Three Capstone Project Proposals - Laura D. Rose 5.6.17

1. A problem at my company is accurately forecasting demand for certain pharmaceutical products. I've looked at the top SKU contributors to our forecast variance over the past six months and realize that if we could improve forecast accuracy with a better model(s) for one or more of these SKUs, we could improve overall forecast accuracy by about at least five percent. My data sources for this project would be company demand, shipment data and government source flu data. At this point, without having extensively studied the data, the explanatory variables for this project could be a large customer's forecast (customers provide us their forecasts of what they expect to buy from us), flu instances as reported by the CDC, and the company's customer sales order or shipment history, which is available from company databases. The flu data would need to be collected or aggregated at the monthly level because the demand data would be at this level. However, another option for the flu variable is a kind of dummy variable where 1 = strong flu season and 0 = not strong flu season. The reason for this is that our customers would presumably buy based on how they perceive the flu season to be, not actual flu cases necessarily. To figure out if a flu season is strong, I would need to look at prior seasons going back several years and compare to see what could constitute a strong season. There are several ways of looking at this, so different methods would need to be considered. For instance, one method could be what is the rate of change between weeks. Another possible explanatory variable is patient consumption data for the drugs for which our products are an ingredient. I could probably obtain the ingredient info from a salesperson. The company marketing department has access to the patient consumption data. At this point, I think that a linear regression model of some type would be helpful for this data, given that it is continuous. However, I'm open to other methods that I learn during the course.
2. I work in supply chain, so predicting product backorder is helpful to the company. Kaggle.com has a dataset involving predicting product backorder. https://www.kaggle.com/tiredgeek/predict-bo-trial This would be an opportunity to employ logistic regression or other techniques beyond the standard linear regression.
3. I work for a pharmaceutical company, so the opiate overdose epidemic is a topic of concern. The following dataset on Kaggle is related to predicting opiate overdoses based on opiate prescripions by prescribing doctor, drug, etc. https://www.kaggle.com/apryor6/us-opiate-prescriptions At this point, I would guess that the data could be modeled with a linear regression, but I'm concerned that could be too simplistic. It's a fairly complex dataset with many drugs and prescribers. Determining how to model this accurately could be outside of the scope of this workshop and may be a project for a more advanced workshop.