# **Milestone 2: Descriptive Stats**

1. Provide a summary of the different descriptive statistics you looked at and WHY.

I started looking at the spread between normal tweets vs mentions and retweets. The reason for this is to get a feel for how often people interact with one another.

Then I looked at the people writing the most tweets to understand if this would then correlate with the people that have been retweeted or mentioned the most.

Afterwards I looked at the people that were mentioned or tweeted the most. This was done to figure out if there is a link between the two. i.e. when someone gets retweets gets favorited as well.

Finally I looked at the people who got replied the most to understand which person was more active communicating over twitter. Together with the person that got replied the most to understand who was creating the most attention from other congressman.

1. Submit 2-3 key points you may have discovered about the data, e.g. new relationships? Aha's! Did you come up with additional ideas for other things to review?

I found out that the person most retweeted was Donald Trump and was the most favorited as well by the twitter community. This result was identical when looking at the favorite tweet count. Sanders was second on both lists. This is not surprising given the amount of media attention they receive.

More interestingly is how different congressmen interact between them. Here I found that Dana Rohrabacher replies the most to people and that Senator Durbin receives the most replied from the other congressmen.

Finally, Republican Don Beyer seems to be the most active person on twitter, but only by a small margin as the top 20 people are within 40 tweets difference.

All these findings come with new ideas to review later like: Is there an evolution in the number of retweets a user has had over time? Or, what is the non congressman user tweeted the most?

1. Did you prove or disprove any of your initial hypotheses? If so, which one and what do you plan to do next?

I managed to prove my initial hypotheses of who was the most influential person on the database (Donald Trump) not surprising.

I also managed to prove that using the replies between congressmen we can find some relationship between them finding who are the most influential people over the platform.

1. What additional questions are you seeking to answer?

I want to deepen on the sentiment analysis of the tweets as well as extract the main topics covered on the tweets. This will give further insights as to what is the main areas of debate between the different parties and if there are any skewness in what one party talks about and the other.