Stat 390 Weekly Report 4 (10/30-11/03)

Progress/Accomplishments:

- Kaylee learned how to do a Prophet model and read about the parameters of the Prophet model and how to tune it
    - She has developed the most basic Prophet model and calculated the initial RMSE
- Blanka has started developing her models, she has developed and tuned ARIMA model and developed AUTO ARIMA model.
- Ava has continued to do feature engineering adding more lag terms to account for a three month pattern she noticed
    - She also added a rolling mean and standard deviation feature
    - Found correlation between engineered lag features
    - She has begun to develop and tune her XGBoost model
    - Her preliminary XGBoost model (untuned) performed relatively well
    - After using RandomizedSearchCV was able to lower RMSE
    - Implemented validation groups instead of just a test set

Problems/Challenges:

- Kaylee's Prophet model has no parameter tuning for a logistic growth model (our model) and only for a more linear model (which is not ours)
    - Kaylee is continuing to learn which hyperparameters to use and tune in a Prophet model
- Blanka's ARIMA model works fine and is pretty straightforward to tune using grid search. However, the AUTO ARIMA model is difficult to tune. She has made a naive model by doing one day shift for predicting values, this base line model has a very low RMSE of 0.06. She was able to achieve the lowest RMSE with ARIMA which was around 0.5, however could not get RMSE lower than 1.6 with the AUTO ARIMA model. She has conducted extensive tuning with different parameters, but it did not provide much more improvement. We believe the low RMSR of the base model is due to the very strong correlation between data points that are one week apart.
- Ava
    - Initially had issues with one-hot-encoding categorical columns, but figured out issue
    - RandomizedSearchCV takes a long time to run on my computer, so it takes a long time for me to move between tuning steps
    - Still trying to figure out a balance between using the suitable amount of lag terms and which to use in the model

Plan/Next Steps:

- Kaylee will continue to learn the Prophet models, develop them and tune them. She is behind thus far (I am acknowledging that) and will work extensively this weekend to catch up
- Blanka will continue working on tuning AUTO ARIMA and developing the other models. She has to figure out why the RMSE will not go down with any tuning for A-ARIMA. Possible approaches she has thought of are taking another look at the data and seeing if there might be a problem there as well as researching what exactly each parameter represents so she can have a better understanding of the tuning process.
- Ava
    - Perform a finer grid search with preliminary findings from broad RandomizedSearchCV tuning
    - Begin to develop keras model
    - Continue to figure out which lag terms will work well in adapting to patterns in the data