THE PREDICTIVE POWER OF DIGITAL TRENDS ON PREDICTIVE ANALYTICS IN THE TOURISM INDUSTRY

BACHELOR THESIS

HARMANAN SINGH PAHWA

STUDENT NUMBER: 430711

TABLE OF CONTENTS

I. ABSTRACT

With the rise in online searches in the tourism industry, the role for effective online marketing strategies has grown in importance. Businesses can make use of consumer trends data to establish meaningful inferences about the usage of their product offerings. This paper identifies the role of Google search trends data to observe if certain queries can predict total number of visitors to the United Kingdom and more specifically to the "British Museum" based in London. The study finds significant predictive power of such search terms on the actual number of visitors. The paper also makes suggestions on how organisations operating within the tourism industry can leverage social platforms to capture trends.

II. INTRODUCTION

With the advent of digital technologies, Customers have now been empowered to conduct a search, evaluate alternatives and make decisions in their purchase journey. A 'Customer Purchase Journey' represents the steps taken by a customer en-route to finally purchasing a product or using a service (Edelman & Singer, 2015).

The various aspects of the Customer journey include awareness where the customer first learns about their need, followed by a search for solutions, gathering information and finally making a choice from all the alternatives available in the market. Companies have long used techniques and methods to understand the preceding journeys of their potential customer to try to capture their interest and provide for better marketing early in the journey (Lemon & Verhoef, 2016).

With the emergence of new search tools, there is a need to explore the scope of predicting consumer trends and behaviours using their digital footprints. Google currently stands out as the world's most used search engine drawing consolidating over 85% market share amongst search engines.

Google trends is a tool that contains search query based time series data that highlights key interests and trends amongst Google users. The query share of a search term is the query volume of a search term divided by the total volume of all queries in a particular geographical location. The queries are matched to ensure that similar terms can fall under the same query name such as "buy car" would also count as a search for the term, "car". If an increasing volume of searches for a particular term is noted, it could correspond to the increasing interest towards the term over a period of time. Google trends captures this data from 2004 onwards and can be sorted to yearly, monthly, weekly, daily and even hourly data that can allow us to identify popular times and seasonality induced effects.

A key to identify whether trends can be used to monitor future sales is also determined by whether the good in question is a search good or an experience good. A search good is a good where potential customers try to review the product characteristics such as quality or price before the purchase and evaluate the utility they would derive from it. Other products are classified as experience goods where the attributes cannot be inferred before the product has been used, hence on such products potential customers may not do a search before the product purchase.

This research paper shall focus on the predictive power of Google Searches on the tourism industry. The United Kingdom saw over 39 million overseas visitors in 2017 (the Office for National Statistics) and shall be the basis of this research. The Office for National Statistics also keeps a record of monthly visits to all National Museums in the United Kingdom, the paper shall observe statistics from "The British Museum" which is by far the most visited museum in the United Kingdom.

The research paper shall also briefly discuss the relevance of the research topic by presenting past academic research. It would be followed by a description of the theoretical framework and methodology used to identify the relationship between the explanatory and response variables. Based on the identified time series model, a regression would list the significance and predictive power of the relationship. Finally, the paper would draw conclusions that are relevant for businesses and initiatives related to the travel/tourism industry. While Google Trends data can potentially provide some predictive power, firms can empower themselves with the right tools

such as TripAdvisor rating and Twitter Analytics to make the right inferences. The scope for Search Engine Marketing is also discussed as it looks at reaching the right target audience based on search history and interests. The results and conclusions shall provide for optimal marketing strategy suggestions based on predictions.

III. LITERATURE REVIEW

Researchers have long tried to predict trends in the travel industry. One such paper by (Lohmann & Danielsson, 2001) looks at how different generations and age groups present different travel behaviours. Observing survey data from Germany, it was believed that as people turn older, their travel tastes and frequency changes, while this was true, they found that the effect of generational effects was far more significant. Travellers who were 60 now seemed to display similar preferences as they did when they were 50.

In order to further these studies, a paper (Burger, Dohnal, Kathrada & Law, 2001) looked at the various time series methods to measure and predict the travels of American tourists to Durban, South Africa. The various methods deployed were ARIMA, moving average, multiple regression and Neural Networks. It also pointed at the unprecedented existence of external effects such as strikes, economic fissures or disease outbreaks. Time series based on Neural networks where input nodes are activated to the output layers by means of gradient descent was identified as the best performing method.

With the advent of the internet, it became important to also look at information online as a means to manage tourism. Information search is a crucial component of the Customer Journey and the subsequent purchase decisions, it strives to not filter out unfavourable choices and in the process, ensures better trip experience. For service providers, it is an opportunity to tap into customer needs; as an informed customer is better able to communicate with available resources. The Internet has also allowed for new marketing channels and avenues for less prominent destinations/experiences to be adequately represented digitally (Buhalis & Law, 2008).

There has been considerable research on how Google trends and other digital data collection tools can be used to predict trends. (Choi & Varian, 2012) list how Google trends data can predict the present as far as aggregate data on economic indicators and travel trends are concerned, due to in-sample forecasting however, it does not look at future trends. Economic indicators' data is often released with a lag, leading to an inability of policymakers to respond (Carrière-Swallow & Labbé, 2011). This study focused on the Chilean automobile market and showcased that an ARMA model using Google search queries as explanatory variables outperformed previous baseline models considerably. The fact that Google search queries are aggregated from a larger demographic, it is more consolidated than data inferred from survey techniques that are prone to biases. Also, (Carneiro & Mylonakis, 2009) had a rather interesting application of trends data to monitor the outbreak of diseases. While the paper mentions that it is best used in conjunction with traditional tool used for surveillance, it identifies trends data as a quicker and broader tool for the identification of diseases.

IV. THEORETICAL FRAMEWORK

i. Research Question

"To what extent do Google search query trends explain and predict monthly number visitors to the United Kingdom and the British Museum in subsequent periods"

Based on the explanatory power of Google trends, firms and agencies can identify the right times and periods to deploy their marketing efforts. The tourism industry is a combination of several services and products such as accommodations, car rentals, tours or flights. Such services have been at the forefront of the digital customer search journey and have been subject to extensive marketing and promotional efforts ever since the early 90s. In recent times, products that were planned at a shorter notice such as museums or events have become quite relevant as far the market dynamics of supply and demand are concerned (Xiang, Magnini & Fesenmaier, 2015). Such "secondary" products once used to be a smaller part of a tourism package, but now have the

ability to diversify consumer demands and are very relevant as far as Search Engine Marketing is concerned. The research question hence focuses on dual aspects, The United Kingdom as a travel destination as well as something more specific such as the visits to the British Museum.

ii. Hypothesis 1

"The volume of Google trends searches has a predictive power on future monthly visits of tourists to the United Kingdom."

To study this hypothesis, the paper shall make use of 5 different explanatory variables and 2 control variables. Google trends allows us to search for different search terms as belonging to one of the many search terms such as "Flights London" or "Hotels UK". "Visit Britain" is the official website of tourism and hence has also been included in the list of variables. The paper restricts the number of explanatory variables to 5 to avoid overfitting.

While it is expected that the search terms' volume grows over a period of time due to increasing accessibility and usage of the internet globally, the Google trends data controls for this by showcasing relative trends with contrast to absolute trends and thus accounts for overarching external effects. Seasonality effects together with Major events are categorical control variables that can explain spikes in the response variable.

DATA TYPE	VARIABLE CODE	VARIABLE
Google Search Query	air	Flights london
Google Search Query	hot	Hotels UK
Google Search Query	guid	UK travel (Youtube)
Google Search Query	dest	United Kingdom
Google Search Query	vb	Visit Britain
Control Variable	event	Events
Control Variable	season	Season

iii. Hypothesis 2

"The volume of Google trends searches has a predictive power on future monthly visits to the British Museum."

Amongst the five explanatory variables are general search terms such as "Things to do in London", the search for information on specific web pages such as "TripAdvisor London" or "Visit London" as these search terms provide the visitor with information on the "British Museum" as a place of interest. Such search terms allow for Information search where the visitor does not exactly know where they want to visit but are rather looking for points of interests that they could visit.

DATA TYPE	VARIABLE CODE	VARIABLE
Google Search Query	bm	British Museum
Google Search Query	ttd	Things to do in London
Google Search Query	lm	London Museums
Google Search Query	tal	TripAdvisor London
Google Search Query	vl	Visit London
Control Variable	event	Events
Control Variable	season	Season

The variables, "Events" and "Season" are both categorical variables. "Events" marks months where a major event such as The London Olympic Games or the Brexit referendum took place whereas "Season" segregates the year into quarters of Spring (February-April), Summer (May-July), Autumn (August-October), Winter (November-January). Weather can play a major influencing factor on tourism. Every major event that could play a significant role in determining the tourism numbers has been added. If any of these events occur during a particular month then the month gets a value "1" while all other months are categorised as the "0" group.

For the first and the second hypothesis, an ARDL model shall be used to conduct the research.

$$Y_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots\ldots\ldots + \beta_i X_{t-i} + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + \ldots\ldots\ldots + \delta_j Y_{t-j} + \varepsilon$$

The response variable in the ARDL model not just depends on the lags of the explanatory variable but also on its own past values. The number of lags and the exact model will be decided based on the Dickey Fuller test as described later in this paper.

In linear equation terms this corresponds to the equation of type: **y = mx + c**,

where y is the response variable that changes based on the "x", the explanatory variable. The explanatory power of "x" is defined by the slope "m". The intercept of "y" is the intrinsic value of y when x=0, this is defined by "c".

To test for hypothesis 1, the following equations are drawn.

$$act\_vis_t = \beta_0 + \beta_1 * air + \beta_2 * hot + \beta_3 * guid + \beta_4 * dest + \beta_5 * vb + \beta_6 * event + \beta_7 * season + \gamma_1 * act\_vis_{t-12} + \varepsilon_0$$

$$act\_vis_t = \beta_0 + \beta_1 * air + \beta_2 * hot + \beta_3 * guid + \beta_4 * dest + \beta_5 * vb + \beta_6 * event + \beta_7 * season + \gamma_1 * act\_vis_{t-1} + \varepsilon_0$$

For hypothesis 2, the equation obtained is as follows:

$$act\_vis_t = \beta_0 + \beta_1 * bm + \beta_2 * ttd + \beta_3 * lm + \beta_4 * d\_tal + \beta_5 * bm + \beta_6 * event + \beta_7 * season + \gamma_1 * act\_vis_{t-12} + \varepsilon_0$$

In these equations, the act_vist at time t takes the role of the y value in the current time periods or for predicting act_vis values in the future periods.

V. DATA & METHODOLOGY

Google trends data is publicly available online. The data for the monthly visits to the United Kingdom and the monthly visits to "The British Museum" are published by the National Office of Statistics. For preliminary research, the descriptive statistics shall be presented to gain a better idea of the distributions and scales. Google trends data is normalised to exclude the effects of factors such as the rising accessibility of internet to the population.

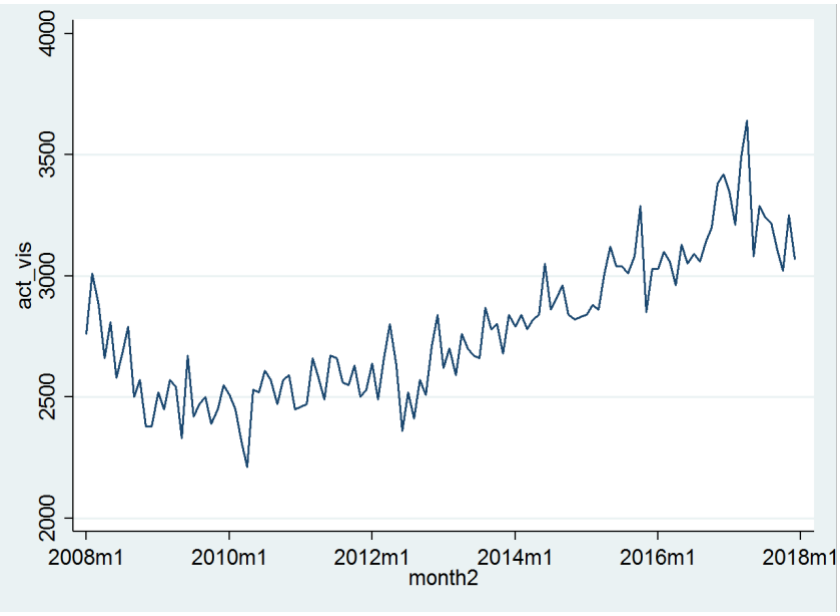


Fig 1. Monthly Visitors to the United Kingdom (2008 - present)

The United Kingdom has seen an upward trend in incoming visitors over the past decade with only 2010 seeing a drop compared to the previous year. The fig 1. shows that these trends can be seasonal, with spikes seen at regular intervals. The UK sees an average of 2, 700, 000 visitors a month (Table 1). With a minimum of 2.2 mn to a maximum of 3.6 mn, the range is seen as quite large and variable, again pointing to the seasonality effects mentioned previously. Amongst the explanatory variables, it is quite evident that the search term "United Kingdom" is the most popular search term with a mean of over 72 and the lowest standard deviation amongst all other variables. The search query here is very broad and captures a large portion of people searching for news or other information about the United Kingdom and hence as expected shows a considerably larger search volume. This in stark contrast to "UK Travel" a search query that is

specific to searches on YouTube, Google's acquired video sharing platform, that shows a mean volume of roughly 31. Due to the differences in popularity of terms, the research allows to check for a more diversified set of terms.

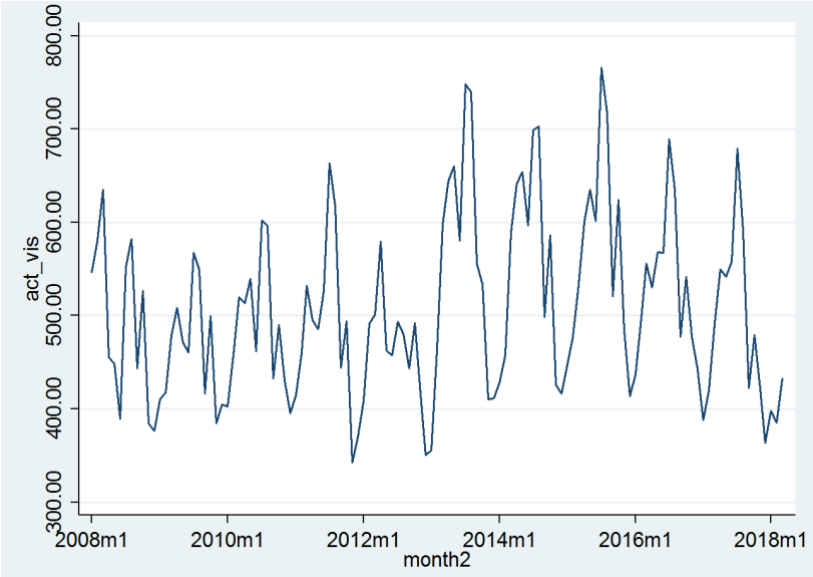


Fig 2. Monthly Visitors to the British Museum.

The British Museum sees a higher standard deviation between yearly maximum and minimum values post 2014 as compared to before 2014. The effects of seasonality are seemingly more pronounced in visits to the British Museum than the visits to the United Kingdom (Fig 2). The British Museum based in London sees the highest footfall in terms of visitors amongst all museums and galleries in the United Kingdom with over 500,000 visitors every month on average (Table 2). The minimum and maximum values vary considerably with a range of 342,000 to 756,000 which is also reflected by the standard deviation of the same. The search query, "Things to do in London" has the highest mean amongst other search queries, this along with "TripAdvisor London" and "Visit London" are general queries that enable visitors to explore tourist spots in London. It needs to be noted then that if the British Museum is the top result to these queries or is rated highly as a destination, it could receive more potential visitors.

For the hypotheses mentioned previously, the research shall make use of time series analysis. At first the Dickey-Fuller test for unit roots shall be tested to observe any stationarity within the

variables. It is important to check for this as non-stationarity often leads to spurious regressions, based on the assumptions of linear regression.

It is also possible that the effect of Google trend searches effects the potential visits to Museums and countries in subsequent periods. It is crucial to measure the lags in such cases. The time series is expected to be of Autoregressive Distributed Lag type (ARDL), where the number of visitors depends on previous visitor numbers and simultaneously also on the Google search trends. To determine the number of lags, a regression analysis will be performed and the optimal number would be chosen based on the Schwarz criterion. Naturally, every lag measurement will see the number of observations drop by one, however to measure accurately it is imperative to have the lowest amount of observations as the parameter for comparison.

VI. RESULTS

To check for non-stationarity and ensure that the linear regression is not spurious, the Dickey-Fuller test for unit roots is performed for all the variables. The statement to be tested here is "*There are no unit roots in the series. The series is stationary*".

It was identified that all the "Google search query" variables pertaining to the first hypothesis are indeed stationary as the p-values for all the search terms were higher than the significance level of 0.05%. The null hypothesis that the there are no unit roots is not rejected. Amongst the variables to tests the second hypothesis, all variables except "TripAdvisor London" also had p-values over 0.05, hence these variables were taken as stationary. The null hypothesis that there are no unit roots in the variables is rejected. To make "TripAdvisor London" stationary, the time series had to be detrended by taking the differences in lags until the differenced variable becomes stationary, on testing for this it is seen that the TripAdvisor London becomes stationary after the first differencing.

To build an ARDL model it is also imperative to know how the response variables' current values correspond to its values in the past years. In this study, the response variables are visits to

the United Kingdom and visits to the British Museum. This can be identified by performing autocorrelations on the values.

Based on Table 3 and 4, it is evident that the auto correlated values are all significant as a consequence of the p-values being lower than 0.05, hence the null hypothesis that the variables are not auto correlated is rejected. Based on the partial autocorrelation values, it can also be seen that "Annual visits to London" are auto correlated with their values until the 3rd lag as after this there are no peaks in the values. Hence it can be said that the values are correlated with one another for 3 quarters. For the variable, "Total annual visits to the British Museum", the partial autocorrelations show peaks only for the first lag (or until the first quarter). The autocorrelation table is hence able to show that the past values of the response variable indeed influence the current variables. The current number of visitors to the United Kingdom are influenced by seasonal factors, hence it is intuitive that the number of visitors in January every year shall have similar values.

Before a time series regression is performed, it is also imperative to look at the number of lags that are most suitable as far as the relationship between the response and control variables is concerned. For the ARDL models, the paper shall look at Distributed Lags from both last period and twelve periods ago and for each of these select the most optimal number of lags for the explanatory variables. The BIC values for both situations are mentioned in Table 5.

Table 5: BIC values for the variables pertaining to hypothesis 1.

Number of Lags	0	1	2	3
BIC (L.1)	1555.397	1549.704	1537.435	1523.465
BIC (L.12)	1443.189	1445.417	1448.727	1440.138

The BIC values are observed for only 3 lags as there is an assumption that the search for flights and hotels pertaining to a destination are usually not searched at a point exceeding 3 months from the actual date of travel. Due to the large variation in time that people actually start

searching for such terms on Google, the paper shall also look at the time that has the largest predictive power. Based on the BIC values, the regression shall proceed with taking 3 period lags for both the cases. The regression values are stipulated as in Table 6 and 7.

Table 6: Time series regression for predicting the power of Google Search trends on actual visits to the United Kingdom with a DL of t-12.

		act_vis
air	L1.	5.435
	L2.	1.92
	L3.	11.304*
hot	L1.	6.357*
	L2.	0.241
	L3.	8.67*
guid	L1.	2.364*
	L2.	3.250*
	L3.	2.754*
dest	L1.	-1.193
	L2.	4.831*
	L3.	3.178
vb	L1.	0.901
	L2.	-1.383
	L3.	0.268
event	0	0
	1	96.11**
season	0	0
	1	-0.103**
	2	0.550
	3	-0.778
act_vis (historical)	act_vis (t-12)	0.331*
_cons		2176.744*
N		108

For the time series regression taking the lag of actual visits to twelve periods ago, it can be specified that the search terms marked with an asterix (*) are significant at a significance level of 0.05 (values marked with ** are significant at the 0.10 level) Hence the null hypothesis that none of the explanatory powers, i.e. none between β_1, β_2, β_3 , β_4 or β_5 are greater than 0 is rejected.

The ways in which people use search engines has been studied extensively to identify consumer behaviours. Based on a paper by Xiang & Pan (2011), the search engines are used in three steps. Users start with entering a search query, based on these queries, search engines give results that are a close match to the query entered. Finally, the user interacts with these results by reading their titles or by viewing them. From a consumer journey perspective, this means that the user first identifies keywords based on experience and intuitiveness and is led towards making a series of decisions based on this presentation of information.

As expected people searching for flights and hotels to a particular destination usually have a longer search journey due to the generally increasing costs with the lapsing of time. For tourism purposes, visitors seem to explore places to visit and explore within the United Kingdom closer to the visit. The YouTube search term, "UK travel" is significant in its predictive power with a lag of one period while the general search term "United Kingdom" also shows significance on the current number of visitors for searches made 2 periods prior. The official website operated by the tourism board of the United Kingdom does not show a significant correlation with the actual number of visitors arriving in the United Kingdom.

Besides, 12 periods ago values of actual visitors significantly determine the current number of visitors as well.

Table 7: Time series regression for predicting the power of Google Search trends on actual visit to the United Kingdom with a DL of t-1.

		act_vis
air	L1.	5.857**
	L2.	4.459*
	L3.	6.969*
hot	L1.	5.966*
	L2.	0.337
	L3.	7.17*
guid	L1.	1.882*
	L2.	1.855**
	L3.	1.641*
dest	L1.	-0.228
	L2.	4.532*
	L3.	-0.535
vb	L1.	1.691
	L2.	-2.399
	L3.	0.336
event	0	0
	1	5.817
season	0	0
	1	-0.085**
	2	0.243
	3	-0.466
act_vis (historical)	act_vis (t-1)	0.476*
_cons		0.477*
N		108

For the time series regression taking the lag of actual visits to one periods prior, it can again be stated that the null hypothesis that none of the potential predictive powers, i.e. none between β_1, β_2, β_3, β_4 or β_5 are greater than 0 is rejected.

Like in the previous case, search for flights and hotels to a particular destination find correlations with future values of actual visits. The YouTube search term, “UK travel” is yet again significant in its predictive power with a lag of one period whereas the search term “United Kingdom” shows significance on the current number of visitors for searches made 2 periods prior like in the previous case. The search for the “Visit Britain” website on Google does not seem to find any correlation with the actual visits made in the future.

Table 8: BIC values for the variables pertaining to hypothesis 2.

Number of Lags	0	1	2	3
BIC (L.12)	1225.993	1242.313	1259.988	1276.362

For Hypothesis 2, the Optimal number of lags is selected only for 12 month lags of the response variable and in this case the lowest BIC value as per the Schwarz criterion is that of 0 lags. The linear time series regression for the same can be found in Table 9.

As can be observed, the Time series regression in this is seen to have significant values for all search variables except for the search term “Visit London” which is the official website by the Tourism ministry based in London. The search term ‘d_tal’ is the detrended variable for the search term “Trip Advisor London” and while it is significant, is seen to have a negative explanatory power on the total visits to the British Museum. Based on the long tail of searches that tourists usually proceed with in their search journeys, it is interesting to note that a broader search term such as “Things to do in London” has lower explanatory power than a term that is specific to the “British Museum”.

Table 9: Time series regression for predicting the power of Google Search trends on actual visits to the British Museum.

	act_vis
bm	6.120*
ttd	4.541*
lm	1.887*
d_tal	-1.493*
vl	-0.381
L12.act_vis	0.222
0.event	0
1.event	34.58*
1.season	0
2.season	-0308*
3.season	0.382*
4.season	-0.672*
_cons	8.448
N	111

When people access a Search Engine or website for more information, the Long Tail effect is the measurement of the more rare "digital assets" on a web page. For the TripAdvisor website for instance, the "Home" page is visited the most, followed by more specific content such as the "British Museum". While some visitors might decide to directly search for the museum's website, others are introduced to it by means of digital advertisements that appear on webpages based on search history and interests of the browser. While the implications of using Search Engine Marketing with context to digital trends shall be discussed in the next section, it has been noted that people introduced to more specific terms in the Long-Tail are more likely to be intrigued than a more general web page (Xiang & Gretzel, 2010). While, long-tail keywords might be visited more seldom than say the "visit" page of an attraction, collectively the long-tail keywords can have a lasting impact on increasing footfall.

VII. USING SOCIAL PLATFORMS TO INCREASE ENGAGEMENT

While Google trends can be an instrumental predictive tool in forecasting actual visitors, social platforms such as Facebook and Twitter or platforms where services are reviewed such as TripAdvisor can provide insights into such predictive mechanisms and effective use of such tools could potentially increase future visitors.

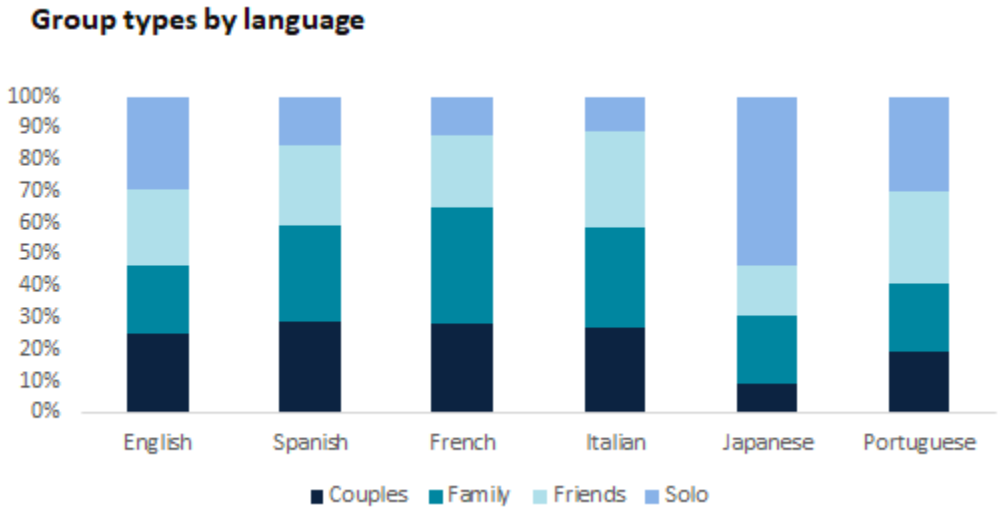
One aspect of this paper was to see how travel related organisations can make predictions based on Search volumes, another aspect is to actually take charge of social platform trends to make people learn about tourism opportunities. If firms are proactive in these domains, they make their services better and also directly reach people who could potentially search them on the Search Engines.

TripAdvisor.com sees over 30 million visitors on its online portals every month, of these 5 million visitors are registered members. It claims to generate over 10 million user generated reviews for over 250,000 travel destinations and hotels per month. It combines the elements of a blog and a social network but maintains that its core role is to collect and project user generated ratings and reviews within the travel industry. Users of this portal can rate their experiences related to visits on a 5-point scale. It also takes surveys to understand consumer dynamics such as their purpose of the trips, whether they were travelling alone and so on.

A study by O'Connor (2008) researched how reviews on hotels can assist managers in improving their services and understand consumer needs. All the hotels taken into consideration had received several reviews and it was seen that this user generated content was viewed and that subsequently had an impact on the purchase decisions made by the consumers.

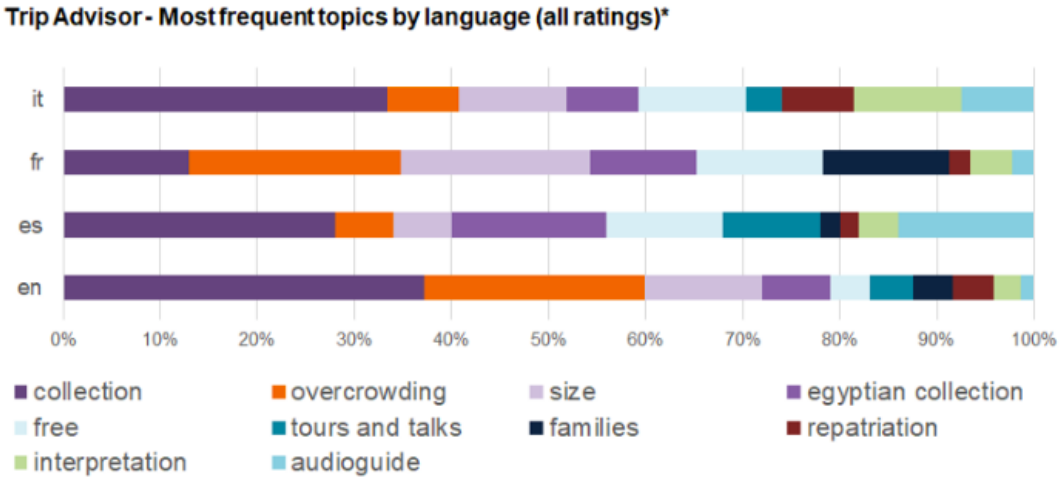
Very specific aspects such as those described in Fig. 3 can enable firms to understand their target audiences and focus on digital marketing.

Fig 3: Type of traveller by language of review (Source: TripAdvisor- The British Museum)



Since visitors to the British Museum come from around the world, their reviews on TripAdvisor can be made in different languages. From a marketing perspective, if the museum was to launch a special family ticket, it could target its Pay-per-Click advertisements to French speakers as they seem to visit the museum with their families the most in contrast to Japanese speaking visitors who are largely solo visitors.

Fig 4: Popular keywords with respect to the British Museum by language of review (Source: TripAdvisor- The British Museum)

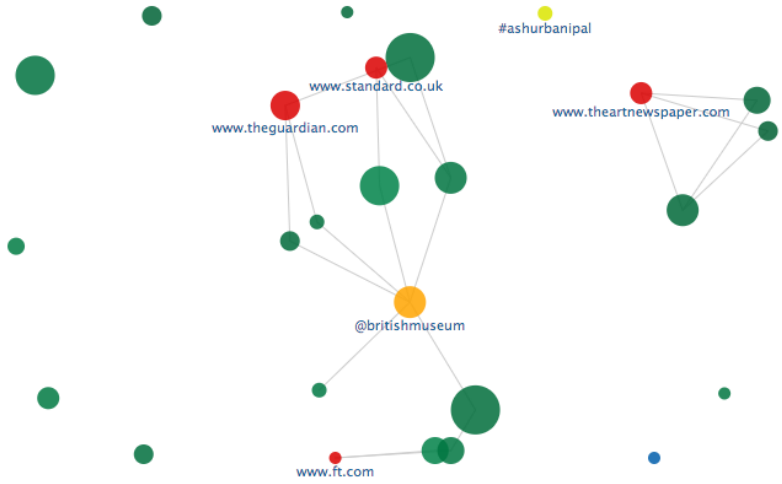


Similarly, greater insights can be derived from the application of Natural Language Processing from reviews. The British Museum's analytics team discovered the most mentioned "buzz words" by speakers of four different languages, i.e. "Italian", "French", "Spanish" and "English". The "family" aspect amongst French visitors is further testified here as one of the main keywords used in their reviews. Spanish speakers for instance also mention 'tours and talks' which provides a ground for developing a Spanish audio guide.

Twitter was classified as a *social awareness stream by (Naaman, Becker and Gravano, 2011).* With Millions of users, Twitter has become a prominent platform for discussions of a wide variety of topics. Twitter in essence allows all users to make messages that are short (up to 140 characters) called Tweets. Twitter enables users to follow other users, hence providing a ground for influencers and institutions to directly broadcast Tweets to a large audience. Tweets that are posted publicly carry a lot of information about public opinion and concerns. This can differ largely with varying geographies and demographic aspects. Trends can be captured on Twitter through "hashtags", #trumpkim, #singapore etc. were popular hashtags in the recently concluded meeting in Singapore between Donald Trump, The President of the United States and Kim Jong Un, The Supreme Leader of the People's Republic of Korea.

The travel industry benefits a lot from referrals and featured posts by prominent internet users and can subsequently influence end consumers to search up information on Search Engines. Using Twitter Sentiment Analysis tools, it is possible to garner real time inferences from how people tweet about a product/service. Fig 5. Highlights prominent words or pages that are most linked to the British Museum in a given period of time.

Fig 5: Key associations and post linkages for the British Museum on Twitter (Date: 19-06-2018)



It seems that a lot of tweets have been made to the handle @britishmuseum through blogposts or featured posts by web pages such as theguardian.com.

Fig 6: Twitter Sentiment for the British Museum (Date: 19-06-2018)

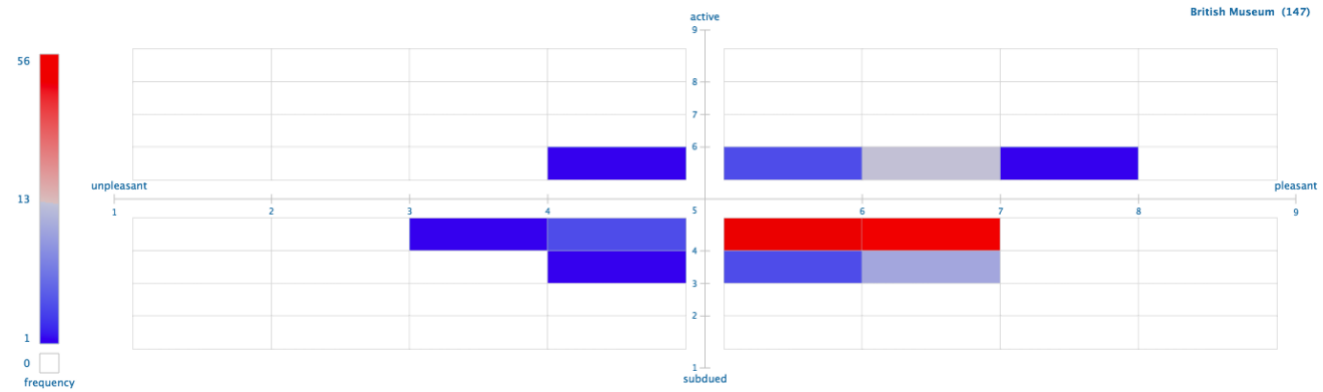
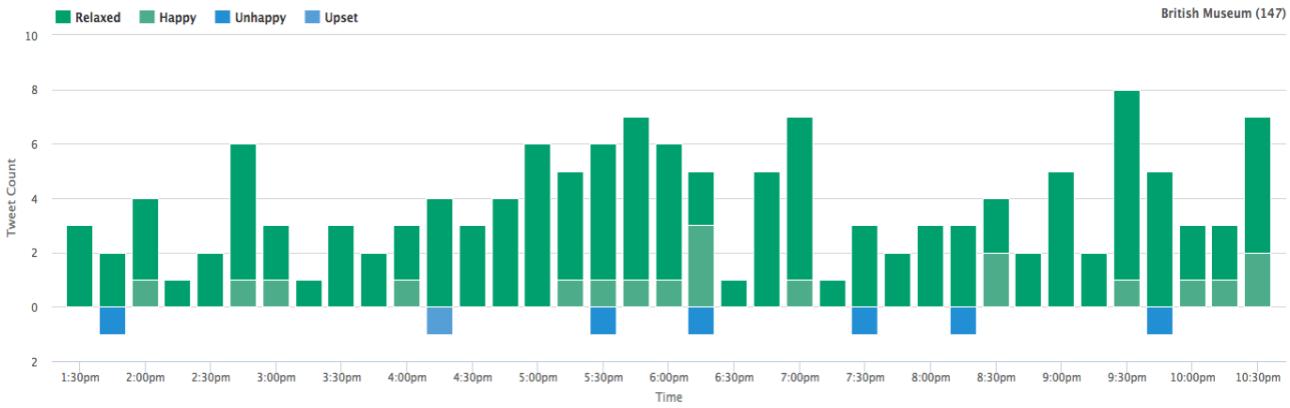


Fig 6. makes use of natural language processing to infer sentiments reflected by tweets made about the British Museum on the 19th of June 2018. The blue cells represent tweet frequency over the median, whereas red represents lower than median. Semantic analysis captures specific words in a tweet such as “bad”, “amazing”, “Interesting” etc. alongside emoticons to place the tweets along the axis.

Fig 7: Twitter Sentiment through the day for the British Museum (Date: 19-06-2018)



Based on the demands of the institution, such semantics can be optimised. In the Fig 7. The most popular times for tweets can also be seen. The British Museum can here choose to make tweets and posts between 17:00 and 18:00 and again at around 21:30 to reach maximum number of viewers.

VIII. IMPLICATIONS

As has been seen, digital trends can predict future footfall for business and institutions operating in the tourism industry. Firms need to capitalise on their 'Search Engine Optimisation' strategy to meet key objectives.

By unlocking search trends, organisations can delve deeper into understanding how people venture into a website in the first place. Webpage traffic can be segmented into a few main types such as direct, referral and organic traffic. Direct visitors are those that are currently aware about their needs and directly type the website name or url in search bars. Organic visitors are those that click on a webpage while searching for products or services that are related to the click. In the example of this study, an organic page visitor is someone who arrives on the British Museum website after searching for "Things to do in London" on the Search Engine. Finally, referral traffic is the traffic that is introduced to a webpage through ads or social networks and may not specifically be on a search journey. Search Engine Optimisation in this regard can enable firms

to find search terms and words where their webpage comes up amongst the first few web pages and higher ranked web pages tend to attract more organic visitors. While appearing as a highly ranked result on a keyword like, “Museums in London” is important, it is also more competitive. Identifying key “Long tail” words and connecting them to specific exhibits for instance can attract a more engaged audience and provide for an optimisation that is unique to the museum.

After the identification of keywords and formulating good content that matches with these keywords, it is crucial to get the right endorsements. Endorsements are based upon the referrals of a website by another website, the more a website is referred by other people or pages, the higher ranks it gets on search results. It is also interesting to note that when the referral is made by another website that is heavily referred itself, the weighted contribution of the referral to the rank is higher. The tourism board or administrators of tourist attractions must try to get featured by social media influencers, bloggers or popular travel to foster more website clicks and in turn boost actual visitor numbers (Ellis, 2017). Google Adwords currently stands as the most prominent platform for search related advertisements. Advertisements are subject to fee schemes such as “Pay per Click” where a payment is made to the placeholder based on the actual number of clicks to the advertisement.

As discussed in the results of this paper, it is evident for places of interest to be engaging on social platforms. Word of mouth or sharing of content on social media is a powerful factor in influencing customer adoption positively. Every tweet or Facebook video adds to the digital landscape and provides an environment conducive to not just feature but alter digital trends.

IX. CONCLUSION AND DISCUSSION

It was determined in this paper that different search terms do provide significant predictive power on the outcomes in the tourism industry. Search Engines provide an opportunity for businesses to market and understand consumer trends. Businesses are increasingly capitalising on Search Engine Optimisation and Search Engine Marketing (SEM) to offer their product and services based on how people search for and obtain information (Xiang & Gretzel, 2010).

For SEM, marketeers try to obtain a high-ranking search result when a query is entered. The competitive landscape does lead to several firms making attempts at gaining the attention of potential consumers, who in their day to day online activity are exposed to a plethora of information. The competition tends to be quite stiff and follows market principles of supply and demand. Online users are increasingly getting more accustomed to narrowing down on the information that appears to be most relevant to them. Search Engines such as Google are now building stronger algorithms to combat irrelevant and spam information by making the ranking order more policed and stringent.

The applications of this study go beyond the travel industry. Based on the specificity of keywords, restaurant owners could look at hourly search trends to predict the number of bookings they would get in the subsequent hours. Firms could hence be better prepared for outcomes. Future research in this regard could look at the predictive applications of search trends across businesses and industries. While this paper chose monthly data due to the availability of actual monthly visits data, there is also scope to study the relevance of search trends for shorter periods such as hourly or weekly data.

Business need to also understand these search queries in a more human context, i.e. they could effectively not just predict business opportunities and turnovers but also truly understand consumer needs. The search queries while being seemingly simple paint a story influenced by several factors including consumer knowledge, experience with using a product/service, stage of decision making etc. (Bing Pan, Zheng Xiang, Law & Fesenmaier, 2010). In industries dealing with a broad customer range, marketeers could observe how people use and search for words to build and develop new product offerings or alter service delivery as deemed efficient.

X. APPENDIX

Table 1: Descriptive Statistics for variables relevant for visits to the United Kingdom.

Actual Visitors	Mean	2776.75
	Std. Deviation	288.83
	Minimum	2210
	Maximum	3640
Flights London (Search Query)	Mean	42.33
	Std. Deviation	12.70
	Minimum	26
	Maximum	100
Hotels UK (Search Query)	Mean	44.05
	Std. Deviation	17.73
	Minimum	21
	Maximum	100
UK Travel (YouTube Search Query)	Mean	31.4
	Std. Deviation	19.58
	Minimum	0
	Maximum	100
United Kingdom (Search Query)	Mean	72.67
	Std. Deviation	9.03
	Minimum	55
	Maximum	100
Visit Britain	Mean	34.43
	Std. Deviation	19.70

	Minimum	8
	Maximum	100

Table 2: Descriptive Statistics for variables relevant to visits to the British Museum.

Actual Visitors (In thousands)	Mean	508.27
	Std. Deviation	95.57
	Minimum	342.73
	Maximum	765.88
British Museum (Search Query)	Mean	40.17
	Std. Deviation	11.90
	Minimum	25
	Maximum	100
Things to do in London (Search Query)	Mean	65.29
	Std. Deviation	12.89
	Minimum	41
	Maximum	100
London Museums (Search Query)	Mean	56.25
	Std. Deviation	14.28
	Minimum	34
	Maximum	100
TripAdvisor London (Search Query)	Mean	54.33
	Std. Deviation	23.87
	Minimum	13
	Maximum	100

Visit London (Search Query)	Mean	59.79
	Std. Deviation	9.28
	Minimum	42
	Maximum	100

Table 3: Autocorrelation values for "Total Visits to London"

LAG	AC	PAC	Q	Prob>Q	-1 0 1 [Autocorrelation]	-1 0 1 [Partial Autocor]
1	0.8664	0.8740	92.344	0.0000		
2	0.8364	0.3866	179.13	0.0000		
3	0.8345	0.3015	266.27	0.0000		
4	0.8033	0.0819	347.71	0.0000		
5	0.7946	0.1462	428.08	0.0000		
6	0.7406	-0.1250	498.53	0.0000		
7	0.7236	0.0401	566.37	0.0000		
8	0.7282	0.1388	635.68	0.0000		
9	0.6700	-0.0727	694.89	0.0000		
10	0.6481	0.0516	750.79	0.0000		

Table 4: Autocorrelation values for "Total Visits to The British Museum"

LAG	AC	PAC	Q	Prob>Q	-1 0 1 [Autocorrelation]	-1 0 1 [Partial Autocor]
1	0.6143	0.6175	47.552	0.0000		
2	0.3500	-0.0376	63.114	0.0000		
3	0.1472	-0.0775	65.89	0.0000		
4	-0.1310	-0.2965	68.106	0.0000		
5	-0.2573	-0.0738	76.735	0.0000		
6	-0.3475	-0.1573	92.602	0.0000		
7	-0.2850	0.1082	103.37	0.0000		
8	-0.1719	0.0077	107.32	0.0000		
9	0.0706	0.2792	107.99	0.0000		
10	0.2108	-0.0140	114.04	0.0000		

BIBLIOGRAPHY

Bing Pan, Zheng Xiang, Law, R., & Fesenmaier, D. (2010). The Dynamics of Search Engine Marketing for Tourist Destinations. *Journal of Travel Research*, *50*(4), 365-377. doi: 10.1177/0047287510369558

Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. *Tourism Management*, *29*(4), 609-623.

Burger, C., Dohnal, M., Kathrada, M., & Law, R. (2001). A practitioners guide to time-series methods for tourism demand forecasting — a case study of Durban, South Africa. *Tourism Management*, *22*(4), 403-409.

Carneiro, H., & Mylonakis, E. (2009). Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks. *Clinical Infectious Diseases*, *49*(10), 1557-1564.

Carrière-Swallow, Y., & Labbé, F. (2011). Nowcasting with Google Trends in an Emerging Market. *Journal of Forecasting*, *32*(4), 289-298.

CHOI, H., & VARIAN, H. (2012). Predicting the Present with Google Trends. *Economic Record*, *88*, 2-9.

Edelman, D., & Singer, M. (2015). Competing on customer journeys. *Harvard Business Review*.

Ellis, M. (2017). *Easy and Practical SEO Wins for Museums - Eventbrite UK Blog*. [online] Eventbrite UK Blog. Available at: https://www.eventbrite.co.uk/blog/easy-seo-wins-museums-ds00/.

Jun, S., Vogt, C., & MacKay, K. (2007). Relationships between Travel Information Search and Travel Product Purchase in Pretrip Contexts. *Journal Of Travel Research*, *45*(3), 266-274.

Lemon, K., & Verhoef, P. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing*, *80*(6), 69-96.

Lohmann, M., & Danielsson, J. (2001). Predicting travel patterns of senior citizens: How the past may provide a key to the future. *Journal of Vacation Marketing*, *7*(4), 357-366.

Medium. (2018). *Invisible Insights: learning from Trip Advisor reviews*. [online] Available at: https://medium.com/mcnx-london/invisible-insights-learning-from-trip-advisor-reviews-b5c825fa4409.

Naaman, M., Becker, H. and Gravano, L. (2011). Hip and trendy: Characterizing emerging trends on Twitter. *Journal of the American Society for Information Science and Technology*, 62(5), pp.902-918.

O'Connor P. (2008) User-Generated Content and Travel: A Case Study on Tripadvisor.Com. In: O'Connor P., Höpken W., Gretzel U. (eds) *Information and Communication Technologies in Tourism 2008*. Springer, Vienna

Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of Forecasting*, *30*(6), 565-578.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, *31*(2), 179-188. doi: 10.1016/j.tourman.2009.02.016

Xiang, Z., Magnini, V., & Fesenmaier, D. (2015). Information technology and consumer behavior in travel and tourism: Insights from travel planning using the internet. *Journal of Retailing and Consumer Services*, *22*, 244-249. doi: 10.1016/j.jretconser.2014.08.005

Xiang, Z., & Pan, B. (2011). Travel queries on cities in the United States: Implications for search engine marketing for tourist destinations. *Tourism Management*, *32*(1), 88-97. doi: 10.1016/j.tourman.2009.12.004