

## **DSCI 510 Final Project**

**May 10, 2022**

### **What do Twitter and Reddit users say about the Beijing Olympics?**

#### **Motivations**

The Beijing Olympics game has come to an end. People have passionately talked about the memorable moments that showed immense inner strength, athleticism and grace on social media. However, I find that there are several negative topics about “boycottBeijingOlympics” on twitter and Reddit. To learn what people say about the Olympics, I am going to collect comments and reviews from these two platforms.

#### **Brief description of data sources**

I want to find what's the most frequent keywords about the Beijing Olympics and the negative “boycottBeijingOlympics”, as well as the difference of discussion between that in Twitter and Reddit. Therefore, I made three word clouds from three dataset. The first one is a collection of tweets with keyword “BeijingOlympics”, the second one is a collection of tweets containing “boycottBeijingOlympics”, and the third one is a relevant post on Reddit.

##### **1. Scrape Twitter without using API**

I scraped Twitter according to the following conditions:

Select ‘BeijingOlympics’ as keywords,

Set time span from February 4th to 20th 2022, which is the time that the Beijing Olympics open.

This dataset consists of 6978 rows and 29 columns. I will only use the column “content”.

##### **2. Scrape Twitter without using API**

I scraped Twitter according to the following conditions.

Select “boycottBeijingOlympics” as keywords,

Set time span from February 4th to 20th 2022, which is the time that the Beijing Olympics open.

This dataset consists of 4975 rows and 29 columns (the number could change when scrawl in different time). I will only use the column “date content”.

##### **3. Scrape Reddit using API**

I collected comments from post “Would you support a boycott of the 2022 Beijing Olympics as part of a “tougher stance” on China?”, which is the most popular post with the “boycott olympic” keyword on Reddit.

([https://www.reddit.com/r/AskAnAmerican/comments/m5yj40/would\\_you\\_support\\_a\\_boycott\\_of\\_the\\_2022\\_beijing/](https://www.reddit.com/r/AskAnAmerican/comments/m5yj40/would_you_support_a_boycott_of_the_2022_beijing/))

## Analysis performed

Based on the tweets and Reddit content, I create three word clouds to see the most frequent word about Olympics, and compare the theme on two platforms.



Figure1: Word Cloud of “BeijingOlympics” Topic on Twitter

The first word cloud below shows that the most frequent words are “BeijingOlympics” and its synonyms. Besides, they talk about other topics as following:

- Specific sports, like “Figure Skating”.
- The people and countries involved, like “athlete”, “women”, “man”, “Canada” and “Russia”.
- Other related words to show people’s interest and care, like “watch”, “win”, “congratulations”.



Figure2: Word Cloud of “boycottBeijingOlympics” Topic on Twitter

The second word cloud above shows that the most frequent words are “BoycottBeijingOlympics” and its synonyms. Besides, the most frequent words are those from political fields, like “China”, “people”, “FreeTibet”, “CCP” and so on. There are not many words about the Olympics itself.

The third word cloud shows that the most frequent words are “Olympics”, “athlete”, “China”, which are very relevant to the Olympics itself. In addition to the political words similar to the previous word cloud it also has broader types of topics, which are absent from that in tweets.

- Words about competition, like “win”, “compete”, “medal” and “beat”.
- Words about feelings, like “care”, “love”, “hard” and “good”.

### **Conclusions drawn**

By comparing the three word clouds, we can draw two conclusions.

Firstly, most of the tweets about the Olympics are focusing on the event itself while the majority of boycotting tweets are talking about politics about China. In the word cloud of “BeijingOlympics”, people mentioned “athlete” and “medal”, and paid more attention to the fact that the Olympic Games is a sports event, but in the negative theme, “Humanright” and some political words have become dominant. In the first figure, there also many information about other countries like Canada, but in the second figure, the topic only focus on China, which is narrower.

Second, the topics of Reddit are more diverse and deeper than that of Twitter. While there are only one-way and negative themes in the tweets of “boycotting”, Reddit users also use keywords such as “good”, “love”and “watching” to express their attention to and affirmation of the Winter Olympics. In Twitter, people use the same terms or synonyms to reinforce their stand, but in Reddit, their expression consists of more personal opinions.