Weather Impact on Accident Traffic Severity

Springboard Capstone 1
Greg Gibson 2020

Overall, according to INRIX, a transportation data firm, the average American spent close to 100 hours in traffic in 2019, losing nearly $1,400 in fuel.

Moreover, INRIX noted that from 2017 to 2019 the yearly average time Americans lost to traffic increased by two hours.

Is there a possible ‘quick win’ to make any sort of reduction?

Can Weather Predict Traffic Severity from Accidents?

Traffic apps and GPS units can measure how long it takes to move through this:

Can data be used to predict how long the mess will be there?

How long should I delay leaving?

The Data

The main source of data is in one file located at:

https://smoosavi.org/datasets/us_accidents

"A Countrywide Traffic Accident Dataset"

Moosavi, Samavatian, Parthasarathy, Ramnath

Nearly 3 Million Accident Records

Across 49 States

2016 through 2019

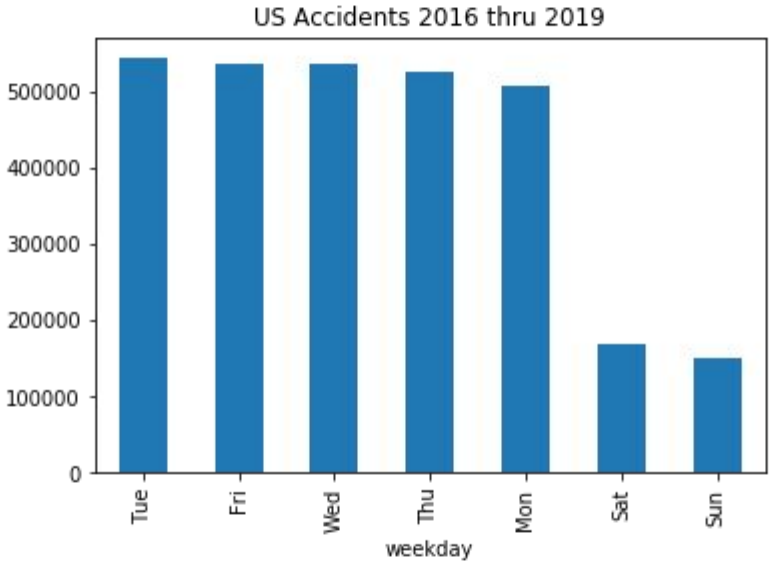
48 Features Including 9 Describing the Weather

Collected from Departments of Transportation, Law Enforcement Agencies, Traffic Cameras and Traffic Sensors

Data Summaries: Accidents by Weekday

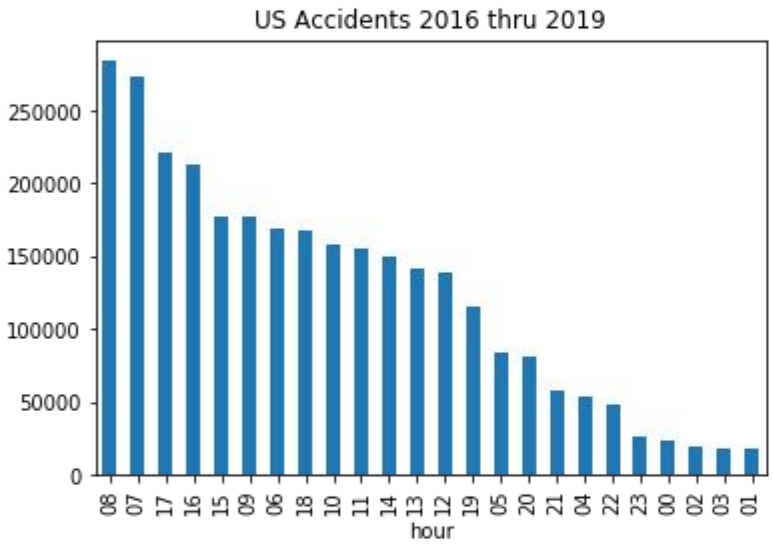
Accidents were spread fairly evenly over the five typical workdays.

Substantially higher than weekend accidents.



Data Summaries: Accidents by Hour

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

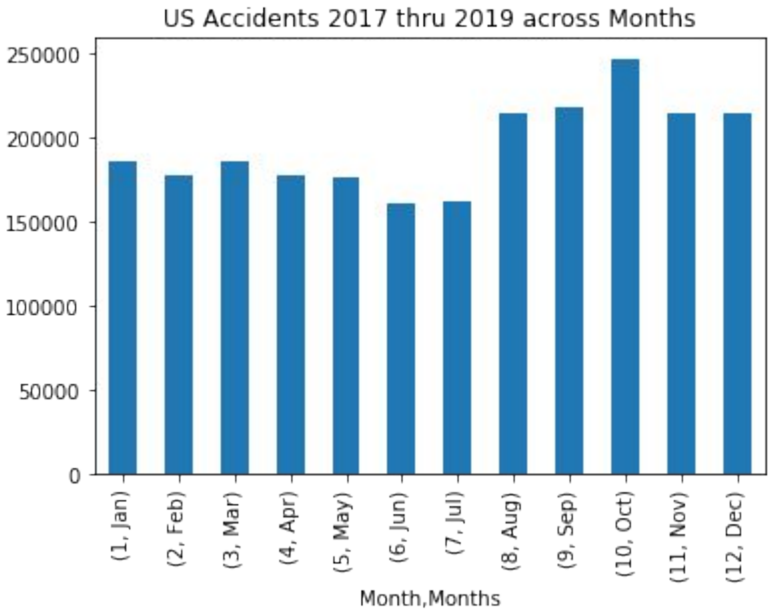


Data Summaries: Accidents by Month

Mid-year has the fewest traffic accidents.

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

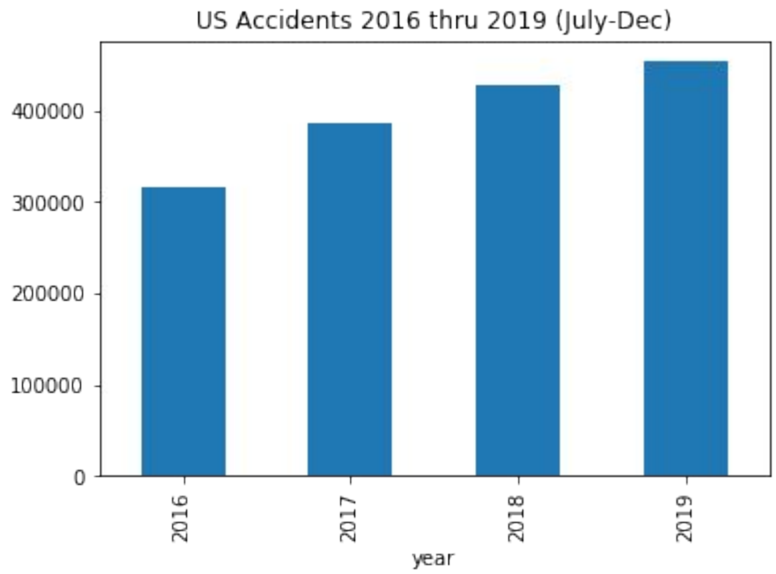
October is the highest, which includes darker morning commute hours, Columbus Day weekend travel, and Halloween.



Data Summaries: 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.



Can Weather Predict Traffic Severity?

	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

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

Simple Classification

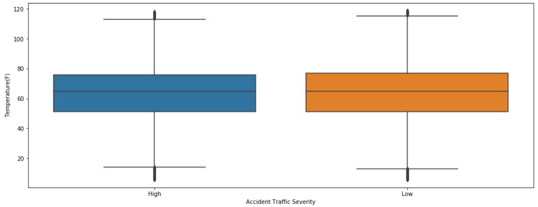
Split the 4 levels of Severity into only Low and High

Compare weather features between Low and High

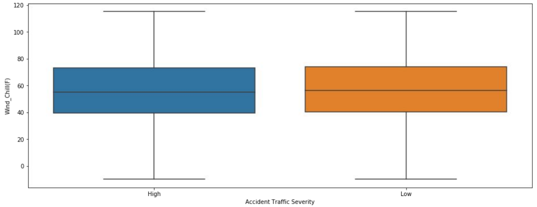
Train classifiers to predict Low or High severity traffic

Class Comparisons

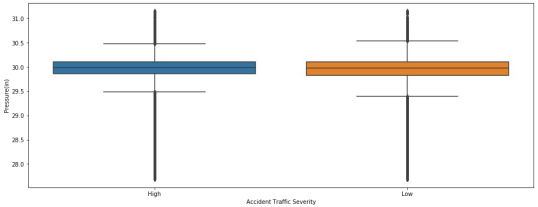
Temperature

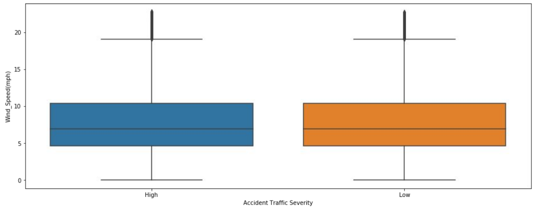


Wind Chill

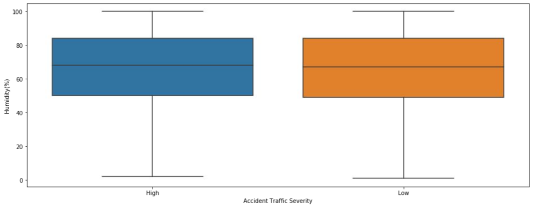


Pressure

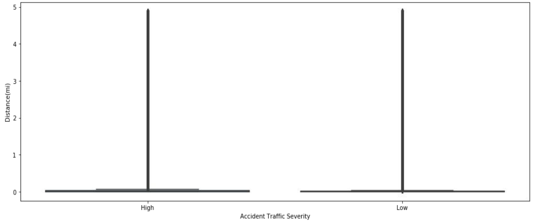




Wind Speed



Humidity



Distance

<u>Visually, all of the weather features appear nearly identical High vs Low!</u>

Logistic Regression Classifier

Throwing all the features in to see what sticks, and it is not the weather at the top!

There's a mix: road features such traffic signals and crosswalks, length of traffic, day, time, and a description "clear".

The line will be drawn at coefficients of at least +/- 0.01 away from zero, retaining 1/3rd of the features

	coef
Traffic_Signal	-0.135802
Distance(Mi)	0.100135
Distancezero	-0.073732
Crossing	-0.061685
Daynight01	0.041559
Weekday_Sat	0.036909
Weekday_Sun	0.032850
Junction	0.032669
Precipitationzero	-0.026913
Hour_8	-0.021010
Pressure(In)	-0.020345
Weather_Group_Clear	-0.017879
Hour_7	-0.017450
Weekday_Tue	-0.017421
Weekday_Wed	-0.017111
Weekday_Mon	-0.015440
Wind_Speed(Mph)	0.014852
Station	-0.013350
Stop	-0.013228
Weekday_Thu	-0.012335

Logistic Regression Classifier II

Confusion Matrix

		Actual	
		Low	High
Predicted	Low	94.3%	5.7%
	High	84.0%	16.0%

Well, that's no good!

How did we do with our top features? *Not well!*

When there was "High" severity traffic after an accident, the classifier only found it 16% of the time.

If this were an application, it would generally tell motorists the traffic would dissipate soon, when in actuality it would be a long delay!

Random Forest Classifier

The Random Forest classifier placed more importance on the weather features.

Confusion Matrix

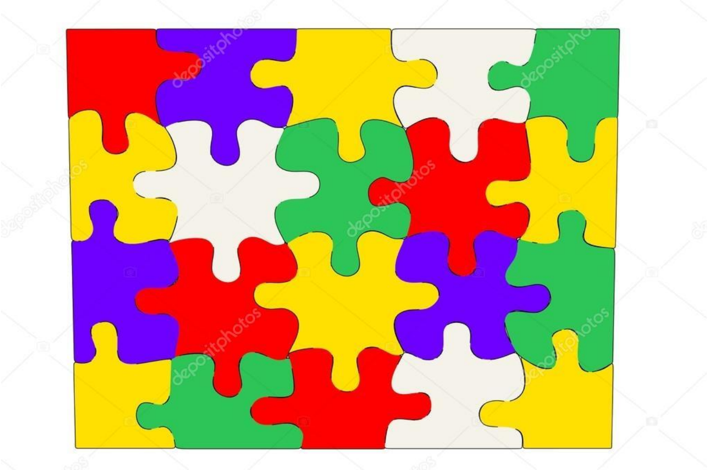
		Actual	
		Low	High
Predicted	Low	64%	36%
	High	27%	73%

This classifier also improved predicting High severity traffic to 73%.

	importance
Duration(m)	14%
Wind_Chill(F)	11%
Pressure(in)	10%
Temperature(F)	10%
Humidity(%)	10%
Wind_Speed(mph)	8%
Distance(mi)	5%
Precipitation(in)	5%
Traffic_Signal	5%
Visibility(mi)	2%
Crossing	2%

Conclusion

After a traffic accident occurs, it is feasible to apply weather conditions to differentiate whether resulting traffic will be severe or not. The added information would be beneficial to motorists to update their travel plans. Such warnings could be sent as alert messages to smartphones and GPS devices, not only by travel applications but also from weather applications which utilize users' location.



This novice exercise with binary classifications showed a potential with nearly 75% accuracy identifying high severity traffic. Other predictive machine learning tools and additional information on roadways, i.e. number of lanes, average speed, volume, etc., should generate more consistent notifications.