**Project Title: An Analysis of the Relationship between Weather and Environmental factors with Motor Vehicle Accidents**

* **Group Members:**
* 220950125041 : JEMIMA JOAQUINA DIAS
* 220950125042 : KAJAL PATEL
* 220950125043 : KAMAL PRAKASH YADAV
* 220950125053 : NAINESH ANIL KHANJIRE
* 220950125054 : NALIN PUSHP
* **Team Lead :** JEMIMA JOAQUINA DIAS
* **Project Supervisor :** ABHISHEK RAJU CHAVAN

Traffic crashes have a significant impact on the world economy and are a leading cause of death and injuries worldwide. More than one-half of all traffic deaths globally occur among people ages 15 to 44, their most productive earning years. This project is developed to understand feature correlations, patterns and analyze trends with the current weather conditions in the area of the accident. To develop this system end to end, first an ETL (Extract Transform Load) pipeline is setup to continuously ingest data from the NYC vehicle crash data set and web scrape weather data with a daily update. The data set contains about 2 million records.

The next step involves transforming the data set by cleaning and filling missing values of zip code, latitude, longitude etc. so that we have comprehensive data about each incident. We also scraped weather data from wunderground for each weather station in New York and this data is then processed and mapped with the crash data set. The workflow pipeline is orchestrated using an Airflow DAG, that runs on a daily schedule. This data is used for visualization and prediction analysis using machine learning models such logistic regression, SVM and Gradient Boosting, to predict the severity of a motor vehicle accident based on 4 classes such as very high, high, medium, low centered on weather and environmental factors such as temperature, precipitation rate, dew, wind speed and high-risk areas that can warn user based on historic data.