**Project:** Boys & Girls Club Program Analysis

**Description:** Executive Director needs to annually report to the state on effect of Boys & Girls Club after-school program. This program helps struggling students and is funded by the state. His immediate goal would be me completing EDA, next goal would be definitively showing effect (or lack thereof), and “moonshot” would be a predictive model that identifies the important features and help identify students with most need.

**Data**

* Exists (.csv) and available to me
* Size is ~1000 students and ~50 features

**Technology**

* Algorithm: Linear Regression or Random Forest
* Lots of Excel

**My take:** Checks business value box, but doesn’t do much for me. I like this as a side project for my resume rather than capstone

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Project:** Chatbots that can emulate personalities

**Description:** Chatbots are usually designed to answer questions or guide/motivate a user through a process. I want to see if I can make a chatbot troll somebody, make a chatbot get offended by somebody.

**Data**

* [Positive/Negative tweets](http://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/) ([source 2](http://www.sananalytics.com/lab/twitter-sentiment/))
* [Question Answering](https://data.quora.com/First-Quora-Dataset-Release-Question-Pairs)
* Urban dictionary (web scrape)

**Technology**

* Algorithm: NLP sentiment analysis
* Big data tech (AWS and such)
* Proof of concept ([Article,](https://www.technologyreview.com/s/608387/an-algorithm-trained-on-emoji-knows-when-youre-being-sarcastic-on-twitter/) [Website](https://deepmoji.mit.edu/))

**My take:** Business value? Not really. Feasible? Its aiming high. Will endlessly amuse me? Priceless.

**Project:** Sentiment Analysis of Speech

Description: Classify customer service calls into positive or negative. I actually found out a company already does this (callminer)

**Data**

* There’s speech datasets but I couldnt find any with labels (like positive/negative)

**Technology**

* Algorithm: NLP sentiment analysis
* Big data tech (AWS and such)

**My take:** I like the NLP field with Watson and sentiment analysis. I want to do something in this field and I’m still just waiting for the right idea/dataset.