Progress report: Utilizing ML to analyze speeding in NYC via NYC traffic dataset

# Project background:

It’s a trouble for government allocating limited resources into the correct location under the right timing for their work. However, we can use Machine Learning technique to analyze the location and timing that most of vehicles are speeding, to have better understanding about the speeding in NYC.

All the data preprocessing work will be done by Python and SQLite and the machine learning work will be done by Matlab.

# Original datasets:

City of New York Department of Transportation provides two datasets. The table below shows the details of it.

Dataset 1: Sensors information

|  |  |  |
| --- | --- | --- |
| linkid | The Unique id of the set sensors | 4616337 |
| linkpoints | The coordinates of each sensor | 40.74,-74.00 …… 41.740,-74.0 |
| EncodedPolyLine | The encoded value of linkpoints | y{swFvavbMjANlGSvQn@fa@f |
| EncodedPolyLineLvls | The encoded value of linkpoints | BBBBBB |
| Transcom\_id | Same as linkid | 4616337 |
| Borough | The location of the sensors | Manhattan |
| linkName | The address of the set sensors | 12th Ave S 57th St - 45th St |
| Owner | Owner who owns sensors | NYC\_DOT\_LIC |

\*We can also get the coordinates of the sensors by decoding EncodedPolyLine and EncodedPolyLineLvls via Google Maps API. The coordinates we get from the api will be the same as the coordinates in linkpoints.

Dataset 2: Traffic records

|  |  |  |
| --- | --- | --- |
| id | The vehicle id | 453 |
| Speed | The speed of the car detected | 34.18 |
| TravelTime | The average time of vehicles passing sensors | 190 |
| date | Last time data was received from sensors | 4/20/2016 2:19:42 PM |
| linkid | The id of the set of sensors | 4456502 |

# Data-Preprocessing

We need to use HERE API to get the speed limits for each coordinate and take a average of it representing the speed limit of a set of sensors. The code below my work described above.

# Project Plan

|  |  |
| --- | --- |
|  | Finish all data preprocessing work |
|  | Use cluster technique |
|  | Use regression (relationship between speeding and time) |
|  | Write final report |