Framework for analyzing blight ticket payment likelihood.

As usual we want to create a few models and compare their performance against one another.

Q: Should I be creating a train test split when I’ve already been provided a training and a test data set?

I think not. I feel like I should be simply using y\_train and X\_train to train my model. Then create predictions using X\_train. And evaluate it against y\_train to obtain a score of 75% and above, before using it on X\_test to create y\_trest\_pred.

Steps for this assignment:

1. ~~Read in training data~~
2. ~~Split training data into target variable (y\_train) and data (X\_train)~~
3. ~~Merge address data~~
4. ~~Merge latlon data~~
5. Subset X\_train to all predictor variables you think will be relevant, numeric or categorical
6. Then fit your first model. Start with logistic. Create logistic model.
   1. clf = fit(y\_train, X\_train)
   2. y\_train\_pred = clf.predict(X\_train)
   3. evaluate: score(y\_train, y\_train\_pred) using AUC
7. Use additional scoring techniques of measurement and simulation where necessary. And refer to past exercises for support.