# DS 740 Final Project Proposal by Matt Allen

The data set that will be used for the final project is the NYC Open Data. It is available on Kaggle at the URL: https://www.kaggle.com/nycopendata/new-york. In particular, the datasets related to NYC's Citi Bike system will be used. There are two files one that documents Citi Bike stations and one that has 30 million Citi Bike trips 2013-present. The data will be used to predict gender based on Citi Bike trips. Two methods to be used in the analysis are linear discriminant analysis, and logistic regression. The model may be used by Citi Bike to understand their customers better. Are there differences in how women and men use the Citi Bike system?