ORIE 4741 Final Project: Traffic Accidents Prediction

Jialiang Wei, jw2684

Jingxuan Li, jl4267

Heru Wang, hw743

Question

Our goal is to analyze US Traffic Accident data and identify the most important elements incurring the traffic accidents. We plan on answering the following questions:

- What kind of elements affects the probability of accidents most? Natural elements (weather, humidity, etc.) or social elements (distance from convenience store, population density etc.)
- Is there any advice for individuals and government to prevent traffic accidents?

Dataset

This is a countrywide traffic accident dataset, which covers 49 states of the United States. The data is continuously being collected from February 2016, currently there are about **3.5 million** accident records in this dataset. For every accident record we have data on source of the accident report, natural language description of the event, severity of the accident, start and end time, latitude and longitude of the start and end point, the zip code in address field, time-stamp of weather observation record, temperature, wind chill, humidity, air pressure, visibility, wind direction and speed, presence of crossing, railway, traffic signal, junction, etc., and the period of the day.

Our data was collected from Kaggle.com
https://www.kaggle.com/sobhanmoosavi/us-accidents

We plan on using various combination of the distributions listed above to predict under what circumstance is one accident most likely to happen.

Project Value

We believe this is a worthwhile project because both individuals and governments will benefit from the analysis. Through our findings, meaningful advice could be given out to prevent and reduce the occurrence of accidents.