**NYPD Project Proposal**

**Problem statement**: In this project, I will be looking at a dataset of motor vehicle collisions in New York City over the last three years. The dataset is hosted on the NYC Open Data website and includes all vehicle collisions recorded by the NYPD since July 2012. This is a live dataset that is updated every weekday with recent accident data. Over 50,000 accidents are recorded in the set, along with details relevant to the accidents: time, data, vehicle type, possible contributing factors, injuries, fatalities, and the precise coordinates of each incident.

**Hypothesis**: My hypothesis is that the most accidents in NYC occur at difficult intersections and congestion points, and not necessarily as a result of speeding.

**Questions to answer**:

1. Top ten causes of injuries in the traffic accidents.
2. Total number of vehicle collisions and daily frequency.
3. Pattern of accidents over 1 day period and extend it to a 7 day period and 30 days period.
4. Which boroughs, year and months were most motorists killed so that they can be more careful and drive prudently. Extend this to pedestrians.
5. Average time for an accident to occur and the average time during the way when it has highest change of occurring.
6. What type of vehicles causes most accident.
7. Create a model to predict the number of accidents that will occur.

**Client**: NYPD and City of New York. I am hoping my analysis will help the city of New York to implement additional measures in place to stop accidents and help NYPD predict the most dangerous areas for accidents so that actions can be taken to avoid such accidents from occurring.

**Data**:https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95

**Deliverables**: Jupyter notebook with all the code and github link, A project report and a presentation.