Capstone 2 proposal

Idea 1: Music Recommender

Suggesting music has become the interest of many music streaming businesses such as pandora, spotify and apple music. This essentially requires the construction of a recommendation system which relies heavily on machine learning techniques:

1) At its core the recommender system would utilize:
    a) Content based filtering
        i) Songs are tagged with different attributes
        ii) Lyrics - opportunity for sentiment analysis
        iii) Genre
    b) Collaborative filtering
        i) Taste profiles of over 1,019,318 unique users
        ii) 48,373,586 user play counts
2) Network analysis can be performed treating users and songs as nodes in a bipartite graph (another method for song recommendation)

Datasets:
https://labrosa.ee.columbia.edu/millionsong/

Idea 2: Feelings About Food

This dataset contains reviews, business info and user data all in json format. Over 5,200,000 user reviews and 174,000 businesses in 11 metropolitan areas. Using this dataset the following could be analyzed:

1) Ratings could be used to profile similar users in order to recommend new restaurants. A user interface could be developed to input a users account information and suggest restaurants based on their history.
2) A sentiment analysis could be performed on the reviews which could be used to predict positive or negative reviews.
3) Network analysis on restaurants could be performed treating customers and restaurants as nodes.

Datasets:
https://www.kaggle.com/yelp-dataset/yelp-dataset

Idea 3: Movie Networks

There are many datasets available on movies. These datasets include lists of actors/actresses, metadata about the movie, images of movie posters and even databases of movie reviews. I think some of this data could be combined to answer some interesting questions:

1) Can we classify a review? Sentiment analysis could be conducted on reviews to attempt to classify them as positive or negative.
2) Using the database of user reviews (from movielens), a recommender system could be built which profiles similar users in order to recommend new movies (using collaborative and content based filtering).
3) A network analysis could be conducted in which nodes are actors and movies or movie genres (bipartite graph). Are certain actors more likely to appear in movies together? Could you classify genre based on the actors in the movie?

Datasets:
https://grouplens.org/datasets/movielens/
https://datasets.imdbws.com/