CMPT732 Final Project

Analysis and Prediction of US Traffic Accidents

Background

- *How Many Car Accident Deaths Are There Per Year in US?*
  1.35 million people are killed
  Additional 20-50 million are injured or disabled

- *Economic impact of traffic accidents in US?*
  About $1 trillion loss per year

Source:
https://www.latimes.com/business/autos/la-fi-hy-economic-impact-of-traffic-accidents-20140529-story.html

https://www.thewanderingrv.com/car-accident-statistics/

Data

US-Accidents: A Countrywide Traffic Accident Dataset

acquired from Kaggle: https://www.kaggle.com/sobhanmoosavi/us-accidents

- **Severity**
- **Location**
- **Time**
- **Weather**
- **Points of interest**

Population and Housing Unit Estimates

obtained from United States Census Bureau: https://www.census.gov/programs-surveys/popest.html

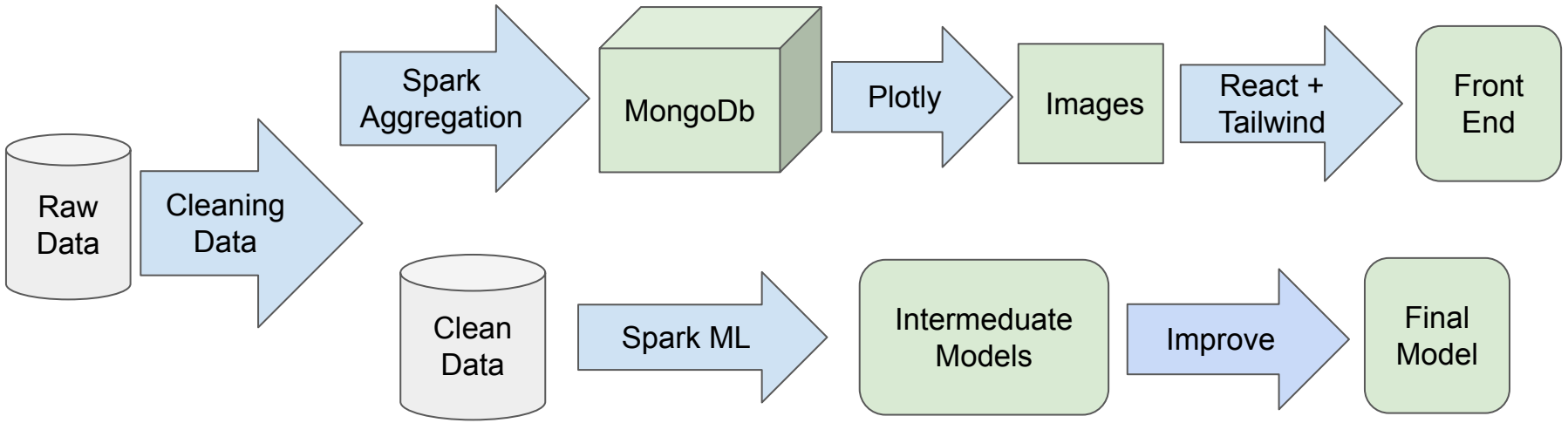
- **Population**

2020 United States Presidential Election Results

generated from Wikipedia: https://en.wikipedia.org/wiki/2020_United_States_presidential_election#Maps

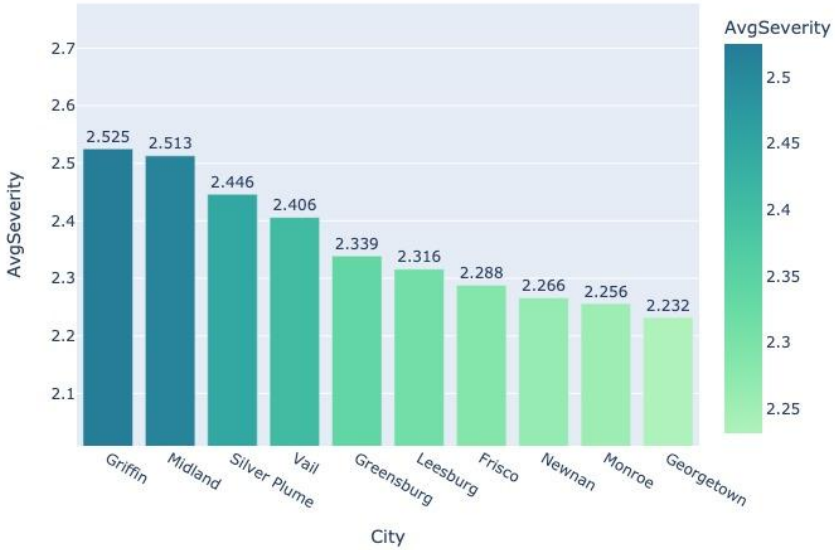
- **Candidate's percentage**

Methodology

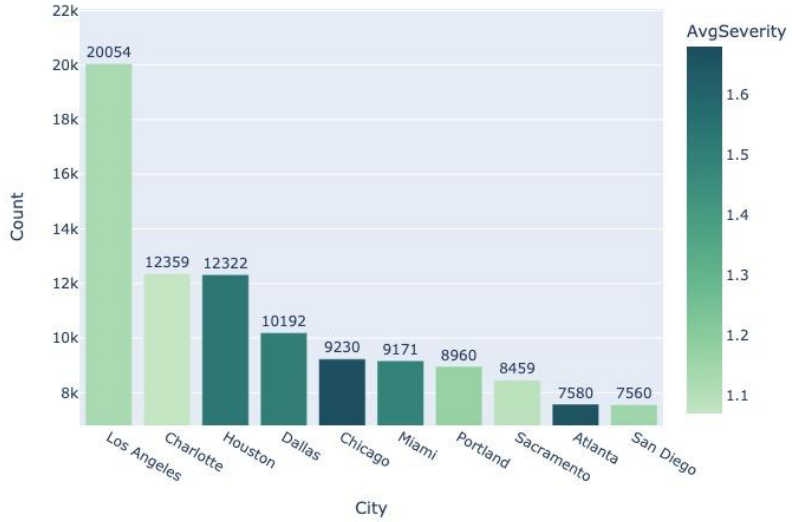


Location Analysis

Cities with most severe accidents

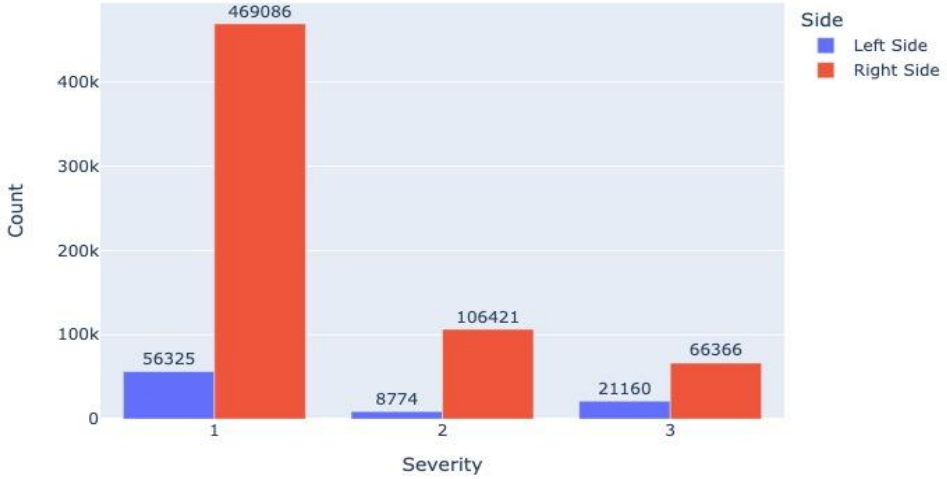


Cities with most number of accidents



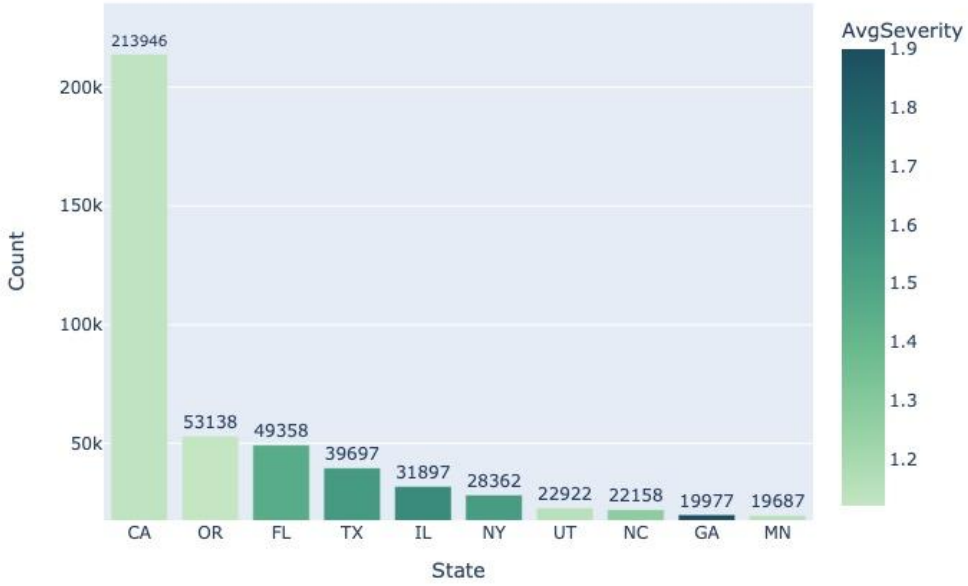
Location Analysis

Road vs Accidents



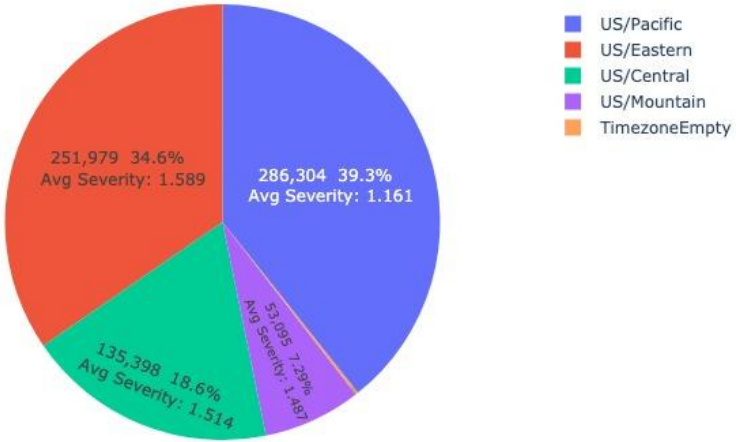
Location Analysis

Number of Accidents by States



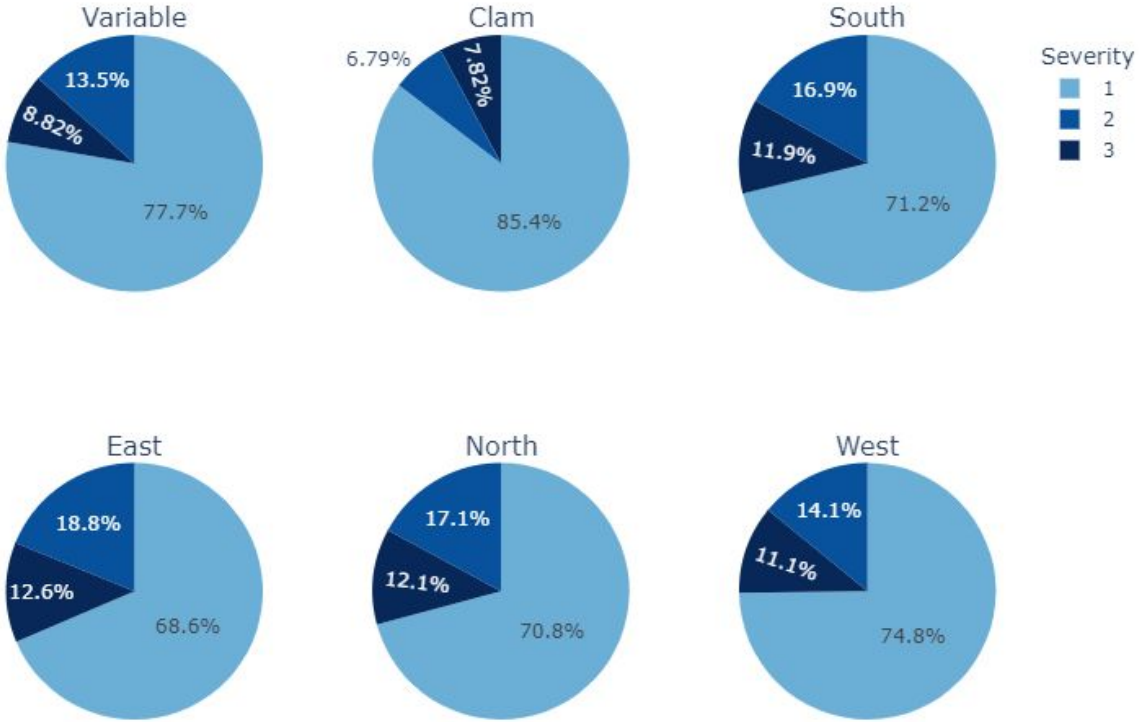
Location Analysis

Accidents Timezone Distribution



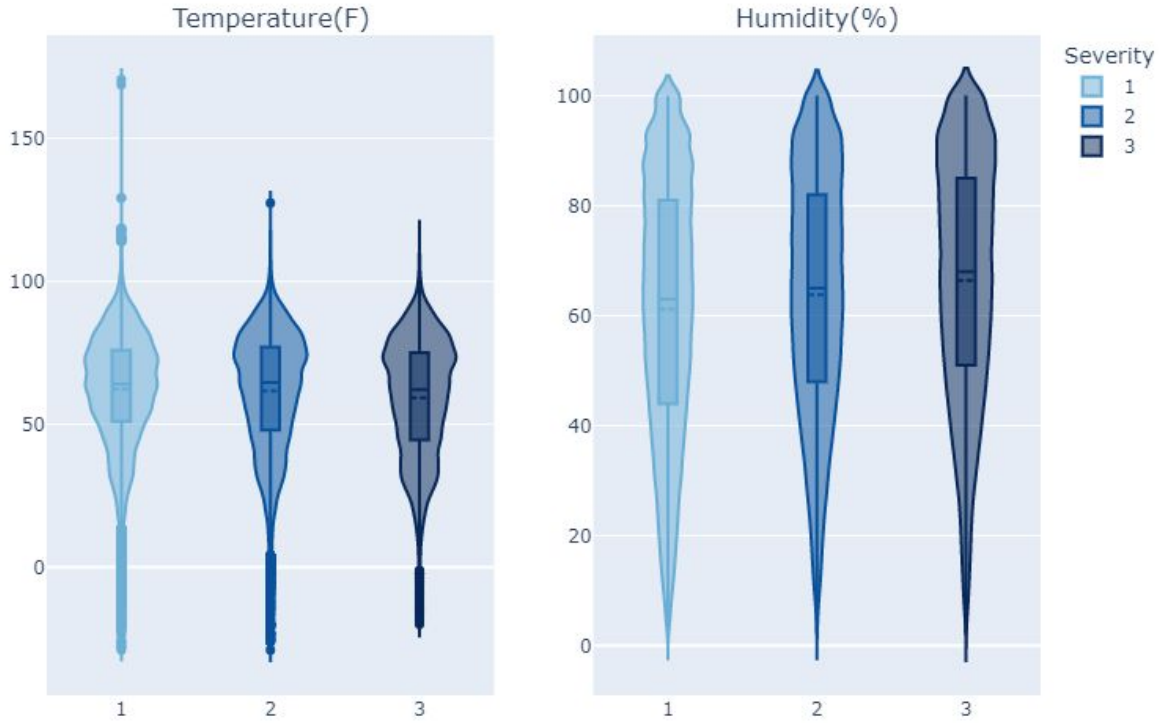
Weather Analysis

Proportions of Accident Severities in Various Wind Directions



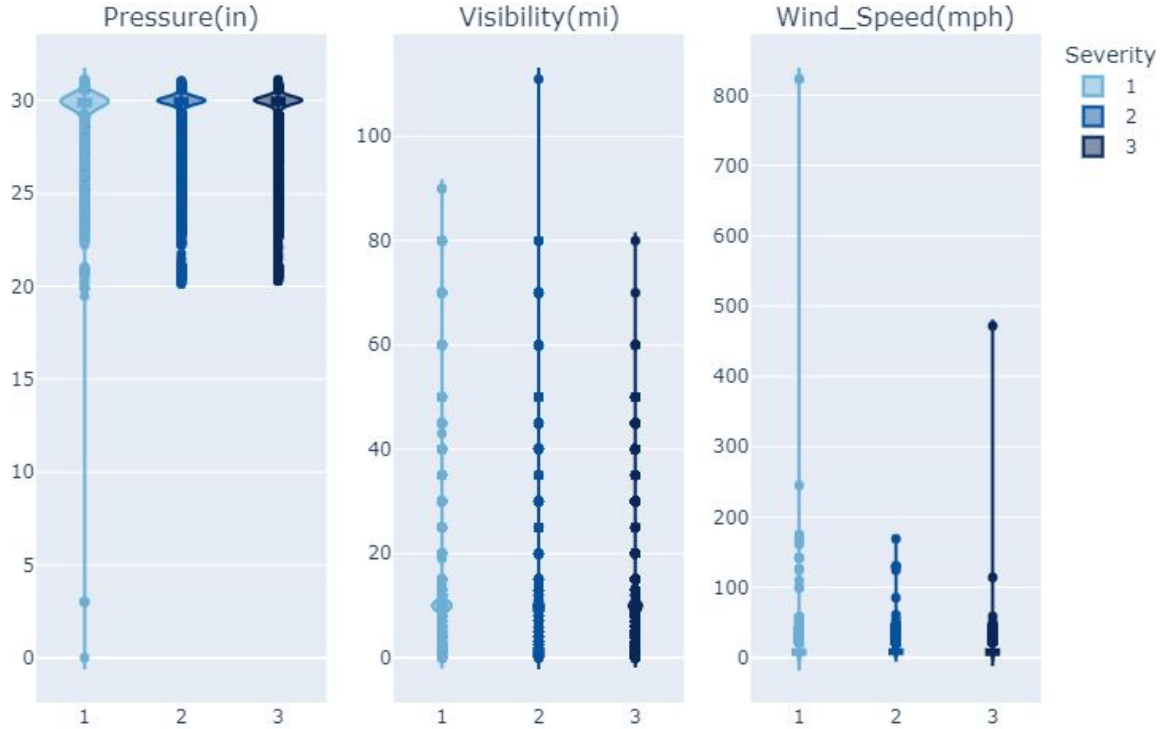
Weather Analysis

Density of Accident Severities in Temperature(F) and Humidity(%)



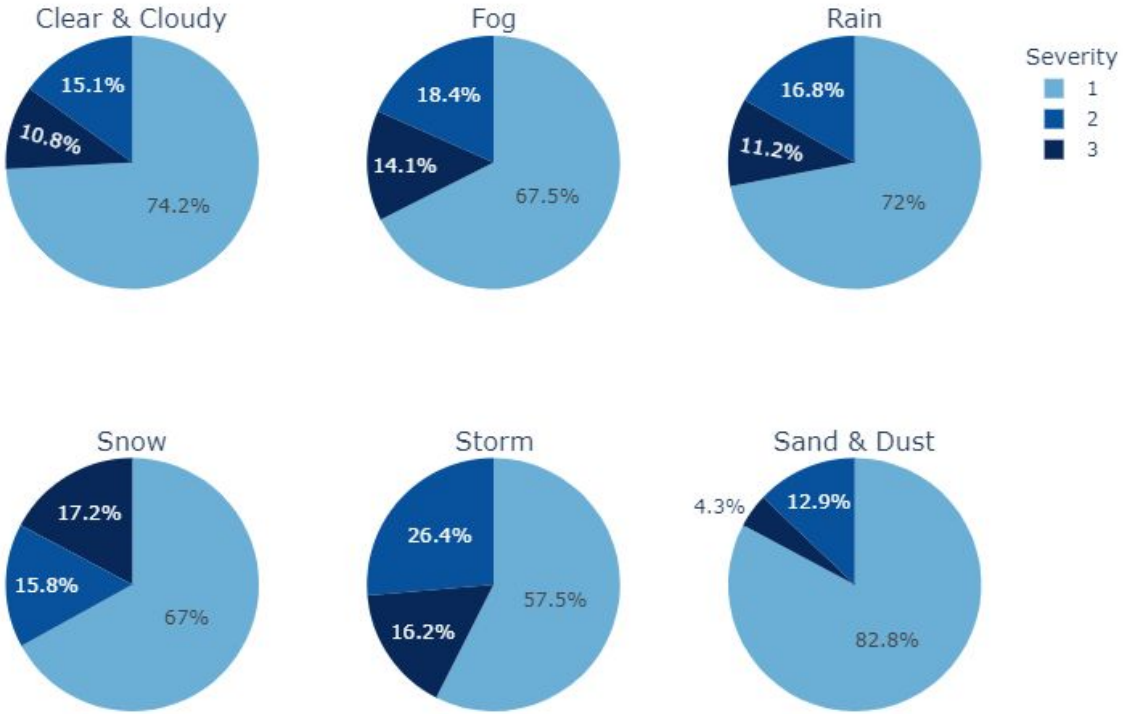
Weather Analysis

Density of Accident Severities in Pressure(in), Visibility(mi), and Wind Speed(mph)



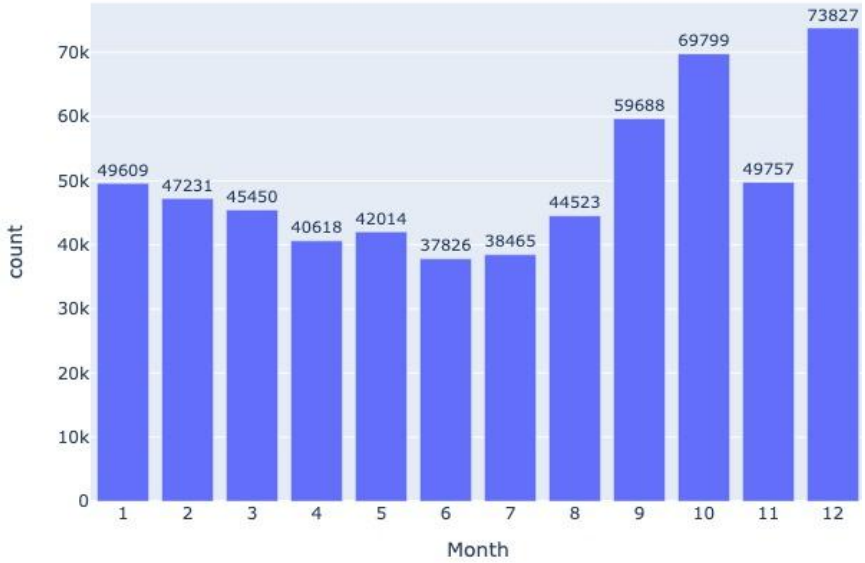
Weather Analysis

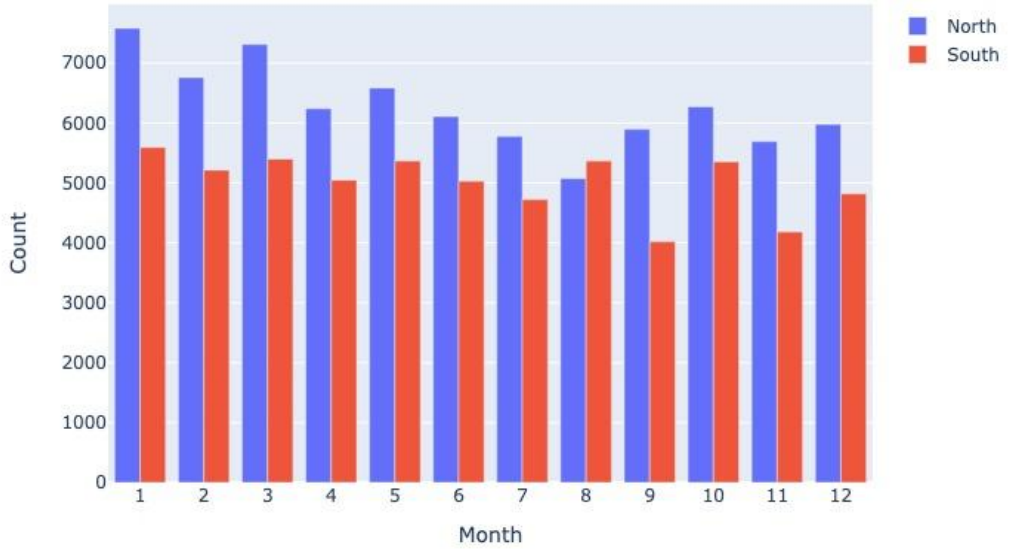
Proportions of Accident Severity in Various Weather Conditions



Car Accident Time Analysis

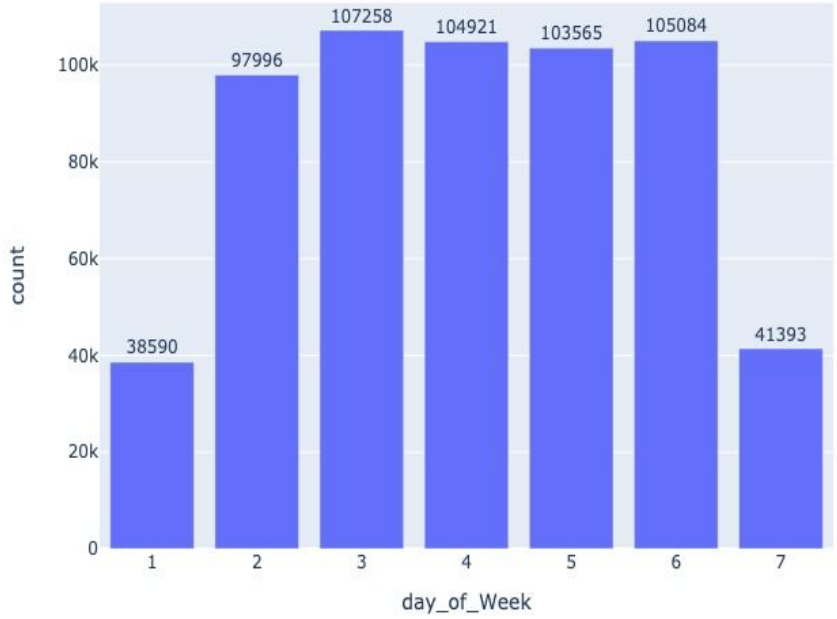
The left graph shows the monthly statistics of traffic accidents, and the right graph shows the monthly statistics of traffic accidents in the north and south (latitude >40 and <35).

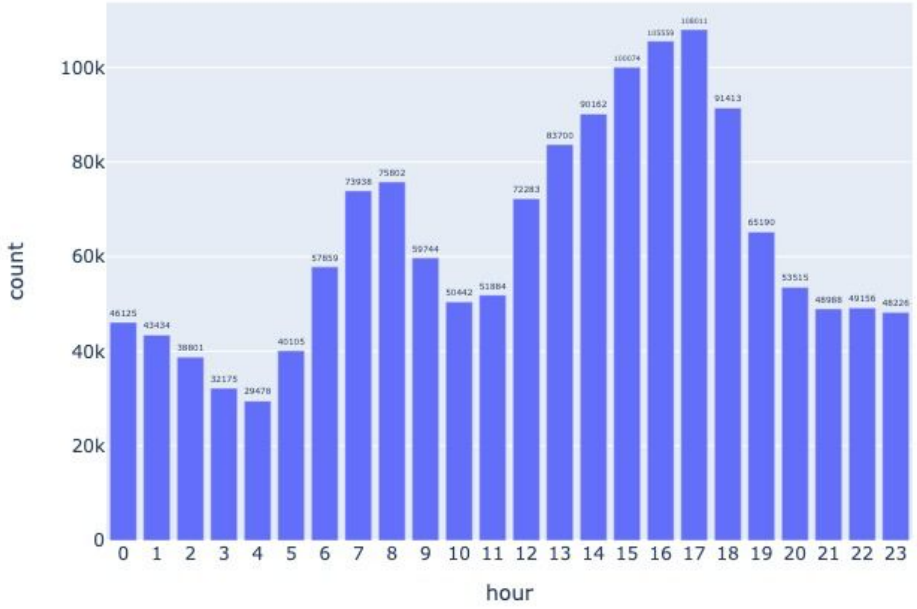




Car Accident Time Analysis

Traffic accident counts by day of week(left) and hourly traffic accident counts(right)





Car Accident Time Analysis

The left table shows that which days has the most accidents in the cities with the most traffic accidents. And the right trale shows that Weather conditions on the day with the highest number of traffic accidents in LA.

date	city	count
2019-12-23	Los Angeles	138
2019-11-08	Los Angeles	124
2019-12-06	Los Angeles	115
2019-12-04	Los Angeles	114
2019-09-13	Los Angeles	107
2019-12-19	Los Angeles	106
2019-10-08	Los Angeles	101
2019-11-01	Los Angeles	100
2019-10-04	Los Angeles	100
2019-09-03	Los Angeles	99
2019-12-23	San Diego	98
2019-10-25	Los Angeles	94

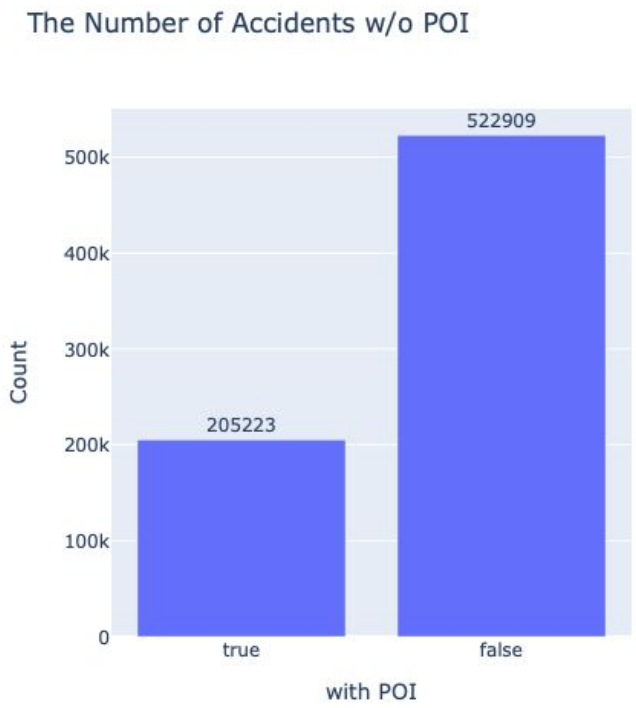
Weather_Condition	count	City	Date
Light Rain	51	Los Angeles	2019/12/23
Cloudy	31	Los Angeles	2019/12/23
Heavy Rain	21	Los Angeles	2019/12/23
Rain	15	Los Angeles	2019/12/23
Mostly Cloudy	8	Los Angeles	2019/12/23
Partly Cloudy	6	Los Angeles	2019/12/23
Fair	4	Los Angeles	2019/12/23
Light Rain / Windy	2	Los Angeles	2019/12/23

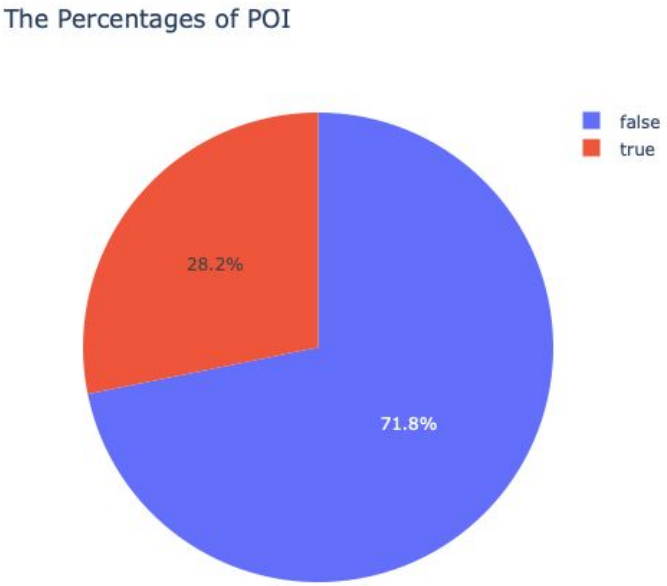
RED vs. BLUE

Rea StatesVS. Blue States：We find out five states (CA, NY, IL, MA, NJ) that Biden won in the 2020 election, and five states (OH, TX, TN, IN, MO) that Trump won. We can see that the number of accidents per capita in blue states is more than twice that of red states.

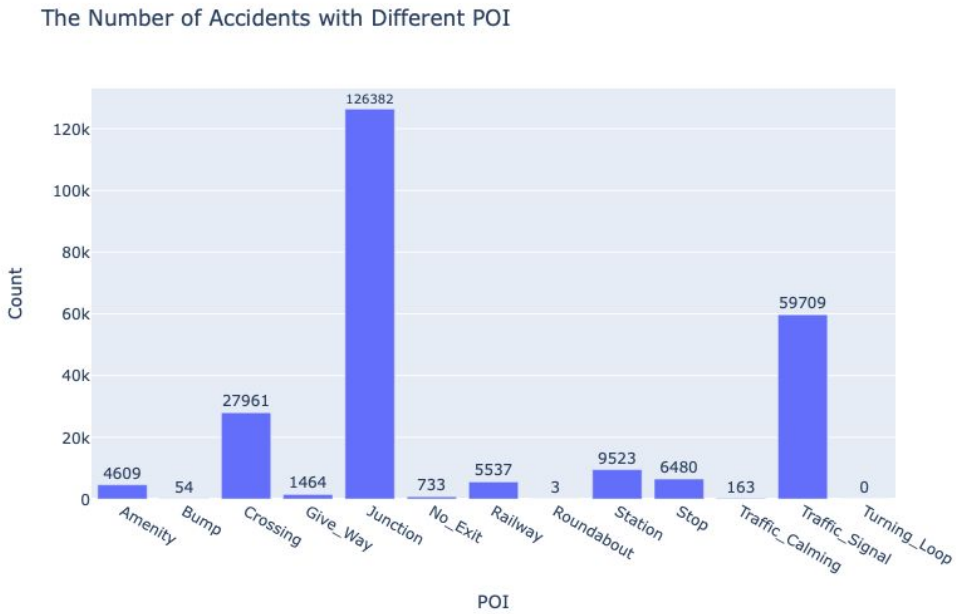
State	population	count	avg_accident
Red States	60383802	61729	0.001022
Blue States	87412298	239559	0.002741

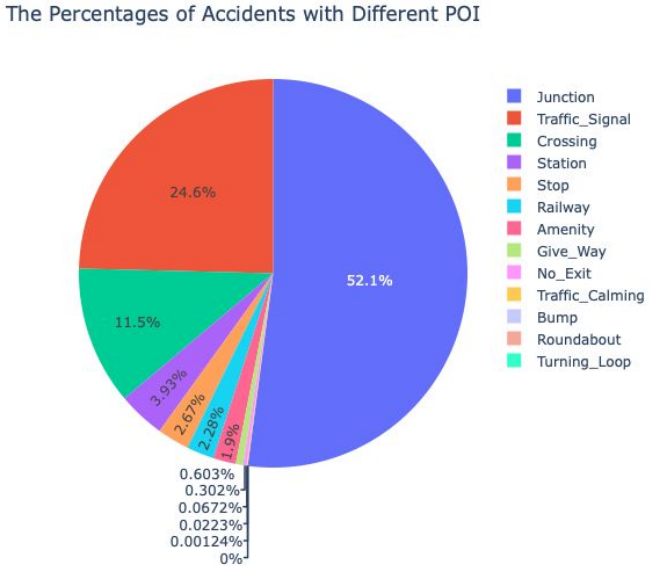
POI Analysis



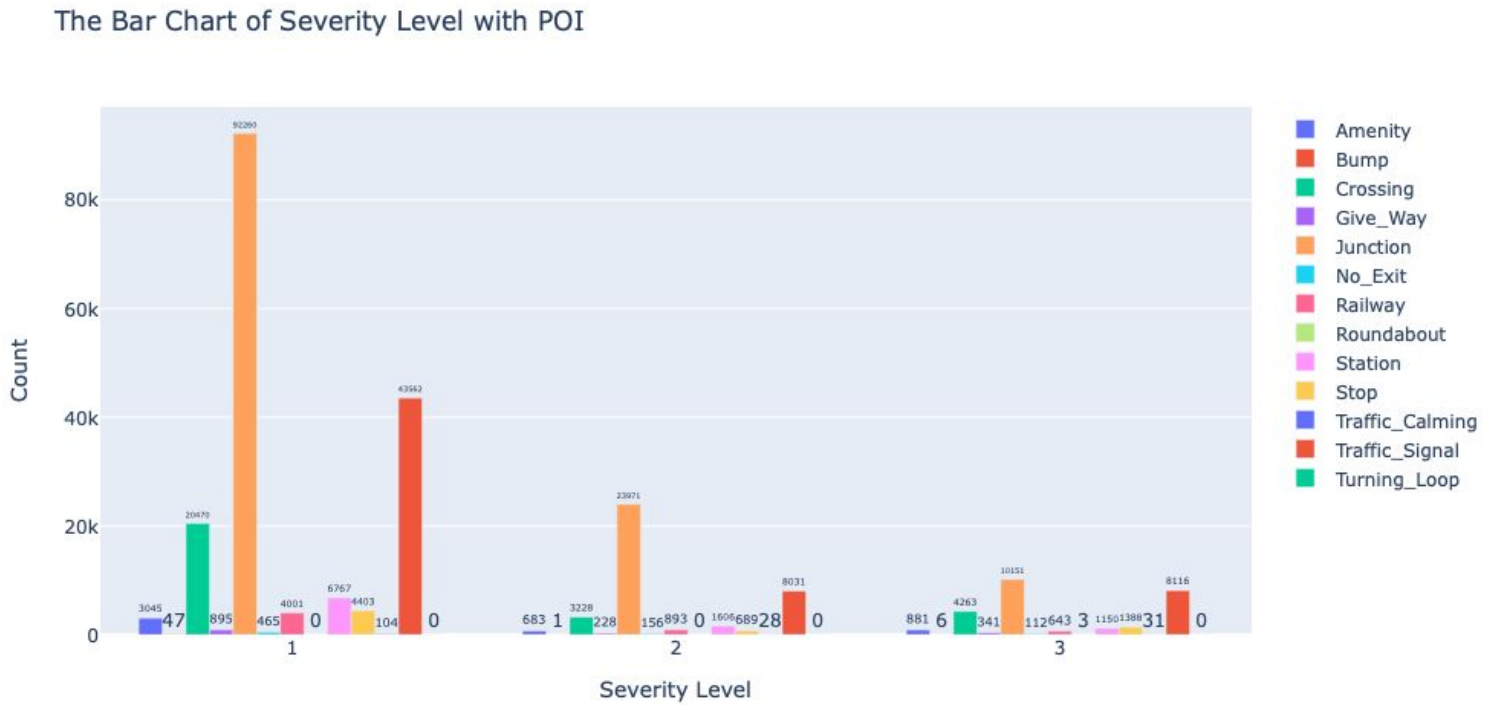


POI Analysis





POI Analysis

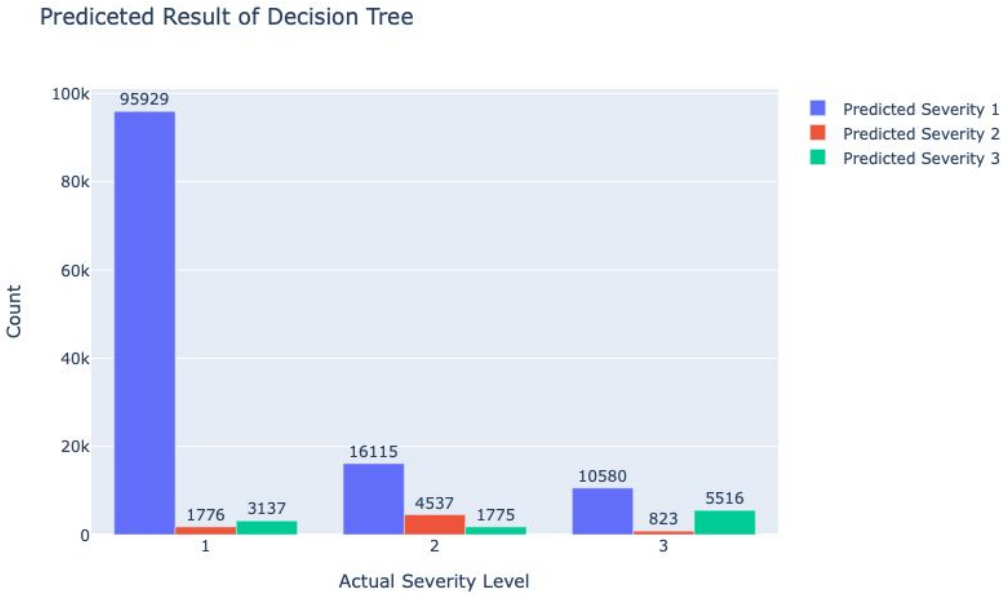


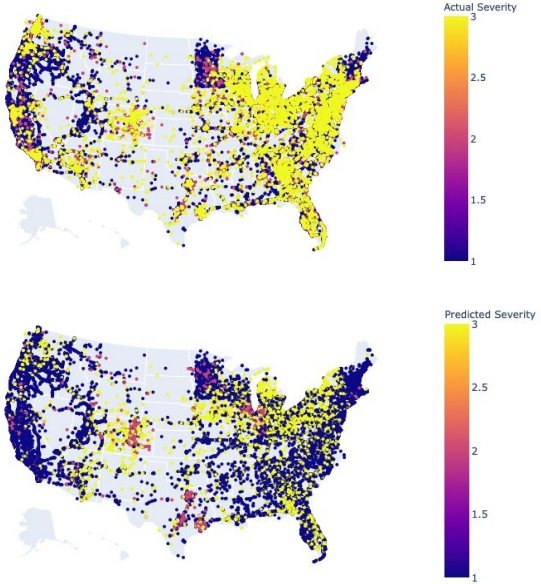
Classification and Prediction with Machine Learning

- Feature Selection
- Model: Decision Tree
- Training
- Evaluation
- Hyperparameter Tuning

Prediction Results

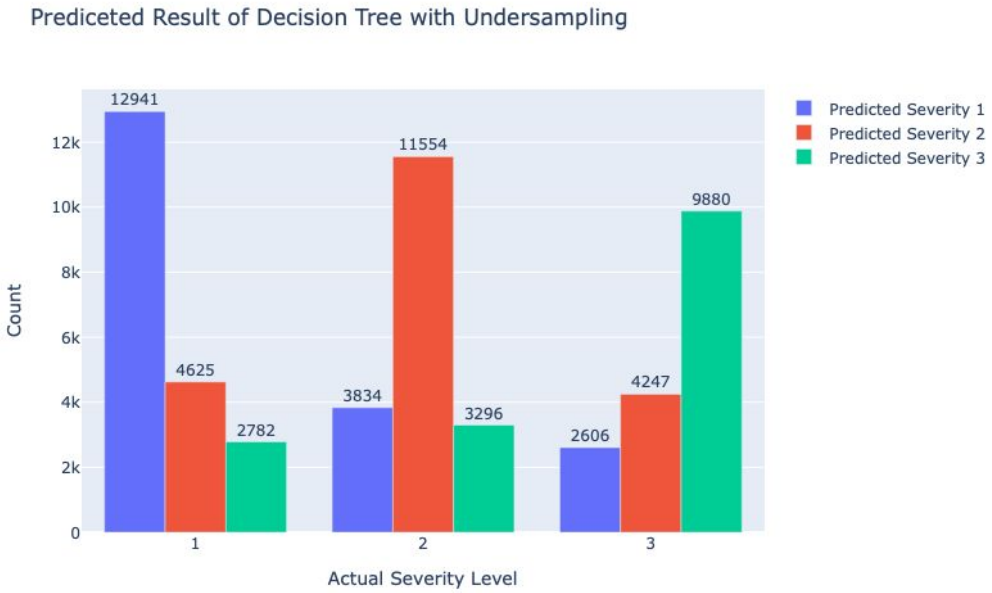
Traning accuracy: 78.04% Validation accuracy: 75.79%

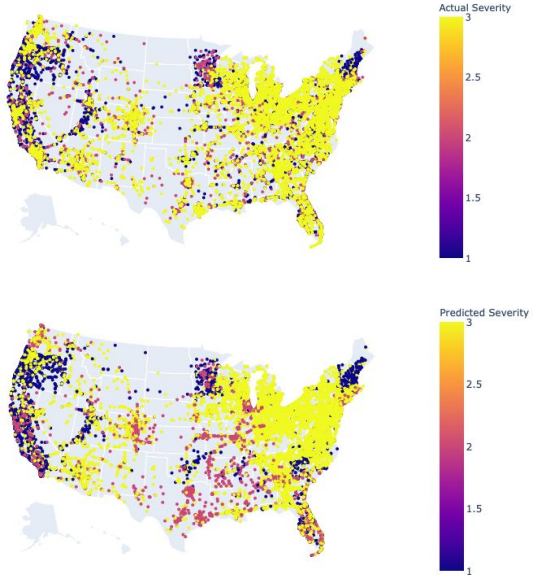




Undersampling

Traning accuracy: 66.16% Validation accuracy: 61.64%





Undersampling and Oversampling

Traning accuracy: 67.30% Validation accuracy: 64.97%