**Capstone Proposal:**

**No Sponsor, CA Healthcare**

1. **Project Information:**

First, I would like to say that I could not find a sponsor. The people that I personally know that work with data all work in fields where discretion is mandatory (healthcare, legal, and insurance). A lot of this material is confidential and thus, I am legally not able to access it. I didn’t have any luck online either. So, I will instead be looking at some data found on data.gov for this project.

The data set I will be using is focused on healthcare, as that is a field that I have professional interest in. I will be looking at a dataset analyzing California hospitals and their quality ratings and mortality ratings regarding six different procedures. Analysis surrounding this data can have real world implications when it comes to understanding the most effective hospitals and which hospital is better at treating certain conditions. This can be invaluable information for those seeking treatment centers as well as for the state itself in analyzing their state-wide medical coverage.

1. **Project Goals:**
2. I plan on finding the most effective health-care center in California.
3. I plan on listing the top healthcare facility when it comes to treating each of the six listed conditions.
4. I, inversely, plan on finding the least effective centers
5. **Data Overview:**

There is quite a lot of data given. At nearly 300,000 pieces of data, I understand this may be too large a data set to use at this point in my data-analysis career. Unless told otherwise I’d still be willing to take a crack at it. It is formatted well, and some data isn’t entirely necessary for my purposes so I could clean it up significantly by removing fields. I’d say the given data is easy to understand on the surface as well so despite its large number of values it could be taken care of. It is available in csv form as well allowing me to plug it into any of our programs I feel comfortable with.