

## 1 Goal of the Project

Predict the Permit Type in advance, so that business owners can have an oracle of where the optimal location is in terms of spatial availability, surrounding business types, and building conditions.

## 2 Model Test Accuracy

- Logistic Regression Model: 74.09%
- Fully Connected Neural Network Model: 71.45%
- Recurrent Neural Network Model: 75.54%

## 3 Model Selection Rationale

- We select logistic regression because it's easy to explain, robust, and it could also be a benchmark to compare non-deep learning model with the deep learning model.
- We select the Fully Connected neural Network Model because it's the simplest deep learning model, thus, good to be a benchmark to compare with other complex neural network structure.
- We select RNN because data is time related. The government controls the approval for type of permits. Thus, if one type is approved so much early, it might be more difficult to get the same type passed in the future. Also, the building resources in the NYC is limited, if a location has been heavily utilized for a type of business earlier, then less other permits will be approved for that location and business types in a short period of time.

## 4 Feature Importance Analysis

We randomly shuffle the values of each feature and test our model on the test set to determine which feature is the most important. It turns out that most important features for the Fully Connected Neural Network Model is the "Work Type".