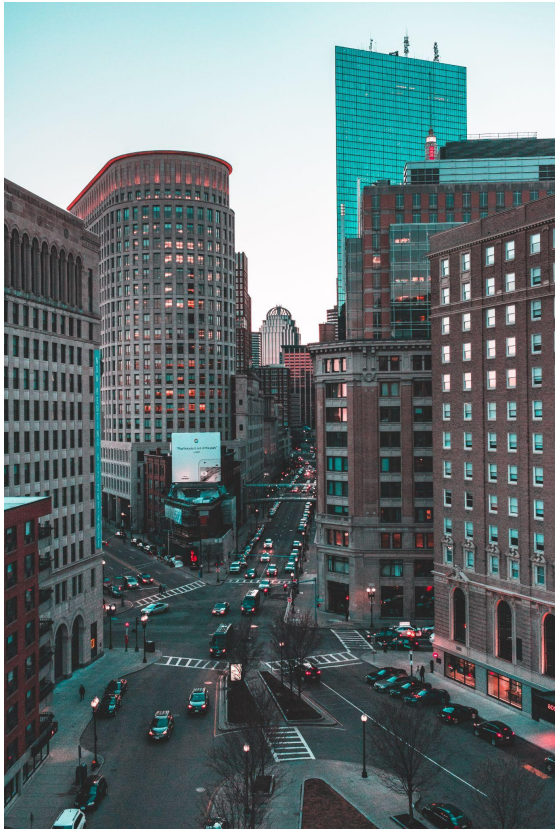
How to Improve the Safety of Driving in MA

Team6: Kaiyu Wang, Jacinto Lemarroy Pipper, Molly Oyer, Yesol(Sally) Lee, Peixuan Liu

Business Problem

- Helping the Massachusetts Department of Transportation analyze car accidents.
- Analyzing the relationship between the frequency and severity of car accidents in relation to various factors
- Understanding the patterns of car crashes in MA
- Joining with another covid dataset to find out the influence of covid on car crashes.



Overview of Datasets

US-Accidents (https://www.kaggle.com/sobhanmoosavi/us-accidents)

- “A Countrywide Traffic Accident Dataset.” from 2019 by Moosavi and others
- “Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights." from 219 by Moosavi and others.

COVID-open-data (https://github.com/open-covid-19/data)

- The title of the data set is "COVID-19 Open-Data: curating a fine-grained, global-scale data repository for SARS-CoV-2" and the site lists "O. Wahltinez and others" as significant contributors.

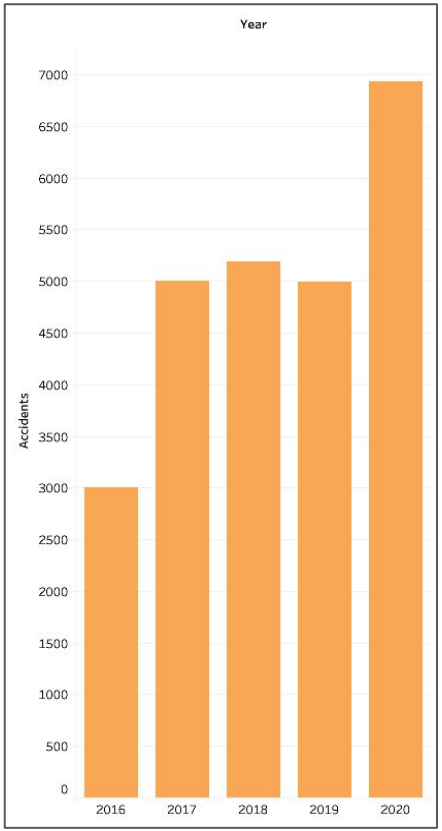
	ID	Severity	Start_Time	End_Time	Start_Lng	Start_Lng	End_Lat	End_Lng	Distance_mi_	Description	...	Roundabout	Station	Stop	Traffic_Calming	Traffic_Signal	Turning_Loop	Sunrise_Sunset	Civil_Twilight	Nautical_Twilight	Astronomical_Twilight
0	A-38249	2	2016-09-01 09:14:56+00:00	2016-09-01 15:14:56+00:00	41.03674	-73.67549	41.08345	-73.6653	3.271	Between CT-120A/King St/Exit 27 and Round Hill...	...	False	False	False	False	False	False	nan	nan	nan	nan
1	A-1251601	2	2017-10-09 15:46:08+00:00	2017-10-09 21:46:08+00:00	41.03674	-73.67549	41.08345	-73.6653	3.271	Between CT-120A/King St/Exit 27 and Round Hill...	...	False	False	False	False	False	False	nan	nan	nan	nan
2	A-1506619	2	2016-09-16 07:21:23+00:00	2016-09-16 13:21:23+00:00	41.03674	-73.67549	41.08345	-73.6653	3.271	Between CT-120A/King St/Exit 27 and Round Hill...	...	False	False	False	False	False	False	nan	nan	nan	nan

3 rows × 47 columns

	location_key	date	place_id	wikidata_id	datacommons_id	country_code	country_name	iso_3166_1_alpha_2	iso_3166_1_alpha_3	aggregation_level	...	cumulative_vaccine_doses_administered_pfizer	new_persons_fully_vaccinated_moderna	cumulative_persons_fully_vaccinated_moderr
0	ES_CT	2021-04-21	ChIJ8_UwhdxbpBIRUMijleD6AAE	Q5705	nuts/ES51	ES	Spain	ES	ESP	1	...	NaN	NaN	Na
1	ES_CT	2020-10-03	ChIJ8_UwhdxbpBIRUMijleD6AAE	Q5705	nuts/ES51	ES	Spain	ES	ESP	1	...	NaN	NaN	Na
2	ES_CT	2021-04-26	ChIJ8_UwhdxbpBIRUMijleD6AAE	Q5705	nuts/ES51	ES	Spain	ES	ESP	1	...	NaN	NaN	Na

3 rows × 701 columns

Overview: Total Car Crashes between 2016-2020 in MA

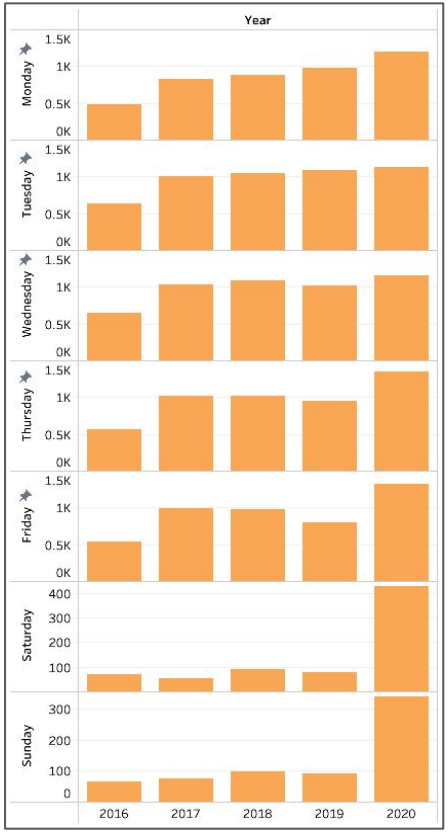


Year	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
2016	484	637	651	558	538	72	66
2017	821	1,006	1,038	1,018	995	52	75
2018	886	1,045	1,082	1,014	977	91	97
2019	971	1,091	1,014	947	793	80	92
2020	1,193	1,131	1,161	1,354	1,325	427	339

▲ The number of accidents by year and day of week (table)

◂ Total number of accidents by year

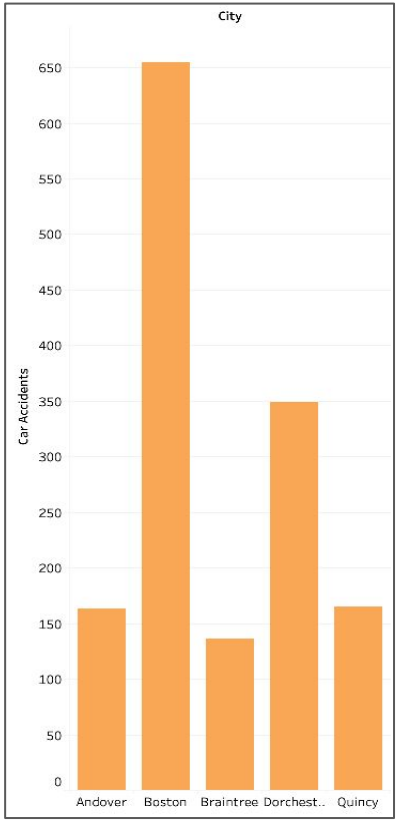
The number of accidents by year and day of week (graph) ►



Car Accidents in 2020

Street	Car_Accidents
I-93 N	459
I-93 S	428
I-95 S	298
I-95 N	266
I-495 S	244

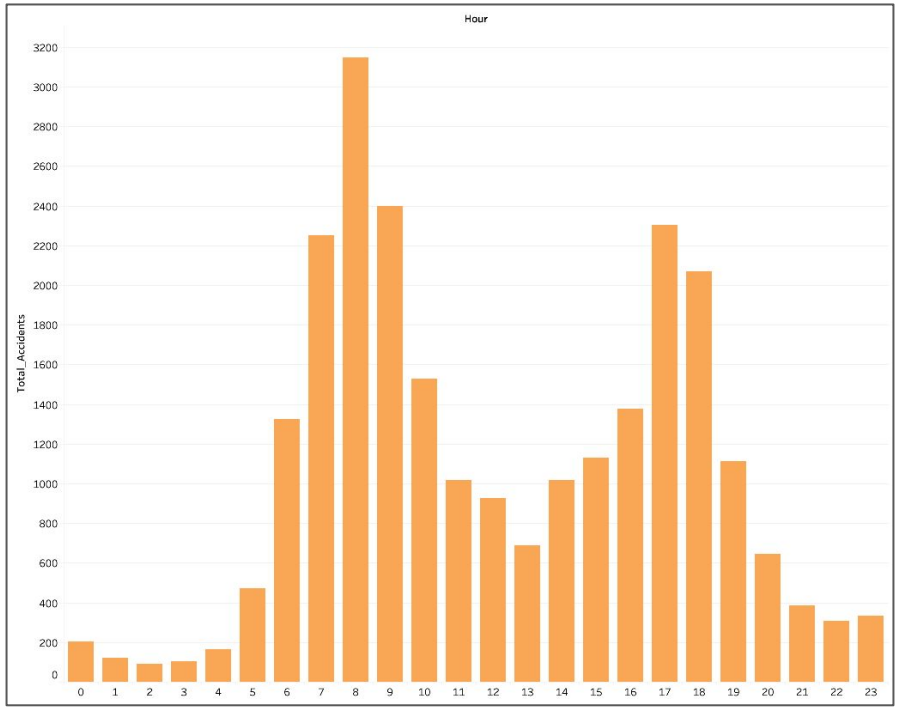
▲ Top 5 streets with most car accidents



▲ Top 5 cities with most car accidents

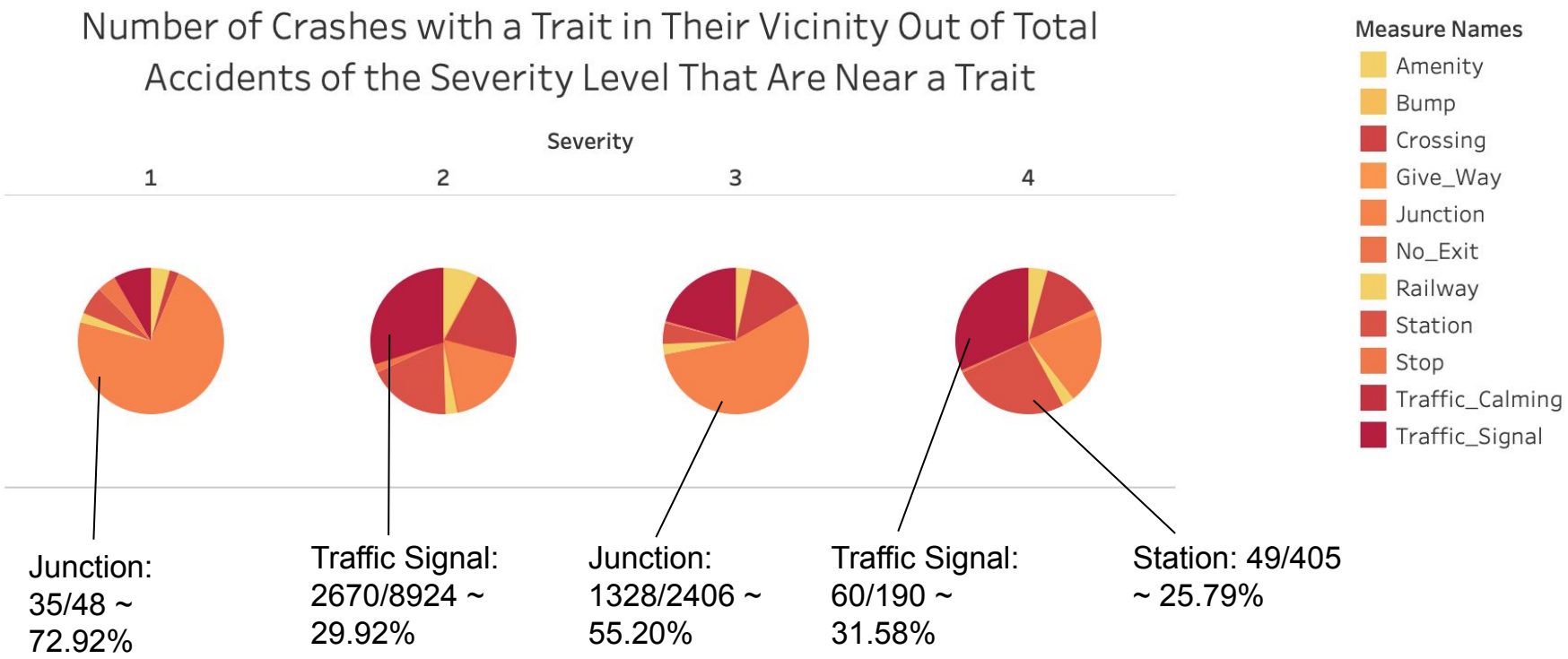
Car Accidents and Time period

Total Accidents	Morning %	Afternoon %	Evening %
25,121	51	30	19



▲ Number of car accidents by hour

How Severe are Accidents Near Each Trait?



Can You Always Trust Your Data?

	Amenity	Bump	Crossing	Give_Way	Junction	No_Exit	Railway	Roundabout	Station	Stop	Traffic_Calming	Traffic_Signal	Turning_Loop
0	794	3	2199	46	2987	8	290	0	1807	200	4	3230	0

- The above table contains the results of counting every time an accident is in the vicinity of each trait
- According to this query, Massachusetts has *NO* accidents at roundabouts
- In Massachusetts, *‘roundabouts’* are referred to as *‘rotaries’*, so these accidents may not be properly accounted for

What Can Go Wrong?

```
%%bigquery

SELECT COUNT(*) num_rotaries

FROM
`ba775-b2-team6.team6.carcrash_MA`

WHERE Street LIKE '%Rotary%'
```

	num_rotaries
0	36

- When we count how many crashes occur on a street including the name ‘Rotary’, we see there were 36 such accidents
- There also could be accidents not taking place *in* the rotary, but taking place *near* it - accidents like this are unaccounted for in num_rotaries
- If something seems unlikely, it may be an error

Car Accidents and Weather Conditions

Part1. Temperature, Wind Speed and Humidity

Average weather conditions

Month of Start_Time	Avg. Humidi..	Avg. Temperatur..	Avg. Wind_S..	Avg. Precipitatio..	Avg. Visibility_..
January	69.80	33.21	9.47	0.007	7.61
February	73.89	34.88	8.65	0.011	6.04
March	72.88	40.86	11.15	0.016	6.75
April	68.08	45.36	12.23	0.017	7.52
May	63.38	58.45	9.76	0.007	9.05
June	64.93	70.75	8.00	0.009	8.92
July	68.57	75.65	7.27	0.022	8.89
August	66.08	74.71	7.83	0.009	9.31
September	72.46	66.28	8.56	0.011	8.63
October	76.37	55.98	8.53	0.011	8.00
November	71.85	45.08	8.83	0.016	8.10
December	73.03	34.89	8.66	0.007	7.53

Coefficient of Correlation

Temperature
0.0969
Wind Speed
-0.6258
Humidity
0.6337

Car Accidents and Weather Conditions

Part1. Temperature, Wind Speed and Humidity

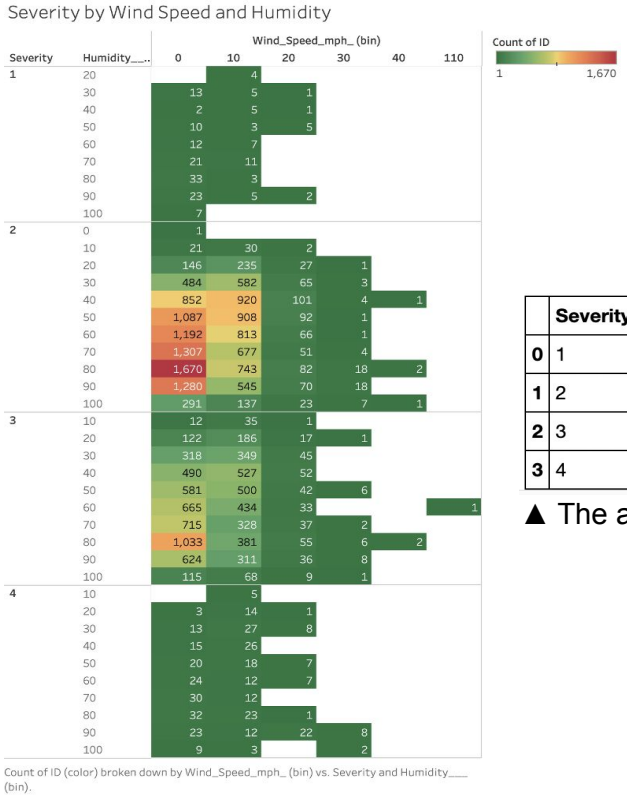
Wind Speed

	Interval_start	Interval_end	num_accident
0	0	10	13348
1	10	20	8944
2	20	30	967
3	30	40	91
4	40	50	6
5	110	120	1

Humidity

	Interval_start	Interval_end	num_accident
0	0	10	1
1	10	20	107
2	20	30	775
3	30	40	1953
4	40	50	3051
5	50	60	3402
6	60	70	3430
7	70	80	3436
8	80	90	4544
9	90	100	3418
10	100	100	771

Interval start: inclusive Interval end: exclusive



	Severity	avg_humidity	stdev_humidity	avg_wind_speed	stdev_wind_speed
0	1	70.87	20.43	7.25	6.12
1	2	66.80	20.51	9.55	5.43
2	3	65.49	20.99	9.54	5.44
3	4	65.16	23.07	11.59	7.10

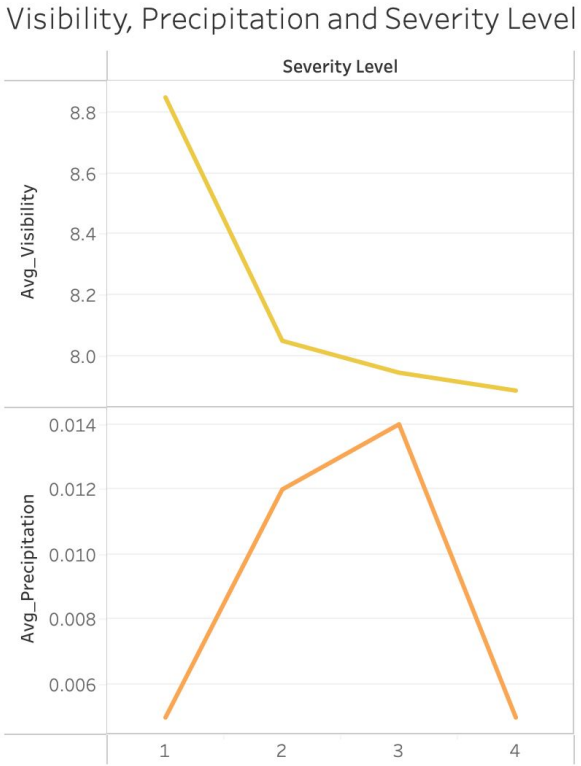
▲ The average and standard deviation of humidity and wind speed by severity of car crash

Car Accidents and Weather Conditions

Part2. Visibility and Precipitation

Severity	Avg_Visibility	Avg_Precipitation
1	8.851	0.005
2	8.049	0.012
3	7.944	0.014
4	7.885	0.005

- The lower the visibility, the more severe the accidents could be.
- When the severity level is less than 4, the larger the precipitation, the more severe the accidents.



Car Accidents and Weather Conditions

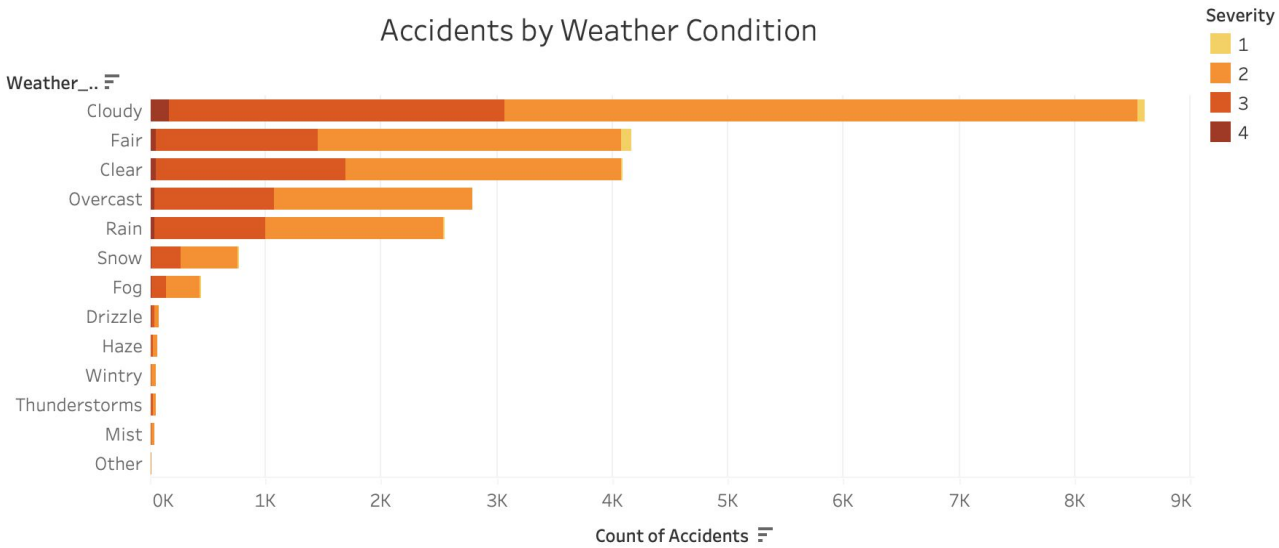
Part3. Weather Conditions

Severity	Count	Fair	Clear	Cloudy	Overcast	Rain	Snow	Fog	Drizzle	Haze	Wintry	Mist	Thunderstorms
1	174	50%	1.72%	31.61%	0%	12.64%	1.72%	2.3%	0%	0%	0%	0%	0%
2	14698	17.92%	16.23%	37.34%	11.63%	10.46%	3.28%	2%	0.35%	0.27%	0.23%	0.12%	0.28%
3	8400	16.62%	19.55%	34.45%	12.35%	11.39%	3.07%	1.49%	0.21%	0.27%	0.17%	0.17%	0.37%
4	388	13.14%	13.66%	43.04%	9.54%	10.05%	3.09%	2.06%	4.64%	0%	0.26%	0.26%	0.26%

- Most of the accidents have a severity level of 2.
- When the severity level is 1, ‘Fair’ is the most common weather condition (50%). For accidents of higher severity levels, ‘Cloudy’ is the most common one.

Car Accidents and Weather Conditions

Part3. Weather Conditions



Count	Weather_Condition
8604	Cloudy
4168	Fair
4083	Clear
2784	Overcast
2555	Rain
755	Snow
431	Fog
82	Drizzle
62	Haze
49	Wintry
46	Thunderstorms
32	Mist

- Based on the results, most of the accidents occur in **‘Cloudy’** weather.
- However, as different weather conditions have different distribution and frequency, we can't conclude that ‘Cloudy’ weather is more likely to cause accidents than other weather.

Car Accidents and Covid 19

	month	monthly_carcrash_number_2020	monthly_carcrash_number_average	difference_percentage	new_confirmed
0	January	454	384	18.23	0
1	February	445	361	23.27	0
2	March	643	309	108.09	6620
3	April	999	264	278.41	55585
4	May	842	309	172.49	34760
5	June	843	326	158.59	11917
6	July	340	351	-3.13	8730
7	August	357	423	-15.60	10921
8	September	417	405	2.96	3583
9	October	506	492	2.85	26460
10	November	560	491	14.05	67556
11	December	524	435	20.46	149046

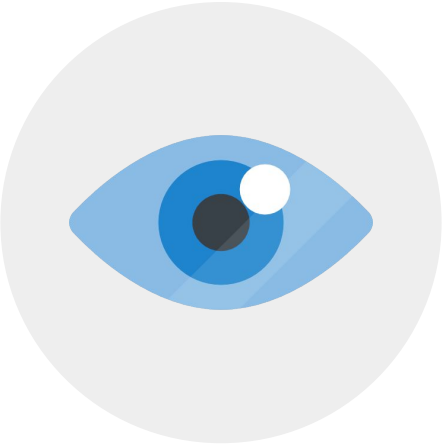
Car Accidents and Covid 19

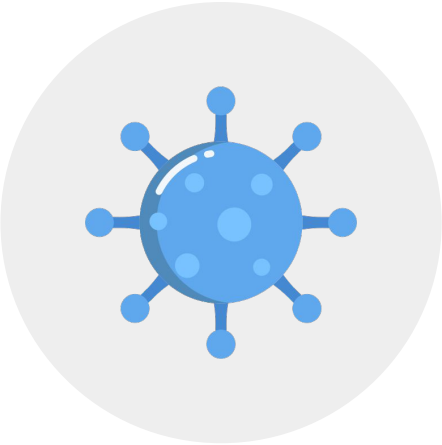
	week	weekly_carcrash_number_2020	weekly_carcrash_number_average	difference_percentage	new_confirmed
0	10	88	73	20.55	125
1	11	179	69	159.42	387
2	12	217	63	244.44	3732
3	13	225	67	235.82	7479
4	14	183	62	195.16	11124

	Slope	Intercept	Correlation
0	0.009114	104.97634	0.652399

Conclusion

- The most common time for a car crash is during commuting hours.
- The lower the visibility, the more severe the car accidents could be.
- COVID led to less cars and faster driving, leading to more accidents.





Suggestions

- More State Police on the roads during commuting hours.
- Additional road screens to remind drivers of speed limit, etc., when the weather conditions are abnormal.
- The government could improve the drainage of roads to decrease flooding.

Restrictions

- Filtering out NaN values could cause the loss of information and impact the accuracy of the results.
- Additional datasets of weather distribution in MA and number of cars on the road would be useful for further analysis.
- U.S. Accidents dataset does not contain a variable for the reasons for car crashes, and the descriptions of vicinity are vague and had omissions.
- Since streets have different lengths, making a comparison is biased.

Thank you for your attention