Data Analytics and Visualization Bootcamp

Sentiment Analysis using Social Media

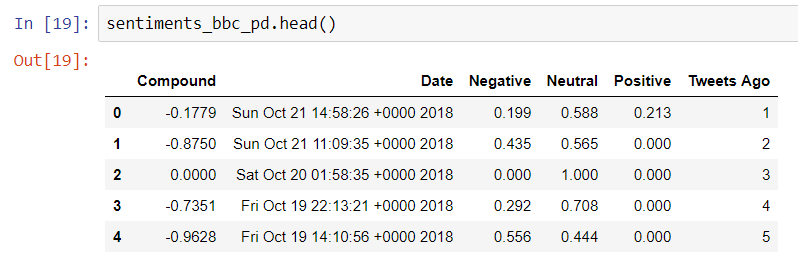
Jeff Olson

10/9/2018

**Summary:**

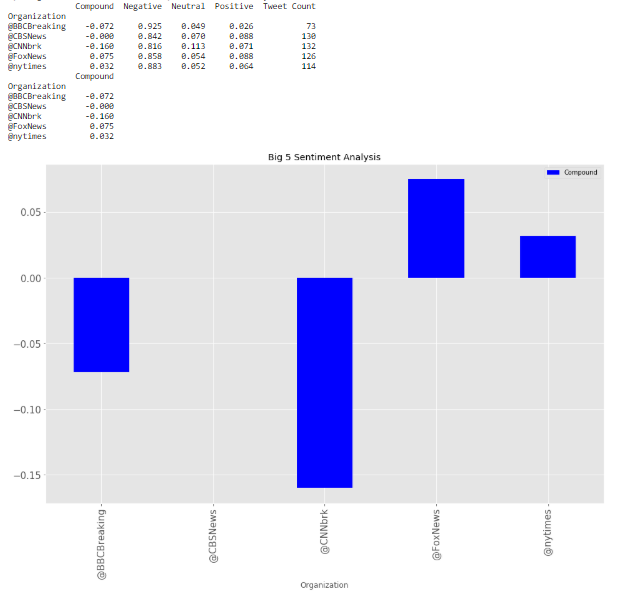
Twitter api registration allows one to do data analytics on tweets as well as construction of twitter bot applications in HEROKU. For this data, 100 tweets were collected from @BBCBreaking, @CBSNews, @CNNbrk, @FoxNews, and from @nytimes using Twitter api methods in a jupyter pandas application to access social media comments. The comments (or tweets) were passed through a VADER sentiment to analyze for positive, negative, neutral and compound ratings (or scores). Then the scores were charted out. CVS files and charts were exported directly from the application. Tweets were analyzed for trends. Figure 1 shows a typical data frame from the api query (BBC news in this case). Compound score is the Negative, n

**Figure 1) Sample DataFrame for BBC breaking news**

****

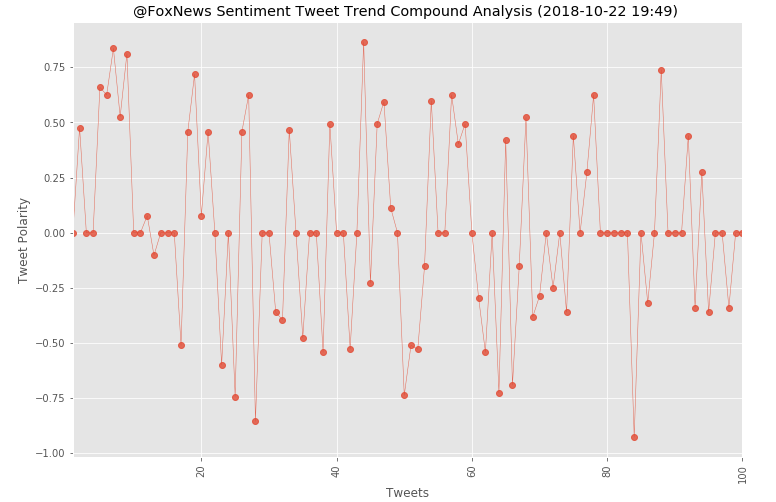
**Trends Discussion:**

1 – Sentiment analysis of all five news organizations was evaluated with the VADER sentiment algorithm. Sample sizes are indicated below. Results show in **Chart 1) Sentiment analysis by organization** that CNN is the most negative compound sentiment while Fox news has the highest. It’s a small sample though compared to annuals.

**Chart 1) Sentiment analysis by organization.**

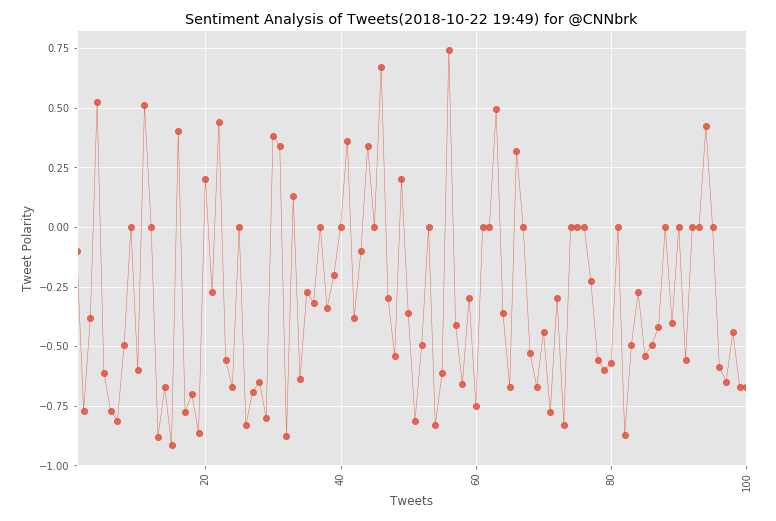
In regard to fox news, this is a sample less than 200, and data may be skewed by most recent tweet activity which shows a slight uptick but again it is a very small sample compared to annual.

**Chart 2) Fox Sentiment analysis by last 100 tweets.**



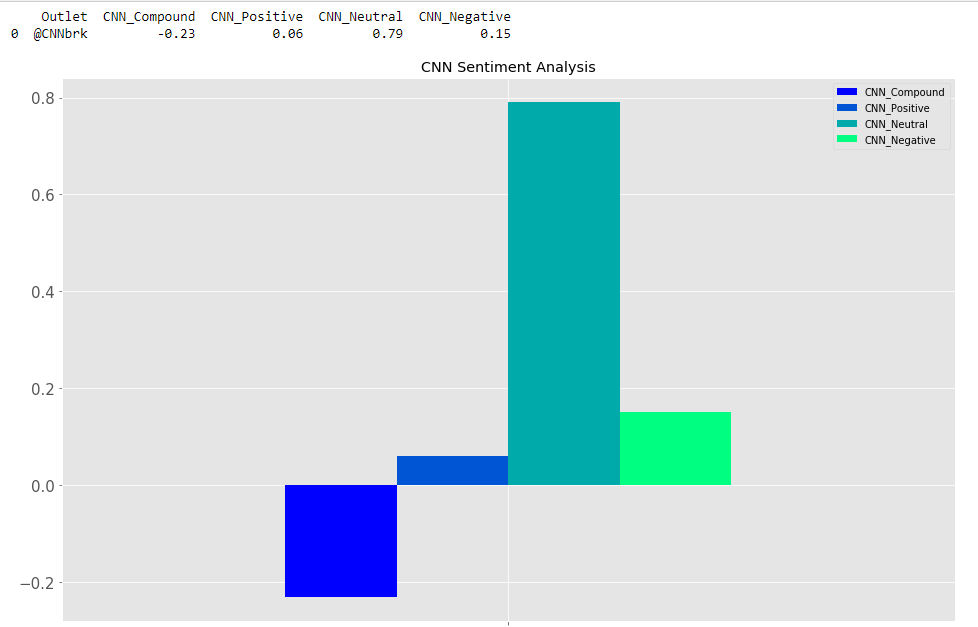
2 – CNN trend is worth looking at due to the findings in section 1. Overall the data in **Chart 3) CNN compound sentiment vs last 100 tweets** does appear to be weighted toward the negative sentaments.

**Chart 3) CNN compound sentiment vs last 100 tweets.**

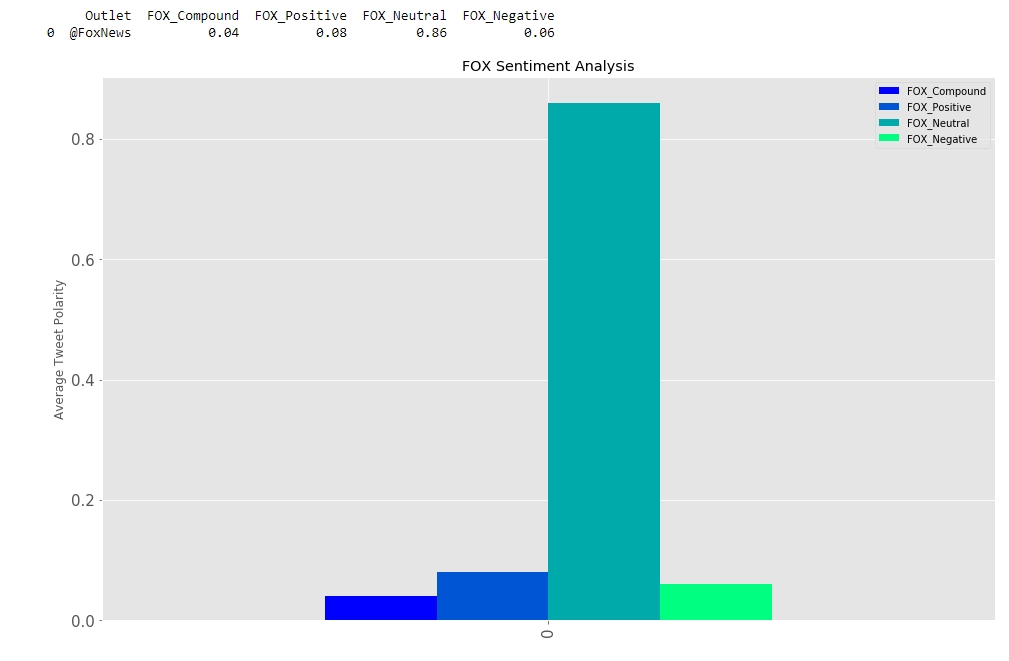


The bar chart comparison shows that fox does have some negative sentiments but they are less than cnn.

**CHART 4 a) CNN NEWS SENTIMENT BREAKDOWN**

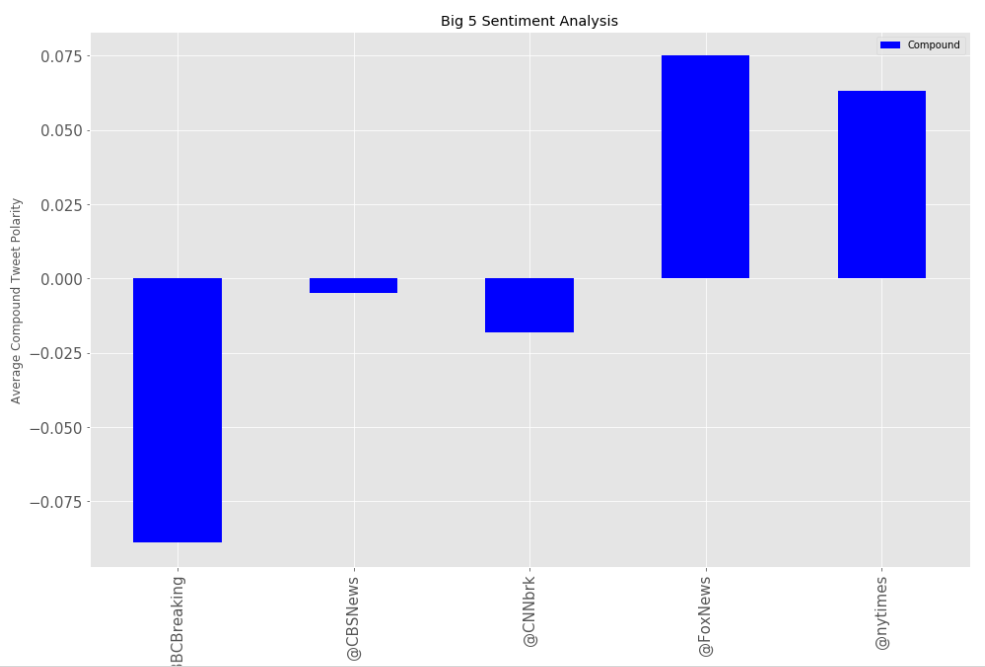


**CHART 4 b) FOX NEWS SENTIMENT BREAKDOWN**



**3** – Strangely enough FOX appears to be the only one with a positive compound sentiment. Not sure if I believe it because I do see negative stuff on there that sometimes is blatantly negative. So it leads me to wonder more about the VADER sentiment algorithm calibration and how it works. But that is beyond the scope of this assignment. I reran the Big 5 tweet sentiments (see CHART 5) REDO of FOX news sentiment breakdown)approximately 20 minutes later and NYT is also shown as a net positive on the compound score.

**CHART 5) REDO of FOX news sentiment breakdown.**



**What are some of the limitations of this dataset?**

First, it is a small sample and more historical data could be reviewed to add validity. Secondly, the data could be analyzed with a statistical significance based on sample size and analysis of variance to compare all 5 against each other (Tukey analysis) . If the distributions were compared with t tests over say an annual data set it may prove a statistical significance and add more credibility for a more concrete data interpretation.