



**IS6051: Business Data Visualization**  
**Prof. Selja Seppala**

**Big Data Analysis Report**

**Presented By:**

**Abhiram Tamraparni (121109050)**

## **Executive Summary:**

Social media has become a big hub which has a lot of information. Useful insights can be extracted from it by carefully segregating the useful content. Many social media platforms like Twitter, Facebook and Instagram have become a playground for everyone. They are filled with influencers, bloggers, and account pages of almost all the companies in the world. They have become a mode of communication between the consumers and the companies. The reviews of many products and services provided by many companies are given by customers and social media influencers. All these information can be extracted, cleaned and analysed to understand the sentiments of the consumers regarding the products and services. This will help the companies to take decisions based on the results to improve and upgrade their products and services to satisfy the needs of customers. By doing this the marketing and operational costs can be managed properly and a better customer service can be provided. Thus increasing the popularity of the company and increasing the revenue to make huge profits.

The task of Trending Now consultancy is to analyse the Twitter social media buzz of Formula1 which is the most expensive motor racing sport in the whole world organised by International Automobile Federation (FIA). The races takes place in 22 countries spread across the whole year. Globally on an average there are 70.3 million viewers per race ([www.formula1.com](http://www.formula1.com), n.d.). This opens up huge opportunities to market and advertise. There are loads of sponsors who market their companies on the F1 cars and boards. The main goal of the sentimental analysis of Formula1 is to gather information and analyse different social media trends and how these trends can be used to market and advertise the Formula1 merchandise to viewers.

For the social media sentimental analysis, Trending Now consultancy has taken data of #Formula1 of Formula1 from Twitter to analyse the trends. A huge amount of data was extracted from Twitter using R programming language. The dataset was then cleaned to remove unwanted and redundant data and converted into usable format. 5 KPI's were formed from looking at the sample tweet data. The KPI's were chosen in a way so that the executives can understand and gain useful insights from it and help them make informed decisions from them. All the trends identified from the KPI's were projected using graphs in Tableau. Dashboards were created in order to better understand the statistics of the analysis which would help in understanding the trends. Some recommendations based on the results were also laid out by the Trending Now consultancy to churn out maximum profits by upgrading the advertising and marketing

strategies. This report contains Introduction, The steps used by Trending Now consultancy to extract, clean, analyse and display the trends of Formula1 on twitter. This report also contains recommendations provided by the consultancy based on the results from the analysis carried out.

## **Introduction:**

Formula1 was started in the year 1950 (Formula One Art & Genius, 2010). It is a series of 20-22 car races held in 20 countries all over the world. This event is taking place every year since its start. Formula1 is the richest motor sporting event in the whole world and has millions of loyal fan following. Formula1 has their own collection of products consisting of t-shirts, jackets, shoes, caps and many more types of apparels. Puma is the company which sells exclusive merchandise of Formula1. By analysing the twitter trends and performing sentimental analysis on them, the F1 merchandise can be marketed better to parts of the world where it is required. This will help in increasing the sales of the apparels and in turn increasing the popularity of Formula1 in those regions.

Digital marketing has become very popular and effective to reach out to the consumers. By analysing the Twitter data, the executives of F1 can upgrade their marketing strategy to gain customer base. Better advertising can also be done by looking at the results of analysis to maximise the sales of the products and increase the popularity of Formula1 across the globe.

## **The Analysis:**

**Identifying relevant data:** In order to carry out the analysis first and most important thing is to identify the data. The hash tag #Formula1 is used to get the tweets from Twitter. This is the official hash tag of Formula1 and this hash tag is used by almost everyone who tweets about it. There are many different hash tags like all the teams hash tags which are the subset of Formula 1. #Formula1 has a broader spectrum of tweets and was chosen for the analysis.

**Extracting data:** R programming was used to extract all the data from Twitter. The library named 'twitterR' present in R was used to access the tweets through its official API. The elevated access of Twitter developer account was set up in order to get API keys and secret keys to access the Twitter data content. The hash tag #Formula1 was passed to the searchTwitter method and the value was

set to get at least 10000 tweets. This increases the number of tweets in-order to perform better and accurate analysis.

**Cleaning the data:** The extracted data was loaded into a newly created data frame to clean it. All the redundant and unwanted fields were removed to make the data more understandable. The repeated tweets were also removed to get clear results of analysis. Finally the data was saved as a .csv file in order to import it into Tableau data visualisation tool for further analysis.

```
# Authentication variables are used for authentication of Twitter API's
api_key <- 'eBqic5FCJ4bzSGEMpsEpoSwi3'
api_secret_key <- 'BuHF8fMdDiyND2t3cx67Eu7ZspTYHutsTCAY5P2n8VZo0i9UVS'
access_token <- '1515635028419657734-C4Kwe6NigUwtSpMInZ1T00fmjMv2gu'
access_token_secret <- 'v1ZDMLsjcMVdpoa7vLfPjc0S9cwvgnUwuYaMom4Gvt3AC'

# Twitter Library is loaded
library (twitter)

#Twitter connection is made via API
setup_twitter_oauth(api_key, api_secret_key, access_token, access_token_secret)
tweets<- searchTwitter('Formula1', n=10000, lang = 'en', since="2021-03-09", until="2022-03-09")

# re-tweets are removed
noretweets <- strip_retweets(tweets, strip_manual = TRUE, strip_mt = TRUE)
summary (noretweets)

#Convert the stored tweets into a multiple data frame
tweetsdf <- twListToDF(noretweets)

#user profiles are found from the above tweets
userInfo <- lookupUsers (tweetsdf$screenName)
userFrame <- twListToDF (userInfo)
summary (userFrame)
userFrame

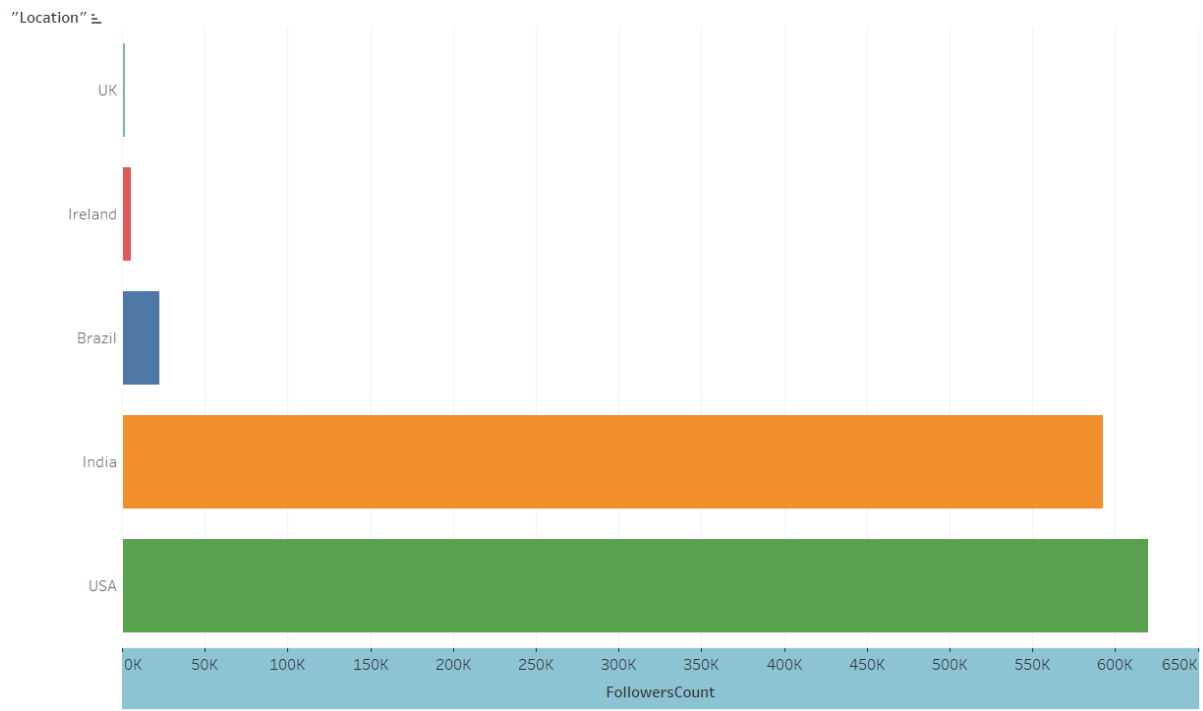
#merging the tweet and user information data frames
finalTweetsdf <- merge(tweetsdf, userFrame, by = "screenName")
finalTweetsdf

#Creating and writing into a csv file
write.csv(finalTweetsdf, "Formula1.Csv")
```

**Data visualisation:** Tableau is the data visualisation tool which was used to display graphs of data by creating a simple dashboard which is easily graspable. The exported csv file from R was imported into Tableau workspace. Necessary filters were added to rows and columns to get a good representation of the data using graphs. The visuals were plotted in a way as to provide good insights and trends present in the data.

## **KPI's used:**

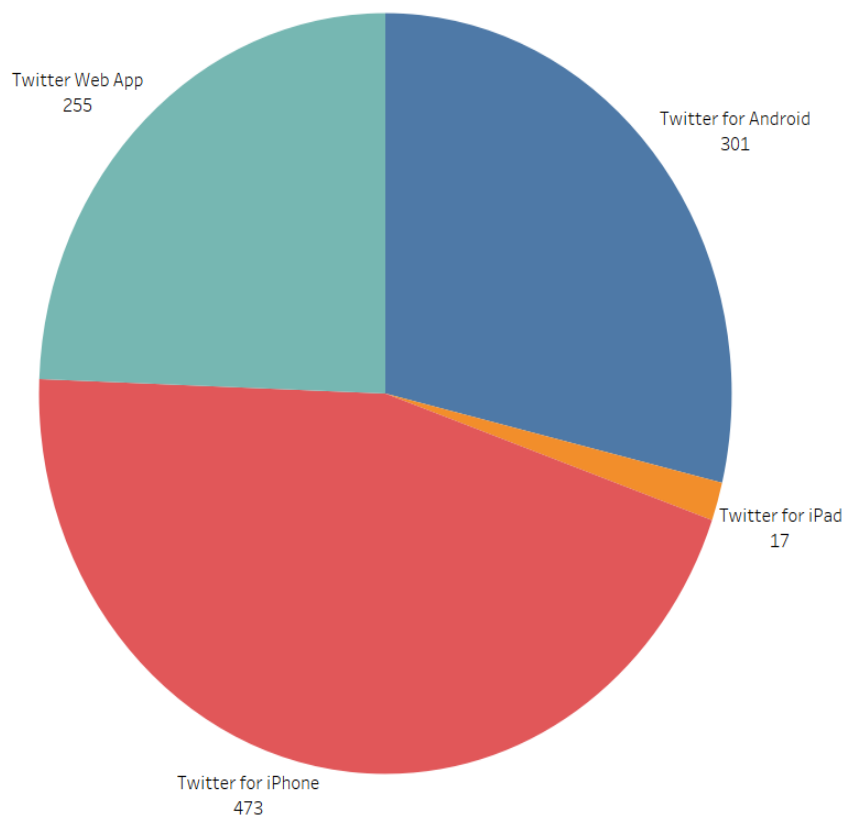
1. **Followers based on location:** A bar graph is used to depict the number of followers of Formula1. The bar graph is sorted in ascending order of number of followers per location.



## **Recommendations:**

It can be identified from the graph that the fan base in UK and Ireland is very less compared to that of India and USA. The marketing team of Formula1 can concentrate on the countries where the popularity of the sport needs to be raised. In-order to do so, a separate marketing campaign must be conducted to reach out to maximum number of people and get them interested in Formula1. More advertising can be done in these regions to increase the number of viewers. Digital marketing can be done on social media platforms like YouTube, Twitter, Facebook and Instagram.

2. **Source of tweets:** There are mainly 4 types of applications from which users use Twitter. They are Android phones, iPhones, I Pad tablets and Twitter web application. A pie chart is prepared to represent the number of users tweeting from these devices. The majority of Twitter users are using iPhones. Android Phones comes second followed by Twitter web and I Pad at last. From this we can see that majority of the users use mobile phones to tweet.

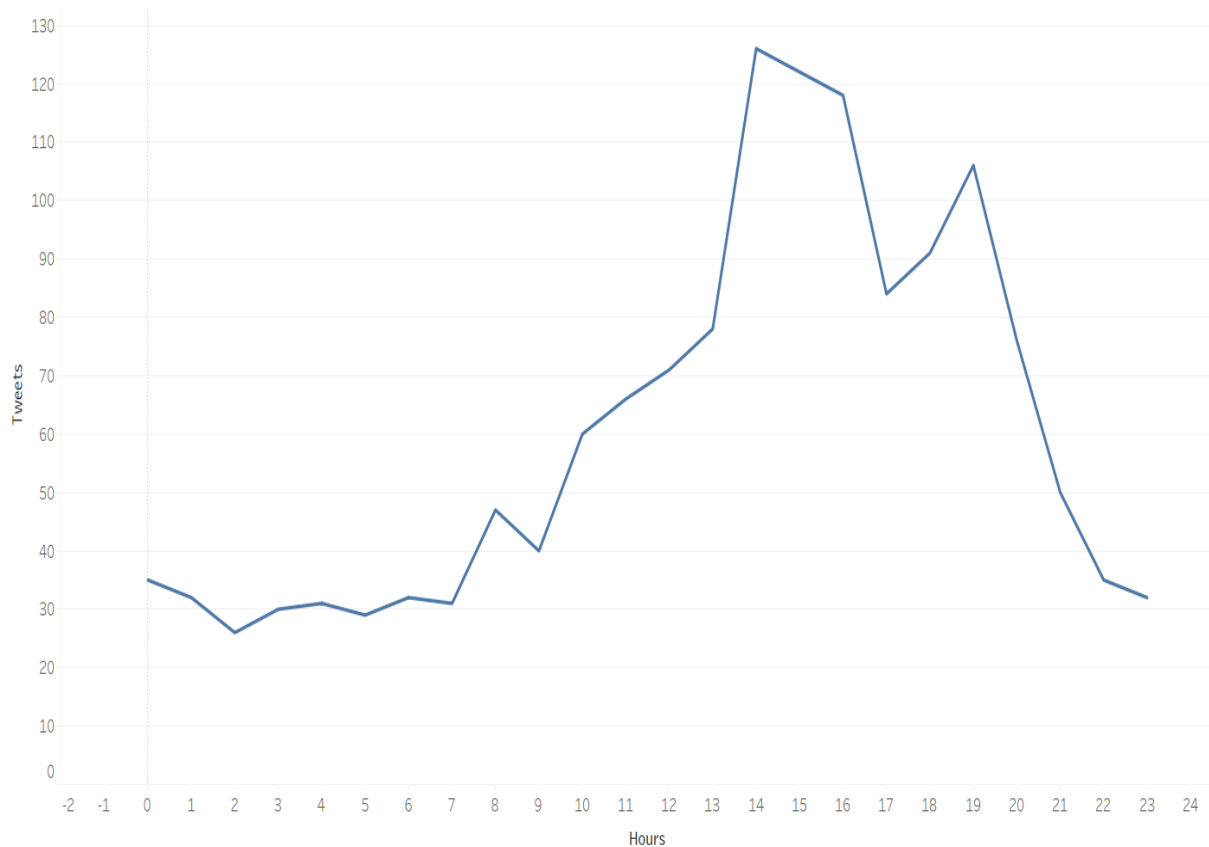


### Recommendations:

Digital marketing and advertising can be concentrated more on android phones and iPhones. The Formula1 merchandise can be advertised more on applications which run on these operating systems. The marketing strategy can be based on these platforms since consumers using twitter web application via PC is very less. More pop up advertisements can be introduced on mobile applications like Facebook and Instagram. F1 merchandise can be recommended on the applications so that users get to know about the products, view them and purchase them. This will help in boosting the revenue of Formula1.

- 3. Activity Time:** In order to find at what time of the day most tweets are posted the created field was used. DATEPARSE function was used to filter out the date in correct format and a new field called timestamp was created. The data was then plotted against the tweets generated. A line graph was used to represent the times when tweets were posted. The line clearly presents the peak and low period of times where the tweets were

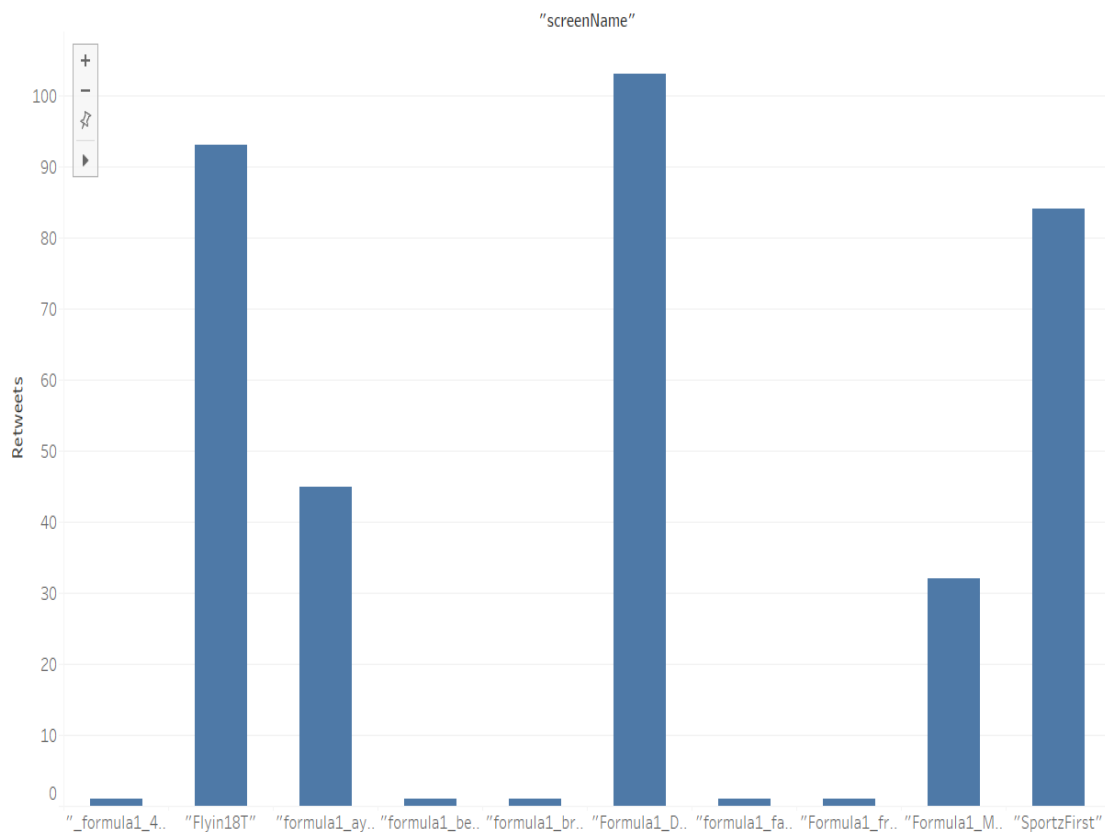
high or low. By examining this graph the peak activity time for tweets regarding Formula1 can be understood.



### Recommendations:

It can be seen from the graph that the peak times when the tweets were posted were between 2pm to 3pm in a day. This tells us that there is a lot of traffic and tweets about Formula1 which are posted in that time period. From this information the marketing strategy can be devised so that consumers receive more advertisements regarding Formula1 and its merchandise during those hours when a lot of users are talking about Formula1 in social media platforms. This will help in reaching out to maximum number of customers.

- 4. Retweets and Influencers:** A bar graph was plotted between the users and the number of retweets which their tweets got. In Twitter it is believed that the number of times a tweet is re-tweeted the more popular it is. By plotting this graph it was understood that which twitter users had the most re-tweets. “Formula1\_Daily”, ”Flyin18T” and “SportzFirst” were the 3 twitter handles which got the most re-tweets. Bar graph is one of the most simple and easiest graphs which can be understood by anyone.

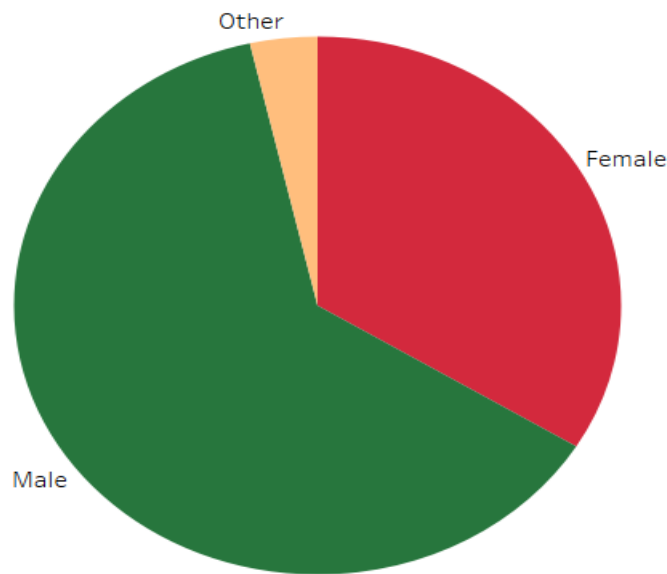


## Recommendations:

It is clear from the graph that the above mentioned 3 twitter handles are very popular among all others. The marketing executives of Formula1 can reach out to their account holders to ask them promote Formula1 to all users of Twitter and on other social media platforms. This will help in spreading the news and latest developments pertaining to Formula1. Influencers are a huge deal in these modern times. People follow them and there evidences where a product got high demand after social media influencers endorsed it. Similar marketing and advertising strategy can be adopted by the executives of Formula1.

- 5. Gender activity:** The number of tweets posted by male, female and other genders was plotted with the help of a pie chart. It can be seen that majority of males tweeted about Formula1 races which female and other genders didn't post as much put together compared to males. This shows that there are less number of females and other genders who are not big fans of Formula1 and they don't follow it as much as the male audience.





### **Recommendations:**

It is very clear that Formula1 is followed largely by male population. In order to attract female and other genders the marketing strategy should be created in such a way as to reach out as much as possible to female and other gender population. The merchandise can be marketed by giving options of size and colours and with different styles of clothing. The executives of Formula1 should work and start a campaign that breaks the stereotype that motor sport only belongs to male audience. This will attract more people to tune in to Formula 1 and buy exclusive Formula 1 merchandise. This will increase the sales and makes the company profitable.

### **Conclusion:**

Formula 1 has a huge potential to reach out and gain fans in every part of the world. Social media is a powerful tool when used wisely. The content present in social media platforms if made use of properly will change the direction in which a company is headed. Data driven approaches can be taken by the companies to excel to maximum heights. Trending Now consultancy have thus analysed Formula1 tweets and have provided recommendations based on the

insights found from the analysis. With the help of results and insights from the Twitter data and KPI's Formula1 can take informed decisions to reach out to maximum number of people and get them interested in the sport. By doing this they can market their merchandise to more people to get good sales and generate excellent revenues and drive the company towards success.

## **References:**

- [www.formula1.com](https://www.formula1.com/en/latest/article.formula-1-announces-tv-race-attendance-and-digital-audience-figures-for-2021.1YDpVJIOHGNuok907sWcKW.html#:~:text=Globally%2C%20the%20average%20audience%20per). (n.d.). Formula 1 announces TV, race attendance and digital audience figures for 2021 | Formula 1®. [online] Available at: <https://www.formula1.com/en/latest/article.formula-1-announces-tv-race-attendance-and-digital-audience-figures-for-2021.1YDpVJIOHGNuok907sWcKW.html#:~:text=Globally%2C%20the%20average%20audience%20per>.
- Formula One Art & Genius. (2010). F1 Origins. [online] Available at: [http://www.f1-grandprix.com/?page\\_id=222](http://www.f1-grandprix.com/?page_id=222).