|  |
| --- |
| WPI LOGO.png |
| [**ofo Bike-Sharing Service Analysis**] |
| DS501 Case Study 1 : Collecting Data from Twitter  9/20/2017 |
|  |
|  |
|  |

**Team 6 Haowen Zhu, Dekun Geng, Yin Hang, Yixin Luo, Weiqing Li**

**(1) What data you collected?**

We choose the topic of bike share and use the Twitter Streaming and Searching API to sample a collection of tweets about this topic in the last few days. 200 tweets have been collected and analyzed. We will produce an overall snapshot on the useful data of users’ age, gender, profession, location, and attitudes to bike-sharing through their tweets. Furthermore, we choose a popular twitter user ’ofo us’ with 15.8k followers and collect the list of all friends and all followers of ‘ofo us’.

**(2) Why this topic is interesting or important to you? (Motivations)**

There are three main reasons for use to choose this topic. First, ofo bike-sharing system has already been established in Seattle first and recently introduced 200 bicycles to Worcester. By the end of 2017, there will be more than 400 bicycles put in Worcester. Currently many cute yellow bikes can be caught parking in the Common Park in Worcester. As Worcester are the city where we live, it makes more sense to know that how the easy-to-use bike-sharing service changes the way we were and what innovations it brings to our city.

Additionally, from the business prospective, doing a research is meaningfully to figure out which city is the next targeted market and what kind of business strategies should be taken in marketing based on the data support we collected from twitter. (This part will be explained more specifically by our analysis later).

Last but not least, from the previously successful business case of ofo bike-sharing system that sharing bikes has become phenomenal and could be easily seen in every corner in China. We fully believe that it is unstoppable trend for this convenient and environmentally friendly transportation to become popular not only in China, but also in US even around the world. Therefore, it is meaningful and innovative to do some data analysis for this topic.

**(3) How did you analyze the data?**

* Creating a Twitter API connection

Before making any requests to Twitter, we created an application to get the authorization to access user's account data. This step helps to tackle the problem of permission of acquiring and analyzing Twitter data.

***Table 1 API Connection***



* Searching for Tweets

This is a basis of a search query to target some tweets with the topic word “bikeshare” after hashtag. We set the data volume 200 and print the results as json file into a local file.

* Streaming API

In our case, we want to maintain a constantly updated view of tweets, thus streaming API is also included to complete the research. We use streaming API to acquire the data for the same topic and save the data results to a local file.

* Frequency Analysis

Computing the frequencies of the words that have been used in these tweets, then a table of the top 30 words with their counts will be plotted.

Also, we plot a table of the top ten tweets that are the most popular among our collected tweets to find the most popular tweets. Most popular Tweet Entities in our collection of tweets will be shown by plotting a table of the top 10 hashtags.

* Getting All Followers and Friends for a User

To get the information about all friends and followers of ‘ofo us’ in twitter, we plot a table which contains the ID numbers and screen names of twenty followers and friends of ‘ofo us’.

* Business plan

We grab the location data of the collected Json profiles and plot the distribution map about the ‘bike share’ topic. ‘Textblod’ library is used to do the sentiment analysis for ‘bike share’ topic in different cities for different bike share brands.Furthermore, we do the weighted analysis for the cities with different bike share brands in order to obtain an intuitive figure of different city’s satisfaction for bike-sharing system.

**(4) What did you find in the data?**

***Table 2 the top 10 tweets that are the most popular among our collection***

**A close up of a newspaper

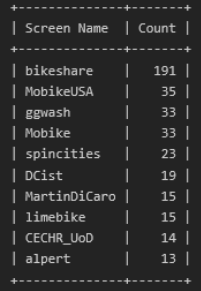
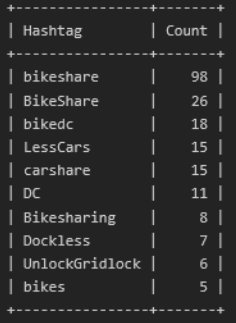
Description generated with high confidence**

***Table 3 the top 30 words with their counts.***

***Table 4 top 10 hashtags with counts.***

***Table 5 top 10 user mentions that are the most***

**A screenshot of a cell phone

Description generated with high confidence**

***Table 6 20 out of the followers, plot their ID numbers and screen names***

***Table 7 20 out of the friends and their ID numbers and screen names***

***Table 8 mutual friends within the two groups, their ID numbers and screen names***

A screenshot of text

Description generated with very high confidenceA close up of text on a black background

Description generated with very high confidenceA screenshot of a cell phone

Description generated with very high confidence

A close up of a map

Description generated with high confidence

***Figure 1 People Attention to share bike (2017)***

We grab the location data of the collected tweets in Json profiles and plot the distribution map about the ‘bike share’ topic. The color ’sky blue’ means this area has zero tweet about bike share, which means the citizen in this area pays little attention to bike share. While the color ‘red’ means this area arises large discussion about bike share. We assume that the deeper of the color, the higher acceptance for the bike share.

As we could see from the figure, Seattle has the deepest color. This is in accordance with the real market: Seattle is the one of the cities that share bikes launched firstly. There is a small red dot corresponding to the Washington city and the adjacent area has green color in the map. We could assume that the bike share business is still being spreading from the Washington city to the adjacent area. Therefore, it is possible that this area could be the targeted market for ofo.

However, there is another problem we need to consider. The area with deep color tends to be competitive in bike share market. There is already three bike-sharing companies in Seattle. The best targeted city for ofo should has relatively high acceptance for bike-sharing system. Meanwhile, the competition should not be that fierce. Eastern U.S might be a potential place to be the next targeted market since the color in this area is median. It could be a great trade-off between acceptance and competition.

For the individuals in middle part of U.S, the concept of bike-sharing system is relatively new. IF ofo decided to expand markets in such area, propaganda and advertisement is needed.

A close up of a map

Description generated with very high confidence

***Figure 2 people’s welcome to dockless share bike***

We do the sentiment analysis for ‘bike share’ topic in different cities for different bike share brand. The deeper the color, the customers are more satisfied with the corresponding brand. And the red circles centering at the different cities means the number of tweets for different bike share companies. We assume that the larger diameter of the circle, the larger market shares that the corresponding company has. Take the Washington city as the example. It has two circles. The bigger one corresponds to ‘Mobile Washington’, the smaller one corresponds to ‘Capital Bike Washington’. Their sentiment scores are 0.2314 and 0.2142.

A close up of a map

Description generated with very high confidence

***Figure 2 people’s welcome to dockless share bike***

We get the weight average of the tweets about different brand of bike-share Company and plot the precious figure again. For this figure, the color represents the individuals’ support for the concept of bike share. The deeper color means the citizens support the bike share more.