

A Popularity and Sentiment Analysis of Starbucks' Frappuccino on Twitter

Oscar Zarate, Goutam Raju, Lida Zhang

Final Report
EECS 510: Social Media Mining
Northwestern University

Abstract:

Social media has entrenched itself as an indispensable component of marketing departments the world over. The moment-to-moment nature of data obtained from social media can reveal patterns that are not yet distinguishable from other information sources. For our project, we collected and analyzed a large set of data from Twitter focused on one of Starbucks' most recognizable beverages - the Frappuccino. In particular, we focused on spatial, temporal, and sentiment patterns to extract meaningful inferences from the data. This report details our process, findings and conclusions.

Code and data can be obtained [here](#).

Introduction:

Social media provides us a window into the ideas and opinions of countless people around the world. It provides us an information channel more instantaneous and encompassing than any traditional reporting source ever could. Of course, this enormous volume of information comes at the cost of limited structure. Put simply, social media can be a highly valuable source of data, but without proper processing it remains largely inaccessible. In this project we chose the Frappuccino drink from Starbucks as the subject of our social media information analysis.

Our goal in choosing this topic was to understand consumer behavior and study the patterns in rise and fall of cultural trends. Starbucks has become an integral part of American culture-although its presence is seen in other countries as well as we will see in further sections. In particular, consumers have transformed dining at Starbucks into a very social-media centered experience. Photography of food and drinks from Starbucks franchises for posting onto various social media such as Instagram, Twitter and Facebook has become a very common occurrence-especially with younger customers. It is also to be noted that with the sustained consumer interest the coffee house enjoys, new beverages and food items introduced by Starbucks gain a lot of attention on social media (and even regular media outlets). Starbucks' online and social media advertising also helps drive this phenomenon. With these factors taken into consideration, we decided that Starbucks would provide an interesting foundation to analyze consumer interaction with a well-known and liked brand. We concentrated on the Frappuccino beverages in lieu of the rapid introduction of new flavors and high consumer interest.

Exploratory Data Analysis:

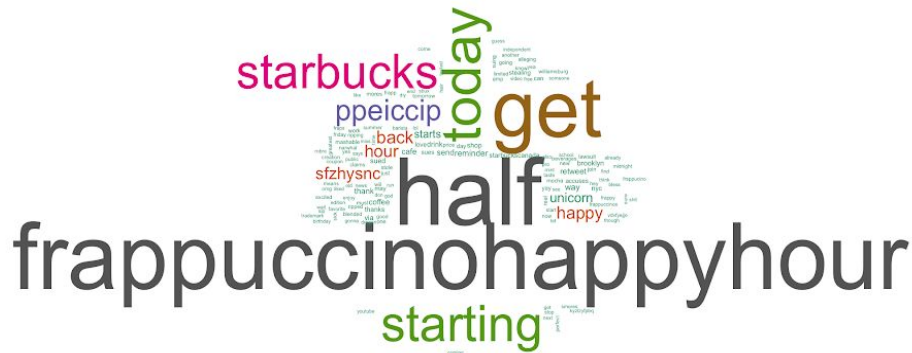
Twitter offers an API for access to tweet data. We utilized a python script that would make calls via the API and store the response from Twitter servers in the form of JSON. The tweet collection was automated such that collection would end at midnight of each day. The data collected in that time period would be stored in an individual JSON file. Tweet collection would resume shortly after the timeout period for the next day.

We initially tested our hypothesis about the popularity of Starbucks on social media by collecting any tweets that contained the word 'Starbucks'. Without filtering the calls to the Twitter API, we obtained tweets on a wide range of topics at a rate that would lead to several gigabytes a day. Our initial idea was to narrow our data collection and analyze customer thoughts about the lines at Starbucks franchises. We assumed that customers standing in lines at Starbucks outlets would have the time to post on social media about their boredom or excitement. Using the location attribute from this tweet data, foot traffic at various Starbucks outlets could also be tracked to provide a picture of the popularity of Starbucks franchises around the country. Below is a word cloud that depicts the most frequent words accompanying 'Starbucks' and 'line':



Although initial results seemed promising with a seemingly wide variety of opinions as can be inferred from the word cloud above, this idea was discarded when we attempted data collection with the query- 'Starbucks + line'. Very few tweets were obtained and it became clear that we would not be able to create a satisfactory dataset.

Following this we decided to contrast customer opinion on various drinks on the Starbucks menu. However after studying a sample of the tweets we had collected, we noticed that there was less explicit mention of older and more staple drinks on the Starbucks menu than the newer releases. Thus we decided to keep our focus on the 'Frappuccino' line of products which was routinely subject to introduction of new varieties. Our initial survey of Frappuccino and Starbucks tweetscape was just before the start of the Starbucks promotion for the beverage line- Frappuccino Happy Hour. As such our word association cloud reflected this:



We selected this search query since the availability of far greater number of tweets and the upcoming promotional event surrounding the beverage line would allow for an interesting analysis of customer response. Along with this, a short while before our tweet collection had begun, Starbucks had introduced a widely publicised Frappuccino variety called ‘Unicorn’. The timing of our tweet collection would allow us to look into the impact of all these factors on social media.

Dataset Details and Software Used:

Our final dataset consists of 71,576 tweets, all of which contain the words ‘Starbucks’ and ‘Frappuccino’. Tweet collection began on May 5th and continued till May 19th (an incomplete file for May 4th was discarded in the analysis phase). There was a break in tweet collection from 20th-26th due to issues with our collection script. Another stretch of data was collected following this between May 27th-30th.

The processed dataset used in our analysis contains a wide range of attributes provided by the Twitter API, including username, time/date, location if provided, language, profile aesthetics, and similar information for the original poster if the tweet is a retweet. In addition to these, the TextBlob Python package was used to determine polarity and subjectivity for each tweet. Polarity is the sentiment of the tweet and subjectivity is a measure of how subjective or objective the content of the tweet is. The ranges for these are:

- Polarity (Sentiment) : -1 to 1 [1 is highly positive and -1 is very negative]
- Subjectivity: 0 to 1 [0 being highly objective and 1, highly subjective]

In the remainder of this report we will refer to polarity in the range of [-1, 0) as negative, 0 as neutral and (0, 1] as positive.

With the exception of sentiment calculation, RStudio and RStudio Server were used for all dataset aggregation and data cleaning tasks. All scripts used in data processing will be made available to the instructors along with this report.

Note: As mentioned above, the tweet collection script was not active for a period of 7 days. Thus the dataset consists of tweets that may be divided by some time interval. However the time lapse was not very detrimental to the overall dataset as collection was stopped only a few days after the end of the Happy Hour promotional event. Sufficient representation of tweet attribute values on non-promotional days is available to perform analysis.

Data Analysis:

In this section we will describe the various trends observed from our collected data. The time-series graphs in this section have a vertical black line demarcating the time-lapse in tweet collection mentioned above. All time series figures use UTC time. **Full size graphs can be found in the appendix.**

An important slice of time that will be referred to in several of the analyses is May 5th-14th (3PM - 6PM) which was the duration of the Happy Hour promotional event during which Frappuccinos were sold half-price at all Starbucks outlets.

Tweet Frequencies and Polarity:

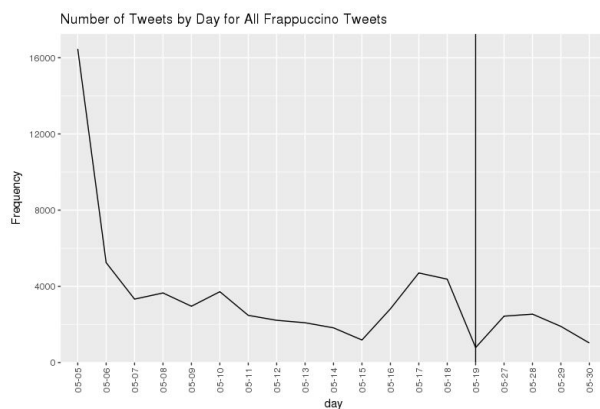


Figure 1

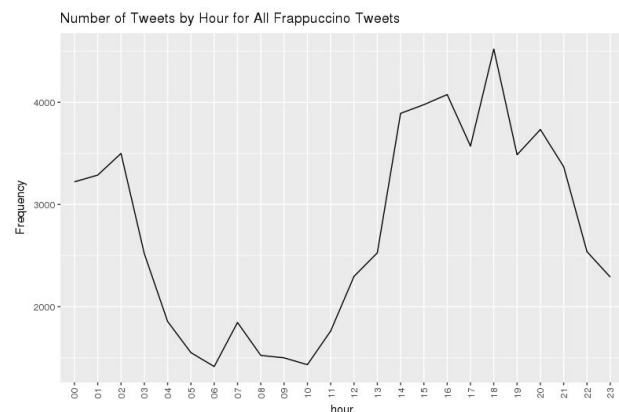


Figure 2

In Figure 1 we can see that there is a large difference in tweet count between May 5th and the 6th. Since the collection script runs for the same amount of time each day, the large difference in tweet numbers must result from generation of greater volume of Frappuccino related tweets on the 5th. We believe this can be explained by several factors. Taking a sample of random tweets from this time slice we can see that many of them are centered around the keywords 'Narwhal Frappuccino'. Other highly referenced topics during this time period are 'Happy Hour' and 'Unicorn'.

Possible reasons for high frequency of tweets on these days:

- Happy hour promotion begins 5th of May.
- New custom Frappuccino dubbed 'Narwhal' went viral on May 4th.
- Limited time offering, Unicorn Frappuccino still being advertised
- Lawsuits against Starbucks for Unicorn Frappuccino

We can see that tweet frequency has a downward trend beyond this time slice. We notice a bump in Frappuccino related tweets on the 16th-17th. Cause: On May 16th the Roasted Marshmallow S'mores Frappuccino was reintroduced into the Starbucks menu. Once again the data corresponds to this news. A large majority of tweets are found to be retweets of the Starbucks official announcement of the drink and customer engagement with this news.

In Figure 2 we can see the tweet frequency averaged over each hour of the day for all 19 days of tweet collection. Since the X-axis is UTC time it is offset from time in USA by -4 to -6 hours depending on time-zone. Accounting for this, we can see that maximum tweet frequency is between 8:00AM/10:00AM and 2:00PM/4:00PM. [UTC 14:00-20:00]. There also a considerable number of tweets between 6:00PM/8:00PM and 8:00PM/10:00PM [UTC 12:00AM-2:00AM]. Most people visit coffee shops in the morning, so the spike in the morning period is expected. The sustained volume of tweets till ~4:00PM could possibly be attributed to Happy Hour. However the spike between 6:00PM and 8:00PM is slightly surprising and may indicate that many customers prefer to visit Starbucks after work.

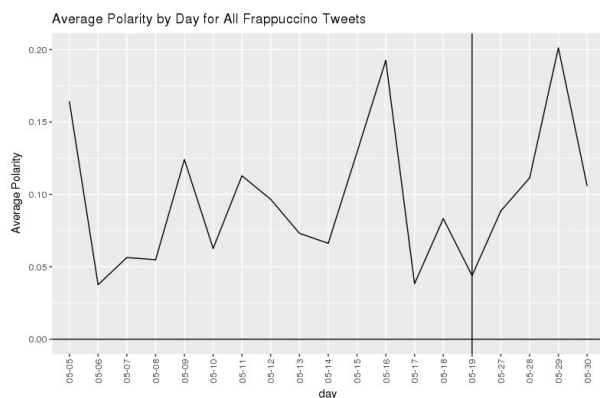


Figure 3

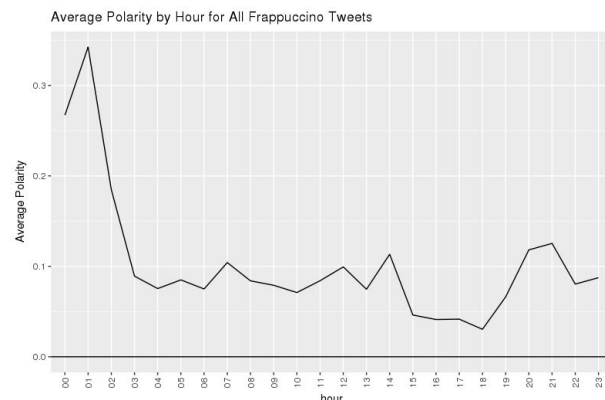


Figure 4

The figures above represent the distribution of average polarity of tweets concerning Frappuccinos over the tweet collection period and the distribution of tweet polarity averaged by hour across all days. We can see that although there are spikes in customer sentiment, average sentiment towards Frappuccinos is largely positive. Once again the patterns that can be observed from Figure 3 can be attributed to the trends mentioned previously:

- Spike before May 5th can mostly be attributed to Narwhal Frappuccino and buzz around Unicorn Frappuccino. Looking at the tweet set, we can also see that a single (positive polarity) ad tweeted by the @frappuccino handle was widely retweeted helping boost the average positive sentiment. Large number of positive tweets about happy hour are also seen during this period.
- Drop in positive sentiment on May 7th: Although there is positive sentiment around the happy hour offer, many of the posts concern the Unicorn lawsuit, which is construed as negative by our sentiment analysis. There are also fewer posts regarding happy hour on this day.
- Large spike on 15/16 is due to large number of positive tweets surround reintroduction of S'mores Frappuccino into menu
- Spike on May 28th is mostly due to widespread retweeting of @ViralFreQ's seemingly positive tweet about the Unicorn Frappuccino. On closer examination we can see that the tweet actually has negative connotation about the Unicorn Frappuccino that would be difficult to detect through text parsing- 'RT @ViralFreQ: DIY Unicorn Frappuccino that ACTUALLY tastes good'

In Figure 3 there is a lot of flip flop in the average polarity. Although it is seen that certain events such as introduction of a new item to the menu or advertising regarding an offer or new beverage temporarily boosts positive polarity, there seems to be a base amount of positive and negative tweets surrounding Frappuccinos at any given time barring any external positive/negative influence. However it is obvious that positive tweets outnumber the negative. Figure 4 shows the hourly distribution of average polarity across all days in consideration. We can see that the polarity is largely uniform aside from the large spike at around 8:00PM. This corresponds with the spike seen in the hourly frequency graph as well. As mentioned before, we hypothesize that customers visit Starbucks in the evening, perhaps after work, or this could be a time slot when many people are free of other obligations and post new tweets. Analysing the contents of tweets posted during this period does not reveal any distinct causes.

Overall Polarity Analysis:

Since we had noted some interesting trends with regards to the polarity, we decided to look at the polarity distribution across all tweets in our dataset.

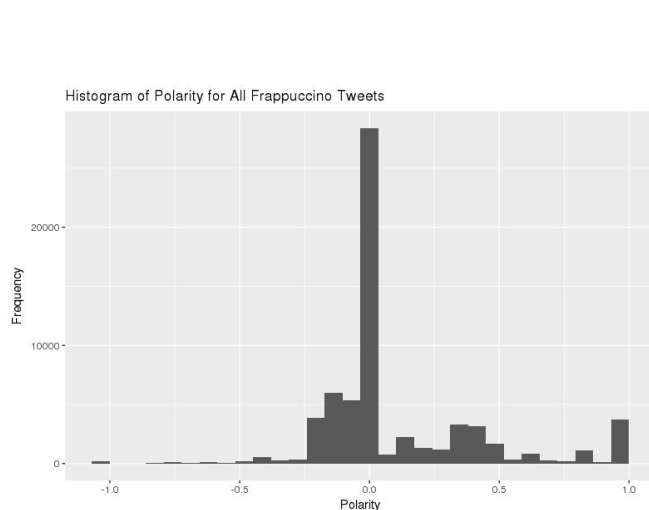


Figure 5

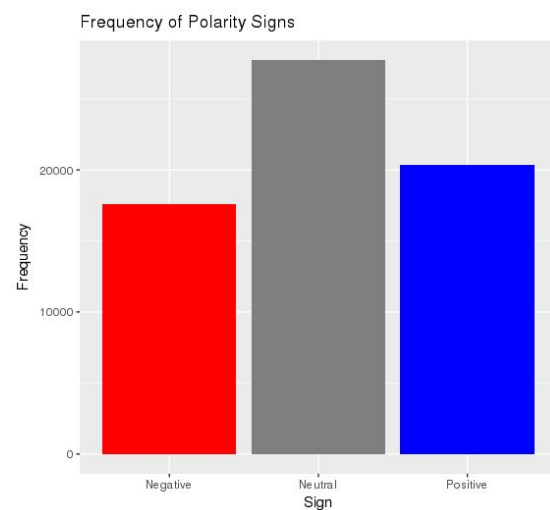


Figure 6

Figure 5 depicts a histogram of polarity for all the tweets collected. We can see that a vast majority of tweets have been classified as neutral. Most of the negative tweets have a polarity ranging between <0.0 and >-0.25 . Looking at the distribution of positive tweet polarity values, we see that there is a much larger variance. Majority of positive tweets have a polarity score between >0.0 and <0.5 . A significant number of tweets are also overwhelmingly positive (score of 1.0). Figure 6 displays a more abstract version of this same information. Leaving aside neutral scores, which should not affect the overall trend, it can be seen that there are >2000 more positive polarity tweets than negative polarity tweets. This is reflective of the polarity trends we previously observed. The overall sentiment towards Frappuccino always remains positive across our observation period. Using this data in conjunction with our previous result lends credence to our inference that there are always a greater base number of positive tweets surrounding this subject than negative ones. This trend is observed in varying degrees based on news surrounding Frappuccino (lawsuits, new beverages, etc.). It is also affected by posts from influential users in this domain.

Analyzing Various Frappuccino Flavors:

Starbucks offers a number of Frappuccino flavors, with new entries being announced at regular intervals. As is to be expected of any retail food product, some of these offerings are seen to be more popular than others. We have chosen 5 popular flavors to examine from our dataset. (Legend: Pink-Unicorn, Brown-S'mores, Gray-Mocha, D.Green-Midnight Mocha, L.Green-Green Tea). Figure 7 and 8 below depict the tweet frequency and polarity by day for each flavor.

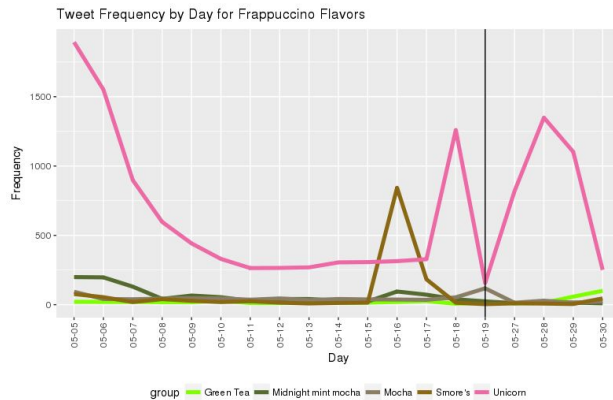


Figure 7

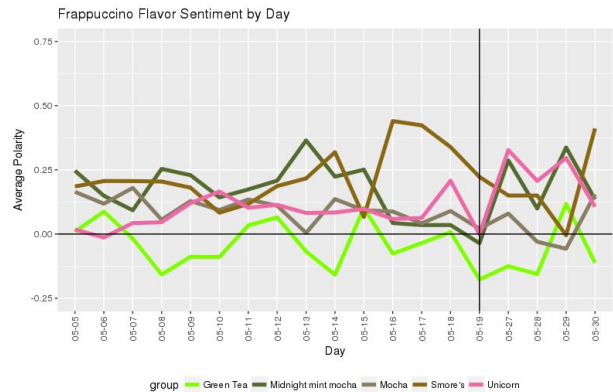


Figure 8

Figure 7 shows that tweet frequency for Unicorn Frappuccino declines steadily after May 5th. Since the lawsuit against the Unicorn Frappuccino was widely reported on this day, it is understandable that interest in this topic declined slowly in the following days. The interest surrounding the drink on May 29th was due to the aforementioned @ViralFreQ who is among the top retweeted users in our dataset. Their tweet was retweeted over 2000 times between May 27th, 28th and 29th. As we mentioned before, S'mores Frappuccino was re-released on the 16th and we can see the corresponding increase in tweets on that day. The other flavors maintain a somewhat stagnant number of mentions across all days. In Figure 8 we see the sentiment for each flavor on each day of the observation period. As we may notice from the Figure 7, some of these flavors are mentioned in tweets far more than others. Green Tea Frappuccino for example can be seen to have at most ~80 tweets mentioning it in total. Hence, sentiment can easily be swayed by frequency in this scenario. A few negative green tea tweets a day could easily cause the negative average sentiment we see on several of the days in Figure 8. Leaving aside average sentiment observed due to normal customer preference (as with the Midnight Mocha, Mocha whose popularity is not affected by external factors such as ads during the observation period), for a flavor such as Unicorn we can see that the sentiment is low on May 5th due to lawsuit related tweets. Following this it maintains a somewhat even trend till the 27th where, as explained before, it is mentioned in retweets of @ViralFreQ- 'DIY Unicorn Frappuccino that ACTUALLY tastes good' alongside other actual positive tweets. S'mores Frappuccino sentiment corresponds directly from the frequency trend seen in Figure 7. (Overall flavor sentiment graph available in Appendix)

Frequency and Popularity Across States:

Figure 9 shows a map of frequency of tweets across states in the USA, normalized by current population of those states. Figure 10 shows average polarity of tweets across states in the USA. The location attribute of tweets was used to draw up this distribution. States shaded gray had no tweets regarding Frappuccinos. These figures reveal that Frappuccinos are a topic of interest all

across the country with greater frequency in the western and southern regions. This could be explained by the warmer climate in these areas and the appeal of Frappuccinos as a cold beverage. The information from Figure 10 shows that Frappuccinos have positive sentiment among customers in most states. However we do see some negative polarity of tweets in the central states (Nebraska and Kansas). The number of tweets with actual location information from each state was quite low. California and Nevada had hundreds of tweets while many others had only a handful. We feel an analysis of this kind would be more accurate if we had a much larger dataset of geotagged tweets.

Normalized Frappuccino Tweet Intensity by State

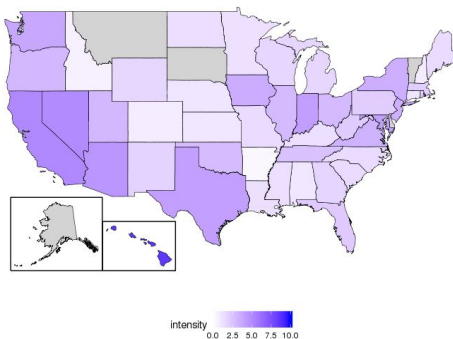


Figure 9

Average Frappuccino Polarity by State

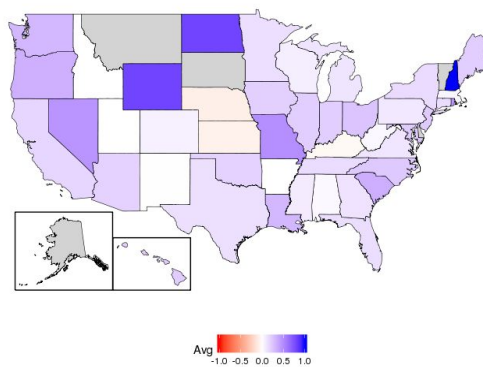


Figure 10

Hashtags, Retweets and Users:

	X	n
1	#FrappuccinoHappyHour	11768
2	#FrappuccinoHappyHour...	2864
3	#スタバ	1645
4	#フラベ	1217
5	#frappuccino	1095
6	#ICYMI	1011
7	#starbucks	894
8	#monday	852
9	#MondayBlues	851
10	#mondaymotivation	850
11	#weekend	740
12	#saturday	739
13	#SocialSaturday	739
14	#SundayMorning	738
15	#SundayFunday	737

Table 1

	User	No.
1	NA	25020
2	frappuccino	10999
3	imtaiki	4038
4	Starbucks	3743
5	ViralFreQ	3020
6	StarbucksMY	1938
7	entabejp	1555
8	Home_Halfway	1287
9	killmefam	731
10	margariet_dh	675
11	itspalmm	640
12	wongnai	524
13	ItsFoodPorn	488
14	StarbucksCanada	401
15	okidkaccutally	358

Table 2

Table 1 shows a list of the most popular hashtags in our tweet dataset. Hashtags related to Happy Hour have highest number of occurrences (which is to be expected due to large number

of retweets of Starbucks Happy Hour ads. These are followed by hashtags containing Japanese characters which speaks to the popularity of Frappuccino beverages and Starbucks in Japan. The list also has hashtags relating to the beginning of the week and end of the week. We can see Frappuccinos are associated with both fun and handling dreariness.


In Table 2 above we can see a list of the most retweeted users. The first row indicates original tweets that were not retweeted. The highest number of retweets were of tweets from the @frappuccino handle which is the official handle for the Frappuccino drink. Followed by @imtaiki who is a popular user from Thailand that advertises various products. The Starbucks official handle is on this list as expected. @ViralFreQ is an account that posts links to interesting articles on their website. They were shown to be responsible for raising Unicorn Frappuccino mentions as mentioned before. Some of the other accounts on this list are journalists, restaurant recommenders (entabejp recommends eateries in Japan, wongnai does the same in Thailand).

	User Screen Name	No.
1	cookingscarlet	208
2	srnorganics1	201
3	Starbucks	155
4	Cafezanmai	138
5	GiftCard_Coupon	134
6	yuracuks	124
7	frappuccino	117
8	Bid_Work	56
9	love_starbucksJ	51
10	redpotterfrah	42
11	eggxte	37
12	MillarMatthews	30
13	bluerangerrose	29
14	5sosraps	22
15	BaristaJill	21

Table 3

Table 3 shows the users who posted the highest number of Frappuccino-related tweets during our observation period. Most of these users do not have many followers, but follow many others and post high volume of tweets. @cookingscarlet who is first on the list posts large number of food related tweets. Interestingly the account has been removed from Twitter at this time. From the rate of posts made we suspect that it could have been a spam account. Many of the other accounts on this list sell various products and also seem to advertise products such as the beverages from Starbucks. A few of these accounts such as Cafezanmai were created very recently and may be trying to gain more followers by posting a large volume of tweets.

We were unable to fit a table of most retweeted tweets in this document due to space constraints, however looking at the top 3 tweets in this list we noticed some interesting results. The most retweeted tweet was: "RT @imtaiki: Starbucks ญี่ปุ่น เริ่มขายเมนูใหม่ Strawberries & cream Frappuccino เพิ่มซอสช็อกโกแลต+เปลี่ยนเป็น Non Fat ฟรี *ใส่... ", a tweet in Thai from Twitter handle @imtaiki advertising a Strawberries & Cream Frappuccino. From the dominance of this tweet in our dataset we can infer that Frappuccinos have a strong presence in Thailand. (The results described below also point to this). Second most retweeted was "RT @Starbucks:



 Get half off any Frappuccino from 3-6 pm!

 #FrappuccinoHappyHour <https://t.co/5PpEi2cciP>" which is very clearly an ad from Starbucks.

 Seeing as this post was retweeted 2043 times, it would seem that the happy hour offering was well received by Starbucks customers. The full list of retweets sorted by retweet frequency is available at the link provided in the abstract.

Tweets Across Countries:

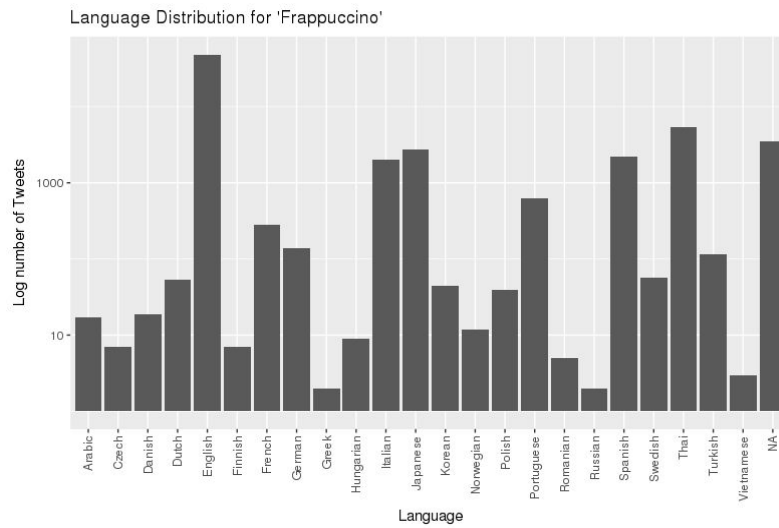


Figure 11

place.country	n
NA	64160
United States	994
Canada	81
Malaysia	67
Republic of the Philippines	53
Thailand	32
日本	32
Brasil	30
United Kingdom	20
Indonesia	16
México	15
Japan	14
ประเทศไทย	12
Peru	10
Argentina	8

Table 4

Figure 11 shows the distribution of tweet languages present in our dataset. We can see that English is the clear majority. This is expected as Starbucks is an American company whose consumer base is mostly English-speaking. This is also reflected in Table 4 which shows the number of tweets originating from different countries (Tweets that did not provide location information fall under NA; a further row for blank entries was removed). We can see 3 English speaking nations in the top 15 countries: 994 tweets from the USA (1), 84 from Canada (2) and 20 from the UK (11). Interestingly the second most prevalent language is Thai, and we can also see that Thailand is 8th in Table 4. This and the information in the previous section about most retweeted users shows that Starbucks has a significant presence in Thailand. We can see that there are a similar number of tweets in Japanese (this bolsters our inference about the popularity of the Frappuccino in Japan from hashtags), Spanish and Italian as well. This does not, however, correspond to the information in Table 4, which shows low tweet counts for these countries. This discrepancy may be attributed to the lack of geo-tagged data from these locations.

Conclusion and Future Work:

Our analysis of Frappuccino related tweets has revealed several interesting patterns regarding customer behavior and preference in relation to the Frappuccino beverage and the Starbucks brand. An interesting topic to expand on this analysis would be to collect a dataset containing several different Starbucks menu items, perform a similar tweet analysis and compare and contrast the results.

Appendix:

All analysis code and data can be downloaded from the link provided in the abstract.

Figure 1:

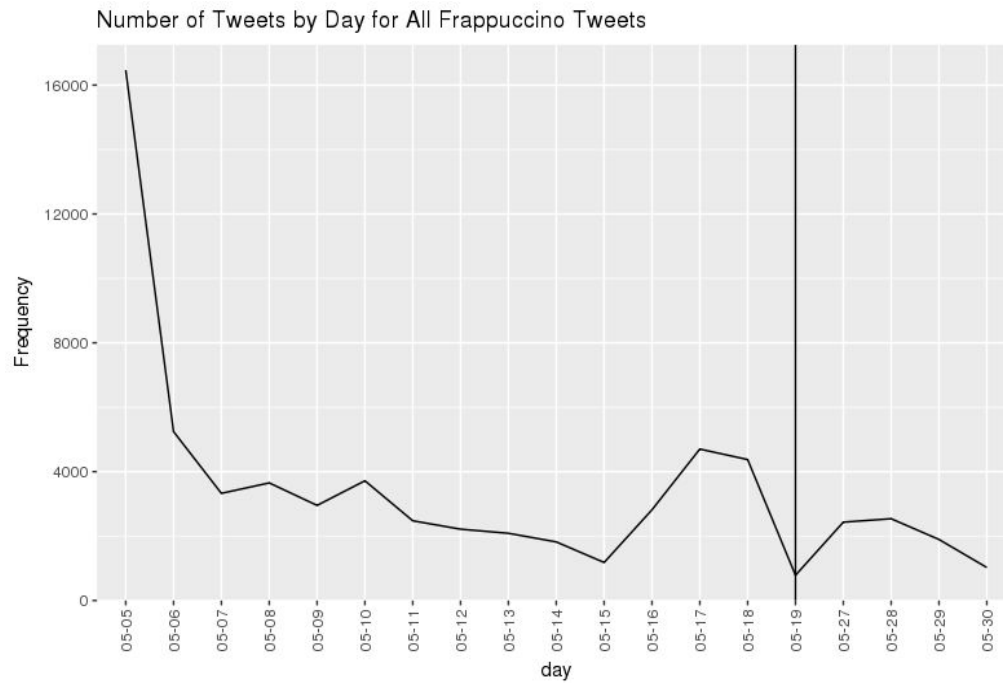


Figure 2:

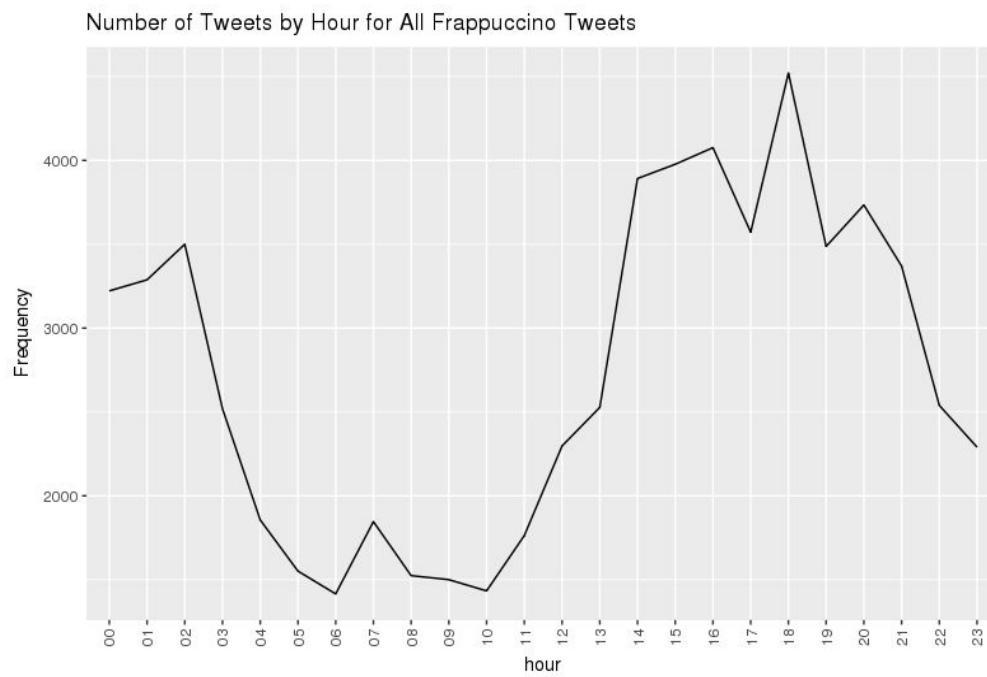


Figure 3:

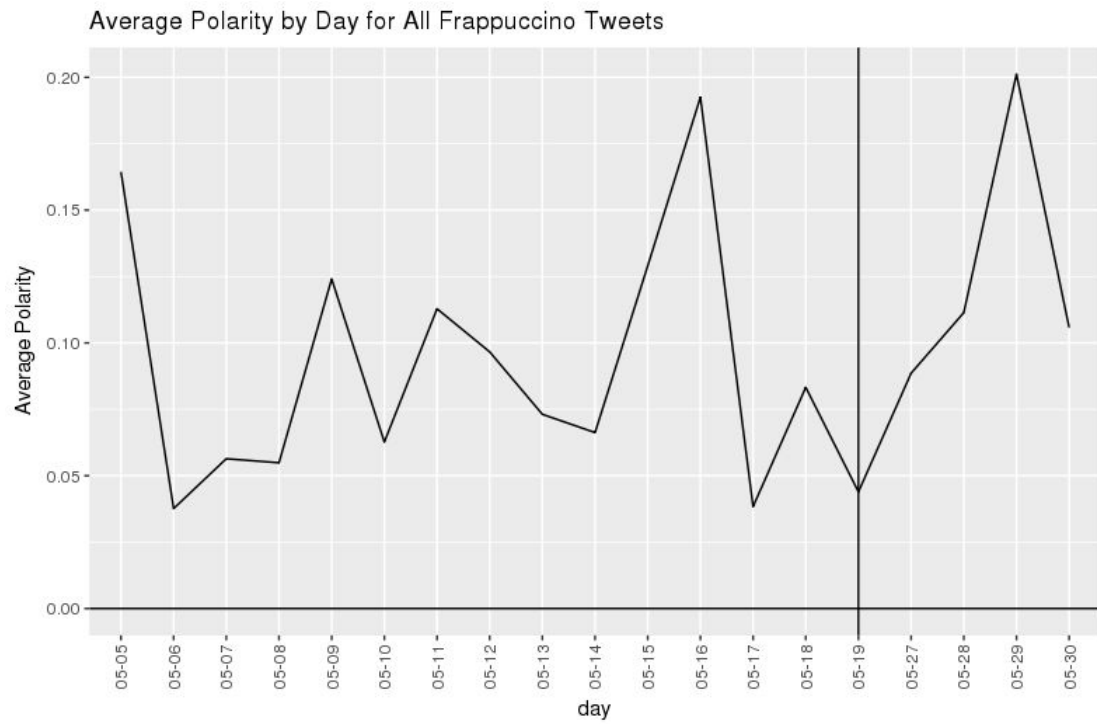


Figure 4:

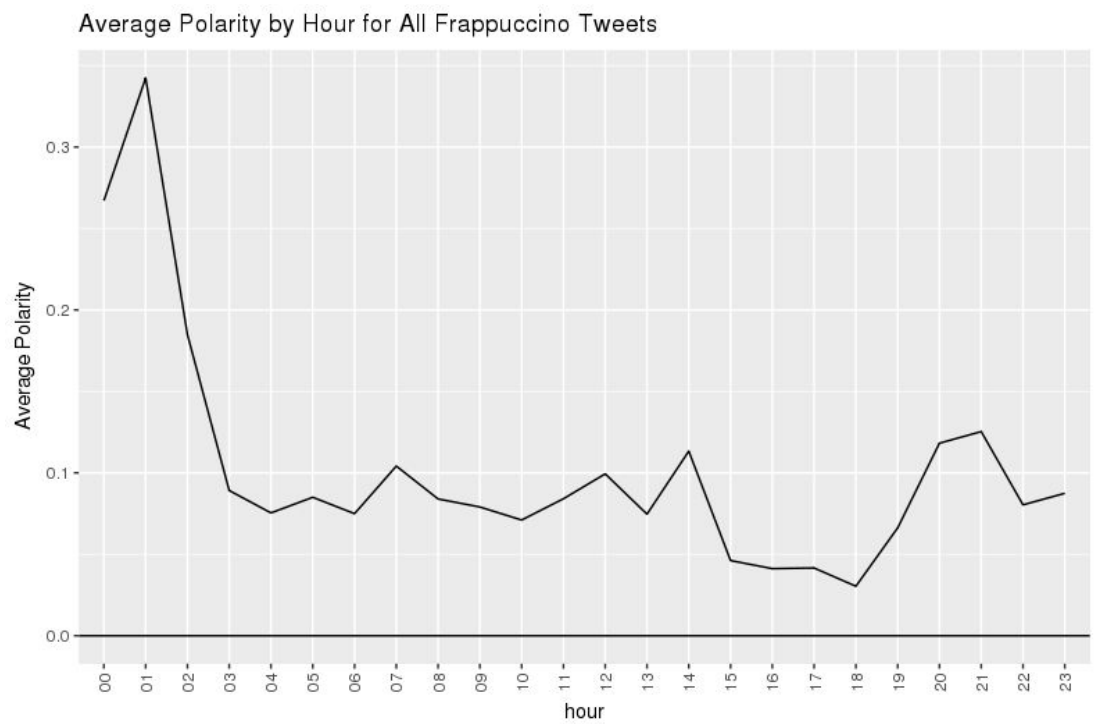


Figure 5:

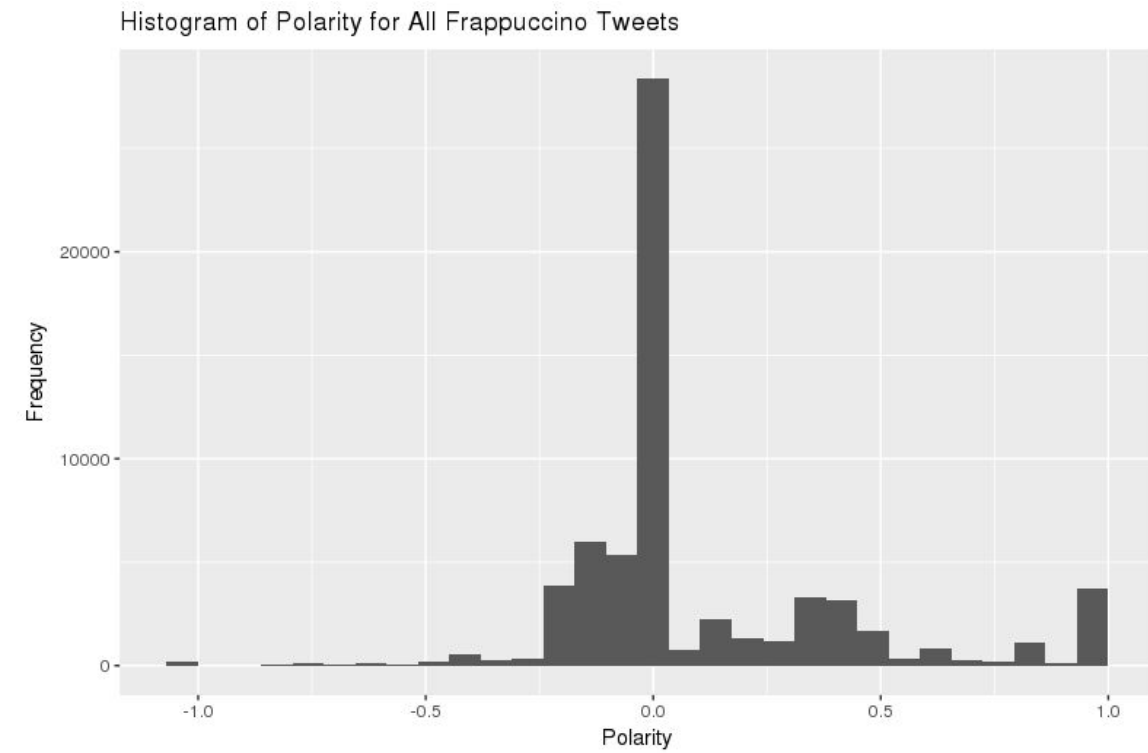


Figure 6:

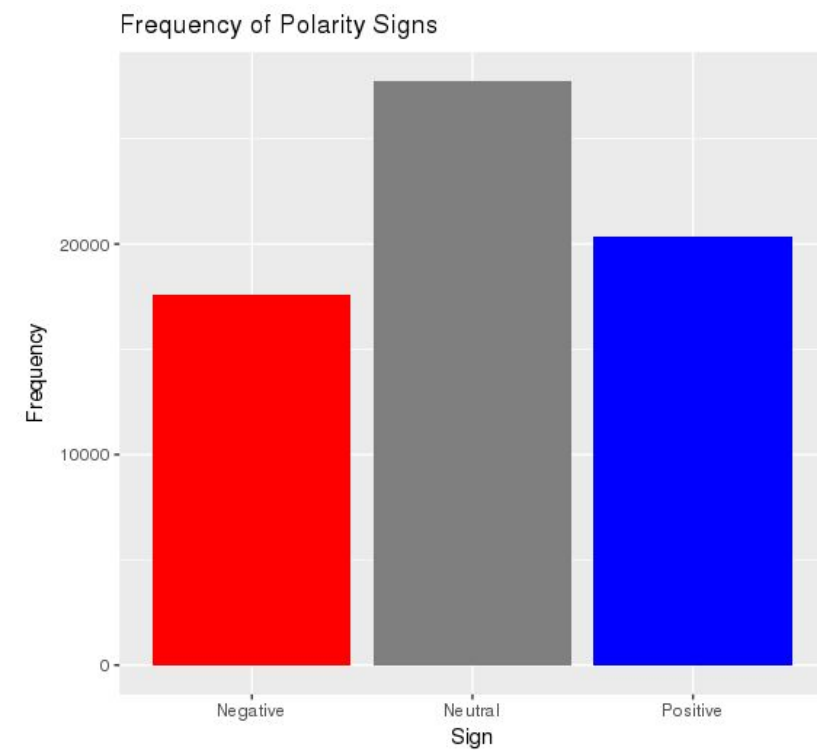


Figure 7:

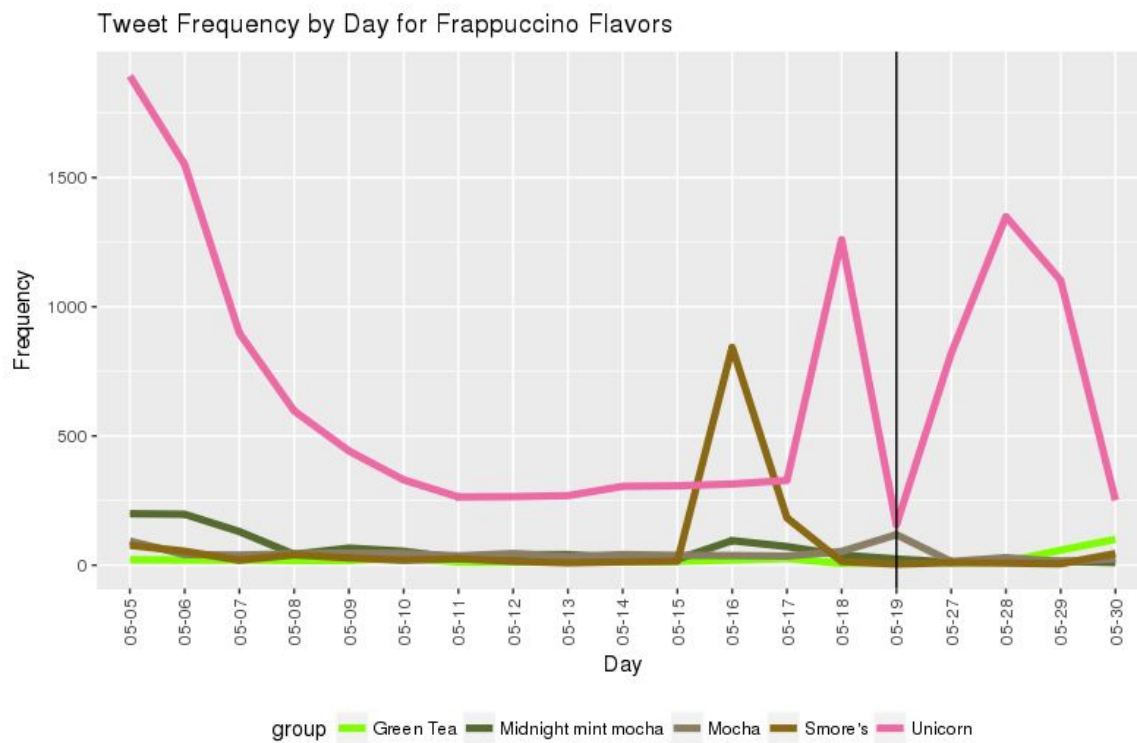
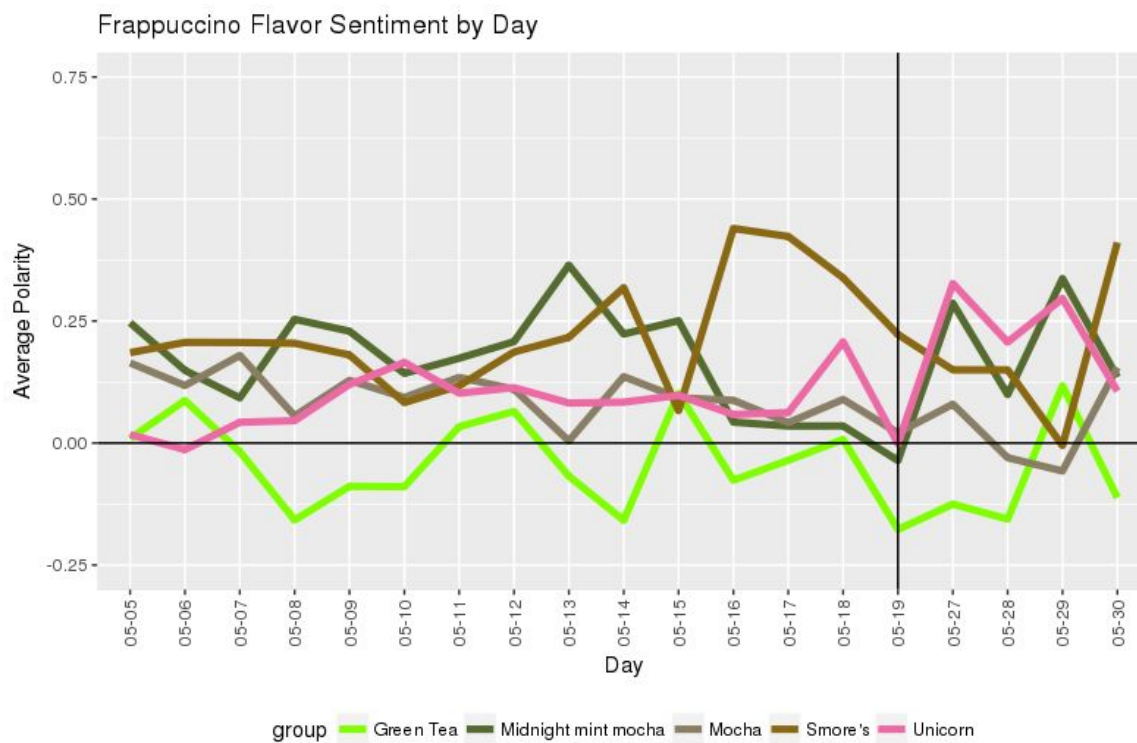


Figure 8:



Overall Flavor Sentiment:

Comparison of Frappuccino Flavor Sentiment

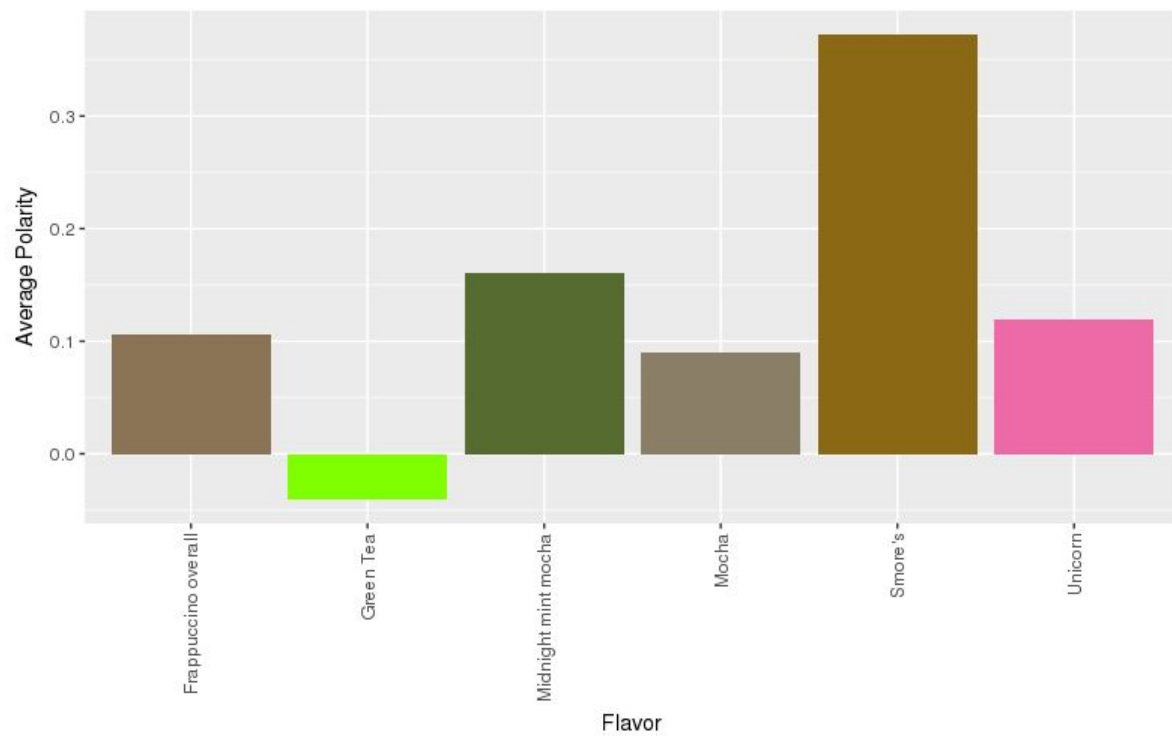


Figure 9:

Normalized Frappuccino Tweet Intensity by State

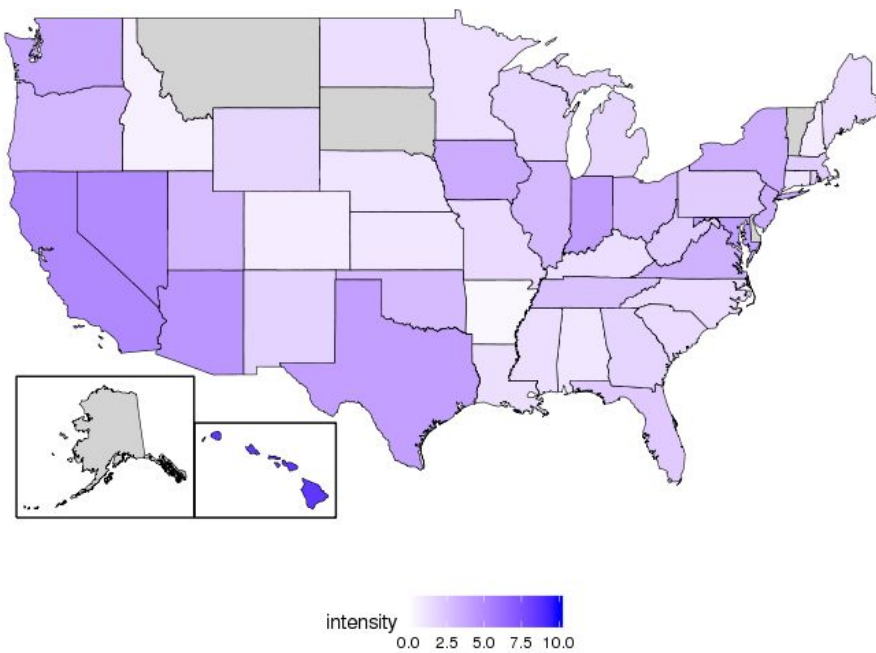


Figure 10:

Average Frappuccino Polarity by State

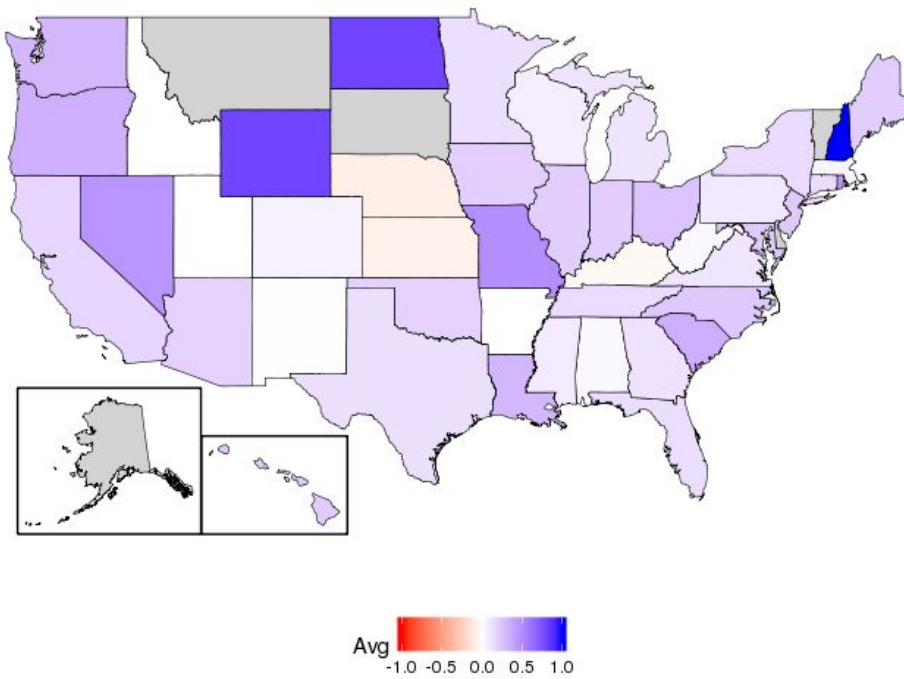


Figure 11:

Language Distribution for 'Frappuccino'

