



US National Traffic Accidents (2016-2023)

MSIS 2607 - Group 3: Aryan Puranik, Connor Reardon, Justin Hsu, Kush Shah

Introduction



Dataset

Data Source Kaggle
[US Accidents](#)
[\(2016 - 2023\)](#)

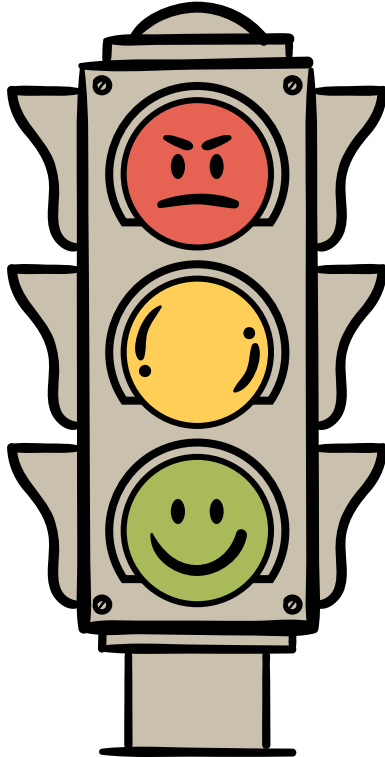
Code

Cleaning Data, Data
Transformation, and
Machine Learning

Managerial Recommendation

Validating
Hypothesis and
Insights

Questions We are Exploring



Question 1

Were there negative impacts of the COVID-19 lockdown on traffic accidents?



Question 2

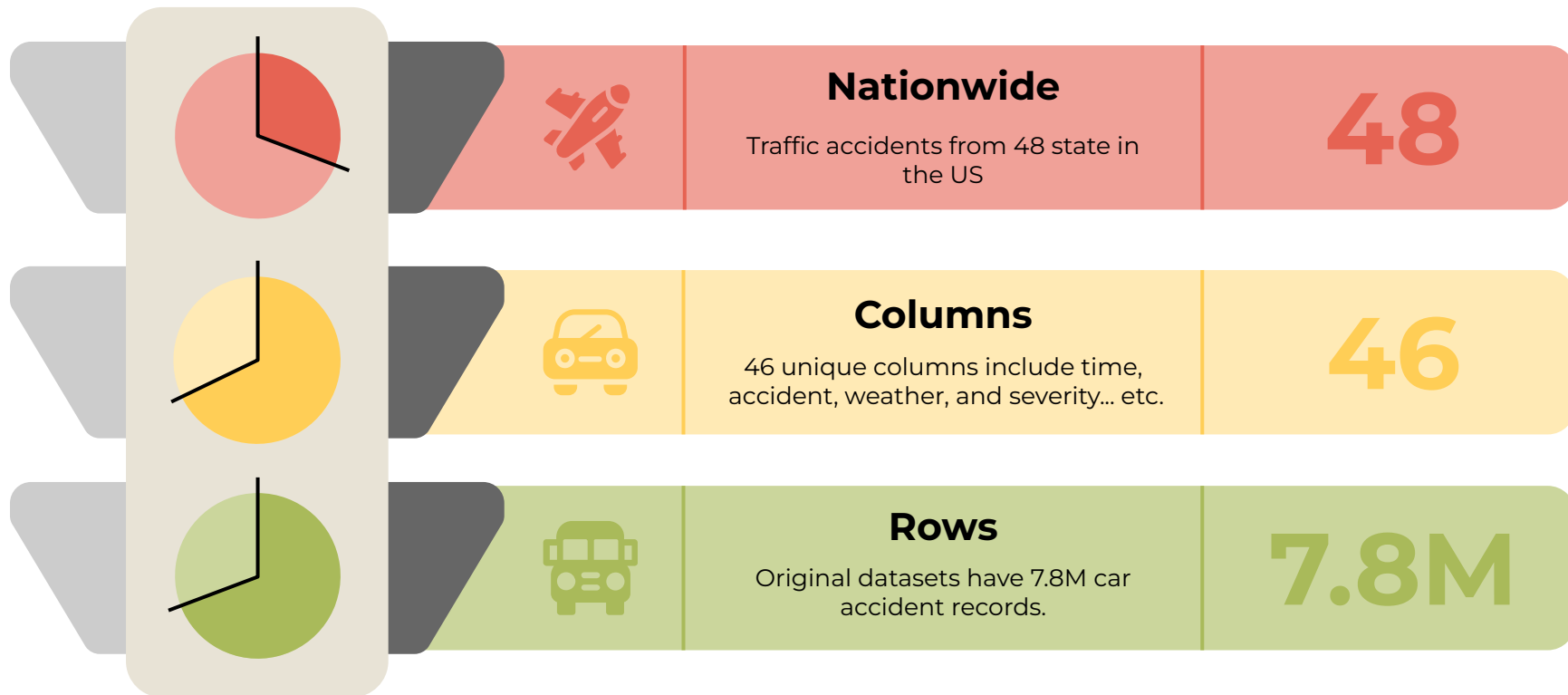
What conditions influence accidents? Where do the most accidents happen in the US?



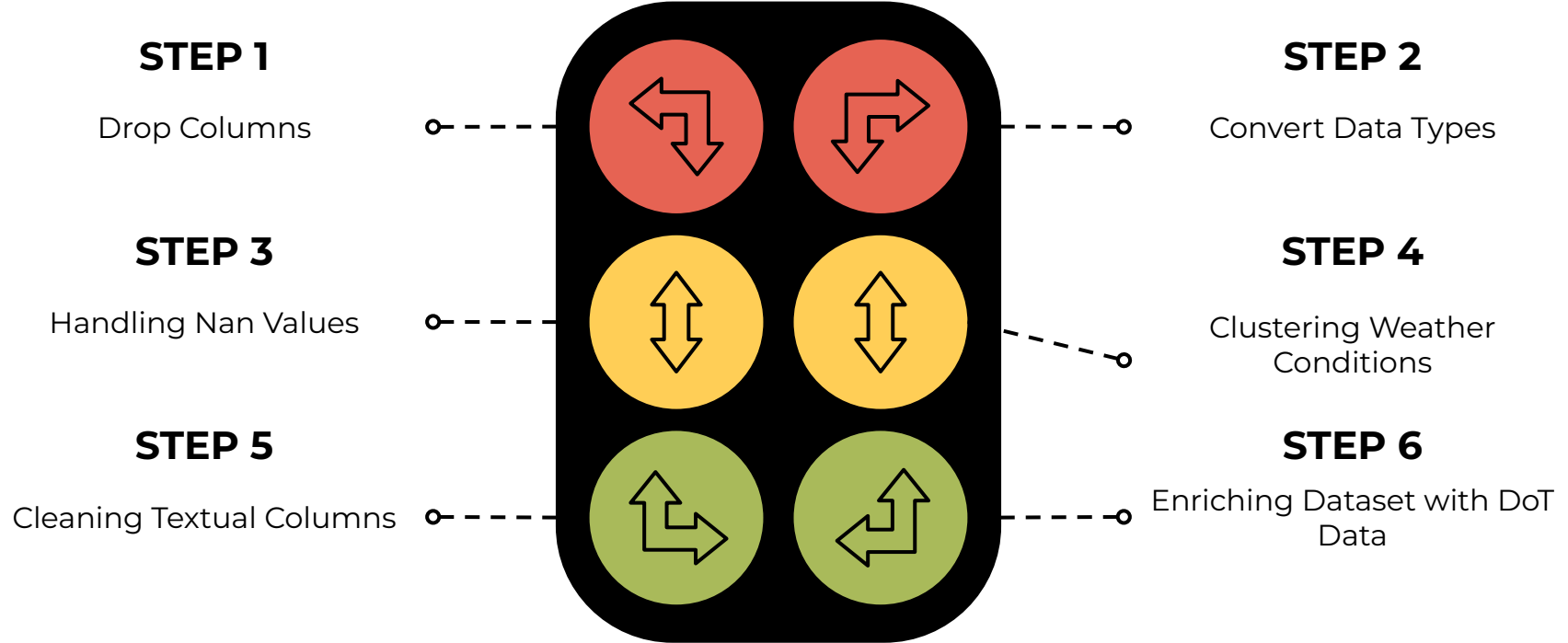
Question 3

Are Californians worse drivers than average? Which state has the best and worst drivers?

Kaggle - US Accidents (2016-2023) Dataset



Steps on Data Cleaning



Dataset Overview

General Information	ID, Source, Severity
GEO Data	Start_Time, End_Time, Start_Lat, Start_Lng, End_Lat, End_Lng, Distance(mi), Description, Street, City, County, State, Zipcode, Country, Timezone, Airport_Code
Weather Data	Weather_Timestamp, Temperature(F), Wind_Chill(F), Humidity(%), Pressure(in), Visibility(mi), Wind_Direction, Wind_Speed(mph), Precipitation(in), Weather_Condition
Road Features	Amenity, Bump, Crossing, Give_Way, Junction, No_Exit, Railway, Roundabout, Station, Stop, Traffic_Calming, Traffic_Signal, Turning_Loop
Period of Date	Sunrise_Sunset, Civil_Twilight, Nautical_Twilight, Astronomical_Twilight

*Red are the columns dropped from table.

Data Type Transformation

Column	Old Data Type	New Data Type
Start_Time	Object	DateTime
End_Time	Object	DateTime
Zipcode	Object	Float
Sunrise_Sunset	Object	Boolean



Zip Codes

Certain Zipcodes were in the format xxxx – yyy, manipulated it to xxxx before converting it to float. Dropped Records with NA Values, about 0.02% of the total records.

Sunrise_Sunset

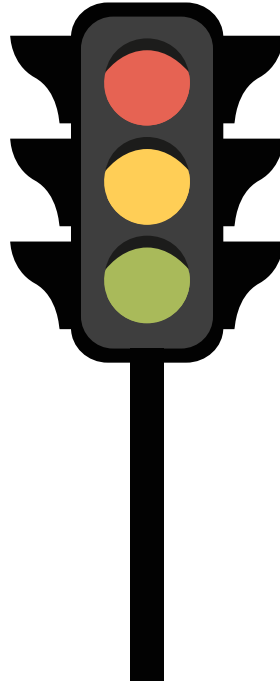
Had only two values, day or night, renamed it as Day, and Assigned 1 for Day and 0 for Night

Handling NAN Values

Columns	Null Values (%)	Manipulation Technique
Street	0.14	Replace with Unknown, to enable analysis on geographic parameters as well.
City	~0	Dropped Rows because of a negligible proportion.
TimeZone	0.10	Dropped Rows because of a negligible proportion.
Airport_code	0.29	Dropped Rows because of a negligible proportion.
Description	~0	Dropped Rows because of a negligible proportion.
Temperature (F)	2.12	Replaced the null with avg value of group based on city, month and day/night, Weather Condition. If still null, deleted.
Wind_Chill (F)	25.86	Replaced the null with avg value of group based on city, month and day/night. If still null, take the average Wind Chill of Weather Condition and replace with null values. If still null, we replace it with Temperature.
Humidity (%)	2.25	Replaced the null with avg value of group based on city, month and day/night. If still null, take the average Humidity of Weather Condition and replace with null values. If still null, drop.
Pressure (in)	1.82	Replaced the null with avg value of group based on city, month and day/night. If still null, take the average Pressure of Weather Condition and replace with null values. If still null, drop.
Visibility (mi)	2.29	Replaced the null with avg value of group based on city, month and day/night. If still null, take the average Visibility of Weather Condition and replace with null values. If still null, drop.
Wind_Speed (mph)	7.39	Replaced the null with avg value of group based on city, month and day/night. If still null, take the average Wind speed of Weather Condition and replace with null values. If still null, drop.
Precipitation (in)	28.5	Replaced the null with avg value of group based on city, month and day/night. If still null, take the average Precipitation of city, Weather Condition and replace with null values. If still null, replace with avg Weather Condition.
Weather_condition	2.24	Dropped Rows because of a negligible proportion.
Day	0.30	Replaces null values with avg time to be called day based on start time.

Cleaning Textual Data

Column
Street
City
State
TimeZone
Airport_Code



Applied the following to the columns:

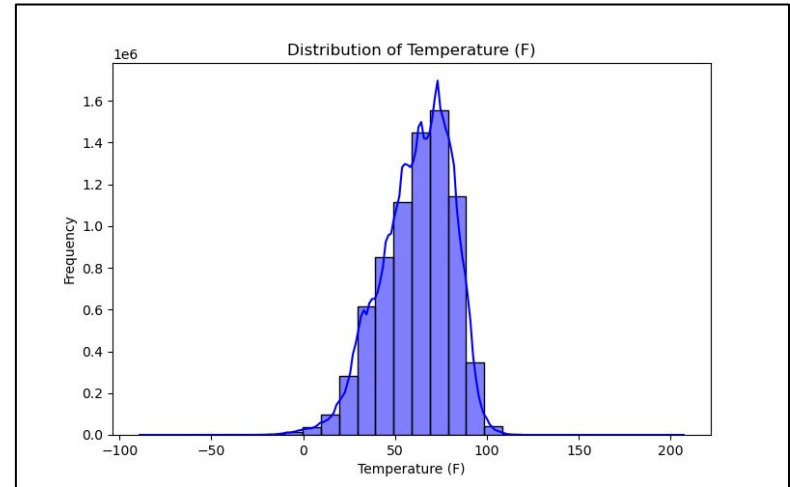
Removed extra white spaces

Applied the following to the columns:

Lower Cased

Identifying Outliers

#	Column	Outlier	Treatment	Supporting Viz
1	Temperature	Minimum: -89 F Maximum: 207 F	Kept only those records with temperature in range -50 to 120 F	Temp_Outlier.png
2	Pressure	Minimum: 0	Pressure can not be practically 0, only kept those greater than 0.	Pressure_Outlier.png
3	Wind Speed	Maximum: 1087 mph	Kept only those records with wind_speed in range [0 to 200]mph	WindSpeed_Outlier.png



Feature Engineering

New Column
Month
Hour_Of_Day
Year
Day_of_Week
Is_weekend
Temperature Category
Distance Category
Humidity Category
Pressure Category
Wind Speed Category
Precipitation Category
Visibility Category



Very Cold : [, -32]
Cold : [-32, 50]
Moderate : [50,70]
Hot : [70,90]
Very Hot : [90,]

Low : [, 5]
Medium : [5,20]
High : [20,]

Low : [,40]
Medium : [40,70]
High : [70,]

Low : [, 29.5]
Medium : [29.5, 30.5]
High : [30.5]

Low : [,10]
Moderate : [10,25]
High : [25,]

None : [, 0]
Light : [0,0.1]
Moderate : [0.1, 1]
Heavy : [1,]

Low Visibility : [, 1]
Moderate Visibility : [1, 3]
High Visibility : [3,]

Clustering Different Weather Conditions

Cluster	Members
Fair	Clear, Fair / Windy, Fair.
Rain	Rain, Light Rain, Heavy Rain, Drizzle, Light Drizzle, Showers in the Vicinity, Heavy Drizzle, Drizzle and Fog, Rain / Windy, Heavy Rain / Windy, Rain Showers, Light Rain Shower, Rain Shower, Heavy Rain Showers, Light Rain / Windy, Rain / Windy, Light Freezing Rain, Heavy Freezing Rain, Freezing Rain, Light Freezing Rain / Windy, Freezing Rain / Windy, Freezing Drizzle, Heavy Freezing Drizzle, Light Freezing Drizzle, Light Drizzle / Windy, Light Rain Showers, Drizzle / Windy, Light Rain Shower / Windy, Heavy Rain Showers, Rain Shower / Windy, Rain and Sleet, Heavy Freezing Rain / Windy, Heavy Rain Shower / Windy.
Cloudy	
Snow	Snow, Light Snow, Snow Showers, Heavy Snow, Sleet, Light Sleet, Snow and Sleet, Light Snow and Sleet, Snow and Sleet / Windy, Heavy Snow with Thunder, Light Snow Shower, Snow and Thunder, Snow and Thunder / Windy, Low Drifting Snow, Drifting Snow / Windy, Blowing Snow, Blowing Snow / Windy, Heavy Blowing Snow, Small Hail, Hail, Heavy Ice Pellets, Light Ice Pellets, Ice Pellets, Heavy Sleet, Heavy Sleet / Windy, Sleet and Thunder, Snow Grains, Light Snow Grains, Light Snow and Sleet / Windy, Light Snow with Thunder, Sleet / Windy, Light Snow Showers, Light Sleet / Windy, Light Snow Shower / Windy, Light Hail, Light Blowing Snow, Blowing Snow Nearby, Drifting Snow.
Fog or Smoke	Fog, Fog / Windy, Mist, Mist / Windy, Light Fog, Patches of Fog, Patches of Fog / Windy, Shallow Fog, Shallow Fog / Windy, Haze, Light Haze, Haze / Windy, Smoke, Light Freezing Fog, Blowing Dust / Windy, Smoke / Windy, Partial Fog, Blowing Sand, Partial Fog / Windy.
Thunderstorm	Thunderstorms, Thunder, T-Storm, Thunderstorm, Light Thunderstorm, Heavy T-Storm, Heavy T-Storm / Windy, Thunder in the Vicinity, Thunderstorms and Rain, Light Thunderstorms and Rain, Heavy Thunderstorms and Rain, Thunder / Windy, T-Storm / Windy, Thunder and Hail, Thunder and Hail / Windy, Thunderstorms and Snow, Heavy Thunderstorms and Snow, Light Thunderstorms and Snow, Thunder / Wintry Mix, Thunder / Wintry Mix / Windy, Light Rain with Thunder, Heavy Thunderstorms with Small Hail.
Extreme	Volcanic Ash, Funnel Cloud, Tornado, Squalls, Squalls / Windy, Sand / Dust Whirls Nearby, Sand / Dust Whirlwinds, Sand / Dust Whirlwinds / Windy, Sand, Sand / Windy, Blowing Dust, Widespread Dust, Widespread Dust / Windy, Duststorm, Dust Whirls, Wintry Mix / Windy, Heavy Smoke, Heavy. .Sleet and Thunder.

Cleaned Dataset

Original No. of Records	7728394
Updated No. of Records	7549676
% dropped	2.31

Data Limitations

Accident Severity Distribution

Class imbalance in severity, likely caused by underreporting of minor accidents (severity one).

Driver Registration

Missing state registration data for drivers. We assume accidents occur primarily between drivers registered in the same state.

Vehicle Information

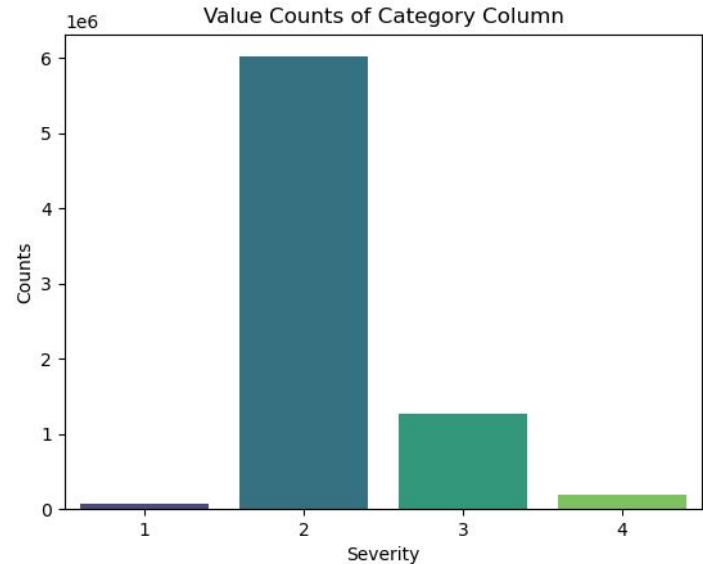
Missing vehicle information, including type, make, and model.

Driver Demographics

Missing driver demographic details, such as age, sex, and race.

Driver Impairment

Lack of impairment data such as DUI.



Q1: Impact of COVID-19 on Traffic Accidents

Did Driving Skills Decline After the COVID-19 Lockdown?

Peak of Summer Surge

The U.S. reached new daily case records, surpassing 70,000 cases per day in mid-July.

State Response

Major states such as California, Texas, and Florida reimplemented shutdown measures.

July- Aug

National accident counts dropped sharply in July and August 2020.

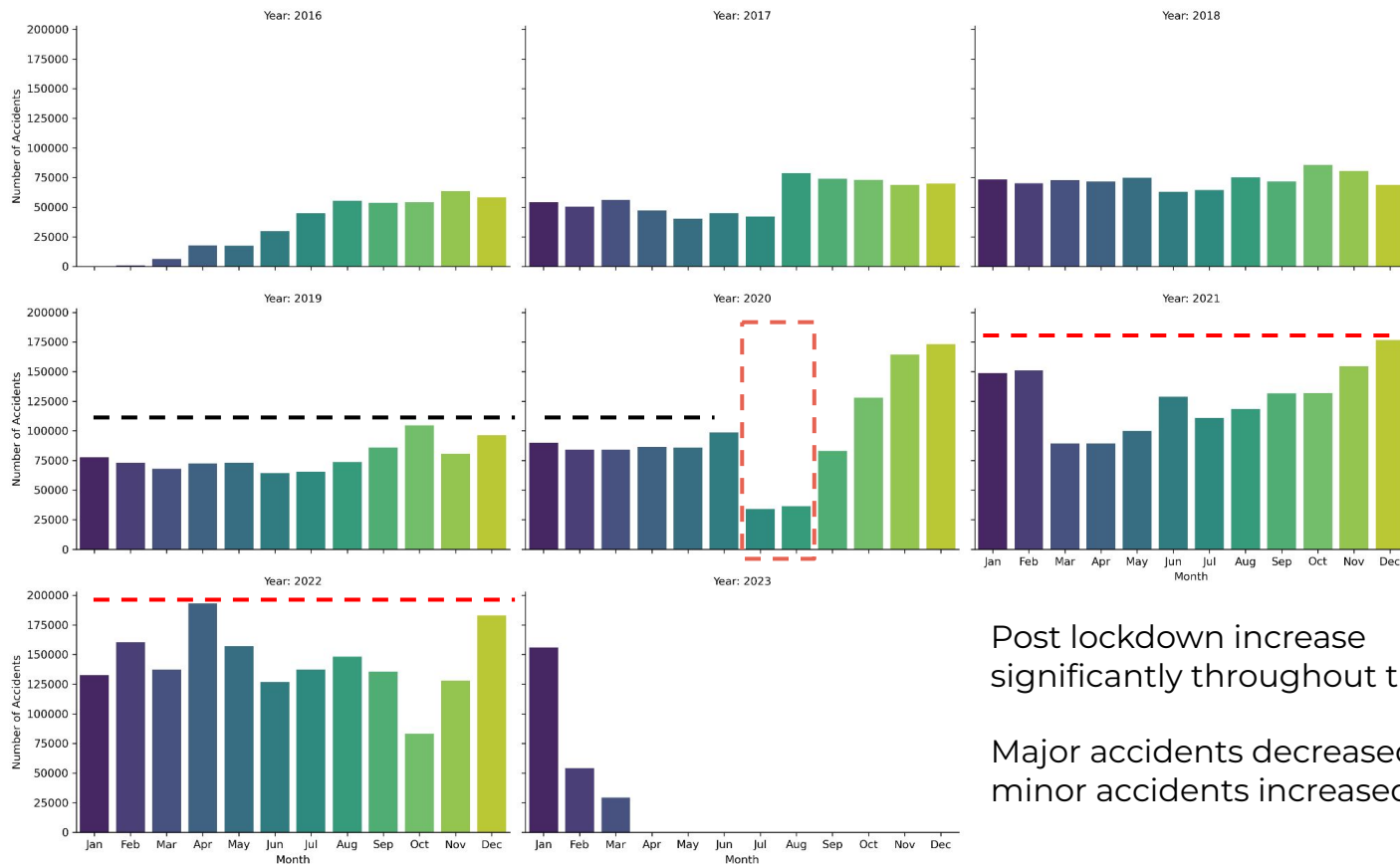
After Covid

The years following 2020, National accident counts are on the rise.



Monthly Accident Counts by Year

Monthly Accident Counts by Year (Shared Scales)



Post lockdown increase significantly throughout the USA.

Major accidents decreased while minor accidents increased.

Accident Severity Description

Minor (Level 1):

- Low-speed incidents with minimal damage, such as fender benders.
- No injuries reported.
- Example: Scraping a parked car or light bumper collision.

Moderate (Level 2):

- Accidents causing noticeable vehicle damage but no serious injuries.
- May involve minor injuries like bruises or cuts.
- Example: Rear-end collisions at moderate speed.

Severe (Level 3):

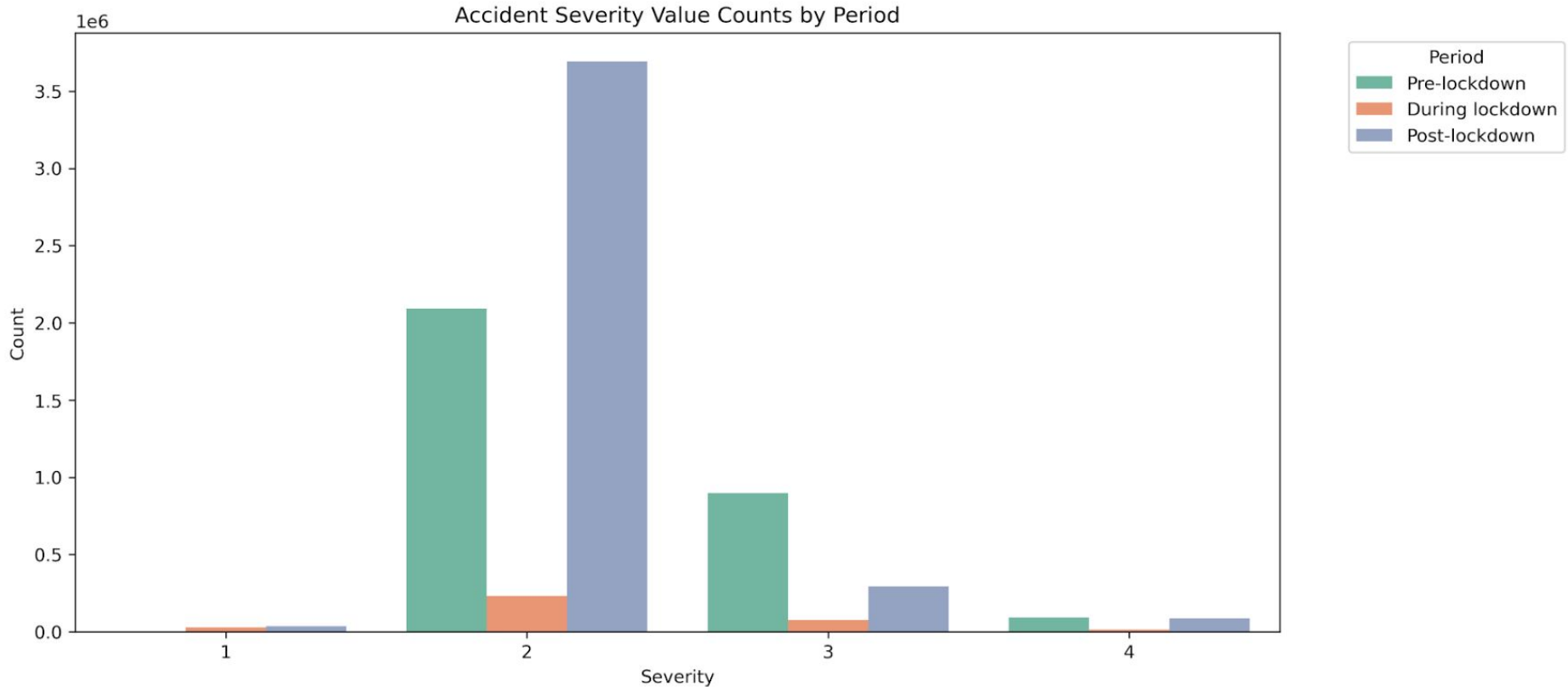
- High-impact accidents with significant vehicle damage.
- Likely involves injuries requiring medical attention.
- Example: Side-impact collisions or crashes on highways.

Critical (Level 4):

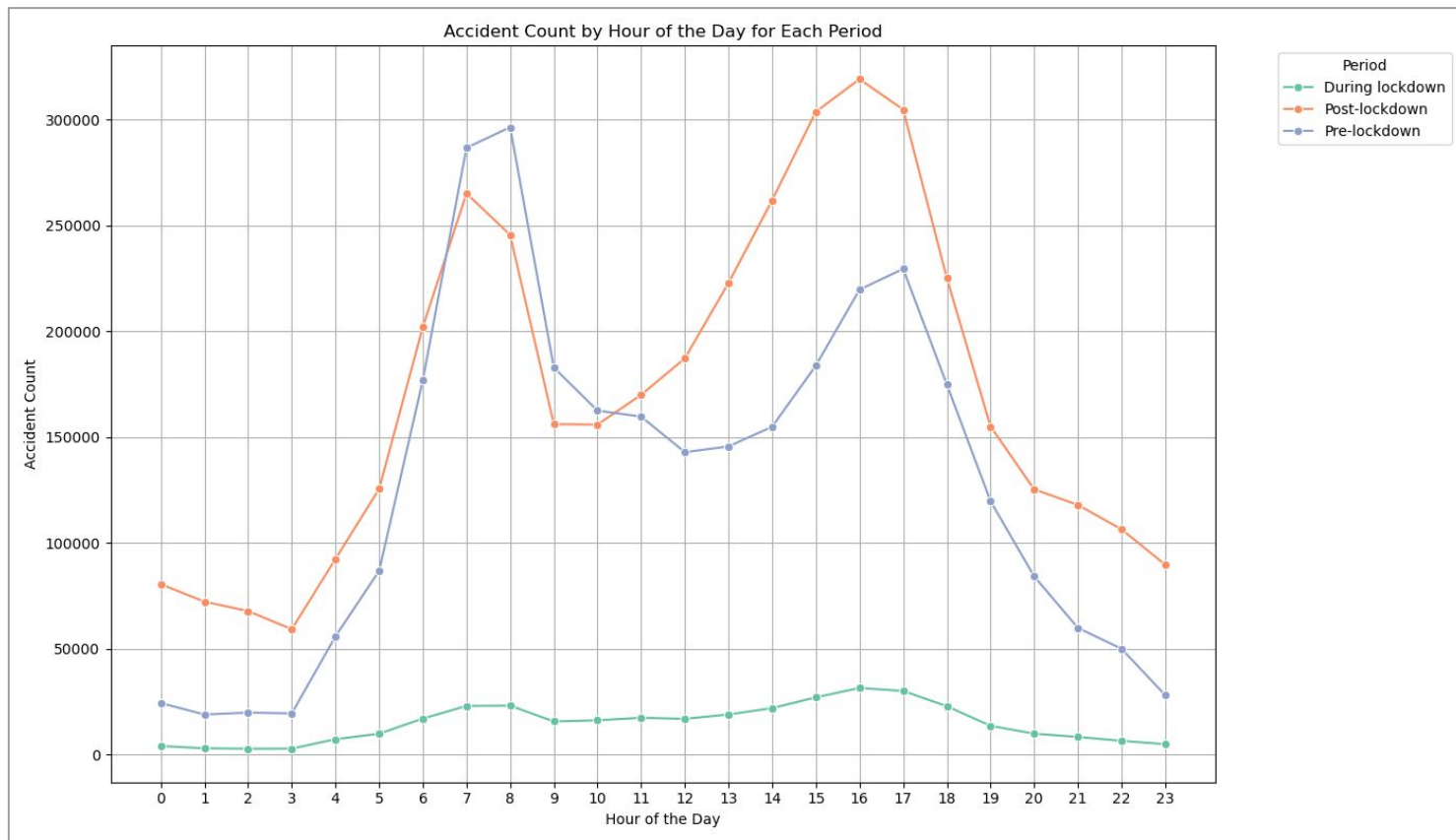
- Major accidents with extensive damage and severe or fatal injuries.
- Often involves multiple vehicles or dangerous conditions.
- Example: High-speed rollovers or head-on collisions.



Pre vs Post Covid Accidents Severity

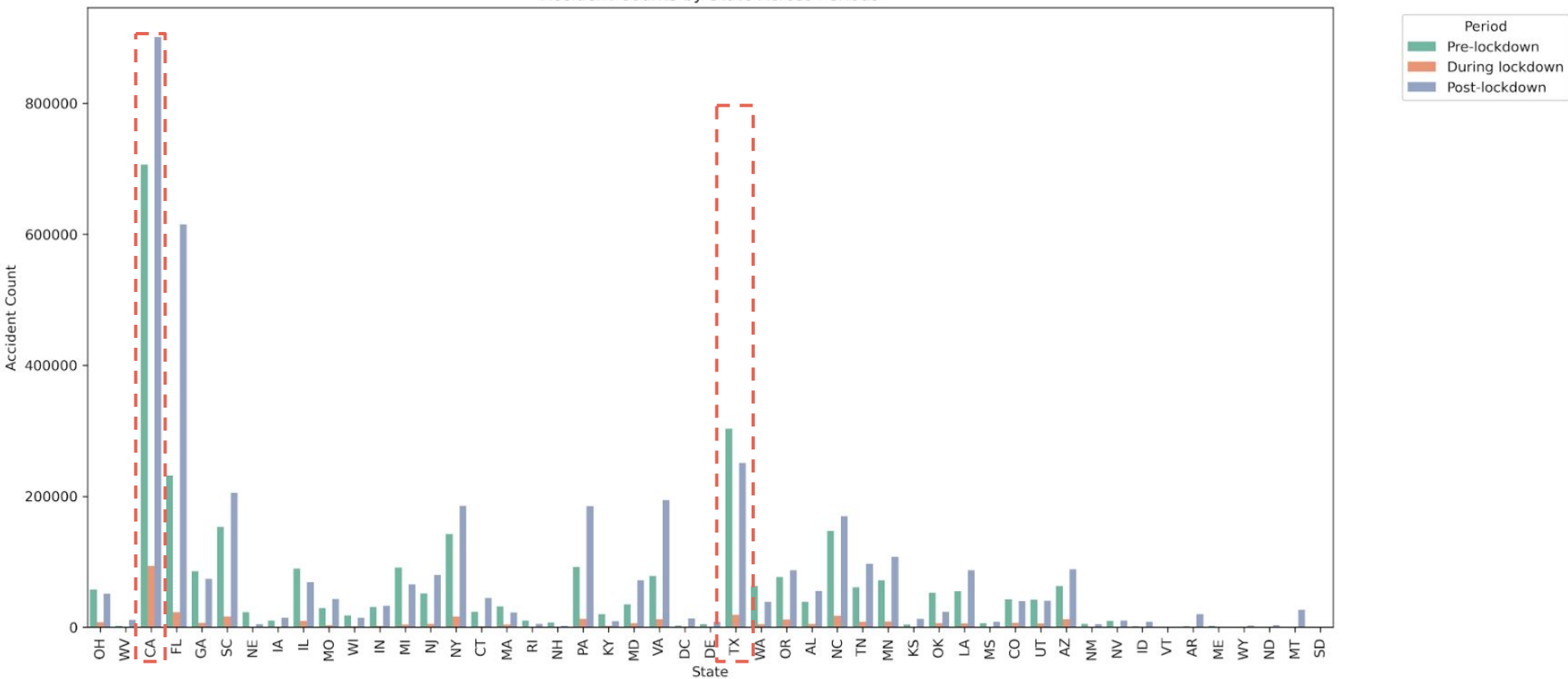


Accident Count by Time of Day



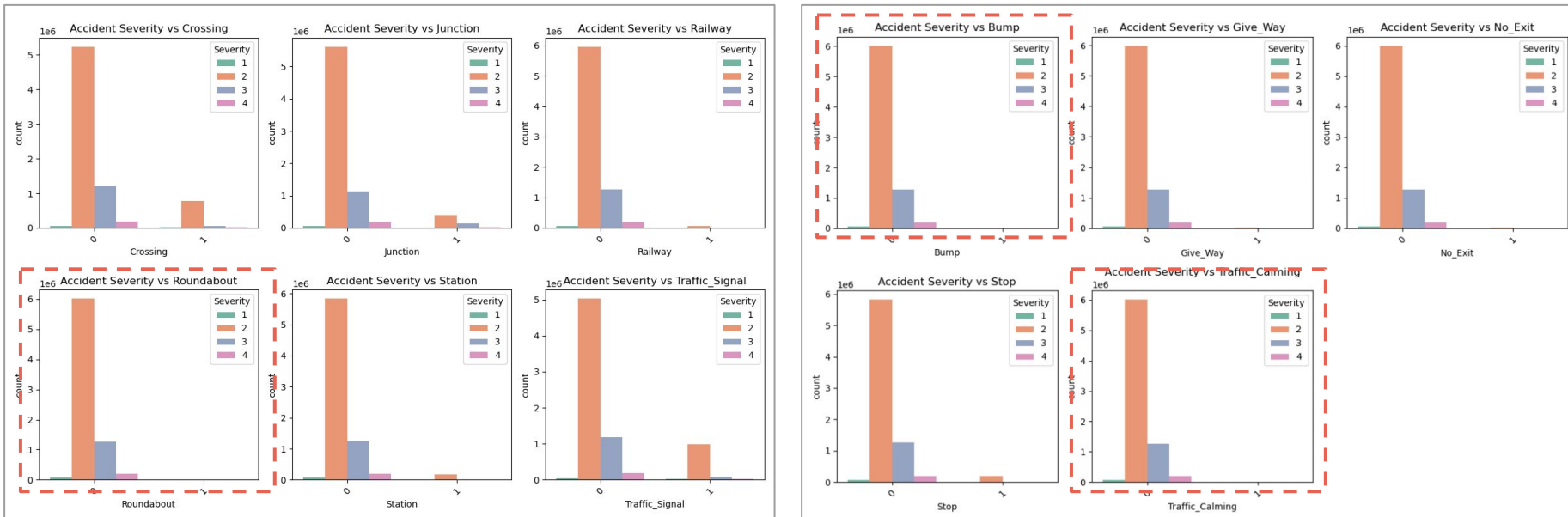
Pre vs Post Covid Accidents

Accident Counts by State Across Periods



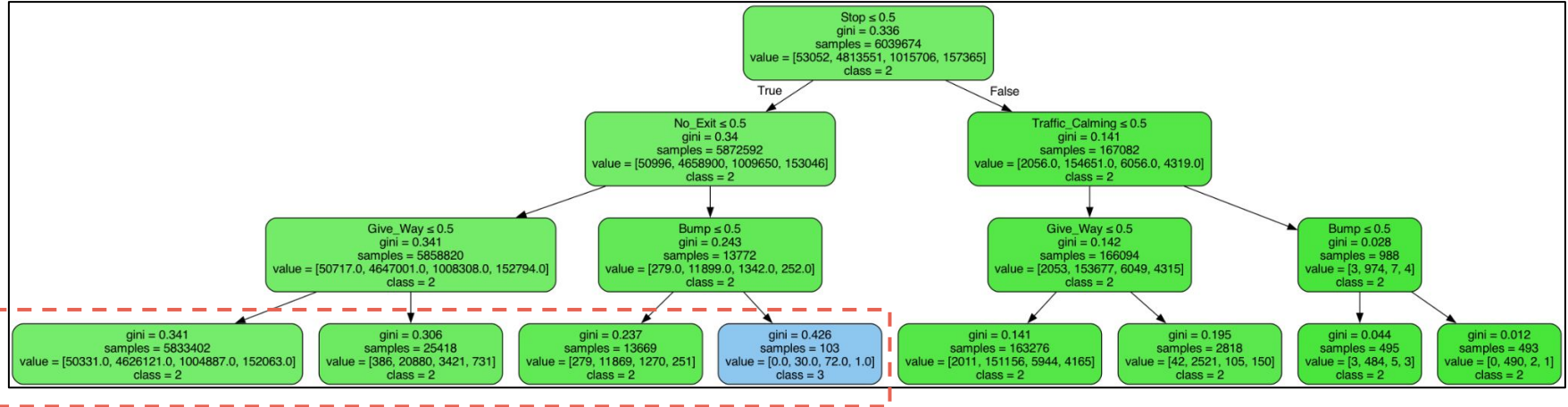
Q2: Conditions Influencing Accidents

Road Features



Least accidents recorded in roundabouts, bumps, and traffic calming features.

Machine Learning



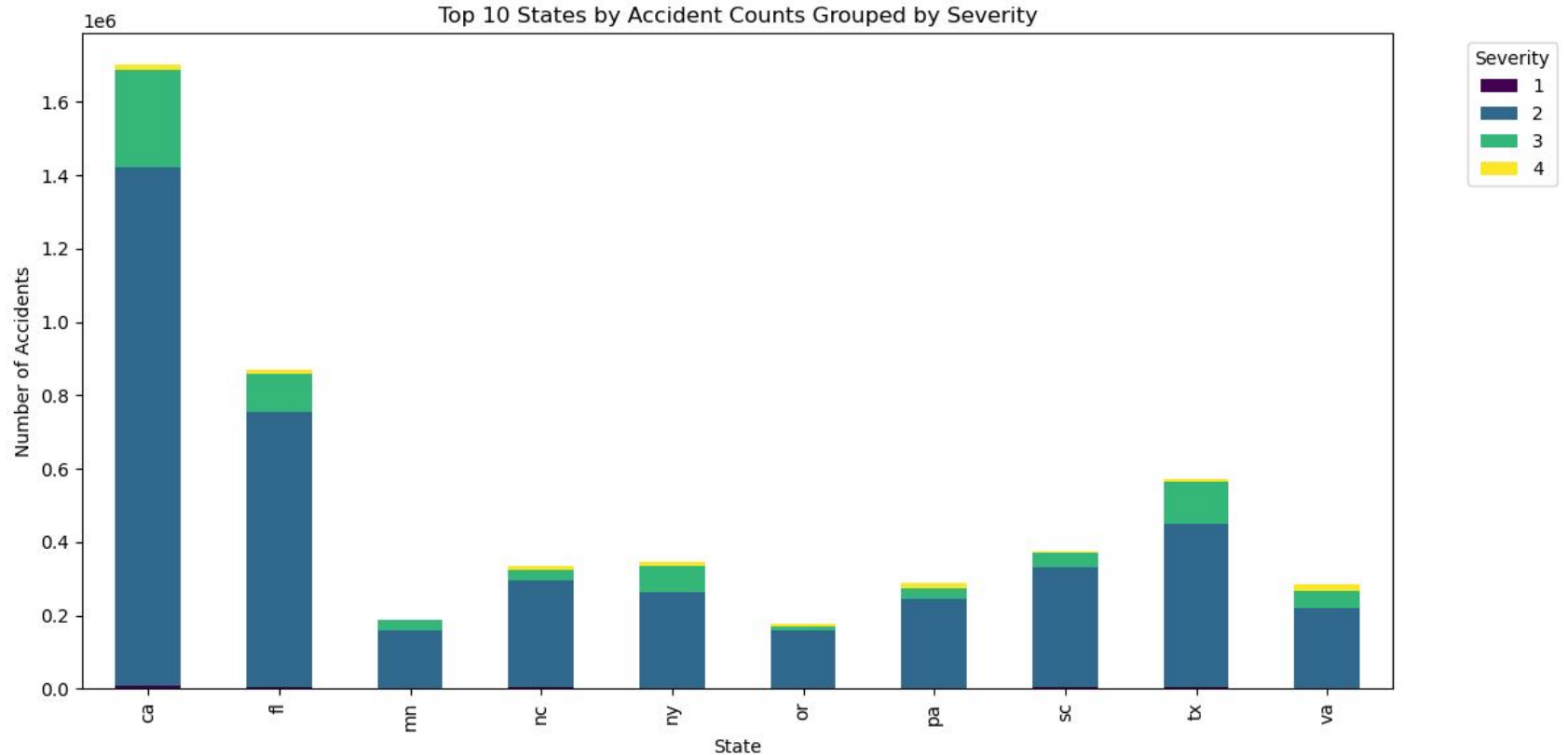
- Locations without a stop sign and where exit signs are not present are more likely to experience accidents.
- If a stop sign is absent, an exit sign is present, and a bump is present, the likelihood of a severe accident increases significantly.
- Areas where a stop sign is present but traffic calming measures are not implemented are also likely to experience accidents.
- 80% accuracy for classifying accidents with severity 2 and 3.

If	Then	Confidence(%)
Crossing	Severity 2	90.9
Junction	Severity 2	72.9
Traffic Signal	Severity 2	88.2
Low Humidity	Severity 2	80.9

*Applied association rule mining, Apriori Algorithm on a sample of data, 5% of total records.

Q3: Are Californians Worse Drivers than National Average?

Are Californians the Worse Drivers?

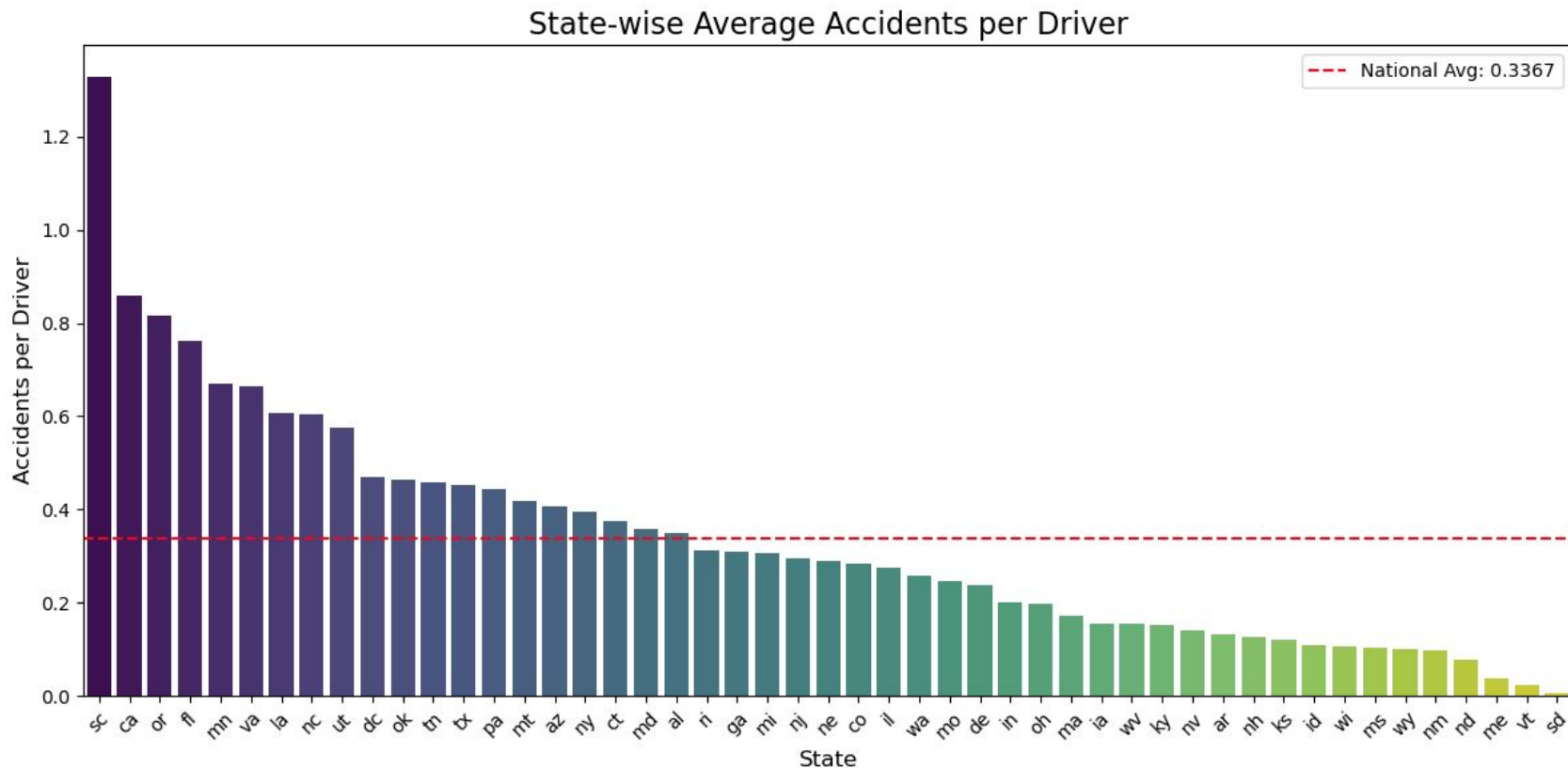


Adding Number of Registered Drivers by State

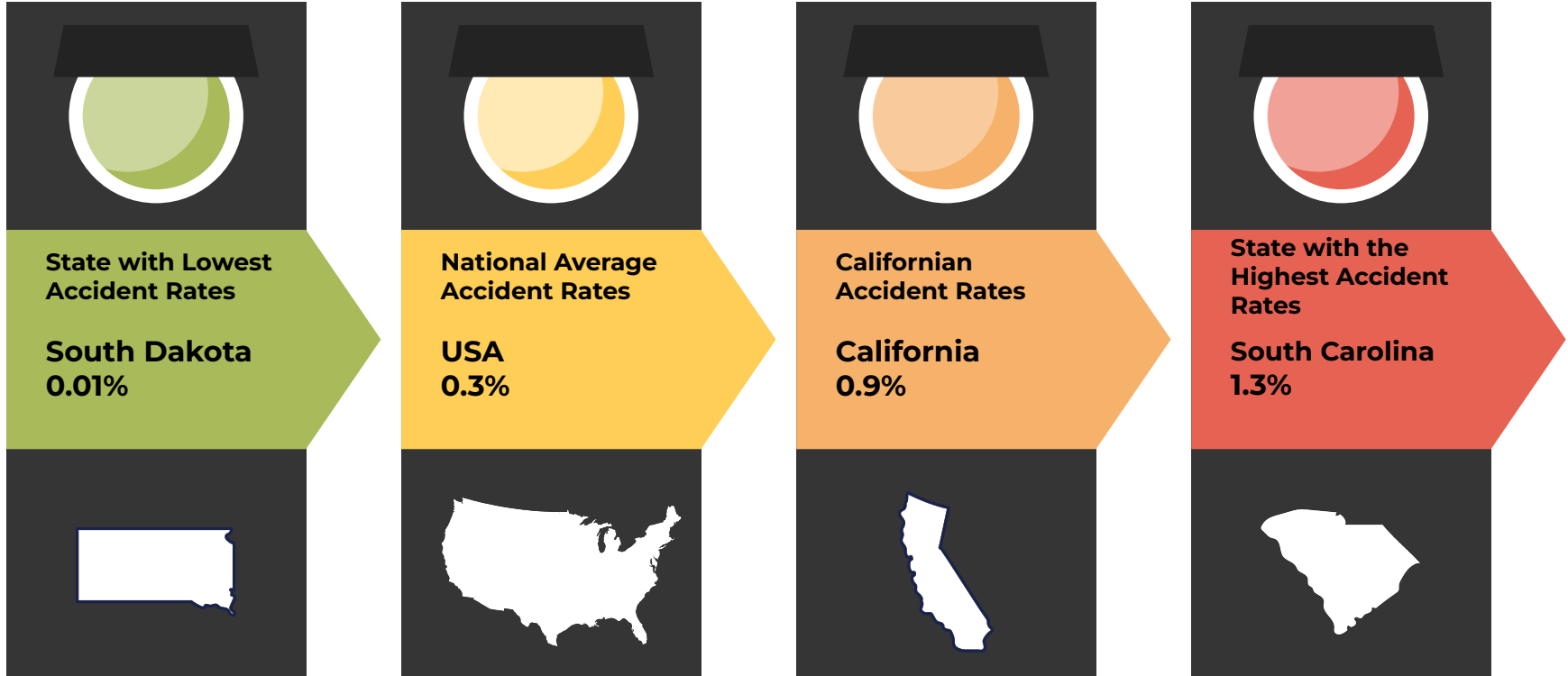
	State	Year	size	Driver_Population	Accidents/Driver
272	sc	2022-01-01	83436	4091650	0.020392
271	sc	2021-01-01	78127	3990909	0.019576
307	va	2022-01-01	94285	5837147	0.016153
270	sc	2020-01-01	61754	3905911	0.015810
62	fl	2022-01-01	260063	16495556	0.015766
174	mt	2022-01-01	13455	870882	0.015450
269	sc	2019-01-01	58119	3877968	0.014987
61	fl	2021-01-01	238418	16144302	0.014768
268	sc	2018-01-01	54780	3846069	0.014243
248	or	2019-01-01	40562	2930701	0.013840

*Data from US
Department of
Transportation

Ratio of Accidents by State & Driver Count



Are Californians the Worse Drivers?



Managerial Recommendation

Utilise Data Insights: Leverage the US Accidents dataset to identify high-risk weather and road conditions impacting traffic safety.

Develop Adaptive Systems: Incorporate decision-tree algorithms to detect hazardous conditions like heavy rain, fog, or snow.

Implement Geo-Locking: Temporarily disable or reroute autonomous services in high-risk areas to minimise accident rates.

Enhance Safety Protocols: Continuously refine autonomous systems using real-time data to improve risk mitigation strategies.

Business Impact: Boost public trust, ensure safer operations, and maintain a competitive edge in autonomous vehicle innovation.



TESLA

Thank You and Drive Safe!

