

Analysis of Motor Vehicle Accidents in New York between 2019-2021

Brief overview of how the project was approached:

This project was approached with a very basic understanding of Python as my background was in Psychology, and I did not have the required skills to start working on this project immediately. However, I was glad we were given enough time to work on it because I spent a week learning how to pre-process data and learn the essential functions of Python. The project was done using Google Collab, and the Pandas and Numpy libraries were used in this project. Once I felt a bit more skilled, the pre-processing didn't take that long, and I started by cleaning the data, removing all the unnecessary columns, removing the NAN values, and ensuring that my data set only related to what was required on the project. After the pre-processing, I got my bar graph done pretty quickly but struggled a lot with the line graph and the pie chart. Learning how to get the graphs from my data was tough, and trying to figure out how to group the needed data based on the parameters for the graphs was challenging. I did a lot of research, read a lot of articles, and watched a bunch of tutorials that really helped me gain some confidence. They helped me realize that the project was not too hard and just required a lot of hands-on problem-solving skills rather than knowing how to use Python professionally. Once I felt comfortable with my Python skills, I encountered the project more confidently and managed to get things done more efficiently.

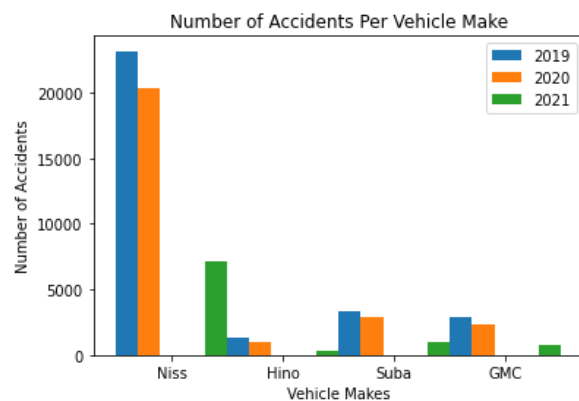
I struggled with finding what function to use and when to use it for the analysis because of my limited knowledge of Python, but the Data Science course this semester helped a lot with this project. The Data Science course went over the basics of Python and provided a lot of the basic functions that potentially helped understand this project better and made it easy to get it done.

Division of Labor

This project was done by myself only, and it took me two weeks to get this project done. The programming part took the longest, and the report took two days to get done.

Analysis of each graph:

1) Analysis of accidents per vehicle make



Based on the bar graph, we can see that Nissan had the most accidents in 2019, 2020, and 2021 compared to all other vehicles. According to an article that showed vehicle sales in the United States over 2019, 2020, and 2021, Nissan sold around 2.9 million vehicles, which is apparent from the bar graph showing that Nissan had the highest accidents. Especially in 2019, when the accidents for Nissan reached approximately 23,000. Nissan sells the fourth most cars in the United States and has the highest overall sales compared to the other vehicles.

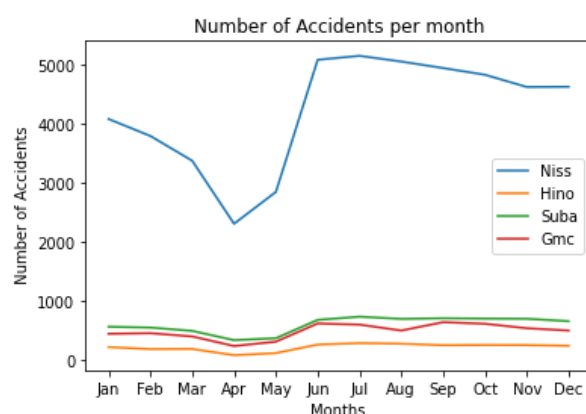
The sales for Subaru were around 1.9 million for the three years. According to the bar graph, the number of accidents for this vehicle was highest in 2019, bringing the total number of accidents to around 7,217 overall.

For GMC, the total sales were around 1.5 million over the three years, and the most accidents for this vehicle were in 2019 as well, bringing the total accidents to about 5,966 over the three years.

The total annual revenue for Hino vehicles was around \$49,000. However, the total number of accidents for Hino was around 2,647, and 2019 was the year with the most accidents for Hino vehicles, just like all the other vehicles. Hino is known for manufacturing commercial and diesel engines, so not a lot of information was found in regards to their sales. This is also the reason why the number of accidents for this make were the lowest as it is usually for commercial purposes and not personal purposes.

All vehicle makes had the most sales in 2019 as this was pre-covid and most people were employed and were living normally. However, after 2019, the accidents and vehicle sales dropped due to Covid-19, as people lost their jobs and could not spend their income on vehicles anymore, hence the drop in sales and accidents.

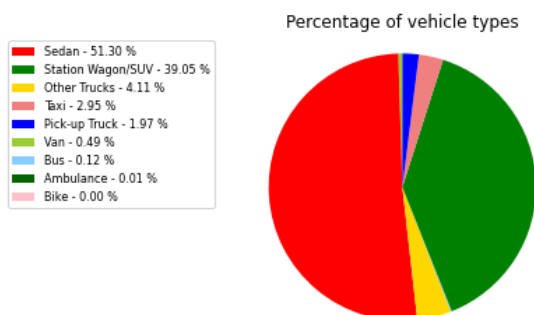
2) Analysis of accidents per month as per the line chart



According to the line graph, most accidents occurred around June and July for all vehicle makes, which could be the case because it is summer and a lot of young adults drive late at

night, and that causes a lot of accidents caused by drinking and driving. Another reason is that more people travel during summer, and that can potentially cause more accidents. There was a dip in the number of accidents around April, which could be because Covid-19 started in December of 2019 and quarantine started, everyone started working from home, and children were studying from home, which meant fewer vehicles were on the road. The number of accidents was pretty constant from July onwards for all vehicle makes, which could be due to fewer young adults on the road late at night as college and school start from august onwards. It could also be due to the cold and icy weather in New York, as fewer people prefer to drive in the snowy conditions.

3) Analysis of the pie chart which shows the percentage of accidents per vehicle type.



As per the pie chart, we can see that most accidents were caused by sedans and Station wagons/SUVs. According to an article by cars.com, the reason for this is that sedans are usually cheaper, financing is more convenient with sedans, and they give better mileage. Also, with the recent inflation and unemployment issues, people prefer to buy cars with lower down payments and instalments. However, that does not mean that station wagons/SUVs

aren't being bought. As it is apparent from the pie chart, the number of accidents for this vehicle type came in second. This is because large families still like to buy larger vehicles that can fit their entire families. The lowest number of accidents were caused by bikes, buses, and ambulances, and this is because a smaller number of bikes are purchased as compared to cars. Also, ambulance and bus drivers have more navigation skills and better driving skills than the average drivers as it is part of their job to ensure patients and students are securely dropped off at the hospital and school.

Three problems that were encountered:

1) Not being able to understand the problem at hand.

The biggest hurdle was not being able to completely understand what was asked on the project and how to ensure that what I was doing was actually right. I struggled with understanding what function to use and when to use it. Just trying to figure out what would work for me was the hardest part. I overcame this issue by researching everything, teaching myself the basics, and ensuring I understood what was asked on the project. I also approached the TAs and emailed them to clear any doubts I had.

2) Not having enough skills to tackle the Python related problems

My limited skills in Python definitely made things harder because not having enough knowledge on how to code and pre-process was challenging, especially when I first started. However, I taught myself a lot during these last two weeks, which helped me a lot in this project. Also, the Data Science course also helped a lot with this project.

3) Finding multiple articles and videos which confused me more

Often, I would find multiple articles and videos on how to solve a particular problem, which confused me more because for a beginner having multiple resources at hand makes things

harder to understand. However, the solution was to search for specific things and issues and not read every single article that came up.

Conclusion:

Overall, the project went well despite having a fear of the unknown. Being from a psychological background and having very limited information about Python, I was genuinely scared that I would not be able to finish the project in time and would face many difficulties. However, knowledge is power, and learning the necessary skills was the key. Despite all the challenges, I was able to finish the project in time and also learned a lot about how large datasets are analysed using python.

Resources:

<https://www.goodcarbadcar.net/>

<https://www.motorbiscuit.com/americas-most-popular-car-brands/>

<https://www.macrotrends.net/stocks/charts/HINOY/hino-motors/revenue>

<https://www.cars.com/articles/why-now-is-the-time-to-buy-a-sedan-not-an-suv-421330/>

<https://www.anchorlegalgroup.com/accidents-increase-summer/>

<https://pubmed.ncbi.nlm.nih.gov/32721782/>

https://learn.org/articles/how_to_be_an_ambulance_driver_requirements_training_program.html