**VA 102 Proposal**

**Project Idea/Goal:**

I am planning to do a NLP analysis on the reviews data from the People Per Hour website to analyze what themes does people’s reviews generally follow. So the question I am trying to answer is this:

What things do people base their reviews on? Thus, what do people care more about?

**Why this project?**

It will help BuckMeUp in making their review algorithm as it would provide them with relevant information about the consumer’s perspective of how they rate things. They thus could refine the review metrics they offer so that people get more detailed reviews, or they could build an automated text theme classifier into our web app for automated review classification. They could then maybe make a recommender system which provides results based on individual or some combination of themes thus making for a better user experience.

**How I will accomplish this?**

1. Collection of the dataset ( from peopleperhour.com). As I have been working with Stan on this since August, the scraper for this has already been built and most of the data has already been collected.
2. Using the NLTK to extract themes and tags from the reviews and classify them. Over the next few weeks( ~4-6 ), I will be looking into NLTK and trying to make a themes extraction algorithm to that would classify the reviews in 3 categories (money-centric, time-centric, quality-centric).
3. Using Tableau, D3,js and other software to visualize the findings. The last stage would be doing some visual analytics with the collected themes to do some structured analysis which I haven’t completely decided yet.