

# Exercise 2 | **Interviewing Trump's tweets dataset**

Author: Ludek Cizinsky

Date: 14. 02. 2022

# TASK 1 | **EDA**

## EDA | Exploration and brainstorming 1

### Properties of the dataset

- There are in total **3 sheets** with data: sheet showing top terms in the tweets sorted by frequency, sheet with original tweets and sentiment analysis, sheet with raw data.
- In total there are **56,571** tweets from which **46,694** are original tweets (the rest are re-tweets). There are **1,234 uni-grams** with corresponding frequency.
- There are three columns describing sentiment: strength of positive and negative sentiment, average of these two. The positive sentiment strength has range from 1 to 5, the negative one has it from -1 to -5. Interestingly, average column in fact is not average of the two columns, but just a sum which makes sense.
- Interestingly, the following tweet would be according to the average column **neutral**:

*My persona will never be that of a wallflower - I'd rather build walls than cling to them. --Donald J. Trump*

I believe that this should have a negative sentiment, but if it was inspected by some NLP system, then it probably makes sense as not having the full context, this does not appear negative.

- By further inspection of tweets, I noticed that some of them have for example high number of retweets, but 0 people made it their favorite. I assume this is perhaps to fill missing value

## EDA | **Exploration and brainstorming 2**

### **Transformation of data**

- I could group the tweets by year and examine evolution of properties like number of retweets or favorites
- I could sort the tweets by popularity – e.g. by number of retweets
- I could filter out flagged or deleted tweets
- I could filter out retweets (which has been already done)
- I could filter tweets for only certain range of sentiment
- I could try to look at outliers such as tweets with high number of retweets and 0 favorites and filter these out

### **Questions**

- How has Trump's popularity evolved over time where popularity can be measured by either the number of favorites or retweets?
- Is there a trend in terms of sentiment of tweets related to Trump?
- Has sentiment of the tweets changed during elections?
- How many tweets were posted on average per day for each year?

## TASK 2 | **Concepting**

## Concepting | **Plot selection**

### **Selected question**

*How has Trump's popularity evolved over time where popularity is measured by number of retweets?*

### **Selected plot**

I first narrowed down my search to plots which are used for showing time series data. One of the candidates was for example **line chart** which is a very popular option due to its simplicity and easy readability. However, I then decided to choose **horizon graph** whose main advantage is that it is capable of highlighting density of time series view at relevant ranges while preserving the resolution of the graph. I believe this should help me better demonstrate the growth of Trump's popularity as the visual appearance, at least to me, looks more robust, and thus is more convincing than for example line chart. On the opposite side, this comes with a tradeoff as for more general audience, this might be a type of a graph that they have never seen before and as such, the information which is being communicated might get lost due to the lack of understanding of the horizon graph and how it works.

## TASK 3 | **Editorial brainstorming**

## Editorial brainstorming | **Plot design notes**

### **Which colors to use**

I would like to use a color which is rather neutral but still relevant to the topic, i.e. Donald Trump. Inspired by the American flag, I decided to use **blue**. This also makes sense from a perspective of Twitter whose main color is also blue.

### **Headline**

The headline should capture shortly but precisely the main point of the figure, i.e., demonstration of evolution of Trump's popularity on Twitter described by # of retweets of posts related to Trump.

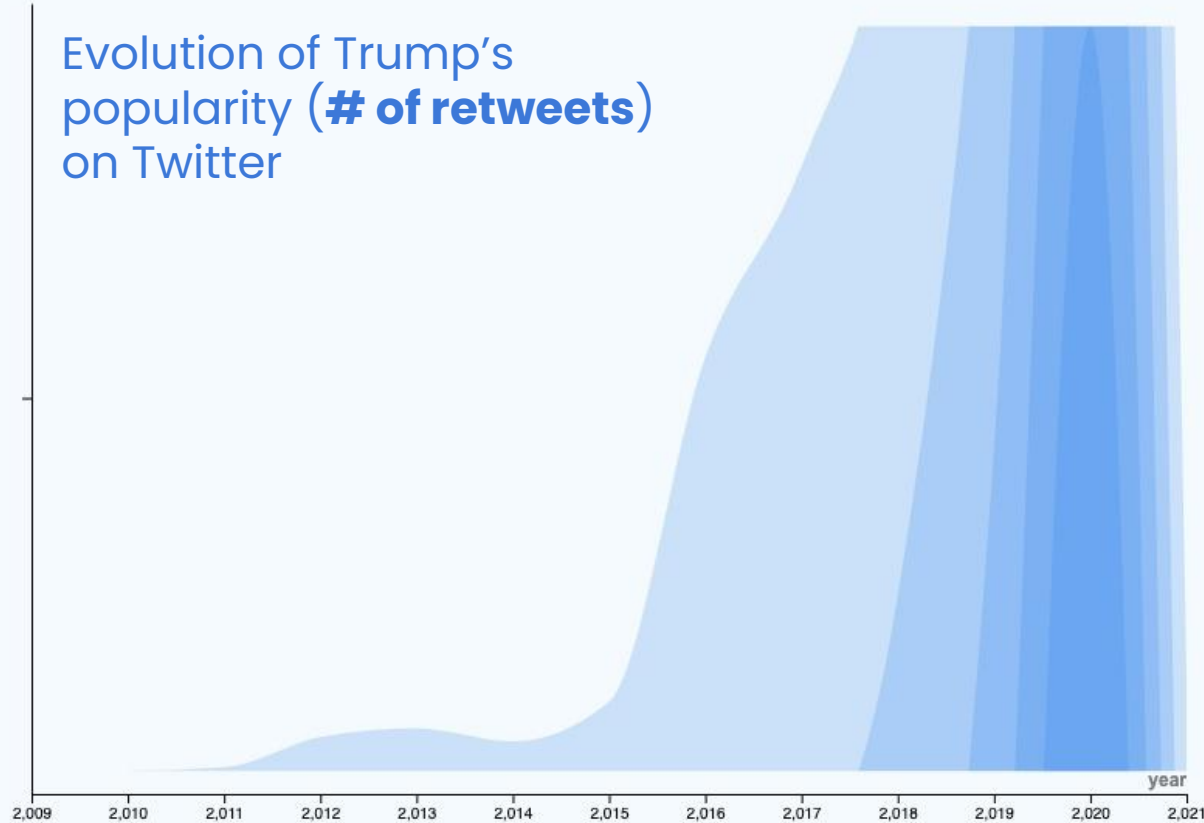
### **Other considerations**

I would like to also add a small section describing the "behind the scenes story" to give the reader a context. Finally, it would be nice to also use some of the available tweets to add more authenticity to the figure.



TASK 4 | **Visualize the graph in rawgraphs.io**

## Evolution of Trump's popularity (# of retweets) on Twitter



### About

Trump announced his presidential candidacy in June 2015. Despite not being taken seriously in the start of his campaign, he quickly gained a popularity. (Wikipedia)

“  
@MalachiNorris: Donald J. Trump is the very definition of the American success story. Through Trump we can bring back the American dream.”