**Project Design Writeup and Approval Template**

**Project Problem and Hypothesis:**

The project seeks to explore the capabilities of natural language processing. I will be analyzing customer service in a number of verticals based on the sentiment of tweets extracted. I will explore the following verticals

* Airlines
* Hotels
* Healthcare
* CPG
* Telecom

Predictions will include sentiment scores, understanding customer pain points, and projecting ways in which changes could be made. It will be a continuous rather than binary outcome. This could have an impact on the way these industries or specific companies approach customer service. I think whether the company has had incidents occur or not could have an impact on values.

**Datasets**

* I will be pulling from the Twitter API with Python wrapper
* <https://github.com/bear/python-twitter/wiki>
  + Status and User information
  + Number of followers
* I will be obtaining information for the following Twitter handles
  + *@united*
  + *@JetBlue*
  + *@Delta*
  + *@SouthwestAir*
  + *@StarwoodBuzz*
  + *@hilton*
  + *@MarriotIntl*
  + *@Nestle*
  + *@ProcterGamble*
  + *@Unilever*
  + *@KraftHeinzCo*
  + *@pepsi*
  + *@TysonFoods*
  + *@ATT*
  + *@TMobile*
  + *@sprint*
  + *@verizon*

**Domain knowledge**

I have very little domain knowledge in this space. Other research efforts in this space include:

* There has been significant research in this space. Some examples include
* Use a quick Google search to see what approaches others have made, or talk with your colleagues if it is work related about previous attempts at similar problems.
* *Include a benchmark, how other models have performed, even if you are unsure what the metric means.*

**Project Concerns**

* Questions:
  + Many companies have multiple accounts –how do I deal with this, for example, what about companies that have dedicated customer support handles?
  + What if the variation within channels is higher than between? What question will I answer then?
* What are the assumptions and caveats to the problem?
  + I wish I had access to more customer support data, as most people do not take to twitter to complain
  + I’d like to know what percentage of complaints occur on twitter
* What are the risks to the project?
  + Will the NLP algorithms accurately capture the information I am looking to extract?

**Outcomes**

The goal of this project is to better understand customer complaint behavior, with the ultimate goal of feeding this information into a customer support bot or into improved processing for an NLP system. The expected outcome would have by-company and by-vertical breakdown of sentiment by a number of key categories. There will be an analysis by:

* Sentiment, both by company and customer
* Content, both by company and customer
* ‘Event Handling,’ which will show responses to key events and outcome using time series
* Identifying key items for change will mark success for this project