**Twitter Sentiment Analysis of Climate Change and Global Warming**

**Introduction**

Climate change is an increasing concern among many people in the world and while some of us believe that it is real, some believe it is just a hoax. I thought it will be helpful to understand the trend of user’s opinions about climate change and global warming using twitter data. My question is, is climate-change really an emergency? I believe it is and we need to act on it. My hypothesis is that there are more people who believe that climate change is real.

**Data collection**

Twitter data was collected and analyzed to answer the question if climate change is an emergency and to test out the hypothesis. The REST API in the tweepy package in Python was used to collect around 2k tweets with the keyword ‘climate change’ on April 27, 2019. Retweets and replies were filtered out and not considered for the data collection. Only tweets in English language were used.

**Data Cleansing**

Every tweet was cleaned up to remove numbers, special characters, punctuations, http/https links, user mentions etc. Text was then converted to lowercase. The re.sub () function was used from the re module to perform the data cleansing. A pandas data frame was created with the username, text and cleaned\_text column after cleansing the tweets.

**Sentiment analysis and classification**

A new column was added to the data frame named ‘sentiment’ to categorize users as believers or deniers of climate change. In order to categorize tweets of people’s views on climate change being real vs not real, I used two lists of search keywords. First list had keywords such as ‘emergency’, ’actonclimate’, ’climateaction’, ’urgency’, ’climateemergency’ etc to categorize people as believers and second list had keywords such as ‘hoax’, ’fake’, ’lies’, ’misinformation’, ’misleading’ to categorize people as deniers. For tweets which did not have any of the above keywords from both the lists, the sentiment column had a value as “Non determinable”.

**Hypothesis Test**

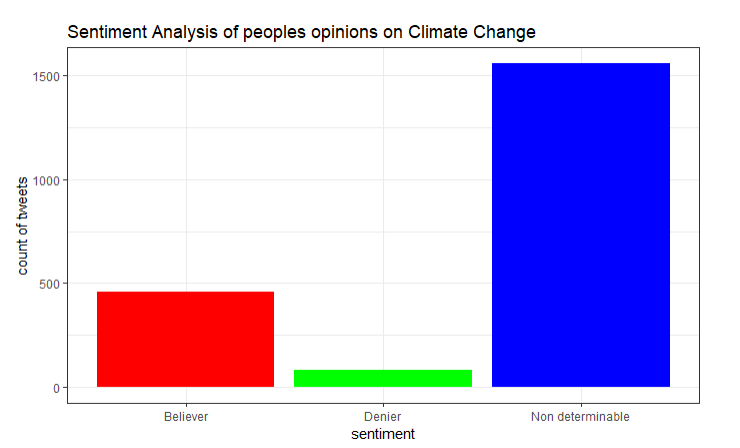
A 1-sample test of proportions was used as following:

Null Hypothesis: - P1 = 0.5, Proportion of twitter users who believe that climate change is real ***is*** 50%

Alternate Hypothesis: - P1 > 0.5, Proportion of twitter users who believe that climate change is real is ***greater than*** 50%.

**Plots**

Bar graph was created to display the sentiment variable from the data frame as shown below. From the bar graph, it can be inferred that most of the tweets are non determinable for believing vs not believing in climate change. For the rest of the tweets, it is clear that people who believe that climate change is real is greater than the people who believe that climate change is not real.



**Conclusion**

Based on the p-value above which is less than 2.2e-16, at the alpha = 0.05 significance level, there is enough evidence to claim that that the proportion of people who believe that climate change is real is **greater** than the proportion of people who believe that climate change is not real. Therefore, climate change is an emergency and we need to act on it to save our planet.