Motor Vehicle Collisions

SQL Project

NOVEMBER 13, 2023

OKOLI NNAMDI NNAEMEKA
09067260556
Danielokoli448@gmail.com

Professional Background

With over three years of vrsatile experience in both office and freelance settings, I have honed my skills as a data analyst. My impactful tenure at Guinness Nigeria was marked by the optimization of supply chain operations through strategic data analysis, resulting in a notable reduction in processing time.

As a National Youth Service Corps (NYSC) graduate, I served as a Data Entry personnel, utilizing Microsoft Excel to navigate extensive datasets and presenting insights through clear Excel tables and PowerPoint presentations. Transitioning to the role of IT Assistant at Delta Broadcasting Service (DBS) Warri, I further strengthened my data management skills over six months.

My commitment to continuous learning is evident in my proficiency in SQL, MS Excel, and Python. Additionally, I bring expertise in visualization tools such as Tableau and Power BI, with ongoing efforts to enhance my skills in R language. Practical applications include the creation of interactive dashboards using Power BI and experimentation with R language in controlled environments.

Notably, my analytical approach focuses on tangible outcomes and quantifiable achievements. At Guinness Nigeria, my data-driven strategies had a direct and positive impact on supply chain efficiency. Eager to contribute this expertise to a remote data analyst role, I am poised to deliver results through a combination of technical proficiency and strategic problem-solving.

Table of Contents

Introduction

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 (New York City). 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. MV104-AN is a form used to report accidents involving motor vehicles.

The main problem is to decrease the rate of motor vehicle collisions in the city.

I used the dataset Motor_Vehicle_Collisions to analyze and answer this problem.

I applied SQL commands to analyse data: JOIN, ORDER BY, AS, WHERE, AND, OR, SUM(), COUNT(), GROUP BY, HAVING.

Also, I used Root Cause Analysis to understand the problem and ask the right questions.

As a result, I have found out crucial insights of the provided dataset, prepared visualisations, and report for my team.

Root Cause Analysis

The problem I want to analyze in this dataset is to how to decrease the rate of motor vehicle collisions by first analyzing the reasons for these crashes, to properly execute this I used root cause analysis to find the root or roots of this problem.

To understand the problem, I need to analyze the existing database of vehicle crashes in the city of New York. I will also present some crucial numbers and visualization of the datasets. So, I decided to ask some questions.

- How many recorded Motor Vehicle Collisions do we currently have in our database?
- What timeframe does this dataset cover?
- How many people lost their lives as a result of vehicle collisions in this dataset?
- How many people got injured as a result of vehicle collisions in this dataset?
- Which borough has the least Motor vehicle collision incidents?

Also, I decided to apply Root Cause Analysis to the problem to figure out the underlying issues in order to identify appropriate solutions.

1. Why is this motor_vehicle_collisions dataset needed?
   Ans. It shows the number of recorded people who are killed or injured during vehicle collisions in New York.

2. Why are there motor vehicle crashes?
   Ans. Driver inattention/ distraction is the biggest cause of motor vehicle collisions. It easily doubles the next known cause following it.

3. Why do motor vehicle drivers not pay attention while driving?
   Ans. These collisions occur more during the day than at night so it could be they tend to pay less attention then.

Insights from the Analysis

I made use of a relational database Motor_Vehicle_Collisions to answer the problem.

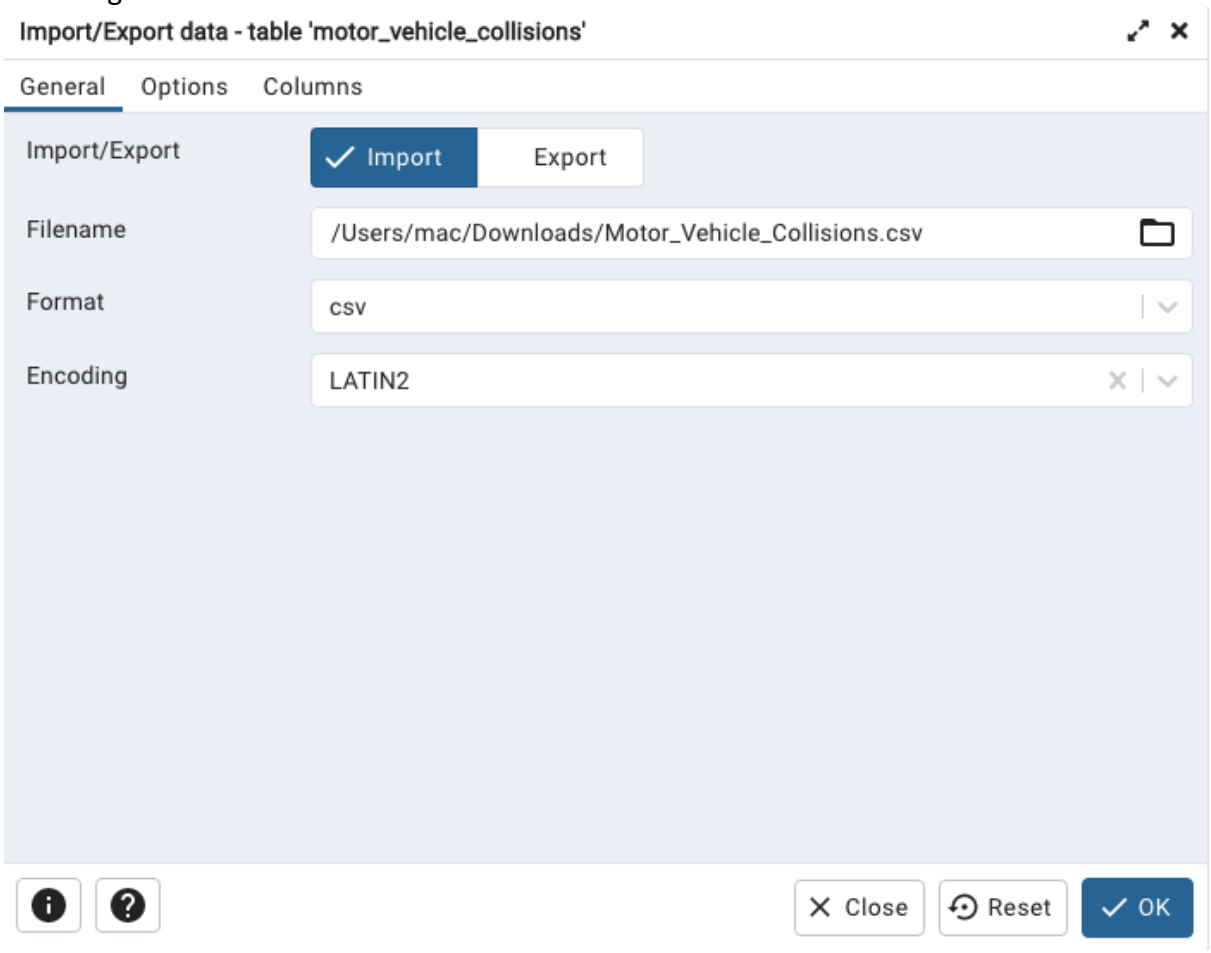
POSTGRESQl Database Management System was used to find out main insights.

Motor_Vehicle_Collisions includes such data:
- CRASH DATE
- CRASH TIME
- BOROUGH
- ZIP CODE
- LATITUDE
- LONGITUDE
- ON STREET NAME
- CROSS STREET NAME
- OFF STREET NAME
- NUMBER OF PERSONS INJURED
- NUMBER OF PERSONS KILLED
- CONTRIBUTING FACTOR VEHICLE 1
- CONTRIBUTING FACTOR VEHICLE 2
- COLLISION_ID
- VEHICLE TYPE CODE 1

A Motor_Vehicle_Collisions table was created using Postgresql.

```
CREATE TABLE motor_vehicle_collisions(
    crash_date DATE,
    crash_time TIME,
    borough TEXT,
    zip_code INTEGER,
    latitude REAL,
    longitude REAL,
    ON_STREET_NAME VARCHAR(100),
    CROSS_STREET_NAME VARCHAR(100),
    OFF_STREET_NAME VARCHAR(100),
    NUMBER_OF_PERSONS_INJURED INTEGER,
    NUMBER_OF_PERSONS_KILLED INTEGER,
    CONTRIBUTING_FACTOR_VEHICLE_1 VARCHAR(100),
    CONTRIBUTING_FACTOR_VEHICLE_2 VARCHAR(100),
    COLLISION_ID SERIAL PRIMARY KEY,
    VEHICLE_TYPE_CODE VARCHAR(100)
);
```

The Motor_Vehicle_Collisions was then imported into the created table with the "LATIN2" Encoding



SELECT statement was used to fetch data from a database.

```
48  SELECT * FROM motor_vehicle_collisions;
```

To find the recorded number of crimes with the COUNT() Function, I used a command like this:

```
57  SELECT COUNT(crash_date)
58  FROM motor_vehicle_collisions;
```

To get the timeframe of the dataset, I used the MIN() and MAX():

```
50  SELECT MAX(crash_date)
51  FROM motor_vehicle_collisions;
52
53  SELECT MIN(crash_date)
54  FROM motor_vehicle_collisions;
```

To find the recorded number of people who died in these vehicle crashes with the SUM() Function, I used a command like this:

```
68  SElECT SUM(number_of_persons_killed)
69  FROM motor_vehicle_collisions;
```

To find the recorded number of people who injured in these vehicle crashes with the SUM() Function, I used a command like this:

```
74  SElECT SUM(number_of_persons_injured)
75  FROM motor_vehicle_collisions;
```

I used the GROUP BY function to know the total number of vehicle collisions in each borough in NYC

```
113 SELECT borough, COUNT(*) AS vehicle collisions, SUM(number_of_persons_killed) AS DIED, SUM(number_of_persons_injured) AS INJURED
114 FROM motor_vehicle_collisions
115 GROUP BY borough
116 ORDER BY COUNT(*) DESC;
```

I used the GROUP BY function to know the top 10 contributing factors of vehicle collisions in NYC

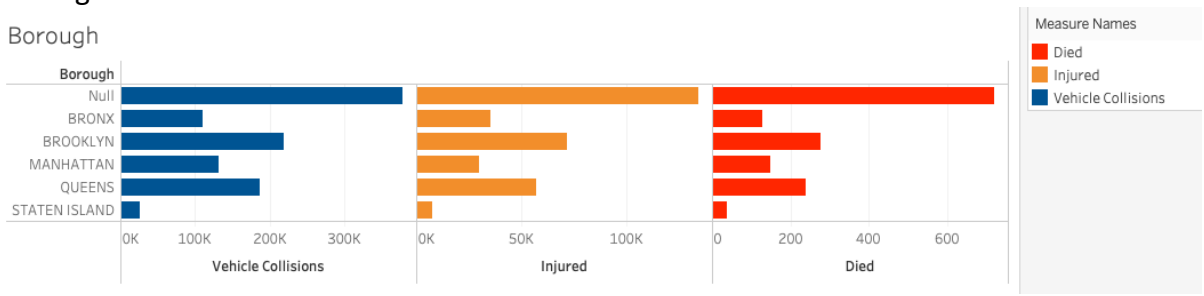
```
125 SELECT contributing_factor_vehicle_1, COUNT(*) AS vehicle collisions, SUM(number_of_persons_killed) AS DIED, SUM(number_of_persons_injured) AS INJURED
126 FROM motor_vehicle_collisions
127 GROUP BY contributing_factor_vehicle_1
128 ORDER BY COUNT(*) DESC
129 LIMIT 10;
```

I used the COUNT function to find the number of vehicle collisions across 6 hour periods of the day

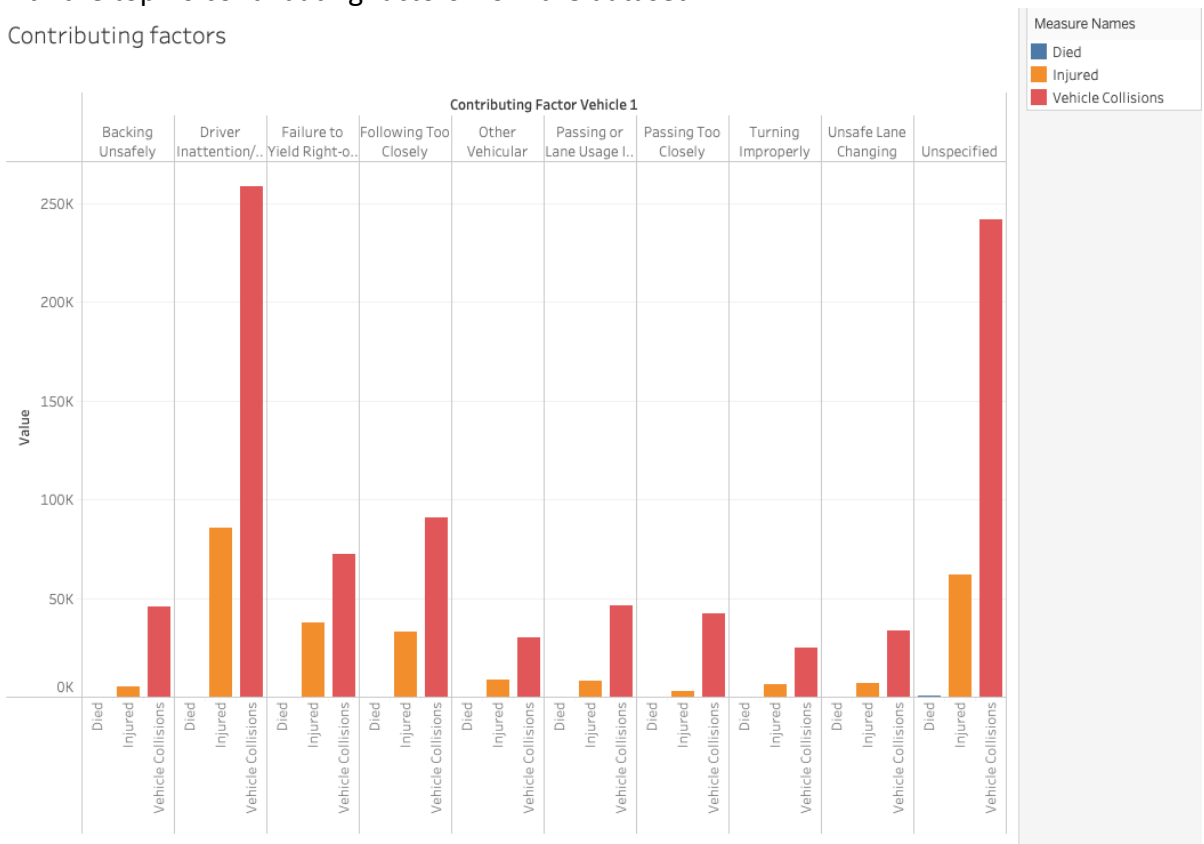
```
84  SELECT COUNT(crash_time)
85  FROM motor_vehicle_collisions
86  WHERE crash_time BETWEEN '00:00:00' AND '05:59:00'
87  AND contributing_factor_vehicle_1 = 'Driver Inattention/Distraction';
88
89  SELECT COUNT(crash_time)
90  FROM motor_vehicle_collisions
91  WHERE crash_time BETWEEN '06:00:00' AND '11:59:00'
92  AND contributing_factor_vehicle_1 = 'Driver Inattention/Distraction';
93
94  SELECT COUNT(crash_time)
95  FROM motor_vehicle_collisions
96  WHERE crash_time BETWEEN '12:00:00' AND '17:59:00'
97  AND contributing_factor_vehicle_1 = 'Driver Inattention/Distraction';
98
99
100 SELECT COUNT(crash_time)
101 FROM motor_vehicle_collisions
102 WHERE crash_time >= '18:00:00'
103 AND crash_time <= '23:59:00'
104 AND contributing_factor_vehicle_1 = 'Driver Inattention/Distraction';
```

Tableaus was used as a very powerful tool for data analysis and Visualization for better understanding of the data.

This Tableau clearly shows the total number of vehicle collisions, injuries and deaths in each borough from the dataset.



This visualization shows the total number of vehicle collisions, injuries and deaths associated with the top 10 contributing factors from the dataset



Findings and Recommendations

Here are the results of the data set analysis:

- The total number of motor vehicle collisions we have in the database is 1,048,575.
- The dataset covers over 11 years motor vehicle collision incidents in NYC between 27/07/2012 and 21/09/2023.
- A total of 1551 people lost their lives as a result of motor vehicle collisions in the city.
- 336,499 individuals were injured due to these collisions.
- Staten Island is the borough with the lowest number of recorded vehicle collisions in New York City.

Record of total vehicle collisions, injuries and deaths across all boroughs in the city

borough	vehicle_collisions	died	injured
NULL	376665	719	134389
BROOKLYN	218586	276	71788
QUEENS	186289	238	57155
MANHATTAN	131637	150	29941
BRONX	110151	129	35527
STATEN ISLAND	25247	39	7699

We can see that Brooklyn has the highest number of known vehicle collisions, deaths and injuries while Staten Island has the lowest. Also, many of these collisions had no official borough recording so were left as null. Thus, a better record should be kept so the locations of these collisions are known.

Record of total vehicle collisions, injuries and deaths of contributing factors.

contributing_factor_vehicle_1	vehicle_collisions	died	injured
Driver Inattention/Distraction	258549	200	85677
Unspecified	241847	410	61874
Following Too Closely	91000	6	33135
Failure to Yield Right-of-Way	72183	141	37439
Passing or Lane Usage Improper	46498	22	8336
Backing Unsafely	45505	22	5385
Passing Too Closely	42256	1	3177
Unsafe Lane Changing	33483	15	6733
Other Vehicular	30199	16	8606
Turning Improperly	24661	14	6461

We can see that the majority of these collisions are caused by driver inattention/Distraction. Moreover, many collisions were left unspecified.

Thus, drivers should be well educated of the importance of paying attention while driving and laws should be put in place to hold faulters accountable when they don't.

Conclusion

I have analysed the dataset motor_vehicle_collisions to decrease the rate of motor vehicle collisions in the city and save lives and properties.

After a thorough analysis of the dataset, I came to these conclusions. I discovered that a lot of these vehicle collisions happen in the city and it results in injuries or worst, loss of life.

However, there are some crucial points that we need to count and decrease the rate of motor vehicle collisions in the city.

Incipiently, I found out that the main recorded cause of vehicle collisions in the city is Driver inattention/distraction. Drivers need to be more educated on this issue! A quick glance at the phone, a noisy passenger, an ill-mannered pet… This is all it takes to distract a driver and cause a crash ending lives or causing injuries. The law enforcement agencies should place strict rules on this so it can be stopped.

Moreover, I observed that many of these crashes occur in Brooklyn. There could be a variety of reasons for this, none that were filled in the dataset provided. More information should be put together for this.

In conclusion, our focus should be on curbing the main contributing factors of these vehicle collisions from the top boroughs they occur. Furthermore, drivers and other vehicle occupants should be properly educated on the importance of being focused while driving to reduce the occurrence of these collisions.