

Analysis of Traffic Fatality Records

This manuscript ([permalink](#)) was automatically generated from [uicedes/project-team-front-row@26dff1d](#) on October 27, 2024.

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 subset data that our project will be focused on is labeled "accidents.csv" 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 and in the below table:

Table 1: Abbreviation Legend.

| 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. |

| Column | Description |
|------------|---|
| LGT_COND | This data element identifies the light condition at the time of the crash, such as daylight, dark, or dusk. |
| 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. |

| Column | Description |
|----------|--|
| RUR_URB | This data element identifies whether the crash occurred in a rural or urban area. |
| 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. Specifically, we will look to implement safety factors that guard against drunk driving (traffic cameras, sensor systems, DUI checkpoints, etc.) 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

The dataset that we have chosen describes the details surrounding motor vehicle crashes that have resulted in at least one fatality in the United States during 2015.

From the data, we have interpreted that 32,166 fatal crashes occurred in 2015. Out of the total number of crashes, 26.78 percent of them involved an intoxicated driver. In the state of Illinois, 264 of 914 crashes involved a drunk driver (28.9%). More statistics are found in the table below.

Table 2: Misc Statistics.

| Desc (2015) | Statistic |
|------------------------------|----------------------------|
| Crashes (US) | 32,166 |
| Total Fatalities (US) | 35,092 |
| Drunk Driver Crashes (US) | 8,617 |
| Drunk Driver Percentage (US) | 26.78% |
| Crashes (IL) | 914 |
| Total Fatalities (IL) | 998 |
| Drunk Driver Crashes (IL) | 264 |
| Drunk Driver Percentage (IL) | 28.88% |
| Most DD Crashes, Dates (US) | 03May, 15Aug, 02Aug, 16Aug |
| Most Crashes, Dates (IL) | 07Mar, 27Jun, 17Apr |

We first looked at a map of the United states to plot the fatal accidents vs the drunk driving fatal accidents. This visualltion shows hotspots for both categories.

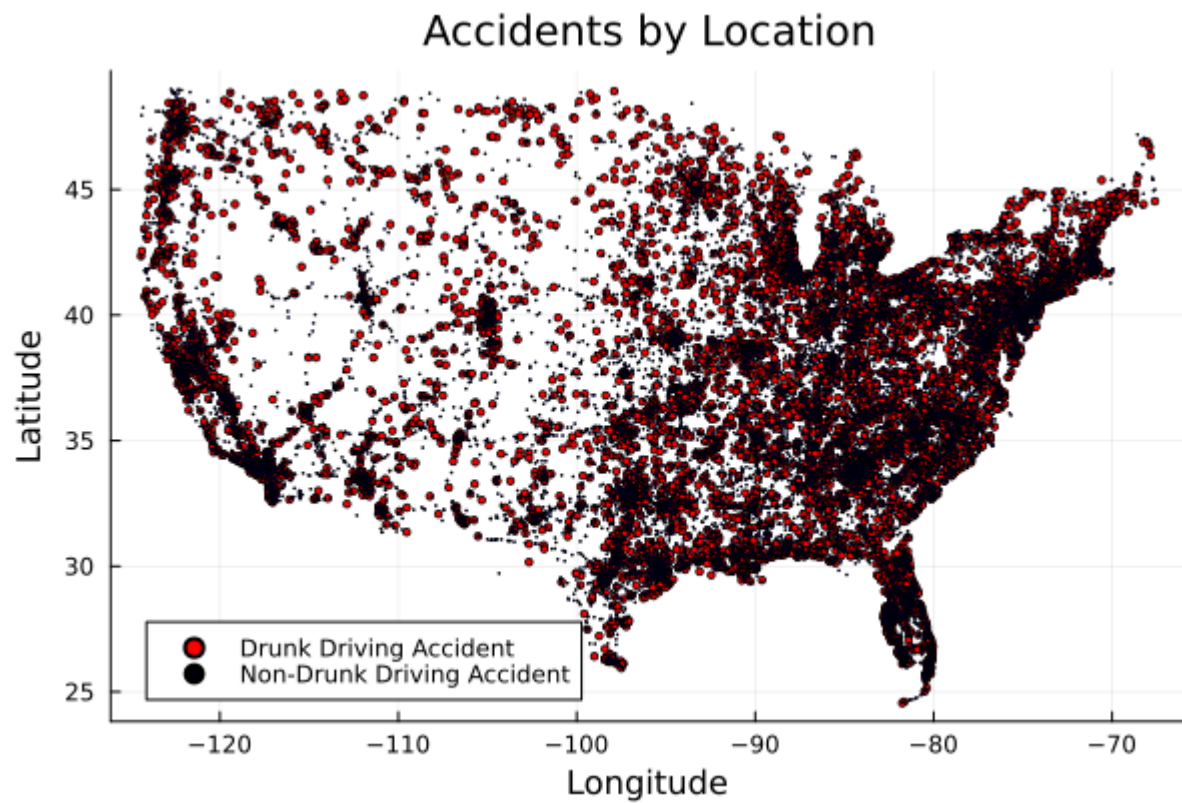


Figure 1: US Map

We then created a scatter plot to visualize the number of accidents and fatalities per accident for the United States. As seen in the figure, most of the crashes result in only one fatality, however, there are a handful of multiple fatality crashes. The scatter plot annotates the amount of crashes for drunk drivers and sober drivers next to the datapoint.

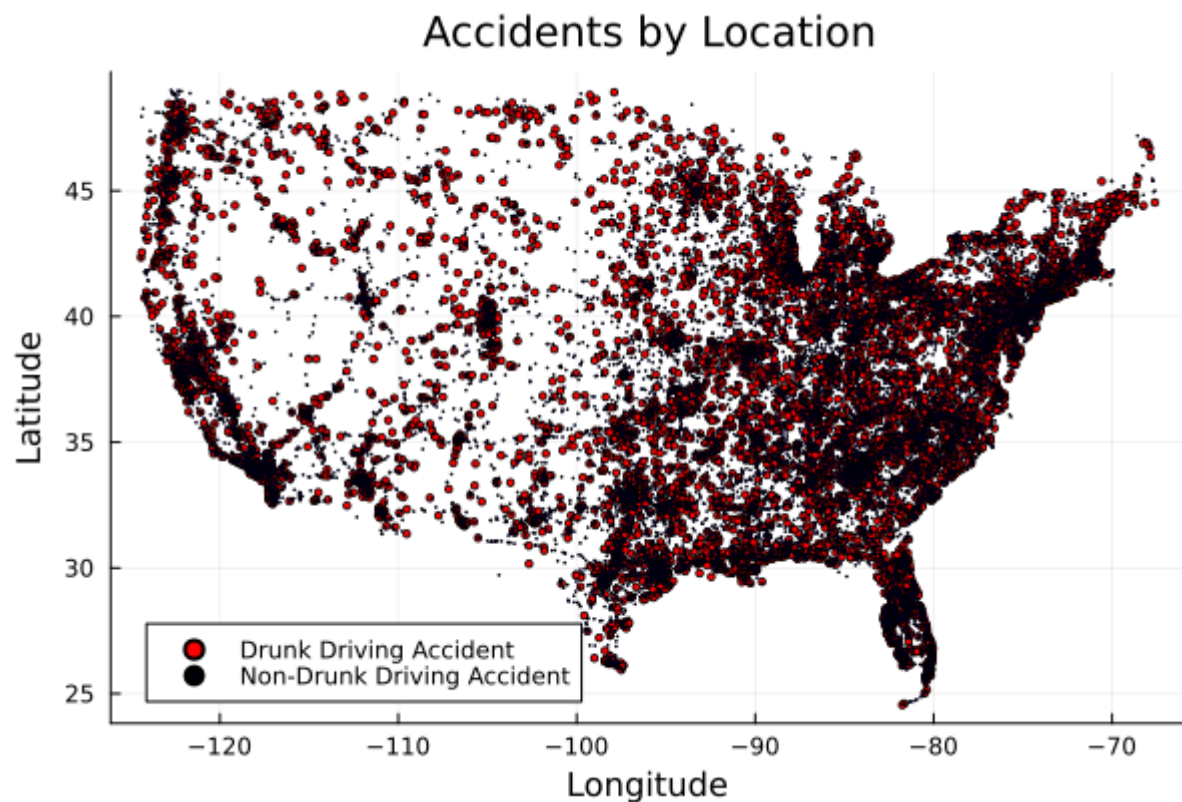


Figure 2: Scatter Plot

The next factor that we analyzed to understand the data was the time of accidents. As seen in the figure below, the highest number of accidents occurred on the weekends. On average, 104 accidents

occurred on a given day of a weekend whereas 82 accidents occurred per weekday. Monday and Tuesday have the lowest number of accidents and as the week progresses, the number of accidents increases.

The portion of accidents due to drunk driving by the day of the week follows similar trends. The average ratio of drunk driving accidents to total number of accidents was 37.47% for the weekend and 21.19% for weekdays. This is most likely due to the fact that drinking is more popular on the weekends. Similarly, there is a figure for crashes vs time of day.

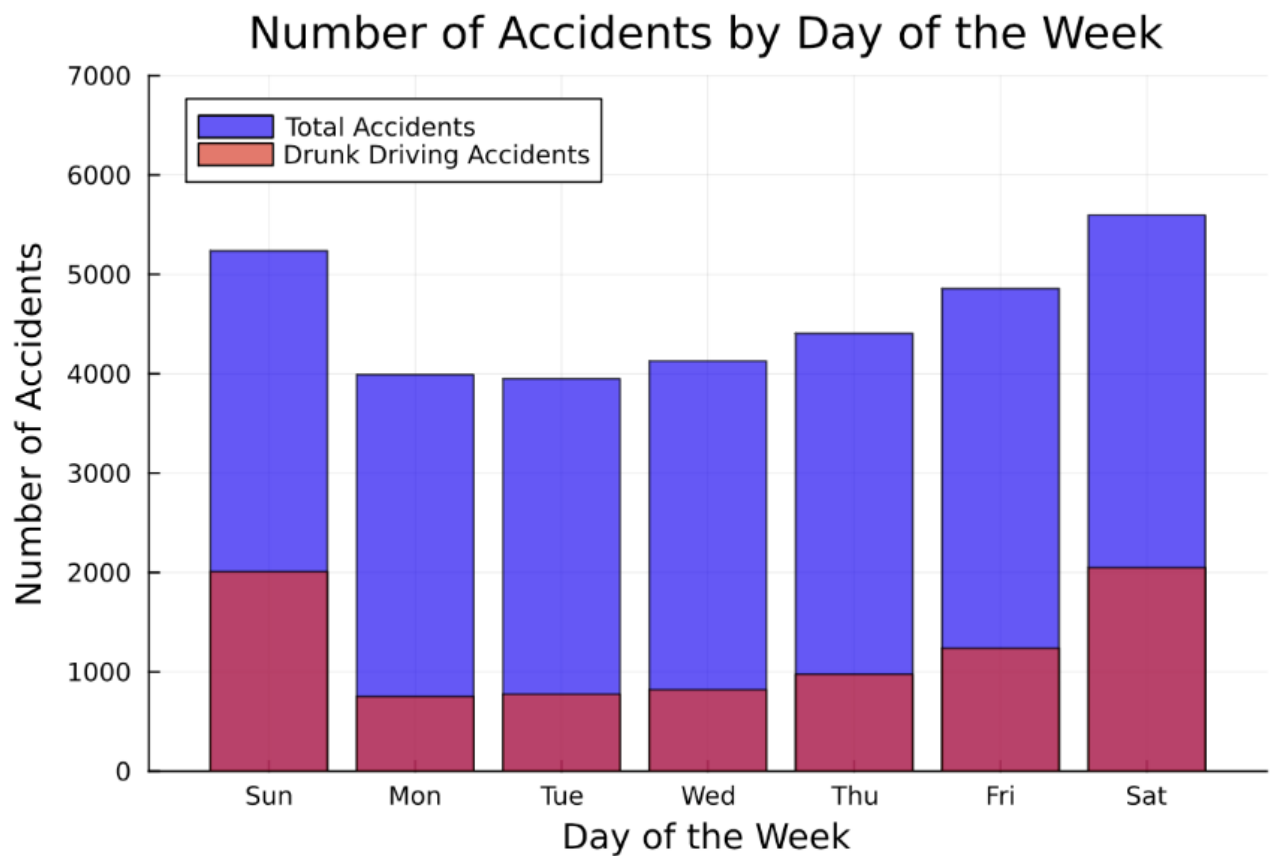


Figure 3: Accidents vs Day of Week

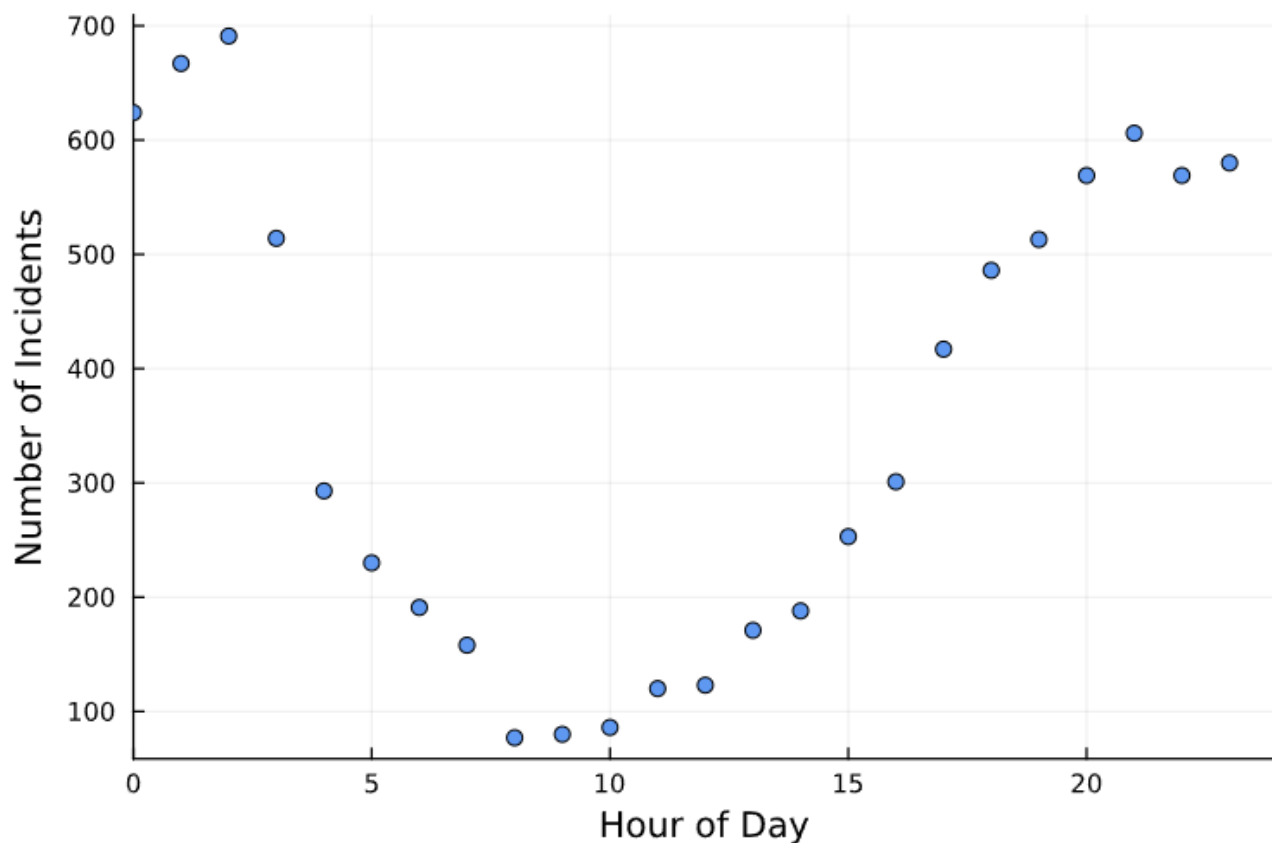


Figure 4: Time of Day

Our team also looked at the number of drunk driver accidents per day of the year. See Table 1 for the most popular days for drunk driver accidents.

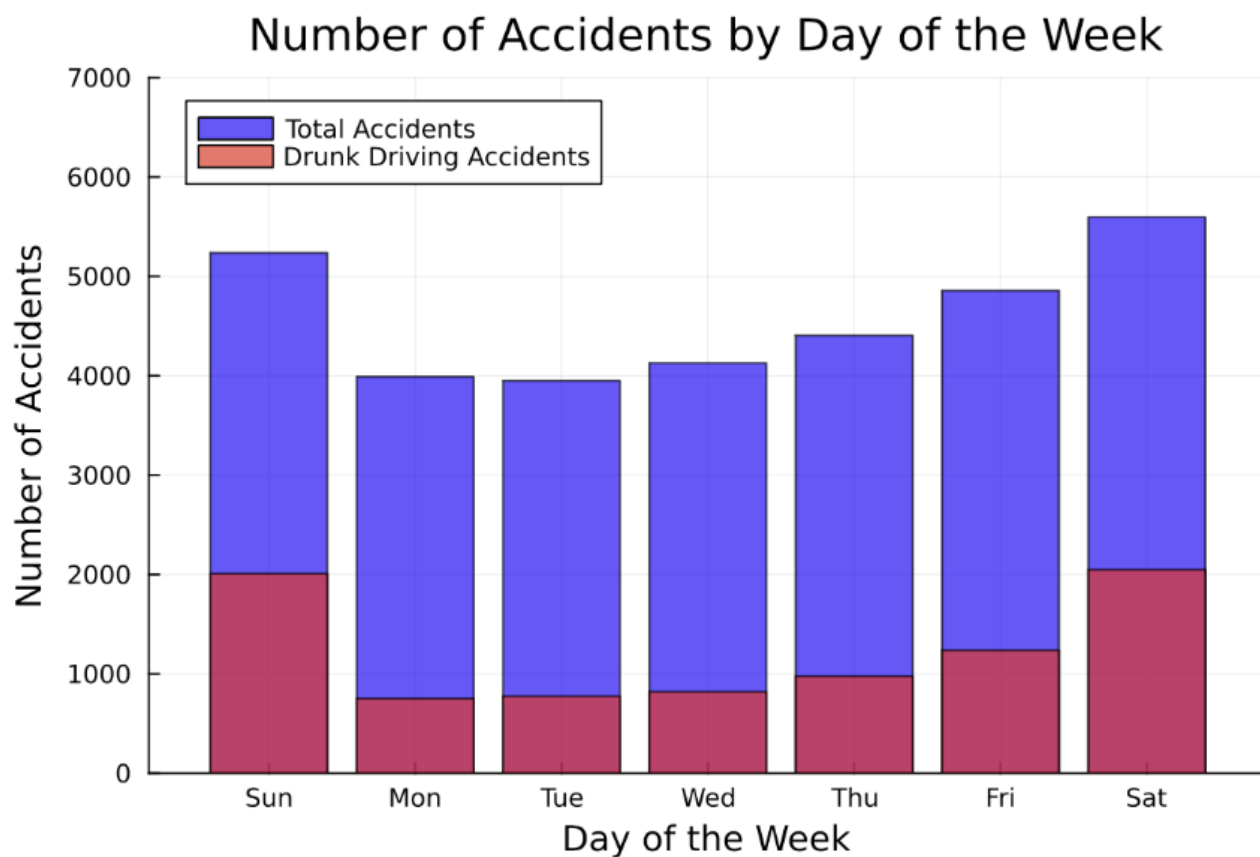


Figure 5: DUI Crashes by Day (US)

As we conducted our Exploratory Data Analysis, we aimed to focus in on the state of Illinois. Below is the same information as Figure 1, but specific to Illinois for 2015. From this visual, we can see that most crashes are in the areas with larger cities (i.e. Chicago).

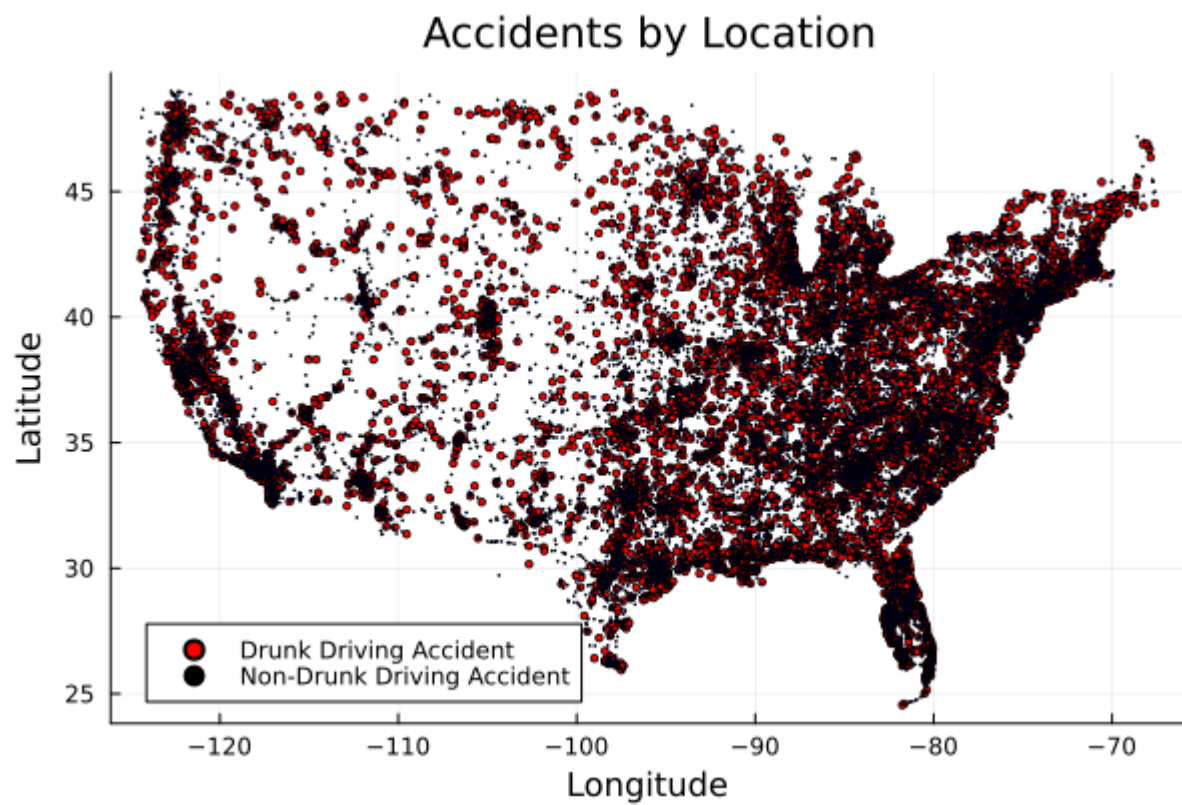


Figure 6: Illinois Map

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.