MSA 8150 Project: **Metropolis Crimes**

Project Type

This problem is considered a **large** problem.

Problem Setup

It is around the year 2044 in the city of Metropolis, where Superman is in charge of helping the people of the city. Due to the population growth, Superman cannot keep track of every crime in the city and has finally come to the conclusion that maybe reporting the incidents to the Metropolis police department (MPD) would be a better idea than flying around!

The MPD has published a record of the crimes over a period of time. Based on this dataset, they are willing to predict the type of crime once they receive a call from Superman. This way, they would know ahead of time what kind of incident they should be expecting and what backup resources they may need.

You are asked to help the MPD develop such predictive models. The input features of the given data file are:

- `incident_datetime:` This is a variable indicating the time and the date that an incident happened.
- `incident_cord_x:` This is the x-coordinate of the location that the incident happened.

- `incident_cord_y:` This is the y-coordinate of the location that the incident happened.
- `num_victims:` This is a variable indicating the number of victims in the incident.
- `location_type:` A variable that indicates the type of location the incident happened, location types are identified by numbers and may correspond to different points such as residential, roads, public gardens, etc. If for a crime the location type has not been reported, **the `location_type` for that incident takes a value of zero**.

The response variable is:

- `Crime_Type`: This is a variable that indicates the type of the crime that took place. This is a categorical variable and takes one of the following values:
  - "AGG ASSAULT"
  - "AUTO THEFT"
  - "BURGLARY-NONRES"
  - "BURGLARY-RESIDENCE"
  - "HOMICIDE"
  - "LARCENY-FROM VEHICLE"
  - "LARCENY-NON VEHICLE"
  - "RAPE"
  - "ROBBERY-COMMERCIAL"
  - "ROBBERY-PEDESTRIAN"
  - "ROBBERY-RESIDENCE"

Modeling Instructions

You would need to use the file `Train.csv` to fit your models. Your model can take `incident_datetime, incident_cord_x, incident_cord_y, num_victims,` and `location_type` as inputs and should be able to predict the class of crime (i.e., crime type), as one of those listed above.

Please make sure to communicate with the instructor and Piazza about any potential ambiguities related to the data.