Project-1

Report

UNIVERSITY OF TEXAS ARLINGTON

Vrunda Vyas | Yash Tobre

(1002052013) (1002074030)

22nd September, 2022

Index

1. Overview

Project Goal:

To analyze the crash report data of New York over the dates 1 September 2019 to 31 Aug 2021. The analysis included 4 Vehicle Makes meaning 4 Vehicle Brands, namely Nissan, BMW, Audi and Subaru.

Project objective and scope:

To analyze the crash count over the period of 24 months and correlate it with an intent to find reasons behind it by analyzing the data. To observe the trend over the 24 months period and find plausible reasons for the different patterns. To observe and analyze if different vehicle types are prone differently to accidents.

Proposed Approach:

We extracted the data within the date range to a dataframe, and counted the values of different accident entries using the combination of dataframe manipulation and different functions. Moreover, we analyzed the queries and found out the corresponding relevant events in the history of different vehicle brands as well as took into consideration the distinct events that may have impacted the traffic collision over the years.

2. File Description

Libraries that are used :

Library	Usage
Pandas	Pandas provide extremely streamlined forms of data representation. We have used pandas for **data cleaning** and **analysis**.
re	We have used regular expressions for **pattern matching**. That helped us to search for a pattern again without rewriting it.
Matplotlib	We have used Matplotlib to create **static,**

	animated, and visualizations in Python.
Seaborn	To **enhance the visual feature** of the graph.

Dictionaries -To associate a value to a unique key, and then to quickly access this value To analyze related factors efficiently by key value pair. With this we used to find a certain python object.

List - To store multiple items in a single variable we have used a list.

Dataframe - To manage multidimensional data in an easy way. It allows to store and manipulate tabular data in rows of observations and columns of variables, as well as to extract valuable information from the given data set.

3. Division of Labor

		Hours
Yash Tobre	Query 1 &2 Programming and Analysis, Report Making	8
Vrunda Vyas	Query 2&3 Programming and Analysis, Report Making	8

Problems Encountered

a. Whitespace Problem: In our data, we encountered some entries, after doing the name cleaning, with 3 letters (like BMW) with the 4th letter being a white space. This was the reason why most of the entries of VEHICLE_MAKE= BMW were not simply not getting detected.

   Solution: We checked the data once again, and tried searching for entries using whitespace. This caused many entries to show up. That led us to figure out the real root of the problem. Hence we used the replace function to replace whitespace with nothing, i.e. “ “ with ‘’. This solved our problem.

b. The Dataframe Jumbling: Due to us creating different data frames for different queries, and running them in a different order, some of our data frames were modified in a different way before we could use them. For example, the name cleaning function was

applied on a dataframe without the segregated year and month columns, which lead to that data being unclean.

Solution: We set a specific order, and updated the variable names so that we could understand which data frame was used for which purpose. This spared us a lot of confusion.

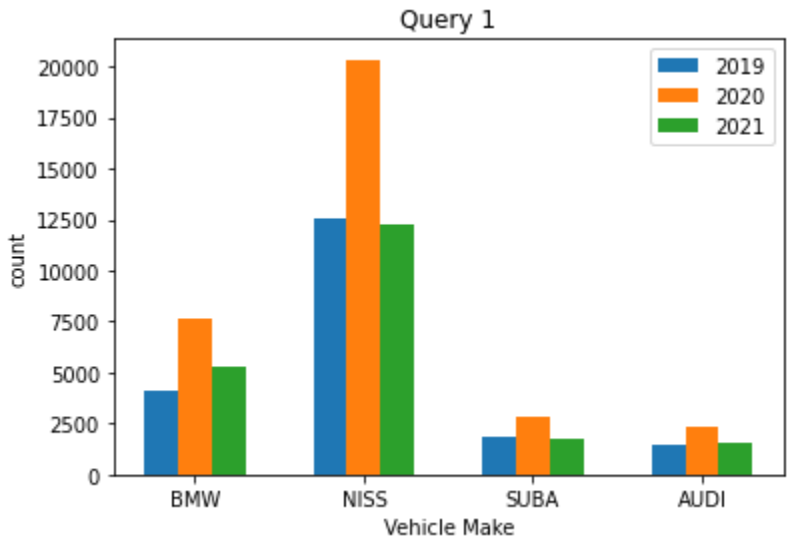
c. Groupby Function: For us, using the groupby function and counting for specific attributes was a bit hard. We tried to understand the groupby function and implemented it, but it was unsuccessful. So we were looking for some way to create a counter.

Solution: To count the specific entries of a specific column, we took the help of a groupby function and a counter column. This enabled us to extract information from a counter column instead of any other multiindex column.

d. Axis of Graph: Our dataset was not perfect 24 months, instead it included 4 months from]2019, 12 months from 2020 and 8 months from 2021. Thus we had a challenge of plotting a continuous plot for a 24 month timeline.

Solution: We created 3 different line graphs, one for each year. This enabled us to understand information and retain it in the same way.

5. Analysis from Query 1



a) It is strikingly clear, the year 2020 had more crashes than any other year. But again the year 2020 included more months than the other two. Therefore we calculate the monthly average of all the months for all 3 years. It is shown in Table:

	BMW	AUDI	NISSAN	SUBA
Year 2019	1024	360	3149.75	470.75
Year 2020	1907.5	586.25	5084.75	718.75
Year 2021	1328	397.5	3066	446

Table: Average Number of Monthly Accidents for each Vehicle Make

Even with the average number of monthly accidents, it is clear that 2020 had a spike in accidents. An Article from nytimes confirms that despite the pandemic the number of accidents increased drastically[1].

b) From the different vehicle makes, Nissan is clearly the one involved in the Majority of Accidents. An article from BusinessInsider lists the top 20 cars to be most likely to be involved in accidents, and Nissan has the most cars in the list[2].

c) Car Suba has an Eyesight Driver Assist Technology, which monitors your position on the road and will alert you if you stray out of your lane. Pre-Collision Braking helps you avoid or reduce frontal impacts by alerting you and applying full braking force in emergency situations. Eyesight also detects objects in front of the vehicle that you are likely to hit. The first generation of the Eyesight system was available on 2013-14 legacy and outback vehicles and the 2014-16 Forester. In recent times, safety features have grown in number and sophistication, doing its part in preventing or mitigating collisions. Moreover the 2019 versions included advanced features, too. Because of this Technology Suba has very less car accidents comparetevely to other cars.[3]

d) The reason BMW has less accidents is because there are plenty of safety features to go around in modern BMW. A big part of staying on the road is accident prevention. Assistant suite includes a frontal collision warning system, active blind- spot detection, lane departure warnings, and speed limit notifications. In addition it also offers: Automotive city collision mitigation, Park distance control, BMW assist eCall, Active Protection System, Fatigue and focus alert. BMW cars were awarded the safety pick award. Which explains the less number of accidents as compared to Nissan. [4,5,6]

e) Audi has introduced features like pre sense, Audi adaptive cruise control with Stop & Go and Traffic jam assist, Vehicle exit assist, Turn assist and collision avoid assist, Night vision

assistance to reduce accidents. Moreover, Audi has maintained a low recall rate throughout the year, which explains its low accident count.[7]

6. Analysis from Query 2

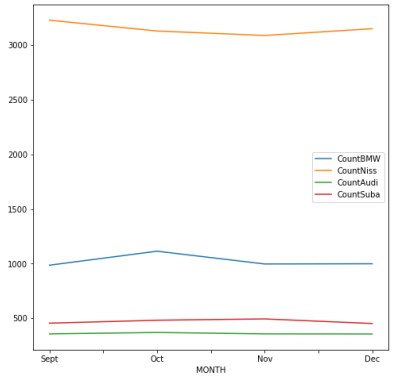


Figure: Line Graph for months against count for Year 2019

a) For Year 2019:
   i) Nissan: Nissan has recalled many of its different models, over the year 2019 due to different reasons like potential fire hazards, reverse camera backup system failure. This explains the marginal gap between the counts of other vehicle makes. Moreover, in November, potentially because of the recall, there is a small dip in the number of accidents.[8]
   ii) BMW: The count of BMW is comparatively less as compared to Nissan, however we see a spike in October. This can be attributed to the BMW Recall in 2020 regarding backup camera concern. After this recall, however BMW went back to maintaining it's previous position.[9]
   iii) Subaru: According to the data generated, Subaru noticed a spike of accidents count until November 2019. This was due to the key component issues observed in a 2019 Subaru Model. The line is increasing linearly until it reached a peak (when the recall was called) after which it went back to it's previous slope.[10]
   iv) Audi: Audi has managed to maintain the lowest recall rate of all the brands over the years 2014-2019. As we observe from the graph, Audi has the lowest accident counts, which can be attributed to advanced safety features.[11]

Moreover, from the data we can see that the number of accidents has less variance through the months September to December. This can be verified, as

July, August and October are regarded as most accident-prone months. So the amount does not see any spike except one in October. This is usually attributed to Vehicle miles going down in winter, and going high in summer.[12]

b) For year 2020:

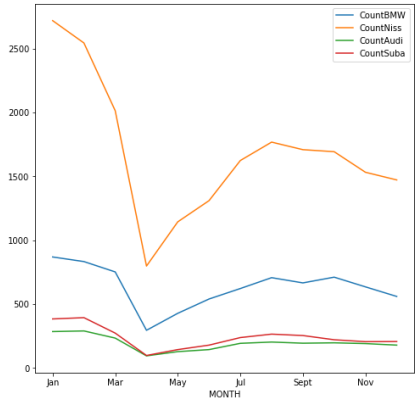


Figure: Line Graph for months against count for Year 2020

For all the Vehicle brands in 2020, we observe a similar pattern for almost all the vehicles. There is a huge drop in April 2020, which is the result of Stay at home restrictions due to COVID-19 global pandemic. Moreover, the seasonal variation trend is also followed, that is July-August being more accident prone to others, and October showing the spike. We can observe a spike in the July to August section, which could correspond to the recall regarding the false bluelight illumination, which decreases after the recall alert and follows the regular trend of spiking up again in October.

c) For the year 2021:

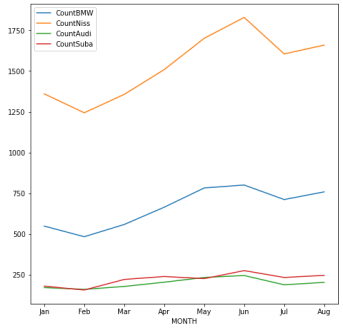


Figure: Line Graph for months against count for Year 2021

As we analyze the data, we see that the accident count of 2021 is very similar to the year 2019.[13] However, we can clearly see an increase in the traffic accidents till the month of June, which was an alarming concern for all the New York Residents[14]. Moreover, this count is more than as compared to 2019[15]. Afterwards, the trend returns to its normal values, and we again see a spike in July- August Patch. We see a slight spike for Audi in May month which can be attributed to the recall for vehicles made after February 2021.[16]

7. Analysis from Query 3

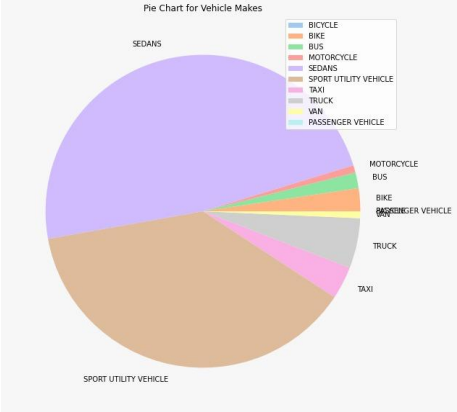


Figure: Pie - Chart for percentage crashes for each vehicle_type

From the pie chart, we can easily say that Sedan vehicle_type had the most accidents occurred compared to other car that is 48.1% from all of the vehicle_types, because due to the smaller car size as compared to other vehicle types like SUV and Trucks, it is more likely to be in the blindspot of bigger vehicles.Moreover, Sport Utility Vehicle had also high percentages of crashes that is 37.9%, although SUVs are safer than Sedan, they pose greater threat to other vehicles while parking and reversing.

Therefore, We can say that from 1st September, 2019 to 31st August,2021 , Sedan and Sport Utility Vehicle types had the most crashes.

References

1. https://www.nytimes.com/2020/05/21/business/suv-sales-best-sellers.html
2. https://www.businessinsider.com/most-dangerous-vehicles-small-cars-death-fatalities-fatal-crash-ratings-2020-6?amp
3. https://www.torquenews.com/1084/all-new-subaru-foresters-now-come-fitted-eyesight-what-s-big-deal
4. https://www.iihs.org/news/detail/2019-bmw-3-series-4-door-sedan-is-a-top-safety-pick-
5. https://www.iihs.org/news/detail/2020-bmw-2-series-gran-coupe-earns-top-marks-for-crashworthiness
6. https://www.iihs.org/ratings/top-safety-picks/2021/all/bmw
7. https://www.nerdwallet.com/uk/personal-finance/the-most-recalled-cars-of-2019/
8. https://www.theguardian.com/business/2019/nov/17/nissan-recalls-nearly-40000-cars-in-us-over-potential-fire-hazard
9. https://bottarolaw.com/blog/vehicle-recall-bmw-recalls-more-than-250000-vehicles-due-to-backup-camera-concerns/