Report On:

**Project I: Python Programming and Data Analysis**

**Team 18**

**Fall 2022**



Report By:

Nikhil Das Karavatt

1002085391

Table of Contents

[Overview: 3](#_Toc114768387)

[File Descriptions: 3](#_Toc114768388)

[Division of Labour: 3](#_Toc114768389)

[Problems Encountered: 3](#_Toc114768390)

[Analysis: 4](#_Toc114768391)

[1. Analysis of number of collisions per year for four specific car manufactures which are BMW, Nissan, Toyota, and Suzuki. 4](#_Toc114768392)

[2. Analysis of number of collisions per month for four specific car manufactures which are BMW, Nissan, Toyota, and Suzuki. 5](#_Toc114768393)

[3. Analysis of different type of vehicles involved in crashes. 7](#_Toc114768394)

# Overview:

While analysing Motor Vehicle Collisions (MVC) data file which has details on each vehicle involved in the crash in New York from 2012 to present (10 years) daily, I was able to explore various domains, applications, and challenges related to Python and its libraries like pandas, NumPy, and matplotlib. From cleaning the data sets by removing Null rows and by slicing the data from September 2017 to August 2019 as per the parameter file to analysing the cleaned data has helped gain more confidence to code in python

# File Descriptions:

There are in total 7 files in which first four are the main files which has helped me with data visualization while other three was just for reference purpose. All files are in .csv format

|  |  |
| --- | --- |
| File Name | File Description |
| Motor\_Vehicle\_Collisions\_-\_Vehicles | The main raw file with motor vehicle collisions details for 10 years in New York |
| MVC | This csv contains two years of data that I had in my parameter to analyse |
| Vehicle\_make\_final | This csv has null free values in Vehicle\_make and also the name of manufacturers was standardised |
| Vehicle\_type\_final | The csv has null free values in Vehicle\_type and also some types of vehicles were merged into 1. |
| YEAR\_COUNT | Vehicle\_make\_final.csv was grouped by vehicle\_make and year which represented the count as per the grouping. |
| MONTH\_COUNT | Vehicle\_make\_final.csv was grouped by vehicle\_make, year, and month which represented the count as per the grouping. |
| Vehicle\_type\_count | Vehicle\_type\_final.csv was grouped by Vehicle\_type and represented count of the types of vehicles. |

# Division of Labour:

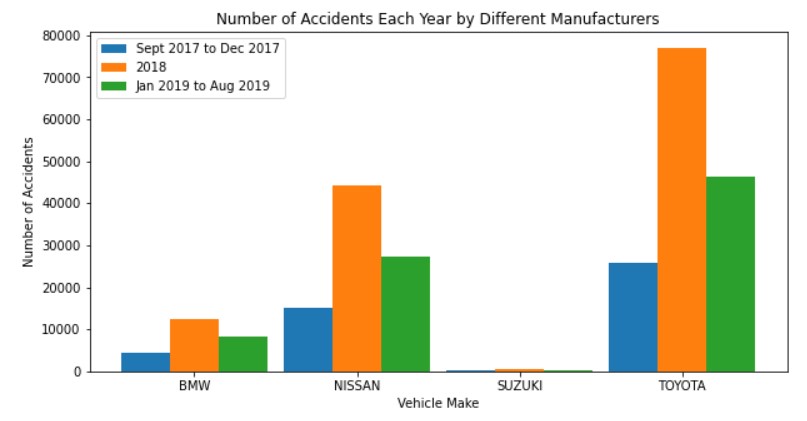
The project was done solely by me and it took me around 30 hours of work to finish it.

# Problems Encountered:

1. While cleaning the name and using the re function some vehicle make had space after the name for example “BMW “which was causing issue to my function to pick all “BMW” counts accurately but later fixed it by using replace function from “BMW “to “BMW”.
2. In pie chart I had many vehicle\_type which wasn’t clearly visible. I fixed it by using explode function.
3. I wasn’t able to do the indexing properly hence wasn’t getting the right values or else an error for my chart but I used loc function which helped me fix the error and understand indexing in an easier way.

# Analysis:

## Analysis of number of collisions per year for four specific car manufactures which are BMW, Nissan, Toyota, and Suzuki.

From the above graph, by the numbers Toyota has the largest count when it comes to number of accidents that took place in New York, USA, while Suzuki has the least. Also, the year 2018 shows the greatest number of accidents but we cannot truly compare this data with other years as for other years (2017 and 2019) we have counted the accidents for only few months unlike the year 2018 where all the months are involved.

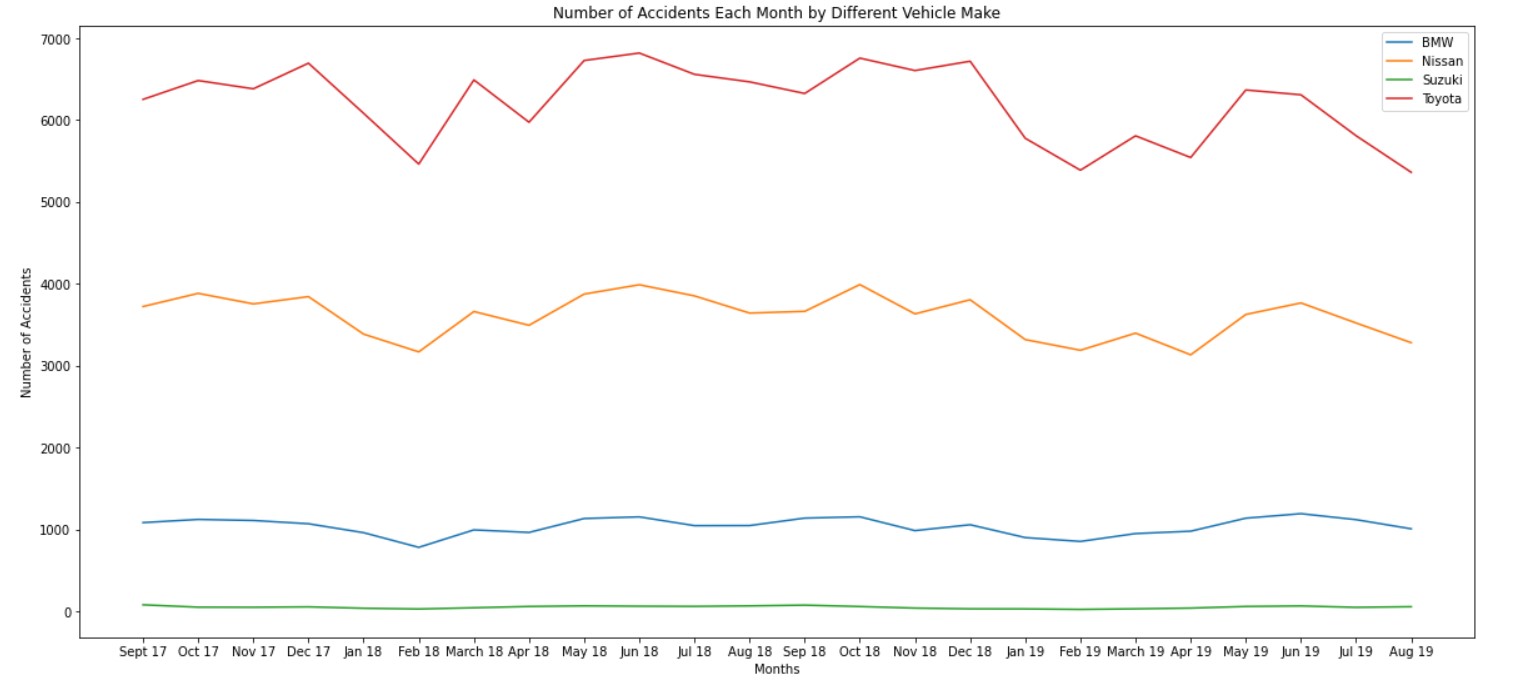
Can we conclude that does it make Toyota the least reliable car manufacturer and Suzuki the most? The answer is clearly no! Toyota has a huge market share in United States’ automobile industry. Rather recently in 2021 Toyota became [America’s top selling automaker](https://www.cnbc.com/2022/01/04/toyota-dethrones-gm-to-become-americas-top-selling-automaker-in-2021.html). In contrast, Suzuki has one of the least market shares in automobile industry in USA.

As per [CarSaleBase](https://carsalesbase.com/us-car-sales-analysis-2019-brands/), In 2019 Toyota sold 3 most cars in USA, followed by Nissan at 6th position and BMW in 10th position. Suzuki wasn’t there in the list. Hence the simple explanation to why the greatest number of accidents for Toyota is simply because of the greater number of Toyota cars in the streets of New York hence making its probability high to get into an accident than Suzuki, BMW or Nissan.   
  
As per [Valuepenguin](https://www.valuepenguin.com/top-deadliest-vehicles), The Toyota Camry, the most popular sedan in the United States, is the fourth-most deadly car. It was involved in 4,734 crashes over their study period. However, the Camry consistently sells more cars than its top competitor, the Honda Accord. This suggests that it's more likely to be involved in a fatal crash than the Accord. The Camry has slightly fewer occupants killed per crash — 0.59, compared to the Accord's 0.65.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Make | Model | Total fatal crashes | Units sold in 2018 | Occupants killed per vehicle in a fatal crash |
| Nissan | Altima | 3,267 | 209,146 | 0.60 |
| Nissan | Sentra | 1,561 | 213,046 | 0.67 |
| Toyota | Camry | 4,734 | 343,439 | 0.59 |
| Toyota | Corolla | 3,430 | 303,732 | 0.64 |

Hence, as per the analysis done above though Toyota has the greatest number of accidents in 2018 when compared to its competitors BMW, Nissan, and Suzuki, it is one of the most reliable and widely used car manufacturer while Suzuki on the other hand less market share in USA hence it has low counts in accidents too.

## Analysis of number of collisions per month for four specific car manufactures which are BMW, Nissan, Toyota, and Suzuki.



From the above line graphs, we can conclude that monthly Toyota has the greatest number of accidents while Suzuki the least. The reason for this we have already analysed in part 1 analysis.

we can also conclude that the number of accidents per vehicle manufacturers isn’t so constant. Rather it changes over the months as per New York’s season.

For Suzuki its comparatively constant that’s because of low count of accidents when compared to others. When we compare each plot line, they show a similar trend. Specially the Nissan and Toyota plot lines can be seen flowing in a similar way.

Basically, there should be some external factors apart from the car manufacturer’s quality which maybe causing a greater number of accidents. Few of these factors are weather and Holidays.

If we see the line graph, we can verify that there is a drop in number of accidents from December to February. There are two main factors for this trend, firstly, the holiday season starts, hence the daily commute of many people is reduced thereby reducing the chances of accidents and secondly, its snowfall time in New York in those months thereby many people avoid travelling until its important and many prefers to travel by public transport if feasible as cars face cold start.

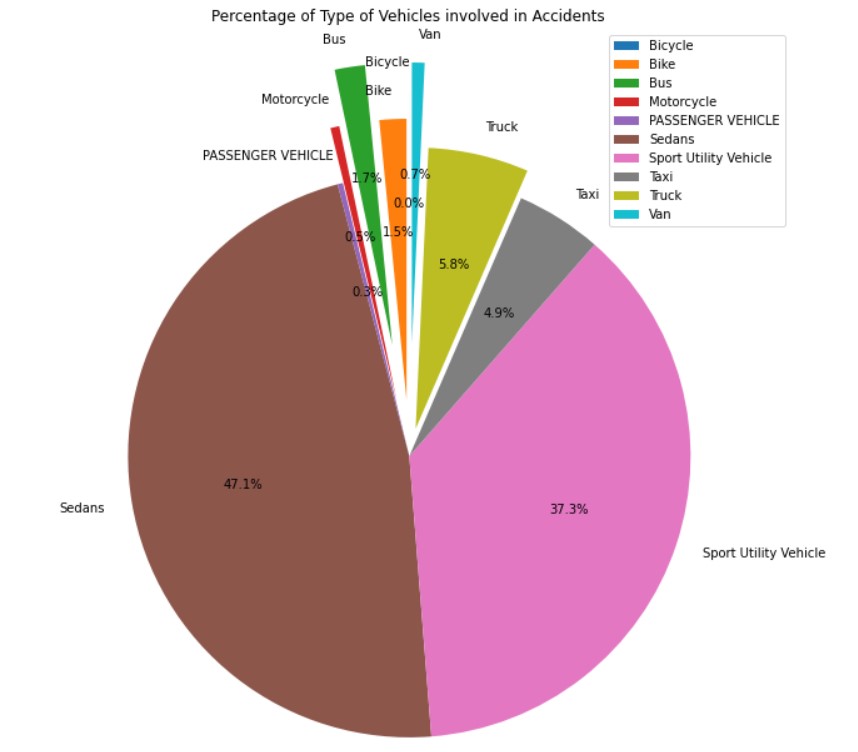
In February month we can see a sharp rise in accidents, this is mainly because of melted snow. A melted snow makes road even more slippery than a snow road. Snow covered roads can be messy, though for most cold and dry snow events the traction is generally uniform, predictable and manageable. There are, however, a variety of meteorological conditions that can produce exceptionally dangerous and slippery road surfaces. Many of these involve some form of icing. Slushy roads that occur when a snow event changes to rain may hide a layer of ice beneath the slush if the road surface temperatures are still below freezing. Which is a favourable condition in New York in the month of February hence a greater number of accidents.

And once the spring starts which is around March, the condition becomes favourable to drive a car, leading to lesser number of accidents but this again takes a higher trend. This can be for multiple reasons. It rains in New York from April to June thereby not making a comfortable drive for many divers.

We can see that in winter months (November-February) generally there is a huge dip in number of accidents while in summer months (June-September) there is a slant decline in the number mainly because of summer vacation in schools and universities. We can see a rise in accidents around September – August one of the contributing factor to it can the schools reopening.

So, we have analysed how weather and holidays can affect not only traffic but the number of accidents that takes place over the year. There are many other external factors that could be contributing to the accidents for example: the trend of substance abuse, natural calamity etc.

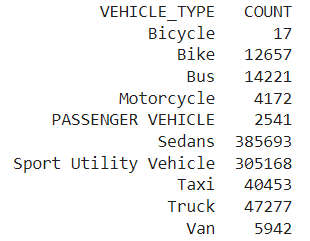
## Analysis of different type of vehicles involved in crashes.



Note: There was no much data for PASSENGER VEHICLE as sedans was already a different category. I have added convertible, limo and pass vehicle types to passenger vehicles.

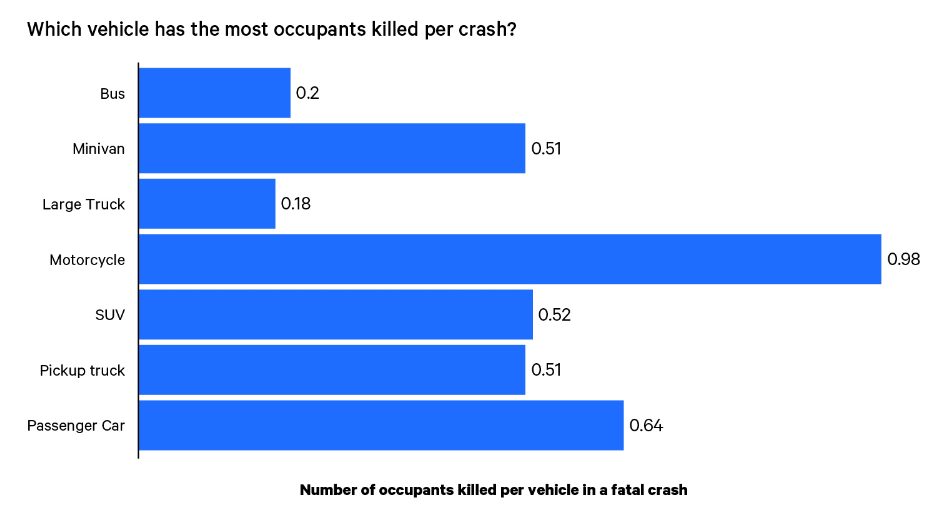
From the pie chart, we can analyse that the sedans followed by Sport Utility Vehicle constituted around 80 percent of the vehicle accidents in New York from September 2017 to August 2019, which is very high.

Though motorcycles and bikes were involved in a comparatively small number of fatal crashes — 12567 and 4172 — they are vastly overrepresented when controlling for the number of vehicles on the road. According to the Bureau of Transportation Statistics, there were 8.7 million motorcycles on the road out of 273.6 million total. That means that motorcycles account for just 3% of vehicles.



Furthermore, the typical number of motorcycle occupants killed per crash is a staggering 0.98. This means that if a motorcyclist is involved in a fatal crash, it's the motorcyclist or their passenger who is killed nearly every time.

Conversely, bus riders are very unlikely to be killed in a crash. Despite the fact that buses can carry a lot of people, the typical bus-involved fatal crash has just 0.20 occupant fatalities. This means that at least 80% of the time, no one on the bus is killed at all.



[Source](https://www.valuepenguin.com/top-deadliest-vehicles) for above Stats.

There by from this we can conclude that though Passenger Vehicles which generally includes Sedans and SUVS apart from other type of passenger vehicles constitute the highest percentage in number of accidents, and bikes and motorcycles which has a very low percentage in number of accidents has the most occupants killed per crash.