Analysis of Traffic Fatality Records

Authors

- **Justin Rebholz** ✉
  · ⭘ jrebholz12
  Department of Civil and Environmental Engineering, University of Illinois

- **Cameron Kimber** ✉
  · ⭘ ckimber2
  Department of Civil and Environmental Engineering, University of Illinois

- **Hannah Daggett** ✉
  · ⭘ hed2
  Department of Civil and Environmental Engineering, University of Illinois

- **Riley Kelch** ✉
  · ⭘ rileykelch
  Department of Civil and Environmental Engineering, University of Illinois

✉ — Correspondence possible via GitHub Issues or email to Justin Rebholz <rebholz4@illinois.edu>, Cameron Kimber <ckimber2@illinois.edu>, Hannah Daggett <hed2@illinois.edu>, Riley Kelch <rjkelch2@illinois.edu>.

Abstract

Description

The dataset that will be used for this project is the Fatality Analysis Reporting System created by the National Highway Safety Administration. The data will be obtained from the NHTSA's FARS database, which is publicly accessible.The FARS dataset is available in the CSV format. The specific dataset that our project will be focused on is labeled "accidents" and includes 32K+ instances and 52 columns. The columns descriptions are described in the Fatality Analysis Reporting System (FARS) Analytical User's Manual 1975-2015. Columns are described in the below table:

Column	Description
ARR_HOUR	This data element records the hour when emergency services arrived at the scene.
ARR_MIN	This data element records the minute when emergency services arrived at the scene.
CF1, CF2, CF3	These data elements record contributing factors to the crash, such as driver behaviors or environmental conditions.
CITY	This data element identifies the city in which the crash occurred.
COUNTY	This data element identifies the county in which the crash occurred.
DAY	This data element records the day of the month on which the crash occurred.
DAY_WEEK	This data element identifies the day of the week on which the crash occurred.
DRUNK_DR	This data element records whether a driver involved in the crash was suspected of drinking alcohol.
FATALS	This data element records the number of fatalities resulting from the crash.
FUNC_SYS	This data element identifies the functional classification of the trafficway segment where the crash occurred.
HARM_EV	This data element records the first harmful event that occurred in the crash sequence.
HOSP_HR	This data element records the hour when the injured were admitted to the hospital.
HOSP_MN	This data element records the minute when the injured were admitted to the hospital.
HOUR	This data element records the hour when the crash occurred.
LATITUDE	This data element identifies the location of the crash using latitude coordinates.
LGT_COND	This data element identifies the light condition at the time of the crash, such as daylight, dark, or dusk.

Column	Description
LONGITUD	This data element identifies the location of the crash using longitude coordinates.
MAN_COLL	This data element identifies the manner of collision, such as rear-end, head-on, or angle.
MILEPT	This data element records the milepoint nearest to the crash location.
MINUTE	This data element records the minute when the crash occurred.
MONTH	This data element records the month in which the crash occurred.
NHS	This data element identifies whether the crash occurred on a National Highway System (NHS) route.
NOT_HOUR	This data element records the hour when the crash was reported to authorities.
NOT_MIN	This data element records the minute when the crash was reported to authorities.
PEDS	This data element records the number of pedestrians involved in the crash.
PERMVIT	This data element counts the number of persons in motor vehicles in transport (motorists) involved in the crash.
PERNOTMVIT	This data element counts the number of persons not in motor vehicles in transport (non-motorists) involved in the crash.
PERSONS	This data element is a count of the total number of persons involved in the crash.
PVH_INVL	This data element is the number of parked or working vehicles involved in the crash.
RAIL	This data element identifies if the crash involved a rail system or crossing.
RELJCT1	This data element identifies the relationship of the crash to a junction, such as intersection or non-intersection.
RELJCT2	This data element provides additional information about the crash's relationship to the junction.
REL_ROAD	This data element identifies the relationship of the crash to the road, such as on the roadway or off the roadway.
RD_OWNER	This data element identifies the entity responsible for the ownership of the road where the crash occurred.
ROUTE	This data element records the type of route where the crash occurred, such as Interstate, U.S. Highway, or State Highway.
RUR_URB	This data element identifies whether the crash occurred in a rural or urban area.

Column	Description
SCH_BUS	This data element identifies if a school bus was involved in the crash.
SP_JUR	This data element identifies if the crash occurred in a special jurisdiction, such as military or Indian reservations.
STATE	This data element identifies the state in which the crash occurred. The codes are from the General Services Administration's (GSA) publication of worldwide Geographic Location Codes (GLC).
ST_CASE	This data element is the unique case number assigned to each crash. It appears on each data file and is used to merge information from the data files together.
TWAY_ID	This data element identifies the primary trafficway on which the crash occurred.
TWAY_ID2	This data element identifies the secondary trafficway associated with the crash, if applicable.
TYP_INT	This data element identifies the type of intersection involved in the crash, if applicable.
VE_FORMS	This data element is a count of all vehicle forms applicable to this crash.
VE_TOTAL	This data element is the number of contact motor vehicles that the officer reported on the PAR as a unit involved in the crash.
WEATHER	This data element identifies additional weather factors at the time of the crash.
WEATHER1	This data element records the primary weather condition at the time of the crash.
WEATHER2	This data element records the secondary weather condition at the time of the crash.
WRK_ZONE	This data element identifies if the crash occurred in a work zone.
YEAR	This data element records the year in which the crash occurred.

Link: https://www.kaggle.com/datasets/nhtsa/2015-traffic-fatalities

Plan and Proposal

Using the FARS dataset we aim to understand the trends in traffic fatalities in a given year and what factors are affecting those trends. We will also look at how the different variables play a role in the severity of the accident and identify geographic regions that are more prone to accidents. The trends in traffic fatalities found through this project can be used to inform policy makers and ultimately decrease the number of traffic fatalities.

Analysis

Exploratory Data Analysis

Predictive Modeling

References

National Highway Traffic Safety Administration. "2015 Traffic Fatalities." Kaggle, https://www.kaggle.com/datasets/nhtsa/2015-traffic-fatalities. Accessed 24 Oct. 2024.

National Highway Traffic Safety Administration. Fatality Analysis Reporting System (FARS) Analytical User's Manual 1975-2015. U.S. Department of Transportation, Aug. 2016.