Binary classification:

* TO predict the severity of the crime binary classification is used.

* The required features for normalisation is latitude, longtitude, closest-station.

* Other features for the dataset are all categorical data & have all been converted into dummy variable. (District, tym-blk, weekday, location descriptions).

* we performed split on the 80;000 sample records, into training set & test set and normalised by themselves.

* Latitude & longtitude info more helpful for classification, If the latitude & longtitude is low, It's more likely that severe crime will happen.

* It is consistent wid the fact that the safety at south chigo is notoriously bad.

* The economic, unemployment rate, age states do provide the expected results. I.e. For regions with lower income & higher unemployment rate, the crimes are going to be more severe.

* As a result if the tym blk is 3 am to 6 am, the proportion of severe crime is obviously higher, which is not the case in tym blk 9 am to 12 pm.

* The crime takes place in an appaw apartment or house, it's more likely to be a severe crime. The crime takes place on a street where everyone can see, it's less likely going to be severe.

* Severe crime accounts for 46% of total crime and non-severe is around 53%.

* Built different classifiers using the training set, & examined the accuracy of all the classifiers using the test set.

* Based on the accuracy comparison of diff models, Baseline model have least accuracy of 54%.

* If classification accuracy is the only criteria to judge whether a model is good or not, a good model have greater than the baseline model.

* Logg compared all of the model logestic regression have the accuracy of 61%.

severe

Arson, Assault, battery, sexual assault, criminal damage, criminal trespass, Homicide, robbery.