# HOMEWORK 3-B

# Step 1:

**Topic for tweet collection:**

**Crime related**

**Geographic region:**

**New York City**

# Step 2:

Data collection strategy

Tweets and blogs talking about crimes and safety related topics

Historical crime records

Focus on:

Tweets and/or blogs talking about crimes

Motivation:

I have chosen the crime related topic for this homework, I wanted to fetch the tweets which are related to crime in the city of New York.

Any police department or a person ,Will tweet about a crime whenever he reads or even gets affected by a certain crime. The tweet will include the information about a crime or even will have a picture of the crime/crime location.

## QUERIES:

THEFT

RAPE

CORRUPTION

VIOLENCE

CRIME

TERROR

FELONY

MURDER

ASSAULT

ATTACK

ROBBERY

HOMICIDE

BURGLARY

KIDNAPPING

Reason for using these queries:

As we always read in any articles or even in newspapers, Any crime related topic will have the above keywords.

I went to twitter, And filtered several tweets from New York Police Departments.

It had tweets with the above keywords.

## Step 3:

By using the REST API, I collected 500 tweets related to my queries and from that I found 259 tweets related to topic and 242 which were not. I made a text file named “500-tweets” which I got from the API and another file named “500-labels” in which I manually label each tweet as positive and negative.

## Step 4:

Random Sample of Data:

I used REST API for collecting the random sample. I used the same location and fetched 1000 random tweets which were not related to crime or any other to my keywords.