Maggie Kahn

Business Analytical Strategy

Dr. Sanchez

30 September 2021

Sentiment Analysis in Twitter

Introduction

As a society, we are constantly creating rivalries and comparing each other in a multitude of ways. Whether its who's football team is better that year or the ongoing conversation regarding vaccines, we are always looking for ways to differentiate them and crown a winner. Comparisons like these don't stop with people or things, they can also include companies as well. For this project, identified a company that I had an interest in and that also had a well known rival. I decided to go with Nike and Adidas, two very popular brands for clothing, shoes, and sporting accessories. Hopefully, I can analyze people's opinions on the two companies through Twitter and see what the general consensus is regarding the companies and their popularity among Twitter users.

Data and Methodology

My first step in analyzing Nike and Adidas was pulling Tweets from Twitter into Jupyter. I imported the necessary packages and created csv files for each company that would later be populated by each users information. Twitter is a massive microblogging and networking service where its users can tweet about a wide range of topics, so I decided it was best to narrow down my search by inputting hashtags into my code. I used #Nike and #Adidas as keywords to help me in my search. Prior to this command, I also inputted code that would help Jupyter read the

information I was typing in. This included several “imports” and downloads that needed to be installed on my computer in order to read the tweets correctly. Most of these commands that helped clean the data and remove unnecessary information were given to the class and are now displayed towards the top of my Jupyter file. One of the more important codes was a sentiment function. I looked at polarity and subjectivity. Polarity is usually displayed with a -1 or a +1. It is better to have a higher polarity or sentiment at +1 than a low one near -1. Subjectivity generally refers to emotion or judgment. Having a higher subjectivity mean that there is an overall higher rating in terms of how people perceive the company. In order to further analyze the Tweets, I was pulling, I inputted code into Jupyter to show me some statistics and summaries for both companies. For Nike, I used the code sumstats=NikeData.describe() and sumstats.round() to print descriptive statistics (Table 1). It lists the mean, standard deviation, minimum and maximum, along with a few other things. I was most interested in the mean because it shows the average for each category and is a helpful way to understand the data. I inputted the same code but for Adidas (Table 2) and noticed that Nike had a higher average for reach and impression. I thought it was very important to store these results in my data table for later use, so I added them and used the code AdidasData.head() and NikeData.head() to display them in my data set (Table 3). This function allows me to look at the first couple of rows so I can see if the code ran correctly.

After performing these steps on each company, I wanted to combine both data sets into one excel file so I could later upload it to Tableau. On one of the Adidas data frame, I inserted a code to add a column that would insert a one is the user information had to do with Nike and a zero if it was for Adidas (*Adidas["Nike"]=0 and Nike["Nike"]=1*) . Before exporting, I combined the data with the code *frames = [Adidas, Nike]NikevsAdidas = pd.concat(frames)* and

double checked that it combined correctly with the code *NikevsAdidas.head()* (Table 4). Finally, I wanted to make a sample plot to hopefully gain some insight into what my plot may look like once I start working in Tableau. I used the code *sns.relplot(data=NikevsAdidas, x="friends",y="followers", col="Nike", kind="scatter")* to display a simple scatter plot for both Nike and Adidas. It shows the followers and friends each user had for both companies (Graph 1). Finally, I exported my file as an excel file so I could download it into Tableau. I used the code *NikevsAdidas.to_excel('NikevsAdidas_tweetdata.xls')* to perform this task.

	followers	friends	impressions	reach	engagement	engagement_rate
count	2990.0	49.0	2.530000e+02	4.900000e+01	49.0	49.0
mean	1205.0	872.0	3.455121e+07	4.451692e+07	0.0	36.0
std	2999.0	2411.0	2.131722e+08	2.432961e+08	1.0	97.0
min	0.0	0.0	0.000000e+00	7.000000e+00	0.0	0.0
25%	47.0	22.0	1.000000e+01	1.722000e+03	0.0	0.0
50%	280.0	51.0	1.722000e+03	1.259200e+04	0.0	1.0
75%	1089.0	363.0	1.644400e+05	2.034280e+05	0.0	5.0
max	78586.0	12173.0	1.691012e+09	1.691012e+09	3.0	500.0

Table 1: Descriptive Statistics for Nike

	followers	friends	favorites	impressions	reach	engagement	engagement_rate
count	6734.0	36.0	0.0	36.0	36.0	36.0	36.0
mean	863.0	471.0	NaN	47.0	1823330.0	0.0	0.0
std	1692.0	1123.0	NaN	72.0	5949536.0	0.0	1.0
min	0.0	1.0	NaN	0.0	11.0	0.0	0.0
25%	88.0	76.0	NaN	3.0	132081.0	0.0	0.0
50%	301.0	76.0	NaN	10.0	853776.0	0.0	0.0
75%	862.0	366.0	NaN	70.0	853776.0	0.0	0.0
max	30021.0	4987.0	NaN	291.0	34534987.0	0.0	3.0

Table 2: Descriptive Statistics for Adidas

```
In [304]: AdidasData.head()
Out[304]:
```

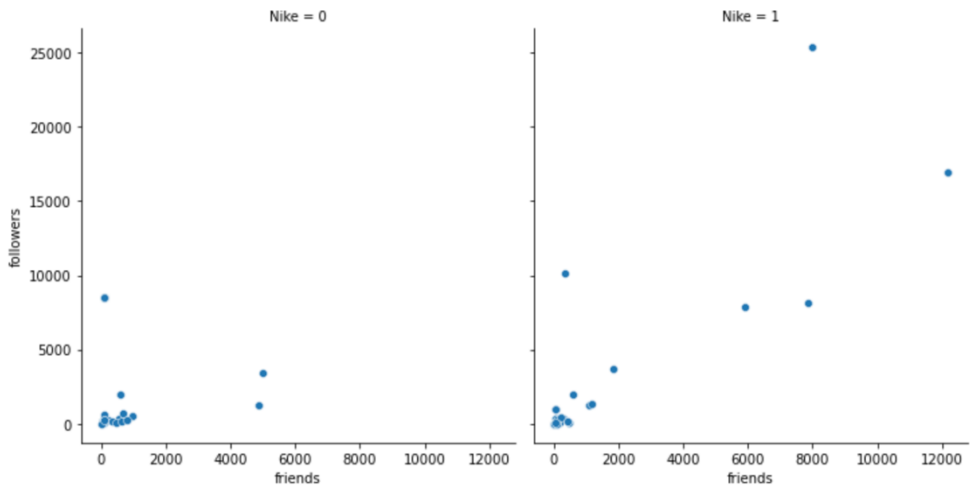
rity	subjectivity	lang	...	followers	friends	hashtags	place	coordinates	favorites	impressions	reach	engagement	engagement_rate
0.0	0.0	en	...	8537.0	76.0	adidasspzl, cpcompany, adidasshoes, trefoil	NaN	England	NaN	70.0	853776.0	0.000082	0.008199
788	0.5	en	...	8537.0	76.0	adidasmccarten, adidasshoes, trefoil	NaN	England	NaN	40.0	853776.0	0.000047	0.004685
375	0.75	en	...	225.0	233.0	adidas, cpcompany	3_stripes4life, LukeShaw23	NaN	NaN	1.0	225233.0	0.000004	0.000444
375	0.75	en	...	8537.0	76.0	adidas, cpcompany	LukeShaw23	England	NaN	121.0	853776.0	0.000142	0.014172
182	0.4772727272727273	en	...	377.0	543.0	NaN	NaN	Delaware, USA	NaN	0.0	377543.0	0.000000	0.000000

Table 3: Displaying the first couple of rows of the table

```
In [354]: NikevsAdidas.head()
Out[354]:
```

polarity	subjectivity	...	friends	hashtags	place	coordinates	favorites	impressions	reach	engagement	engagement_rate	Nike
0.0	0.0	...	76.0	adidasspzl, cpcompany, adidasshoes, trefoil	NaN	England	NaN	70.0	853776.0	0.000082	0.008199	0
878787878788	0.5	...	76.0	adidasmccarten, adidasshoes, trefoil	NaN	England	NaN	40.0	853776.0	0.000047	0.004685	0
0.375	0.75	...	233.0	adidas, cpcompany	3_stripes4life, LukeShaw23	NaN	NaN	1.0	225233.0	0.000004	0.000444	0
0.375	0.75	...	76.0	adidas, cpcompany	LukeShaw23	England	NaN	121.0	853776.0	0.000142	0.014172	0
818181818182	0.4772727272727273	...	543.0	NaN	NaN	Delaware, USA	NaN	0.0	377543.0	0.000000	0.000000	0

Table 4: Displaying the first couple of rows of the table



Graph 1: Displaying Adidas on the right and Nike on the Left Friends vs. Followers.

Results

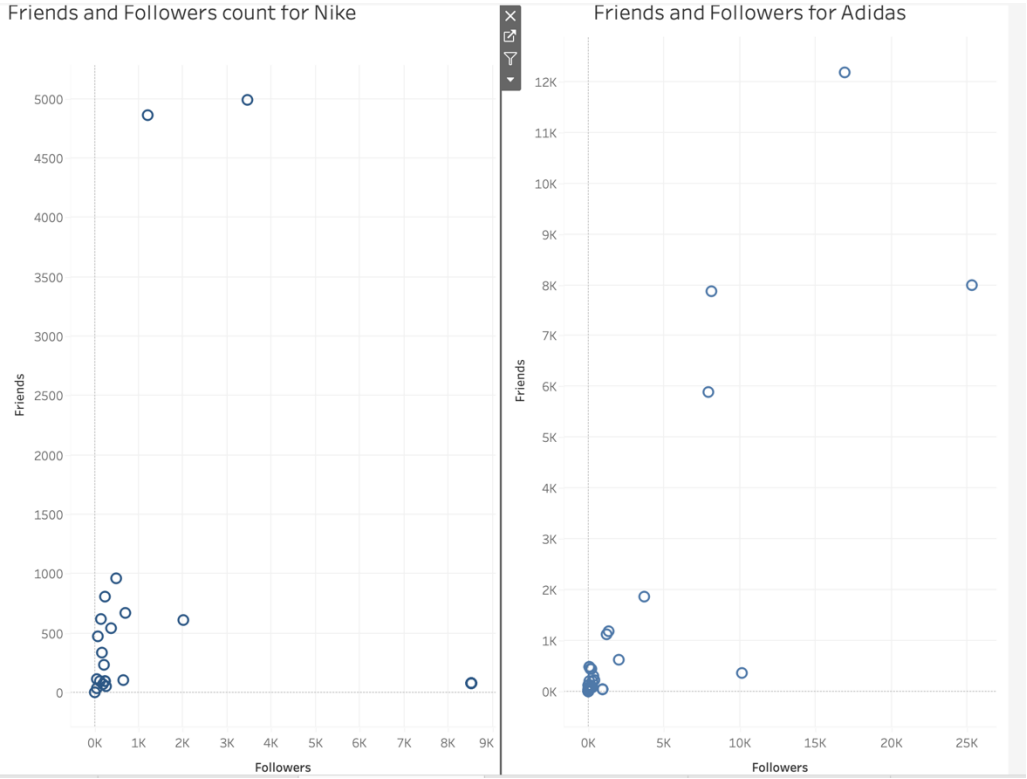
After performing the statistics and summaries, I uploaded my excel file to Tableau and created visualizations to help with coming to a conclusion regarding the two companies and which one received the most attention. My first visualization looks at the followers that engaged with my hashtags (#Nike or #Adidas) and looked at how many friends they had. I created this graph in hopes to see which company possessed followers that had a high number of friends. Having accounts that have a high number of followers could possibly correlate with engagement. and impressions. From the results in graph two, that those who interacted with the Adidas hashtag tended to have a higher number of followers and therefore more friends.

My second visualization looked at all the users and their overall reach for Nike and Adidas. I thought it was best to separate these by company so it would be easier to compare the two. By looking at graph two, more users under Adidas had more reach than those under Nike. Reach can be so important when it comes to selling products. If people aren't aware of the

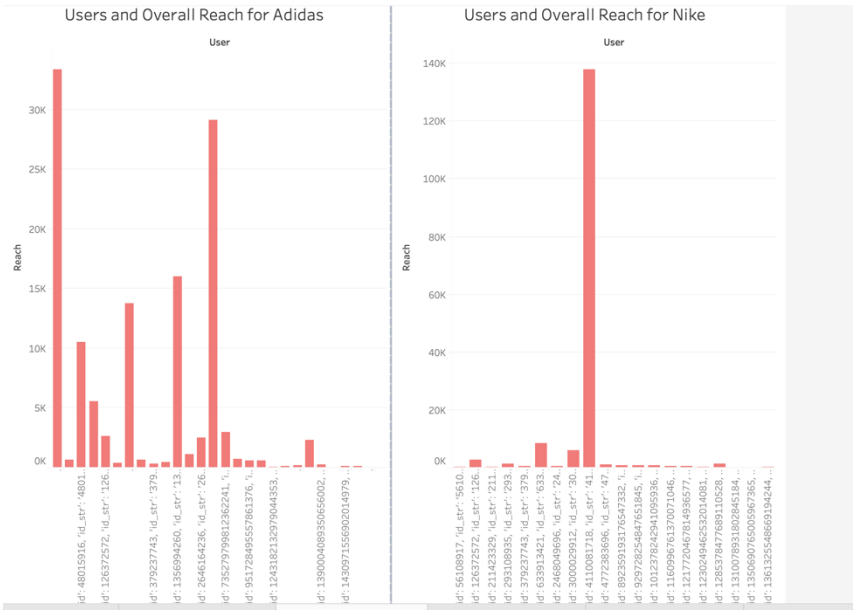
company, then their products won't be sold. The more people who know, the bigger the audience would be and hopefully that would lead to more sales.

My third visualization looked at the users and how many followers they have. It also analyzed their engagement with the hashtags for each company. By looking at graph four, you can see that users with a smaller following tended to interact with the hashtags more than users with more followers. Overall, Adidas had a higher engagement rating than Nike.

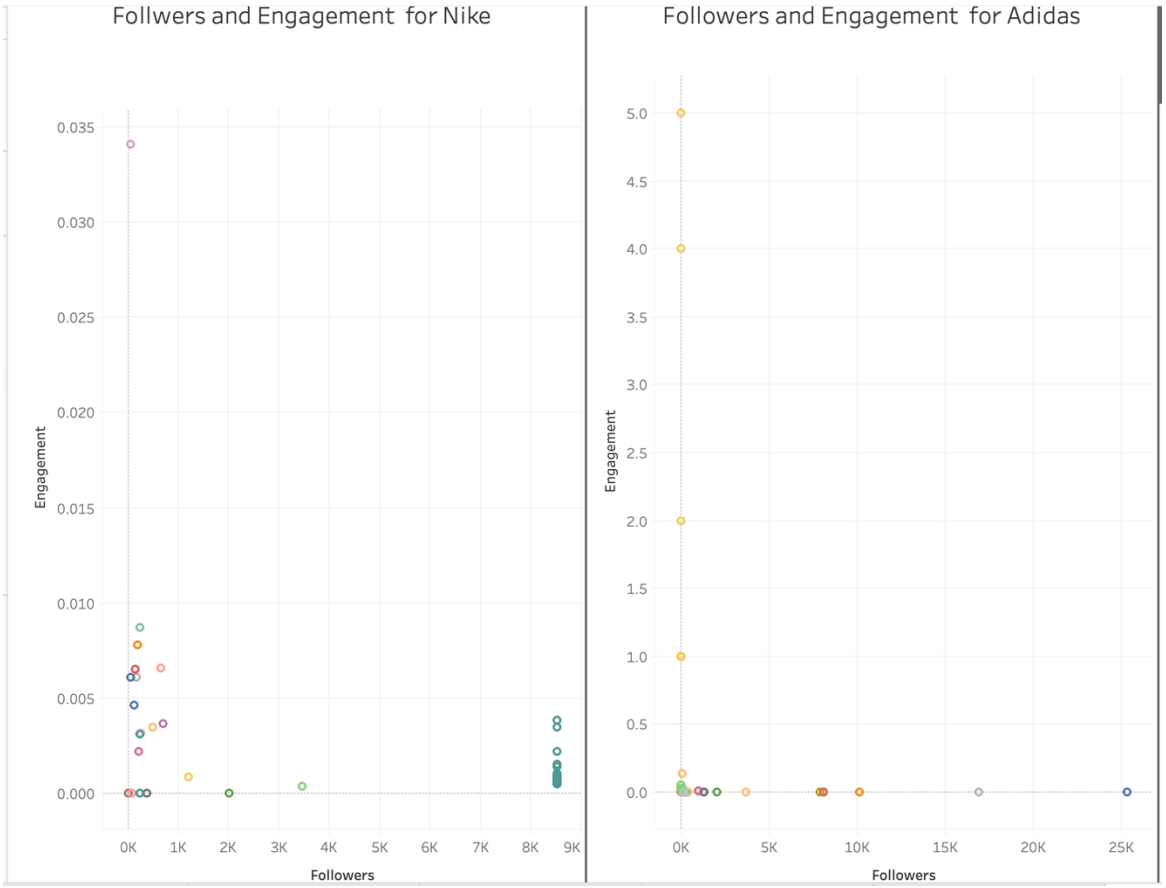
The fourth and final visualization (graph five) analyzed the impression and reach for both companies. Surprisingly, Nike had an overall higher impression but Adidas tended to do better with reach, which was clearly shown in graph three.



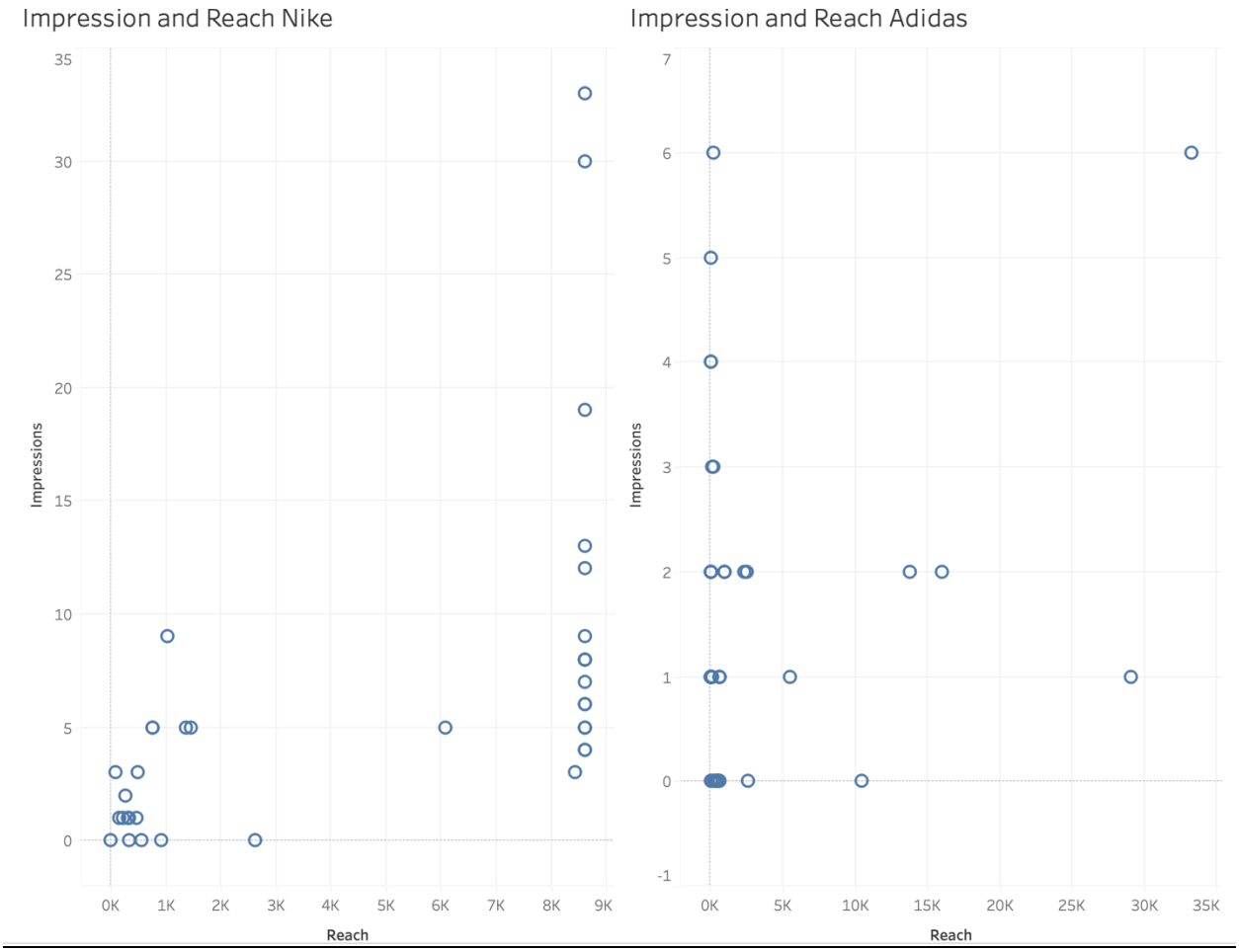
Graph 2: Friends and Followers count for Adidas and Nike



Graph 3: Users and Overall Reach for Nike and Adidas



Graph 4: Followers and Engagement for Nike and Adidas



Graph 5: Impression and Reach for Nike and Adidas

Connections and Conclusions

After analyzing the data in Jupyter and creating visualizations in Tableau, I was very surprised by the results. I previously thought that Nike would have had higher results for all four comparisons because of their popularity among my generation, in sports, and in the United States. Recently, Adidas has become increasingly more popular among foreign countries because their athletic gear for sports like soccer and very well perceived. I also forgot to take into account the sponsorships that certain teams or other companies might have with Nike or Adidas that would increase their reach and engagement. Through this entire process, I was able to realize that

companies can express their success and popularity in more ways than one. While most of my project went smoothly, there were some limitation along the way. Columns in the data frame were left unpopulated or the information was hard to display. For example, the user column had each user name as a huge paragraph that included much more than their names. This made it hard to display them In Tableau on some of the graphs. Another problem was with inputting the code into excel. Sometimes the code would come up with a warning or wouldn't work at all. Luckily, with some assistance, I was able to fix the code and get it to run normally. In the future, I hope to further analyze the Twitter data I pulled to look into what specifically makes the company so popular among their audiences. I also want to continue looking into overall engagement and how impression could change over time.