

Name: Shahad Alameel

Email: Shahad.a@outlook.com

Github: ShahadAlameel

Project Proposal

Background:

The goal of this project is to work with New York Police Department to try to reduce the number of car accidents in New York City by finding how can metros impact the number of car accidents. By finding the relation between the rate of car accidents and the traffic of the metro stations, we could help encourage more people to use the metro rather than using cars. Which will lower the rate of accidents and cut the costs and efforts of NYPD in handling the accidents.

Data Description:

1. MTA Dataset:

The New York subway MTA turnstile data is a series of data files containing cumulative number of entries and exits by station, turnstile, date and time. Data files are produced weekly. Data records are collected typically every 4 hours with some exceptions.

Sample size is 3 months (from May to July 2019) to find the relation between car accidents and metro traffic without the impact of covid-19.

The dataset has 2480677 rows and 11 columns. Column names are:

- **C/A** = Control Area (e.g., A002)
- **unit** = Remote Unit for a station (e.g., R051)
- **SCP** = Subunit Channel Position represents an specific address for a device (e.g., 02-00-00)
- **station_code** = C/A + unit, locating a station
- **turnstile** = C/A + unit + SCP, locating a turnstile
- **Station** = Represents the station name the device is located at
- **date** = Represents the date (MM-DD-YY)
- **time** = Represents the time (hh:mm:ss) for a scheduled audit event
- **datetime** = date + time (MM-DD-YY hh:mm:ss)
- **desc** = Represent the "REGULAR" scheduled audit event (Normally occurs every 4 hours)
- **entries** = The cumulative entry register value for a device

- **exits** = The cumulative exit register value for a device

2. The Motor Vehicle Collisions Crash Dataset:

The Motor Vehicle Collisions crash table contains details on the crash event. Each row represents a crash event. The Motor Vehicle Collisions data tables contain information from all police reported motor vehicle collisions in NYC. The police report (MV104-AN) is required to be filled out for collisions where someone is injured or killed, or where there is at least \$1000 worth of damage.

Sample size is 3 months (from May to July 2019) to find the relation between car accidents and metro traffic without the impact of covid-19.

The dataset has 2480677 rows and 29 columns. Column names are:

- **CRASH DATE:** Occurrence date of collision
- **CRASH TIME:** Occurrence time of collision
- **BOROUGH:** Borough where collision occurred
- **ZIP CODE:** Postal code of incident occurrence
- **LATITUDE:** Latitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)
- **LONGITUDE:** Longitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)
- **LOCATION:** Latitude, Longitude pair
- **ON STREET NAME:** Street on which the collision occurred
- **CROSS STREET NAME:** Nearest cross street to the collision
- **OFF STREET NAME:** Street address if known
- **NUMBER OF PERSONS INJURED** : Number of persons injured
- **NUMBER OF PERSONS KILLED:** Number of persons killed
- **NUMBER OF PEDESTRIANS INJURED:** Number of pedestrians injured
- **NUMBER OF PEDESTRIANS KILLED:** Number of pedestrians killed
- **NUMBER OF CYCLIST INJURED:** Number of cyclists injured
- **NUMBER OF CYCLIST KILLED:** Number of cyclists killed
- **NUMBER OF MOTORIST INJURED:** Number of vehicle occupants injured
- **NUMBER OF MOTORIST KILLED:** Number of vehicle occupants killed
- **CONTRIBUTING FACTOR VEHICLE 1:** Factors contributing to the collision for designated vehicle
- **CONTRIBUTING FACTOR VEHICLE 2:** Factors contributing to the collision for designated vehicle
- **CONTRIBUTING FACTOR VEHICLE 3:** Factors contributing to the collision for designated vehicle
- **CONTRIBUTING FACTOR VEHICLE 4:** Factors contributing to the collision for designated vehicle
- **CONTRIBUTING FACTOR VEHICLE 5:** Factors contributing to the collision for designated vehicle

- **COLLISION_ID:** Unique record code generated by system. Primary Key for Crash table.
- **VEHICLE TYPE CODE 1:** Type of vehicle based on the selected vehicle category (ATV, bicycle, car/suv, ebike, scooter, truck/bus, motorcycle, other)
- **VEHICLE TYPE CODE 2:** Type of vehicle based on the selected vehicle category (ATV, bicycle, car/suv, ebike, scooter, truck/bus, motorcycle, other)
- **VEHICLE TYPE CODE 3:** Type of vehicle based on the selected vehicle category (ATV, bicycle, car/suv, ebike, scooter, truck/bus, motorcycle, other)
- **VEHICLE TYPE CODE 4:** Type of vehicle based on the selected vehicle category (ATV, bicycle, car/suv, ebike, scooter, truck/bus, motorcycle, other)
- **VEHICLE TYPE CODE 5:** Type of vehicle based on the selected vehicle category (ATV, bicycle, car/suv, ebike, scooter, truck/bus, motorcycle, other)

Tools:

- Technologies: SQL, SQLite, python, Jupyter notebook
- Libraries: Numpy, Pandas, Pandas profiling, Matplotlib, Seaborn, Sklearn