Project 3 Proposal:

The group includes Eliot Chern, Sean Galloway, Ian Koenes, and Cora Micsunescu. This project will use the IMDB data from Kaggle (<https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset?select=IMDb+movies.csv>) with 85,000 records to create a Flask based application that performs a few different analyses. The base data source is the CSV files. If needed within the flask app, an additional data store will be MongoDB. The flask app is being deployed to Heroku.

One analysis will take the data and use a few different machine learning tools to learn the movie's profitability based upon the genre and the description. The anticipated models are Logistic Regression, Random Forest, and Kera Neural Networks. Accuracy is the measure used to determine the model's performance. The webpage will hold graphs on the model's accuracy and the calculated profitability from a user-selected genre and narrative. A flask application will generate profitability on the saved models, while javascript handles displaying the information.

A second analysis will test the hypothesis that the most profitable movies are English language movies. The models used will be Linear Regression, Logistics Regression, and Neural Networks. Since it is not known initially what model works best, the models covered in class will be used.