# Weather Impact on Accident Severity

Springboard Capstone 1
Greg Gibson

# Data Source:

Published December 2019: https://smoosavi.org/datasets/us_accidents

Description: This is a countrywide traffic accident dataset, which covers 49 states of the United States. The data is continuously being collected from February 2016, using several data providers, including two APIs which provide streaming traffic event data. These APIs broadcast traffic events captured by a variety of entities, such as the US and state departments of transportation, law enforcement agencies, traffic cameras, and traffic sensors within the road-networks. Currently, there are about **3.0 million** accident records in this dataset.

Acknowledgements: Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. "A Countrywide Traffic Accident Dataset.", arXiv preprint arXiv:1906.05409 (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.

# Can Weather Predictably Impact Traffic Severity from Accidents?

There are many available travel routing applications, such as from Google, Apple, Mapquest, and Waze, to name a few.  They conveniently provide shortest distance and fastest route planning, as well as amenities such as fuel and restaurant locations.

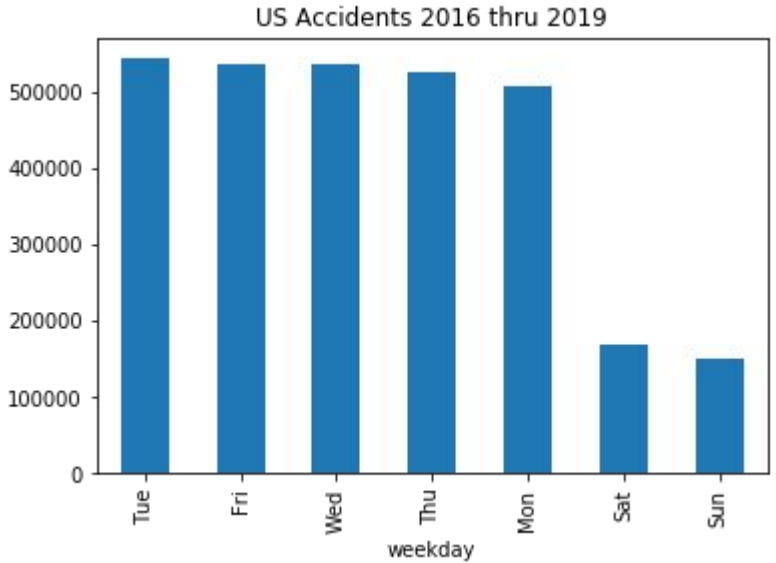
In addition, these applications can warn drivers of delays due to accidents and congestion, offering updates on when you can expect to reach your destination and alternate routes.

Is it possible to also use weather information to predict how long traffic will be impacted by the accident?  This additional information could provide drivers choices on whether to stay in the traffic for an expected short duration, or how long to delay their travel based on a prediction of when the accident will be cleared.

# Accidents by Day of Week

Accidents were spread fairly evenly over the five weekdays.  Surprisingly, Tuesday was the highest and Monday was the lowest at 93% of Tuesday's accident volume.

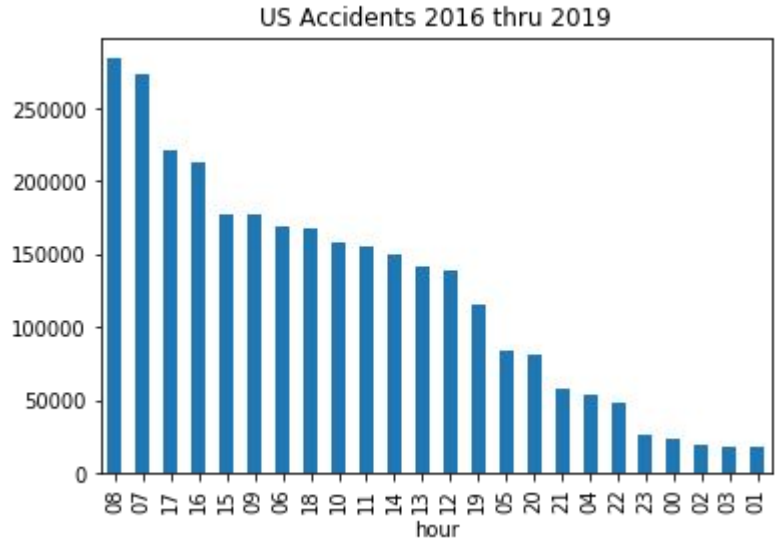
Is it because we are refreshed on Mondays after a weekend, but then stressed after starting a new workweek with less sleep?



# Accidents by Hour

Less surprising, most accidents are occuring during common commuting hours, led by 8 and 7 AM, followed by 5 and 4PM.

Is it sleep related again that morning hours cause 28.5% more accidents than the commute home?

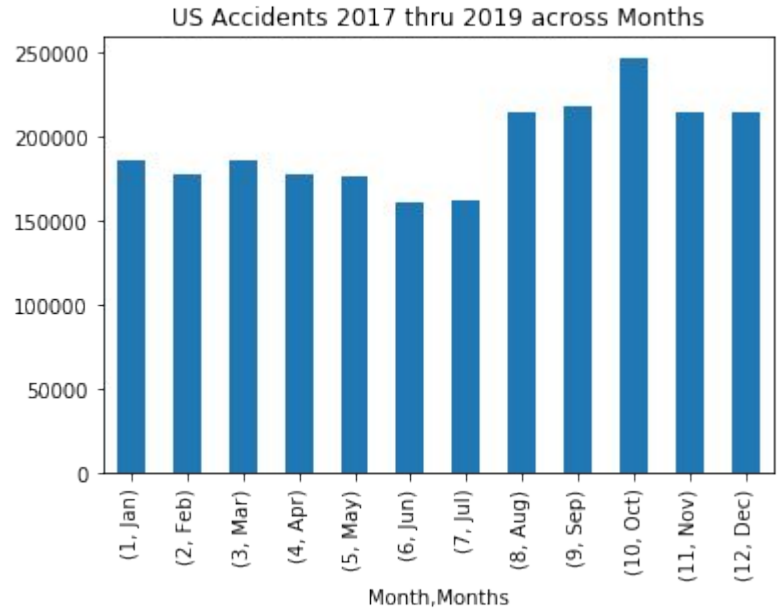


# Accidents by Month

Early to mid summer in the middle of the year have the lowest number of accidents. Key factors may be warm and dry weather, longer periods of sunlight, lighter commuting volume, and no school activity.

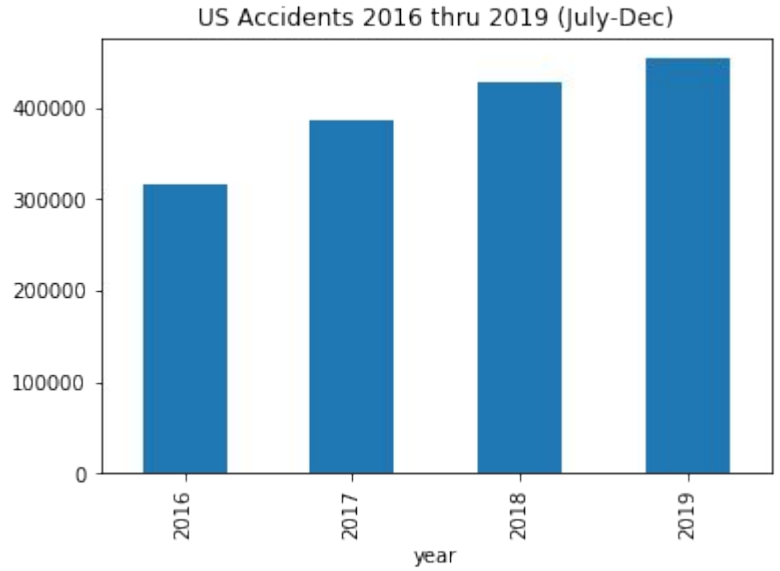
The last five months of the calendar year average 23% higher than the first five months.

October stands out as the highest month for accidents. October is 11% of total yearly accidents, as well as a 15% increase over surrounding months. Contributing factors may be adjusting to the darker commute hours, Columbus Day weekend travel, and pedestrian accidents during Halloween.



# Accidents by Year

A visible upward trend in yearly accident counts. This dataset only partially collected the first half of 2016, all years are comparing the second half, only. One theory is the increase in distracted driving due to smartphone use.



# Can Weather Predictably Impact Traffic Severity from Accidents?

A correlation between weather metrics and severity of accident traffic is not immediately apparent.

The first three rows of Severity (a scale of 1-4 identified by the dataset creators), Duration, and Distance have essentially zero correlation with weather metrics. There is a low correlation between accident traffic Distance (length) & Duration (time), as well as Severity & Distance.

| | Sev | Dur | Dist | Temp | WChill | Hum | Pres | Vis | WSpd | Precip |
|---|---|---|---|---|---|---|---|---|---|---|
| Severity | 1 | 0.03 | 0.18 | -0.02 | -0.03 | 0.02 | 0.04 | -0.01 | 0.03 | 0.03 |
| Duration | 0.03 | 1 | 0.26 | 0.00 | -0.02 | -0.03 | 0.02 | 0.02 | 0.04 | 0.03 |
| Distance | 0.18 | 0.26 | 1 | -0.05 | -0.05 | 0.02 | 0.01 | -0.01 | 0.03 | 0.02 |
| Temp. | -0.02 | 0.00 | -0.05 | 1 | 0.83 | -0.33 | -0.21 | 0.21 | 0.00 | 0.06 |
| Wind_Chill | -0.03 | -0.02 | -0.05 | 0.83 | 1 | -0.14 | -0.27 | 0.15 | -0.11 | 0.04 |
| Humidity | 0.02 | -0.03 | 0.02 | -0.33 | -0.14 | 1 | 0.03 | -0.41 | -0.16 | 0.11 |
| Pressure | 0.04 | 0.02 | 0.01 | -0.21 | -0.27 | 0.03 | 1 | 0.04 | -0.01 | 0.06 |
| Visibility | -0.01 | 0.02 | -0.01 | 0.21 | 0.15 | -0.41 | 0.04 | 1 | 0.03 | -0.12 |
| Wind_Speed | 0.03 | 0.04 | 0.03 | 0.00 | -0.11 | -0.16 | -0.01 | 0.03 | 1 | 0.04 |
| Precip. | 0.03 | 0.03 | 0.02 | 0.06 | 0.04 | 0.11 | 0.06 | -0.12 | 0.04 | 1 |

The strongest correlation at 0.83 is seen between Temperature and Wind Chill. Also intuitive is a moderate, negative correlation between Visibility and Humidity, such as difficulty driving through fog. Another interesting point lies in the moderate, negative correlation between Temperature and Humidity. Science websites note the warmer air becomes, the more water vapor it can hold.