Data Analysis

Summer Project

Gencer, Alper Sukru

October 2, 2020

Data Analysis

Here, I run the analysis:

There are two questions that I try to answer:

1. If the president's tweets show statistically different sentiment score before/after the issue date of executive order.
2. If the reaction (likes/mentions) that president's tweets receives before/after the issue date of executive order are statistically different than the montly average.

```
rm(list = ls())
gencer_TrumpTweets_3 <- read_csv("C:/Users/alper/OneDrive/Belgeler/GitHub/python_summer2020/Summer_Proj

####  There are some problems regarding python/R inconsistency. Let me correct them:
library(tidyverse)
gencer_TrumpTweets_3 <- gencer_TrumpTweets_3 %>%
  mutate(sentiment_bin = ifelse(
    sentiment_bin == "[[ True]]", 1, 0)
)
gencer_TrumpTweets_3[,ncol(gencer_TrumpTweets_3)-0] # Cool!
```

```
## # A tibble: 22,727 x 1
##    sentiment_bin
##            <dbl>
##  1             1
##  2             1
##  3             1
##  4             1
##  5             1
##  6             1
##  7             0
##  8             0
##  9             0
## 10             0
## # ... with 22,717 more rows
```

```
####  Let's correct the sentiment format:
gencer_TrumpTweets_3 <- gencer_TrumpTweets_3 %>%
  mutate(sentiment = gsub("\\[|\\]", "", sentiment))
gencer_TrumpTweets_3[,ncol(gencer_TrumpTweets_3)-1] # Cool!
```

```
## # A tibble: 22,727 x 1
##    sentiment
##    <chr>
##  1 0.53970174
##  2 0.53066443
##  3 0.50350911
##  4 0.53035888
##  5 0.50053568
##  6 0.50876087
##  7 0.49854986
##  8 0.48930895
##  9 0.49862007
## 10 0.47559445
## # ... with 22,717 more rows
```

```
gencer_TrumpTweets_3$sentiment <- as.numeric((gencer_TrumpTweets_3$sentiment))
```

Now let's find monthly like and retweet averages:

```
# Let's first correct the date format:
library(lubridate)
gencer_TrumpTweets_3$created_date2 <- as.Date((gencer_TrumpTweets_3$created_date), tryFormats = c("%m/%
gencer_TrumpTweets_3 <- gencer_TrumpTweets_3 %>%
  mutate(month = month(created_date2), year = year(created_date2))
gencer_TrumpTweets_3$monthyear <- paste(gencer_TrumpTweets_3$year, gencer_TrumpTweets_3$month, sep = "-


gencer_TrumpTweets_3 <- gencer_TrumpTweets_3 %>%
  group_by(month, year) %>%
  # Monthly like and retweet averages:
  mutate(favorite_monthavrg = round(mean(favorite_count))) %>%
  mutate(retweet_monthavrg = round(mean(retweet_count))) %>%
  # Deviation from the Monthly like and retweet averages:
  mutate(diff_favorite = favorite_count - favorite_monthavrg) %>%
  mutate(diff_retweet = retweet_count - retweet_monthavrg)
```

And lastly, let's find monthly sentiment averages:

```
gencer_TrumpTweets_3 <- gencer_TrumpTweets_3 %>%
  group_by(month, year) %>%
  # Monthly like and retweet averages:
  mutate(sentiment_monthavrg = (mean(sentiment))) %>%
  mutate(sentiment_bin_monthavrg = (mean(sentiment_bin))) %>%
  # Deviation from the Monthly like and retweet averages:
  mutate(diff_sentiment = sentiment - sentiment_monthavrg) %>%
  mutate(diff_sentiment_bin = sentiment_bin - sentiment_bin_monthavrg)
```

Hypothesis 1:

Let's check our first hypothesis that if the president's tweets show statistically different sentiment score close to the issue date of executive order:

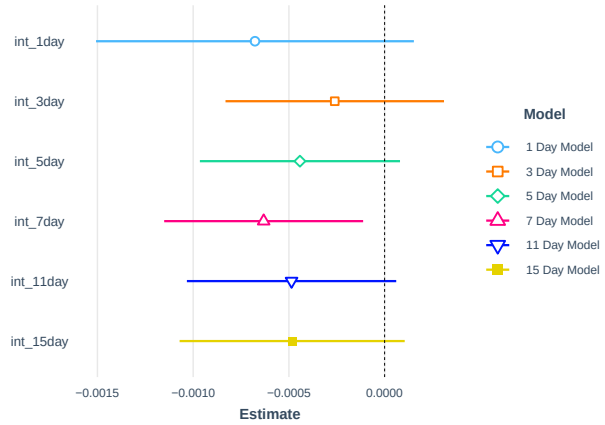

Figure 1: DV = Sentiment

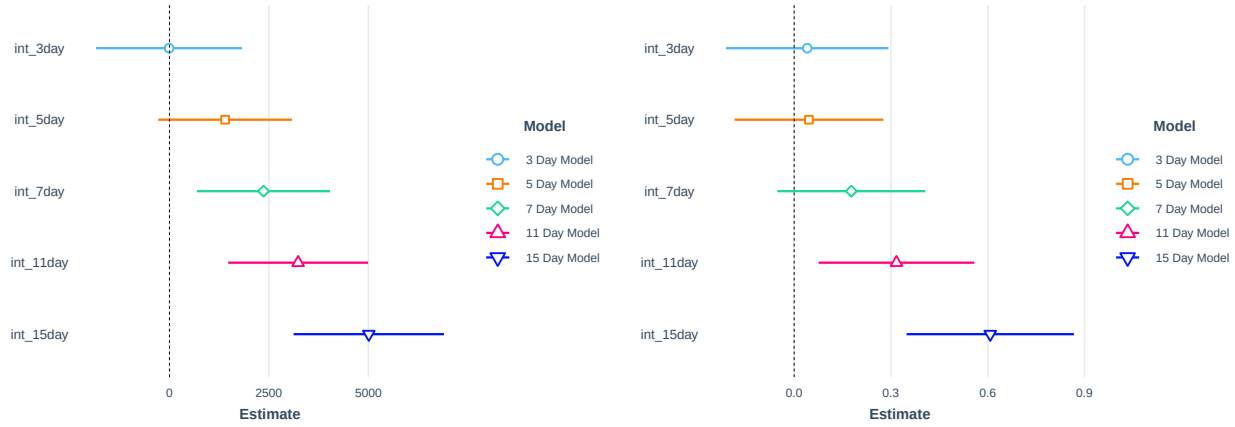

Figure 2: DV = Number of Likes

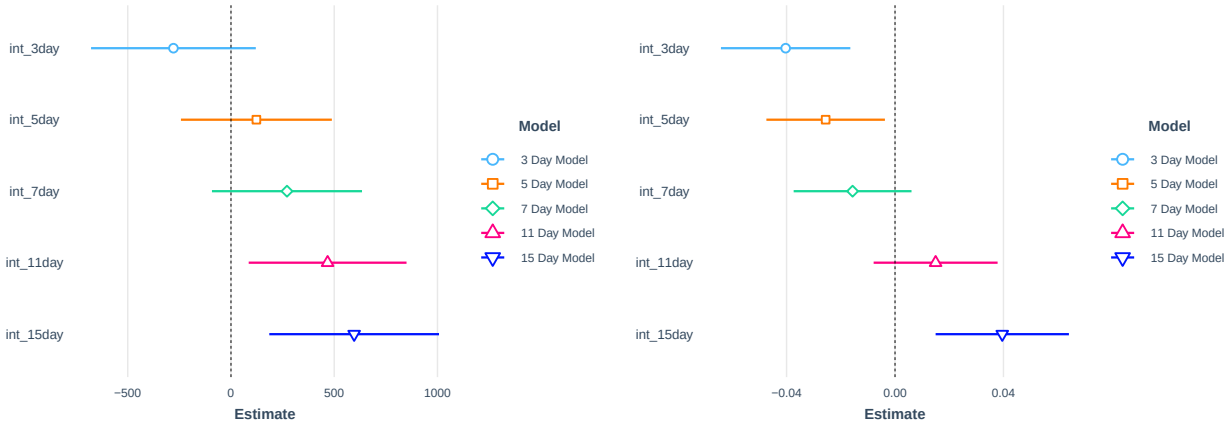

Figure 3: DV = Number of Retweets