,18
	Big Ben	5			
Less Popular	Statue of Liberty	5	26.074	8991,86	0,28
	London Eye	24			

Source: Author's own elaboration

In conclusion, the complete data set process can be observed with the simplified diagram below.

Figure 10: Flowchart for Final Dataset



5 RESULTS

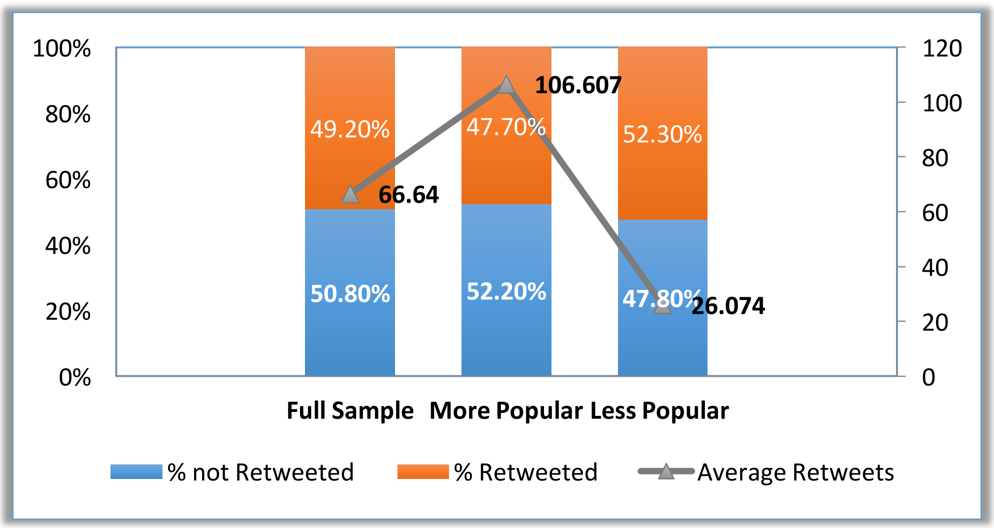
In order to build a better understanding the similarity or difference between popular and less popular locations in Twitter, this research started by conducting the descriptive analysis of the data between the two.

5.1 DATA DESCRIPTION

As referred to in previous sections, the three main variables that are relevant to the objective of this study are the number of followers, the number of retweets and the sentiment score. In this section descriptive statistics were used to characterize the data set in terms of the main variables.

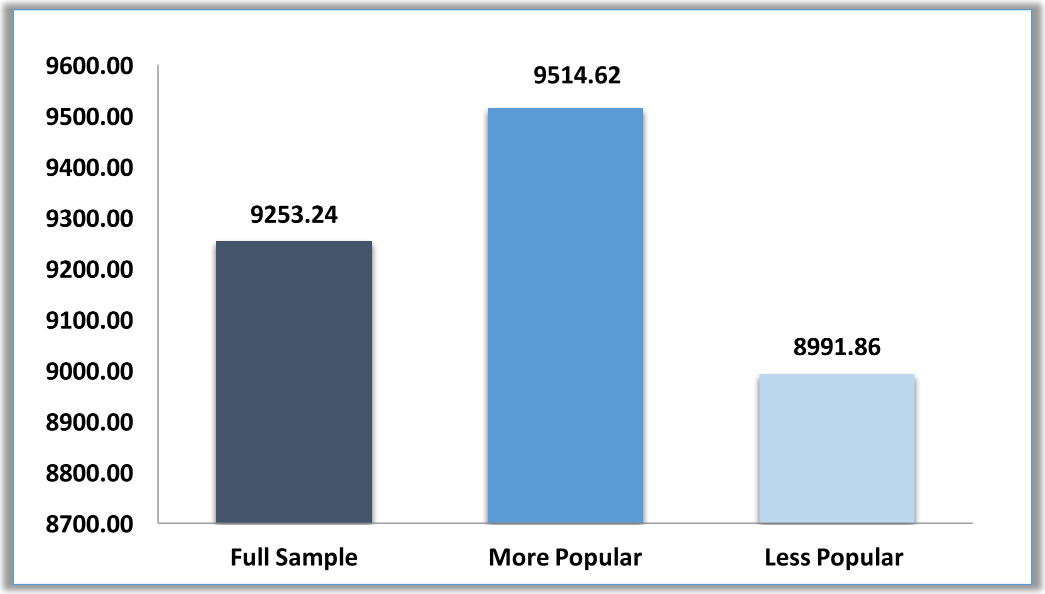
Retweets. In this study, the retweet rate is defined as the percentage of tweets that have been retweeted within a given sample and independent of the frequency of retweets. A cross tab analysis was run to determine how the retweet rate differs according to the popularity level of the touristic location. Figure 11 below showcases the results, the sample has a closely split retweet rate (49,20%). In terms the touristic attractions segregated by group, popular locations have a lower retweet rate (47,7%) with a higher average retweets (106,607), whilst the less popular locations have a higher retweet rate (53,30%) but lower average retweets (26,074). More and less popular attractions significantly differ in terms of the retweet rate as reported by the chi-square $\chi^2(1, N = 4000) = 8.102$, $p < 0.005$ (Appendix 1).

Figure 11: Retweet Rate and Average



Followers. The more popular locations have a higher average (M = 9514,62; SD = 98411.93) than less popular locations (M = 8991,864,; SD = 73660,50). Figure 12 below illustrates the average number of followers:

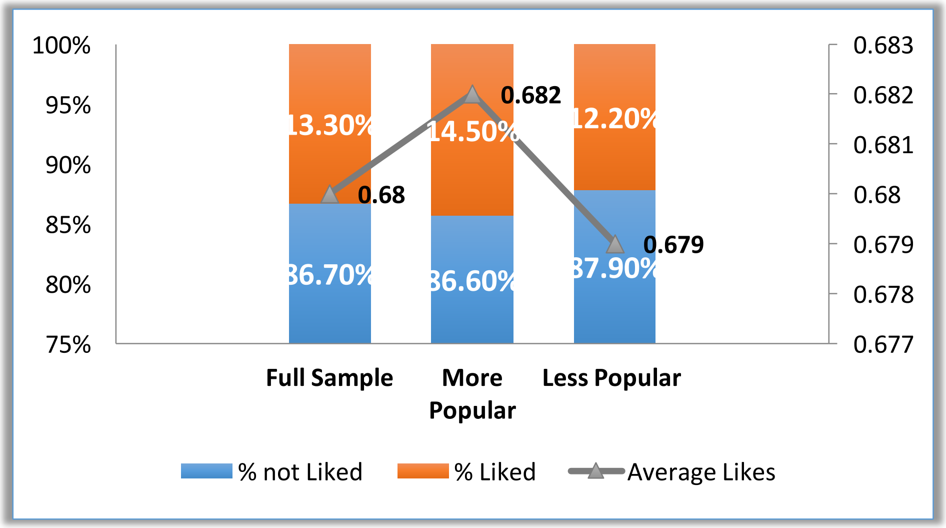
Figure 12: Average Number of Followers:



Likes. In this study, the like rate is defined as the percentage of tweets that have been liked within a given sample and independent of the frequency of likes.

A cross tab analysis was run to determine how the like rate differs according to the popularity level of the touristic location. Figure 13 below showcases the results, in the sample; the majority of retweets have not been liked. In terms the touristic attractions segregated by group, popular locations have a both a higher like rate and average with 14,50% of tweets being liked. More and less popular locations, significantly differ terms of the like rate as reported by the chi-square $\chi^2(1, N = 4000) = 4.588, p < 0.005$ (Appendix 2).

Figure 13: Likes rate and average



Tweet Polarity and Sentiment Score.

The tweet Polarity category, is defined by the sentiment score range, Table 5 below, indicates what are the sentiment score ranges for each type of polarity across our sample.

Table 5: Sentiment Score range and respective Polarity

	Sentiment Score Range		
Documents & Expressions	*Negative*	*Neutral*	*Positive*
	[-1.8463; -0.15]	**[-0.0554; 0.2084]**	**[0.1021, 2.25]**

Table 6 below is a summary table sampling a different tweet polarity and the example of the relevant text for each location.

Table 6: Tweet Text Semantic Scores and Polarity Matrix

City/Country	Attraction	Sentiment Rank Mean	Valence	An example of the Tweet Text
London, United Kingdom	Big Ben	1.5	Positive	London is such a beautiful place! Walked about 8 miles exploring this city. Taking it all in! #BigBen
		0	Neutral	What time is it? #BigBen says it's time to rock!
		-0.174	Negative	#BigBen is actually kinda small... and fat.
	London Eye	1.6	Positive	Iconic, unusual and with spectacular views! The #LondonEye is such a special venue for weddings
		0	Neutral	Take me back to #londoneye
		-0.3	Negative	Rainy London #Londoneye
New York City, United States of America	Statue of Liberty	1.6	Positive	The #statueofliberty is very impressive - one day I will climb the #crown!
		0.11	Neutral	It's so #foggy I can't even see the #statueofliberty. Kinda makes me feel like I am #spying on her from behind a #curtain. haha
		-0.553	Negative	#StatueOfLiberty? More like #StatuteOfLunacy
	Central Park	2.25	Positive	So beautiful! #centralpark #nights @ Central Park North
		0	Neutral	Where do the ducks go in the winter? #ducks #centralpark
		-1.85	Negative	Too damn cold! BALLS! #centralpark

Source: Author's own elaboration

A cross tab analysis was run to determine how the polarity is distributed across the full sample and the locations segregated by popularity The sample is evenly split between positive and neutral tweets (46%). Negative tweets represent only 7% of the sample. For more popular locations, the majority of the tweets are neutral (57,7%) whereas positive and negative are 17,2% and 6,4% respectively. In the case of less popular locations, the majority of tweets are positive (45,7%) neutral and negative representing 33,7% and 8,4% respectively. The difference in the polarity rate in terms of popularity was found to be significant as reported by the chi-square $\chi^2(3, N = 4000) = 242.627, p < 0.001$ (Appendix 3).. The figures below showcase these distributions.

Figure 14: Full Sample Tweet Polarity Distribution

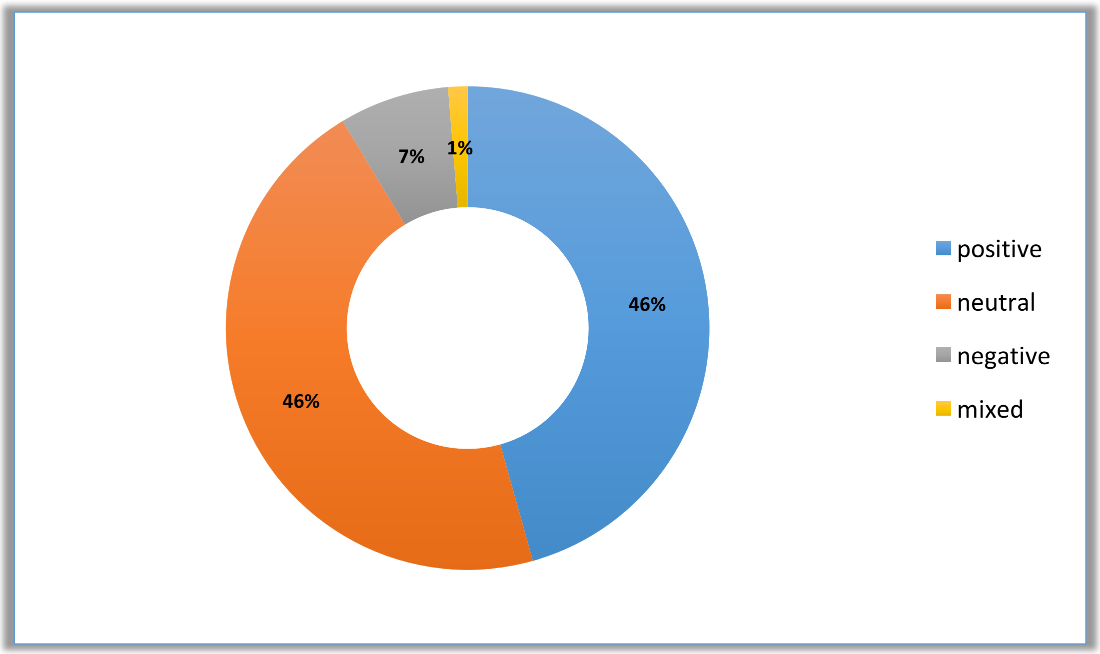


Figure 15: More Popular Tweet Polarity Distribution

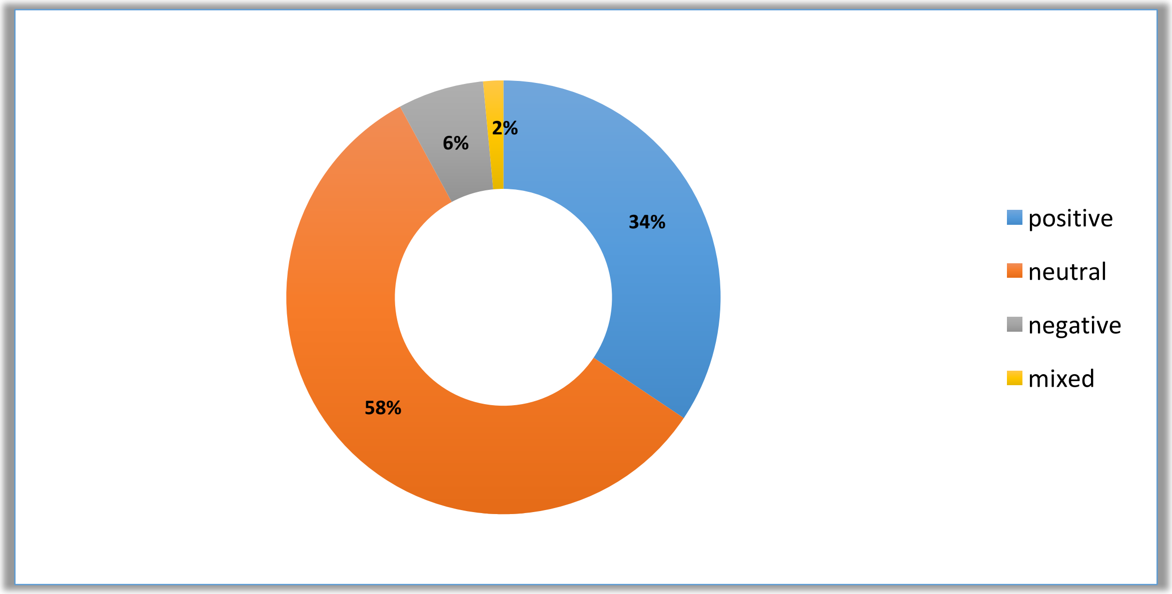


Figure 16: Less Popular Tweet Polarity Distribution

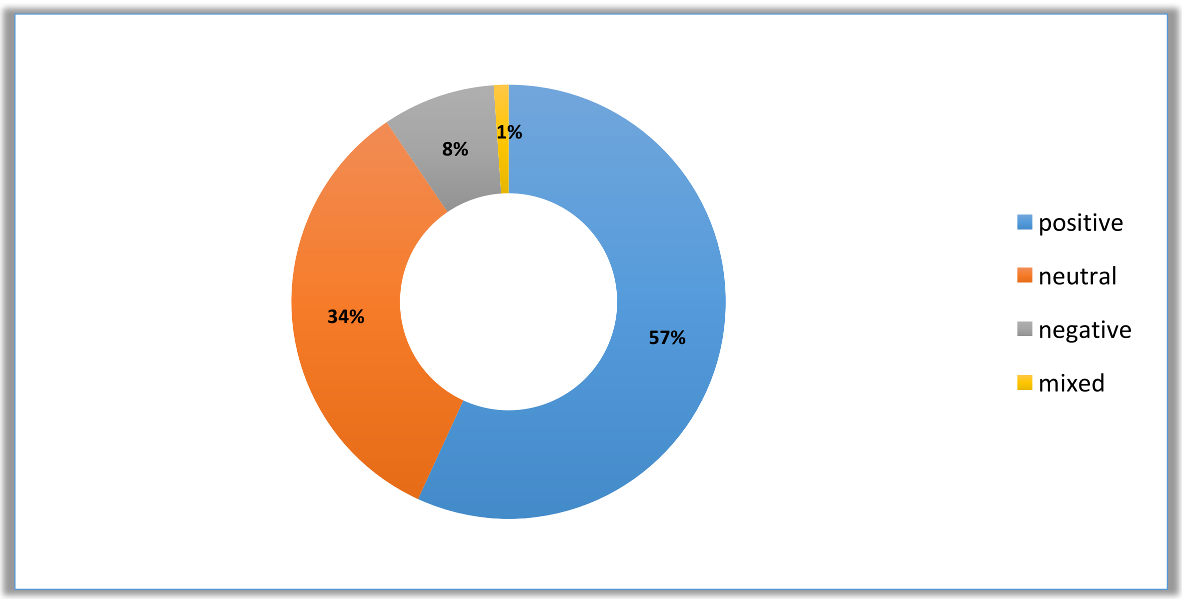
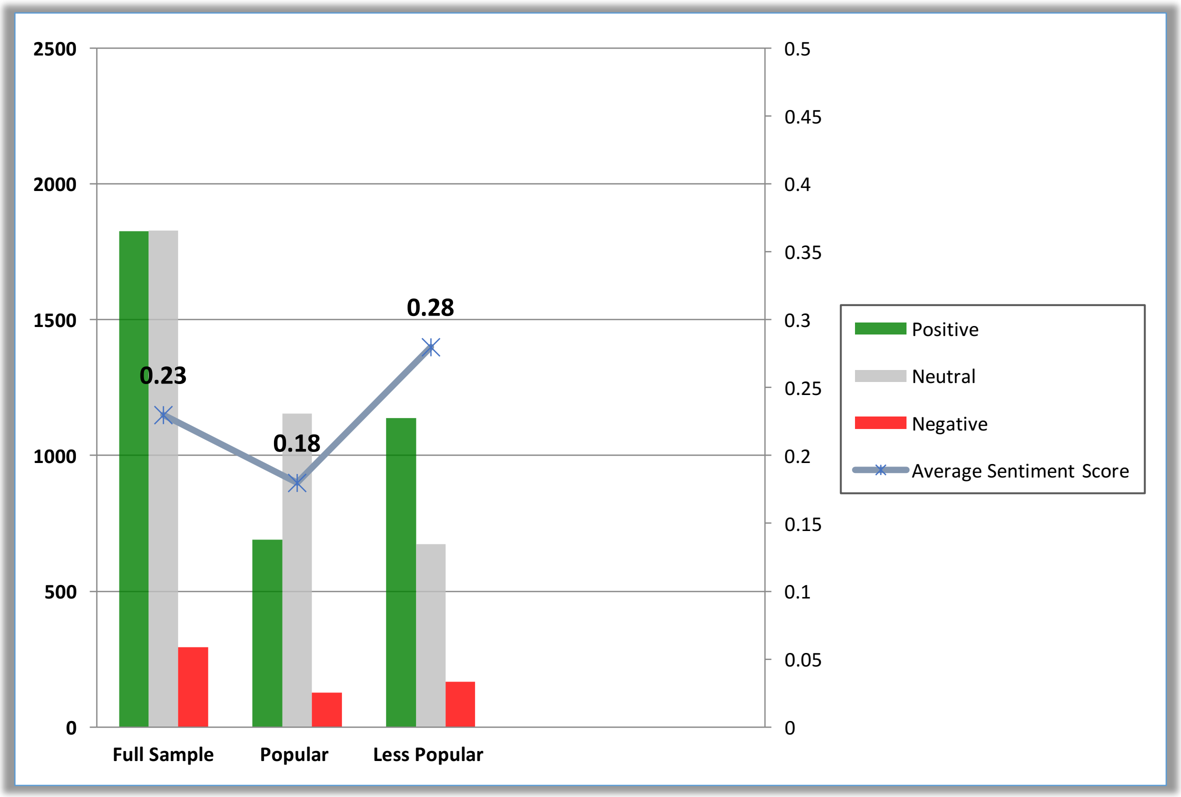


Figure 17: Tweet Polarity Frequency per Location



5.2 EMPIRICAL TESTING

5.2.1 Testing Hypothesis 1

H1. In Twitter, tweets about more popular touristic attractions, generate a higher number of retweets than less popular ones.

In this hypothesis, we would like to understand if message diffusion is related to popularity by testing weather the number of retweets, is higher for more popular locations.

Results:

A one way anova was run to determine the nature of the relationship between the number of retweets and For hypothesis **H1a**, the hypothesis of homogeneity of variances is violated. Therefore, the Welch's adjusted F ratio is reported. There was statistically significant difference between groups means at the 0.05 alpha level [$F(1, 2266.426) = 6.967$ $p<0.01$, (or, $p<0.05$) (Appendix 4). More popular touristic locations (M=106.607), achieved a higher amount of retweets than less popular locations (M=26.074). Hence, evidence is found to support hypotheis H1a that more popular locations generate higher amount of retweets.

5.2.2 Testing Hypothesis 2

H2a. In Twitter, for more popular locations, the number of followers has a stronger significant impact on the number of retweets than for less popular locations.

The hypothesis we are trying to test, is meant to comprehend whether the number of followers, in other words network size, can imply a higher level of direct exposure of a message and therefore be reflected in the number of retweets. In order to test this hypothesis, the sample is segregated by the popularity level.

Results

This hypothesis is tested through correlations on each sample of more and less popular locations respectively in order to determine the how the relationship between the number of retweets and followers differs across more and less popular touristic attractions. Since data is not normally distributed (Appendix 5.1), a non-parametric correlation test was used. The Kendall-Tau was preferred over the Spearmen test, as both variables are measured in a continuous scale and they do not exhibit a monotonic relationship violating the assumptions of the Spearmen test.

The results of both correlations do not support hypothesis H1. For more popular locations a very weak positive significant correlation was found ($\rho = 0.049$, sig. < 0.01) (Appendix 5.2). For less popular locations a very weak positive significant correlation was also found ($\rho = 0.075$, sig. < 0.00) and it was the relationship was slightly stronger (Appendix 5.3).

Therefore, hypothesis 2 is rejected as there is no evidence to support the claim that the number of followers, for more popular locations, has stronger effect on the number of retweets than for less popular ones.

5.2.3 Testing Hypothesis 3

> **H3.** In Twitter, tweets about more popular touristic locations, carry a higher sentiment score than less popular ones .

The premise for this hypothesis is to test whether the general sentiment associated with more popular locations may be the driver for their greater popularity.

Results

A one-way ANOVA was performed to compare how the popularity of location, influences the sentiment expressed (measured by the sentiment score) in tweets relating to it. A statistically significant difference between the mean sentiment scores was found at a 1% alpha level as reported by ANOVA [$F(1, 3998) = 77.168$, $p = 0.000$] (Appendix 6). Tweets about less popular locations (M=0,2834) had a higher average sentiment score than tweets about more popular locations (M=0,1832)

Therefore, hypothesis **H3** is rejected as there is no evidence to the claim that more popular locations generate a higher average sentiment.

5.2.4 Testing Hypothesis 4

H4. In Twitter, negative posts about popular touristic attractions are the most retweeted.

In order to test this hypothesis, the level of popularity segregates the sample. Therefore this hypothesis is tested on two levels, first for more popular locations and second for less popular locations.

Results

A one-way ANOVA Tests was run to determine if for more popular locations, there is a significant difference among the sentiment polarity groups and their respective number of retweets. The hypothesis of homogeneity of variances is violated. Therefore, the Welch's adjusted F ratio is reported. There was statistically significant difference between groups means at the 0.05 alpha level [$F(3, 429.209) = 18.926$ $p<0.00$, (or, $p<0.05$) (Appendix 7.1).

A Games Howell post hoc test revealed that for more popular locations negative tweets (43.05 ± 161.16, $p= 0.024$) were less retweeted than mixed tweets. There was no other significant difference between negative sentiment tweets and other groups. The test also revealed that neutral tweets were significantly more retweeted than positive (9.46 ± 27.71, $p= 0.007$) and mixed tweets (1.96 ± 3.26, $p= 0.004$) (Appendix 7.1).

A one-way ANOVA Tests was run to determine if for less popular locations, there is a significant difference among the sentiment polarity groups and their respective number of retweets. The hypothesis of homogeneity of variances is violated. Therefore, the Welch's adjusted F ratio is reported. There was statistically significant difference between groups means at the 0.05 alpha level [$F(3, 591.164) = 50.973$ $p<0.00$, (or, $p<0.05$) (Appendix 7.2).

A Games Howell post hoc test revealed that for less popular locations negative tweets were significantly more retweeted than mixed tweets (0.318 ± 0.5679, $p= 0.00$) and neutral tweets (3.616 ± 10.7449, $p= 0.00$) . There was no other significant difference between negative sentiment tweets and other groups. The test also revealed that neutral tweets were the most retweeted group, being significantly more retweeted than positive (9.46 ± 27.71, $p= 0.007$) and mixed tweets (1.96 ± 3.26, $p= 0.004$) (Appendix 7.2).

In regards to more popular locations, no evidence was found to support the hypothesis. Concerning less popular locations, the hypothesis is validated, there is evidence to support the claim the negative retweets are more retweeted.

Below is a Table highlighting the results of the hypothesis.

Table 7: Hypothesis results summary

Hypothesis	Results
H1	Valid
H2	Not Valid
H3	Not Valid
H4a	Not Valid
H4b	Valid

Source: Author's own elaboration

6 DISCUSSION & CONCLUSION

Social media in the travel information search domain has witnessed significant growth (Xiang and Gretzel, 2010). Companies that have invested in social media engagement have seen a strong return on investment by improving the strength and durability of their brand (Dholakia & Durham, 2010; Cruz & Mendelsohn, 2010). In the tourism industry, social media has been growing in relevance for the consumer decision process and positive eWOM spread can be an important contributor to brand awareness and sales (Hudson and Thal, 2013). Additionally, research on the impact of social media on hotel performance by Duan et al. (2015), not only highlighted the role of social media in generating awareness but also its importance as a source of insights for marketing strategy in different industries. The authors provided evidence that online customers generate perceptions on different dimensions and thus emphasizing the importance of analyzing UGC. Furthermore, the authors also found that a higher number of reviews lead to more future reviews stressing on the relevance of a social media strategy.

Scharl et al. (2008) distinguished the impact that estimation and web-based social networking spread has on touristic area's image. The authors argue that cities have been seen to take advantage of the advancements in technological infrastructure to attract tourists and advertise services and as such, platforms like Twitter can be influential tools for building on city and brand reputation.

In addition, Twitter can offer countries and other tourism related entities, a platform for interaction between their assets and current or potential clients in a more individual manner. Travelers' or guests' tweets about their experiences could decidedly or contrarily influence the host nations' or establishments' image, future notoriety, and capacity to pull in more guests (Arns, 2015).

Taking into consideration the role that social media has on the tourism industry, this study attempted to explore whether there were any differences in general perceptions amongst more and less popular touristic locations. For this purpose, a sample of tweets mentioning popular and less popular touristic attractions was used to answer the research question focusing on the content and level of message diffusion.

The findings of this study suggest that there is a difference in terms of exposure and affection in regards to more and less popular touristic attractions on social media.

Previous literature on social media network structures, has some divergence with what contributes more to exposure particularly in regards to network size, measured by the number of followers. The main argument centers around how effectively a higher followers base expands the networks that a message is exposed to.

A study on over 60 million tweets by Suh et. al (2010), found that the followers and followees have a significant effect on predicting weather a tweet is retweeted. Contrarily to these findings Cha et al. (2010), argued that using the number of followers, as a proxy for user network reach can be misleading, and that the retweet measure is usually more significant for practical applications. Corroborating Chas´s research, Benevenuto et al. (2010) and Bakshy et al. (2011), studies further emphasized that the size of network is not effective contributor to propagation of a message due to differences in activity of a user base. The authors contended that a smaller group of more active followers may be more effective.

The results of this study are in line with previous research in regards to the ineffective role of the follower's base. There was no evidence to suggest that the size of network has significant influence in the propagation of a message. These opposing views on the relationship between network size and message propagation, may be explained by an etiquette phenomenon on social media, where users tend to follow their followers and rendering the followers base less reliable as means of influence (Avnit et al., 2009). Nonetheless, Hong et al. (2011) and Zhou et al. (2011) found retweets to be reliable in providing tractable information for diffusion models.

This study´s results in respect to the use of the number of retweets as measure of exposure, are convergent with the external popularity ranking used of TripAdvisor where more popular locations exhibited a higher average number of retweets.

Besides the aspect of diffusion, the content of the message is of particular importance. Hence, research by Cha et. al (2010), found that retweets are mainly driven by the content value of the tweet, supporting the growing importance of further analysis on the content of tweets.

The main issue with the proposed task revolved around the difficulty in deriving the valuable insights underlying these large volumes of unstructured data. In this regard, sentiment analysis is regarded as tool of choice as it can offer more value than just relevant information, enabling the capture of feelings and opinions that are embedded within a message (Anbananthen et al., 2013). In this study, semantic analysis was conducted through an automated tool, based on machine learning techniques. The tool of choice was Semantria as previous studies had attested its reliability (Gao et al. 2015; Peisenieks & Skadiņš, 2014; Serrano-Guerrero et al. 2015).

Half of the sample sentiment polarity was neutral, which implied that half of our tweets did not capture strong emotions. Several studies, mostly on the valence of online reviews, have chosen to remove messages with neutral sentiment as they believed them to be insignificant. In the specific case of this study author argues otherwise on the basis of Cocker´s (2014), which implies that sharing carries motives. Consequently, due to the network structure of twitter, a mention of a touristic attraction, despite its neutral sentiment, should have a higher diffusion impact.

The results confirmed this assumption, that neutral tweets have some significance and relation to the aspects of message diffusion, as they were found to be the most retweeted sentiment polarity category. Despite its ambiguity, this finding may be explained with the retweeting action, being an act of approval or recognition and carrying the implicit sentiment of the retweeter (Boyd et. al 2010; Wong et. al 2016).

Confirming the significance of neutral sentiments, research by Tang et al. (2014) on online reviews, found that neutral sentiments have a mixed effect on sales. The authors defined neutral reviews as either mixed or indifferent with both sets of categories generating a positive and negative effect on sales respectively.

Despite significance, previous research also addresses how differences in sentiments polarity trigger sharing and diffusion. Previous studies have demonstrated that users tend to follow a self promoting role and generate messages with a more positive sentiment (Sobel et al., 2009; Berger & Milkman, 2010). In the case of this study, and given the characteristics of the sample mentioned above, neutral polarity sentiments seemed to have a predominant role. When analyzing the sample collectively, neutral sentiment tweets stand out as the most retweeted. When looking at the sample segregated by popularity, for more popular locations most retweets are neutral while for less popular locations, most retweets were negative. Hence, positive tweets were not found to be catalyzers of retweets.

Nonetheless, the impact of affection in the popularity of tourist attractions is somewhat unclear. The sentiment score of the tweets messages revealed ambiguous results. Messages relating to less popular locations were found to have a higher average sentiment. However, these results lack in relevance as for both sets of locations the mean sentiment was of neutral polarity. Additionally, when the neutral tweets are removed to test for emotional divergence, the two sets of locations are found not to differ.

Supporting these findings, study on sentiments of comments of three monuments in Spain mining data from TripAdvisor by Valvadia et al. (2017) found the number of neutral polarities to be of highest frequency. In addition, the study observed that the distributions of polarities were divergent with the user ratings suggesting that user rating should not be used as a rating for sentiment

However, previous research on online reviews has demonstrated the relevance valence in addition to the volume that can have effect on business performance. According to Kostrya et al. (2015), positive valence of online reviews, has strong effect on customer's choice and volume moderates its impact. Similarly, a study by Kim et. al (2015), on the effect of social media on post-trip experience, found that sharing positive experiences on social media, result in more positive evaluations.

In this study, no such relationship was found for positive sentiment messages. Nonetheless, when analyzing the attributes of positive sentiment messages, it is important to consider how an online review is intrinsically related to the product in hand while a comment in social media can have a much less direct relation.

However, for less popular locations, our results show that the findings are synonyms with Barbagallo et al. (2012) study of tweets in the tourism domain, negative tweets were more retweeted than positive tweets. This finding may offer some direction in terms of the relationship between message sentiment and popularity. A study by Floyd et. al (2014), found that negative sentiment reviews also had an effect on costumers choice by dissuading potential customers from purchasing. Taking this study into consideration, although the question is intrinsically in the relation remains, the difference in popularity may be explained by the fact that negative messages were more diffused.

Previous research, found that the extent to which experiences are shared, can affect travelers emotions and perceptions towards an actual experience (Moore, et al. 2012; Garnefeld et al. 2011). This study aimed at understanding weather more activity in social media explained the difference in popularity of touristic attractions. Results showed that network size has no relevant effect on diffusion, the more popular locations had a significant higher message diffusing activity measured by the number of retweets with most of them being of neutral polarity. Less popular locations on the other hand, had a higher diffusion of negative messages but a higher average sentiment.

Furthermore, sentiment analysis of social media is of great utility for the tourism industry as it is effective in capturing real time information and public buzz (Philander et al. 2016).

6.1 THEORETICAL IMPLICATIONS

There are two main contributions to the tourism destination's theoretical literature that were found in this study. The first aspect represents that the perception of the social media users on the iconic attractions regardless if the location has been visited or not, can be different for the popular and less popular destinations. Primarily, it's possible to measure exposure of the information on locations through number of retweets, with application of which, this study found that more popular locations are tend to experience higher average information diffusion than less popular, given that there's no significant influence message propagation by number of followers. However, previous studies suggested that not only information exposure but also message content is crucial for the research of perception on tourist destinations. In this line, second perspective of the study consists of identifying the affection by applying sentiment analysis tool, and measuring the relevant valence associated to popular and less popular locations. Even though there was a number of previous studies on valence of Twitter posts given the specific touristic location, they did not include the comparison of several famous locations considering the difference in the rank of popularity.

6.2 PRACTICAL IMPLICATIONS

With the increasing popularity of tourism destinations, knowledge of the perception of the social media users on iconic attractions ultimately contributes to the marketing destination planning. Firstly, the social platforms such as Twitter further builds consumers recognition of the destination and thus influence their decision in planning their trips. Therefore, it is crucial for the DMOs to constantly monitor the posted information and take a managerial action for further changes. For instance, less popular locations tend to have more negative retweeted posts than positive, suggesting that the timely action on improvement strategy by DMOs could enhance the opinion of both potential and real visitors. Furthermore, the sentiment analysis in this study revealed that such tool could examine not only the increasing popularity of tourism destinations, but also to evaluate the difference, by combining popular and less popular groups, between variously ranked brands, products, restaurants, hotels and other industries. For DMO's, monitoring public opinions and comparison of exposure of the retweets between popular locations, could also serve in development of marketing strategies, recommending which destination needs more investment in promotion. This could be applied by both leveraging positive location sentiments, as well as responding to a negative perception on time. Moreover, by analyzing the themes and categories of negative sentiments an investigation could be done into what factors influence a negative opinion about a certain location. For instance, for some destinations themes on pollution or long lines may help to form a better business decision in a long run.

6.3 LIMITATIONS OF THE STUDY

The study identifies certain limitations, that may also represent opportunities for future research. The information on Twitter and TripAdvisor is altogether different in both structure and significance. Twitter, as other online networking sites, can be viewed as microblogging stage that enables sharing contemplations and insights about any theme. TripAdvisor on the other hand, is more directed towards the tourism business, encouraging the sharing of surveys about travel related spots. Thusly, obtaining pertinent information on Twitter is considerably more complex process. For instance, during the collection period, between the period of March to May, London Eye had 9,340 reviews on TripAdvisor while the sample on the study only captured 2,644 tweets with the #LondonEye. This was evident across a variety of popular locations, that were initially planned to be included in the study but did not retrieve a sufficient data set of tweets thus limiting the amount of locations in use.

Another main imitation was processing capacity of the Semantria trial version as it only allows for a total of 4,000 thousand transactions in this context, referring to the analysis of each tweet text. Not every tweet retrieved represents a sentiment expressed about the location, there are also advertisements about products or informative links. Thus, if there was no limitation in number of transactions and research timeframe, reviewing the categories or themes assigned to certain keywords could eliminate irrelevant information from the sample and generate more reliable results. Moreover, this exemplifies the main limitation of semantic analysis identified by this study. Notwithstanding the sophistication of machine learning tools such as Semantria, semantic analysis it is still not robust enough to detect implicit feelings without the aid of stronger wording. Thus human intervention may be necessary for the construction of a more effective lexicon but on the other hand, greatly limits the processing ability of vast amounts of unstructured data that can uncover new knowledge.

Therefore, in terms of data mining, a wider time span should be given to the data collection process, in order to ensure that sufficient data is gathered. Therefore next studies may include a wider range of locations and use the TripAdvisor or any other popularity ranking more orderly. For instance listing the top 3 locations in several different cities.

Additionally, it is proposed to compare popular and less popular destinations, see if there is an underlying sentiment that influences the choice of visitors, study occasions such as holidays (Christmas, New Year and etc.). That would allow tourism operators to leverage these occasions for promotion of their businesses, or in situations such as terror attacks understand how to minimize the negative sentiments and its impact on number of visitors. Moreover, it will be interesting to see the same study by using daily data of visitors in tourist attractions and tracking geo-tagged locations, images and posts.

REFERENCES

Akehurst, G. 2009. User generated content: the use of blogs for tourism Organizations and tourism consumers. ***Service Business,*** 3(1): 51-61.

Anbananthen, K., Elyasir, A. 2013. Evolution of opinion mining. ***Australian Journal of Basic and Applied Sciences***, 7(6): 359-370

Anger, I. and Kittl, C. 2011. Measuring Influence on Twitter. I-KNOW, ***Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies art***. 31.

Antoniadis, K., Vrana, V., & Zafiropoulos, K. 2014. Promoting European Countries Destination Image through Twitter. [Online] Available at: https://cld.pt/dl/download/42cb39a2-9d80-4eae-8f59-a9e1713d1e70/Proceedings%20Book%20ITC%2713/Papers/P43-13.pdf [Accessed 1 June 2017].

Antoniadis, K., Zafiropoulos, K., & Vrana, V. 2015. Communities of Followers in Tourism Twitter Accounts of European Countries. ***European Journal of tourism, Hospitality and Recreation***, 6 (1): 11-26.

Arns, J., 2015. ***Annual Review of Cultural Heritage Informatics***. Rowman & Littlefield Publishers, Inc. Maryland.

Avnit, A. 2009. ***The Million Followers Fallacy***. Internet Draft, Pravda Media. [Online] Available at: http://tinyurl.com/nshcjg [Accessed 26 September 2017].

Ayeh, J., Au, N., & Law, R. 2013. "Do We Believe in TripAdvisor?" Examining Credibility Perceptions and Online Travelers' Attitude towards Using User-Generated Content .***Journal of Travel Research,*** 52(4): 437-452.

Bach, M.P., Schatten, M., & Marusic, Z. 2013. Data Mining Applications in Tourism: A Keyword Analysis. ***Central European Conference on Information and Intelligent Systems***. [Online] Available at: http://bib.irb.hr/datoteka/648698.paper.pdf [Accessed 11 December 2016].

Bakshy, E. & Hofman, J. M. & Mason, W. A. & Watts, D. J. 2011. ***Everyone's an influencer: quantifying influence on twitter.*** In Proceedings of the fourth ACM international conference on Web search and data mining (WSDM '11). ACM, New York, NY, USA: 65-74

Barbagallo, D., Bruni, L., Francalanci, C. and Giacomazzi, P. 2012. An Empirical Study on the Relationship between Twitter Sentiment and Influence in the Tourism Domain. Information and Communication Technologies in tourism. doi: https://doi.org/10.1007/978-3-7091-1142-0_44

Bayram, M. and Arici, S. 2013. ***Destination Marketing Organizations' Social Media Usage: A Research on Balkan Countries.*** International Conference on Economic and Social Studies (ICESoS'13), 10-11 May, Sarajevo.

Bekk, A. 2016. Digital Divide in Tourism. Differences among Generation X and Y towards Online Travel Reviews Writing. [Online] Available at: http://www.etourism-students.com/blog/digital-divide-tourism/ [Accessed 3 December 2016].

Benevenuto, F. & Cha, M. & Gummadi K.P. & Haddadi, H. 2010. ***Measuring user influence in Twitter: The million follower fallacy.*** In Proceedings of the 4th International AAAI Conference on Weblogs and Social Media. Pp. 10-17, Association for the advancement of artificial intelligence.

Bennett, T. 2014. 7 Advantages of Tourism in an Economy. [Online] https://blog.udemy.com/advantages-of-tourism/ [Accessed 28 October 2016]

Berger, J. & Milkman, K. 2010. ***Social transmission, emotion, and the virality of online content.*** Wharton Research Paper.

Berry, M.W. & Kogan, J. 2010. ***Text Mining: Applications and Theory.*** USA: John Wiley and Sons.

Blackshaw, P., & Nazzaro, M. 2006. **Consumer-generated media (CGM) 101: Word of mouth in the age of the web-fortified *consumer***. New York: Nielsen BuzzMetrics.

Blal, I., and M. Sturman. 2014. The Differential Effects of the Quality and Quantity of Online Reviews on Hotel Room Sales. ***Cornell Hospitality Quarterly***, 55(4): 365–75.

Boyd, D., Golder, S. & Lotan, G. 2010. ***Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter***. In System Sciences (HICSS), 2010 43rd Hawaii International Conference on System Sciences, 5 – 10 Jan., Honolulu.

Calheiros, A. C., Moro, S., & Rita, P. 2017. Sentiment Classification of Consumer-Generated Online Reviews Using Topic Modeling. ***Journal of Hospitality Marketing & Management***, 1-19.

Cha, M. H. (2010). Measuring user influence in twitter: The million follower fallacy. ***ICWSM***, (pp. 10-17).

Christian. 2015. Economic Contribution of Travel and Tourism. ***Association of Accredited Public Policy Advocates to the European Union.*** [Online] Available at: http://www.aalep.eu/economic-contribution-travel-and-tourism [Accessed 2 December 2016].

Cocker, B. (2014). ***Advertising primer. A background to Advertising.*** Internet Marketing subject slides. The University of Melbourne.

Cox, C., Burgess, S., Sellitto, C., & Buultjens, J. 2009. The role of user-generated content in tourists' travel planning behavior. ***Journal of Hospitality Marketing & Management***, 18(8): 743-764.

Crump, J.2011. What Are the Police Doing on Twitter? Social Media, the Police and the Public. ***Policy & Internet***, 3(4) Art. 7.

Crotts, J.C., Mason, P.R., & Davis, B. 2009. Measuring Guest Satisfaction and Competitive Position in the Hospitality and Tourism Industry. ***Journal of Travel Research,*** 48 (2): 139-151.

Deloitte. 2013. ***Tourism: Jobs and Growth. The Economic Contribution of the Tourism Economy in the UK.*** [Online] Available at: http://www.tourismni.com/globalassets/facts-and-figures/research-reports/tourism-performance-statistics/economic-impact-of-tourism/economic-impact---deloitte--tourism---jobs--growth.pdf [Accessed 3 December 2016].

Dijkmans, C., Kerkhof, P., Beukeboom, C.J. 2015. A Stage to Engage: Social Media Use and Corporate Reputation. ***Tourism Management***, 47: 58-67.

Duan, W.,Yu, Y., Cao, Q. & Levy, S. 2015. Exploring the Impact of Social Media on Hotel Service Performance: A Sentimental Analysis Approach. ***Cornell Hospitality Quarterly***. 57. 10.1177/1938965515620483.

Floyd, K., Freling, R., Alhoqail, S., Cho, H., Freling, T. 2014. How Online Product Reviews Affect Retail Sales: A Metaanalysis. ***Journal of Retailing***, 90 (2): 217–32.

Gao, S., Hao, J., & Fu, Y. 2015. ***The Application and Comparison of Web Services for Sentiment Analysis in Tourism.*** 12th International Conference on Service Systems and Service Management (ICSSSM) (pp. 1–6). IEEE.

Garnefeld, I., S. Helm, and A. Eggert. 2011. Walk Your Talk: An Experimental Investigation of the Relationship between Word of Mouth and Communicators' Loyalty. ***Journal of Service Research,*** 14 (1): 93–107.

Gemar, G. & Jimenez-Quintero, J.A. 2015. Text Mining Social Media for Competitive Analysis. ***Tourism and Management Studies,*** 11(1): 84-90.

Girardin, Fabien, Francesco Calabrese, Filippo Dal Fiorre, Assaf Biderman, Carlo Ratti, and Josep Blat. 2008b. ***Uncovering the Presence and Movements of Tourists from User- Generated Content.*** [Online] Available at: http://www.girardin.org/fabien/publications/girardin_tourism_statistics2008.pdf [Accessed September 27, 2016].

Girardin, Fabien, Josep Blat, Francesco Calabrese, Filippo Dal Fiore, and Carlo Ratti. 2008a. Digital Footprinting: Uncovering Tourists with User-Generated Content. ***Pervasive Computing***, 7 (4): 36-43.

Gloor, P.A., Colladon, A.F., Miller, C.Z., & Pellegrini, R. 2016. Measuring the Level of Global Awareness on Social Media. [Online] Available at: http://www.ickn.org/documents/COINs16_SocialMediaGlobalAwareness_v9.pdf [Accessed 25 March 2017].

Gretzel, U. and Yoo, K. 2008. Use and impact of online travel reviews. ***Information and Communication Technologies in Tourism,*** Springer, Wien: 35-46.

Gretzel, U., Yoo, K.H., Purifoy, M. 2007. Role and Impact of Online Travel Reviews. [Online] Available at: http://www.tripadvisor.com/pdfs/OnlineTravelReviewReport.pdf [Accessed 8 December 2016].

Gretzel, U., Yuan, Y. and Fesenmaier, R. 2000. Preparing for the New Economy: Advertising Strategies and Change in Destination Marketing Organizations. ***Journal of Travel Research,*** 39: 146-156.

Gu, Z., Zhang, Y., Chen, Y., Chang, X. (2016) Analysis of Attraction Features of Tourism Destinations in a Mega-City Based on Check-in Data Mining—A Case Study of Shenzhen, China. ***International Journal of Geo-Information***, 5 (1): 210.

Hassan, S.B. 2013. Social media and destination positioning: Egypt as a case study. ***European Journal of tourism, hospitality and recreation***, 4(1): 98-103.

Hay, B. (2010, 2). ***Twitter – But who is listening? A review of the current and potential use of Twittering as a tourism marketing tool***. CAUTHE 2010 20th International Research Conference: Challenge the Limits. University of Tasmania, Australia.

Hays, S., Page, S.J. & Buhalis, D. 2013. Social Media as a Destination Marketing Tool: Its Use by National Tourism Organizations. ***Current Issues in Tourism,*** 1693: 211-239.

Hong, L. a. 2011. Predicting popular messages in twitter. ***Proceedings of the 20th international conference companion on World wide web***, pp. 57-58.

Horn, C., 2010. ***Analysis and Classification of Twitter messages.*** Graz: Graz University of Technology.

Hudson, S. & Thal, K. 2013. The Impact of Social Media on the Consumer Decision Process: Implications for Tourism Marketing. ***Journal of Travel and Tourism Marketing,*** 30, (1-2): 156-160.

Hvass, K. A., & Munar, A. M. 2012. The takeoff of social media in tourism. ***Journal of Vacation Marketing***, *18*(2): 93-103.

Jenders, M., Kasneci, G., & Naumann, F. 2013. Analyzing and Predicting Viral Tweets. ***International World Wide Web Conference Committee (IW2C2).*** [Online] Available at: https://hpi.de/fileadmin/user_upload/fachgebiete/naumann/publications/2013/Analyzing_and_Predicting_Viral_Tweets.pdf [Accessed 22 March 2017].

Kim, J., Fesenmaier, D. 2015. Sharing Tourism Experiences: The Post-trip Experience. ***Journal of Travel Research***. (56).

Kostyra, D., Reiner, J., Natter, M., Klapper, D. 2016. Decomposing the Effects of Online Customer Reviews on Brand, Price, and Product Attributes. ***International Journal of Research in Marketing***, 33 (1): 11-26.

Lange-Faria, W. & Elliot, S. 2012. Understanding the Role of Social Media in Destination Marketing. ***Tourismos: An International Multidisciplinary Journal of Tourism***, 7 (1): 193-211.

Leung, D., Law, R., Van Hoof, H., & Buhalis, D. 2013. Social media in tourism and hospitality: A literature review. ***Journal of Travel & Tourism Marketing***, *30*(1-2): 3-22.

Lu, W., & Stepchenkova, S. 2015. User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software. ***Journal of Hospitality Marketing & Management***, *24*(2): 119-154.

MasterCard. 2016. Global Destinations Cities Index. [Online] Available at: https://newsroom.mastercard.com/wp-content/uploads/2016/09/Global-Destination-Cities-Index-Report.pdf [Accessed 15 June 2017].

Miguens, J., Baggio, R. & Costa, C. 2008. Social Media and Tourism Destinations: TripAdvisor Case Study. ***Advances in Tourism Research***: 1-6.

Miguens, J., Baggio, R., & Costa, C. 2008. Social Media and Tourism Destinations: TripAdvisor Case Study. ***Advances in Tourism Research.*** [Online] Available at: http://www.iby.it/turismo/papers/baggio-aveiro2.pdf [Accessed 25 July 2017].

Milano, R., Baggio, R. and Piattelli, R. 2011. The effects of online social media on tourism websites. ***Information and Communication Technologies in Tourism***, Springer, Wien: 471-483.

Moore, S. G. 2012. Some Things Are Better Left Unsaid: How Word of Mouth Influences the Storyteller. ***Journal of Consumer Research***, 38 (6): 1140–54.

Moro, S., Rita, P., & Coelho, J. 2017. Stripping customers' feedback on hotels through data mining: the case of Las Vegas Strip. ***Tourism Management Perspectives***, 23: 41-52.

Moro, S., Rita, P., & Vala, B. 2016. Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. ***Journal of Business Research,*** 69(9): 3341-3351.

Mukherjee, A. & Nagabhushanam, M. 2016. Role of Social Media in Tourism Marketing. ***International Journal of Science and Research (IJSR),*** 5 (6): 2026-2033.

Mukherjee, S. 2012. Sentiment Analysis – A Literature Survey. [Online] Available at: http://www.cfilt.iitb.ac.in/resources/surveys/SA-Literature%20Survey-2012-Subhabratam.pdf [Accessed 2 March 2016].

Munar, A. & Jacobsen, J.. 2014. Motivations for Sharing Tourism Experiences through Social Media. ***Tourism Management***, 43: 46-54.

Nguyen, V. and Wang, Z. 2012. ***Practice of Online Marketing with Social Media in Tourism Destination Marketing. The case study of VisitSweden.*** Sweden: Södertörns University.

Pang, B. & Lee, L. 2008. Opinion Mining and Sentiment Analysis. ***Foundations and Trends in Information Retrieval***, 2 (1-2): 1-135.

Peisenieks, J., & Skadiņš, R. 2014. Uses of Machine Translation in the Sentiment Analysis of Tweets. Human Language Technologies-The Baltic Perspective. ***Proceedings of the Sixth International Conference Baltic HLT,*** 268: 126–131.

Philander, K., Zhong, Y. 2016. Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. ***International Journal of Hospitality Management***, 55: 16-24

Pike, S. 2016. ***Destination Marketing: Essentials***. New York: Routledge.

Rossi, L. and Magnani, M. (2012). Conversation Practices and Network Structure in Twitter. ***Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media,*** 563-566.

Russell, J. A. 1980. A circumplex model of affect. ***Journal of personality and social psychology***, 39(6):1161– 1178.

Saif, H., He, Y., & Alani, H. (2012) Semantic Sentiment Analysis of Twitter. [Online] Available at: http://iswc2012.semanticweb.org/sites/default/files/76490497.pdf [Accessed 28 March 2017].

Santana, G., 2001. Globalisation, safety and national security. In: Wahab, S., & Cooper, C. (ed.) Tourism in the age of globalisation. London: ***Routtledge***.

Santos, C., Rita, P. & Guerreiro, J. (2017) Improving international attractiveness of higher education institutions based on text mining and sentiment analysis. ***International Journal of Educational Management*** (forthcoming).

Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. 2015. Sentiment analysis: A review and comparative analysis of web services. ***Information Sciences,*** 311: 18–38.

Sevin, E. 2013. Places going viral: Twitter usage patterns in destination marketing and place branding. ***Journal of Place Management and Development***, 6 (3): 227 – 239.

Scharl, A., Dickinger, A. & Weichselbraun, A. (2008). Analyzing News Media Coverage to Acquire and Structure Tourism Knowledge. ***Information Technology & Tourism***,10(1), pp. 3-17

Smith, A.N., Fischer, E., & Yongjian, C. 2012. How Does Brand-Related User-Generated Content Differ Across YouTube, Facebook, and Twitter? ***Journal of Interactive Marketing,*** 26: 102-113.

Sobel, K. & Jansen, B. J. & Zhang, M. & Chowdury, A. 2009. Twitter power: Tweets as electronic word of mouth. ***Journal of the American Society for Information Science and Technology,*** 60 (11): 2169-2188.

Sotiriadis, M.D. & Zyl, C. 2013. Electronic Word-of-Mouth and Online Reviews in Tourism Services: The Use of Twitter by Tourists. ***Electronic Commerce Research***, 13 (1): 103-124.

Sparks, B. A., Perkins, H. & Buckley, R. 2013. Online Travel Reviews as Persuasive Communication: The effects of Content Type, Source, and Certification Logos on Consumer Behavior. ***Tourism Management***, 39: 1-9.

Srivastava, A.N. & Sahami, M. 2009. ***Text Mining: Classification, Clustering, and Applications***. New York: CRC Press.

Stankov, U., Lazic, L., & Dragicevic, V. 2010. The extent of use of basic Facebook user-generated content by the national tourism organizations in Europe. ***European Journal of Tourism Research,*** 3(2): 105.

Stepchenkova, S., Kirilenko, A., & Kim, H. 2013. Grassroot Branding with Twitter: Amazing Florida. ***Information and Communication Technologies in Tourism***: 144-156.

Stevenson, A. & Hamill, J. 2012. Social Media Monitoring: A Practical Case Example of City Destinations. In Sigala, M., Christou, E., & Gretzel, U. (Eds.), ***Social media in travel, tourism and hospitality: Theory, practice and cases***. Ashgate Publishing, Ltd.

Suh, B., Hong, L., Pirolli, P., and Chi, E. H. (2010). Want to be retweeted? large scale analytics on factors impacting retweet in Twitter network. ***In 2010 IEEE 2nd International Conference on Social Computing, pages 177–184. IEEE***.

Tang, T., Fang, E. and Wang, F. 2014. Is Neutral Really Neutral? The Effects of Neutral User-Generated Content on Product Sales. ***Journal of Marketing***, 41 (78): 41– 58.

TripAdvisor. 2016. Everything you need to know about the TripAdvisor Popularity Ranking Algorithm. [Online] Available at: https://www.tripadvisor.com/TripAdvisorInsights/n2701/everything-you-need-know-about-tripadvisor-popularity-ranking-algorithm [Accessed 27 March 2017].

TripAdvisor. 2017a. Everything you need to know about the TripAdvisor Popularity Ranking Algorithm. [Online] Available at: https://www.tripadvisor.com/TravelersChoice-Destinations [Accessed 28 March 2017].

TripAdvisor. 2017b. Everything you need to know about the TripAdvisor Popularity Ranking Algorithm. [Online] Available at: https://www.tripadvisor.com/Attractions-g60763-Activities-New_York_City_New_York.html [Accessed 28 March 2017].

TripAdvisor. 2017c. Everything you need to know about the TripAdvisor Popularity Ranking Algorithm. [Online] Available at https://www.tripadvisor.com/Attractions-g186338-Activities-London_England.html [Accessed 27 July 2017].

UNWTO. 2016. Why Tourism? http://www2.unwto.org/content/why-tourism [Accessed 26 October 2016].

Valdivia, A., Luzon, V., Herrera, M., Francisco. 2017. Sentiment Analysis in TripAdvisor. ***IEEE Intelligent Systems***, (32): 72-77.

Wang, Y., Chan, S.C., & Ngai, G. 2012. Applicability of Demographic Recommender System to Tourist Attractions: A Case Study on TripAdvisor. ***IEEE International Conferences on Web Intelligence and Intelligent Agent Technology.***

Wong, M., Tan, F., Sen, C., Chiang S, Mung. (2016). Quantifying Political Leaning from Tweets, Retweets, and Retweeters. ***IEEE Transactions on Knowledge and Data Engineering***, 28: 1-1.

Xiang, Z. & Gretzel, U. 2010. Role of Social Media in Online Travel Information Search. ***Tourism Management,*** 31:179-188.

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.

Yoo, K.H. & Gretzel, U. 2012. Use and Creation of Social Media by Travelers. In Sigala, M., Christou, E., & Gretzel, U. (Eds.), ***Social media in travel, tourism and hospitality: Theory, practice and cases***. Ashgate Publishing, Ltd.

Yooo, K.H. & Gretzel, U. 2008. What Motivates Consumers to Write Online Travel Reviews? ***Information Technology and Tourism***, 10(4): 283-295.

Zeng, B. & Gerritsen, R. 2014. What do We Know About Social Media in Tourism? A Review. ***Tourism Management Perspectives,*** 10: 27-36.

Zhou, Z. a. 2011. Information resonance on Twitter: watching Iran. ***Proceedings of the first workshop on social media analytics***, pp. 123-131.

APPENDICES

Descriptive Statistics

Appendix 1: Crosstab (Retweet rate, Popularity of Location) SPSS Output

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Retweet_Dummy * Popular	4000	100.0%	0	0.0%	4000	100.0%

Retweet_Dummy * Popular Crosstabulation

			Popular		Total
			More Popular	Less Popular	
Retweet_Dummy	Not Retweeted	Count	1061	971	2032
		% within Retweet_Dummy	52.2%	47.8%	100.0%
		% within Popular	53.1%	48.6%	50.8%
		% of Total	26.5%	24.3%	50.8%
	Retweeted	Count	939	1029	1968
		% within Retweet_Dummy	47.7%	52.3%	100.0%
		% within Popular	46.9%	51.4%	49.2%
		% of Total	23.5%	25.7%	49.2%
Total		Count	2000	2000	4000
		% within Retweet_Dummy	50.0%	50.0%	100.0%
		% within Popular	100.0%	100.0%	100.0%
		% of Total	50.0%	50.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	8.102[a]	1	.004		
Continuity Correction[b]	7.923	1	.005		
Likelihood Ratio	8.105	1	.004		
Fisher's Exact Test				.005	.002
Linear-by-Linear Association	8.100	1	.004		
N of Valid Cases	4000				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 984.00.

b. Computed only for a 2x2 table

Appendix 2: Crosstab (Like rate, Popularity of Location) SPSS Output

Likes_Dummy * Popular Crosstabulation

			Popular		Total
			More Popular	Less Popular	
Likes_Dummy	Not Liked	Count	1711	1757	3468
		% within Likes_Dummy	49.3%	50.7%	100.0%
		% within Popular	85.6%	87.9%	86.7%
		% of Total	42.8%	43.9%	86.7%
	Liked	Count	289	243	532
		% within Likes_Dummy	54.3%	45.7%	100.0%
		% within Popular	14.5%	12.2%	13.3%
		% of Total	7.2%	6.1%	13.3%
Total		Count	2000	2000	4000
		% within Likes_Dummy	50.0%	50.0%	100.0%
		% within Popular	100.0%	100.0%	100.0%
		% of Total	50.0%	50.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4.588[a]	1	.032		
Continuity Correction[b]	4.390	1	.036		
Likelihood Ratio	4.593	1	.032		
Fisher's Exact Test				.036	.018
Linear-by-Linear Association	4.586	1	.032		
N of Valid Cases	4000				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 266.00.

b. Computed only for a 2x2 table

Appendix 3: Crosstab (Sentiment Polarity, Popularity of Location) SPSS Output

Sentiment Polarity * Popular Crosstabulation

			Popular		
			More Popular	Less Popular	Total
Sentiment Polarity	positive	Count	689	1137	1826
		% within Sentiment Polarity	37.7%	62.3%	100.0%
		% within Popular	34.4%	56.9%	45.7%
		% of Total	17.2%	28.4%	45.7%
	neutral	Count	1154	674	1828
		% within Sentiment Polarity	63.1%	36.9%	100.0%
		% within Popular	57.7%	33.7%	45.7%
		% of Total	28.8%	16.9%	45.7%
	negative	Count	127	167	294
		% within Sentiment Polarity	43.2%	56.8%	100.0%
		% within Popular	6.4%	8.4%	7.4%
		% of Total	3.2%	4.2%	7.4%
	mixed	Count	30	22	52
		% within Sentiment Polarity	57.7%	42.3%	100.0%
		% within Popular	1.5%	1.1%	1.3%
		% of Total	0.8%	0.5%	1.3%
Total		Count	2000	2000	4000
		% within Sentiment Polarity	50.0%	50.0%	100.0%
		% within Popular	100.0%	100.0%	100.0%
		% of Total	50.0%	50.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	242.627[a]	3	.000
Likelihood Ratio	245.269	3	.000
Linear-by-Linear Association	98.851	1	.000
N of Valid Cases	4000		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 26.00.

Empirical Testing:

Appendix 4: Hypothesis 1 SPSS Output

Descriptives

Retweets

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum	Between-Component Variance
More Popular		2000	106.607	1320.7996	29.5340	48.686	164.528	.0	22106.0	
Less Popular		2000	26.074	342.3695	7.6556	11.060	41.087	.0	7656.0	
Total		4000	66.340	965.5326	15.2664	36.410	96.271	.0	22106.0	
Model	Fixed Effects			964.8131	15.2550	36.432	96.249			
	Random Effects				40.2668	-445.297	577.978			2777.3902

Test of Homogeneity of Variances

Retweets

Levene Statistic	df1	df2	Sig.
25.375	1	3998	.000

ANOVA

Retweets

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6485644.62	1	6485644.62	6.967	.008
Within Groups	3.722E+9	3998	930864.250		
Total	3.728E+9	3999			

Robust Tests of Equality of Means

Retweets

	Statistic[a]	df1	df2	Sig.
Welch	6.967	1	2266.426	.008

a. Asymptotically F distributed.

Appendix 5: Hypothesis 2 SPSS Output

Appendix 5.1: Test for normality

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Retweets	.473	4000	.000	.039	4000	.000
Followers	.458	4000	.000	.061	4000	.000

a. Lilliefors Significance Correction

Appendix 5.2: Correlation More Popular locations sample:

Nonparametric Correlations

Correlations

			Retweets	Followers
Kendall's tau_b	Retweets	Correlation Coefficient	1.000	.049**
		Sig. (2-tailed)	.	.003
		N	2000	2000
	Followers	Correlation Coefficient	.049**	1.000
		Sig. (2-tailed)	.003	.
		N	2000	2000

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix 5.3: Correlation For less Popular Locations sample:

Nonparametric Correlations

Correlations

			Retweets	Followers
Kendall's tau_b	Retweets	Correlation Coefficient	1.000	.075**
		Sig. (2-tailed)	.	.000
		N	2000	2000
	Followers	Correlation Coefficient	.075**	1.000
		Sig. (2-tailed)	.000	.
		N	2000	2000

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix 6: Hypothesis 3 SPSS Output

Descriptives

SentimentScore

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	Between-Component Variance
						Lower Bound	Upper Bound			
More Popular		2000	.183216357	.375811882	.008403409	.166735999	.199696715	-1.8463016	2.25000000	
Less Popular		2000	.283430193	.345034554	.007715207	.268299504	.298560883	-.92309022	1.80000007	
Total		4000	.233323275	.364170969	.005758049	.222034290	.244612260	-1.8463016	2.25000000	
Model	Fixed Effects			.360751587	.005703983	.222140287	.244506263			
	Random Effects				.050106918	-.40334549	.869992036			.004956336

Test of Homogeneity of Variances

SentimentScore

Levene Statistic	df1	df2	Sig.
.027	1	3998	.869

ANOVA

SentimentScore

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10.043	1	10.043	77.168	.000
Within Groups	520.307	3998	.130		
Total	530.349	3999			

Robust Tests of Equality of Means

SentimentScore

	Statistic[a]	df1	df2	Sig.
Welch	77.168	1	3969.163	.000

a. Asymptotically F distributed.

Appendix 7: Hypothesis 4 SPSS Output

Appendix 7.1: ANOVA More Popular locations sample:

Descriptives

Retweets

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	Between-Component Variance
						Lower Bound	Upper Bound			
positive		689	9.469	27.7194	1.0560	7.395	11.542	.0	200.0	
neutral		1154	174.318	1735.0136	51.0740	74.110	274.526	.0	22106.0	
negative		127	43.055	161.1604	14.3007	14.754	71.356	.0	752.0	
mixed		30	1.967	3.2641	.5959	.748	3.185	.0	15.0	
Total		2000	106.607	1320.7996	29.5340	48.686	164.528	.0	22106.0	
Model	Fixed Effects			1319.3954	29.5026	48.748	164.466			
	Random Effects				63.0466	-94.035	307.249			6810.0339

Test of Homogeneity of Variances

Retweets

Levene Statistic	df1	df2	Sig.
9.156	3	1996	.000

ANOVA

Retweets

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	12633547.7	3	4211182.55	2.419	.064
Within Groups	3.475E+9	1996	1740804.18		
Total	3.487E+9	1999			

Robust Tests of Equality of Means

Retweets

	Statistic[a]	df1	df2	Sig.
Welch	18.926	3	429.209	.000

a. Asymptotically F distributed.

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Retweets

	(I) Sentiment Polarity	(J) Sentiment Polarity	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tukey HSD	positive	neutral	-164.8492*	63.5221	.047	-328.177	-1.521
		negative	-33.5863	127.4115	.994	-361.186	294.014
		mixed	7.5021	246.0759	1.000	-625.208	640.212
	neutral	positive	164.8492*	63.5221	.047	1.521	328.177
		negative	131.2629	123.3516	.711	-185.898	448.424
		mixed	172.3514	243.9986	.895	-455.017	799.720
	negative	positive	33.5863	127.4115	.994	-294.014	361.186
		neutral	-131.2629	123.3516	.711	-448.424	185.898
		mixed	41.0885	267.8319	.999	-647.560	729.737
	mixed	positive	-7.5021	246.0759	1.000	-640.212	625.208
		neutral	-172.3514	243.9986	.895	-799.720	455.017
		negative	-41.0885	267.8319	.999	-729.737	647.560
Games-Howell	positive	neutral	-164.8492*	51.0849	.007	-296.279	-33.419
		negative	-33.5863	14.3396	.094	-70.916	3.744
		mixed	7.5021*	1.2126	.000	4.372	10.632
	neutral	positive	164.8492*	51.0849	.007	33.419	296.279
		negative	131.2629	53.0383	.064	-5.175	267.701
		mixed	172.3514*	51.0775	.004	40.940	303.763
	negative	positive	33.5863	14.3396	.094	-3.744	70.916
		neutral	-131.2629	53.0383	.064	-267.701	5.175
		mixed	41.0885*	14.3131	.024	3.824	78.353
	mixed	positive	-7.5021*	1.2126	.000	-10.632	-4.372
		neutral	-172.3514*	51.0775	.004	-303.763	-40.940
		negative	-41.0885*	14.3131	.024	-78.353	-3.824

*. The mean difference is significant at the 0.05 level.

Appendix 7.2: ANOVA Less Popular locations sample:

Descriptives

Retweets

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	Between-Component Variance
						Lower Bound	Upper Bound			
positive		1137	37.831	453.0793	13.4367	11.468	64.195	.0	7656.0	
neutral		674	3.616	10.7449	.4139	2.803	4.428	.0	106.0	
negative		167	40.054	54.6362	4.2279	31.707	48.401	.0	123.0	
mixed		22	.318	.5679	.1211	.066	.570	.0	2.0	
Total		2000	26.074	342.3695	7.6556	11.060	41.087	.0	7656.0	
Model	Fixed Effects			342.2285	7.6525	11.066	41.081			
	Random Effects				11.6435	-10.981	63.128			173.5039

Test of Homogeneity of Variances

Retweets

Levene Statistic	df1	df2	Sig.
4.970	3	1996	.002

ANOVA

Retweets

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	544347.857	3	181449.286	1.549	.200
Within Groups	233772234	1996	117120.358		
Total	234316582	1999			

Robust Tests of Equality of Means

Retweets

	Statistic[a]	df1	df2	Sig.
Welch	50.973	3	591.164	.000

a. Asymptotically F distributed.

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Retweets

	(I) Sentiment Polarity	(J) Sentiment Polarity	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	positive	neutral	34.2154	16.6366	.168	-8.561	76.991
		negative	-2.2228	28.3607	1.000	-75.144	70.698
		mixed	37.5130	73.6659	.957	-151.897	226.922
	neutral	positive	-34.2154	16.6366	.168	-76.991	8.561
		negative	-36.4382	29.5819	.607	-112.499	39.623
		mixed	3.2975	74.1446	1.000	-187.343	193.938
	negative	positive	2.2228	28.3607	1.000	-70.698	75.144
		neutral	36.4382	29.5819	.607	-39.623	112.499
		mixed	39.7357	77.6207	.956	-159.842	239.314
	mixed	positive	-37.5130	73.6659	.957	-226.922	151.897
		neutral	-3.2975	74.1446	1.000	-193.938	187.343
		negative	-39.7357	77.6207	.956	-239.314	159.842
Games-Howell	positive	neutral	34.2154	13.4431	.054	-.371	68.802
		negative	-2.2228	14.0862	.999	-38.458	34.012
		mixed	37.5130*	13.4373	.027	2.941	72.085
	neutral	positive	-34.2154	13.4431	.054	-68.802	.371
		negative	-36.4382*	4.2481	.000	-47.461	-25.415
		mixed	3.2975*	.4312	.000	2.187	4.408
	negative	positive	2.2228	14.0862	.999	-34.012	38.458
		neutral	36.4382*	4.2481	.000	25.415	47.461
		mixed	39.7357*	4.2296	.000	28.759	50.712
	mixed	positive	-37.5130*	13.4373	.027	-72.085	-2.941
		neutral	-3.2975*	.4312	.000	-4.408	-2.187
		negative	-39.7357*	4.2296	.000	-50.712	-28.759

*. The mean difference is significant at the 0.05 level.