Capstone 2 Presentation

How the Department of Transportation Can Reduce Fatal Vehicle Crashes by 10% in Calendar Year 2024?

Terry A. Meyer

February 4, 2024

Problem Statement Worksheet (Hypothesis Formation)

How can the US Department of Transportation reduce fatal roadway crashes by 10% in calendar year 2024? Is the National Roadway Safety Strategy, the President's Bipartisan Infrastructure Law, and Safe Streets and Roads for All program ($6 Billon over five years) the right approach or government waste?

H

1 Context

The Biden Administration declared a "crisis on America's roadways" after the release of the 2021 National Highway Traffic Safety Administration (NHTSA) data showed a 10.5% increase in deaths in motor vehicle traffic crashes in the United States from 2020-2021. Since then, over $6 Billion USD was appropriated in the President's signature Bipartisan Infrastructure Law to tackle the issue. In 2022 a 0.3% reduction occurred and government forecasts for the first half of 2023 show a potential 3.3% reduction. The National Roadway Safety Strategy (NRSS) established 43 Action Titles to monitor progress from 2022-2023. All of which appear to be qualitative allowing for maximum political bias and maneuvering.

2 Criteria for success

1. Independent assessment of facts and assumptions made in the NRSS using the NHTSA's own data. Does it target the features that contribute most to motor vehicle traffic crashes?

2. Train a machine learning model on a cleaned NHTSA data set to predict fatalities and forecast the 2023 estimated deaths using feature set from the NRSS and the feature set from my own analysis of the data. Goal is a combination that offers a path to a 10% reduction.

3 Scope of solution space

This initiative will focus on examination of NHTSA data from 1975-2021 and statistics from the National Center for Health Statistics (part of the Centers of Disease Control and Prevention (CDC).

4 Constraints within solution space

Resources available. This is an academic capstone project with set time constrains and no additional funding for independent data gathering.

5 Stakeholders to provide key insight

Dept. Secretary of US Dept of Transportation – Validate Policy
NHTSA Officials – Key Findings / Datasets Recommendations
News Outlets / Public – Awareness / Further Studies

6 Key data sources

NHTSA File Downloads | NHTSA

Workbook: NRSS Dashboard (dot.gov)

Implementing the National Roadway Safety Strategy | US Department of Transportation

FastStats - Leading Causes of Death (cdc.gov)

NCSA | Tools, Publications, and Data (dot.gov)

H D E I P

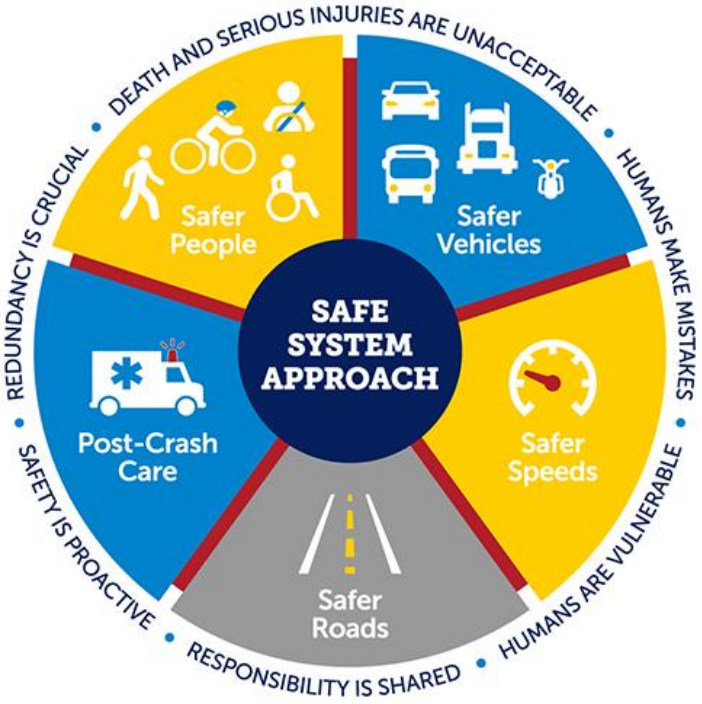
PURPOSE OF THE STUDY AND BACKGROUND

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.

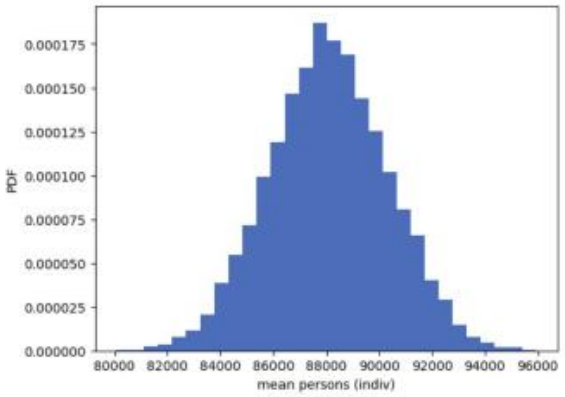
This study uses the NHTSA data from 1975-2021 to take a data-driven look at these claims, assess the appropriateness of these programs, and recommend areas for further analysis.



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 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.

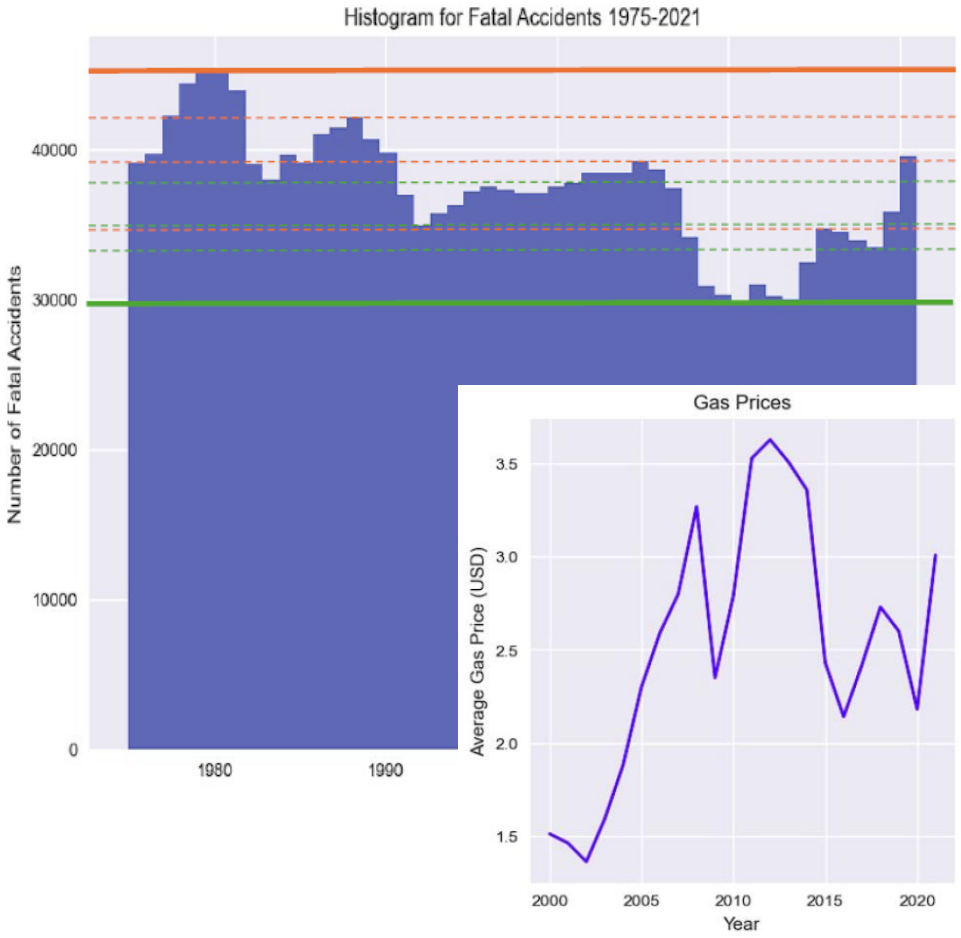
- The percentage increase in fatalities from 2020-2021 is 9.04% and does equal the largest percentage increase in the dataset (though it is less than the claimed 10.5%).

- 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.

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		

DOES THE DATA SUPPORT THE CLAIM?

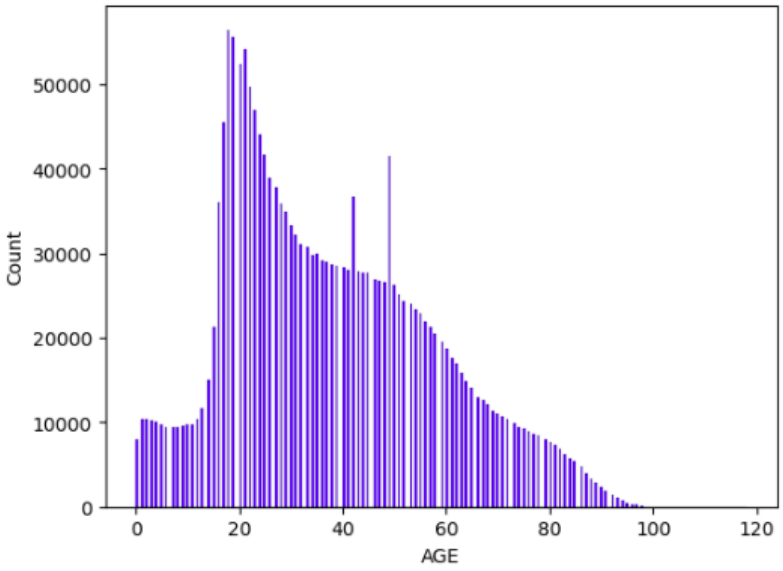
- The increase in fatal accidents from 2020-2021 is significant. It is the largest single increase year to year since 1975.
- The 2021 total, however, is significantly less than the highs of 1980-1981 and 1989. The total just slightly exceeds that of 2005.
- There are still two levels of resistance before the current rising trend in fatal accidents, that began in 2014, hits an all time high.
- COVID-19 and gas prices higher than $3/gal. may explain the 2000-2021 pattern.

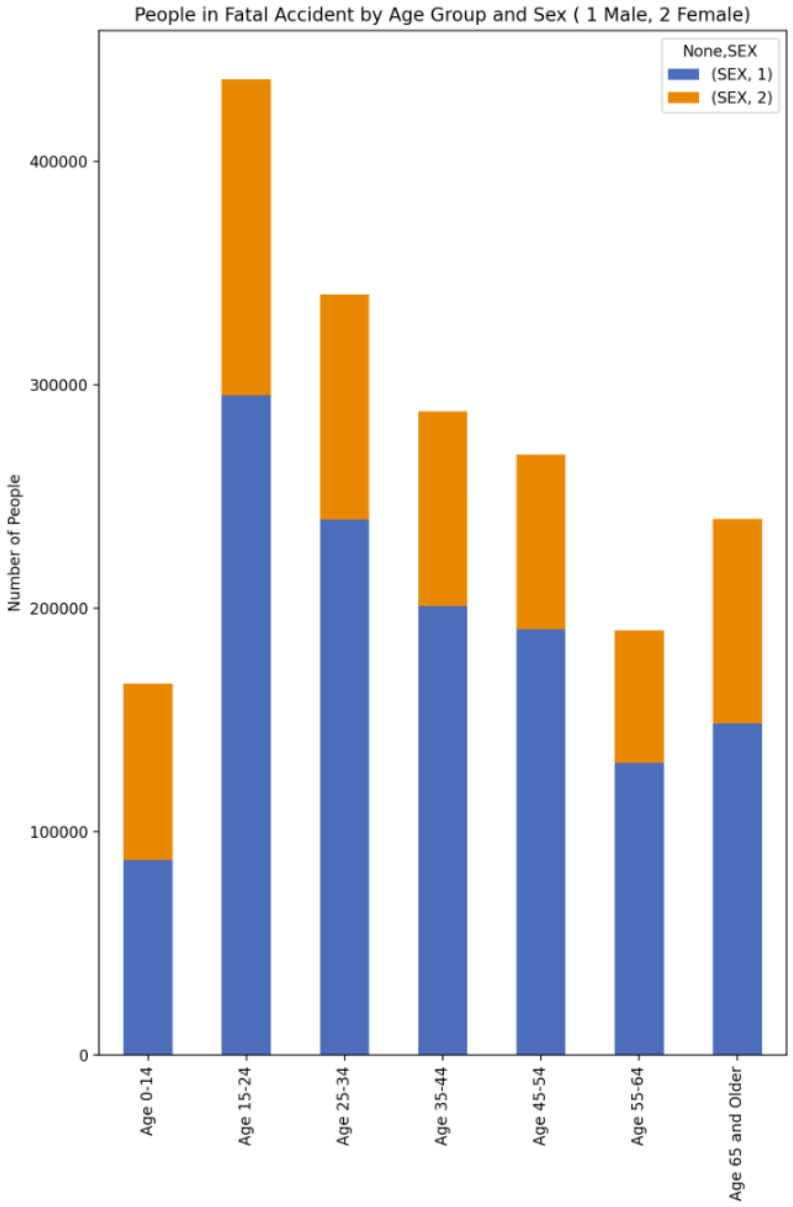


43% OF PEOPLE IN FATALITY CAUSING VEHICLE ACCIDENTS DIE

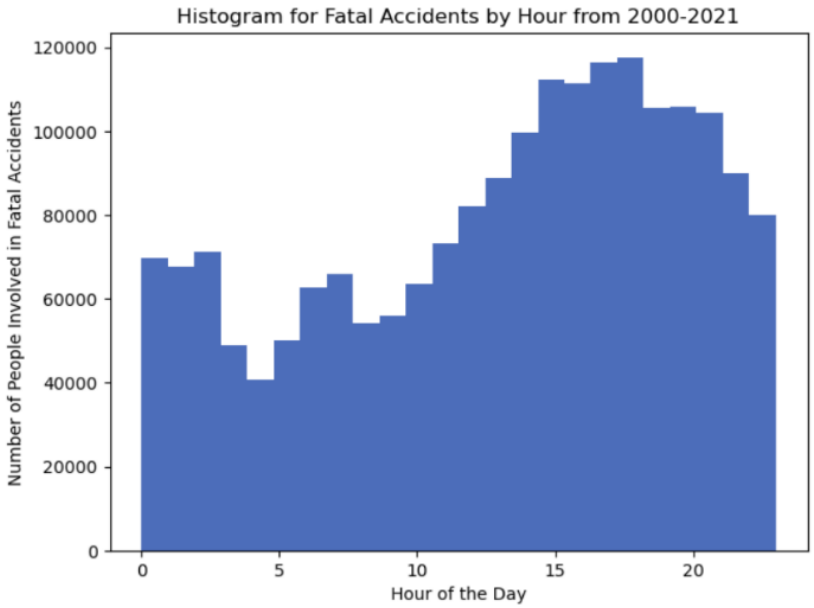
- From 2000-2021, 842,445 people died from a population of 1,938,242 involved in fatality causing vehicle accidents.

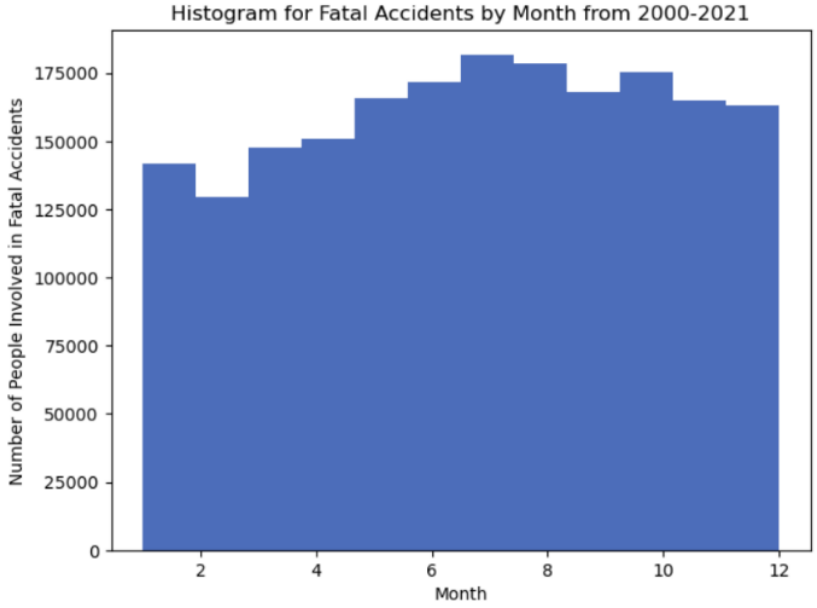
- Both of these charts show that policies and investments targeting males between the ages of 15-24 could yield a 10% reduction in accidents, but does this correlate to behaviors and fatalities?





TIME OF DAY AND MONTH DRIVING PATTERNS

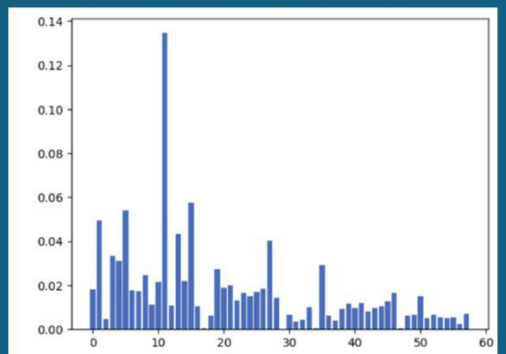




IDENTIFYING THE MOST IMPORTANT FEATURES IN THE DATA FOR ANALYSIS

- feature coefficient values of four different models
- The top ten common features were: ALC_RES, REST_USE, AGE, INJ_SEV, PER_TYP, SEAT_POS, EMER_USE, EJ_PATH, DRINKING, and VE_FORMS.
- I selected MAKE as well. It was the second most important feature in 3 of the 4 models.

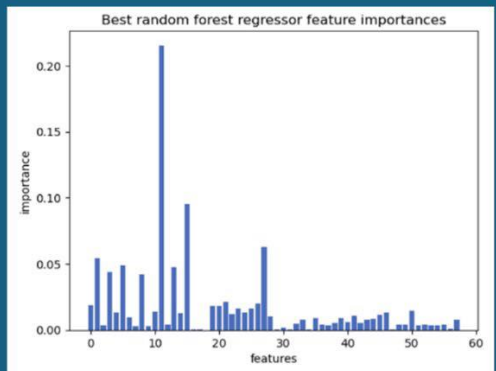
RandomForestClassifer



RandomForestClassifier (unscaled) results

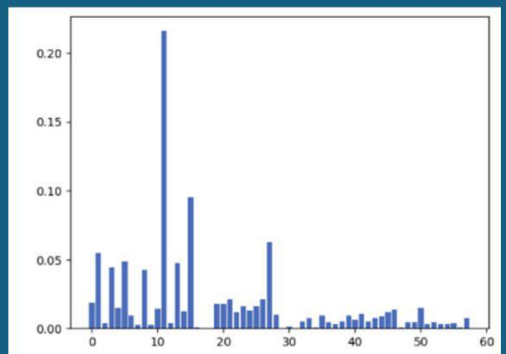
1. 11 - ALC_RES (0.14664)
2. 15 - MAKE (0.06631)
3. 5 - REST_USE (0.05422)
4. 1 - AGE (0.04929)
5. 13 - INJ_SEV (0.04339)
5. 27 - VE_FORMS (0.04047)
6. 3 - PER_TYP (0.03344)
7. 4 - SEAT_POS (0.03201)
8. 35 - SP_JUR (0.02918)
9. 19 - EMER_USE (0.02743)
10. 8 - EJ_PATH (0.02451)

RandomForestRegressor



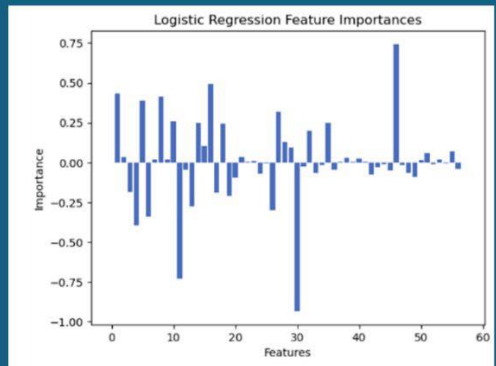
RandomForestRegressor (scaled) results

1. 11 - ALC_RES (0.21555)
2. 15 - MAKE (0.09546)
3. 27 - VE_FORMS (0.06294)
4. 1 - AGE (0.05431)
5. 5 - REST_USE (0.05422)
6. 13 - INJ_SEV (0.04766)
7. 3 - PER_TYP (0.04420)
8. 8 - EJ_PATH (0.04185)
9. 21 - MOD_YEAR (0.02104)
10. 26 - MINUTE (0.02026)



DecisionTreeClassifier (scaled) results

1. 11 - ALC_RES (0.21578)
2. 15 - MAKE (0.09508)
3. 27 - VE_FORMS (0.06264)
4. 1 - AGE (0.05472)
5. 5 - REST_USE (0.04888)
6. 13 - INJ_SEV (0.04763)
7. 3 - PER_TYP (0.04449)
8. 8 - EJ_PATH (0.04231)
9. 21 - MOD_YEAR (0.02076
10. 26 - MINUTE (0.02074)



LogisticRegression (scaled) results

1. 30 - SCH_BUS (-0.93326)
2. 46 - CITY (0.74244)
3. 11 - ALC_RES (-0.72795)
4. 16 - BODY_TYP (0.49188)
5. 1 - AGE (0.43174)
6. 8 - EJ_PATH (0.04231)
7. 4 - SEAT_POS (-0.39386)
8. 5 - REST_USE (-0.39386)
9. 6 - AIR_BAG (-0.33808)
10. 27 - VE_FORMS (0.31859)

DecisionTreeClassifier

LogisticRegression

BLOOD ALCOHOL (BAC) LEVELS

- Blood Alcohol levels (ALC_RES in the dataset) had the highest positive coefficient for all models evaluated.

- Average BAC for all age groups is higher in men (1) by 3%.

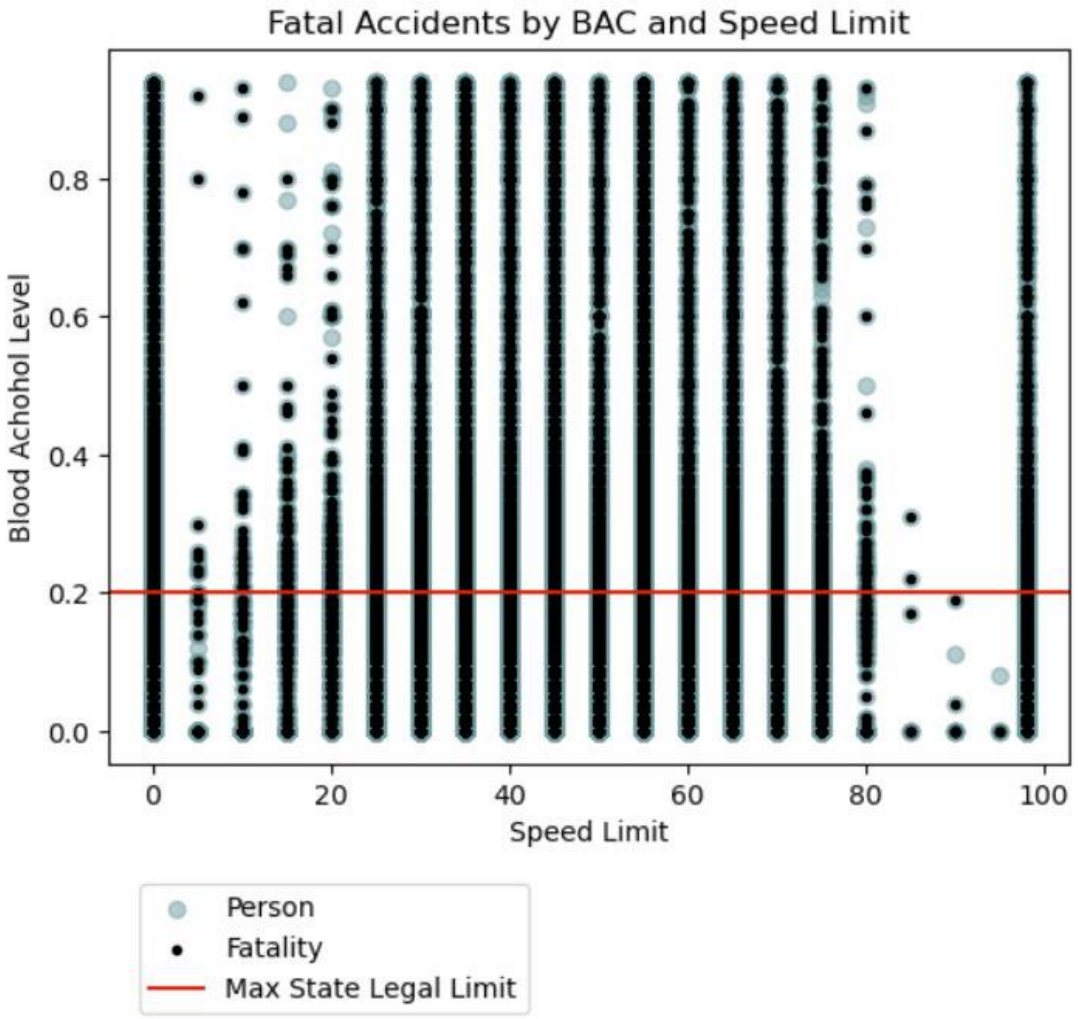
- The highest average levels are highest in ages 25-54. The highest ages group is 25-34.

- 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.

	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

BLOOD ALCOHOL (BAC) LEVELS

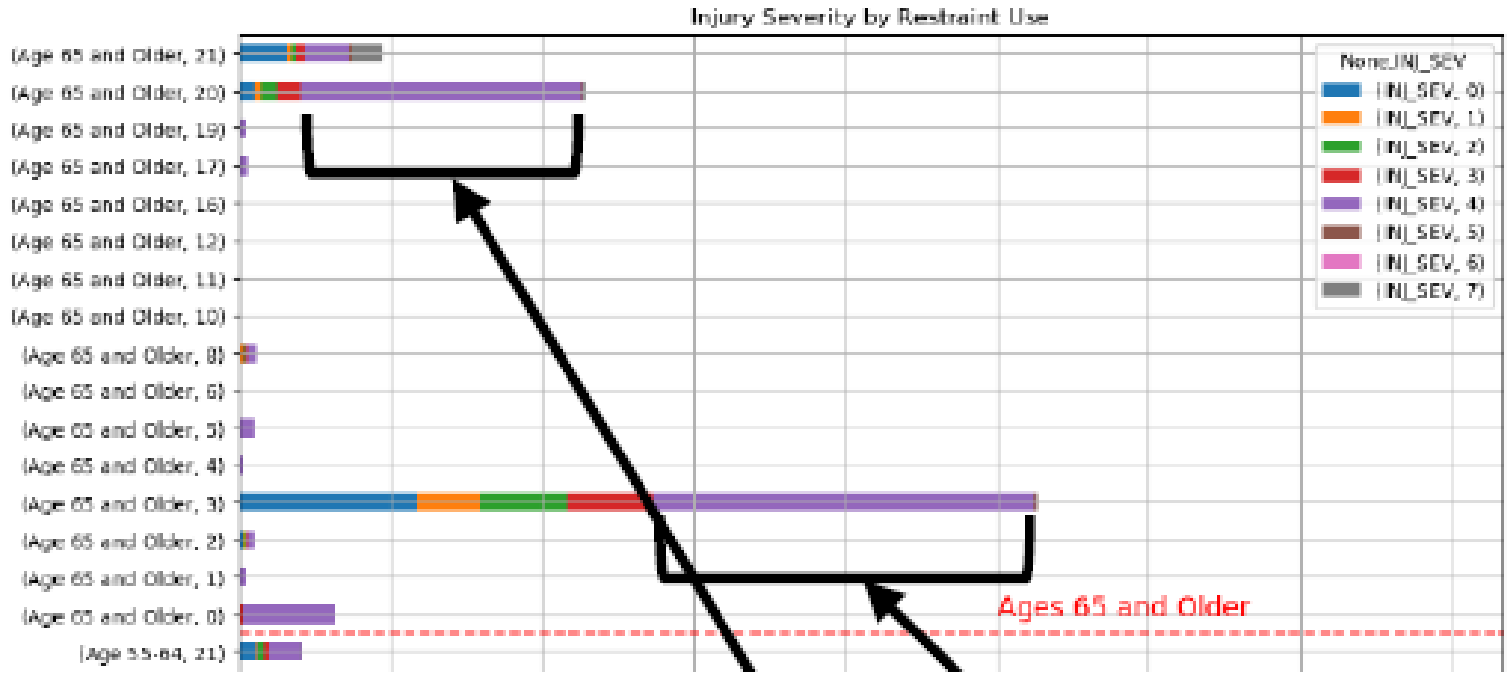
- Drivers over the BAC legal limits in all states are driving around in neighborhoods (25-35 speed limit), rural roadways (40-55), and highways (55-70).
- The dense line at 0, is for No Statutory Limit/Non-Trafficway or Driveway Access.
- These include entrance and exit ramps, neighborhood streets, and off-road areas.



RESTRAINT (SEAT BELT, CHILD SEAT, HELMET) USE

% 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

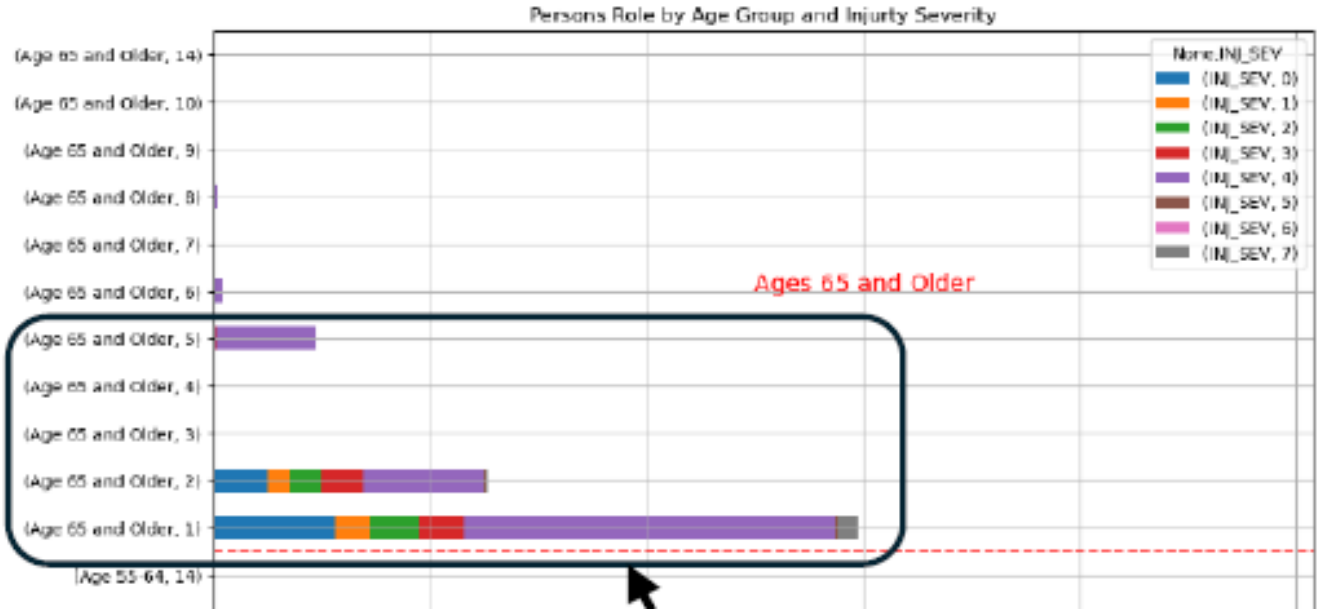
Use of seatbelts became law in 1968



ROLE OF THE PERSON INVOLVED IN THE CRASH

- Following up on the age analysis the driver to passenger ratio does narrow at age 65 and above (more are passengers), but the majority are still driving.

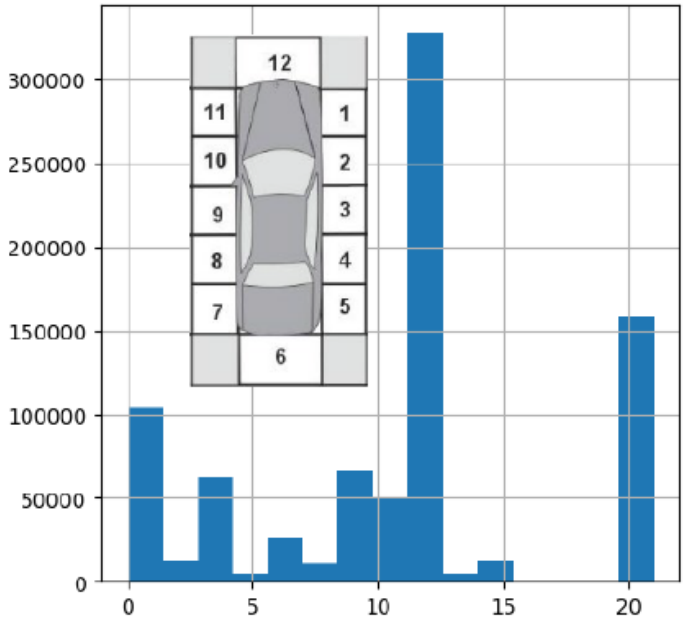
- In both cases, the fatality ratios are the highest of all age groups to include those that are riding bicycles.



SEAT POSITION

- Sitting further back and in the middle seat of the vehicle is a good strategy. T
- Most 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%

```
% 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
```



VEHICLE MAKE

Involved in an Accidents with Fatalities by Make

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

- 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


UNLEASH THE BEAST IN YOU
DURANGO
CHARGER
CHALLENGER
NITRO
The Dodge Fast & Fierce Weekend
From June 10 to 12, choose the best beast for you.

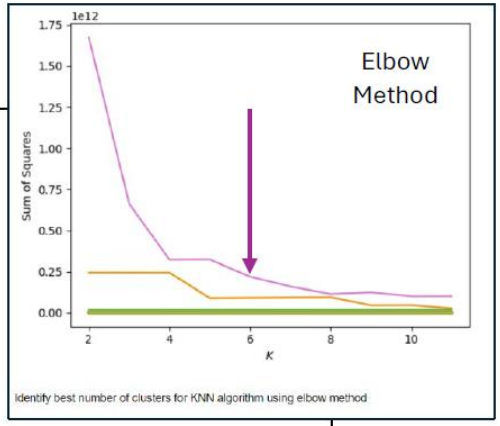
MODELING – VEHICLE ACCIDENT FATALITIES ARE PREDICTABLE

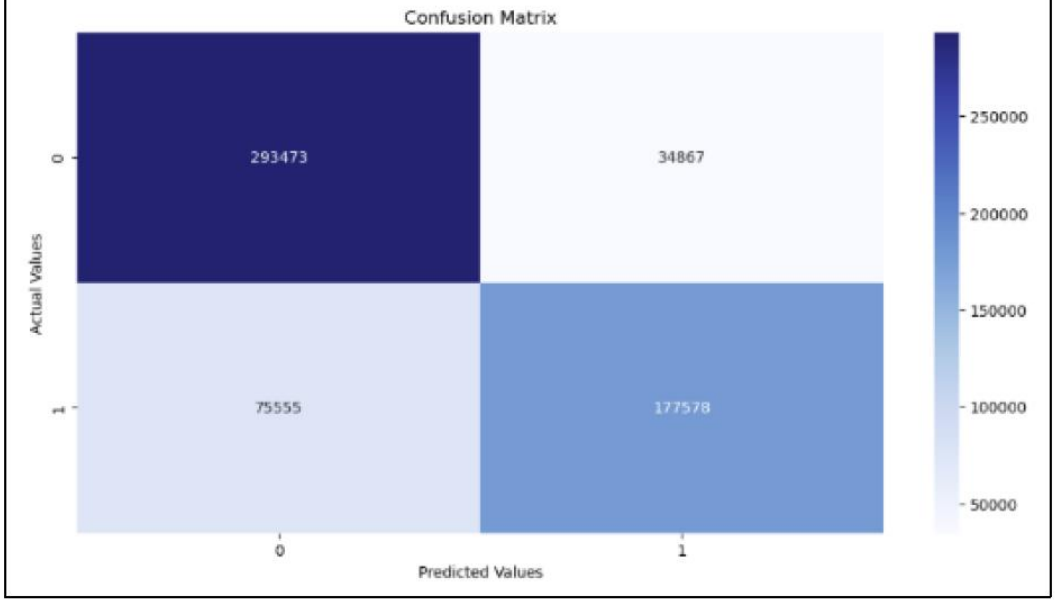
Best Predictive Model (KNN) Scores

```
The Accuracy for Test Set is 81.00995231076938
[[293473  34867]
 [ 75555 177578]]
              precision    recall  f1-score   support

           0       0.80      0.89      0.84    328340
           1       0.84      0.70      0.76    253133

    accuracy                           0.81    581473
   macro avg       0.82      0.80      0.80    581473
weighted avg       0.81      0.81      0.81    581473
```





CONCLUSIONS

- The Department of Transportations claim of a crisis appears premature.
- 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.
- The increase in 2021 may be a result of the law of large numbers and a return to the norm instead of a crisis.
- The percentage of fatal injuries decreased from a high in 2020 of 41.59% to 40.84% in 2021 despite the increases in fatal accidents and persons involved.

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

RECOMMENDATIONS

- 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%.

- 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

- 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 alcohol, the max. legal limit in the United States is 0.20 BAC. Of those who died, 94,803 had a blood alcohol content of 0.20 or higher.

- 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 alcoholic drinks is undeniably a choice, it is not a mistake. This population accounts for 7.92% of the fatalities.