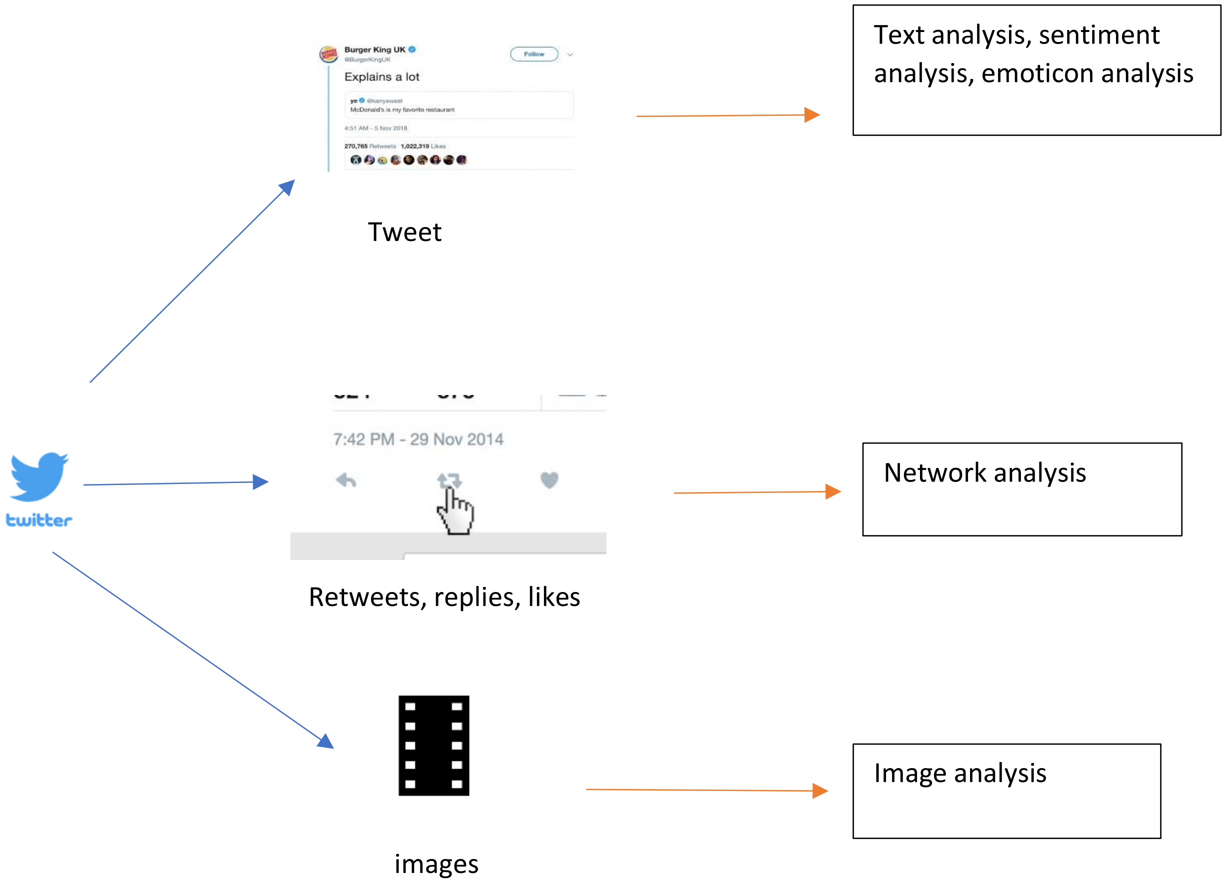

Figure 8: A visualization of the methodology.

3.1 Text analysis:

The tweet text consists of three parts – the text, hashtags, and the mentions. The hashtags can be defined to be a label of content. Although, currently hashtags are used almost in every social networking site like Facebook and Instagram, yet they can be said to gain popularity via Twitter. Hashtags are created to draw attention towards a certain topic, discussion, or event. The right sidebar on Twitter shows all the trending hashtags throughout the world. The hashtags are generally a keyword phrase beginning with the '#' sign. There are no spaces between the words in a hashtag. The travel bloggers always tend to use some generic as well as their niche-specific hashtags to draw the attention of their followers. For the 'Meet South Africa' campaign, there is a niche hashtag '#meetsouthafrica', which not only helped it earn a massive fan following but also created a niche for the campaign. In the text analysis part, I will analyse the tweets, both based on hashtags as well as the user profiles to find out the characteristics in the text which makes the tweets stand out or become more trending and visible. Furthermore, the text analysis will also reveal the generic topics or themes that the travel bloggers choose to talk about the tweets.

3.1.1. Text analysis based on hashtags –

In this section, I have extracted the tweets based on the hashtag '#meetsouthafrica' to find the below insights.

(i) The most used hashtags along with '#meetsouthafrica' –

Since hashtags are labels for content, to make a blog or tweet stand out among other blogs, the travel bloggers use many other hashtags. If a potential traveller tries to find any information about a destination and he has no idea about the official twitter user for the destination, then he might just try to browse using certain characteristics of that place. Hence, using multiple hashtags carries more probability of making the blog visible than a single hashtag since the hashtags are case sensitive. The below hashtags show how the travel bloggers of the campaign have used a variety of hashtags to describe their experiences of South Africa. The more hashtags they use the more is the probability of the tweet being popping out whenever any search related to these topics is carried out. The most used hashtags along with '#meetsouthafrica' are-

```
Selecting by n
# A tibble: 11 x 2
   hashtag                n
   <chr>              <int>
 1 #SouthAfrica          37
 2 #meetSouthAfrica      30
 3 #MeetSouthAfrica      24
 4 #WowSouthAfrica       13
 5 #safari               12
 6 #SofaSafari           12
 7 #travel               10
 8 #Travel               10
 9 #Holiday               8
10 #ItsMySouthAfrica      7
11 #southafrica           7
```

Figure 9: Top hashtags of #meetsouthafrica tweets.

Key insight – These hashtags can be used by the potential traveller to find out information about South Africa. Besides, an upcoming travel blogger who wants to write a blog or article about South Africa, can use these top hashtags to make their tweets more popular. This also highlights the fact as to how travel bloggers generate many hashtags out of the same topic since they are case sensitive and use it in their tweets to increase the chances of being visible.

(ii) The top mentions in the tweets of '#meetsouthafrica' –

The travel bloggers can tag businesses or other travel bloggers, either for promotion or engagement. Depending on settings of the account, the people who are mentioned in the tweets might get notified immediately and hence increases the probability of more replies, likes, and retweets. This shows how travel bloggers refer to other bloggers in their tweets to increase the engagement with other bloggers. The top mentions are as follows-

```
Selecting by n
# A tibble: 11 x 2
   mentions                n
   <chr>               <int>
 1 @SouthAfrica           27
 2 @VisitSAUK             14
 3 @SATravelTrade          8
 4 @ConSambo               7
 5 @TshepangLobete         7
 6 @WisaniMusengi          7
 7 @GoToSouthAfrica        6
 8 @travelopulent          5
 9 @VisitSAIndia           5
10 @Mtpatourism            4
11 @visitgauteng           4
```

Figure 10: Top mentions of #meetsouthafrica tweets.

'@SouthAfrica' is the official screen name of the 'Meet South Africa' campaign. These mentions indicate the travel bloggers who are mostly tagged in the tweets, replies, and retweets in this campaign.

Key insight – These top mentions are obviously the people who can create more engagement in their tweets and hence an upcoming travel blogger, who wants to tweet about South Africa, can tag these mentions in order to increase the engagement of their tweets too.

(iii) The top bloggers associated with this hashtag '#meetsouthafrica' –

The travel bloggers who use this hashtag the most might be the travel bloggers collaborating to promote the campaign or any traveller who has already travelled to South Africa. Any upcoming travel blogger wanting to blog about South Africa can mention these top bloggers to make their blogs or tweets more engaging. These top travel bloggers are as follows-

```
Selecting by n
# A tibble: 13 x 2
   screen_name          n
   <chr>            <int>
 1 @Frenchtouropera    28
 2 @WithersClarke      12
 3 @RSouth_Africa      10
 4 @SafariKZNMark       6
 5 @SouthAfrica         4
 6 @berghouse           3
 7 @hacktourism         3
 8 @Campx4_Outdoor      2
 9 @hermanusjewels      2
10 @KatGledhill         2
11 @needleslodge        2
12 @savenuescom         2
13 @TandemFlightCo      2
```

Figure 11: Top bloggers of #meetsouthafrica tweets.

Key insight – These people are the travel bloggers who have tweeted related to this campaign, the most. These bloggers can be referred to by a potential traveller to find out more information about South Africa. Also, an upcoming travel blogger can refer these bloggers in their tweets of South Africa in order to increase the engagement of their tweets and make them more visible in the communities of the mentioned bloggers as well.

(iv) The most liked tweets using the hashtag '#meetsouthafrica' –

The most liked tweets are mostly the tweets that either contain all the flavours that a follower likes, it contains the right information, an emotional throwback to some memory of travel, or a very attractive media. An upcoming travel blogger can go through these most liked tweets to get the design of an ideal tweet or blog. The most liked tweets associated with this hashtag are as follows –

```
   created_at          screen_name     text                                                                              favorite_count
   <dttm>              <chr>           <chr>                                                                                      <int>
 1 2020-04-29 04:19:32 BrentLeoSmith   "A magic moment of a 9 month old cheetah watching her reflection! #safariwithbrent #pai~            81
 2 2020-04-23 16:00:11 SouthAfrica     "One of South Africa's most beautiful inland provinces is the Free State, home to numer~            42
 3 2020-04-28 13:49:28 SouthAfrica     "Yesterday, in the heart of Times Square in NYC, @NASDAQ joined us in inviting everyone~            22
 4 2020-04-30 16:00:53 SouthAfrica     "South Africa's Cradle of Humankind is a transformative experience. Only here can you w~            21
 5 2020-04-30 10:05:07 SafariKZNMark   "Planning a visit to the probably the best all round province in #SouthAfrica? #KwaZulu~            11
 6 2020-04-27 08:49:15 SafariKZNMark   "Don't like it, manic or blue, with a new moon. It must be #Monday so lets create a #Mo~            10
 7 2020-04-29 20:21:18 carmensluxtrvl  "I'm celebrating #WineWednesday at home by looking back on my trip to gorgeous South Af~             9
 8 2020-04-26 12:27:02 berghouse       "Early morning in the #Drakensberg #mountains #SouthAfrica #meetsouthafrica #visitsouth~             9
 9 2020-04-26 13:55:04 Frenchtourope~  "The ultimate tubing #adventure experience takes place in the Plaatbos Nature reserve &~             8
10 2020-04-28 10:15:55 deckchairhq     "Dreaming of the stunning #sunrise to #sunset view of Camps Bay &amp; The Lion's Head f~             8
```

Figure 12: Most liked tweets of #meetsouthafrica tweets.

Key insight – The most liked tweets can be further analysed to find out specific characteristics like the topic/topics, emoticons, hashtags and mentions which have made the tweet a favourite to its followers.

(v) The most like retweets associated with hashtags using '#meetsouthafrica'-

The most liked retweets indicate the type of tweets that make the readers or followers engage the most. Occasionally, people also type 'RT' to indicate that they are retweeting or reposting someone else's tweet. These tweets are either so appropriate or informative that the followers either forward them for others to read or they repost on these tweets, their own experiences, and recommendations. Retweeting is also one of the ways to engage in some conversation and become more visible among the followers for a travel blogger. The most engaging retweets associated with this campaign or hashtag is as follows-

```
Selecting by retweet_count
# A tibble: 23 x 4
   created_at          screen_name    text                                                                             retweet_count
   <dttm>              <chr>          <chr>                                                                                    <int>
 1 2020-04-23 16:00:11 SouthAfrica    "One of South Africa's most beautiful inland provinces is the Free State, home to numero~            10
 2 2020-04-28 13:49:28 SouthAfrica    "Yesterday, in the heart of Times Square in NYC, @NASDAQ joined us in inviting everyone ~             8
 3 2020-04-30 16:00:53 SouthAfrica    "South Africa's Cradle of Humankind is a transformative experience. Only here can you wa~             8
 4 2020-04-29 04:19:32 BrentLeoSmith  "A magic moment of a 9 month old cheetah watching her reflection! #safariwithbrent #pain~             5
 5 2020-04-28 07:46:38 Frenchtourope~ "Connect with us on our social media for more updates, #TravelTips #bestplacestovisit #S~             3
 6 2020-04-26 13:55:04 Frenchtourope~ "The ultimate tubing #adventure experience takes place in the Plaatbos Nature reserve &a~             3
 7 2020-04-29 18:09:13 KatGledhill    "@travelopulent @GansbaaiTourism Q5 LOVE visiting the endangered penguin colony in #Bett~             3
 8 2020-04-30 14:57:04 KaapstadTours  "Let's focus forwards. Check out our Motorcycle Tours and Escorted Safari Tours.  Let's ~             2
 9 2020-04-28 09:56:58 SafariKZNMark  "If you'll just come with me you'll see the beauty of Tuesday afternoon...Variety everyd~             2
10 2020-04-29 08:58:35 SafariKZNMark  "And the vision that was planted in my brain, still remains...\nWhat is the vision here?~             2
```

Figure 13: Most retweeted tweets of #meetsouthafrica tweets.

> **Key insight** - These retweets can be further analysed to give a hint to the upcoming travel blogger about the characteristics which make the followers repost the tweets more. Hence, it will highlight the features of a tweets that makes it more engaging.

3.1.1.1. Text mining:

The above insights are the most common analysis that I can perform on a tweet based on hashtags. Now, I will use the various tools and techniques present in R, in order the prepare the data, run diagnostics, and mine the deeper insights buried in the tweets containing the hashtag '#meetsouthafrica'.

After the authentication and extracting of Twitter data depending on the hashtag '#meetsouthafrica', I cleaned and pre-processed the text to make it appropriate for the rest of the text mining task. This pre-processing and cleaning involved substituting unnecessary special characters, trimming and removing white spaces, converting to lower case, removing stop words, emoticons, and numbers. Then I converted the cleaned text into a corpus for further processing. In Natural Language Processing (NLP), a corpus means a collection of structured texts. This corpus then undergoes tokenization, in which the text is split into tokens or words. The type of text mining that I will implement, is called a "bag of words". In the "bag of words" type of analysis, each token or word is treated as a unique feature of a document. This bag of words analysis provides us with a term document matrix (tdm) where each word represents the rows while the columns are indicated by the documents.

(i) Word frequencies:

The frequency of the top words that occur in the tweets is as shown in the below figure. The top words give us an idea as to which word has been mentioned frequently in the tweets or the trending topics about South Africa. I can see that the most frequently used word is 'canyon'. Currently, people are discussing their desires to travel after the lockdown and the Blyde Canyon in South Africa is one of the most popular destinations among travellers. Also, the Blyde Canyon is the third-largest reserve in the world and hence the rest of the frequently mentioned words justify it. Furthermore, a Pan-African online trading tool has been launched in South Africa and thus the remaining top mentioned words. Therefore, if an upcoming travel blogger wants to tweet about South Africa and become popular, then they should probably tweet about these trending topics.

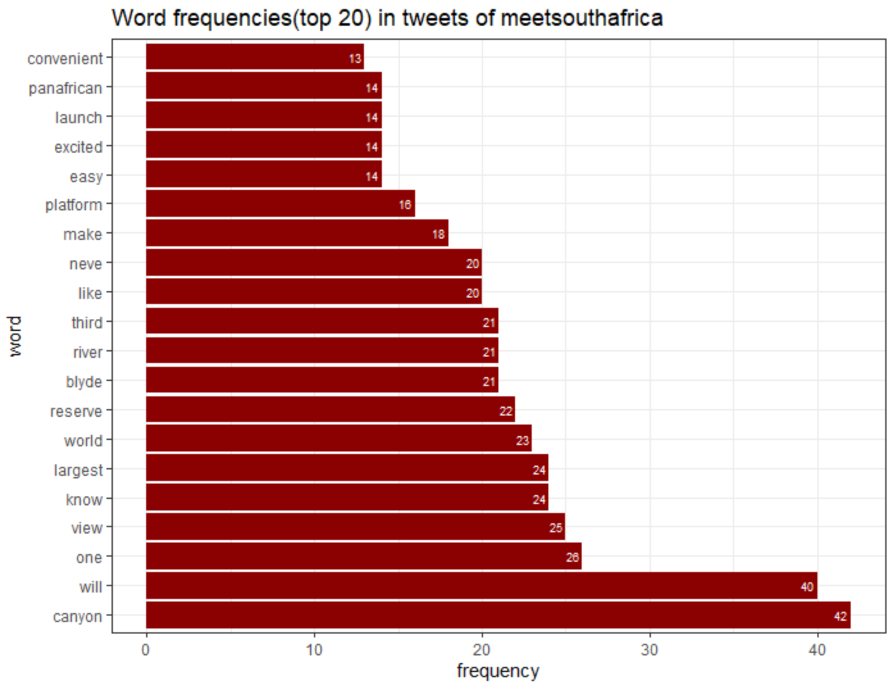

Figure 14: Word frequencies in #meetsouthafrica tweets.

Key insight – The words most frequently mentioned indicate the most trending topic in the tweets. Hence, if an upcoming travel blogger wants to gain instant popularity, then they can tweet about these trending topics.

Word associations in NLP illustrate the words that are most likely to relate with a word. In our case, this word association will be calculated based on its association with the most frequently appearing words in the tweets i.e. the trending topics. The below figure shows the most associated words with the word ‘canyon’ in the tweets of “#meetsouthafrica”. The findAssocs() function in R gives the correlation values of these words with the word ‘canyon. The figure below shows how the word ‘Blyde’, ‘river’, and ‘third’ are associated very strongly with the word ‘canyon’. The values in the figure below indicate the correlation value with the word ‘canyon’. Thus, the words ‘third’, ‘river’ and ‘Blyde’ having a correlation value of 1 illustrates that these words always appear together. This gives us more details about trending topics.

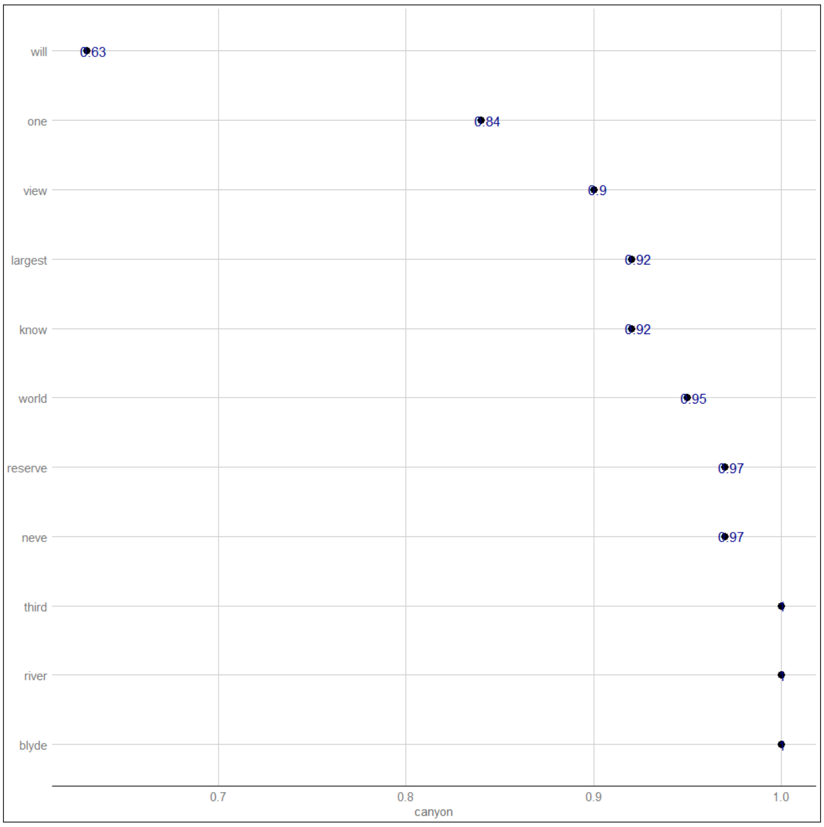

Figure 15: Word associations in #meetsouthafrica tweets.

Key insight – The word associations tell us a bit more about the trending topics. For example, in the previous section, I have seen how ‘canyon’ is the most frequently mentioned word, hence from the word associations, I can see how the word ‘canyon’ is strongly correlated with the word ‘Blyde’ ,’river’ and ‘third’ , indicating the most trending topic to be ‘Blyde Canyon’ which is the third largest river canyon in the world. Thus, an upcoming travel blogger can get some more details about the trending topics regarding South Africa.

The most frequent words appearing in a tweet indicate the words which are used to describe a certain topic. By having a look at the word cloud of the most frequently appearing words, we can get an overall idea about the tweets trending in South Africa. This word cloud is made by words which have more than frequency 10.

Figure 16: Words that have occurred more than 10 times in #meetsouthafrica tweets.

Key insight – This is like most frequently mentioned words and lets an upcoming travel blogger know about all the trending topics.

(iv) Term Frequency distribution:

The term frequency is a join of the frequency of each token and the total number of that token in all documents i.e. tweets in our case. After counting the frequency of terms in each tweet and sorting, we get the below statistics:

```
   term        n
   <chr>    <int>
 1 will       40
 2 one        25
 3 view       25
 4 know       24
 5 largest    24
 6 world      23
 7 reserve    22
 8 blyde      21
 9 canyon     21
10 river      21
# ... with 427 more rows
```

Figure 17: Sorted frequency of words in #meetsouthafrica tweets.

The distribution of term frequency has been shown in the figure below. Along X-axis is the number of times a term has occurred by the total number of terms. For example- the statistics show that there 250 terms which have frequency 1 and the total number of terms is 1247. Hence, the tallest bar in the histogram is of the terms which appear with a frequency of 1. Furthermore, this proves how tweets contain common words as well as unique words like documents.

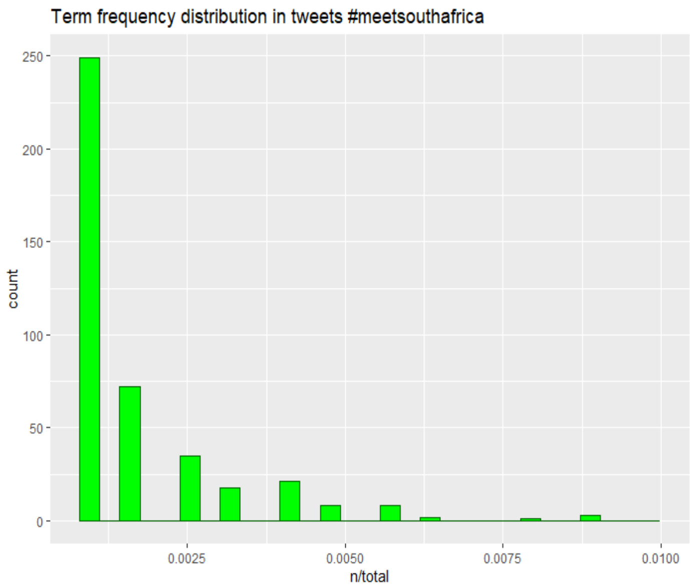

Figure 18: Term frequency distribution of #meetsouthafrica tweets.

This long-tailed distribution is a very typical term frequency distribution whenever analysis of documents is involved. The relationship between a term's frequency and its rank is called Zipf's law. Zipf's law states that the probability of occurrence of words always start high and then tapers off. This relates to the fact of how stop-words occur very frequently and hence their rank is low. From the chart above, we can say that the word 'will' tends to occur very often and hence its importance is insignificant.

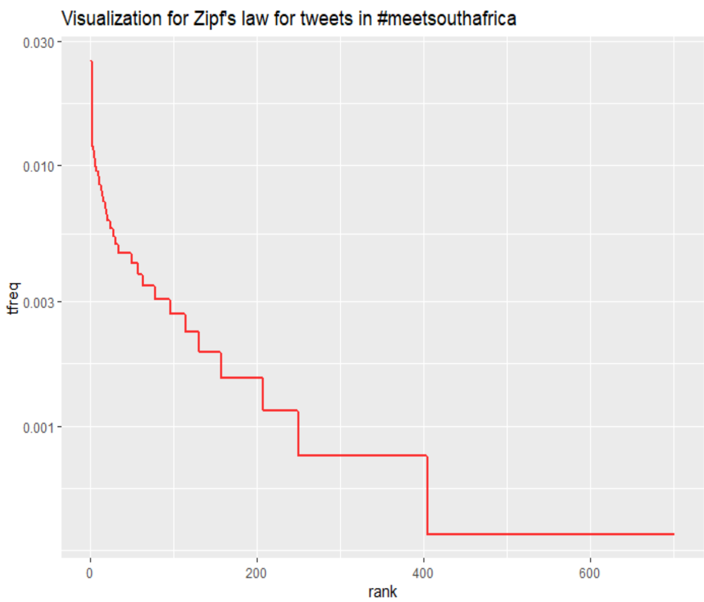

Figure 19: Projection of Zipf's law in #meetsouthafrica tweets.

Key insight – This proves how the tokens or words distribution in tweets follow the power law in which the frequency of a word is inversely proportional to its rank. This explains why I removed the stop words or the most common words appearing in the cleaning phase of the tweet text like 'the', 'is' and 'that' as their rank or importance is negligible relative to the tweet. Furthermore, while blogging or tweeting about a trending topic, the upcoming travel blogger can have an idea as to which words they should not be using more related to that topic.

(v) Term Frequency – Inverse Document Frequency:

In term frequency-inverse document frequency (tf-idf), each word or term is given a weightage to quantify each tweet or document. This weightage is the importance of the word to the whole tweet or document. The frequency of a term that appears in a document or tweet is the term frequency or tf part of tf-idf. So, if a word appears too many times in a document then it is termed as important but at the same time, if that word appears in several documents then it loses its importance. Therefore, this weighting system must decrease the importance of a word if it appears in too many tweets or documents. Applying tf-idf to our tweets does the following-

```
<<DocumentTermMatrix (documents: 156, terms: 437)>>
Non-/sparse entries: 1247/66925
Sparsity           : 98%
Maximal term length: 14
Weighting          : term frequency - inverse document frequency (tf-idf)
Sample             :
     Terms
Docs  blyde canyon know largest one reserve river view will   world
  1       0      0    0       0   0       0     0    0    0 0.00000
  107     0      0    0       0   0       0     0    0    0 0.00000
  114     0      0    0       0   0       0     0    0    0 0.00000
  117     0      0    0       0   0       0     0    0    0 2.76184
  14      0      0    0       0   0       0     0    0    0 0.00000
  15      0      0    0       0   0       0     0    0    0 0.00000
  5       0      0    0       0   0       0     0    0    0 0.00000
  71      0      0    0       0   0       0     0    0    0 0.00000
  89      0      0    0       0   0       0     0    0    0 0.00000
  90      0      0    0       0   0       0     0    0    0 0.00000
```

Figure 20: Tf-idf scores for terms in #meetsouthafrica tweets.

From the figure above, I can see that only the term 'world' has some weight associated with it. This signifies that this word has relative importance to a tweet.

Key insight – The tf-idf scores indicates the important details related to the most trending topics. So, an upcoming traveller can get a hint as in which context the most trending topic has not been talked about much and hence can write about it to become more visible. For example, in one of the tweets, only the word 'world' has a tf-idf score. Thus, in the retweets and replies of that tweet, a travel blogger can talk about the Blyde Canyon in context of the world, since it has not been talked about much.

(vi) Text Clustering:

Whenever I hear the word 'Text Clustering', the first example that comes to our mind is the Google search engine. When I search for a word in Google, it returns all the web pages which contain that word. So, the Google algorithm breaks down the unstructured data from the web pages into a matrix model in which the web pages are tagged by the keywords and then they are used by the algorithm to get the web results (Team, 2020). Text clustering can be done based on term frequency as well as term frequency-inverse document frequency. T-sne or T-distributed stochastic neighbor embedding will give each token or word, a location in 2D space. It is this way that t-sne can gather the similarities and assist in cluster analysis. Clara Clustering is used in large applications to allot the data points to k clusters by reducing the computational time and the RAM storage issues. These clusters can be used to find the unique words and the most frequently used words in each cluster and thus have an idea about the topic people are tweeting about. This concept is like Topic Modelling which I will analyse in the later section of the report. I can see how clustering based on term frequency yields a lesser number of clusters than clustering based on tf-idf. This can be due to the fact that term frequency provides lesser information than tf-idf. Term frequency only indicates the frequency of each term whereas tf-idf also highlights the importance of each term to the tweets.

Key insight – The text clustering will give us the generic topics or subjects that the travel bloggers might be talking about in their blogs or tweets. These topics can be derived by finding the unique words or the most frequently mentioned words in the clusters. Thus, if an upcoming travel blogger can choose the topics or subjects that they want to write about from these lot.

A: By term frequency:

Clustering based on the term frequency gives 7 as the optimal number of clusters. Unique words from these clusters or the term frequency will give us an idea about the topics that people are tweeting about.

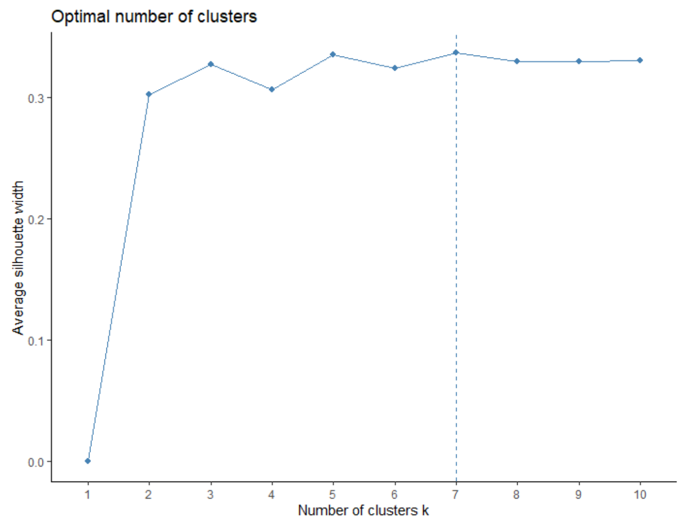

Figure 21: Optimal number of clusters given by term frequency.

B: By tf-idf:

Clustering based on tf-idf gives 10 as the optimal number of clusters. It can be noticed that the optimal number of clusters given by tf-idf is bigger than that by term frequency. This might be since tf-idf takes the importance of the word relative to each document or tweet and hence it is supposed to generate some more topics.

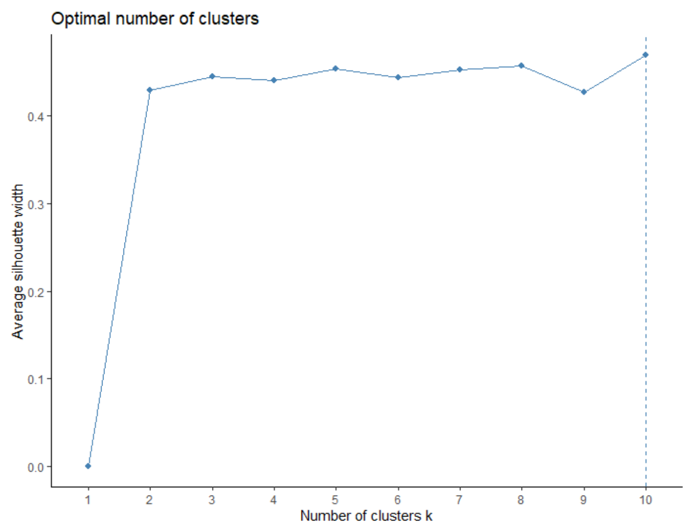

Figure 22: Optimal number of clusters given by tf-idf.

Furthermore, it can also be said that using more unique words that are relevant to the topic/topics of the tweet will increase the tf-idf and will eventually give out more details about the topics. I will derive the important words from each cluster to highlight the generic topics in the next section.

(vii) Word importance from clusters:

The clusters that I derived earlier can be given some context by giving importance to the words in the clusters. I quantified the importance by tf-idf. Tf-idf increases the importance of unique words while decreases the weights of most commonly used words.

The tf-idf values of most commonly used terms in the #meetsouthafrica tweets like 'one,'view', and 'know' are lesser than the words 'blyde','canyon' because these are the words that appear in almost all the tweets. Hence, the more unique words the travel bloggers use in the tweets, the more is the probability of the blog being liked by the followers.

```
   cluster term         n total    tf   idf tf_idf
   <fct>   <chr>    <int> <int> <dbl> <dbl>  <dbl>
 1 6       blyde       21   250 0.084  2.30  0.193
 2 6       canyon      21   250 0.084  2.30  0.193
 3 6       know        21   250 0.084  1.61  0.135
 4 6       largest     21   250 0.084  1.61  0.135
 5 6       one         21   250 0.084  1.20  0.101
 6 6       reserve     21   250 0.084  1.61  0.135
 7 6       river       21   250 0.084  2.30  0.193
 8 6       third       21   250 0.084  2.30  0.193
 9 6       view        21   250 0.084  1.20  0.101
10 6       world       21   250 0.084  1.20  0.101
# ... with 494 more rows
```

Figure 23: Tf-idf scores of terms in cluster 6.

The figure below shows the words having the highest tf-idf scores in each cluster. From these words, the most common topics that the travel bloggers talk about in the #meetsouthafrica tweets can be generated. The top 10 tf-idf scores of the terms in each cluster are as shown below. I can see that clusters 9 and 10 have the highest tf-idf scores out of all the clusters. I can analyse the clusters further to find more insights about the topics in the tweets.

Figure 24: tf-idf scores distribution of terms in 10 clusters.

(viii) Bigrams:

Bigrams are a pair of words that always appear together. They indicate the relationship between words. The unnest_tokens() function in R breaks the tweets into a sequence of words based on n-grams. These bigrams give us more details about the topics in the clusters. The table below shows the bigrams from cluster 2 and 7 and their tf-idf scores.

```
   cluster bigram                   n    tf   idf tf_idf
   <fct>   <chr>                <int> <dbl> <dbl>  <dbl>
 1 2       changes day              7 0.143  2.30  0.329
 2 2       day earth                7 0.143  2.30  0.329
 3 2       earth forever            7 0.143  2.30  0.329
 4 2       forever friends          7 0.143  2.30  0.329
 5 2       friends good             7 0.143  2.30  0.329
 6 2       good like                7 0.143  2.30  0.329
 7 2       like nowhere             7 0.143  2.30  0.329
 8 7       african beadwork         6 0.103  2.30  0.238
 9 7       beadwork beautiful       6 0.103  2.30  0.238
10 7       beautiful culture        6 0.103  2.30  0.238
# ... with 471 more rows
```

Figure 25: Tf-idf scores of bigrams in #meetsouthafrica tweets.

The tf-idf value indicates the importance of the pair of words that occur in the tweets. The lower the value, the lower the importance of the bigrams. If I consider the top 10 tf-idf scores of the bigrams in the respective clusters then I can make more sense of the words in the clusters. This can help me understand the topics better for the most important words in a tweet.

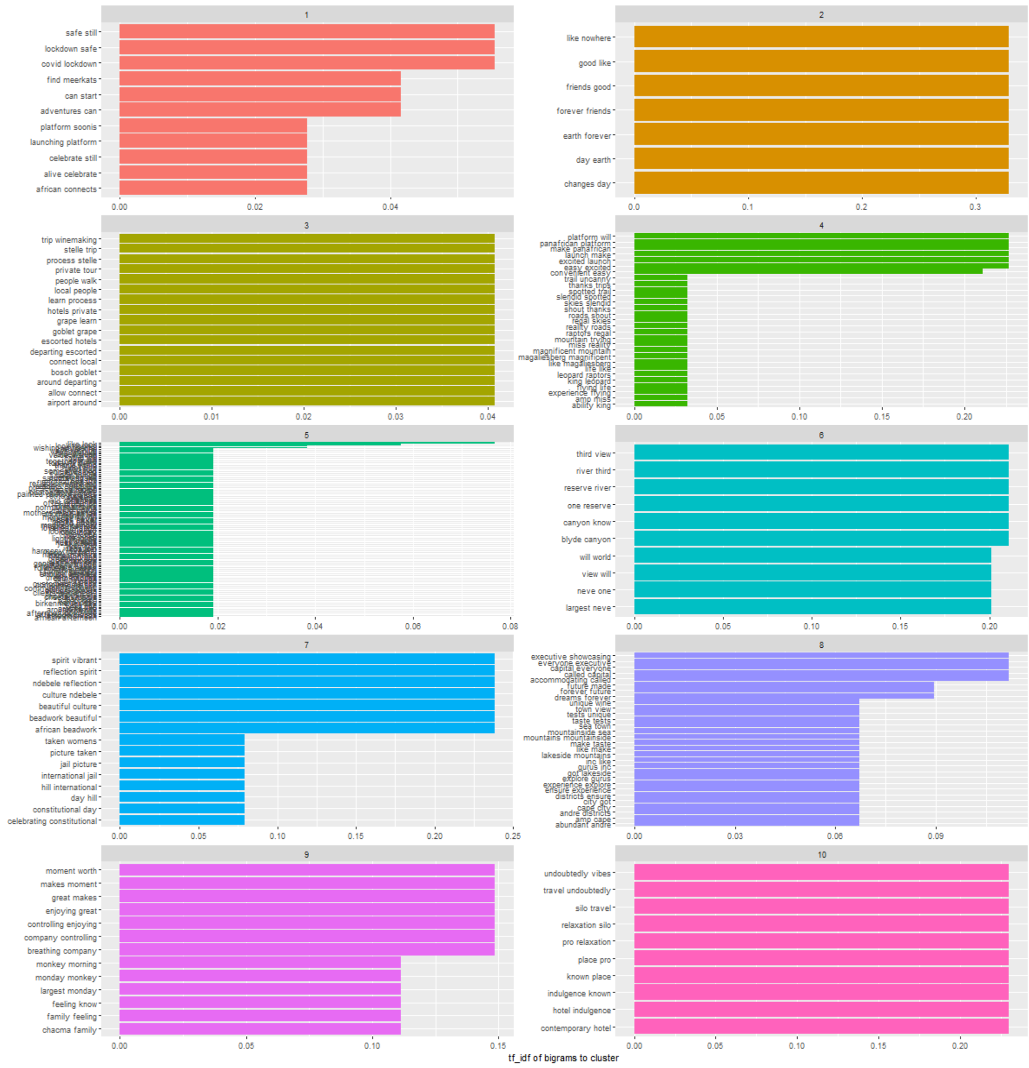

Figure 26: Tf-idf scores distribution of bigrams in 10 clusters.

The igraph package in R helps us to arrange the bigrams into a network. The graph_from_data_frame() function creates a data frame of edges with columns "from" (the vertices, clusters), "to" (the edges, bigrams), and a weighting variable i.e. which is the frequency. I have considered the weighting variable to be 10 in the below example, then I see that bigrams only from clusters 4 and 6 have been selected. These bigrams occur more than 10 times among all the tweets.

```
 [1] 6->blyde canyon       6->canyon know       6->know largest      6->one reserve       6->reserve river
 [6] 6->river third        6->third view        6->largest neve      6->neve one          6->view will
[11] 6->will world         4->easy excited      4->excited launch    4->launch make       4->make panafrican
[16] 4->panafrican platform 4->platform will    4->convenient easy
```

Figure 27: Bigrams which have occurred more than 10 times in #meetsouthafrica tweets.

This igraph object can be converted to a graph with the help of ggraph package. This gives us a visual of the co-occurring words in a tweet containing the hashtag '#meetsouthafrica'. In the figure below, I have considered only those bigrams which appear more than 10 times.

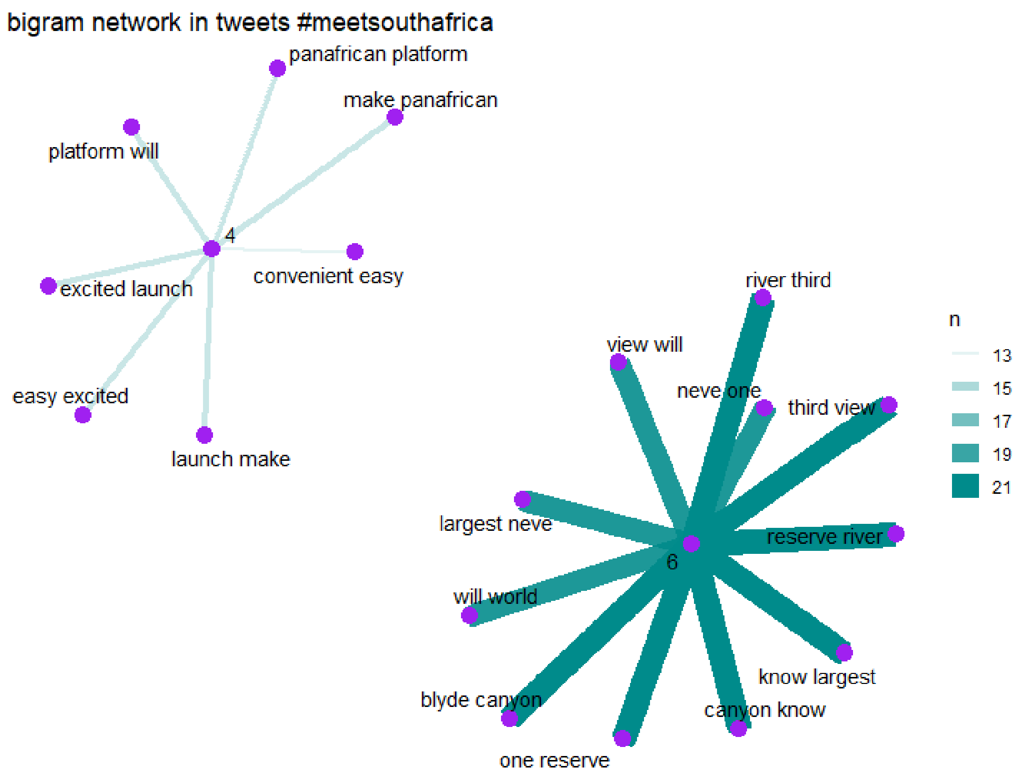

Figure 28: Bigram network of bigrams occurring more than 10 times in #meetsouthafrica tweets.

These bigrams perform the same task as the word associations do. The bigrams also give an idea about the most trending topics.

(ix) Topic Modelling:

Just like text clustering can be used to determine topics of the tweets, there is another alternative known as topic modelling. Topic modelling identifies topics from a cluster of words that always appear together. It illustrates the topics that people might be talking about in tweets. I will use Latent Dirichlet allocation (LDA) algorithm to perform topic modelling to our corpus of tweet text. One of the main hurdles of topic modelling is determining the number of optimal topics. Here, I will use a mathematical approach in which the harmonic mean is calculated for various values of k where k eventually yields the optimal number of topics. The optimal number of topics is the topic corresponding to the highest harmonic mean.

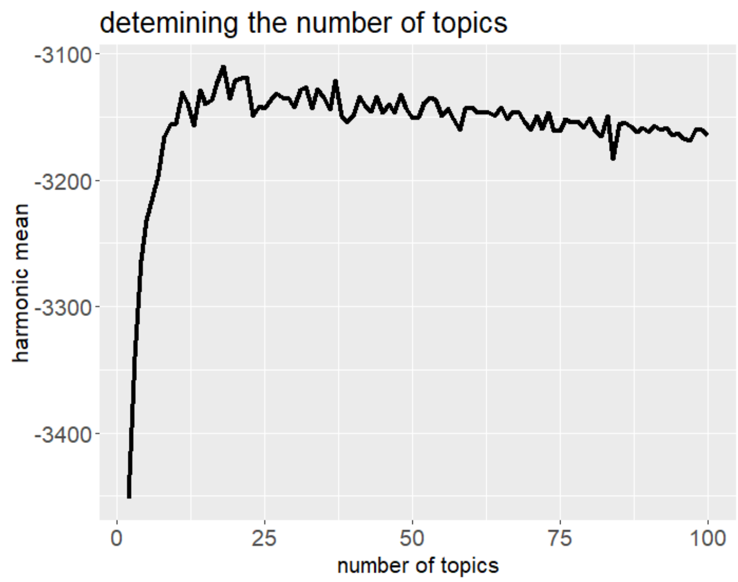

Figure 29: Elbow point showing the optimal number of topics in topic modelling.

Here, the elbow point represents the optimal number of topics. In our case, it comes to be 18. The below figure shows the top 10 terms from all the 18 topics. These topics, like the clusters, give an idea about the subject that people might be discussing in the tweets.

```
      Topic 1       Topic 2      Topic 3         Topic 4     Topic 5     Topic 6    Topic 7     Topic 8     Topic 9       Topic 10
 [1,] "got"         "showcasing" "everyone"      "reality"   "can"       "cape"     "just"      "think"     "morning"     "like"
 [2,] "happy"       "soonis"     "accommodating" "african"   "will"      "around"   "learn"     "covid"     "meerkats"    "uff"
 [3,] "right"       "executive"  "getting"       "covid"     "white"     "city"     "stelle"    "look"      "future"      "around"
 [4,] "hotels"      "ago"        "melanie"       "look"      "process"   "still"    "safe"      "capital"   "goblet"      "walk"
 [5,] "magnificent" "payout"     "tune"          "planets"   "board"     "white"    "never"     "vets"      "will"        "watch"
 [6,] "tonight"     "gave"       "done"          "sea"       "gentle"    "zebras"   "mountain"  "frugal"    "dreams"      "date"
 [7,] "worlds"      "watch"      "maximum"       "hours"     "look"      "trip"     "trip"      "coach"     "may"         "shout"
 [8,] "excited"     "either"     "respectively"  "done"      "normal"    "giants"   "boatbased" "transfers" "scenic"      "cheetah"
 [9,] "comments"    "mini"       "roads"         "roads"     "searching" "memories" "microbus"  "roads"     "trip"        "magic"
[10,] "afternoon"   "trips"      "agree"         "wonderful" "tours"     "family"   "miss"      "can"       "experiences" "bloukrans"
      Topic 11      Topic 12     Topic 13  Topic 14     Topic 15     Topic 16     Topic 17       Topic 18
 [1,] "ranger"      "amp"        "flying"  "called"     "today"      "memories"   "amp"          "still"
 [2,] "find"        "black"      "forever" "lockdown"   "life"       "bosch"      "mountains"    "like"
 [3,] "travel"      "alive"      "every"   "grape"      "forever"    "capital"    "mountainside" "celebrate"
 [4,] "made"        "got"        "watch"   "field"      "keep"       "get"        "miss"         "seater"
 [5,] "investors"   "show"       "stripes" "executive"  "weekday"    "adventures" "view"         "join"
 [6,] "mrs"         "tune"       "shout"   "wherever"   "frugal"     "start"      "magaliesberg" "tour"
 [7,] "philippine"  "comment"    "normal"  "doubting"   "stripesuf"  "seater"     "town"         "railways"
 [8,] "volunteers"  "winemaking" "choice"  "see"        "close"      "sedan"      "massive"      "exhilarating"
 [9,] "amp"         "keeping"    "old"     "process"    "launch"     "family"     "follows"      "thanks"
[10,] "thanks"      "together"   "view"    "winemaking" "definitely" "holiday"    "feature"      "trips"
```

Figure 30: The 10 topics from each of the 18 topics revealed by topic modelling.

Key insight – The insights that I derived from topic modelling is similar to that of text clustering. Either one of the techniques can be used to find out the generic topics of the tweets. In the case of the tweets, topic modelling can be preferred more over text clustering as it revealed 18 topics while text clustering generated only 10 clusters based on tf-idf scores. More the number of topics, more is the number of choices to select from by an upcoming travel blogger.

The word cloud below indicates the top 10 unique words used in all the 18 topics. This can also be interpreted as the unique words used in all the tweet text. These are the words that the travel bloggers used to describe their perspective about their stories about South Africa travel or it can also be said to be the terms with the highest tf-idf scores for each tweet. However, with the word clouds, I can get a clearer picture of the unique words for each topic. The unique words from topic 1 to 5 are as below-

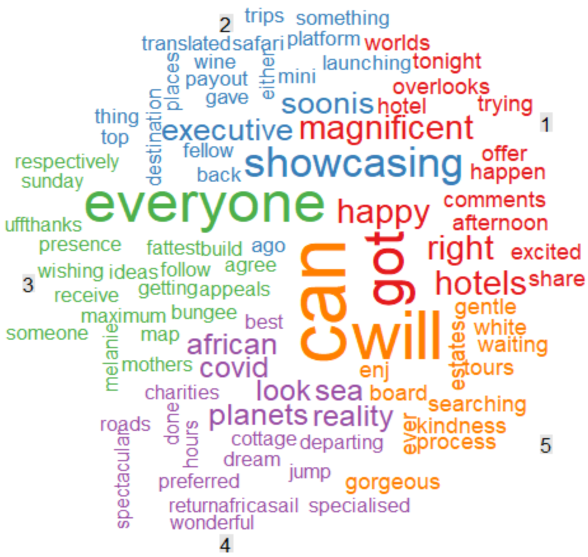

Figure 31: Unique words from topics 1 to 5.

The unique words from topic 6 to 10 can be seen in the word cloud below.

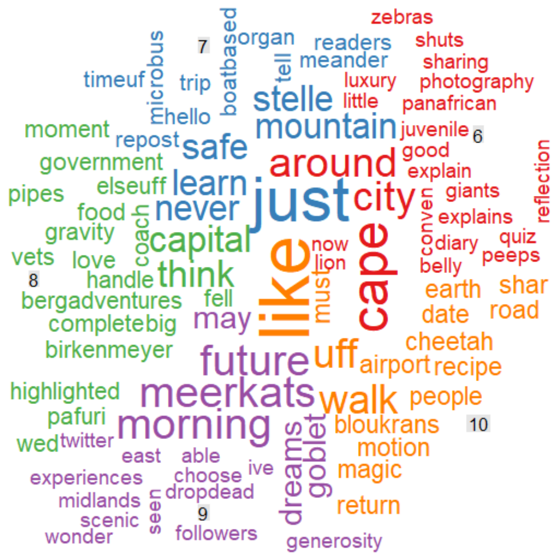

Figure 32: Unique words from topics 6 to 10.

The unique words from topic 11 to 18 can be seen in the word cloud below.

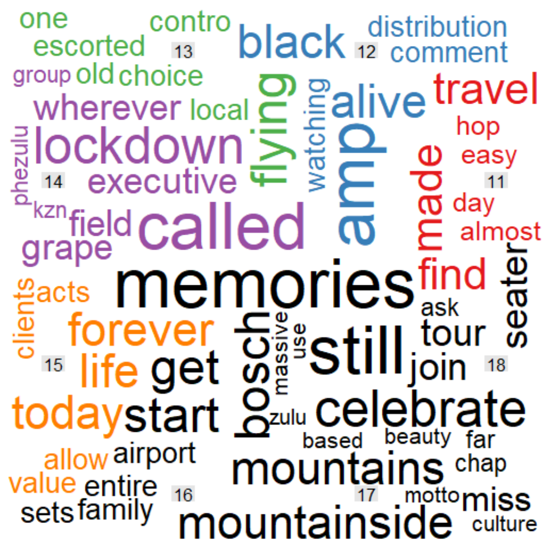

Figure 33: Unique words from topics 11 to 18.

The term distribution in topic modelling via LDA outlines that only 5-8 words lead to the probability of each topic which also means that since a small number of words define a topic, it cannot be certainly said that the topic applies to the whole tweet.

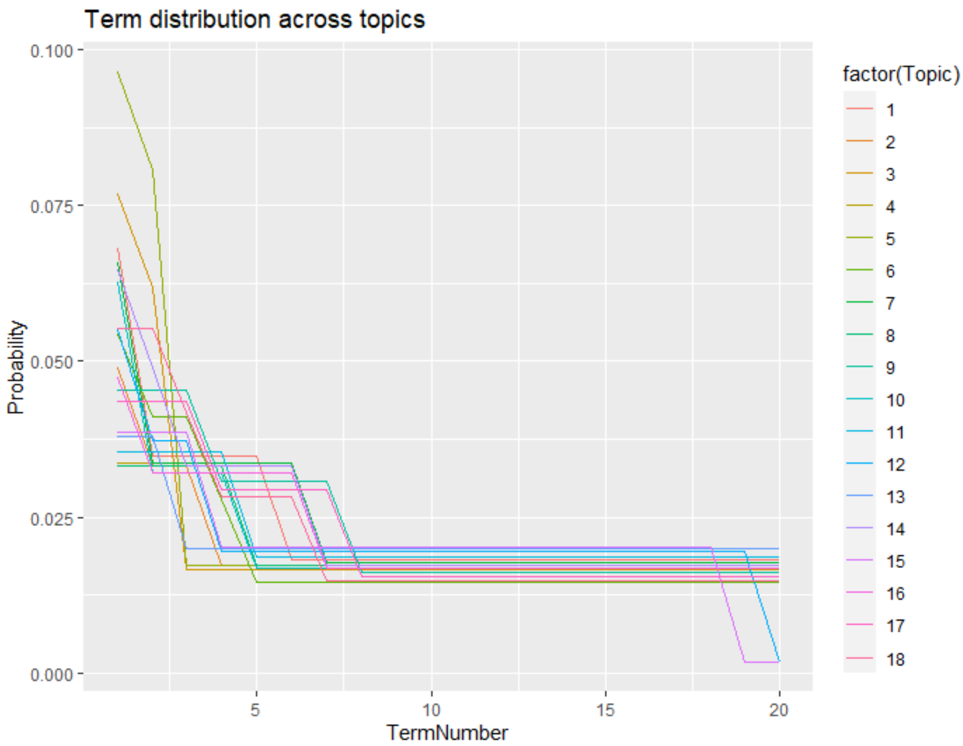

Figure 34: term distribution across topics in topic modelling.

To have a better understanding of the topics, combining more words from each topic might give a clearer idea about the topic. To create a label for the topics, I can rank the terms and concatenate the first three terms to derive a better understanding of the topics.

	Label
1	got happy right
2	showcasing soonis executive
3	everyone accommodating getting
4	reality african covid
5	can will white
6	cape around city
7	just learn stelle
8	think covid look
9	morning meerkats future
10	like uff around
11	ranger find travel
12	amp black alive
13	flying forever every
14	called lockdown grape
15	today life forever
16	memories bosch capital
17	amp mountains mountainside
18	still like celebrate

Figure 35: Labels created by ranking top three terms from 18 topics.

In LDA, each topic is assigned to a document with a probability. These probabilities indicate how a tweet can be a mixture of many topics. For example, one of the highest probabilities occurs for document 1 in topics 2 and 3. This shows how I can combine topics according to probabilities and get a better understanding of the topic of the tweet.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
1	0.05996473	0.07583774	0.07583774	0.05996473	0.04409171	0.04409171	0.05996473	0.04409171	0.04409171	0.04409171	0.09171076	0.04409171
2	0.04708098	0.06403013	0.04708098	0.04708098	0.04708098	0.04708098	0.04708098	0.06403013	0.04708098	0.06403013	0.06403013	0.08097928
3	0.08237548	0.06513410	0.06513410	0.04789272	0.04789272	0.04789272	0.04789272	0.04789272	0.06513410	0.04789272	0.04789272	0.04789272
4	0.04553734	0.04553734	0.04553734	0.04553734	0.06193078	0.07832423	0.04553734	0.06193078	0.04553734	0.06193078	0.04553734	0.09471767
5	0.05241090	0.05241090	0.05241090	0.05241090	0.05241090	0.07127883	0.05241090	0.05241090	0.05241090	0.05241090	0.05241090	0.07127883
6	0.04789272	0.04789272	0.04789272	0.04789272	0.06513410	0.04789272	0.09961686	0.04789272	0.06513410	0.04789272	0.04789272	0.06513410

	V13	V14	V15	V16	V17	V18
1	0.04409171	0.05996473	0.05996473	0.04409171	0.05996473	0.04409171
2	0.06403013	0.04708098	0.06403013	0.04708098	0.06403013	0.04708098
3	0.04789272	0.06513410	0.04789272	0.04789272	0.06513410	0.06513410
4	0.06193078	0.04553734	0.07832423	0.04553734	0.04553734	0.04553734
5	0.05241090	0.05241090	0.05241090	0.05241090	0.05241090	0.07127883
6	0.04789272	0.06513410	0.04789272	0.06513410	0.04789272	0.04789272

Figure 36: LDA assigning probabilities of belonging to 18 topics, to all tweets.

Community detection can be performed to find correlated topics. For our 18 topics, community detection reveals the below graph. It illustrates how topics 13 and 15 are very strongly related. The thickness of the lines indicates the strength of the correlation.

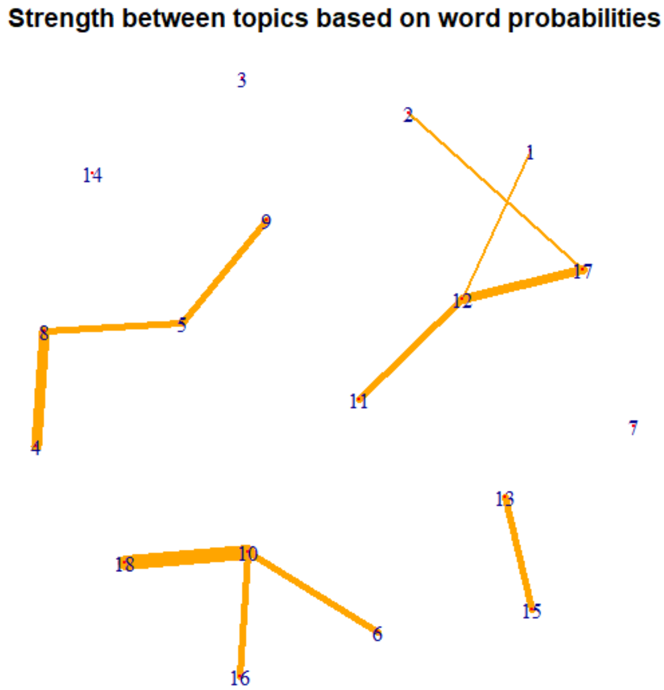

Figure 37: Strength between topics among all 18 topics.

(x) Sentiment analysis:

Using text analytics, I can classify the emotions used in a tweet. Sentiment analysis can be very powerful to observe followers' reactions to tweets or blogs. This reaction can be captured via the words used in the tweet or via the emoticons used along with the tweet. Here, I performed sentiment analysis based on the tweet text. In R, there are several packages to deal with sentiments. I used the Tidytext package for our analysis, which has three lexicons. Lexicons can be defined to be a word book.

These three lexicons are Bing, AFINN, and NRC. Bing segregates words into positive, negative, or neutral whereas AFINN assigns -5 to 5 scores with the signs indicating negative and positive sentiments. The NRC lexicon consists of eight basic emotions like anger, fear, anticipation, trust, surprise, sadness, joy, and disgust. I used the Bing lexicon for analysing the sentiments which categorises words into positive, negative, or neutral.

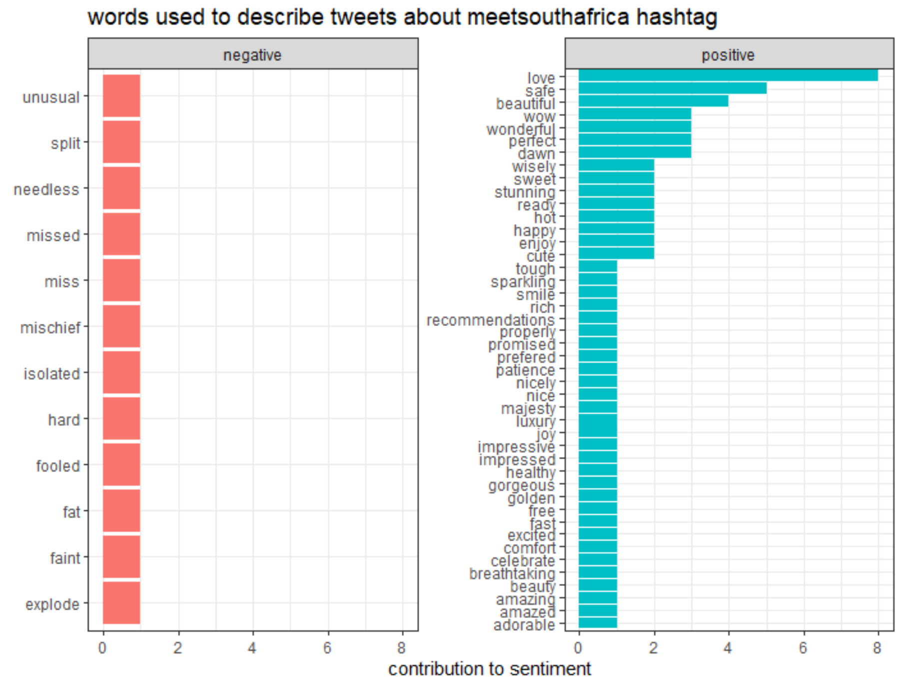

Figure 38: Sentiment analysis of #meetsouthafrica.

This indicates that travel bloggers mostly use positive sentiments or words in the tweets. Hence, the upcoming travel bloggers should maintain a positive polarity in the tweets as positivity always makes people hang around the most.

Key insight – An upcoming travel blogger must always maintain a positive sentiment in the tweets and blogs.

(xi) Emoticon analysis –

The emojis or emoticons often serve as a creative form of language. Humans tend to get attracted to pictures and graphics more than words. Studies show that followers engage more with tweets with emojis than just texts. According to Marissa Window from the Twitter business, emojis can be defined as an elevator pitch for business (Marissa Window, 2020). The chart below shows the words with which the emoji has been linked the most in the '#meetsouthafrica' tweets. The 'heart' and 'sun' emoji indicate that people who travel to South Africa mostly love their experience and the bright and sunny climate of South Africa respectively. These statistics show how the bloggers have associated positive emojis with the tweets of #meetsouthafrica to highlight their travel experiences.

```
# A tibble: 73 x 3
   emoji     words      frequency
   <fct>     <chr>          <dbl>
 1 red heart love               2
 2 red heart incredible         1
 3 red heart journey            1
 4 red heart sharing            1
 5 red heart thank              1
 6 sun       like               3
 7 sun       looks              3
 8 sun       great              2
 9 sun       lovely             2
10 sun       photos             2
# ... with 63 more rows
```

Figure 39: Usage frequency of emoticons in #meetsouthafrica tweets.

The statistics below show that the use of emojis also differ during the days of the week compared to weekends as the frequency of tweeting overall can be anticipated to differ as well. In the later section of user account analysis, I will see whether this difference stands true or not. Nevertheless, the upcoming travel bloggers must use emojis along with the text to draw more attention to their tweets or blogs.

```
  weekday       n
  <chr>     <int>
1 Tuesday     278
2 Wednesday   268
3 Thursday    235
4 Friday      222
5 Monday      216
6 Saturday     24
7 Sunday        6
```

Figure 40: Usage frequency of emoticons in a week.

Key insight - An upcoming travel blogger can use emoticons related to the topic of the tweet and must try to use positive emojis as much as possible.

3.1.2 Text analysis based on user accounts:

The user accounts considered for this purpose have been made based on the number of followers, or the unique niche of the travel bloggers. Generally, in the social media world, popularity is often quantified by the number of followers. The text has been extracted from the Twitter accounts of popular travel bloggers. I will also consider the official Twitter account for the 'Meet South Africa'campaign i.e. South Africa. The screen names used in Twitter are denoted by any name beginning with '@'.

(i) The most frequently used hashtags by the account 'South Africa'-

Key insight - These hashtags can be used by any travel blogger who wants to blog about South Africa or to find information about travel in South Africa. So, it can be used in either way.

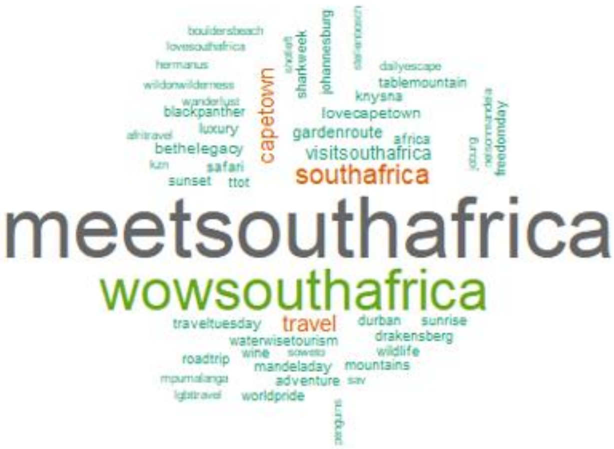

Figure 41: Top hashtags in tweets posted by user account 'SouthAfrica'.

(ii) Most replied to by the account 'South Africa' –

Most travel bloggers or users reply to business or other bloggers whom they want to engage in their tweets. These are either some of the travel bloggers who are associated with the '#meetsouthafrica' campaign or some other blogger who has collaborated with South Africa Tourism for marketing purposes or the businesses who might be interested in the tourism of South Africa.

Figure 42: Most replied to users by user account 'SouthAfrica'.

Key insight – These mentions can be tagged in the tweets by any upcoming travel blogger to create more engagement among other businesses to collaborate with.

(iii) Types of tweets by the account 'South Africa' –

The 'organic' tweets denote the tweets generated by the user. The organic tweets make up most of the tweet percentage for the account followed by replies to tweets and retweets. This indicates how travel bloggers frequently tweet to become visible to the followers. This shows how an upcoming travel blogger with a desire to stand out from the crowd must tweet frequently to remain visible.

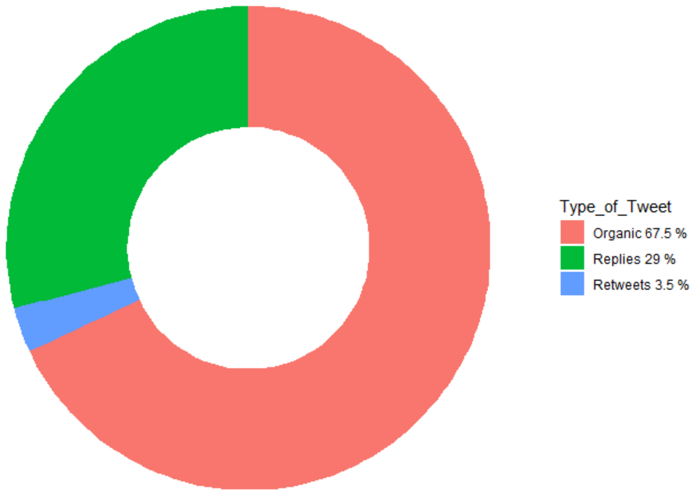

Figure 43: Type of tweets posted by user account 'SouthAfrica'.

Key insight – In order to earn followers and remain visible among them, an upcoming travel blogger must mostly tweet frequently rather than replying or retweeting.

(iv) The types of devices from which these tweets are generated from the account 'South Africa'-

The figure below shows that the majority tweets around 71% are generated or blogged via Hootsuite, which is a social media marketing and management tool. It helps to respond instantly to customer's feedback and questions and lets one track many social media channels at the same time. Hence, Twitter users who are promoting any brands prefer to use Hootsuite. Almost all popular travel bloggers can be seen to use Hootsuite to post tweets. It is also very interesting to see that iPhones are used the least for tweeting although mobile phones are the most used and reachable devices available.

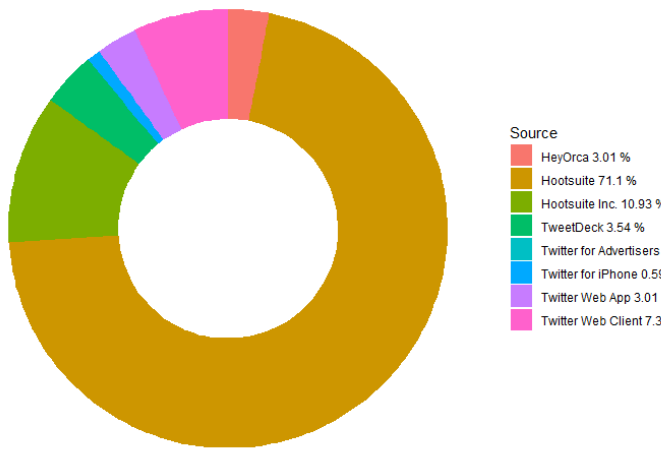

Figure 44: Types of devices from which tweets were posted by user account 'SouthAfrica'.

This figure also indicates the social media tools that an upcoming travel blogger might use to maintain their multiple social media channels.

Key insight – An upcoming travel blogger can use other social media management tools like Hootsuite in order to manage their several network channels.

(v) The frequency of tweets aggregated by year –

The campaign was launched in 2013 and it proved to be very successful in the social media marketing world. The figure below indicates how the number of tweets has decreased in the years followed. In the later section, I will also see how this decline in the frequency of tweeting also affects the tweets being liked by the followers. This is not very ideal in the case of a popular travel blogger as the frequency of tweets must be maintained to be always visible among the followers.

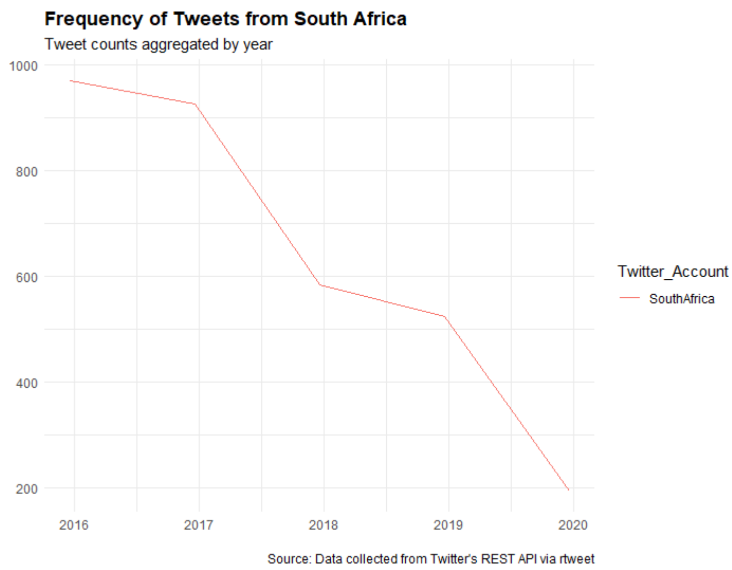

Figure 45: Frequency of tweets by user account 'SouthAfrica'.

(vi) The frequency of favourited by users-

Since the tweeting frequency has gone down in the years, thus the favouriting frequency by followers has decreased significantly as well. Hence, I can say that to engage the followers and like the tweets or to increase the number of followers, the bloggers need to maintain the pace of tweeting.

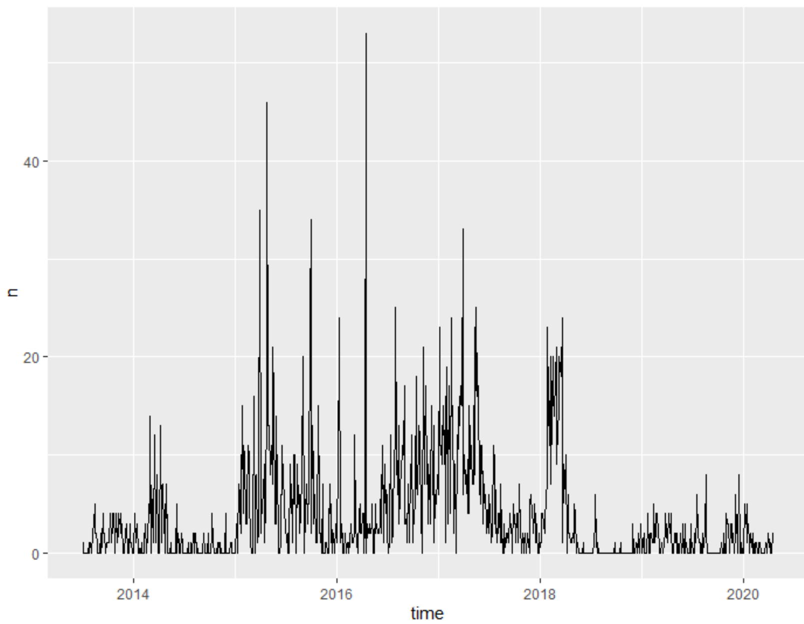

Figure 46: Frequency of likes to tweets posted by user account 'SouthAfrica'.

(vii) The frequency of tweets on a usual week –

For the account South Africa, this frequency can be said to be very low during weekends while quite good during weekdays. This proves the rare usage of emoticons on weekends than on weekdays.

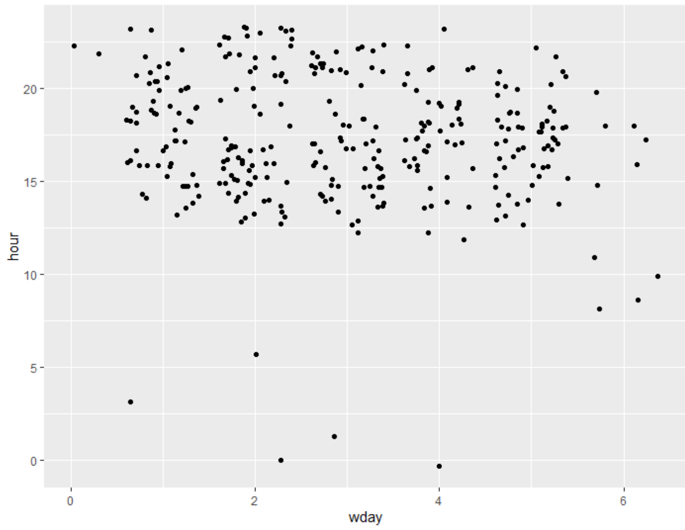

Figure 47: Frequency of posting tweets by user account 'SouthAfrica' in a week against hours.

If I compare this frequency of tweets to that of a very popular travel blogger with millions of followers, then their frequency of tweeting is much higher, and it lasts throughout the week irrespective of weekends and weekdays. They always try to keep themselves noticeable. The plot below shows the frequency of tweeting of Richard Branson, who is a verified Twitter travel blogger with the blue checkmark badge and millions of followers. The blue tick indicates that the account is authentic.

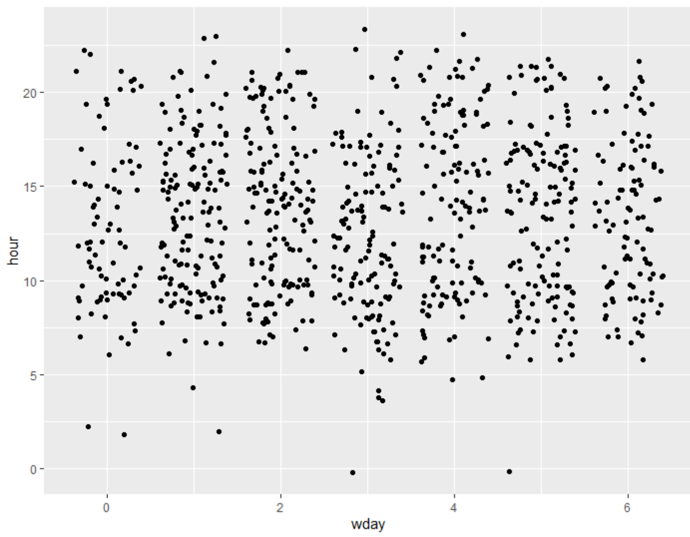

Figure 48: Frequency of posting tweets by user account 'richardbranson in a week against hours.

Key insight – In order to stand out among other travel bloggers, to earn a greater number of likes by their followers and remain visible, an upcoming travel blogger should maintain a constant pace of tweeting irrespective of weekdays and weekends.

(viii) Sentiment analysis of the tweets generated by 'South Africa' account –

The lexicon used for this analysis is NRC and it illustrates how the travel bloggers always post positive tweets followed closely by trust, anticipation, and joy, which also indicates positivity. The NRC lexicon lets me see more details about the emotions used in the tweets posted by the account. Hence, I preferred the NRC lexicon over the Bing lexicon, for analysis in this section to explore some more emotions that can be embedded in the tweets.

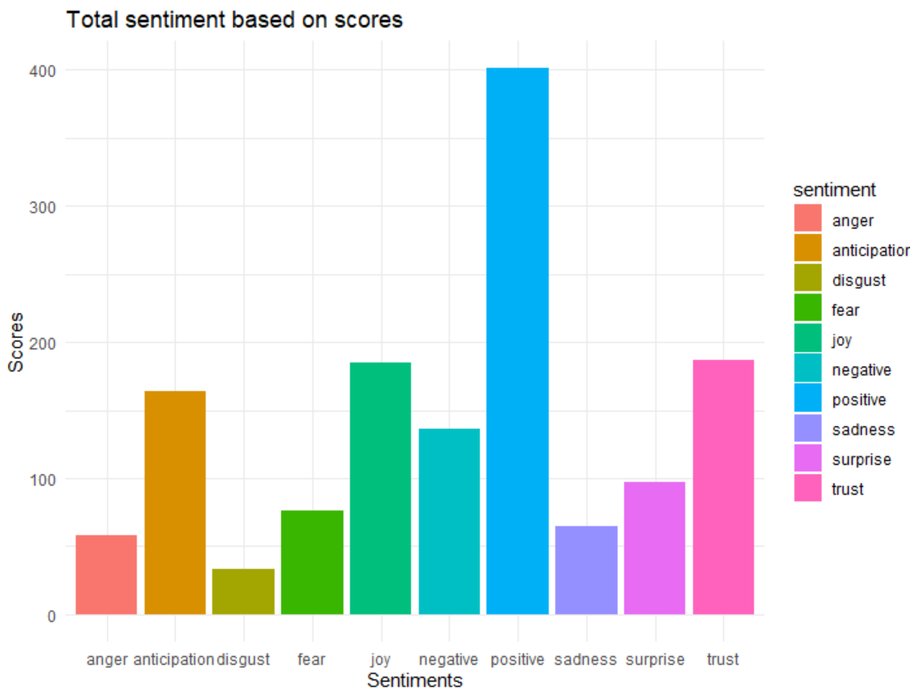

Figure 49: Sentiment analysis of all the tweets posted by the user account 'SouthAfrica'.

(ix) Sentiment polarity with time –

The sentiment polarity can also be seen to be maintained to be positive over time. Even in case of poor global circumstances, like the Covid-19 pandemic, the travel bloggers always seem to post positive blogs or tweets except in Feb 2020. These positive sentiments can be surprise, trust, joy, anticipation, or just positive, in case I use the NRC lexicon. Hence, the upcoming travel bloggers must maintain a positive polarity in their tweets, irrespective of circumstances.

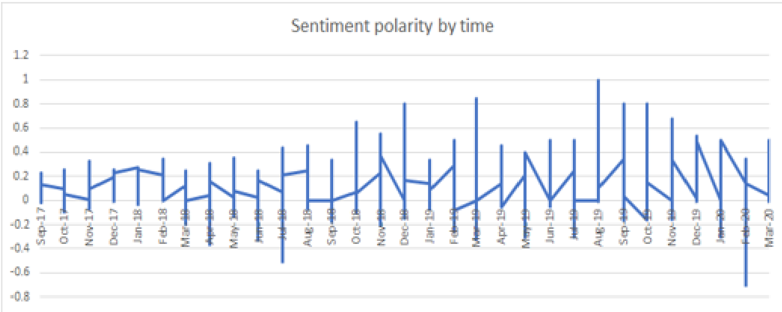

Figure 50: Sentiment polarity of tweets with time posted by user account 'SouthAfrica'.

(x) Sentiment polarity of the account description-

The captions or descriptions that people use to describe their user accounts on Twitter also attract the attention of followers. They are like taglines or one-liners that describe the user or the account. The sentiment analysis of the account description also indicates that people tend to use more positive words in the description. These descriptions have been extracted from 30 popular travel bloggers from Twitter.

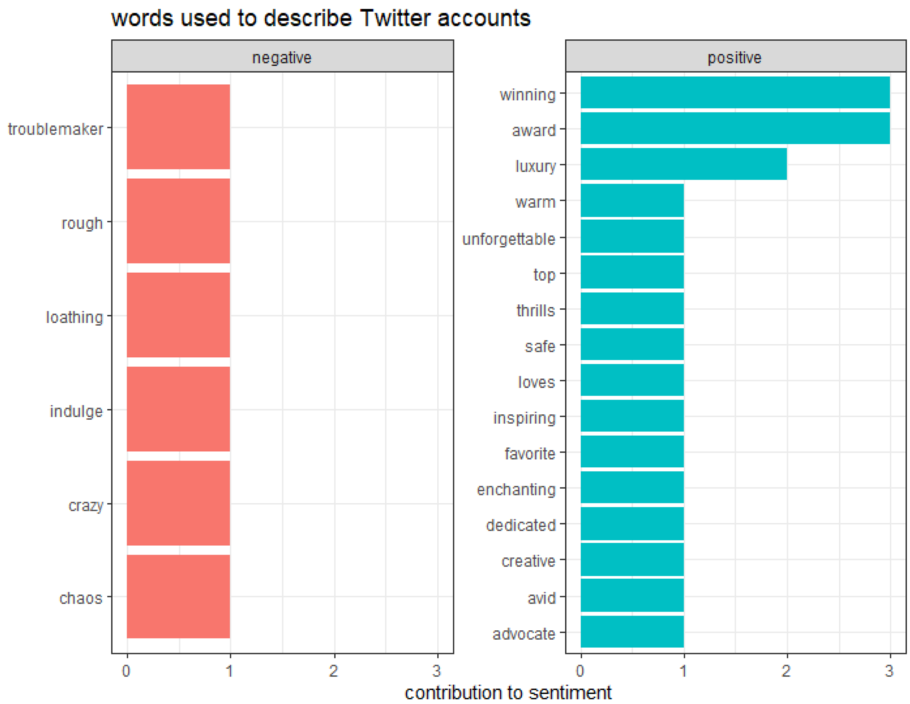

Figure 51: Sentiment polarity of user account description of the top 30 travel bloggers.

Key insight – Not only a positive sentiment must be maintained in the tweets, but an upcoming travel blogger must overall maintain the positivity be it the account description or any other tweets generated by the account.

(xi) Modelling based on fundamental attributes of a Twitter account –

The Twitter user profile has some fundamental attributes like the number of friends, the number of followers, status count, account age, account description, screen name, and the average tweet. The average tweet is the total number of status count divided by the account age. I have collected the data from the top 30 travel bloggers and tried to create a linear model with the followers as the response variable to understand if the other fundamental attributes contribute towards the number of followers. I have performed a log transformation on the model to get a better model.

```
Call:
lm(formula = log(Followers) ~ status_count + Friends + Account.age +
    Average.tweets, data = caps)

Residuals:
    Min      1Q  Median      3Q     Max
-3.7584 -1.0069 -0.1958  0.6246  4.6874

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     5.646e+00  1.684e+00   3.354  0.00301 **
status_count   -3.394e-05  4.129e-05  -0.822  0.42038
Friends        -4.542e-05  1.028e-04  -0.442  0.66321
Account.age     1.351e-03  5.025e-04   2.688  0.01376 *
Average.tweets  1.480e-01  1.421e-01   1.042  0.30948
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.071 on 21 degrees of freedom
Multiple R-squared:  0.318,	Adjusted R-squared:  0.1881
F-statistic: 2.448 on 4 and 21 DF,  p-value: 0.07802
```

Figure 52: Linear regression model for user account parameters of Twitter.

Key insight - Although only 19% of the variance is explained by the model, yet it shows that account age can be a significant predictor in determining the number of followers. The residuals distribution indicates a poor fit of the model illustrating the probability of other underlying factors.

3.2 Network analysis:

The network analysis of a social network illustrates the social structure with the help of network and graph theory. The individual people, organizations, or other entities of interest represent the nodes and their relationship is denoted by the edges. It helps us understand a community by the relationship and ties among them. I will use nodeXL Pro and Gephi for social network analysis. To get the social network of '#meetsouthafrica', I have used nodeXL Pro because of the requirement of historical data. These included the retweets, replies, and likes involved with the hashtag. The communities have been created in Gephi by applying modularity. The communities have been labelled with different colours so that I can distinguish them from each other. The nodes belonging to one community indicate that those nodes tend to connect more densely compared to nodes of other communities. This shows how the nodes belonging to the same colour tend to retweet, reply, or like the tweets and blogs of the other nodes of the same colour. The nodes have been labelled according to the size of the community and hence I can conclude that 'gotosouthafrica' is the biggest community. The upcoming travel bloggers who want to blog about their travel stories about South Africa can mention this user to get more engagement and likes.

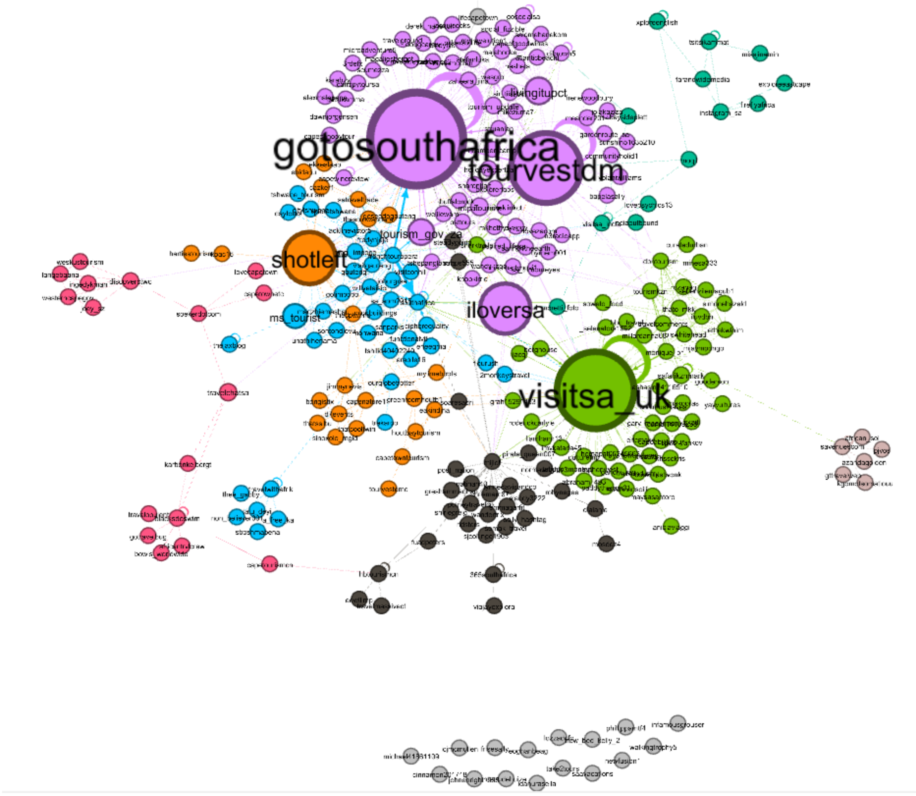

Figure 53: Community detection of retweets, replies and likes of tweets containing #meetsouthafrica.

In Gephi, I can create graphs from online streaming data from the Twitter Importer plugin. I kept the plugin open for a short time period like 30 minutes to capture the communication among the top 20 travel bloggers. I will measure the popularity based on the number of followers. From the created graph, I then selected the giant component so that the bloggers who did not engage at all within that time period, is filtered. From the figure below, I can see that the travel blogger community is quite strongly connected. They often engage in each other’s blogs, tweets, and communications.

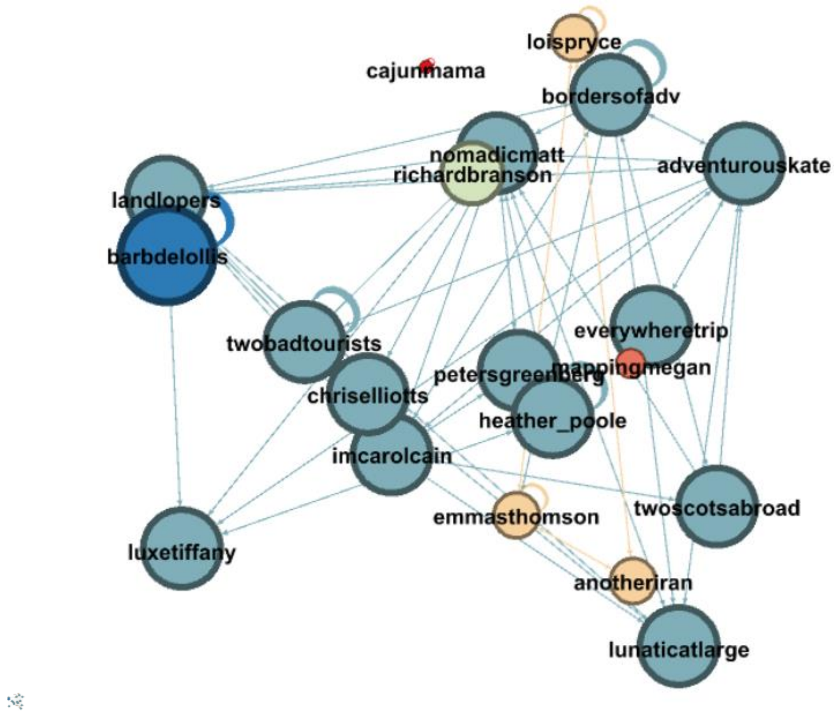

Figure 54: A strongly connected network of the top 20 travel bloggers.

I have also performed community detection on the top 20 travel bloggers (in terms of the number of followers) to find their communities. The nodes' labels represent the Twitter screen names of the bloggers.

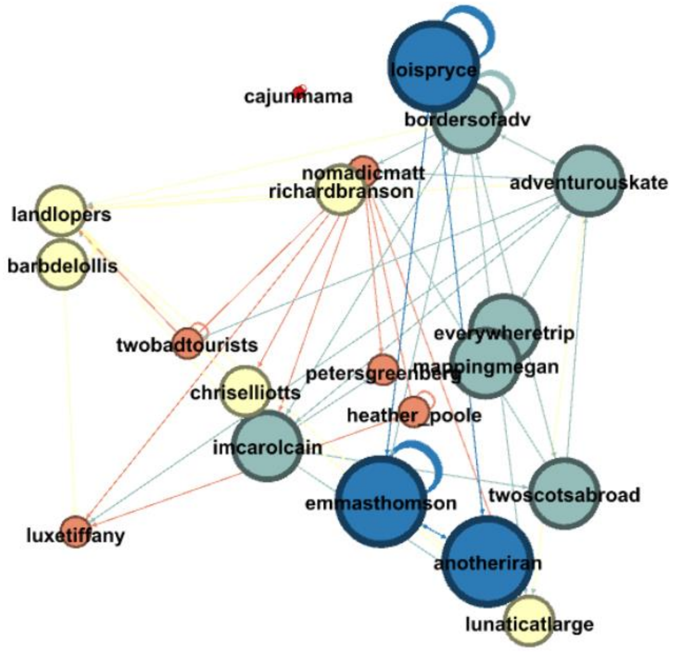

Figure 55: Community detection of the top 20 travel bloggers.

Key insight – The mentions, which we mentioned earlier saying that mentions tend to increase the engagement of tweet, is associated with the networks of the tagged people. The bigger the community of the mentioned people, greater is the probability of engagement. Hence, while mentioning other travel bloggers, an upcoming travel blogger should mention users who have bigger communities. Moreover, an upcoming travel blogger should reply and retweet frequently to their peers in order to maintain a strongly connected network.

3.3 Image Analysis:

Images or visuals always draw more attention to any content compared to only text. But the images shared in social media must be shareworthy and engaging enough to call for the loyalty of the followers and communicate vital details as well. If the images look unreal, then the main intention of the image is never achieved. Since a travel blogger is always trying to connect to a potential traveller's emotion and induce a genuine feeling of being in the same space themselves, the photos must look real. There are many characteristics of a good photo, the most common being a display of emotions, subtle storytelling, perfect lighting, and colour and capturing unique and iconic moments. From a layman's perspective, it must be well lit, easy to interpret and look real. Moreover, studies and analysis by marketing websites have revealed how images boost the engagement on Twitter (AdEspresso by Hootsuite,2020).

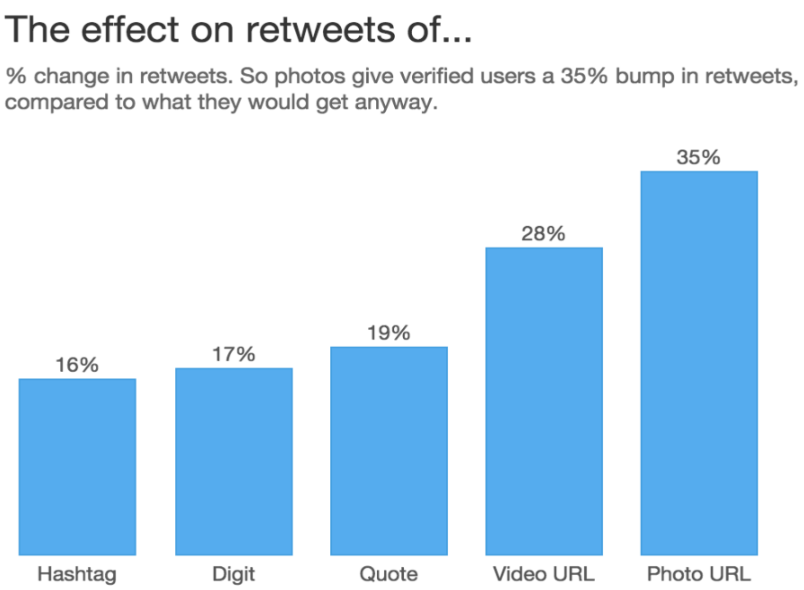

Figure 56: The effect of images on increasing engagement via retweeting (Kevan Lee, 2020).

The main issue with social media images is that they look unreal to most. The images put up by travel bloggers sometimes do not make the potential traveller engage or mentally put themselves up in that picture because they do not connect emotionally to the image. I have examined a random sample of 20 images from popular Twitter users. These images have been preliminarily accessed to check if they are photoshopped or not or whether they are real and trustworthy. I will try to verify by looking at the average colour composition and looking for spots that don't fit. Principal Component Analysis has been performed on the image. The projection of the image in the 2nd principal component gives a grayscale image where white or bright spots signify being photoshopped. Some remarks and statistics of the image analysis is as follows-

Sr.No	Screen Name	Type of account	Images analysed	Remarks
1	SouthAfrica	Blogs based	10	5 images with minute editing, 5 images with no editing at all.
2	EarthPics	Images based	5	Only official logo added with very minute editing.
3	NATGEO	Images based	5	Only official logo added with very minute editing.

Table 1: Remarks on the analysis done on images from Twitter.

Actual Image 1: The below image is a very popular image related to the '#meetsouthafrica' campaign because it was able to generate 31,500 views within the first 12 hours of its uploading.

Figure 57: The famous image of #meetsouthafrica campaign which earned 31,500 likes within the first 12 hours of uploading.

The image's 2nd principal component reveals –The bright white colour of the sky represents being edited. The rest of the image has not been hampered with at all.

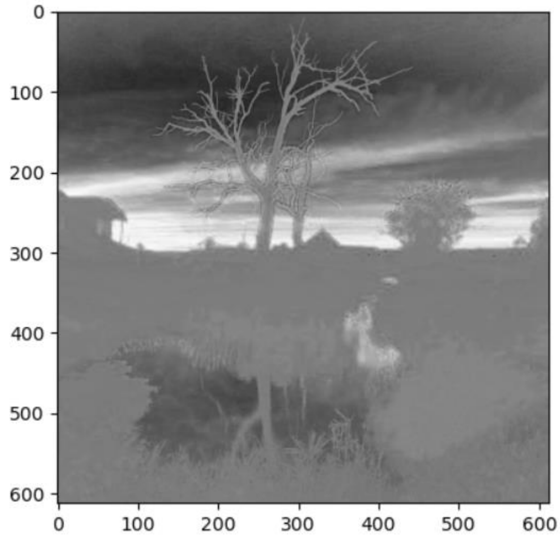

Figure 58: Bright white spots indicating the parts being photoshopped.

Actual Image 2:

Figure 59: A random image from #meetsouthafrica tweets.

The image's 2nd principal component reveals –

The below image shows that the picture has not been tampered or photoshopped at all.

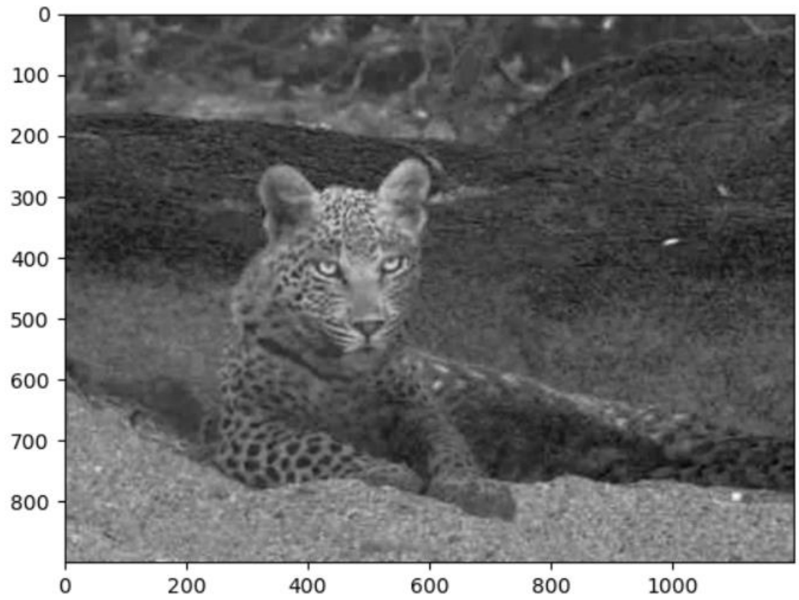

Figure 60: No bright white spots to indicate being photoshopped.

These images have been extracted from the official Twitter account of the '#meetsouthafrica campaign. Hence, I can conclude from the above analysis that the images are mostly real with minute touches of editing. There are many Twitter users like EarthPics and NATGEO, which promote image specific travel blogging and can be followed by any potential traveller who wants to choose a destination based on the pictures and visuals, which by our analysis have revealed that they are authentic and real.

Key insight – Since images play a significant role in capturing the attention of followers to tweets and blogs, an upcoming travel blogger must be very careful with the amount of editing that they perform on their images. The images must be photoshopped up to the extent where they still look real.

3.4 Online survey:

The online survey was designed to capture the opinions of the people-centred around travel blogs. The survey gave us the responses of a random sample of 130 participants from all over the world. Since convenience sampling has been performed to get the online survey data, I cannot assume that it represents the actual population. Nevertheless, it gives us a picture of the perspective of the people about travel blogs and their requirements for information in the planning phase of a holiday.

(i) Why do people travel?

The visualization shows that people mostly like to travel for leisure and exploration followed closely by adventure. Hence, the travel bloggers can write their blogs centered around such topics so that the potential traveller can get the desired information from the travel blogs.

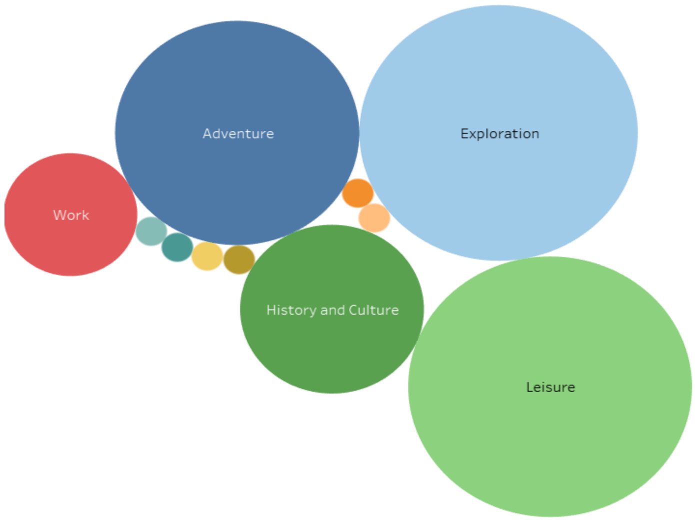

Figure 61: Reasons for which people love travelling.

Key insight – An upcoming travel blogger can select any of these topics like adventure, history and culture or leisure related to destinations while writing blogs or tweets.

(ii) Which category of occupation attract which type of holidays?

The figure below illustrates how budget holidays are preferred over other types of holidays no matter what the occupation is. Out of all the occupation types, employed people like to travel more followed by students. Thus, the travel bloggers can concentrate more on budget holidays rather than exotic ones. The niche that the travel bloggers want to adopt for their blogs can target a certain type of occupation and write about their requirements regarding travel as well.

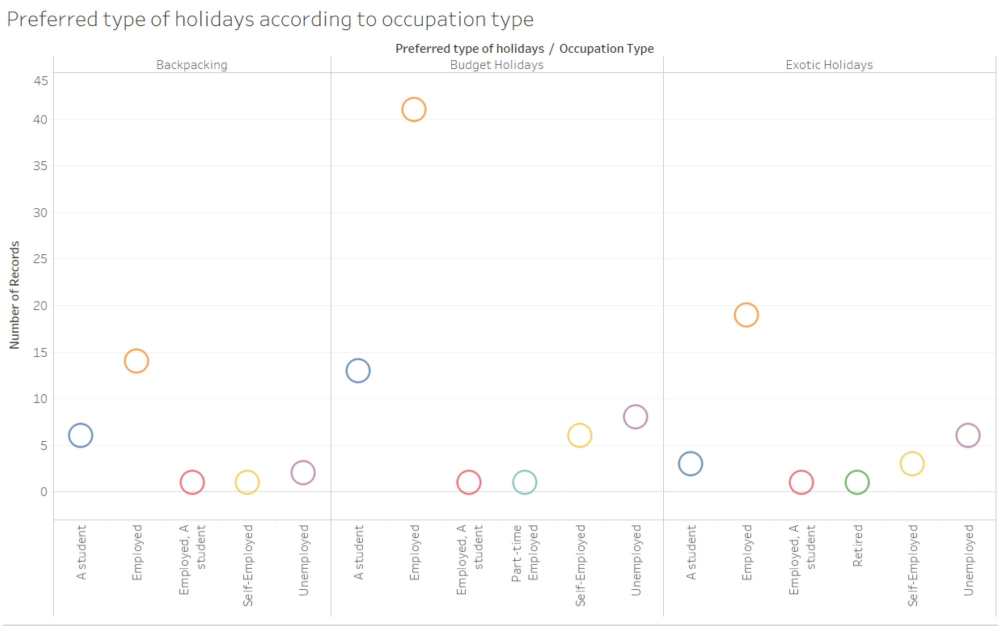

Figure 62: Preferred type of holidays according to occupations.

The statistics of the different occupation types are as shown in the table below. Thus, I can also conclude that employed people dominate in all types of holidays as most of the participants in the survey are employed.

Occupation types	Count
A student	22
Employed	75
A student, employed	3
Part-time employed	1
Retired	1
Self-employed	10
unemployed	16

Table 2: Counts of different occupation types participating in the survey.

Even in budget holidays, the travellers have choices. Some like renting apartments, while some prefer budget hotels. The adventurous people like to accommodate in camps and hostels. The figure below shows the popularity of renting apartments and budget hotels among the budget holidays loving people. The travel bloggers can blog about the specifications of such type of accommodations to gain popularity among its followers.

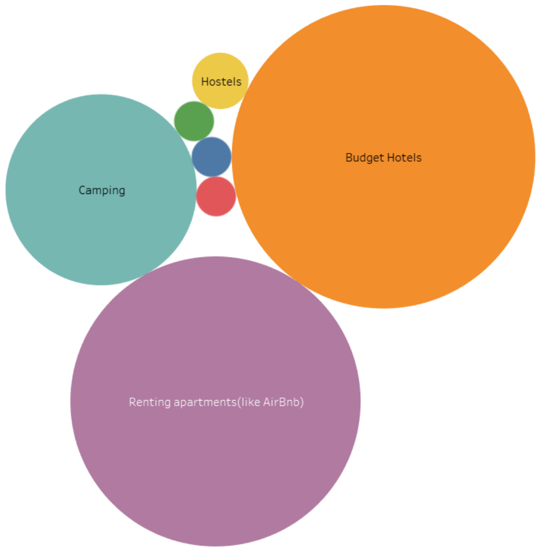

Figure 63: Preferred type of accommodation styles within budget holidays.

> **Key insight -** When providing information about the destinations in the blogs or tweets, the travel bloggers must include budget details. Although people like exotic holidays as well but budget accommodation and budget holidays are preferred more among all categories of professions.

(iii) What type of information source is used during the planning phase of a holiday?

The survey shows that around 90% of the people like to research before going on a holiday. While conducting the research, different types of people look for different details. Some travellers like to explore unique places while some like to mingle with the local culture and people. There are many sources to get different types of information. There are official websites of the destinations which give information about each aspect of the destination while there are websites like Trip Advisor and Lonely Planet which have blogs about the different types of travel, food, and explorations, inclusive of the reviews of previous travellers. The survey shows how people like following travel blogs on social media the most, followed closely by word of mouth from friends and peers and the other possible sources of information.

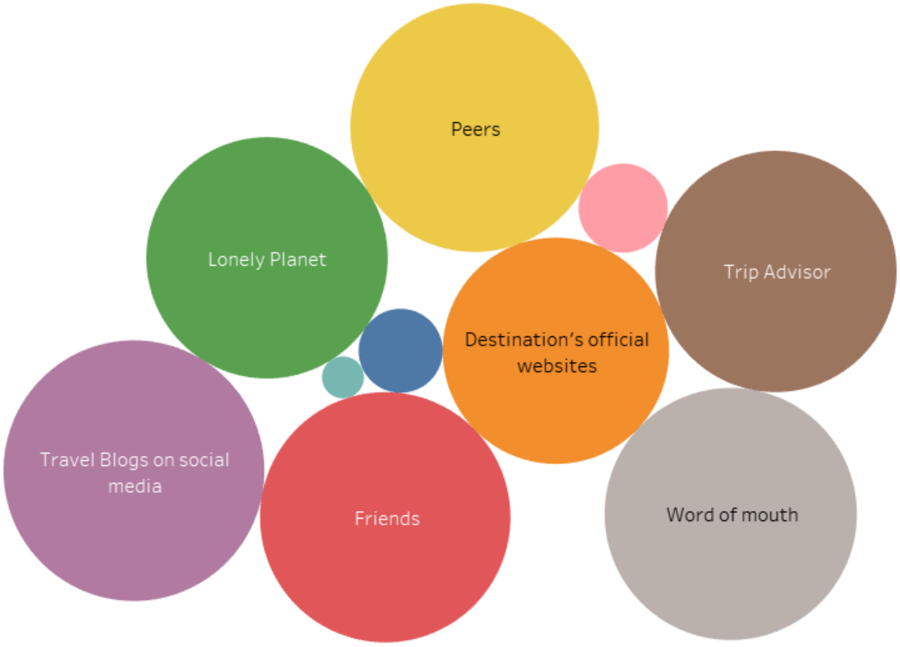

Figure 64: Resources used by people in the planning phase of a holiday.

However, there exist people who do not like to conduct any sort of research before a vacation or holiday and opt for the other options available. Most of such people prefer availing of the customized travel packages of the travel companies. These companies have standard packages as well as customized ones for almost all destinations. These packages are perfect for people who just concentrate on leisure and exotic holidays.

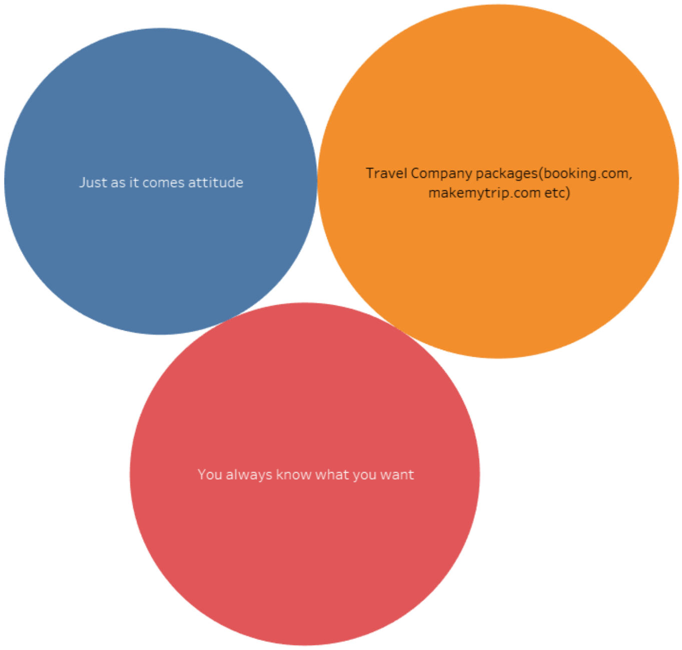

Figure 65: Resources for people who don't like to research.

Key insight – Since the survey shows that travel blogs do play a significant role in the planning phase of a holiday, hence the travel bloggers must be creative and informative enough in order to provide all the necessary information regarding the destinations and to serve the main purpose. Furthermore, the survey also shows how some potential customers prefer travel packages more and hence the bloggers can collaborate with the travel companies to design customised packages for the travellers as well.

(iv) While travelling, what do people expect to enjoy the most?

The figure below shows how people get tempted to choose a destination because of its images. The real power of images lies in the fact that people tend to put themselves mentally in the image and feel magical in it. Hence, the images or pictures in travel blogs carry a lot of weightage in tempting people to choose a destination. Food and wine as well as history and culture contribute a lot towards the decision of choosing a destination. Thus, these can also be a niche in the travel bloggers' blogs.

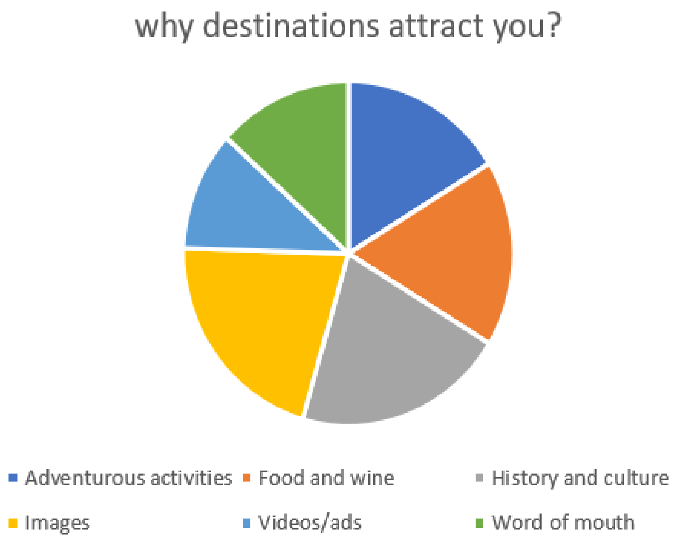

Figure 66: Facets of a destination that attracts people.

Key insight – The survey shows how important images are to people as far as travelling is concerned. Thus, an upcoming travel blogger must post pictures related to the destinations to draw more attention from the followers.

(v) Characteristics of a travel blog that attracts a potential traveller –

The survey indicated that around 60% follow travel bloggers on social media. Different people have different criteria for selecting their favourite travel blogger. Some followers love the creativity and perspective of a travel blogger whereas some like the way they capture and present their stories of travel. The survey showed how people mostly like the storytelling feature of a travel blogger the most followed closely by the perspective. The videos of travel also manage to grab the attention of a follower.

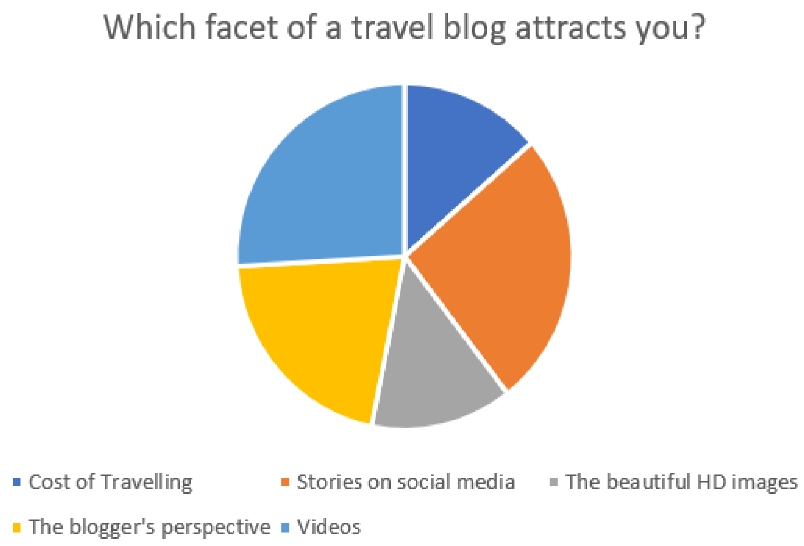

Figure 67: Facets of a travel blog that attracts people.

The people who do not follow any travel bloggers and mostly prefer travel packages have several reasons for disliking the blogs. Some of them are as follows. The most popular reason for disliking blogs is unreal pictures. Other reasons are the blogs are unnecessarily long with no important information and without creative storytelling and, they seem too fancy and hence undoable. Thus, the travel bloggers, while writing blogs must keep in mind these factors of dislike among the followers.

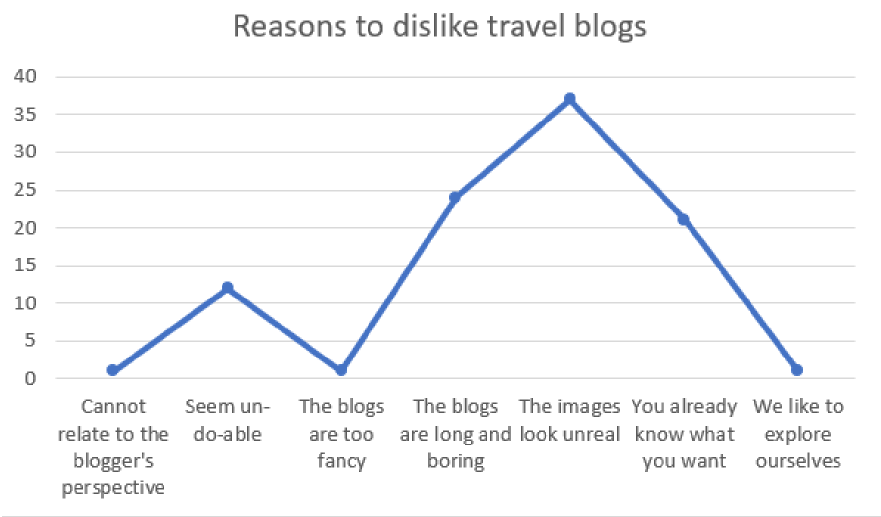

Figure 68: Facets of a travel blog that people dislike.

Key insight – The likes and dislikes regarding travel blogs indicate that an upcoming travel blogger must not only write creative blogs but also upload images with the appropriate amount of editing.

4 Conclusion:

To summarize, analysis of the blogs of the popular travel bloggers highlighted the different facets that make the followers engage more with their tweets. The text mining revealed how the travel bloggers utilise hashtags and mentions to make their blogs more visible. The sentiment analysis indicated how the travel bloggers always maintain a positive sentiment throughout in their tweets. Also, the text clustering and the topic modelling revealed the general subjects or topics that bloggers talk about in their blogs. Besides, the frequency of tweeting always plays a major role in making a blogger stand out. Hence, I can conclude that an ideal tweet by a travel blogger would be about some creative topic, a positive sentiment attached to it, have an image with the appropriate amount of editing and several hashtags and mentions included as well. The network analysis of the travel bloggers related to the campaign as well as that of the other popular travel bloggers of Twitter revealed that an upcoming travel blogger should use mentions of those users with giant communities of followers. Also, the network analysis highlighted that a travel blogger should also engage in other bloggers' tweets to maintain a strongly connected network. Furthermore, the online survey gives us an idea about the likes and dislikes of a potential traveller regarding travel blogs as well as travelling. It illustrated the preference of budget travelling and images to decide on the destinations. However, there are certain limitations involved in the dataset as Twitter allows only the latest 3500 tweets per user, to be extracted. In addition, another limitation is the sampling technique used in the online survey as convenience sampling might not represent the actual population. I recommend further analysis into the networks of popular travel bloggers to get insights about the effect of communities in the creation of social influence. From the entire analysis and project, an upcoming travel blogger can have an idea about the tentative features of a blog that they can include to make their travel blogging profession successful.