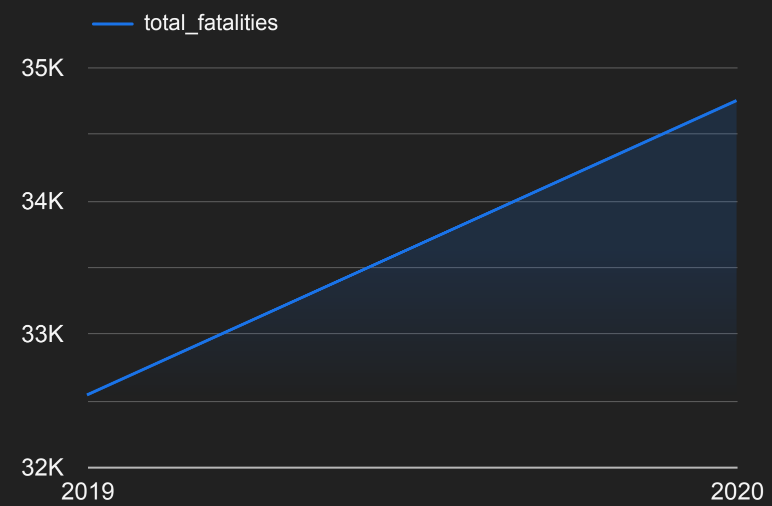
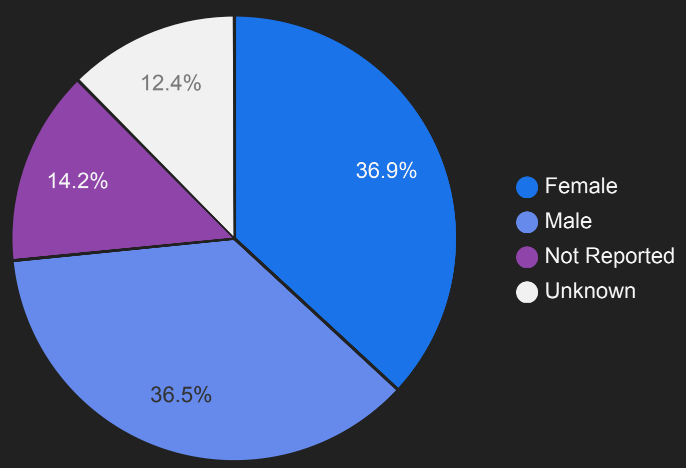
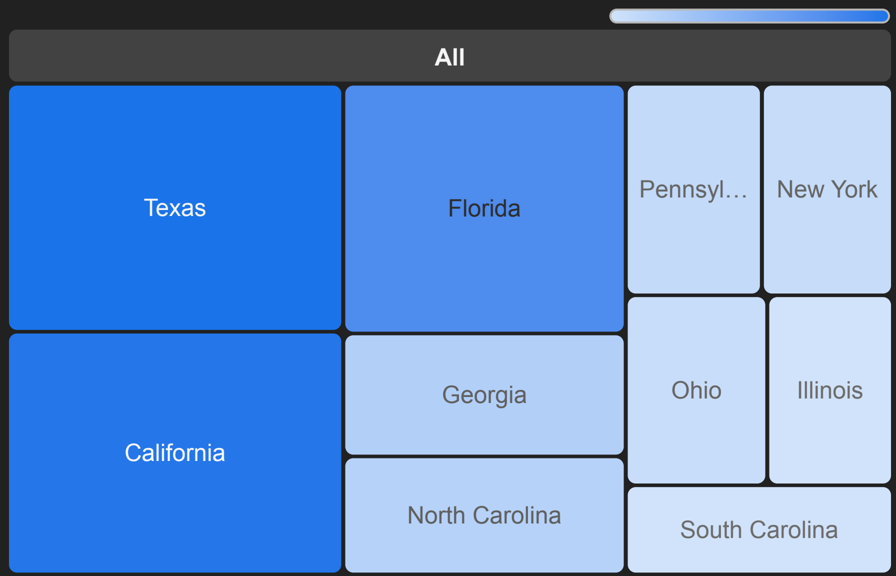
Data-Driven Insights into U.S. Traffic Fatalities: A BigQuery-Powered Analysis



total_fatalities
81,405





Detailed Overview

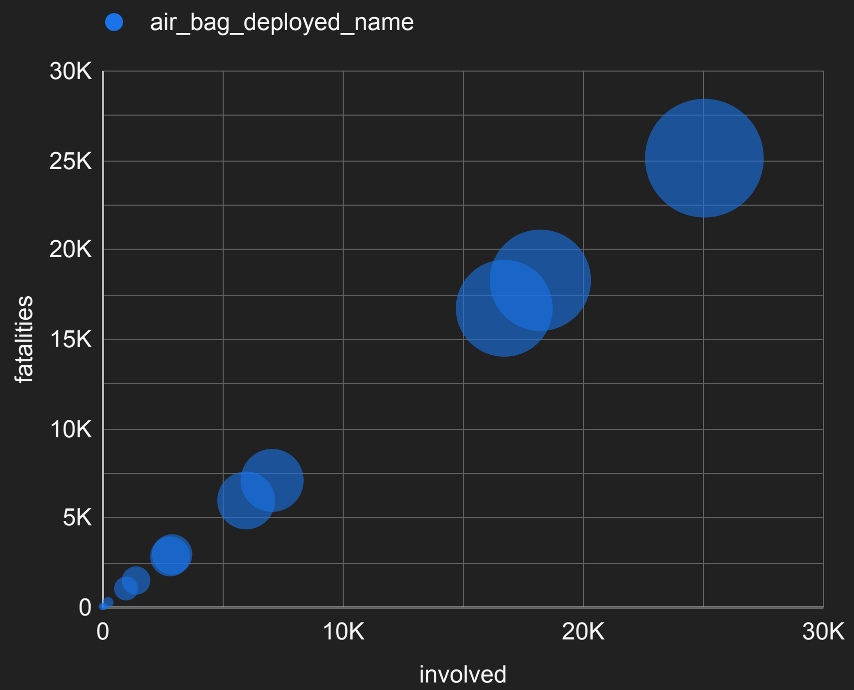
Data-Driven Insights into U.S. Traffic Fatalities: A BigQuery-Powered Analysis

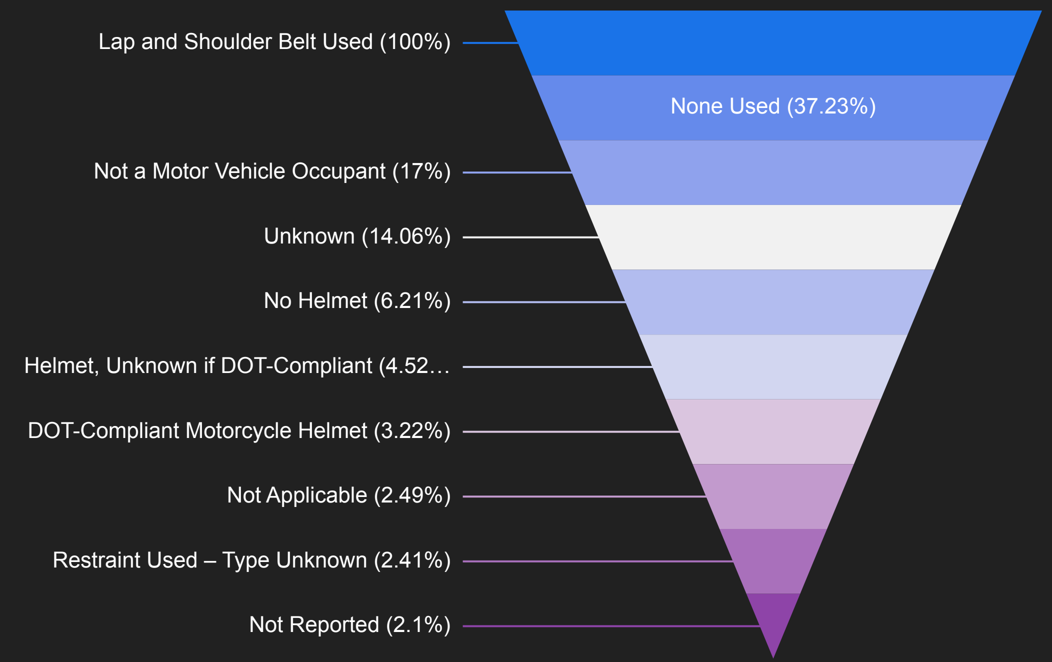
Overview

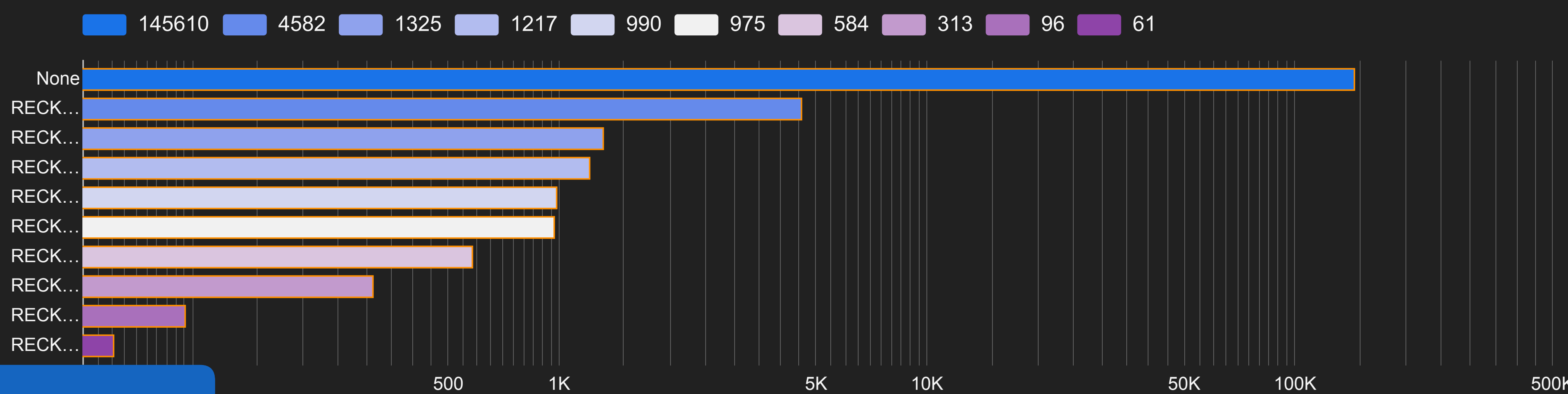
Safety

Behaviour & Timing

Vehicle & Environment







Detailed Overview

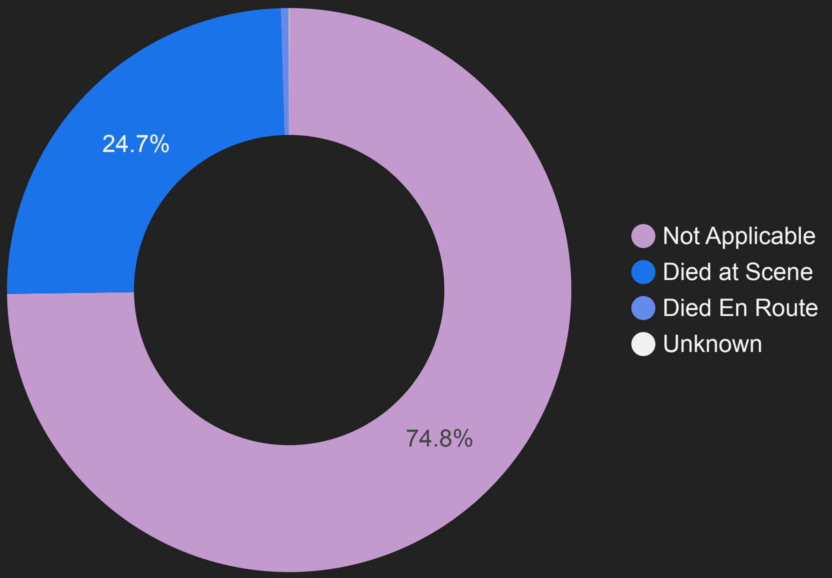
Data-Driven Insights into U.S. Traffic Fatalities: A BigQuery-Powered Analysis

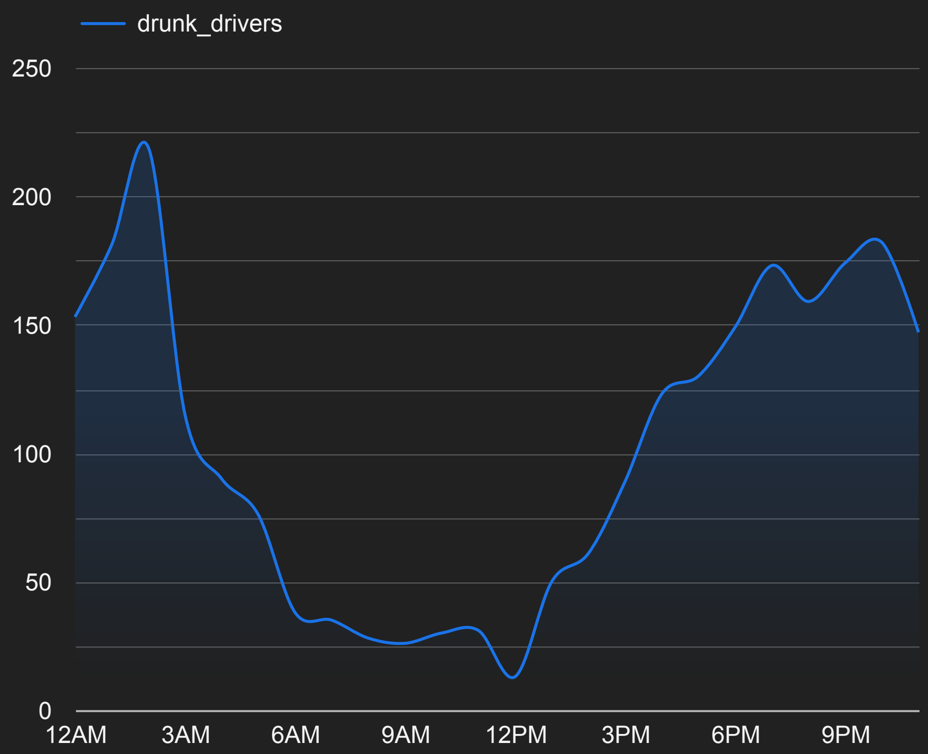
Overview

Safety

Behaviour & Timing

Vehicle & Environment





Detailed Overview

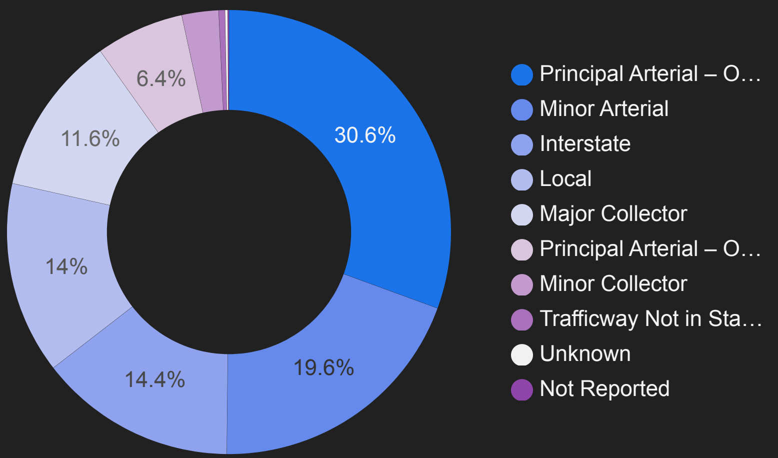
Data-Driven Insights into U.S. Traffic Fatalities: A BigQuery-Powered Analysis

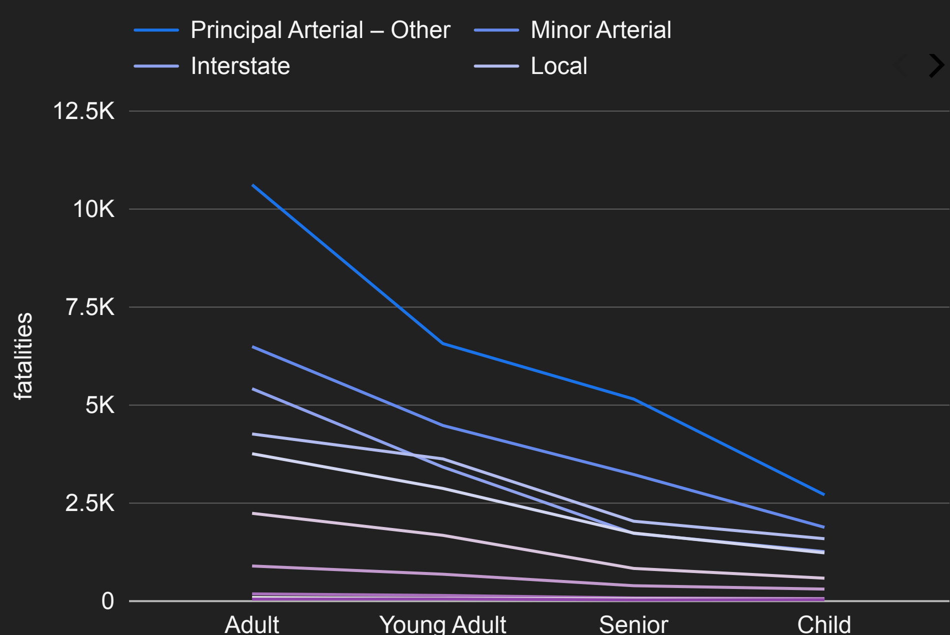
Overview

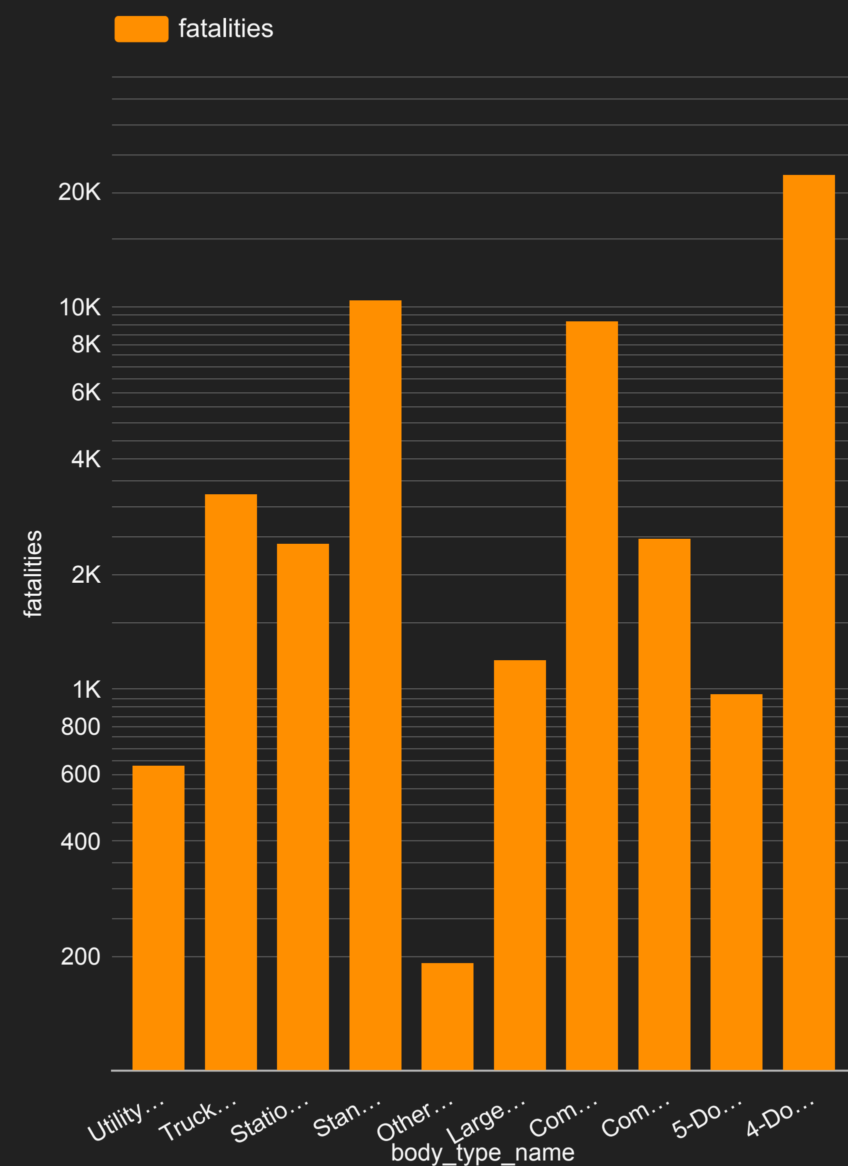
Safety

Behaviour & Timing

Vehicle & Environment







Detailed Overview

Introduction

Why traffic safety?

Why this dataset?

"Motor vehicle crashes are one of the leading causes of preventable death in the U.S. This project explores patterns and risks across multiple factors — from drunk driving to restraint usage — to uncover life-saving insights."

What research says so far

NHTSA 2022: Drunk driving accounts for ~30% of crash fatalities.
CDC: Wearing a seatbelt reduces risk of death by 45%.
WHO: Speeding and nighttime driving are global risk factors.
Nighttime and weekend crashes are statistically more deadly.
Environmental factors like poor visibility, rain, and road curvature significantly increase crash risk.

While individual factors like alcohol or speed are well-documented, our approach explores their interactions using a holistic dashboard

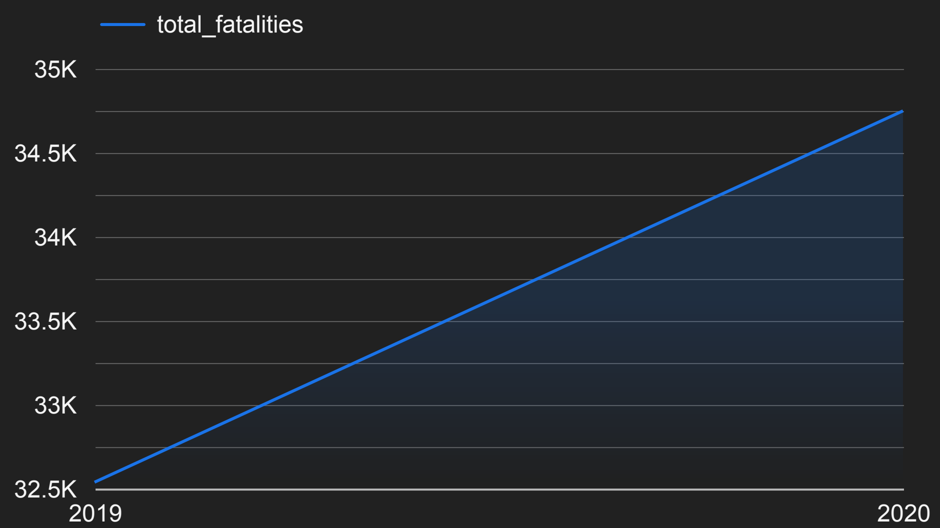
Data Provenance

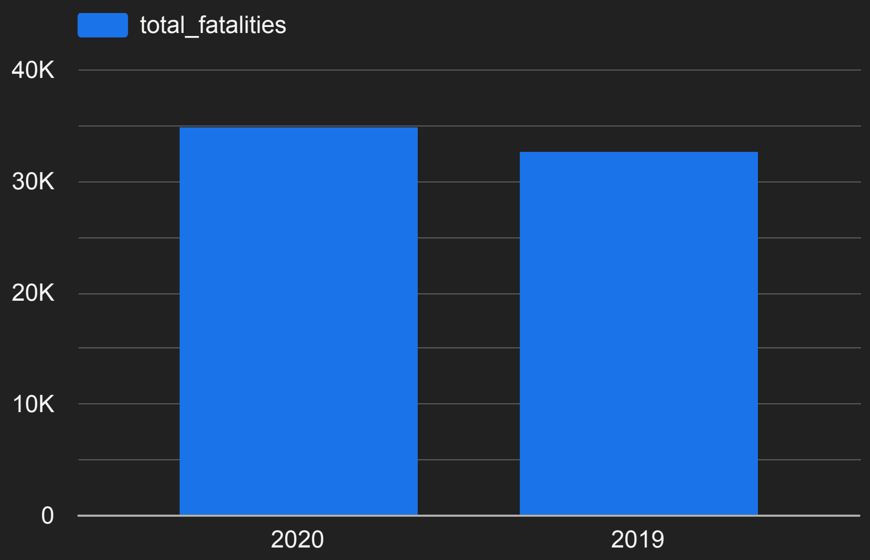
The data for this project comes from the National Highway Traffic Safety Administration (NHTSA), specifically from its Fatality Analysis Reporting System (FARS), which is made available publicly through Google BigQuery under the dataset bigquery-public-data.nhtsa_traffic_fatalities. This dataset contains comprehensive, standardized information on motor vehicle crashes involving fatalities in the United States.

We ensured semantic consistency by using well-documented foreign keys like consecutive_number and state_number to accurately join across tables. Each table is provided and updated by NHTSA and adheres to a strict reporting standard from local law enforcement, medical examiners, and transportation departments.

Total Fatalities - How deadly are U.S. roads?

total_fatalities over time





This section provides a high-level view of total traffic fatalities across the years. It highlights annual trends and serves as a foundation for deeper analysis.

Fatalities increased from around 32,000 in 2019 to 34,000+ in 2020, despite fewer vehicles on the road during pandemic lockdowns.

This suggests riskier behavior during COVID, such as speeding on emptier roads or reduced law enforcement presence.

total_fatalities

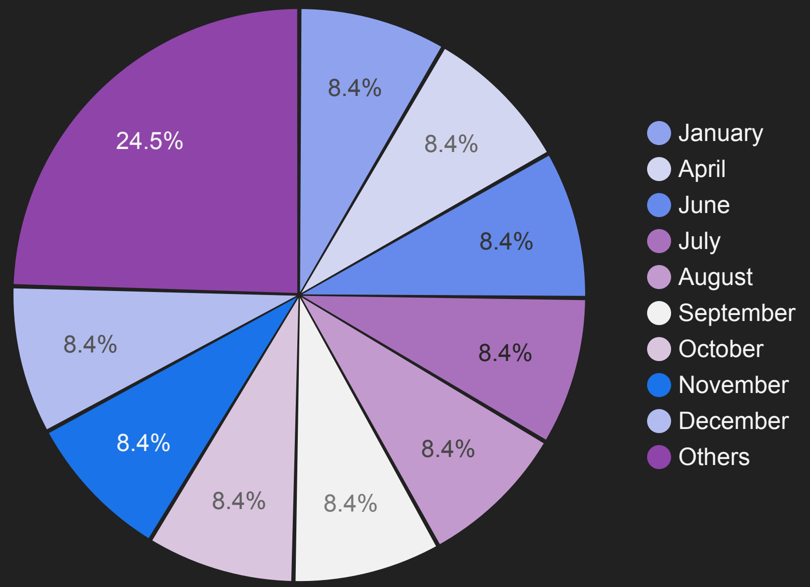
total_fatalities
67,286

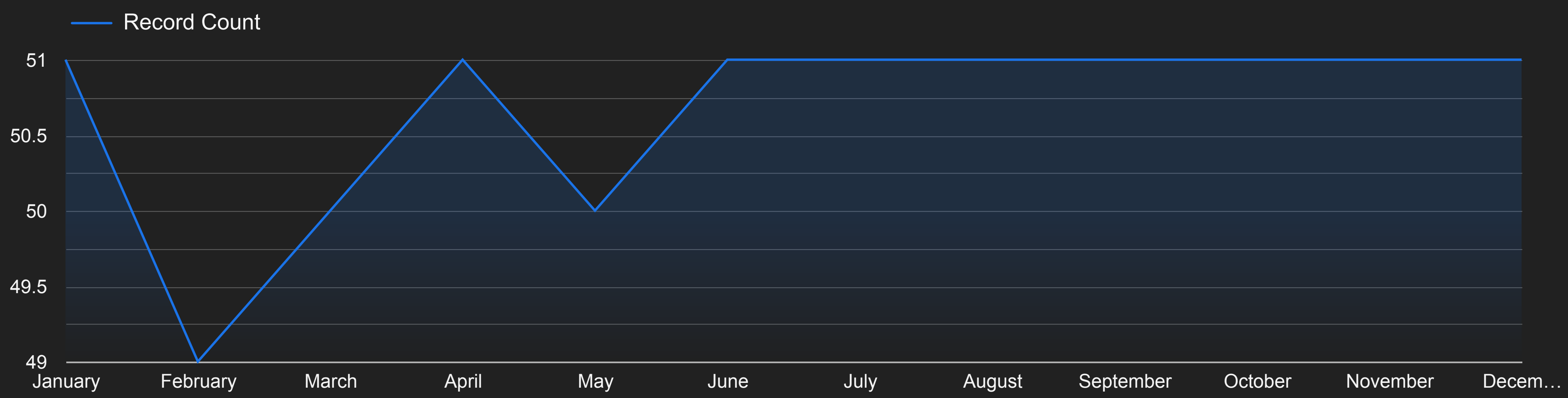
Fatalities by Month (from accident tables)

When do most fatal crashes occur?

Here, we examine the seasonality of traffic fatalities. Visualising crashes by month helps identify high-risk periods and informs time-based prevention strategies.

Fatalities tend to spike during summer and late fall, particularly in July, October, and November.
These months align with holiday travel periods and seasonal changes, indicating the need for targeted awareness campaigns during these times.





Record Count
608

total_crashes
32,538

Fatalities by State (from accident tables)

	state_name	Record Count ▾
1.	California	3,540
2.	Texas	3,427
3.	Florida	2,935
4.	Georgia	1,424
5.	North Carolina	1,348
6.	Pennsylvania	1,088
7.	Ohio	1,053
8.	Illinois	1,000
9.	New York	981
10.	Michigan	981
11.	Alabama	979
12.	Tennessee	962
13.	South Carolina	941
14.	Missouri	872
15.	Arizona	856
16.	Indiana	776
17.	Kentucky	763
18.	Virginia	722
19.	Louisiana	704
20.	Oklahoma	628

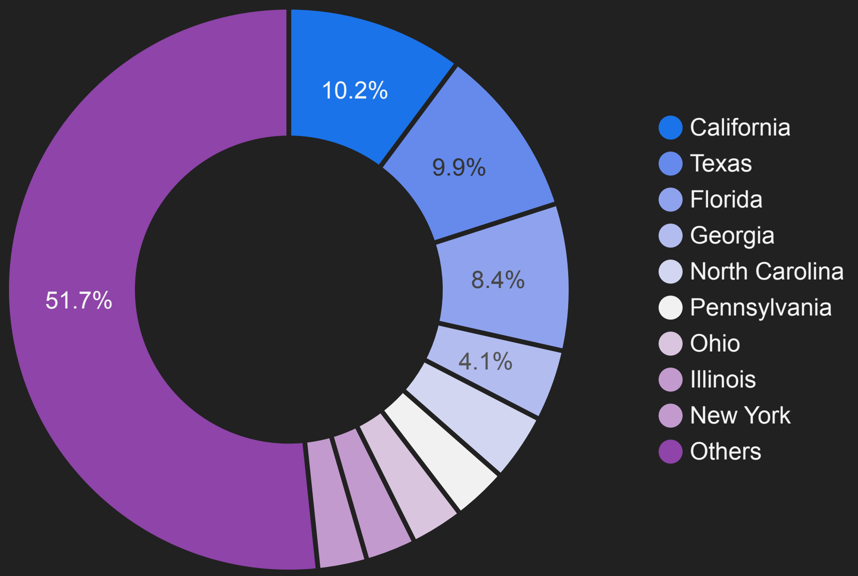
1 - 51 / 51

Where do the most fatalities occur?

This page breaks down fatal crashes by state, helping identify geographic hotspots. States with higher fatality counts can be prioritized for policy intervention.

California, Texas, and Florida lead in total fatalities, which is consistent with their high population and vehicle density.

These states are high-priority regions for policy reform and traffic enforcement initiatives.



Victim Demographic Profile

	sex	fatalities
1.	Not Reported	1
2.	Unknown	1
3.	Unknown	2
4.	Not Reported	2
5.	Unknown	4
6.	Female	171
7.	Female	311
8.	Female	211
9.	Male	1
10.	Male	5
11.	Male	859
12.	Male	214
13.	Male	199
14.	Female	2
15.	Female	1
16.	Female	125
17.	Female	391
18.	Male	207
19.	Female	707
20.	Female	659
21.	Female	692
22.	Female	678

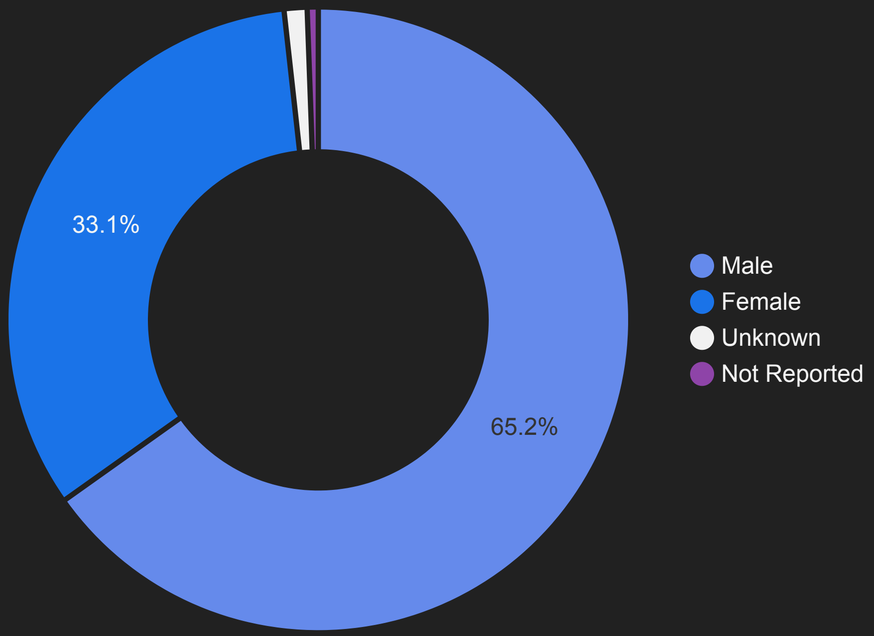
1 - 100 / 206

Who are the victims of fatal crashes?

We analyse the demographic distribution (age, sex) of fatalities to uncover vulnerable populations and tailor safety campaigns.

The majority of fatalities are among males, who make up over 65% of deaths.

Combined with behavioral insights, this suggests that risk-taking behavior (e.g., speeding, DUI) is more prevalent among men, especially younger males.



Who gets hurt the most? - Injury Severity and Role

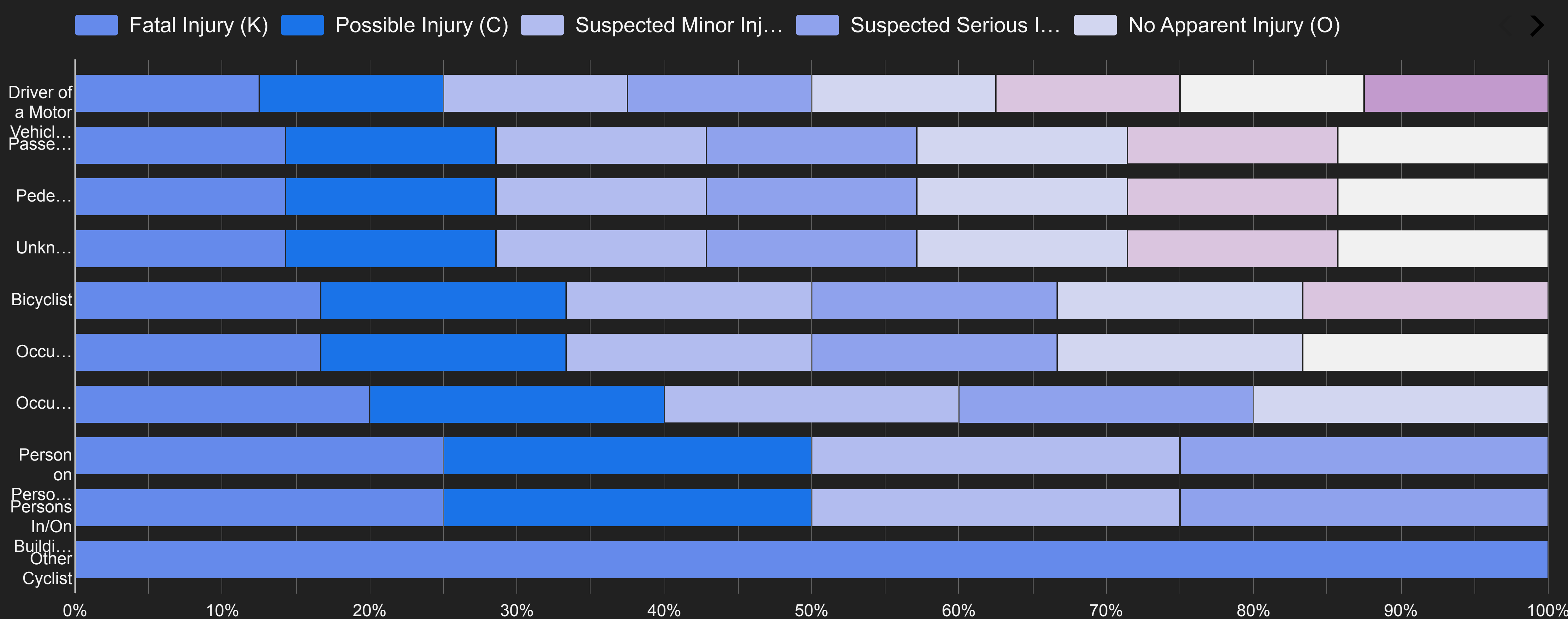
	person_type_name	injury_severity_name	count
1.	Driver of a Motor Vehicle In-Transport	Died Prior to Crash	2
2.	Driver of a Motor Vehicle In-Transport	Fatal Injury (K)	22,350
3.	Passenger of a Motor Vehicle In-Transport	Fatal Injury (K)	6,504
4.	Pedestrian	Fatal Injury (K)	5,494
5.	Bicyclist	Fatal Injury (K)	828
6.	Person on Personal Conveyances (Since 2007)	Fatal Injury (K)	160
7.	Unknown Occupant Type in a Motor Vehicle In-Transport	Fatal Injury (K)	74
8.	Occupant of a Motor Vehicle Not In-Transport	Fatal Injury (K)	32

1 - 56 / 56

Who gets hurt the most and how badly? This section links injury outcomes with person type (e.g., drivers, passengers, pedestrians), giving insight into crash impact across different roles.

The highest fatality rates are among motor vehicle drivers, followed by pedestrians and bicyclists.

This indicates that non-occupant road users are still at substantial risk, warranting better urban design and crosswalk safety measures.



Seatbelt / Helmet Use vs Fatalities

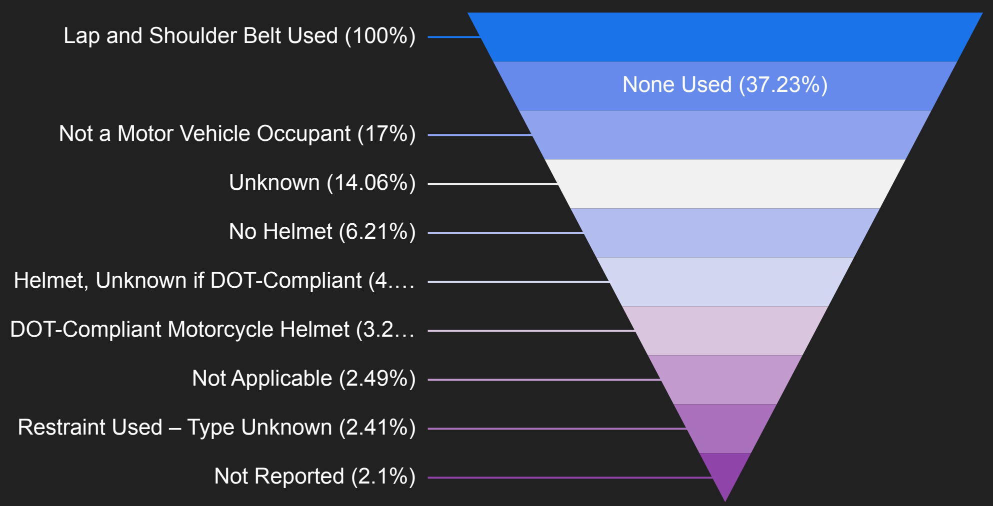
Do safety measures save lives?

We compare fatality rates across different restraint types. This analysis demonstrates the life-saving importance of helmets and seatbelt usage.

A large number of fatalities involved people not wearing seatbelts or helmets.

Only 41,560 fatalities involved known use of lap and shoulder belts. 15,474 deaths occurred when no restraint was used at all.

Stronger enforcement and education around seatbelt and helmet use can significantly reduce fatalities.



	restraint_system_helmet_use_name	total_victims ▾
1.	Lap and Shoulder Belt Used	41,560
2.	None Used	15,474
3.	Not a Motor Vehicle Occupant	7,064
4.	Unknown	5,843
5.	No Helmet	2,581
6.	Helmet, Unknown if DOT-Compliant	1,879
7.	DOT-Compliant Motorcycle Helmet	1,338
8.	Not Applicable	1,034
9.	Restraint Used – Type Unknown	1,000
10.	Not Reported	874
11.	Child Restraint Type Unknown	838
12.	Lap Belt Only Used	602
13.	Child Restraint System – Forward Facing (Since 2008)	539
14.	Booster Seat	247
15.	Shoulder Belt Only Used	196
16.	Child Restraint System – Rear Facing (Since 2008)	188
17.	Helmet, Other than DOT-Compliant Motorcycle Helmet	168
18.	Unknown if Helmet Worn	125
19.	Other	70

1 - 19 / 19

Airbags and Survival

Do airbags reduce fatalities?

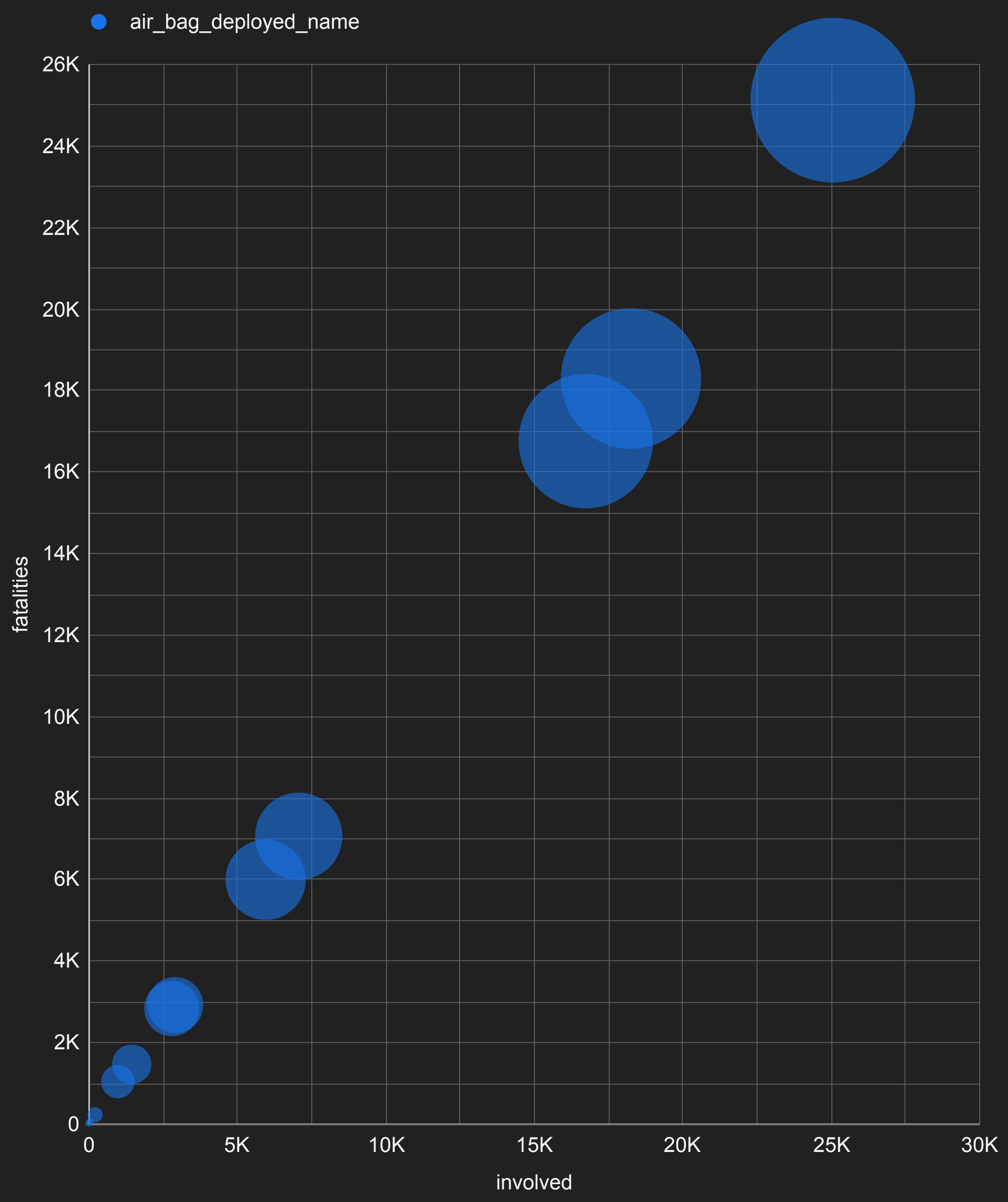
This section explores the relationship between airbag deployment and crash outcomes. It helps assess the effectiveness of airbag systems.

Over 25,000 fatalities occurred in vehicles where airbags did not deploy, compared to lower counts in vehicles with front or curtain airbags.

This implies that proper airbag deployment can reduce fatal outcomes, underlining the importance of modern, functioning restraint systems.

	air_bag_deployed_name	fatalities
1.	Not Deployed	25105
2.	Not Applicable	18287
3.	Deployed: Front	16729
4.	Not a Motor Vehicle Occupant	7064
5.	Deployed: Combination	5968
6.	Deployment Unknown	2908
7.	Deployed: Unknown Location	2825
8.	Deployed: Side (Door, Seatback)	1437
9.	Not Reported	1025
10.	Deployed: Curtain (Roof)	209
11.	Deployed: Other (Knee, Air Belt, etc.)	50
12.	Switched Off	13

1 - 12 / 12



Death Location (at scene vs en route)

Where do victims die—on scene or en route?

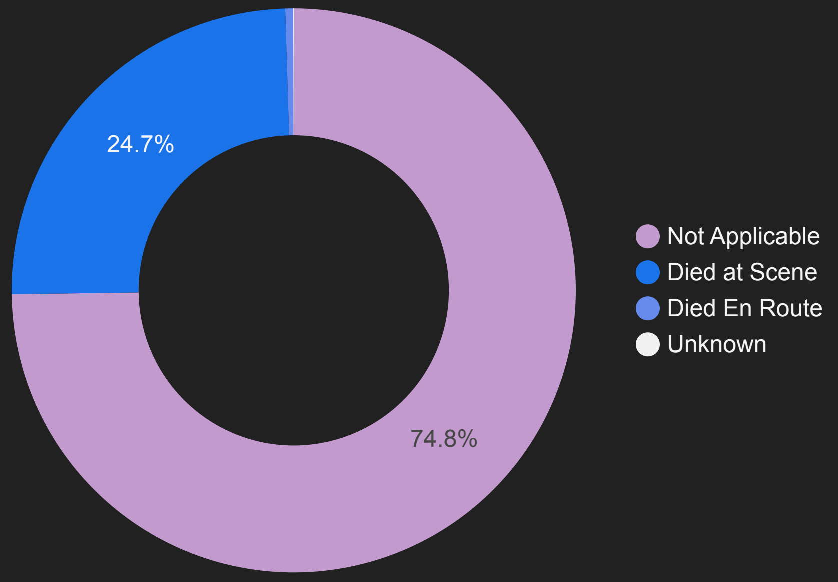
This page breaks down death locations to better understand emergency response effectiveness and crash severity.

About 24.7% died at the scene, while only a small fraction died en route.

This may point to the severity of crashes or delays in emergency response. Improvements in EMS infrastructure could save more lives.

	died_at_scene_en_route	total_deaths ▾
1.	Not Applicable	61,039
2.	Died at Scene	20,188
3.	Died En Route	369
4.	Unknown	24

1 - 4 / 4



Which states have the deadliest crashes per event?

	state_name	crash_…	avg_fatali…
1.	Utah	258	2.76356589…
2.	Washington	499	2.74749498…
3.	District of Columbia	23	2.73913043…
4.	Nevada	297	2.72390572…
5.	Arizona	811	2.69543773…
6.	Rhode Island	41	2.68292682…
7.	Hawaii	85	2.65882352…
8.	Texas	3190	2.64388714…
9.	Michigan	896	2.63504464…
10.	Alaska	60	2.63333333…
11.	Florida	2699	2.61689514…
12.	California	3123	2.60838936…
13.	North Dakota	111	2.60360360…
14.	New Mexico	269	2.59851301…
15.	New York	1061	2.54571159…
16.	New Jersey	521	2.54318618…
17.	Wyoming	129	2.54263565…
18.	Minnesota	375	2.54133333…

1 - 50 / 50

Rhode Island, Michigan, and Alaska had the highest average fatalities per crash (>2.6), despite having fewer total crashes.

This indicates that crashes in these areas, though less frequent, are more severe, possibly due to rural roads, speed, or emergency service delays.

Fatality Risk by Vehicle Type

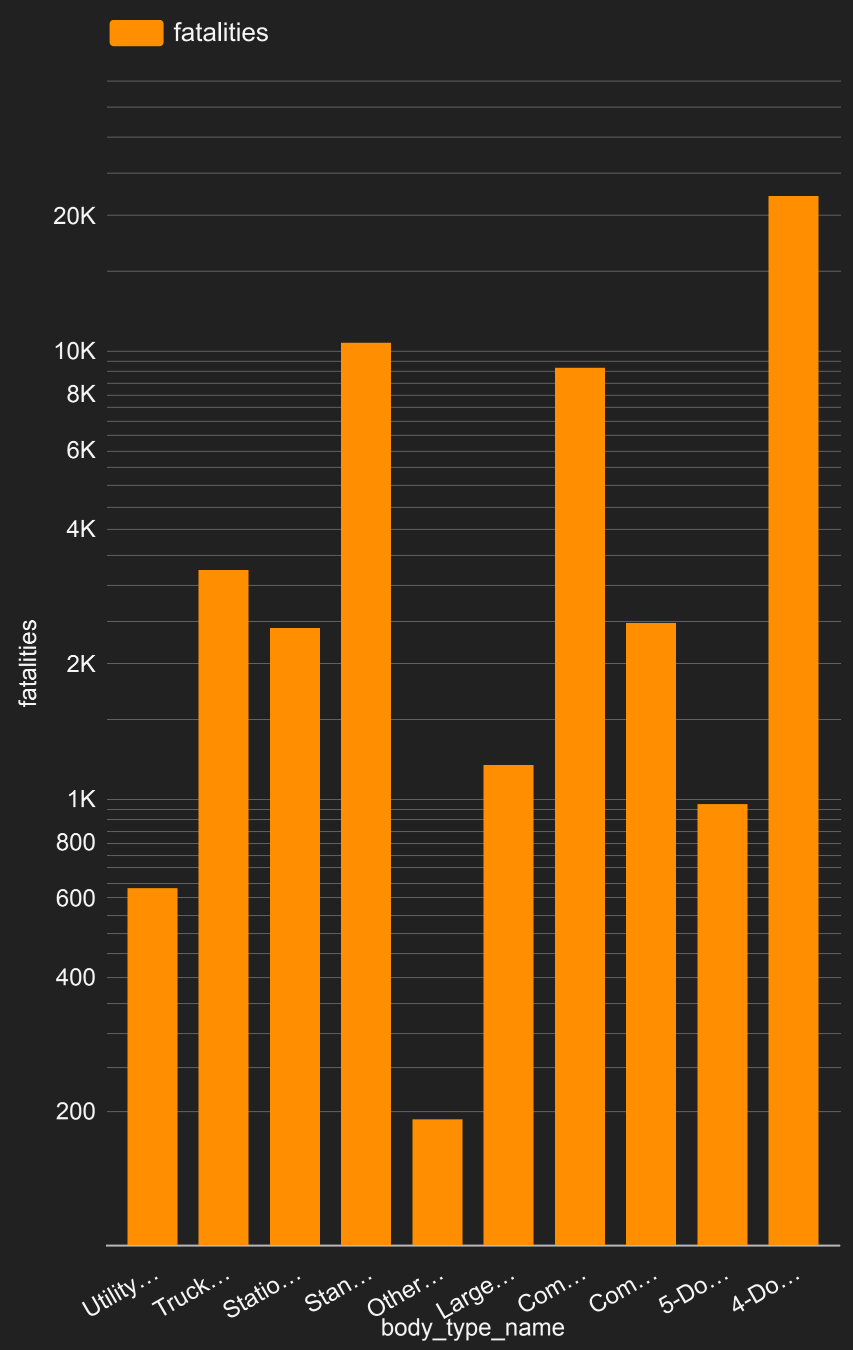
This page shows which vehicle types are most associated with fatalities.

~4-door sedans lead with over 22,000 fatalities.
~ Pickup trucks and compact utility vehicles also show high numbers.

While these vehicles are common, the data suggests potential design vulnerabilities or usage patterns (e.g., rural driving, lack of seatbelt use).

	body_type_name	fatalities
1.	Station Wagon (Excluding Van and Truck-Based)	2,393
2.	Utility Station Wagon	629
3.	Compact Pickup (Gross Vehicle Weight, GVWR, < 4,500 lbs)	2,462
4.	5-Door/4-Door Hatchback	967
5.	Other or Unknown Automobile Type	191
6.	4-Door Sedan/Hardtop	22,047
7.	Standard Pickup	10,369
8.	Compact Utility (ANSI D-16 Utility Vehicle Categories “Small” and “Midsize”)	9,117
9.	Truck/Tractor (Cab Only, or with Any Number of Trailing Units: Any Weight)	3,221
10.	Large Van – Includes Van-Based Buses	1,187

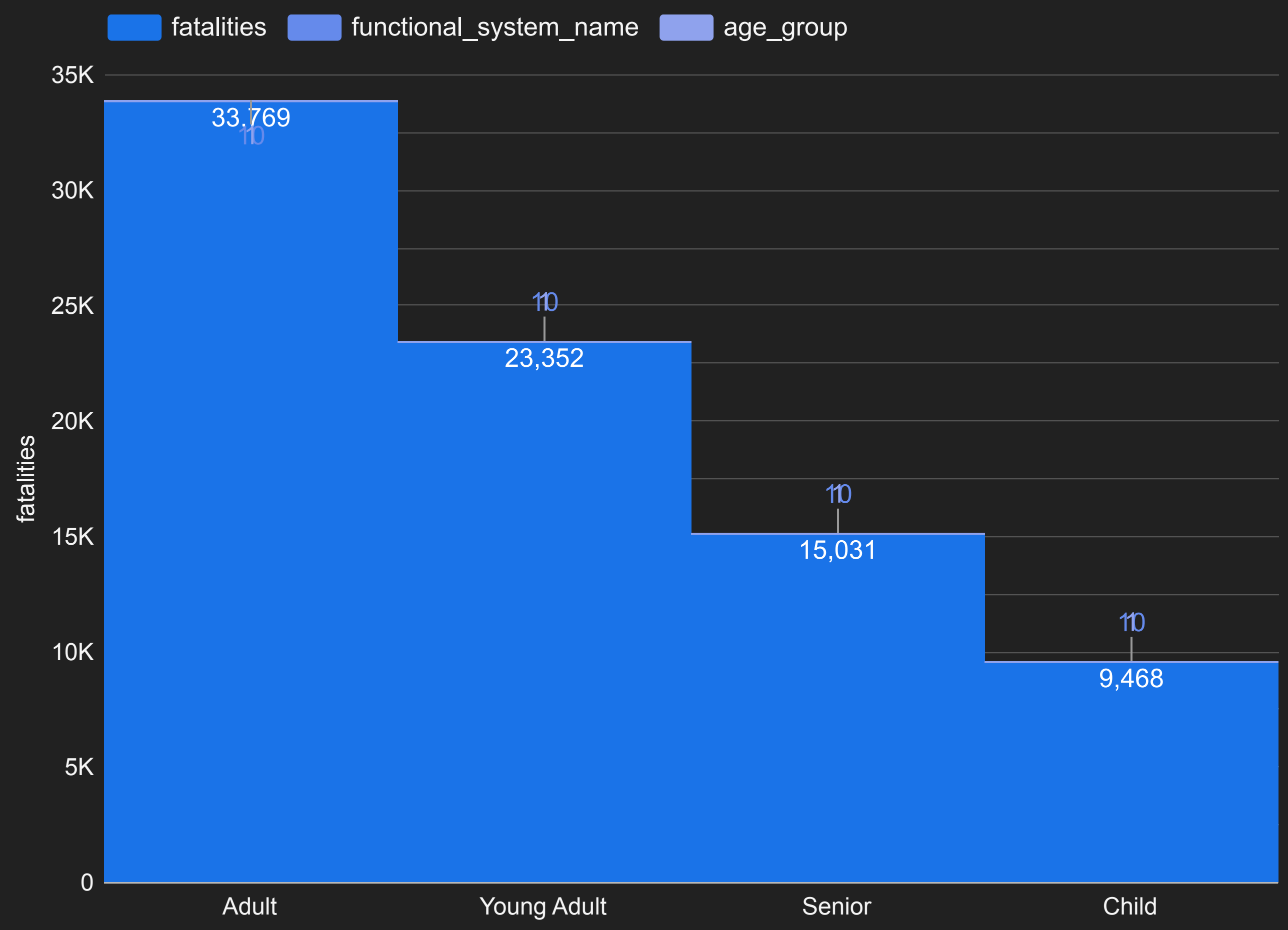
1 - 10 / 10

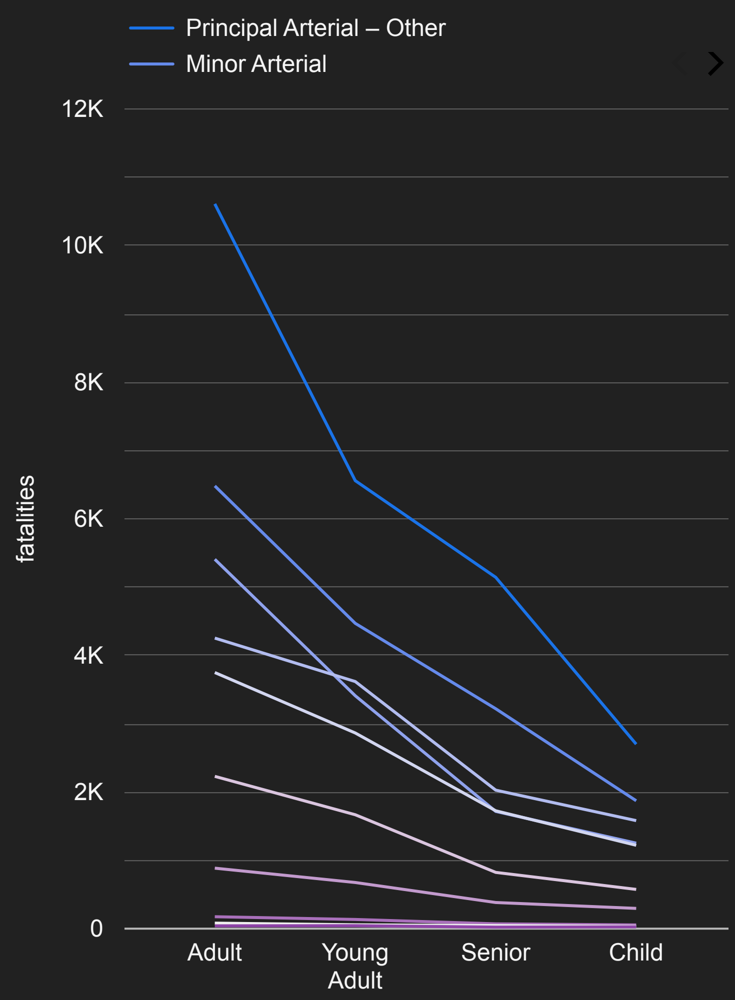


Are young or old people more at risk on certain roads?

This section links road classification to age group risk.

Young adults and adults are the most affected across all road types.





Principal arterials and interstates see the most fatalities.

High-speed roads are inherently riskier, and education campaigns targeting younger drivers could have a significant safety impact.

What visibility issues lead to higher fatalities?

This section looks at how driver vision impairments—like glare, weather, or road design—correlate with fatal outcomes.

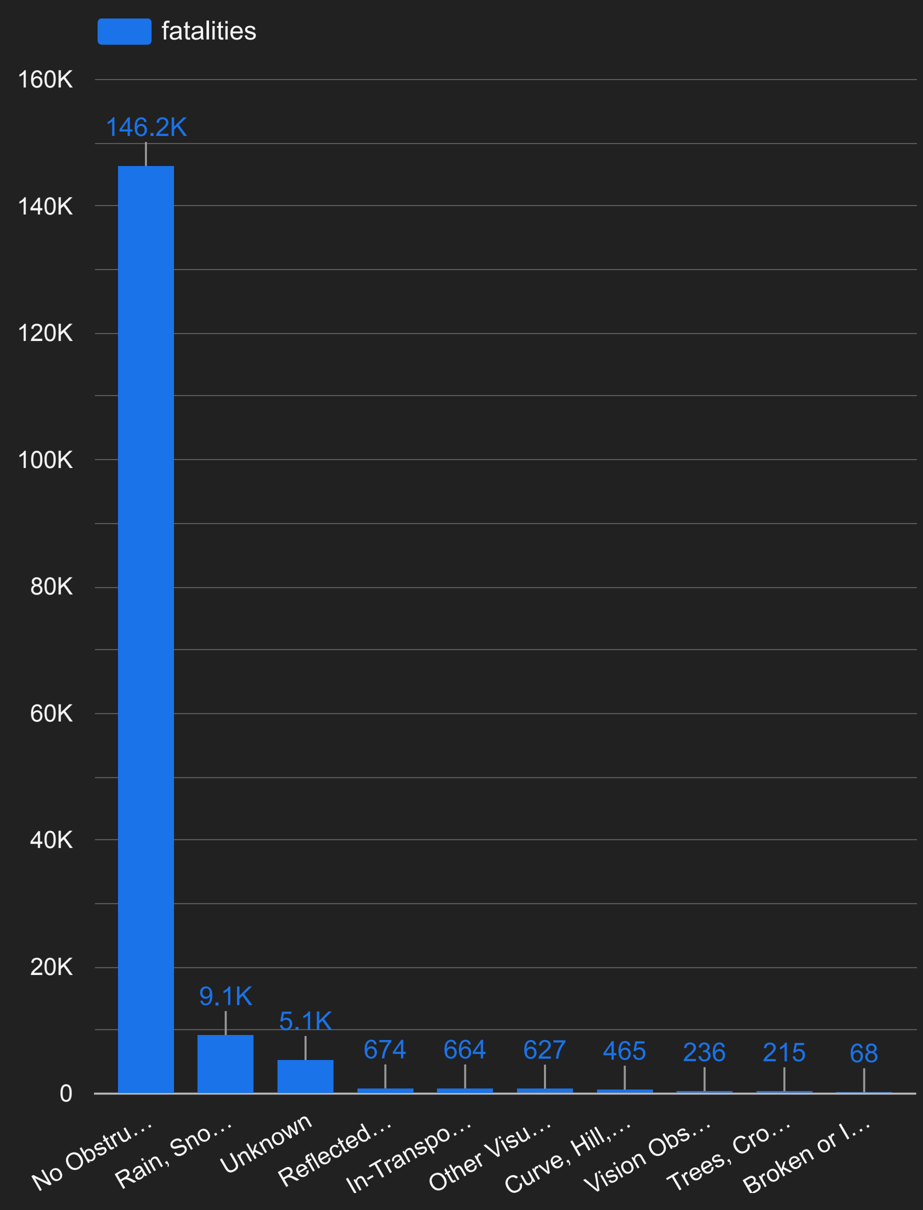
While most crashes report “No Obstruction,” over 9,000 fatalities occurred in conditions like fog, rain, snow, and dust.

Sun glare and poorly cleaned windshields also contributed.

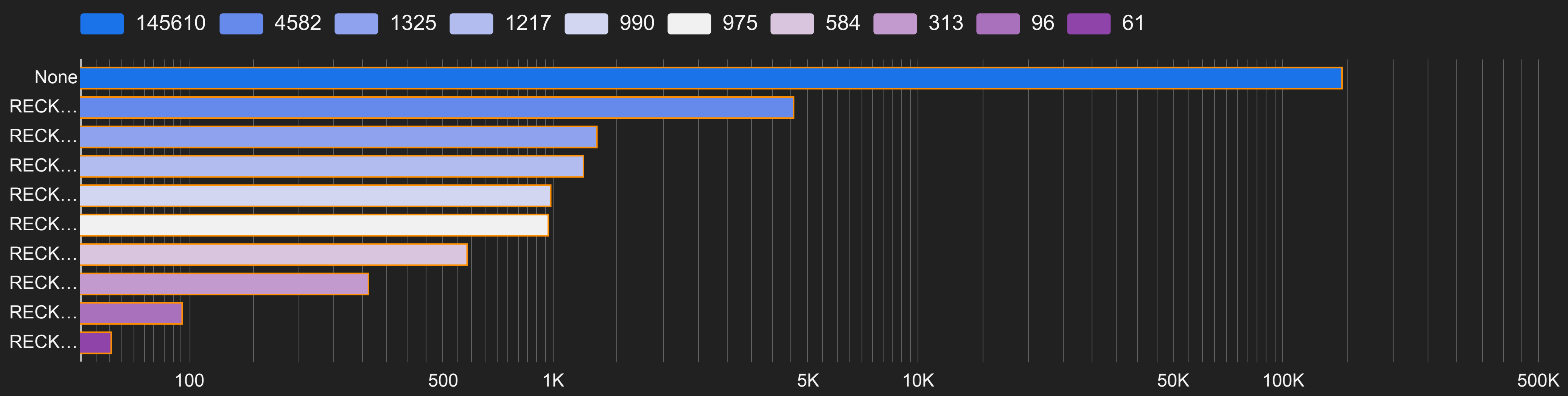
Enhancing vehicle visibility systems and driver training can mitigate risk under poor visual conditions.

	drivers_vision_obscured_by_name	crashes
1.	Broken or Improperly Cleaned Windshield	68
2.	Trees, Crops, Vegetation	215
3.	Rain, Snow, Fog, Smoke, Sand, Dust	9083
4.	Other Visual Obstruction	627
5.	Unknown	5146
6.	No Obstruction Noted	146207
7.	Reflected Glare, Bright Sunlight, Headlights	674
8.	Vision Obscured – No Details	236
9.	Curve, Hill, or Other Roadway Design Features	465
10.	In-Transport Motor Vehicle (Including Load)	664

1 - 10 / 10



Which violations are most associated with fatal crashes?



We explore the link between traffic violations (e.g., reckless driving, hit-and-run) and fatal crashes, identifying high-risk behaviors.
Most fatalities had "No Violation Reported," but thousands occurred under reckless driving, hit-and-run, fleeing police, or manslaughter charges. A significant portion of fatal crashes involve criminal or high-risk behavior, which supports stricter traffic law enforcement.

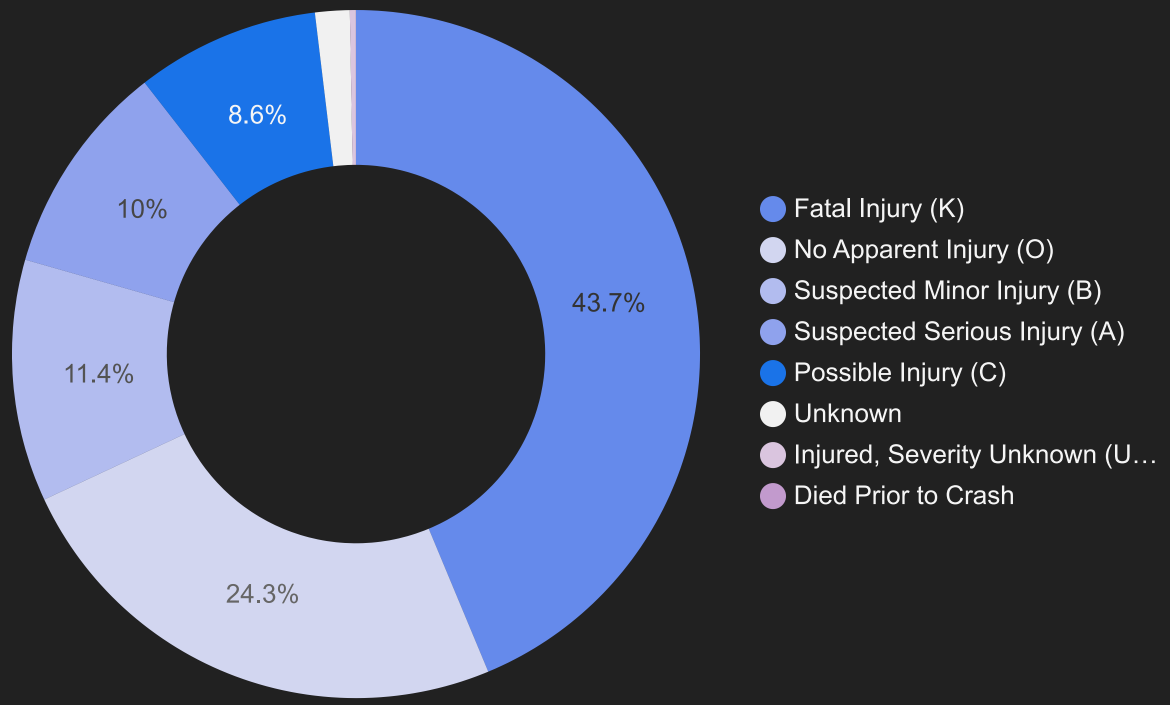
	violations_charged_name	fatalities
1.	RECKLESS/CARELESS/HIT-AND-RUN OFFENSES, Fleeing or Eluding Police	96
2.	RECKLESS/CARELESS/HIT-AND-RUN OFFENSES, Fail to Give Aid, Information, Wait for Police after Crash	313
3.	RECKLESS/CARELESS/HIT-AND-RUN OFFENSES, Serious Violation Resulting in Death	584
4.	RECKLESS/CARELESS/HIT-AND-RUN OFFENSES, Hit-and-Run, Fail to Stop After Crash	975
5.	RECKLESS/CARELESS/HIT-AND-RUN OFFENSES, Manslaughter or Homicide	4,582
6.	None	145,610
7.	RECKLESS/CARELESS/HIT-AND-RUN OFFENSES, Unsafe Reckless (Not Willful, Wanton Reckless) Driving	990
8.	RECKLESS/CARELESS/HIT-AND-RUN OFFENSES, Fail to Obey Police, Fireman, Authorized Person Directing Traffic	61
9.	RECKLESS/CARELESS/HIT-AND-RUN OFFENSES, Inattentive, Careless, Improper Driving	1,217

1 - 10 / 10

What's the injury distribution across crashes?

	injury_severity_name	person…
1.	Fatal Injury (K)	37806
2.	No Apparent Injury (O)	21040
3.	Suspected Minor Injury (B)	9835
4.	Suspected Serious Injury (A)	8686
5.	Possible Injury (C)	7456
6.	Unknown	1403
7.	Injured, Severity Unknown (U) (Since 1978)	245
8.	Died Prior to Crash	3

1 - 8 / 8



This page shows the severity of injuries across incidents to help quantify the burden beyond just fatalities.

Fatal injuries account for over 43% of recorded outcomes. Only 24% involved no apparent injury, and others were serious or minor.

This highlights the lethal nature of these incidents and emphasizes why proactive prevention is essential.

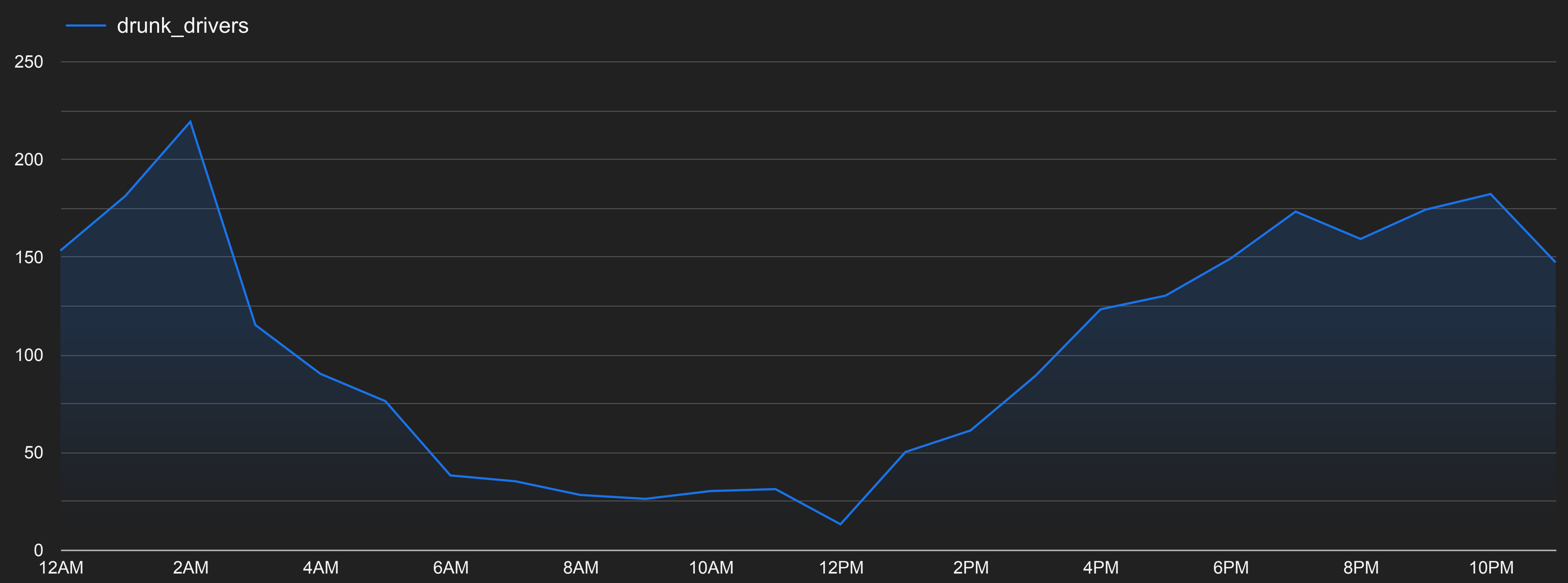
Drunk Drivers vs Time

Can we predict severe crashes?

Using logistic regression in BigQuery ML, we built a model to predict whether a crash will result in multiple fatalities based on crash characteristics.

The model uses features like drunk driving, day/time, vehicle type, and location. Evaluation results show promising accuracy, demonstrating the power of machine learning in anticipating high-risk crashes.

This model can support early warning systems or automated crash triage tools in traffic safety infrastructure.



Conclusion

Alcohol and time of day are the strongest predictors of fatal crashes — especially weekends and late nights.

Proper use of seatbelts, helmets, and airbags significantly reduces fatalities.

Certain vehicle types (e.g., trucks, buses) and road types (e.g., interstates) carry higher risk — especially for elderly drivers.

Vision obstructions like glare, rain, or road curves sharply increase crash risk.

Future Work

Real-Time & Geographic Risk Scoring:

Build a live crash risk dashboard for cities/states Integrate with real-time feeds (e.g. Waze, DOT sensors)
Score roads or intersections for proactive enforcement and emergency readiness