

Traffic Violations

Hypothesis: Accident frequency can be predicted for automobiles based on their color and the month of the year.

EDA

-Can accidents be predicted based on the month?

-Is there a connection between certain colors of automobiles being in more accidents due to their color, or is it more a popularity of that color leading to more of them to have more accidents? Percent out of total of that color only in accidents, may give most accurate analysis here.

-Can a recommendation be made on certain colors being safer automobiles?

-What caused the year 2017 to have more accidents?

- How significantly did alcohol affect the amount of accidents?

Future questions - Can accidents be predicted based on the day of the week (weekday, weekend)?

Can conclusions be made that driving a certain color car on a certain day of the week is more likely to get in an accident vs other car colors?

Problem Solving Outline:

Columns to be used will be; VehicleType, Color, Date of Stop, Accident

Using the rows that all had 'Yes', identify all types of vehicles listed in the 'Vehicle Type' and their value_counts. Decided to only include 4 wheeled as this is the market we are looking at.

Identify if there are any 'null' values for 'Color', in the '02 Automobile' category. Eliminated those.

Access only 'yes', for being in an accident. Access only 'no', for being in an accident.

Identify all colors in this study. Noticed 24 colors were identified in accidents, vs. 26 showing up not in accidents. Eliminate the bottom rung of low producing statistics as they may not be relevant.

Create a deliverable to show number of accidents in each color automobile, broken down by day of the week; time-series plot.

Clients:

The clients would be car dealerships, manufacturers, and auto insurance agencies.

Dealerships can use this data to change up their inventories to reflect selling safer vehicles, and based on their area statistics, to boost sales numbers.

Auto manufacturers can use this data for designing future body styles and colors.

Insurance companies can have data to back up rates that are higher or lower on certain vehicles.

Data:

<https://data.montgomerycountymd.gov/Public-Safety/Traffic-Violations/4mse-ku6q>

Traffic Violations in Montgomery County, Maryland

Deliverables:

of total of that vehicle type vs. # of each color vs. # in accidents broken down by color

A Pandas DataFrame of accident outcomes for each color by percentage.

Graph to represent the percentage of each color automobile were in an accident.

Graph to represent the percentage of each color automobile that were in NOT an accident.

Time series graph to show the number of automobile accidents for each day of the week.

[rubric](#)

Notes

----think about machine learning; supervised(predictions for the accident column, unsupervised(grouping similar or like)

----text wrangling

----categorical features