**BI ASSIGNMENT 3**

**GROUP MEMBERS:**

* Muhammad Sumair 22995
* Khansa Junaid 23074

**ABOUT THE DATASET:**

The dataset we have chosen from Kaggle is NYC Traffic accidents <https://www.kaggle.com/datasets/melodyyiphoiching/nyc-traffic-accidents>.

Accidents involving vehicles that were reported to the New York City Police Department between January and August 2020. Each record corresponds to a specific collision and contains information about the involved vehicles, victims, and contributing variables as well as the date, time, and location of the accident (town, zip code, street name, latitude/longitude).

**DATA CLEANING & PREPERATION:**

To prepare the chosen dataset for further analysis we will first have to clean and prepare the dataset. For cleaning dataset, we have used python and power bi.

There are many anomalies in the dataset like:

• Null records

• Date format

• Duplicate records

• Mismatched column

Data cleaning is performed in the following steps:

* Null Records

Depending on the category, consider the percentage of null records. Some of the dataset's null entries have been removed and replaced with appropriate values. The columns which have missing values more than 70% were removed like VEHICLE TYPE CODE 3, VEHICLE TYPE CODE 4, VEHICLE TYPE CODE 5, CONTRIBUTING FACTOR VEHICLE 3, CONTRIBUTING FACTOR VEHICLE 4, CONTRIBUTING FACTOR VEHICLE 5 and OFF-STREET NAME. Also, columns like BOROUGH’s missing values were replaced with mode value and ZIPCODE was grouped with BOROUGH and missing values were replaced. The remaining columns were also treated respectively.

* Date Format

Since information was gathered from a variety of sources, the date format was inconsistent. Data must be organized logically for the study, so all date formats were changed to "MM/DD/YYYY." In Power BI, the calculated field was made to clear up the date column.

* Duplicate Records

The dataset contained many duplicate records, which were eliminated to increase accuracy.

* Mismatched Column

When data is imported into Power BI, the data type of each column is automatically determined; however, Power BI was unable to locate some of the columns in our dataset. These column datatypes were fixed as part of the data cleaning procedure.

After all these steps were performed the dataset was fully cleaned and ready to be used for further analysis and visualization.

**POWER BI DASHBOARDS:**

1. **PROBLEM STATEMENT:** RISKY LOCATIONS IN NYC

**About the Problem Statement:** Identifying and analyzing the locations where there are the greatest number of accidents. To dig deep into the problem, we are also identifying the time and the month when the accidents take place to provide better and accurate solutions.

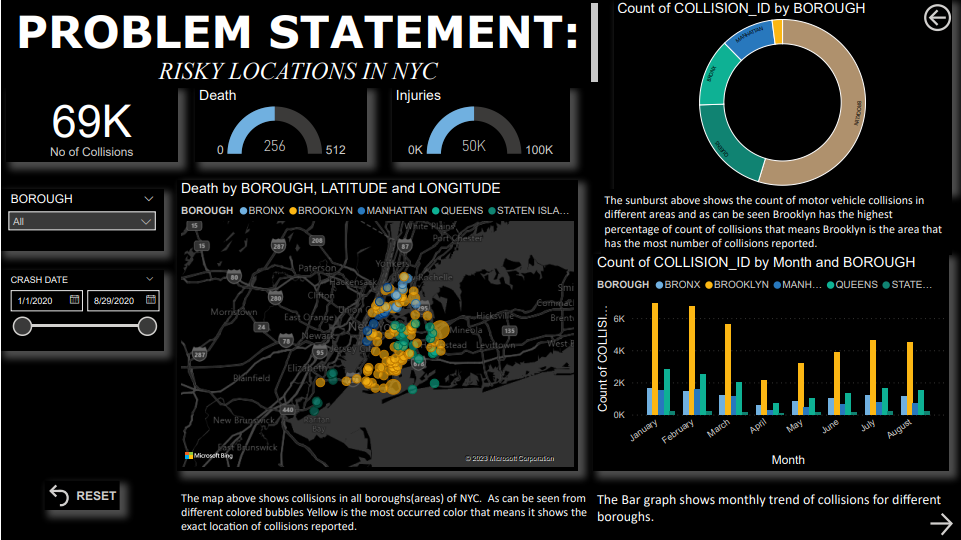
**Solutions:** We can observe from our analysis that most of the accidents are occurring in Brooklyn and that too at night, so a simple thing that can help avoid this is to deploy more streetlights so that the visibility on the roads is much better and any bumps or sharp turns are easily visible to the drivers.

Secondly, Brooklyn should have greater security or check posts so that people avoid over speeding, breaking the laws like not following the signals and drunk driving. If there is higher security on the roads people will avoid doing these things to refrain from paying heavy fines.

Moreover, there should be awareness programs by the government on the causes of the accidents, injuries, and deaths so that people know how dangerous problem this is, and they try to avoid this, and they take this seriously.

Furthermore, the US government should take different measures to limit the number of vehicles on the road, this can be done by implementing new projects for public transportation like bullet trains or more metros in that area so that chances of road accidents are minimized, especially on Brooklyn.

**Dashboard:**



1. **PROBLEM STATEMENT:** FACTORS CONTRIBUTING MOST TO MOTOR VEHICLE COLLISIONS

**About the Problem Statement:** Identifying the causes of the accidents or analyzing why the accidents are happening. First, we figured that the greatest number of accidents are taking place in Brooklyn, and now in this problem statement we are looking for the factors that are causing the accidents in NYC and more specifically in Brooklyn.

**Solutions:** We can observe from our analysis that the most prominent factor that contributes to the accidents is the inattention or distraction of the drivers.

Firstly, the government should spread the awareness amongst the people regarding the road sense and presence of mind while driving, simultaneously government should share the dangerous outcomes, this can lead to if the driver carelessly or irresponsibly drives on the roads like severe injuries and deaths.

Secondly, the government should make licensing a little tough and there should be more tests on the presence of mind of the driver. More strictness should be shown by the government officials, in case of lower presence of mind of the driver, the license should not be issued. If it will be hard to get the license, then people will take this matter more seriously.

Moreover, if a person is involved in an accident, a proper investigation should be run against the person and the license should be suspended for a little time so that the driver can get a lesson out of it.

Furthermore, we see that inattention of the drivers lead to majority of the accidents, it can be possible that underage people are driving and over speeding on the roads which lead to the accidents, so there should be a little tight security to check whether the driver is underage or not and whether the person has the license or not, because it is a possibility that under aged drivers have less attention span and hence low attention on the road while driving, for example youngsters are involved in using mobile phones will driving etc, which can lead to accidents. Hence, tighter security and heavy fines can help NYC minimize the accidents.

**Dashboard:**

