I’d like to brief my high-level scrutiny about cap-stone projects.

Unfortunately, there is likely a sort of juxtaposition between practicality and large beautiful datasets. For instance, nature based studies have likely the best data, like volcanic or earthquake studies, but less jobs in the market. On the contrary, revenue/churn/acquisition predictions are very practical, especially from my perspective, however it’s more difficult to find an original data. I likely want to stick with practicality of what I do, so I’m more close to the former though.

Overall there are 3 domains and I have one or two candidates for each:

* Regression: Kaggle 1C Company competition data or Walmart for revenue prediction
* Classification: Telecom operator churn (Kaggle) or earthquake damage (DrivenData) or volcanic eruption (Kaggle) predictions
* Text Mining: Customer comments of social media per two or three mediums, batch or stream, a telecom company preferably

For the 2nd one, I may go for one of the first two buckets above, where the last bucket likely makes a 3rd cap-stone project I believe. In the end, I want to make sure that I spend my time wisely and I’d like refrain from anything that does not contribute to my finding a job journey.

To be more specific, a churn analysis looks like a better fit for the 2nd cap-stone project. KDD movement favors the famous 2009 competition for this particular domain, however the original Orange Telecom data is not there anymore. The closest option looks like the one here in this link, with 7K+ rows and 21 features: <https://www.kaggle.com/mmsant/telcom-customer-churn-exploratory-analysis/data>. One the supporting sources that can be utilized is Data Science for Busines (Provost et al, 2013).

Lastly, I’d like to thank my mentor, Nik S., for his continuing guidance for this.