1. PURPOSE OF THE STUDY AND BACKGROUND

The purpose of this study was to assess the legitimacy of National Highway Traffic Safety Administration (NHTSA) claims of a crisis in 2021. The NHTSA observed a 10.5% increase from 2020 in deaths in motor vehicle traffic crashes. The NHTSA claims, "This is the highest number of fatalities since 2005 and the largest annual percentage increase in the Fatality Analysis Reporting System's history." The U.S. Transportation Secretary is using these claims to support taxpayer infrastructure spending.

According to U.S. Transportation Secretary Pete Buttigieg,

"We face a crisis on America's roadways that we must address together. With our National Roadway Safety Strategy and the President's Bipartisan Infrastructure Law, we are taking critical steps to help reverse this devastating trend and save lives on our roadways."

According to the NHTSA,

"The Bipartisan Infrastructure Law places a strong emphasis on improving safety and includes the new Safe Streets and Roads for All program, which opened its first round of applications just this week. The program, the first of its kind, invests up to $6 billion over five years to fund local efforts to reduce roadway crashes and fatalities. The Bipartisan Infrastructure Law now being implemented also advances Complete Streets policies and standards, requires updates to the Manual on Uniform Traffic Control Devices, which defines speeds, lane markings, traffic lights and more on most roads in the country, and sharply increases funding for the Highway Safety Improvement Program, which helps states adopt data-driven approaches to making roads safer."

The NHTSA Safe System Approach targets five areas;

- Safer People
- Safer Vehicles
- Safe Speeds
- Safer Roads
- Post-Crash Care

And six themes

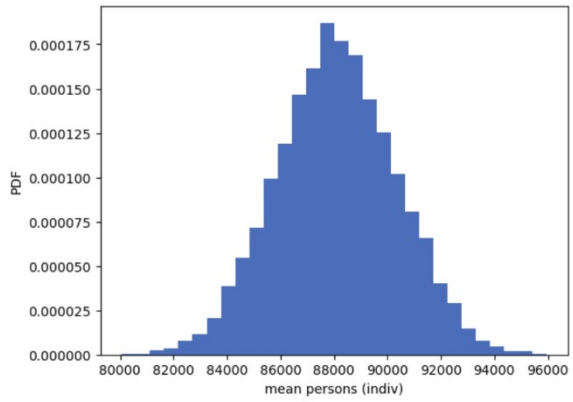
- Death and Serious Injuries are Unacceptable
- Humans Make Mistakes
- Humans are Vulnerable
- Responsibility is Shared
- Safety Is Proactive
- Redundancy is Crucial

This study uses the NHTSA data from 1975-2021 to take a data-driven look at these claims, assess the appropriateness of these programs, and recommend areas for further analysis.

2. DOES THE DATA SUPPORT THE CLAIM?

"This is the highest number of fatalities since 2005 and the largest annual percentage increase in the Fatality Analysis Reporting System's history."

The average number of persons involved in motor vehicle accidents from 2000-2021 is 88,201. In 2021, the total number of people involved in motor vehicle accidents was 96,747. Based on a bootstrap analysis and clearly identifiable in a histogram, the 2021 figure of 96747 persons involved in fatal accidents in the United States is statistically significant and an outlier value.



The 39508 fatalities in 2021 does exceed the previous high recorded in 2005. The percentage increase in fatalities from 2020-2021 is 9.04% and does equal the largest percentage increase in the dataset (though it is less than the claimed 10.5%). The percentage chance of being a fatality in a fatality causing accident however declined from a high of 41.59% to 40.84% from 2020-2021. The 2021 value is the second highest recorded value since 2000.

The claims above appear supported by the data.

But is this a crisis?

YEAR	Deaths (Yes=1)	Accidents	Persons	% That Died
2000	0	27354	100716	37.25
2000	1	37526		
2001	0	27498	101175	37.42
2001	1	37862		
2002	0	27582	101784	37.82
2002	1	38491		
2003	0	27527	101862	37.77
2003	1	38477		
2004	0	27422	100760	38.15
2004	1	38444		
2005	0	27701	101262	38.76
2005	1	39252		
2006	0	26924	98356	39.29
2006	1	38648		
2007	0	25707	94338	39.68
2007	1	37435		
2008	0	23262	84510	40.44
2008	1	34172		
2009	0	21012	76510	40.34
2009	1	30862		
2010	0	20873	74863	40.47
2010	1	30296		
2011	0	20519	73364	40.71
2011	1	29867		
2012	0	21245	76436	40.56
2012	1	31006		
2013	0	20866	74331	40.63
2013	1	30202		
2014	0	20865	73711	40.78
2014	1	30056		
2015	0	23075	81620	39.87
2015	1	32538		
2016	0	24731	86474	40.18
2016	1	34748		
2017	0	24645	85840	40.26
2017	1	34560		
2018	0	24432	84344	40.22
2018	1	33919		
2019	0	24124	82843	40.42
2019	1	33487		
2020	0	25435	86396	41.59
2020	1	35935		
2021	0	28262	96747	40.84
2021	1	39508		

Crisis

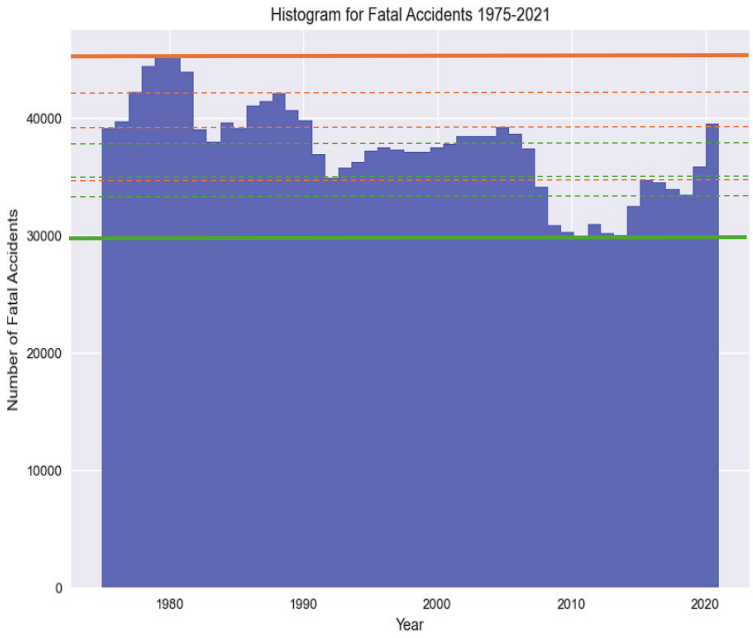
- a time of intense difficulty, trouble, or danger
- a time when a difficult or important decisions must be made

3. DOES THE DATA SUPPORT THE CLAIM (CONTINUED)?

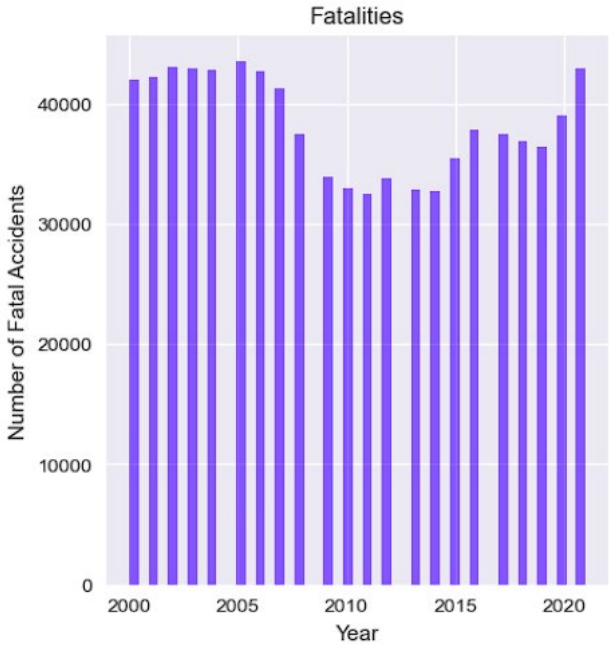
"We face a crisis on America's roadways that we must address together."

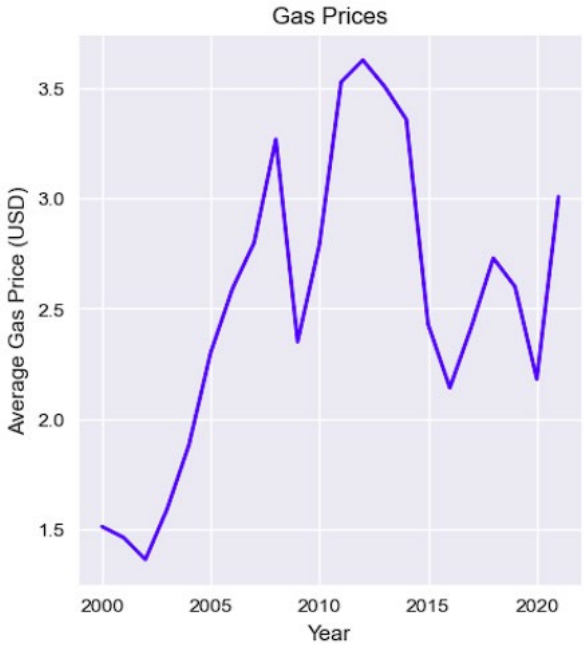
- U.S. Transportation Secretary Pete Buttigieg

The increase in fatal accidents from 2020-2021 is significant. It is the largest single increase year to year since 1975. The 2021 total, however, is significantly less than the highs of 1980-1981 and 1989. The total just slightly exceeds that of 2005. In investment speak, there are still two levels of resistance before the current rising trend in fatal accidents, that began in 2014, hits an all time high. The reduced volume of traffic in the United States as a result of the COVID-19 pandemic from 2019-2020 could explain the unusual increase. The histogram above does show a decline in 2019 followed by large increases in 2020 and 2021.



Another explanation is the price of a gallon of gas. When gas prices are higher than 3 USD/gal, fatal accident totals decline. The increasing fuel prices under the Biden Administration may contribute more to reducing fatal accidents than the National Roadway Safety Strategy and the President's Bipartisan Infrastructure Law. Declaring a crisis on US roadways may be premature.

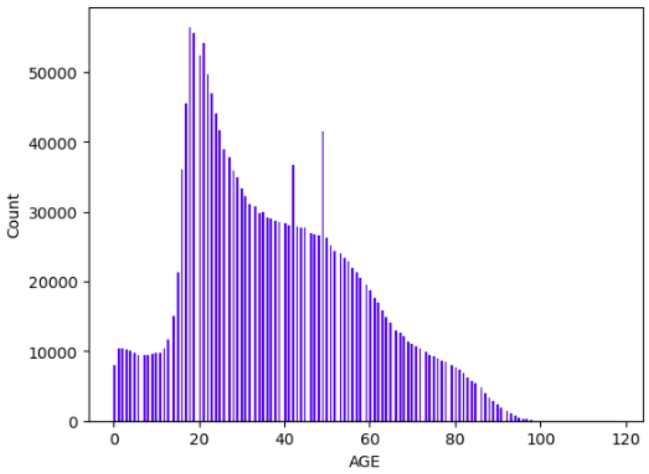


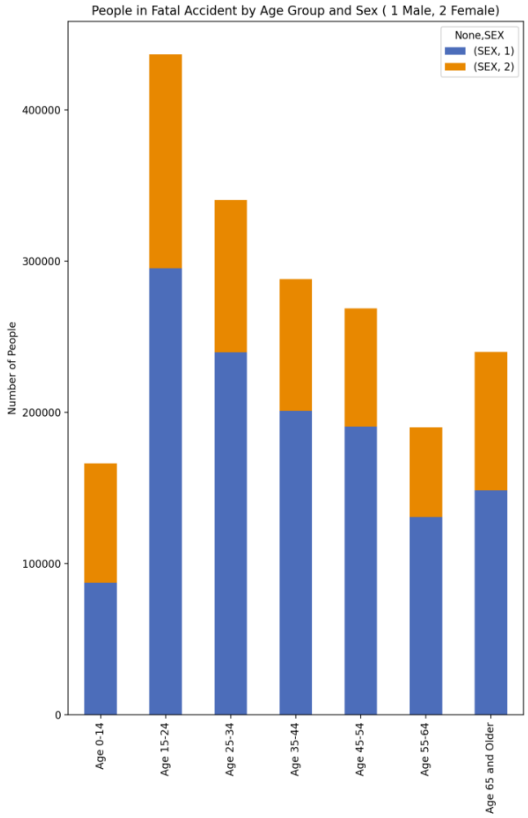


4. FORTY-THREE PERCENT OF PEOPLE IN FATALITY CAUSING VEHICLE ACCIDENTS DIE

From 2000-2021, 842445 people died from a population of 1938242 involved in vehicle related accidents. This means 57% of people survived them. What are the characteristics of these accidents and can targeted policy or investments in traffic safety reduce the fatalities?

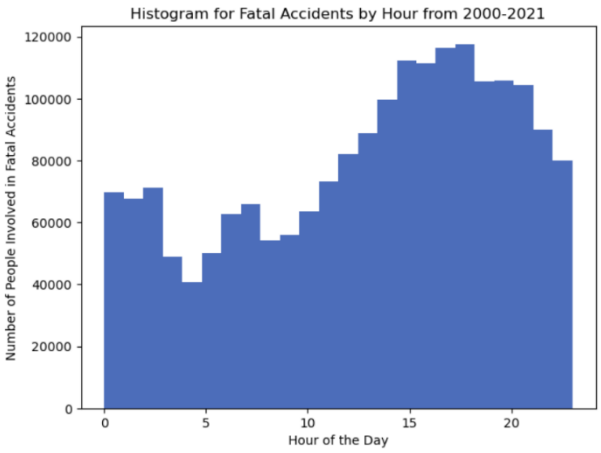
Segmenting this population by age and gender we see noteworthy characteristics. First, after age 25 the number of people involved trends strongly downward. No wonder, that at age 25 the price of US auto insurance drops significantly. The second is that males make up a larger proportion of this population than females.

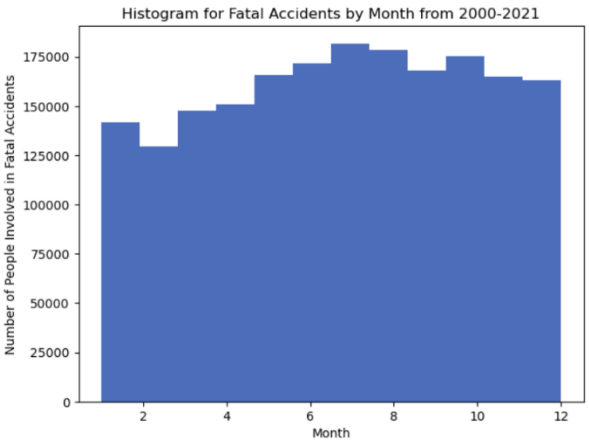




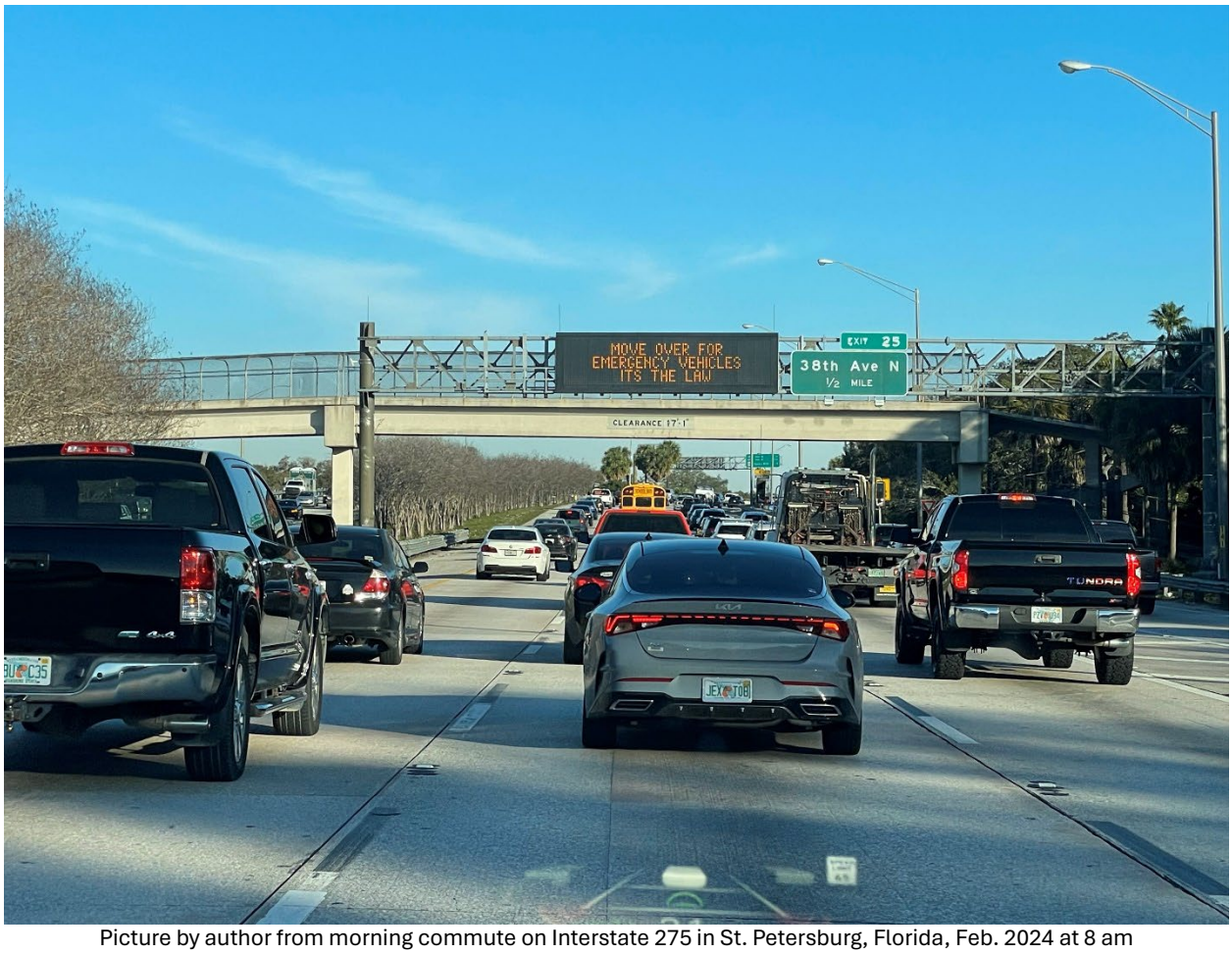
Both of these charts show that policies and investments targeting males between the ages of 15-24 could yield a 10% reduction in accidents, but does this correlate to behaviors and fatalities?

The number of people in fatal accidents increases as the typical US workday commute begins at 5 am (5) and then tapers off after 6 pm (18). The peak between 5 am to 8 am is morning rush hour. Note how the accidents take place over a 10-hour period of time and then remain at their highest levels for four hours during the evening rush hour from work. Then another three hours of evening activity before 10 pm (20). From 10 pm to 3 am alcohol may be a factor. Blood alcohol levels had the largest positive coefficient during KNN modeling indicating a strong influence on accident fatality prediction.





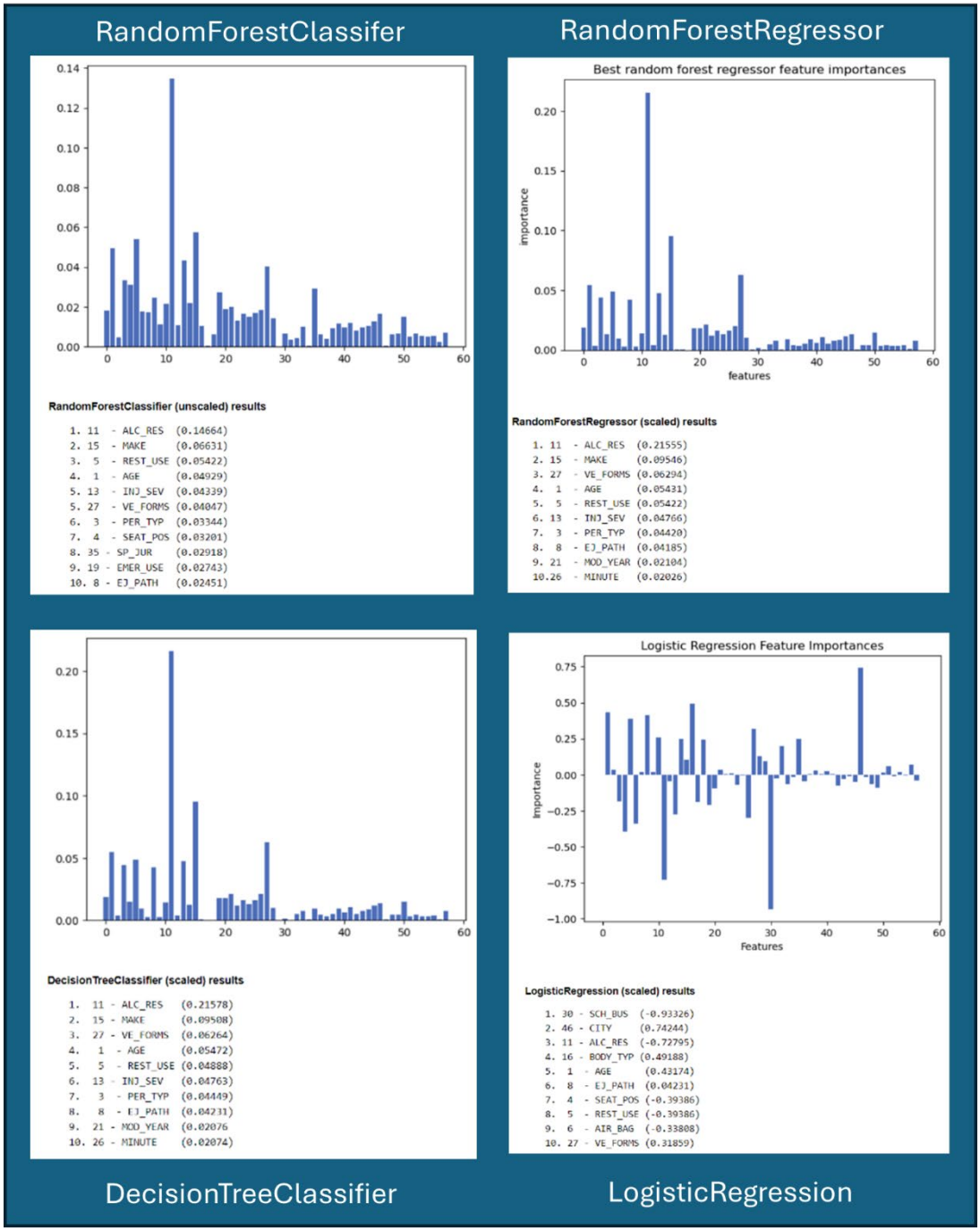
The distribution by month is predictable. The coldest months of the year in the United States are January through March. Recreational and mountain roadways close due to snow and ice accumulation and lack of snow clearing equipment. With the arrival of Spring the pattern increases and peaks during the primary summer vacation months of July and August. The high value in October could be related to either the onset of the rainy season or Halloween where a lot of people are out walking with children at night on neighborhood streets during peak driving hours. There are significant efforts made to provide alternatives for Halloween celebrations during the day and or inside schools and churches. Another explanation is people in northern states migrate to southern states in October and then back again in April. November and December may plateau because of families driving to their relatives to celebrate Thanksgiving and Christmas and increases in package delivery services. Additional research could validate these assumptions. Vacation and holiday driving public awareness outreach may be worthwhile.



Picture by author from morning commute on Interstate 275 in St. Petersburg, Florida, Feb. 2024 at 8 am

4. IDENTIFYING THE MOST IMPORTANT FEATURES IN THE DATA FOR ANALYSIS

We have looked at four of the fifty-eight features; YEAR, MONTH, AGE and SEX. To reduce the scope of having to examine every one of the other features I examined the feature coefficient values of four different models. By rank ordering the results for each and then comparing them I identified the most influential features that predict the fatalities. The top ten common features were ALC_RES, REST_USE, AGE, INJ_SEV, PER_TYP, SEAT_POS, EMER_USE, EJ_PATH, DRINKING, VE_FORMS. I will also include MAKE in this report. MAKE ranked number twelve overall as it only appeared in three of the four models explored, but in the three that it did it was the second most important.



5. BLOOD ALCOHOL (BAC) LEVELS

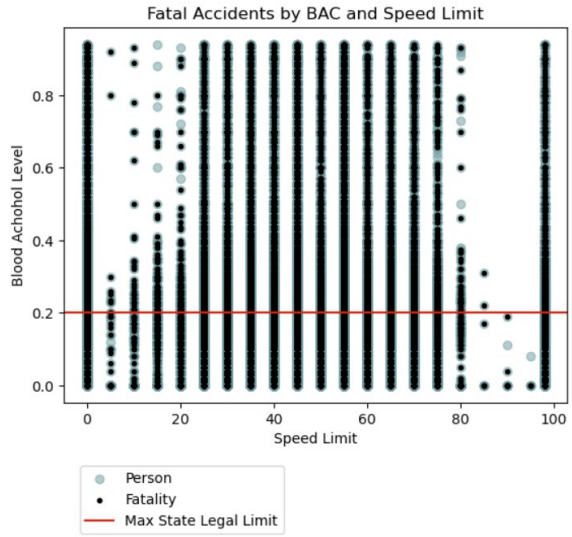
Blood Alcohol levels (ALC_RES in the dataset) had the highest positive coefficient for all models evaluated. The ALC_RES feature needed correction as many two digit values exist from 2000-2014 when the FARS standard encoding was between 0-99.

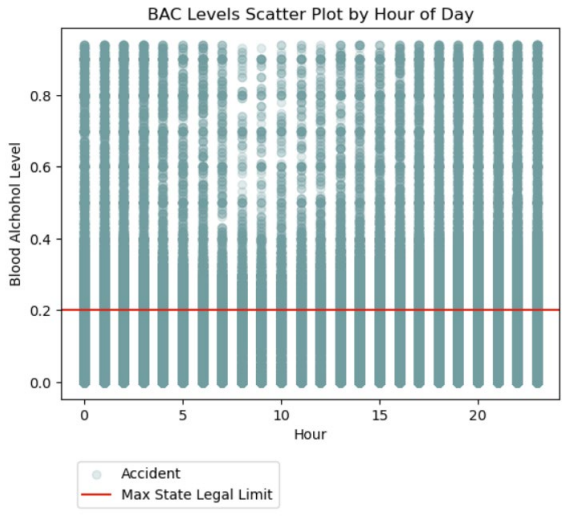
	ALC_RES	
SEX	1	2
AGE_GROUP		
Age 0-14	0.010072	0.007786
Age 15-24	0.077031	0.050755
Age 25-34	0.101713	0.074298
Age 35-44	0.095289	0.068799
Age 45-54	0.085389	0.055168
Age 55-64	0.066428	0.031965
Age 65 and Older	0.029617	0.010203

The legal blood alcohol concentration (BAC) limit by state varies from 0.00 to 0.20. Average BAC for all age groups is higher in men (1) by 3%. The highest average levels are highest in ages 25-54. The highest ages group is 25-34.

The scatter plot below on the right reveals that drunk drivers participate in fatal vehicle accidents at all hours of the day. There is noticeable variation between 4 am and 4 pm meaning fewer occurrences (people are at work). Between 7 pm and 3 am the individual dots for each occurrence form almost solid lines. Clearly not the safest time period to be on the road. According to the USDOT Federal Highway Administration (FHA), over 40% of speeding drivers in fatal crashes were considered to be alcohol impaired. This data set does not contain a speeding feature to verify this claim.

The scatter plot on the left reveals that drivers over the BAC legal limits in all states are driving around in neighborhoods (25-35 speed limit), rural roadways (40-55), and highways (55-70). The dense line at 0, is for No Statutory Limit/Non-Trafficway or Driveway Access. These include entrance and exit ramps, neighborhood streets, and off-road areas. The values over ninety-five are Unknown. There are 842,445 deaths out of 1,938,242 involved persons in this dataset.





Eliminating people where authorities did not record a BAC level the total deaths plotted in the lefthand chart is 531,738 and the total persons involved is 700,239. Few green circles that indicate less than a fatal injury are visible. Policies and information campaigns that target people, ages 25-54, commercial de-glamorization of alcohol in the United States, and increased enforcement of drunk driving and speeding laws could potentially reduce fatal accidents.

6. RESTRAINT (SEAT BELT, CHILD SEAT, HELMET) USE

The bar chart to the left depicts reported injuries by age group, what type of restraint system used, and the resultant severity of the injury. Each bar represents a restraint system code by age group. Purple color = Fatality.

The longest bar by age group is code 3, a shoulder and lap belt used. These people are complying with the law and driving vehicles with shoulder and lap belts.

The total number drops as expected as people age. The sixty-five and older age group breaks this trend; 47.62% (62,783 people) were fatally injured in a vehicle accident. This group merits further research. Are they riding in vehicles driven by younger drivers or more vulnerable?

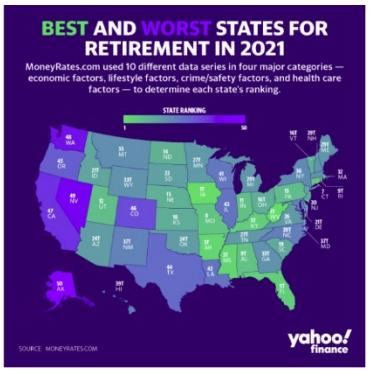
The second longest bar is code 20, None Used/Not Applicable.

```
% Chance of Fatal Injury w/o Use of Restraints
Age 65 and Older: 81.26
Age 55-64: 79.23
Age 45-54: 74.85
Age 35-44: 69.79
Age 25-34: 64.67
Age 15-24: 55.6
Age Under 15: 37.16
```

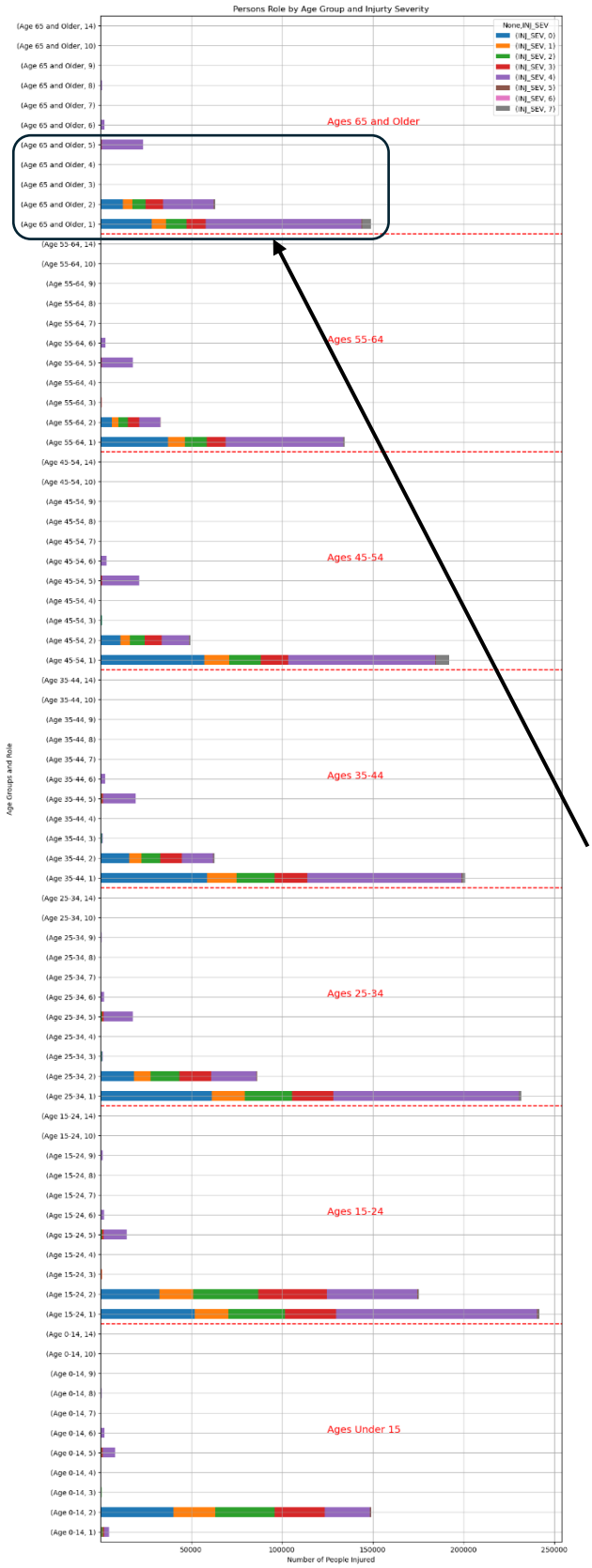
Use of seatbelts became law in 1968

Seat belt laws in the United States vary by state. In 2021 the most fatal accidents occurred in Texas, California, and Florida. These are also the most populated states in the United States; California, Texas, Florida. Texas and California require seat belt or child restraint use for all ages in all seats.

Florida requires all ages in front seats and under the age of eighteen in all seats. Florida should consider a change to all ages in all seats based on this bar plot.



7. ROLE OF THE PERSON INVOLVED IN THE CRASH



The bar chart to the left shows the role of the person in the accident by age group and severity of injury. Each bar represents a person type code by age group. Purple color = Fatality.

The top four roles defined by the number of persons ordered highest to lowest are driver (1), passenger (2), pedestrian (5), and bicyclist (6).

Nearly, all pedestrians and bicyclists involved in motor vehicle accidents died. Few survived with severe injuries.

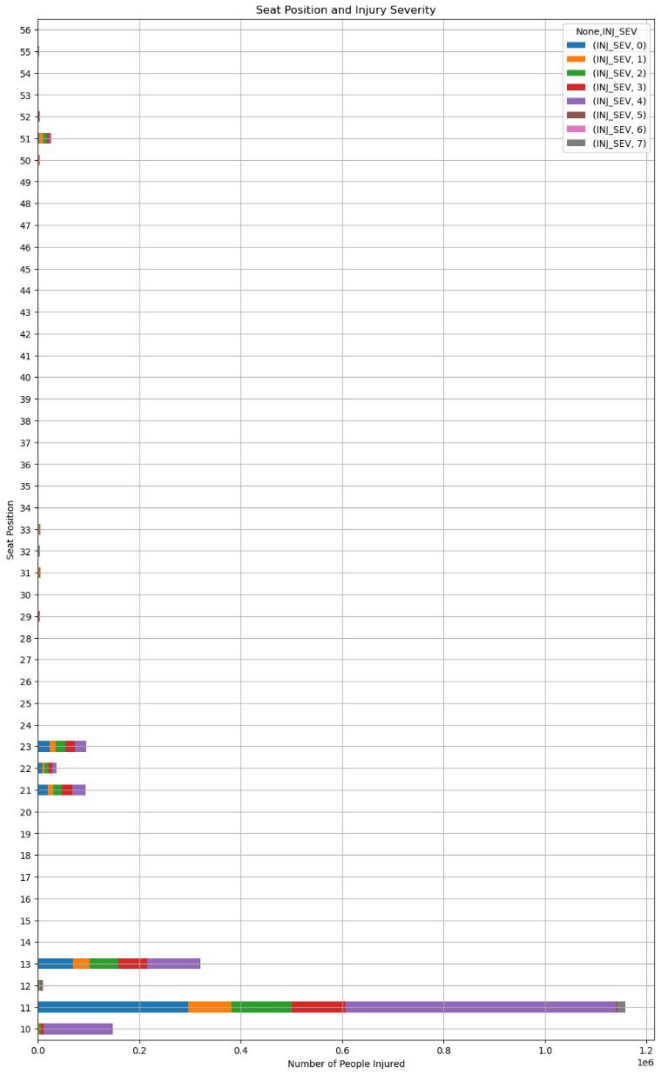
Drivers and passengers of motor vehicles make up the majority of people in these accidents. The ratio of drivers to passengers is consistent from age 25 to age 64. The first two age groups (under 15 and 15-24) contain children legally not allowed to drive. Note that those under age 15 who were driving, died in the accident.

Following up on the age analysis from the previous page we see here that the driver to passenger ratio does narrow at age 65 and above (more are passengers), but the majority are still driving. In both cases, the fatality ratios are the highest of all age groups to include those that are riding bicycles.

A pedestrian or bicyclist becoming a fatality in an accident with a motor vehicle is not surprising. Raising driver awareness through road markings, signage, traffic controls, high visibility clothing/lights, and legal penalties are all good investments.

The ratio of drivers fatally injured appears slightly higher than passengers. The overall patterns, however, are similar. The fact that there are over twice as many drivers as passengers in motor accidents is easily explained by people using their vehicles to go to and from work as individuals and not in carpools.

8. SEAT POSITION



The bar plot to the left is the seat occupied in the vehicle or not in the vehicle at all and severity of injury. Purple color = Fatality.

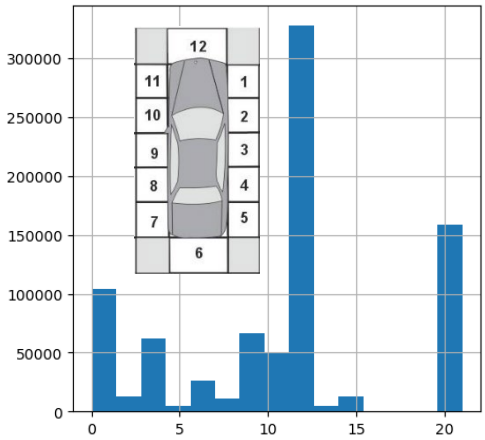
The top seven seat positions are front left driver (11), front right passenger (13), not a motor vehicle occupant (10), second row left (21), second row right (23), second row middle (22), and enclosed passenger or cargo area (51).

Same as in the roles bar chart, most non-occupants of a vehicle involved in motor vehicle accident died or were severely injured.

The front driver bar (11) and front passenger bar (13) both show the same injury proportions as the driver and passenger roles in the previous plot. This shows consistency in the data set.

Computing the percentage of fatalities for each seat position provides an estimate of the percentage chance a person will be a fatality in an accident. The chart on the bottom left of this page shows the computed percentages from this dataset.

```
% Chance of Fatal Injury by Seat Position
First Row Left: 46.02
First Row Middle: 20.96
First Row Right: 32.23
Second Row Left: 26.39
Second Row Middle: 19.21
Second Row Right: 22.63
Third Row Left: 16.8
Third Row Middle: 16.92
Third Row Right: 16.47
Enclosed Cargo Area: 11.5
```



Sitting further back and in the middle seat of the vehicle is a good strategy. The histogram on the right shows the majority of fatal accident collisions happen in the front or side. Values other than clock positions (1-12) are non-collision (0), undercarriage (14), underride (15), or hit by something else (20). Interesting that riding in the cargo area is the best at 11.5%.

9. VEHICLE MAKE

Involved in an Accidents with Fatalities by Make in the United States (2000-2021)

1.	Ford(12)	305,431
2.	Chevrolet(20)	280,677
3.	Toyota(49)	124,045
4.	Dodge(7)	120,679
5.	Honda(37)	113,529
6.	Nissan(35)	74,814
7.	Pontiac(22)	45,090
8.	Jeep(2)	43,252
9.	Suzuki(53)	21,032
10.	Yamaha(76)	16,611

The number of makes of vehicle are extensive in the United States. The top ten list of vehicle manufacturers in fatal accidents is shown on the left. This list is interesting considering in 2020 the list by market share is: General Motors (17%), Toyota (15%), Ford (14%), Honda (9%), Nissan (8%), and Dodge (1%). Chevrolet and Pontiac are together in the General Motors figure. I was unable to locate market share figures for Jeep, Suzuki, or Yamaha. Suzuki and Yamaha sell two-wheel motorcycles in the United States market. If we assume the market share has not changed dramatically from 2000-2021 then it follows that drivers of Ford and Dodge vehicles participate in fatal accidents more than expected.

Using 2020 4 Door Sedan data and the equation: (Make Deaths – Market Share*Total Deaths)/Make Death. We can get a sense of which driver/vehicle combinations are the most dangerous. In this case, Dodge, Honda, and Nissan take the top three spots. I used the 4 Door Sedan because the most people die in that body type by make and it eliminates motorcycles from the Honda data.

- Dodge: 88% greater
- Honda: 40% greater
- Nissan: 35% greater
- Ford: 21% greater
- Chevrolet: 20% greater
- Toyota: 13% greater

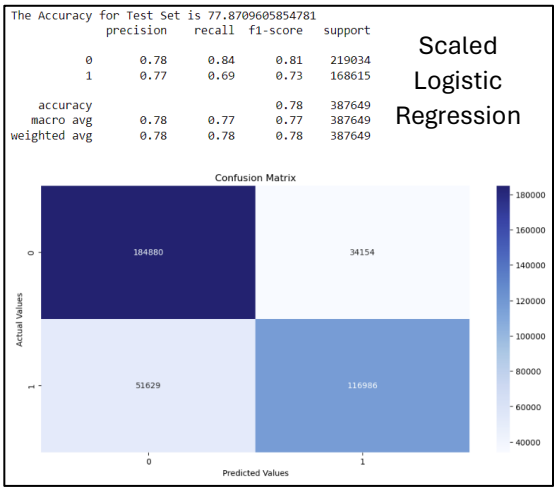
Rank Ordered (Make / Body Type / # Deaths>5000)

1.	Chevy	4-Door Sedan	28813
2.	Ford	4-Door Sedan	23968
3.	Toyota	4-Door Sedan	23493
4.	Honda	4-Door Sedan	20689
5.	Ford	Standard Pickup	20228
6.	Chevy	Standard Pickup	20066
7.	Honda	2W Motorcycle	17386
8.	Nissan	4-Door Sedan	16227
9.	Ford	Compact SUV	15831
10.	Suzuki	2W Motorcycle	14356
11.	Jeep	Compact SUV	14332
12.	Chevy	2-Door Sedan	13917
13.	Chevy	Compact SUV	12249
14.	Yamaha	2W Motorcycle	12166
15.	Pontiac	4-Door Sedan	11007
16.	Dodge	4-Door Sedan	9693
17.	Ford	2-Door Sedan	8987
18.	Ford	Compact Pickup	8826
19.	Dodge	Standard Pickup	8435
20.	Honda	2-Door Sedan	8298
21.	Ford	Light Pickup	6951
22.	Chevy	Compact Pickup	6765
23.	Chevy	Light Pickup	6530
24.	Pontiac	2-Door Sedan	6497
25.	Chevy	Full Size	6284
26.	Dodge	Minivan	5831
27.	Toyota	Compact SUV	5816
28.	Toyota	Compact Pickup	5284

Throughout this study it is evident that human behavior is the primary cause of fatal accidents. Driving is a human activity. Safer people is the key. Most vehicles and motorcycles manufactures on the rank ordered list to the left promote their products as exciting to drive, fast, and aggressively styled to attract buyers. Marketing that reinforces safer driving habits could be more helpful here.

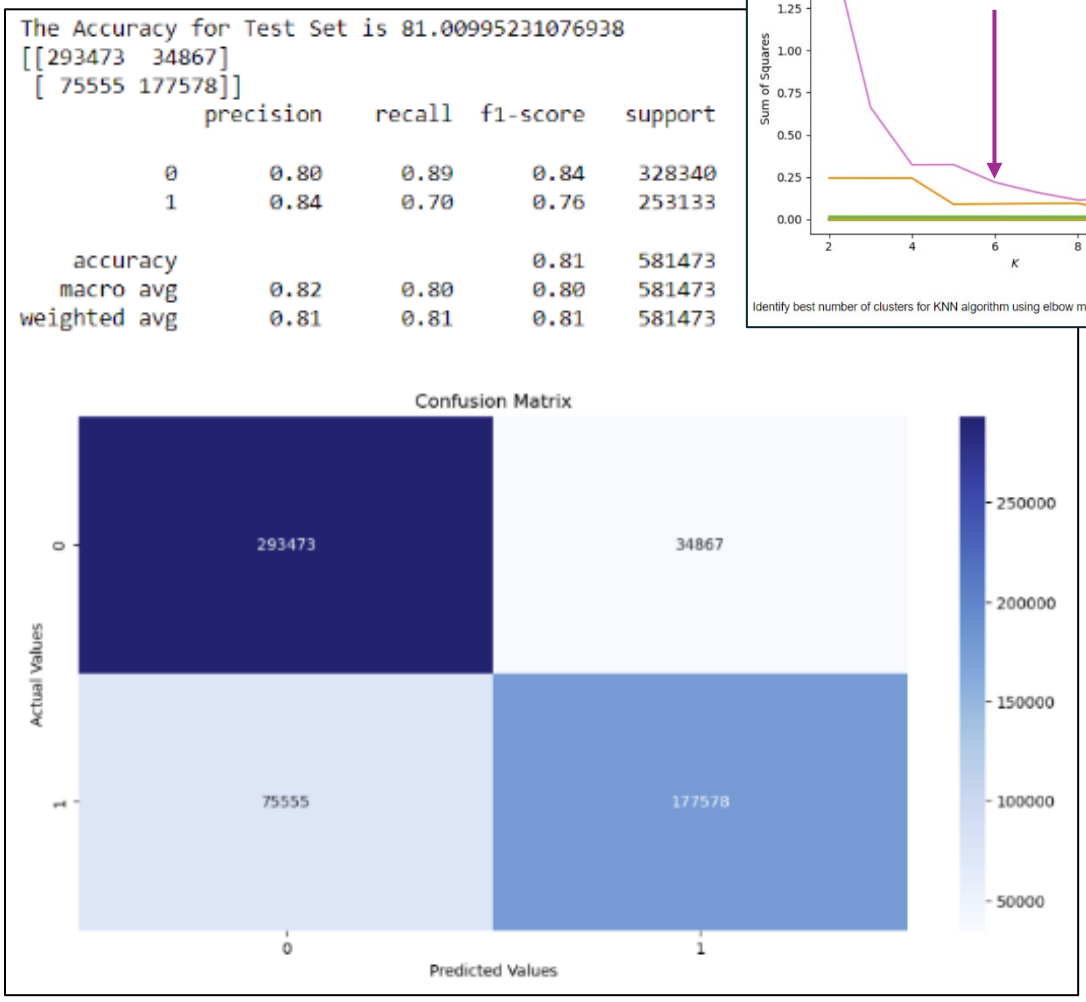
10. MODELING – VEHICLE ACCIDENT FATALITIES ARE PREDICTABLE

Using a K-Nearest Neighbor (KNN) algorithm provided the highest accuracy (**81%**), precision, and recall scores of all. I tuned the algorithm using a *n* (*neighbors*) value of 6 selected using the elbow method technique. I reduced the dataset from fifty-eight features to eleven features, selecting those derived from the feature coefficient modeling and the analysis previously discussed. These adjustments reduced the run time from over three hours (KNN, n=6 with all fifty-eight features) to less than five minutes and increased all scores. The next best model was a scaled, logistic regression with accuracy score of 78. Shown here on the right.



Scaled Logistic Regression

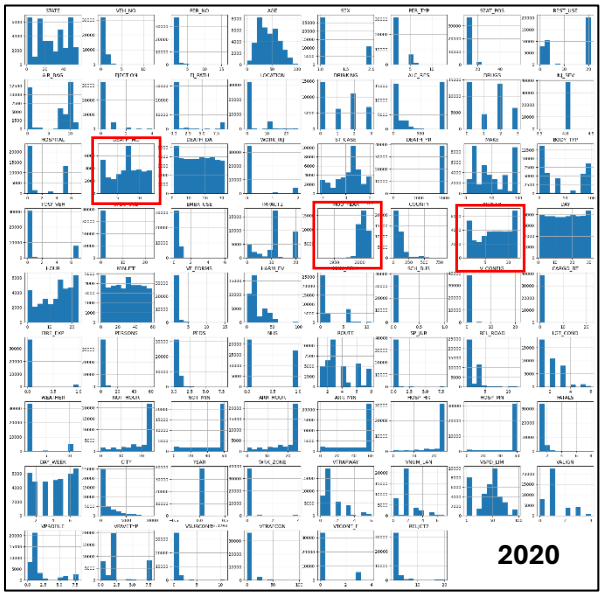
Best Predictive Model (KNN) Scores

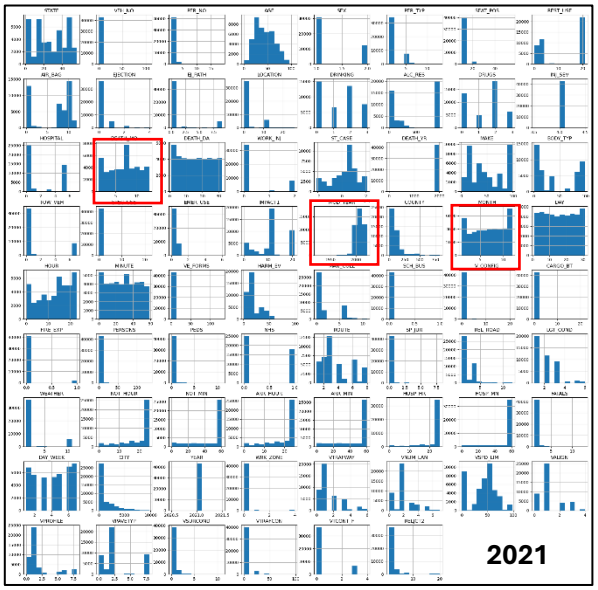


Elbow Method

11. CONCLUSIONS

The feature histograms below are consistent between 2020 and 2021. The patterns are what is important to see here. The increase in fatal accidents in 2021 does not appear to be due to any change in the features detailed in this report and prioritized by the models. The features highlighted with red squares due exhibit a noticeable pattern change from 2020-2021. From left to right these features are month of death, vehicle model year, and month of accident. A simple conclusion is more people were on the road during the winter months in 2021 and driving newer model vehicles.





This leads to the possible conclusion that more drivers returned to the road post-COVID 19 lock downs and the subsequent reduction in work from home arrangements. The increase in fatal accidents in 2021, though outside the norm, is not a result of unsafe vehicles, statutory speed limits, road design, infrastructure upkeep, or post crash care.

Fatality causing vehicle accidents from 2000-2021 appear to have happened for the same reasons, in the same types of vehicles, under the same conditions, with the same results. A typical fatal accident includes sober people ages 18-30 years old, driving four door sedans and pickup trucks and motorcycles, from Summer to Winter between 3 pm and 10 pm, on dry, paved, multi-lane rural highways (speed limit 55 mph), in clear weather conditions.

```
Lanes
% 2 Lane: 67.63
% 4 Lane: 11.95
% 3 Lane: 9.02
```

```
Pavement
% Asphalt: 73.31
% Concrete: 8.4
```

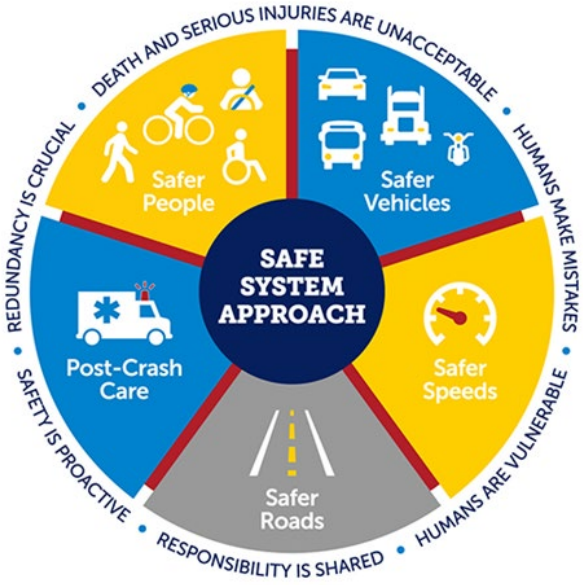
```
Conditions
% Dry: 80.08
% Wet: 11.57
```

```
Weather
% Clear: 81.36
% Raining: 7.45
% Cloudy: 7.33
```

Plotting the person count involved in fatal accidents (see page 3) from 1975-2021 we can see that the increase in 2021 may be a result of the law of large numbers and a return to the norm instead of a crisis. In fact (see page 2), the percentage of fatal injuries decreased from a high in 2020 of 41.59% to 40.84% in 2021 (see page 2), despite the increases in fatal accidents and persons involved. The Department of Transportations claim of a crisis appears premature.

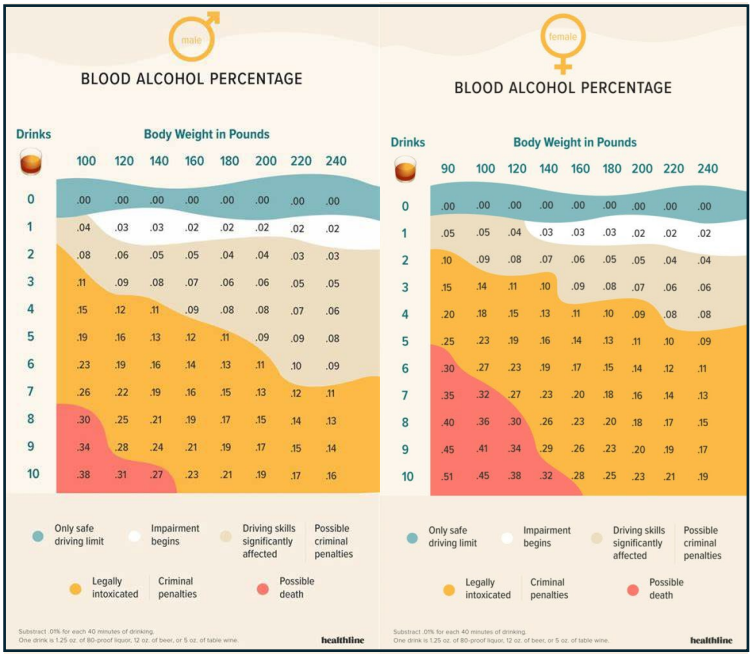
12. RECOMMENDATIONS

To reduce fatalities by 10% the Department of Transportation should focus on collaboration with other departments of government to find approaches that influence safer driving behaviors and less on the other elements of the Safe System Approach. Five of the six statements that surround the image, according to the data, are helpful. People are the problem; the other four wedges of the image are symptoms. The statement that death and serious injuries are unacceptable is naïve. They are a facts, supported by the data, and acceptable to every person that chooses to drive, walk, or bicycle on or near roads in the United States. A more accurate statement is death and severe injury are a choice. Additionally, it is accurate that humans make mistakes, but choosing not to wear a seatbelt and driving extremely drunk are not mistakes.



This study shows that when people of all ages choose to wear seatbelts and not drink and drive, we can reduce the total number of fatalities by more than 10%. From 2000-2021, 842,445 people died out of 1,938,242 people involved in fatality causing accidents. This works out to 43%. Of those who died, 79,874 were not wearing a set belt. This accounts for 9.48% of the fatalities.

If we consider alchohol, the max. legal limit in the United States is 0.20 BAC. Of those who died, 94,803 had a blood alchohol content of 0.20 or higher. Look at the charts to the right. The smallest male must drink 6 drinks and the smallest female 4 drinks to reach this level. Driving a vehicle after consuming this number of alcholic drinks is undeniably a choice, it is not a mistake. This population accounts for 7.92% of the fatalities.



These two human choices, not mistakes, account for 17.4% of the fatalities in this dataset from 2000-2021. It is not a crisis, but it is time to invest less in infrastructure and more to **change our behaviors**.