Discussion: Connecting Data and Society
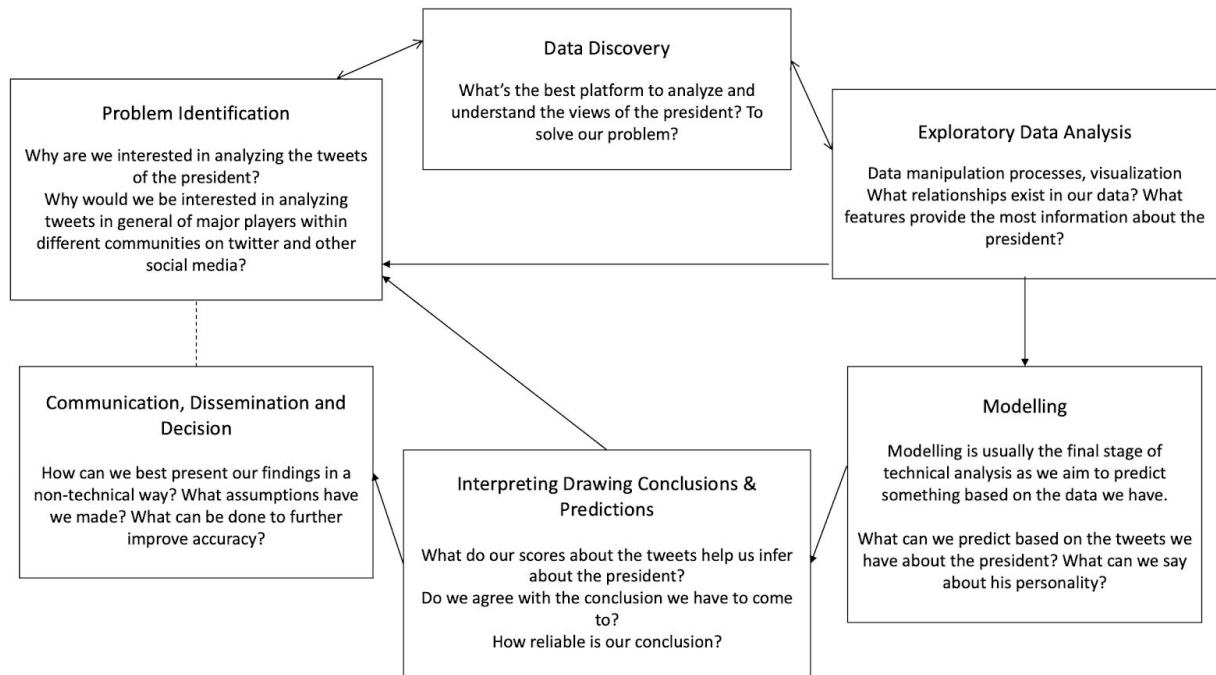**Sentiment Analysis on Trump's Tweets**

1. Understanding the Data Science Life Cycle

	Data Discovery	
Problem Identification	What's the best platform to analyze and understand the views of the president? To solve our problem?	**Exploratory Data Analysis**
Why are we interested in analyzing the tweets of the president? Why would we be interested in analyzing tweets in general of major players within different communities on twitter and other social media?		Data manipulation processes, visualization What relationships exist in our data? What features provide the most information about the president?
Communication, Dissemination and Decision	**Interpreting Drawing Conclusions & Predictions**	**Modelling**
How can we best present our findings in a non-technical way? What assumptions have we made? What can be done to further improve accuracy?	What do our scores about the tweets help us infer about the president? Do we agree with the conclusion we have to come to? How reliable is our conclusion?	Modelling is usually the final stage of technical analysis as we aim to predict something based on the data we have. What can we predict based on the tweets we have about the president? What can we say about his personality?

2. Sentiment Analysis and **what it represents**

Representation: The way in which one thing is made to "stand for" another. Technologies create representations of people and of social/natural phenomena that do particular work in the world and acquire a life of their own, refiguring the identity and agency of the represented person/phenomena.
   a. What are the different uses of sentiment analysis that you can think of - feel free to go beyond the context of political discourse. Eg. Consumer insight for businesses
   b. Reflecting on the homework you completed. What was the purpose of using sentiment analysis? What did it allow you as the data analyst to do that you would not have been able to without the tool?

3. **Power** implications

Power: The asymmetric capacity of an agent to structure or alter the behavior and decisions of other agents, populations, or systems. Technological (computational) power is intertwined with political power.
   a. How might someone in a position of power use their sentiment analysis to influence the decisions, behaviour or ideologies of others.

b. Is the power of sentiment analysis more positive or negative - does it differ based on context?

4. Based on your understanding of how VADER works, what are some problems with this tool? (Hint: Think bias, aggregation techniques, power and influence)

Sociotechnical Imaginaries: An organization in which people and technology interact and work together such that human and technical agency is complexly intertwined and distributed. Large and highly complex sociotechnical systems distribute risks and responsibilities widely and unevenly, and are difficult to regulate. When they fail it is often difficult or even impossible to identify a single human or mechanical cause.

a. How might this relate to the idea of **sociotechnical imaginaries**? If VADER makes errors, is it a human, mechanical or unidentified cause?
    i. If you're interested in learning more about the VADER lexicon: https://github.com/cjhutto/vaderSentiment

Reflecting on Representation as a Whole
Apart from sentiment analysis, some examples that use some sort of scoring or numerical method to represent information are:
- Risk Scores (eg. Health)
- Ratings for various experiences
- Integer representation through one-hot encoding

a. What are some uses and some limits of the abstraction that occurs in the process of generating scores?
b. Do numerical, objective scores always accurately represent what they set out to define? Why or why not? (Hint: think context, objectivity vs. subjectivity)