Full name and semester: Saranjeet Singh Saluja, Spring 2022.

Title: Sentiment Analysis on Tweets related to Cancel Culture (1350 words)

Introduction:

Cancel Culture is a phenomenon in which people, brands or shows are 'cancelled' due to what some people feel is offensive behaviour or problematic ideologies. This subject has been a polarizing topic of debate over recent years. While many feel that it is a way of calling out for accountability for people's actions, many argue that it has devolved into a kind of social media mob rule. There have been studies to understand the process of 'cancelling' [1] and to identify cancel culture as a form of social activism [2]. Through the study [3], the effects of 'cancel culture' on the ostracism of contrarians and the downfall of intellectual discussions within the domain of political science have been studied. This proposed study aims to contribute to this body of work by analyzing the various ways in which people react (positive, neutral, and negative) to Twitter posts and determining whether a specific outlook becomes more prevalent over time than others.

Research Questions:

This study addresses the following questions:

- How do people react to the term 'cancel culture'?
- How has the reaction to the term changed over the years?

Methods:

a. **Data Collection:**
The data for this study was collected from Twitter. The tweets were collected over the years from 2019 to 2021. The hashtags relevant to the subject – 'cancel culture', i.e., #cancelculture, #cancelled, #cancel, #yourecancelled were used to download the tweets from Twitter. For the same, four different datasets were created, each for one hashtag. The library 'snscrape' was used to collect the data. This library bypasses Twitter's restrictions and rate limits. Additionally, one does not require a Twitter developer account when one uses 'snscrape'. As I had to create a large dataset, I felt that 'snscrape' is the ideal choice for scraping the data.

b. **Data Cleaning:**
After scraping the data, I had 428738 data points. However, there were data points that were repeated, promotional, or some not which were not relevant to the subject of the study at all. These posts had to be removed as the aim of the study was to know people's opinions on 'Cancel Culture'.
To conduct the above, I found the usernames with the highest number of tweets and manually checked the nature of their tweets (i.e., if they were relevant posts or spam posts, promotional posts, or not relevant to the subject). The irrelevant posts were filtered out as much as possible.

To clean the data further, the posts containing less than three words, only hashtags and/or emojis were filtered out using regex to avoid reducing the quality of the analysis. Emojis tend to increase the neutral score, which in turn might lead to potential skewness. The tweets which were not in English, along with duplicate tweets were also filtered out. After cleaning the data, a random sample of a hundred tweets was taken each month from 2019 to 2021 using the sample() method of the pandas library in python. Overall, 3600 tweets were consolidated into a CSV file to carry out sentiment analysis.

c. **Analysis:**

For the study, the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool was used to carry out the sentiment analysis of the tweets. VADER is a lexicon and rule-based sentiment analysis tool that is particularly well suited to social media text. It combines lexical features with emotion intensities to calculate the sentiment score. The range of this sentiment score (which is also called the compound score) is between 1 (most positive) and -1 (most negative). As VADER is not only popular but also equally effective, I decided to use this tool to carry out the sentiment analysis.
The entire process was carried out in the following steps:

i. **Loading the data:**

The CSV file containing 3600 tweets was loaded as a data frame by using the pandas library. The same library was used to carry out aggregation and feature selection in the further steps. The data frame to be worked upon consists of 'Datetime', 'Tweet Id', 'Text', 'Username', 'Language' as shown in Fig.1.

	Datetime	Tweet Id	Text	Username	Language
0	2019-01-21 13:05:54+00:00	1087335477584248833	And tnis is who we want to buy \n#Cancelled	Amobbss	en
1	2019-01-07 08:28:37+00:00	1082192269413937152	And once I get to #tottenhamhale it's even wor...	BradChuck	en
2	2019-01-20 21:48:02+00:00	1087104489566453760	All money ain't good money #yingyang Ying Yang...	sweeeeeetiePie	en
3	2019-01-11 22:54:38+00:00	1083859759835557889	Just cause u put SYRUP on Shit, don't make it ...	KooLBeats5	en
4	2019-01-05 23:00:04+00:00	1081686798202519553	NoMo X-mas for whites. #Africans #cancelled #...	FARWPAS	en

Fig.1. Data Frame of the Tweets

ii. **Labelling using Sentiment Intensity Analyzer:**

The SentimentIntensityAnalyzer was used to analyze the tweets. The SentimentIntensityAnalyzer is accessible from nltk.sentiment.vader module. The analysis of the tweets yields a compound score which indicates if a tweet is positive, negative, or neutral. To label this compound score, I have considered the labels for the following scores:

- Greater than 0.05: Indicates a Positive score
- Less than -0.05: Indicates a Negative score
- Between 0.05 and -0.05: Indicates a Neutral score

	Datetime	Tweet Id	Text	Username	Language	compound	sentiment
0	2019-01-21 13:05:54+00:00	1087335477584248833	And tnis is who we want to buy \n#Cancelled	Amobbss	en	0.0772	positive
1	2019-01-07 08:28:37+00:00	1082192269413937152	And once I get to #tottenhamhale it's even wor...	BradChuck	en	0.2500	positive
2	2019-01-20 21:48:02+00:00	1087104489566453760	All money ain't good money #yingyang Ying Yang...	sweeeeeetiePie	en	-0.3412	negative
3	2019-01-11 22:54:38+00:00	1083859759835557889	Just cause u put SYRUP on Shit, don't make it ...	KooLBeats5	en	-0.8655	negative
4	2019-01-05 23:00:04+00:00	1081686798202519553	NoMo X-mas for whites. #Africans #cancelled #...	FARWPAS	en	0.0000	neutral

Fig.2. Labelling every tweet

iii. Aggregation of Data:

The pandas library was used to extract the Year from Datetime. The data frame was aggregated by Month and Year to see the monthly count throughout the years 2019 to 2021. This aggregation allows seeing the change in sentiment monthly per year as shown in Fig.3.

	sentiment	Year_Month	monthly_count	Year	monthly_count_percentage
0	negative	2019-01	37	2019	37.0
1	neutral	2019-01	33	2019	33.0
2	positive	2019-01	30	2019	30.0
3	negative	2019-02	39	2019	39.0
4	positive	2019-02	33	2019	33.0

Fig.3. Aggregation by Month and Year

Another data frame was created which was aggregated by Year only so that the change in sentiment over the years could be analyzed as shown in Fig.4.

	sentiment	Year	yearly_count	yearly_count_percentage
0	negative	2019	445	24.72
1	positive	2019	438	24.33
2	neutral	2019	317	17.61
3	negative	2020	495	27.50
4	positive	2020	413	22.94
5	neutral	2020	292	16.22
6	negative	2021	601	33.39
7	positive	2021	353	19.61
8	neutral	2021	246	13.67

Fig.4. Aggregation by Year

iv. **Plotting graphs to interpret results:**
An essential component of understanding how people's opinions have changed during the established timeline is graph plotting. Four graphs were plotted using 'seaborn' and 'matplotlib', two of the easiest to use libraries for elegant plotting of the graphs. The first three graphs are used to compare the month-wise percentage change in positive and negative sentiment for every year i.e., 2019, 2020 and 2021 as seen in Fig.5. The final graph is a bar plot that shows the percentage change in sentiment over the years.

Results:

a. Percentage change per month:

Fig.5 shows the percentage change per month and year from 2019 to 2021. In this case, as a sample of one hundred tweets was taken for each month, the monthly count is essentially the monthly percentage. By observing the chart, we can identify how the sentiment of the people is changing on a monthly basis.

Fig.5 Change in Sentiment by Month and Year

Initially, in 2019, the negative and positive sentiments percentage is approximately the same, however, the positive sentiment soars to a high of about 50% mid-year. Overall, it appears that the negative and positive sentiments are equally balanced. It is in 2020, however, that we get to observe a clearer trend where the negative sentiment percentage reaches a high of about 55%. It is in 2021, though, that we can observe the sheer difference in the positive and negative sentiment percentage, where the negative sentiment reaches an all-time high of around 60%.

b. Percentage change per Year:

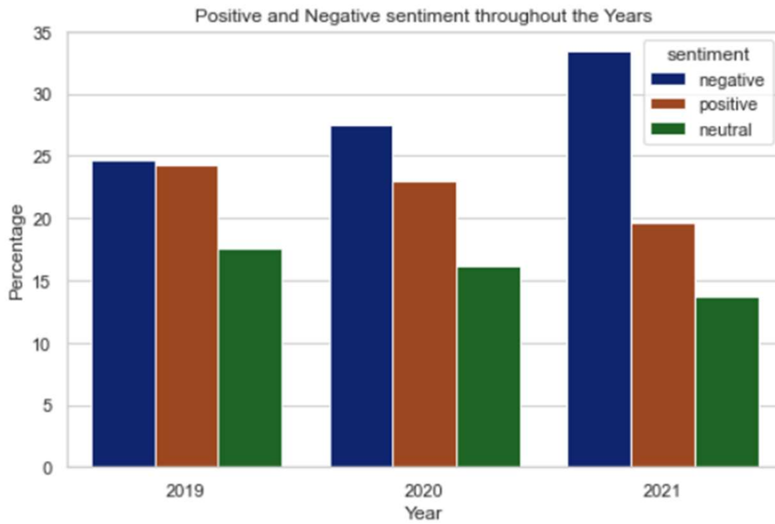

Fig.6. Annual Change in Sentiment

The observations and inferences in **Results (a)** are further solidified upon observing the annual percentage of positive and negative sentiments. The negative and positive sentiments percentages are nearly the same in 2019. However, in 2020, the negative sentiment percentage increases by about 4%, while both the positive and neutral sentiment showed a decrease of about 2%. In 2021, is when the negative sentiment rises to around 33%, an increase of 6%. In the same year, the positive sentiment percentage went down to about 20%, a significant decrease of 6% compared with 2019.

Limitations:

While the best efforts were made to reduce noise by removal of irrelevant posts, there is a possibility that the entire noisy data might not have been excluded. For example, manual checking of the tweets of the users with the most number of tweets was used to identify the noise in the data. However, there might be a single tweet that might not be relevant to the subject of cancel culture, which was included in the sample. This in turn might affect the quality of the data and hence, the study.

Conclusion:

In this research, sentiment analysis of 3600 tweets was conducted, to understand how people's perception of the subject 'Cancel culture' has changed over time. From 2019 to 2021, the negative sentiment toward 'Cancel Culture' or 'Cancelling Someone' has grown significantly. This potentially displays that the movement which started as means to stand up for the truth against the powerful has transformed into a toxic culture.

References:

[1] Samantha Haskell, “Cancel Culture: A Qualitative Analysis of the Social Media Practice of Cancelling”, Boise State University, 2021.

[2] Korri E. Palmer, “KancelKulture: An Analysis of Cancel Culture and Social Media Activism Through the lens of Minority College Students”, The College of Wooster, 2020.

[3] Pippa Norris, “Closed minds? Is a ‘cancel culture’ stifling academic freedom and intellectual debate in political science”, Harvard University, 2020.