

Capstone Project 1 Ideas:

Idea 1: Predict popularity based on Yelp reviews/pictures/reviewerNetworks.

Use Yelp data to predict future popularity of an establishment (restaurant/bar/etc.). My adviser says I would have to define a metric that measures popularity. I'm still struggling on figuring out what realistic impact this analysis would have on a client. Perhaps inform the marketing strategy for establishments?

<https://www.yelp.com/dataset/challenge>

Idea 2: Predict Web Traffic to Wikipedia Articles.

Use article features to predict web traffic to wikipedia articles. This analysis could help clients figure out what drives traffic to certain websites.

<https://www.kaggle.com/c/web-traffic-time-series-forecasting>

Idea 3. Predict NYC Taxi Trip Duration.

Predict the trip duration of New York City taxi rides. I'm interested in predicting commuter times, and I think Kaggle has a nice clean data set for this. A lot of users have also submitted supplemental data sets. I am a little concerned about whether I can add anything new to this problem.

<https://www.kaggle.com/c/nyc-taxi-trip-duration>