# ***Capstone Proposal***

Name : SIDALI BENABDI

Cohort : DAB July 2025

**Project Information:**

The dataset contains information from all police reported motor vehicle collisions in NYC.

Some of the real-world applications my analysis could provide is helping policy and decision makers identify then prioritize areas where infrastructure redesign is needed, focusing the efforts of law enforcement and advocacy groups towards the most lethal driver behavior and provide clear evidence on which road users are affected the most by fatal crashes, paving the way for new legislation like speed limit reductions or increased fines.

**Project Goals:**

With this analysis I am trying to pinpoint high risk locations and time of day and identify historical changes in the kill/serious injury rate (KSI). Identify the 3 to 5 most frequent contributing factors associated with fatal or serious crashes. Link the specific conditions and relative risk under which the most vulnerable road users (cyclists and pedestrians) are injured or killed to ensure their safety.

**Data Overview:**

This dataset consists of one table in a csv format that has 29 columns and above 1 million rows. Each row represents a crash incident. The table handles accidents dating all the way back to the year 2017.

Key variables include geographic location like street name and location, date, time, contributing factor, counts of person, pedestrian and cyclists injured or killed.

Data quality considerations are the sheer size of the data set (not everything will be used , that will depend mainly on the chosen time period for analysis , most probably post covid, and the number of columns, some of them will be combined for ease of analysis) also potential text inconsistencies (e.g., street names) and the subjective nature of the Contributing Factor categorization by the reporting officer. Missing values, field formatting and the standard issues with publicly available data.