



# Clare Drive

By the Road Prophet Team

James Opacich  
Natarajan Vijayaraman  
Jonathan Andrews  
Jaco Lau



# Problem Statement

- 3 Million people are injured in auto accidents.
- Many of these accidents happen in problem areas
- Help bring awareness and attention to high-risk areas.
- Reduce risk for family and community.
- A tool for everyone to make smarter choices around driving.

Clarevoyance for the road

# Process

1. Gather Data
2. Clean Data
3. Explore Data
4. Feature Creation and Selection
5. Creating Deployment and Visualization
6. Model Creation
7. Model Tuning
8. Model Deployment on Streamlit

Clarevoyance for the road

# U.S. Accident Data

- **For Modeling**
  - Over 4 Million Samples
  - Over 4GB of Data
  - Inspired Real World Use-Case
  - Kaggle
- **For EDA and Plotting**
  - U.S. Census Data
  - Scraped from Wikipedia

# Accident Severity Rating Scale (2-4)

Severity 2



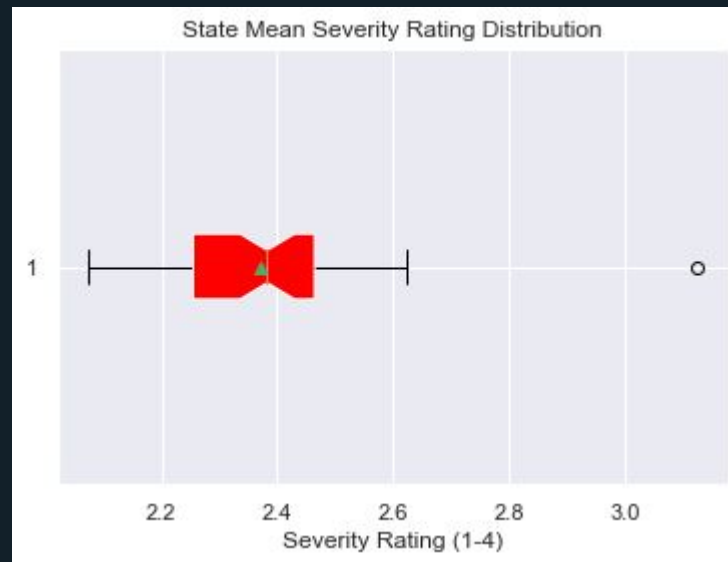
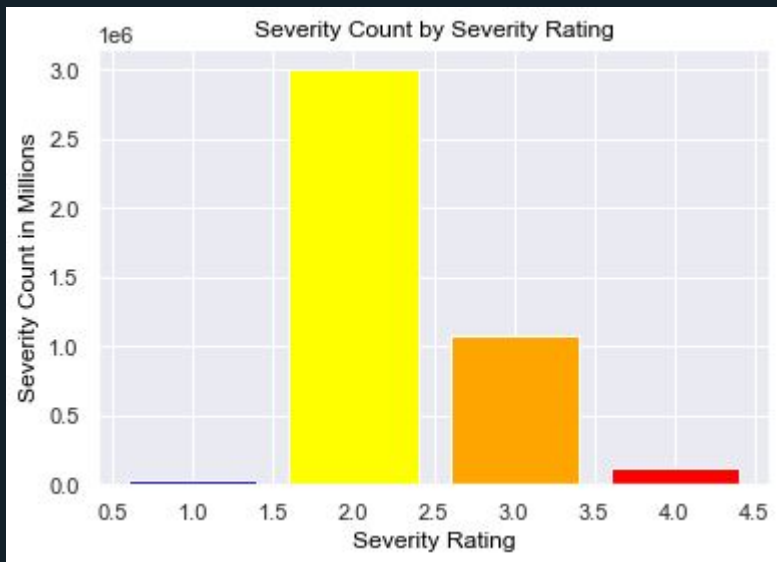
Severity 3



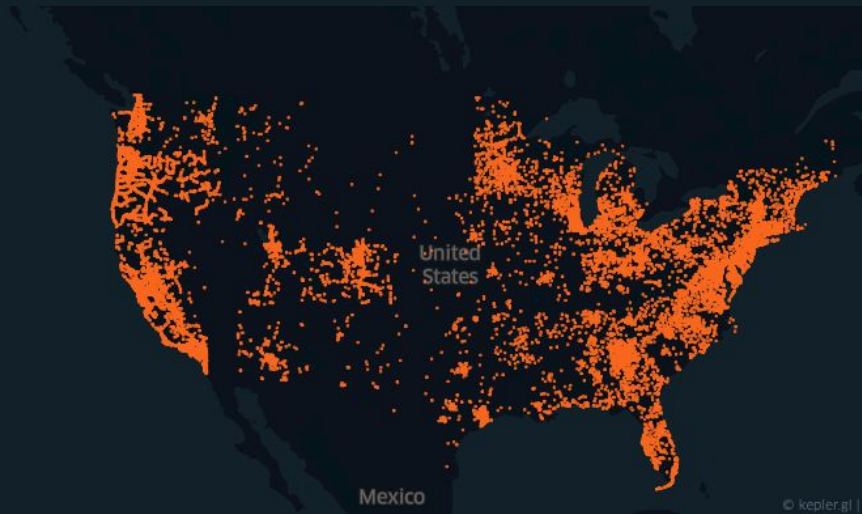
Severity 4



# Severity Rating



# Accidents on Left Side vs Right Side of Road



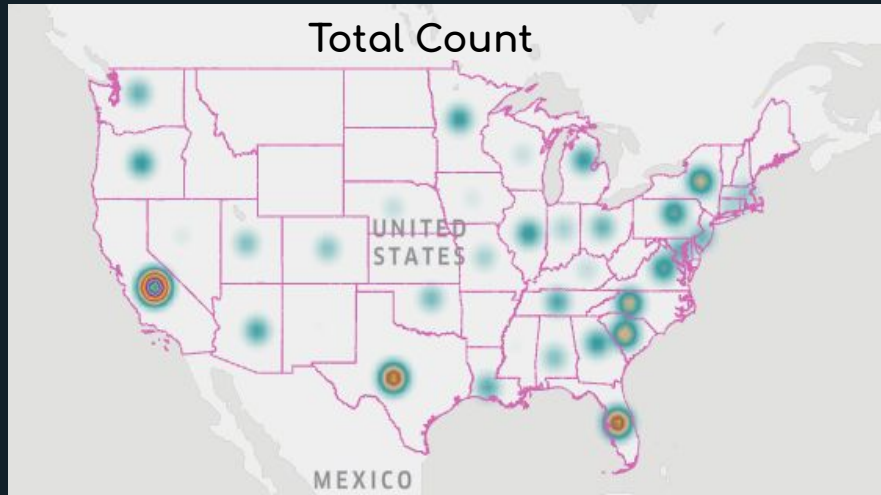
LEFT SIDE



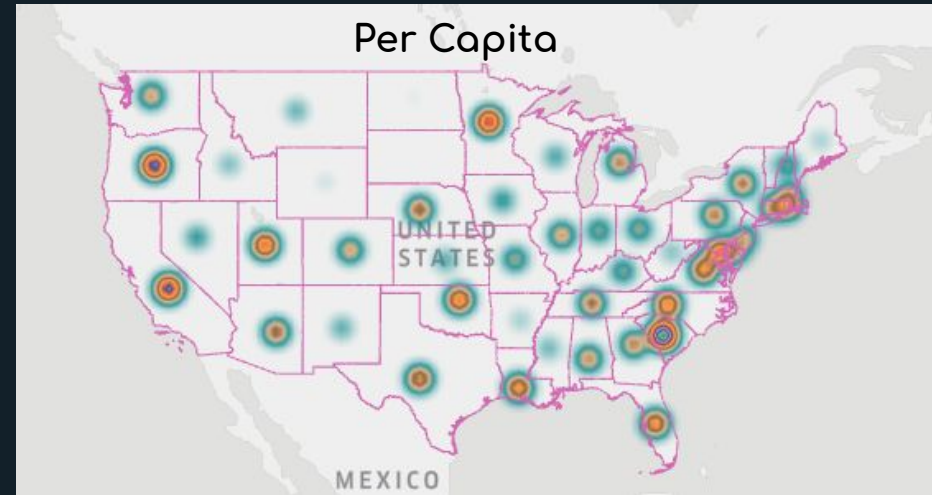
RIGHT SIDE

Where do you see a pattern?

# ACCIDENTS PER STATE



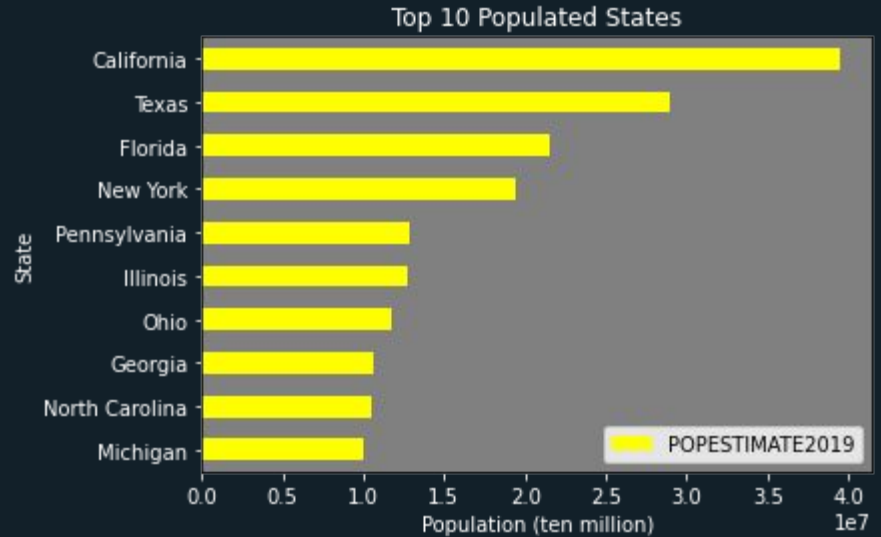
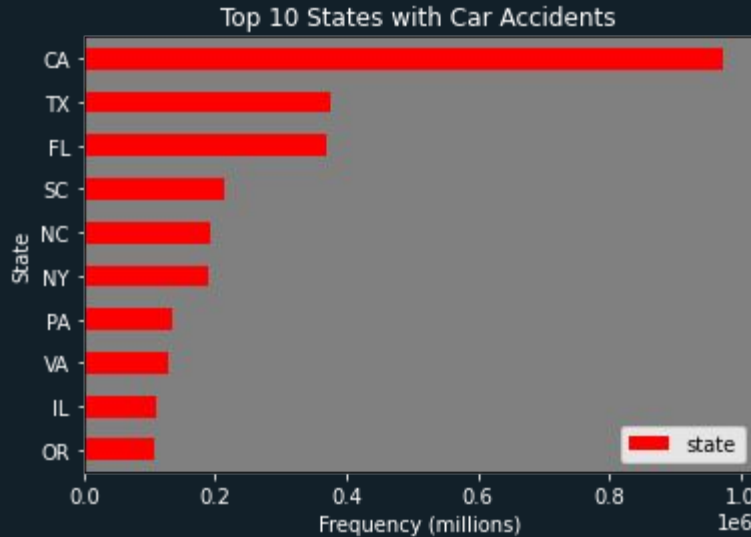
1. California
2. Texas
3. Florida



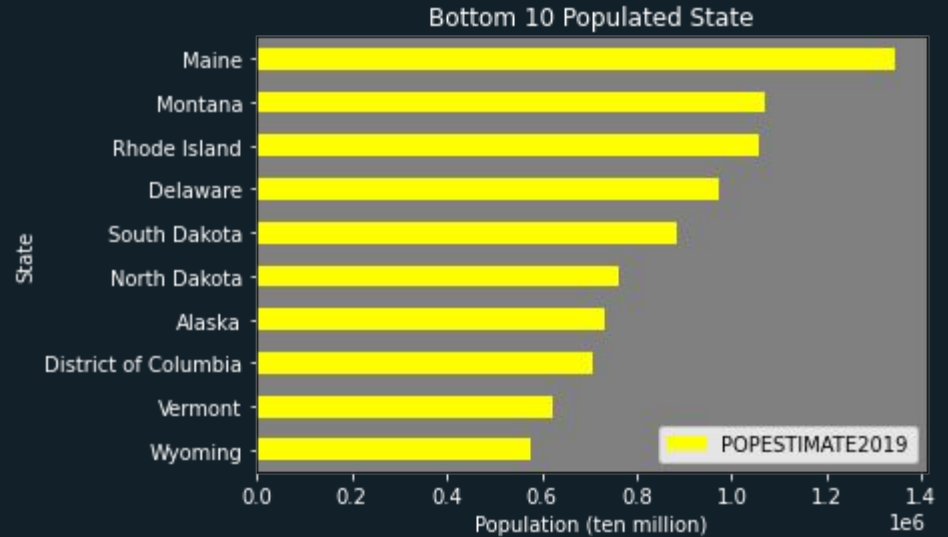
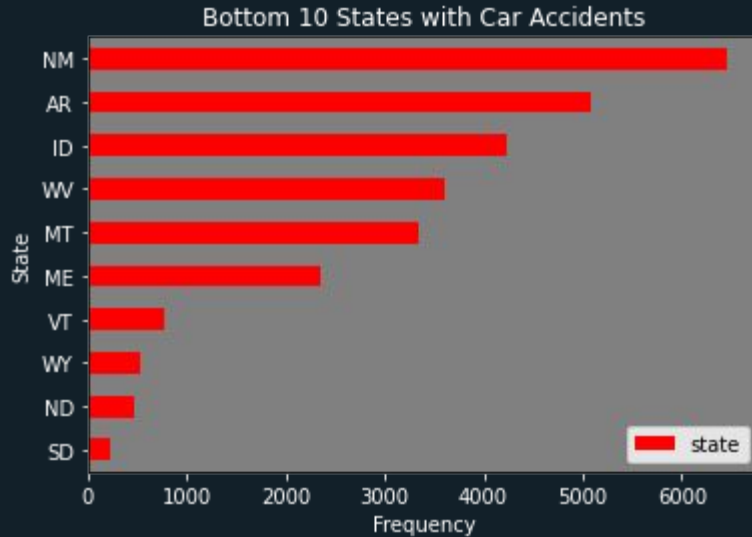
1. South Carolina
2. Oregon
3. California



# State Car Accidents vs. Population

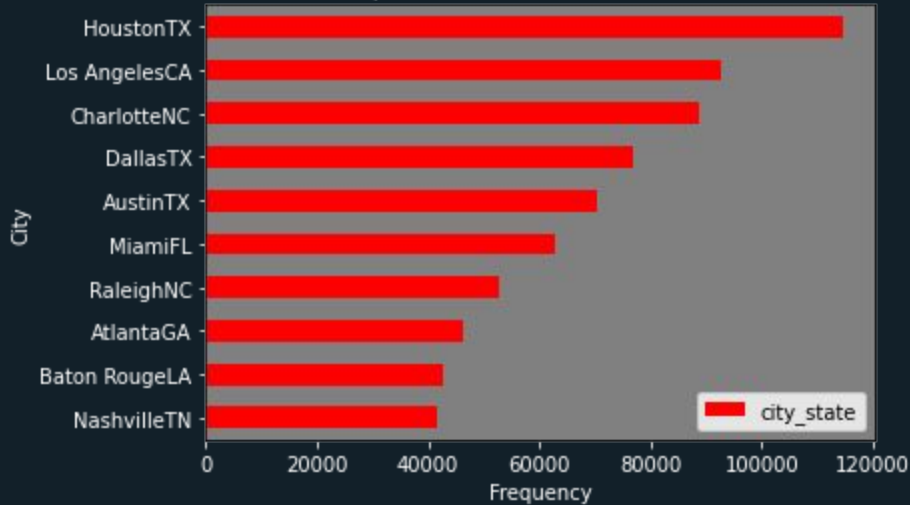


# State Car Accidents vs. Population

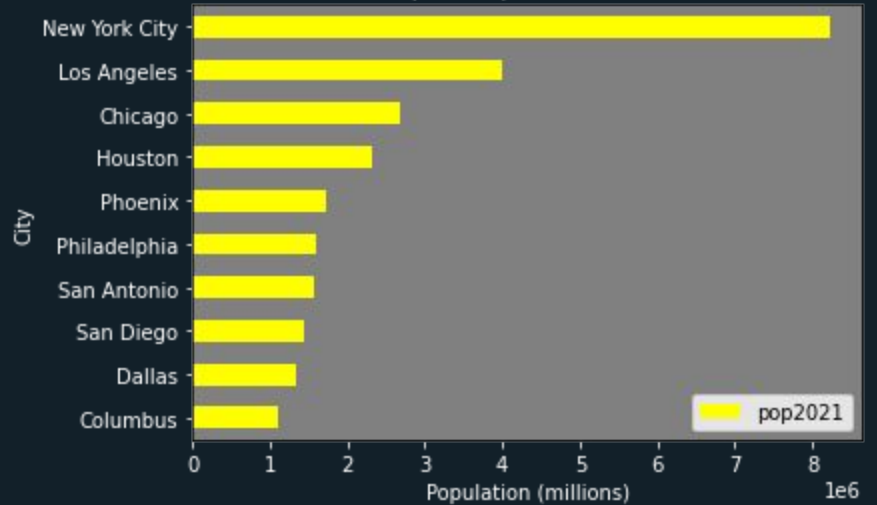


# City Car Accidents vs. Population

Top 10 Cities with Car Accidents

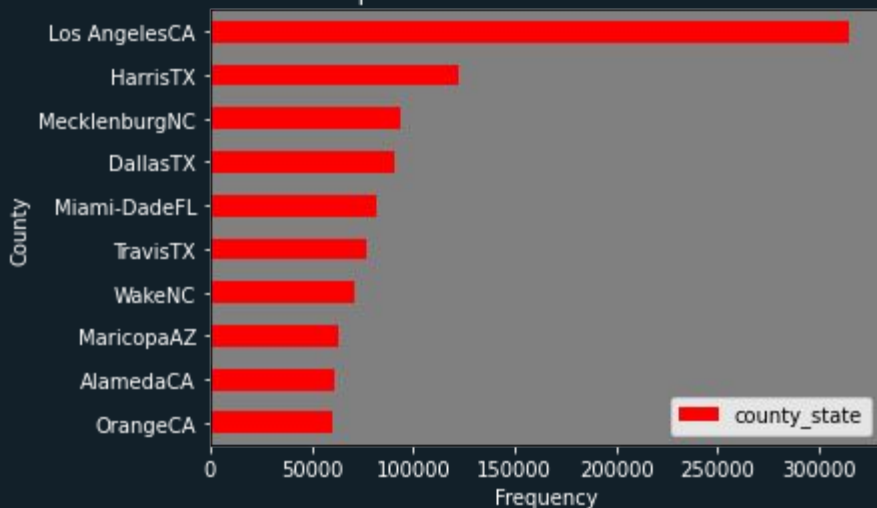


Top 10 Populated Cities

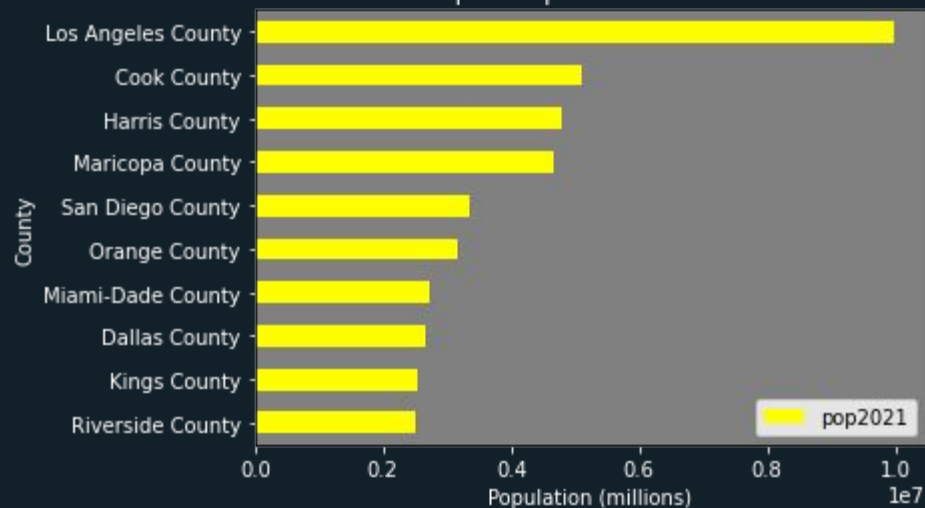


# County Car Accidents vs. Population

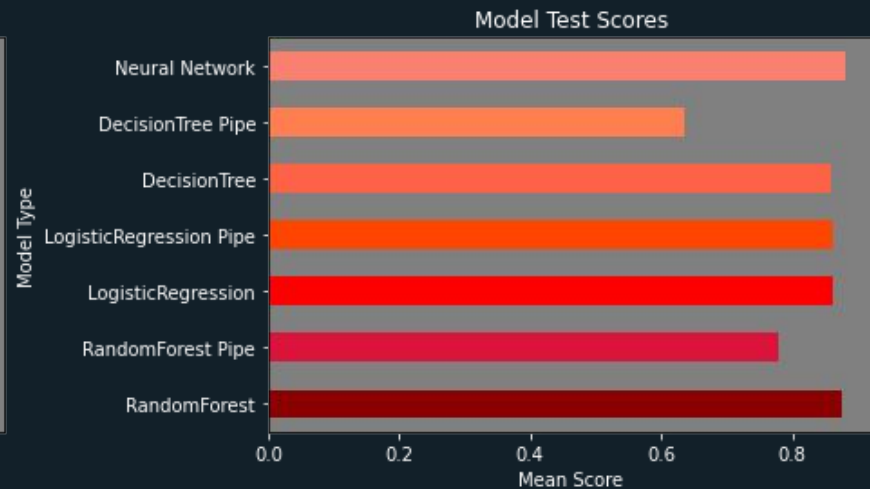
Top 10 Counties with Car Accidents



Top 10 Populated Counties



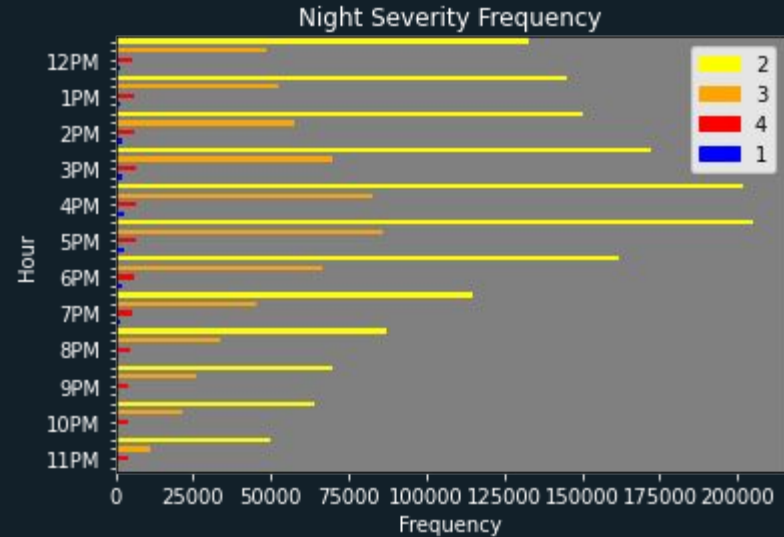
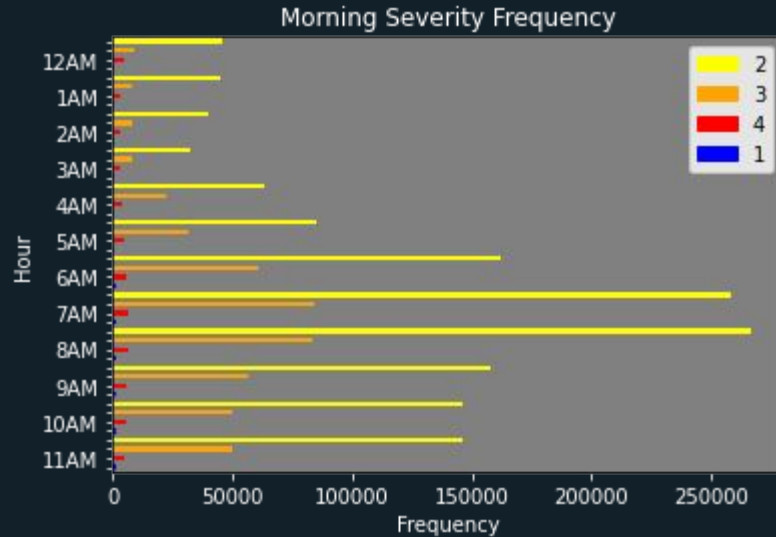
# Severity Models



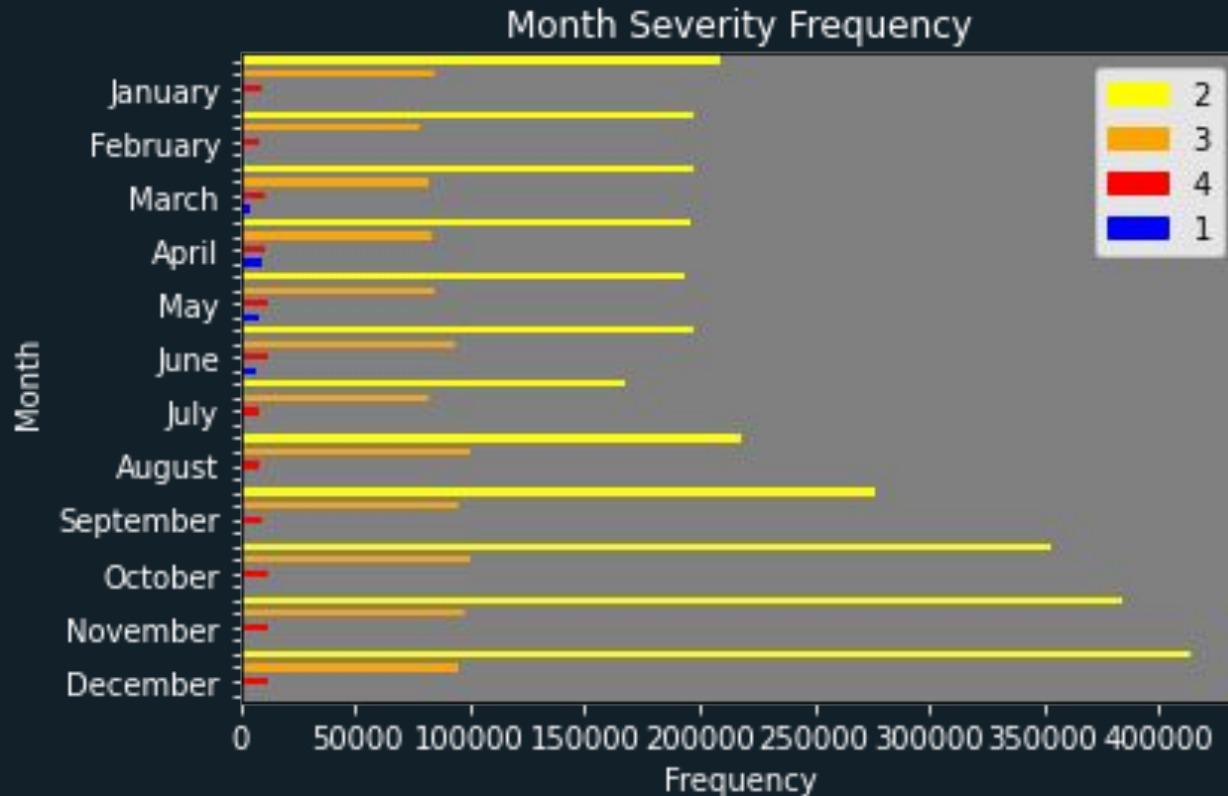
# Feature Importance



# Accidents by Hour



# Accidents by Month





# Predicting Location of Accidents

Using Primarily Weather and area informational data we fit a regression model to predict the location of accidents across the U.S.

## Model Accuracy

Lasso Regression - 3%

Linear Regression - 7%

Ridge Regression - 7%

Sequential Neural Network - 0.000023% accuracy

# Classification of Accidents by State

After the Failure of the Regression model and a substantial amount of frustration we moved on to a much simpler classification model to place the accidents in states.

## Model Accuracy

LogisticRegression - 22%

KNN - 21%

Neural Network 89%

# Extending the Project

If the project was to be extended getting data on speed limits for the accidents would probably be quite informative.

Road type would also probably be quite informative.

The Models would probably also work better with more substantial weather information such as snowfall around where the accidents occurred.

# Conclusion

The road can be a much more dangerous place than many people give it credit for

Certain conditions can lead to significantly less safe driving

- More Severe Accidents Happen in more Rural Areas
- Based on Kepler Data More Severe Accidents Happen on FreeWays and Interstates
- Accidents increase around the Holidays and in more Disruptive Weather

Being informed of the risk they are taking while driving will allow people to be more conscious of their choices and generally more able to reduce risk on the road.

# Clare Drive