US Accidents Analysis 2016 - 2021

A Countrywide Traffic Accident Dataset

Sydney Lieske & Isabella Morales

December 2022

Introduction

This project is an analysis of the US Accidents dataset which covers 49 states of the USA. The data is from Feb. 2016 to Dec. 2021 which is collected from multiple APIs that stream traffic incident data.

With these APIs, they broadcast the traffic data captured by various entities such as the US and state DoT, traffic cameras and sensors within road-networks, and law enforcement agencies. Currently, there are about 2.8 million accident records in this dataset.

Motivation & Goals

- The objective of our data analysis on U.S. accidents was done in order to find preventative measures. Preventative measures can afterwards be used in order to reduce the amount of accidents that occur. Understanding this dataset would allow for the ability to limit the amount of crashes that occur during a year

- The US Accidents dataset, can be used for real-time car accident prediction, casualty analysis, studying the impact of environmental stimuli on the occurrence of accidents, and even car accident hotspot locations. It also can be used to study the Covid-19 pandemic on traffic accidents and behavior

Research Questions

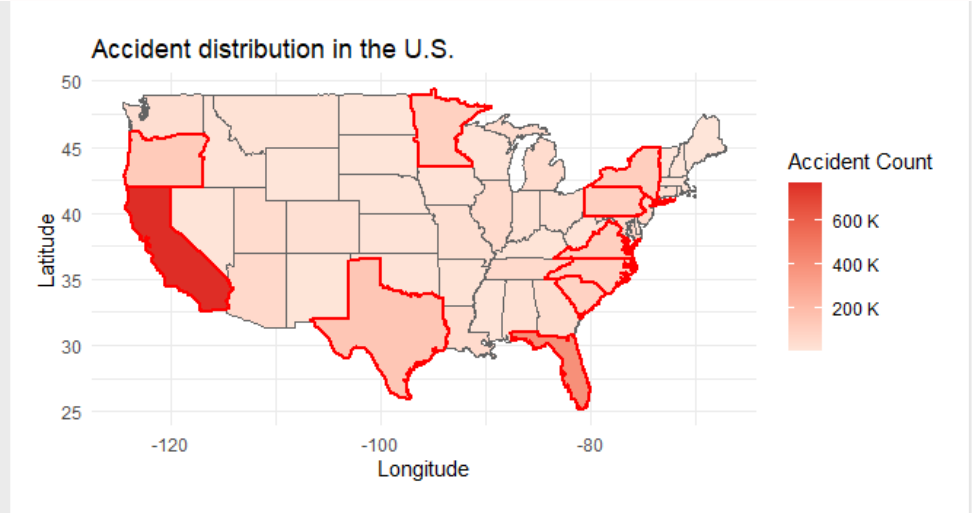
- What is the distribution of accidents across the country?
- Which states have the most cases of accidents? Least?
- Which streets are the most accident prone?
- Which months have the most accidents?
- What time and day are safer to travel?
- What factors are most responsible for these accidents?
- During what weather conditions do the most accidents occur?
- What can be implemented to help reduce these accidents?
- How could these accidents be minimized?
- What kind of IoT devices were used to report the incidents?
- Was an ATMS used?

Analysis Plan

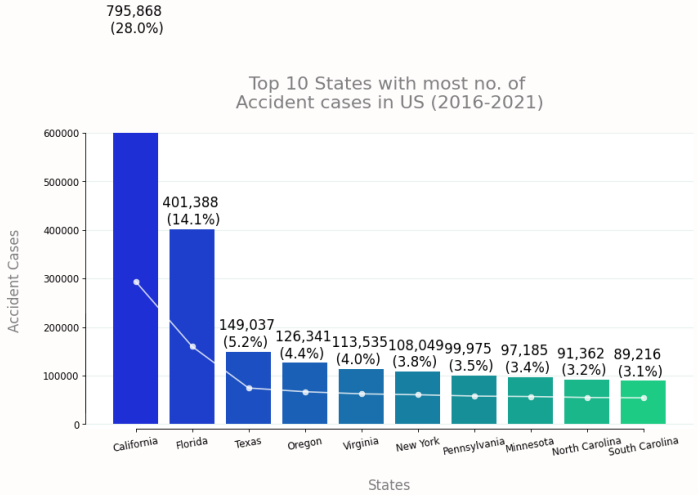
- Research
- Data cleaning
- Analyzing data
  - Location
  - Time
  - Road Conditions
  - Weather

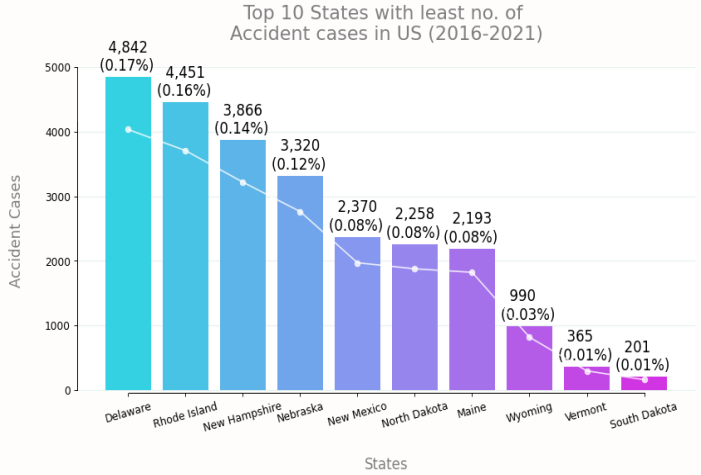
- R / Python
- Rstudio
- Google Colab notebooks

Accident Distribution

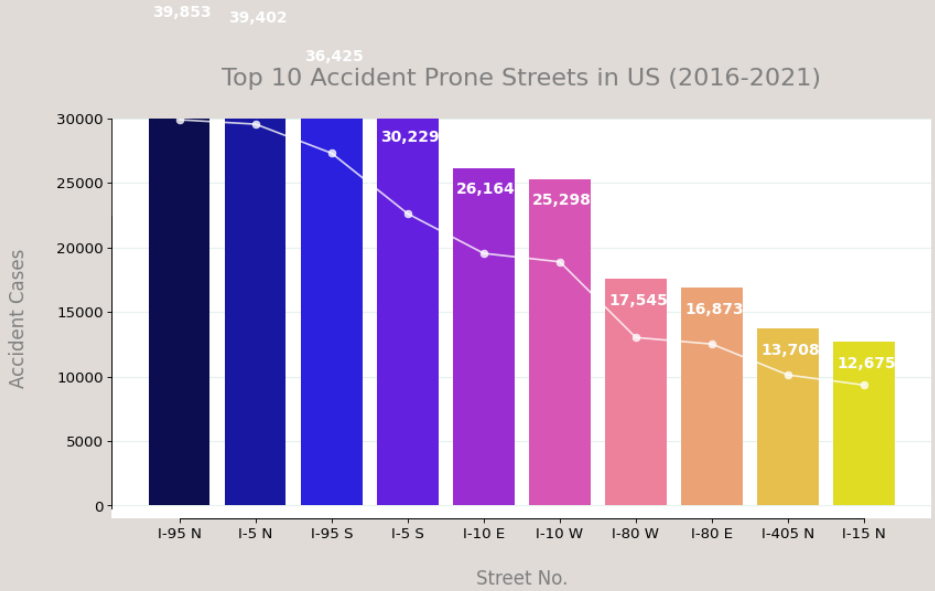


Top 10 States



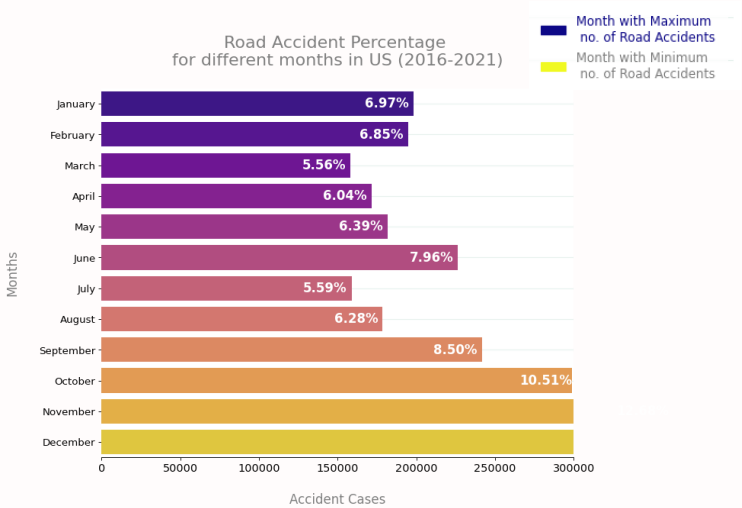


Analysis

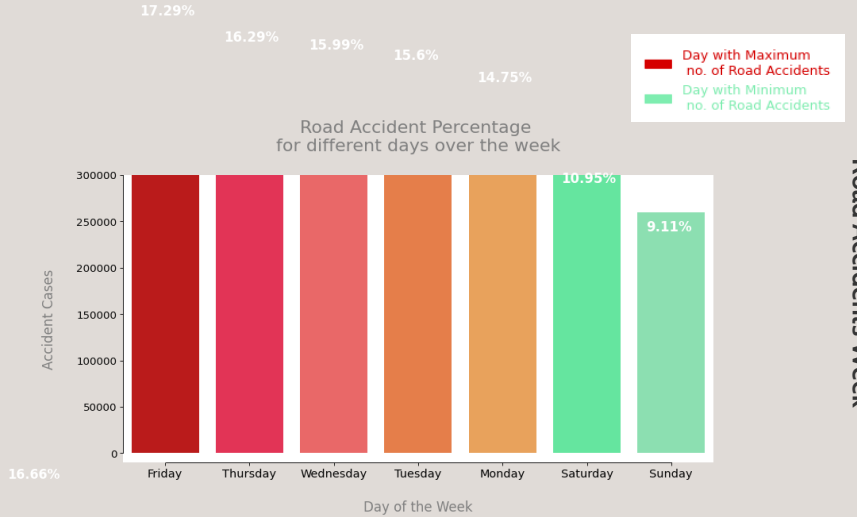

Top 10 Accident Prone Streets in US (2016-2021)
Accident Cases
Street No.
30000
25000
20000
15000
10000
5000
0
39,853
39,402
36,425
30,229
26,164
25,298
17,545
16,873
13,708
12,675
I-95 N
I-5 N
I-95 S
I-5 S
I-10 E
I-10 W
I-80 W
I-80 E
I-405 N
I-15 N

Road Accidents

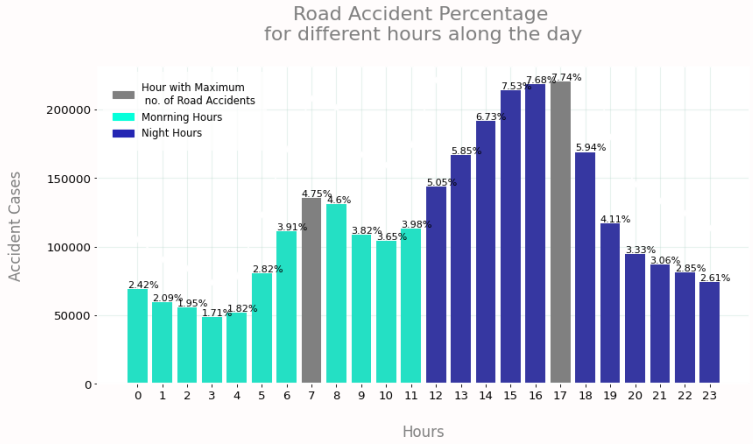
Road Accidents Months

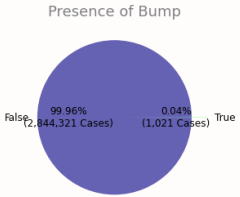


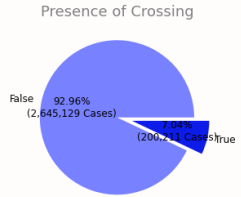
Road Accidents Week

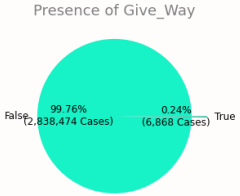


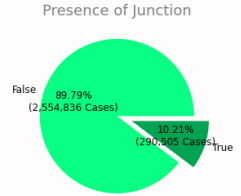
Analysis

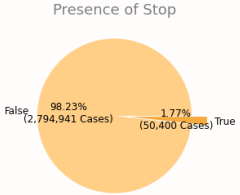

Road Accident Percentage
for different hours along the day
Hour with Maximum
no. of Road Accidents
Monrning Hours
Night Hours
Accident Cases
Hours
0
50000
100000
150000
200000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
2.42%
2.09%
1.95%
1.71%
1.82%
2.82%
3.91%
4.75%
4.6%
3.82%
3.65%
3.98%
5.05%
5.85%
6.73%
7.53%
7.68%
7.74%
5.94%
4.11%
3.33%
3.06%
2.85%
2.61%

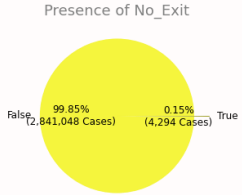

Presence of Bump
False
99.96%
(2,844,321 Cases)
0.04%
(1,021 Cases)
True

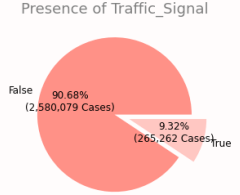

Presence of Crossing
False
92.96%
(2,645,129 Cases)
7.04%
(200,211 Cases)
True

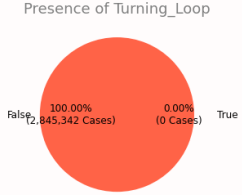

Presence of Give_Way
False
99.76%
(2,838,474 Cases)
0.24%
(6,868 Cases)
True


Presence of Junction
False
89.79%
(2,554,836 Cases)
10.21%
(290,505 Cases)
True

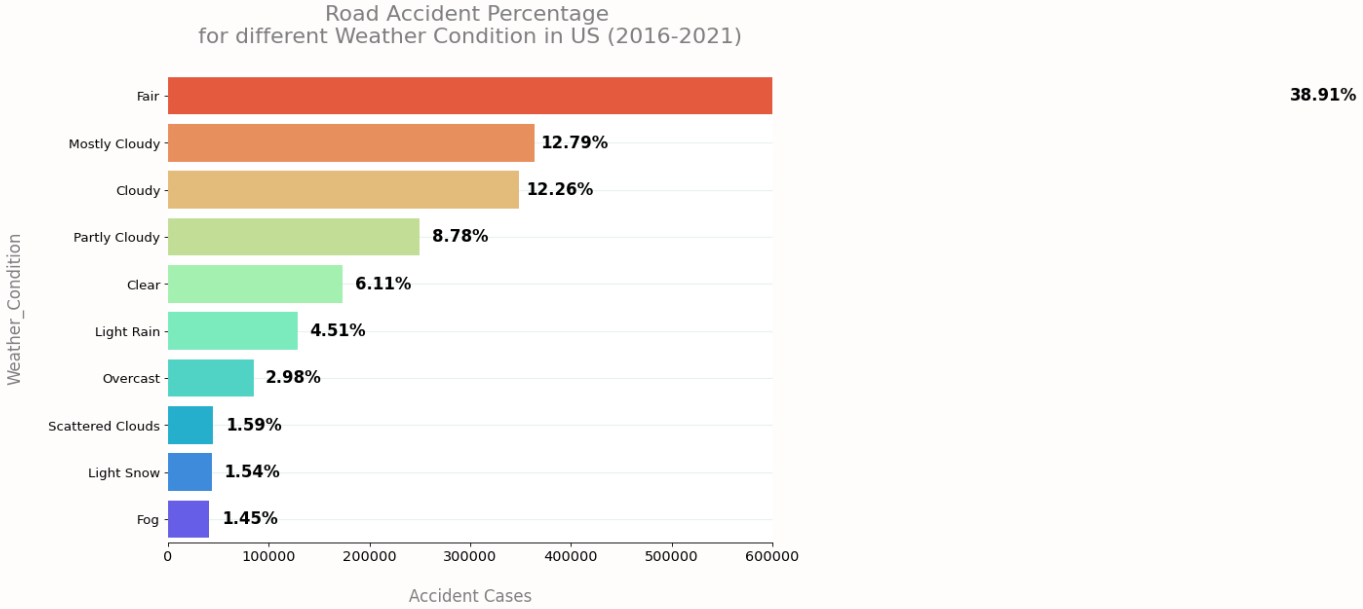

Presence of Stop
False
98.23%
(2,794,941 Cases)
1.77%
(50,400 Cases)
True


Presence of No_Exit
False
99.85%
(2,841,048 Cases)
0.15%
(4,294 Cases)
True


Presence of Traffic_Signal
False
90.68%
(2,580,079 Cases)
9.32%
(265,262 Cases)
True


Presence of Turning_Loop
False
100.00%
(2,845,342 Cases)
0.00%
(0 Cases)
True

Weather Analysis



Results & Conclusion

Through the analysis of our chosen dataset, we were able to answer a majority of our research questions such as:

- Fridays are the most accident prone with cases at 17.29%, vs Sundays 9.11%, the lowest of the days of the week. Hours 3-5pm are the most in the evening time with 5pm being the highest, and 6-8am with 7am at the most in the morning.
- The presence of traffic signals are just one of the preventive measures that we were able to analyze within this data set.
- Accidents occur more on highways or state road than on local streets, preventive measures like speed limits
- Roadside messages and if the vehicles are equipped with ADAS would help to minimize accidents.

Sources for Dataset

Using the dataset:

- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. “A Countrywide Traffic Accident Dataset.”, 2019.
- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. "Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights." In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019.

Questions