Fatal Traffic Accident Analysis

Project 1

Ernesto Beltran, Jessica Nugent, Ivy Tiongco, Mike Whitmore

UCSD Extension Data Science Bootcamp 3/24/2020

Ernesto

Purpose

- Uncover patterns in fatal traffic accidents around San Diego county from 2010-2020
  - Accidents involving a death during or after the accident
- Personal motivation
  - Ernesto has been in 2 accidents as a motorcyclist

Questions

- As we explored the data, we were most interested in asking:
  - Is there a relationship between fatal traffic accident frequency and the following factors?
    - Drunk driving
    - Day/time of the accident
    - Age
    - Number of survivors

Ernesto (next Mike)

Data Sources

- National Highway Traffic Safety Administration (NHTSA) Crash Viewer
  - Fatality Analysis Reporting System (FARS) API
- US Census API
- Google Maps API
- San Diego County Data Portal

Mike

Data Cleanup & Exploration

- Scrolled through every variable in 2010-2020 crash data to decide which ones to explore
- Combined data from 3 data sources into one dataframe

Functions and tasks done to prepare data for graphing:

- API calls
- JSON formatting
- Dataframes
- Functions
- Shapely
- Scipy stats
- Merge dataframes
- Read/write csv
- Groupby
- Binning
- Convert value types
- Excel

Mike

Difficulties, Limitations, & Bias

- Difficulties
  - Researched traffic dataset with historical traffic information, but a free dataset could not be found.
  - Gathering the data required many API calls
  - Duplicates in crash data initially
  - We initially thought we were analyzing accident cases, but we realized we were looking at people so we revised our interpretations
- FARS API dataset looks only at traffic accidents with at least one death, so cannot generalize about traffic accidents with no deaths.
- We focused on San Diego county, so cannot generalize about other cities.

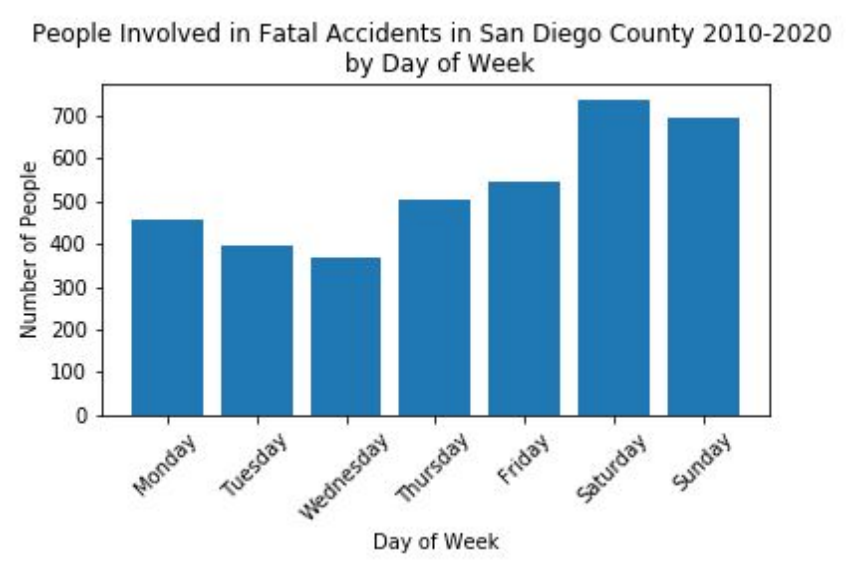
Mike (next Jessica)

Question 1

Is there a relationship between the total number of people in fatal accidents and drunk driving?

Jessica

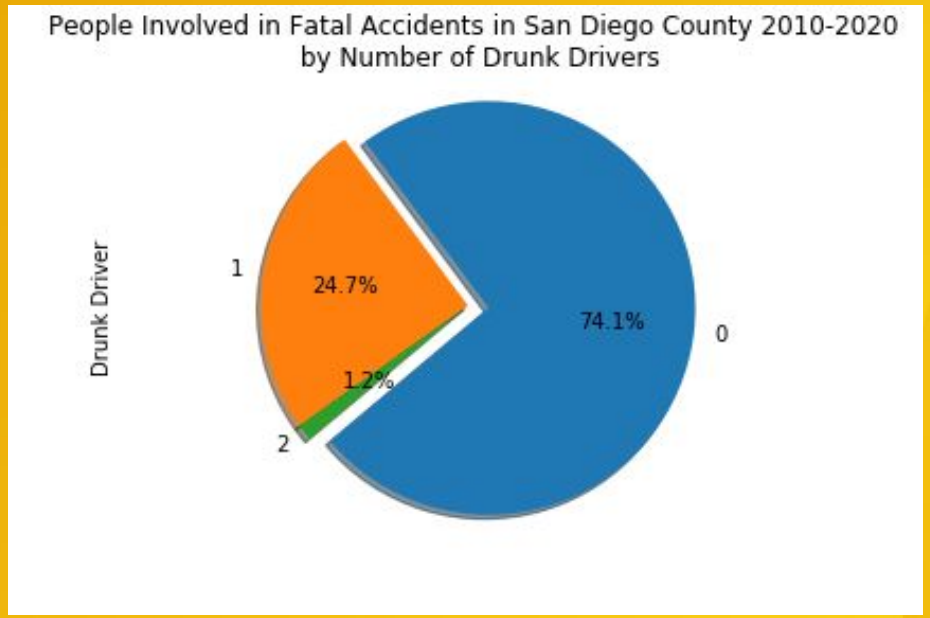
What are your odds of being in a fatal accident by day of the week?



Seeing higher numbers on the weekend, we suspected that accidents involving a death often involved drunk driving.

Jessica

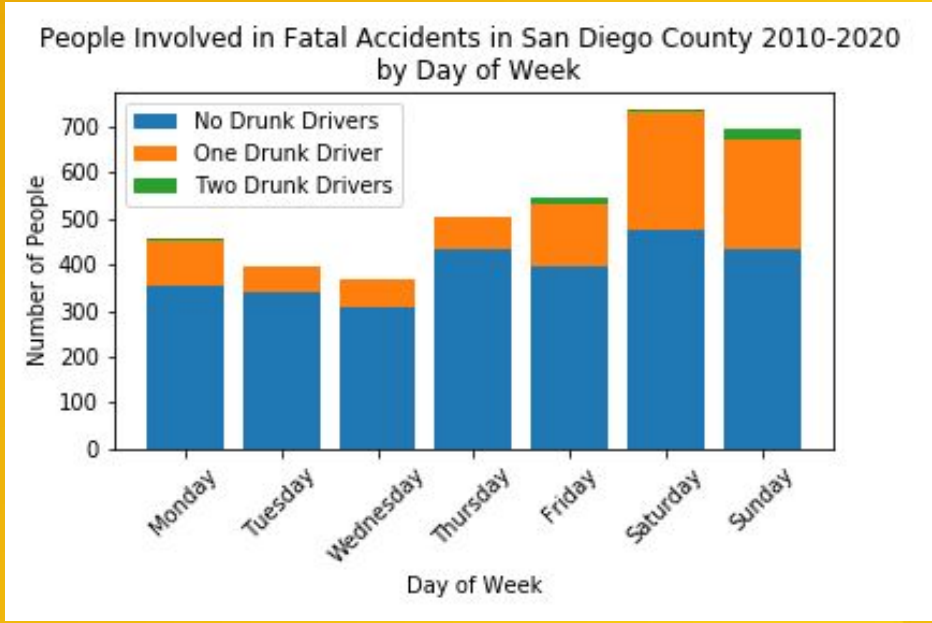
If you were involved in a fatal accident, odds are drinking was not involved.



Initially we did not focus on drunk driving, but data showed it was an important factor in these cases. It was somewhat surprising that majority of fatal accidents do not involve any drunk drivers.

Jessica

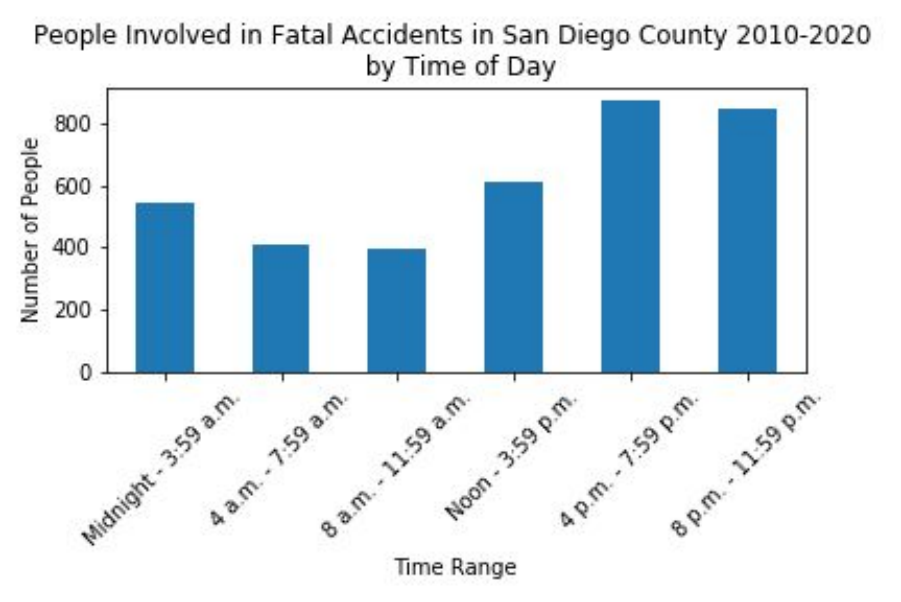
Are weekends more dangerous for drunk driving?



Your chance of being in a fatal drunk driving accident is higher on the weekend.

Jessica

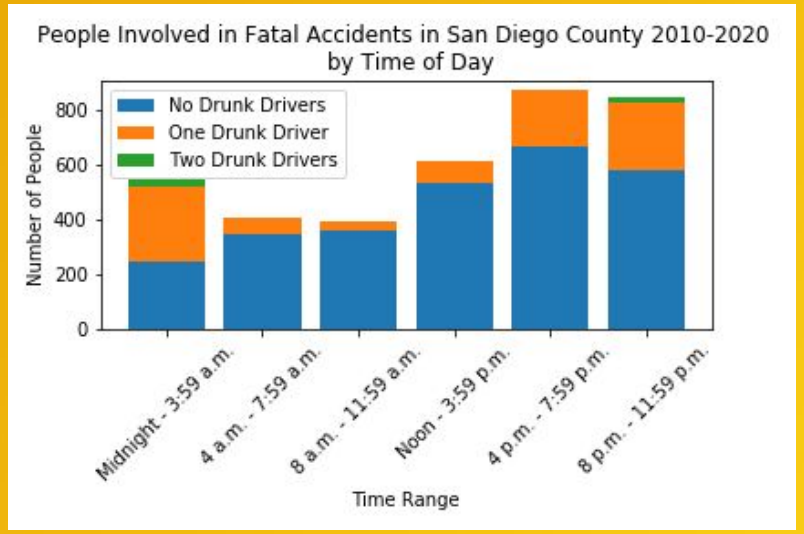
What are your odds of being in a fatal accidents by time of day?



Your odds of being in a fatal accident increase in the evenings and at night.

Jessica

Are nights and evenings more dangerous for drunk driving?



Your chances of being in a fatal accident involving drunk driving are higher in the evening and at night, which is in line with common times for drinking (dinners/parties).

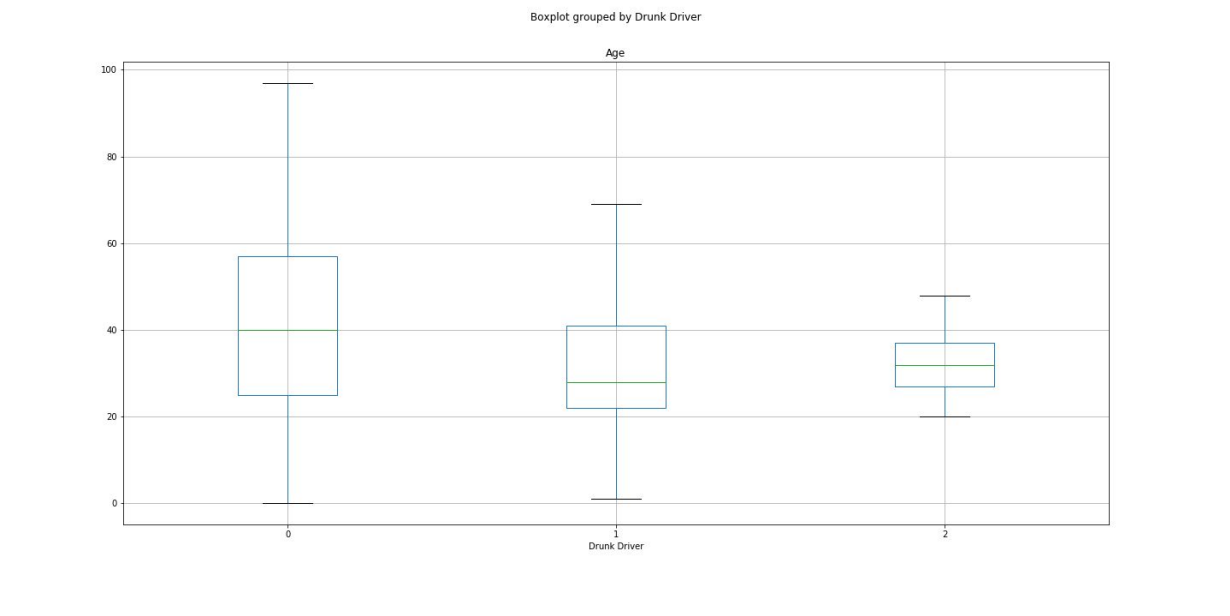
Jessica (next Ivy)

Question 2

Is there a relationship between age and drunk driving in fatal accidents?

Ivy

Boxplot - Age of people involved in fatal accidents by number of drunk drivers



Median of 0 drunk drivers group is higher than the groups with 1 or 2 drunk drivers.
Thus, the older you get, the less likely you are to be in an accident with a drunk driver.

Ivy

ANOVA - Age of people involved in fatal accidents by number of drunk drivers

```
from scipy import stats as st

# Extract individual groups
none = people_df_age[people_df_age["Drunk Driver"] == 0]["Age"]
one = people_df_age[people_df_age["Drunk Driver"] == 1]["Age"]
two = people_df_age[people_df_age["Drunk Driver"] == 2]["Age"]
```

```
# Perform the ANOVA
stats.f_oneway(none, one, two)
```

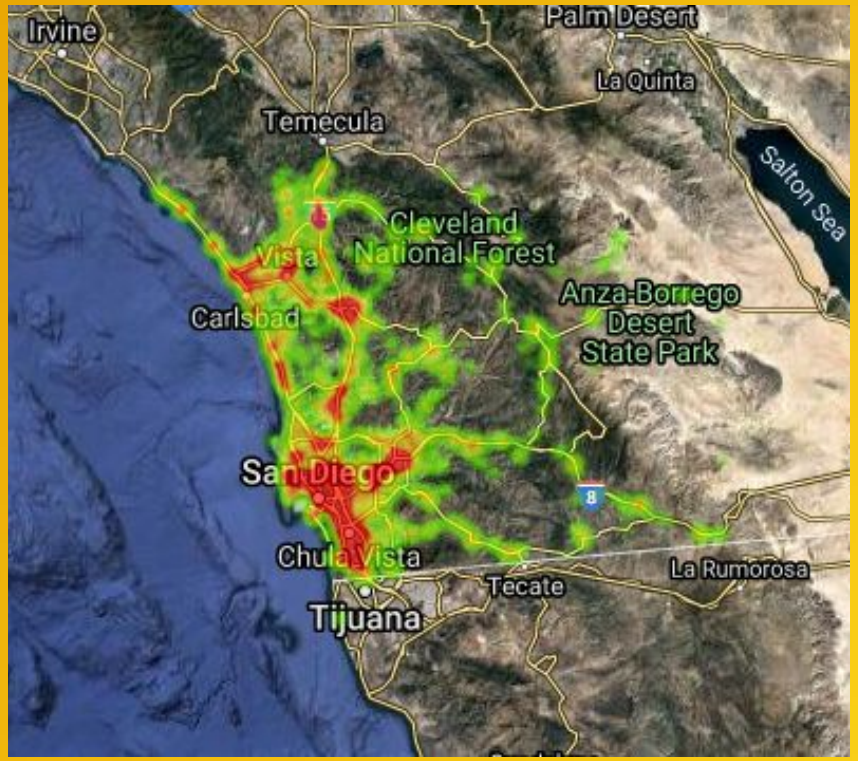
```
F_onewayResult(statistic=27.15407081179687, pvalue=3.4162595646894186e-12)
```

ANOVA analysis has very small p-value of 3.4e-12< 0.05 which shows that we can reject the null hypothesis that mean age is the same for people in fatal accidents with 0, 1, and 2 drunk drivers. We conclude that there is a statistically significant difference among the average age per group.

Question 3

Where in San Diego county are you more likely to be involved in a fatal accident?

Heatmap of number of people involved in fatal accidents



More people are involved fatal accidents in and around the city of San Diego compared to the rest of the county.

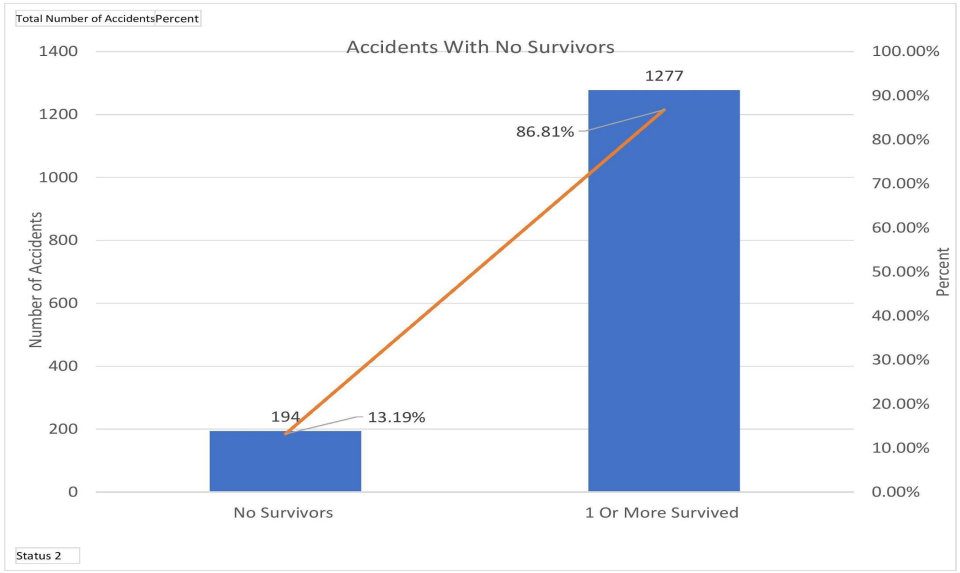
Ivy (next Ernesto)

Question 4

What is the number of accidents where all individuals die compared to accidents that have 1 or more survivor?

Ernesto

Accidents Resulting in No Survivors



13 percent of fatal accidents in San Diego result in no survivors.

Ernesto

Additional Questions

If we had funding and more time, we could analyze:

- Traffic pattern at time and location of accident
- Race
- Road type (freeway vs. street)
- Death circumstances (who died, when, where)
- Vehicle type (i.e. motorcycle vs. car)

Ernesto

Conclusions

- If you are involved in a fatal accident, chances are alcohol is not as likely to be involved as we expected.
- Alcohol-related accidents tend to occur more on the weekends and in the evening/night.
- The older you get, the less likely you are to be involved in a drunk driving accident.
- If you are involved in a fatal accident, you are more likely to be a survivor than not.

Ernesto

Questions?

Thanks for watching and listening!