Final Project Proposal

DS 710

For your final project, you will collect and analyze data from Twitter to answer a question that interests you.

1. What is your question? You might be interested in a brand, an organization, a meme, a trend, a news story, or a place. What do you want to learn about? Do you have predictions or hypotheses about what you will find?
   * For the final project, you will need to summarize your analysis and results in a single page, so keeping your question fairly simple is a good idea. For example, comparing tweets on the top 30 brands of bottled water is n­­­­­­ot better than the top 3.
   * Choose a question for which the answer is not already obvious. For example, a project that compares how many times the hashtag #Nike is used in Oklahoma vs. Texas will obviously find more tweets in Texas, because the population is so much higher! Likewise, the #Packers hashtag will be more common in Wisconsin than Florida, and a project would need to do a deeper analysis in order to be interesting.
2. What is your audience? Who is interested in the answer to your question?
   * To help you focus your writing in the executive summary, this should be a specific audience for a specific reason, not something like, “Myself and my friends, because this seems like an interesting data science question.”
3. What information will you gather about each tweet in order to address your research question?
   * Some possibilities include the text of the tweet, the time, the username of the Twitter user, the location, etc. However, if you don’t need all of these to help answer your question, then it may be beneficial to save only the information you need, to prevent your file sizes from becoming unwieldy.
   * Many tweets do not have location tags. You are welcome to choose a research question involving location data, but you may want to think of a back-up plan in case you’re not able to gather enough tweets with the information you need.
4. How will you collect the data? Will you use the REST or Streaming APIs, or a combination of both? What specific keywords, hashtags, or other features will you search for?
   * Choose a data collection plan that will allow you to gather at least 1000 tweets.
   * The REST API returns tweets from the past 7 days. Choose a research question that can be addressed in the time you have available between now and when the final project is due.
5. How will you analyze the data? Will you compute any proxy variables (such as number of exclamation points as an indicator of emotion, like in homework 11)? Choose at least one pair of specific hypotheses to test. What type of hypothesis test will you use?