DASC-5300 Motor Vehicle Collisions
Group 50

Overview: In this project, data on vehicle collisions are extracted, cleaned, tested, and analyzed based on numerous criteria and conditions. The data file has been mounted on Colab and uploaded to Google Drive. All the rows in the data file with blank or unknown values were removed using a pandas Data frame. I carried out additional research after moving the cleansed data file to a new location in Google Drive. Data frame is used to organize the records, run analysis, discover correlations, and visualize the data using Pyplot. While using pandas and pyplot, we had a few challenges, but we were able to get beyond them by consulting Google and discussing with each other.

File descriptions: After cleaning the data contained in the file **Motor Vehicle Collisions - Vehicles.csv**, we used it for additional analysis. We mostly utilized data structures such as Data Frames, Python Lists, and Numpy Arrays.

Work Distribution: Based on the task's difficulty and projected completion time, we equally divided the analytic, programming, and documentation portions. We both programmed by breaking down tasks into manageable pieces and brainstorming ideas for how to organize or analyze a graph, etc. The coding process took a total of 36 hours, and the report took 15 hours.

Work Division:

• Rohit Nayakanti (1002024866): The initial project understanding component was co-worked. worked on both analyses 1 and 2.

• Sasank Kinnera (1001874178): In the first phase of the project, we worked together. performed report analysis and analysis 3 tasks.

Problems Encountered: We have encountered a few problems while working with Pyplot and pandas, which are mentioned below.

• It took some time for us to come up with a solution to the problem statement.

• Since then, we've encountered a few minor obstacles, including date-to-time conversion, using functions to shorten static code and process variables dynamically, etc.

• The technique presented in the classroom session helped sort the vehicle type names to clean the data.

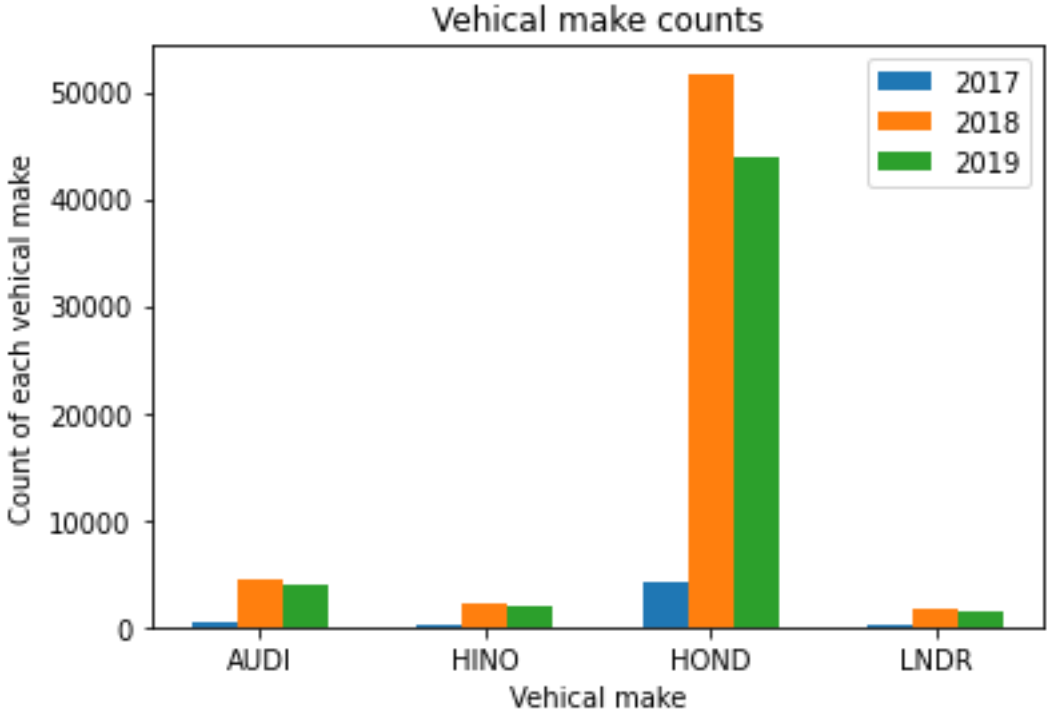
• Each time a problem arose, the group effectively worked through it by discussing it. Also used the internet to better comprehend and put into practice new ideas.

• First, we wrote a long piece of code using conditions such as for loops and if statements. Then, we researched how to shorten the long piece of code.

Analysis #1:

In this analysis, we've drawn a bar graph with vehicle make on the X-axis and the number of collisions each of those cars causes on the Y-axis. Three time periods were chosen: from January 1 to December 31, from January 1 to December 31, and from January 1 to November 30, 2019, and the corresponding colors were allocated.

In the created graph, it can be seen that there were much more accidents involving vehicles made by Honda than any other vehicle makes throughout all three intervals. The vehicle brand AUDI ranks second in terms of the total number of accidents, while the other two, HINO and LNDR, have nearly identical totals.



Conclusions:

- It can be observed that the most accidents have been recorded in the year 2018 for each vehicle made as below.
    - HOND – 51704
    - AUDI – 4512
    - HINO – 2082
    - LNDR – 1680
- The global sales of vehicle_make Honda (HOND) and Land Rover (LNDR) between the years 2017 to 2019 are as follows

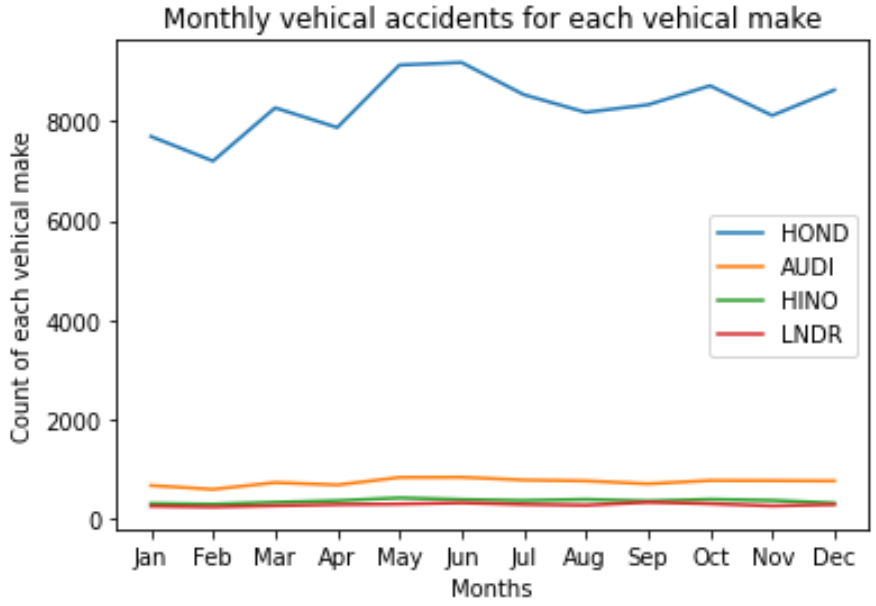
	HOND	LNDR
2017	502.8 billion	44.2 billion
2018	519.9 billion	41.2 billion
2019	532.3 billion	39.6 billion

- As seen in the preceding table, the ratio of sales for both vehicle makes is quite low, as is the number of accidents.
- Over analyzing the data on the web, it has been clearly observed that the Main cause of accidents and crashes is due to **Distracted Driving** which is a human error**.**

Analysis #2:

In this analysis, we've drawn a line graph with the average of each month over all the years on the X-axis and the total number of accidents in each of those vehicles on the Y-axis. We assigned the appropriate colors based on the vehicle makes allotted to our group—namely, HOND, AUDI, HINO, and LNDR.

On the generated graph, it is clear that Honda has experienced a significantly higher number of accidents compared to the other four vehicle makes. The vehicle brand AUDI ranks second in terms of the total number of accidents, while the other two, HINO and LNDR, have pretty much identical numbers.



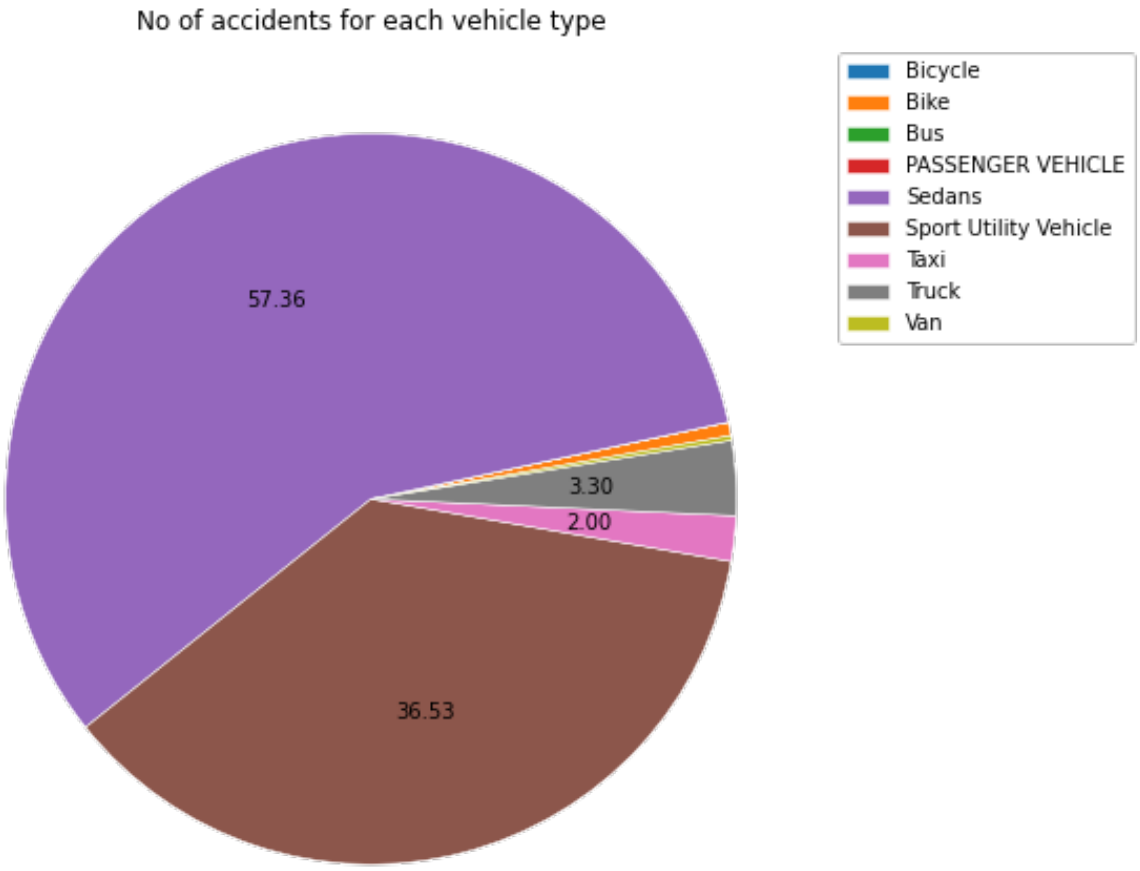
Conclusions:

- Over the years, it has been noted that **Honda** has been involved in the majority of accidents.
- After carefully examining the line graph above, it is clear that for all vehicle makes, May and June had the highest number of accidents, and that the summer season is what caused the big rise. There are more cars on the road during these warmer months. Unfortunately, this may result in more collisions and tragic traffic-related incidents. According to studies, the two riskiest months to drive in New York City are May and June. During these months each year, hundreds of people suffer injuries in vehicle accidents.
- On the other side, February was the month with the fewest accidents across all vehicle makes. People are simply not on the roads as much in February. It makes sense why one wouldn't want to be traveling by car during this time of year, when Northern and Midwestern states may have erratic winter storms. Additionally, more long-distance travel is made in the fall around the holidays, which might contribute significantly to the data for later months. Additionally, because February was a shorter month than January, fewer car accidents happened.

Analysis #3:

In this analysis, the percentage of accidents that occurred for each type of vehicle over the years 2017, 2018, and 2019 is represented by a pie chart. The HOND, AUDI, HINO, and LNDR car types were represented with the appropriate colors in the resultant pie chart, which was created by taking into account all vehicle makes.

It is clear from the generated pie chart that, out of the 10 different vehicle categories, sedans account for a significantly high percentage of accidents. Sport utility vehicles are the second most common type of vehicle in terms of the proportion of accidents and are followed by trucks, taxis, and other vehicles.



Conclusions:

- Most accidents have been reported for the vehicle type Sedans over the other vehicle kinds.
- Most of the Sedans would be utilized for personal use, and below are the primary reasons why the accident rate is high
    - Distracted Driving
    - Accidents Caused by Speeding
    - Drunk Driving Accidents
- The pie chart shows that the vehicle types with the lowest percentage of accidents are passenger vehicles, buses, etc. It is clear that the drivers of these vehicles have the necessary training to ensure that they are never distracted.