Frequency and Severity of Car Accidents in the US

By Denis Azarov

04/29/2023

Email; LinkedIn

CS 06 Intro to Data Viz w Python

Stanford Continuing Studies

Lecturer: Cathal J Flanagan

Goals & Sources

GOALS

The dataset of car accidents in the US was analysed using different visualizations to understand the frequency, severity, and distribution of car collisions across the United States and in selected cities. These visualizations can provide information that can save human lives and lower insurance costs.

DATA SOURCES

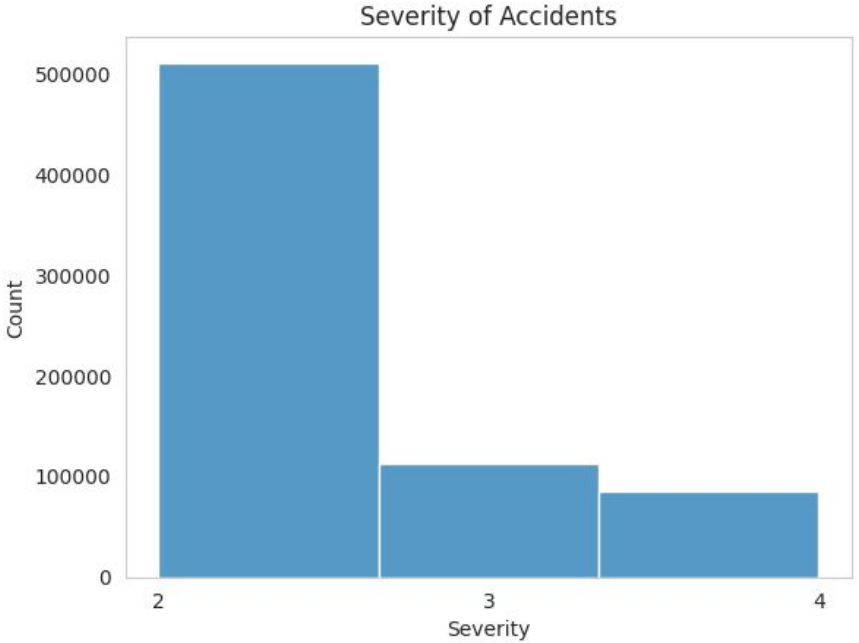
The countrywide car accident dataset was sourced via Kaggle. (To access the data a registration on Kaggle is required). The accident data is collected from February 2016 to Dec 2021, using multiple APIs that provide streaming traffic incident (or event) data. A subsets of data for San Francisco and Los Angeles were used for GEO visualizations. The dataset was also reduced to the timeframe from 2016 to 2019 to exclude possible COVID-19 effects.

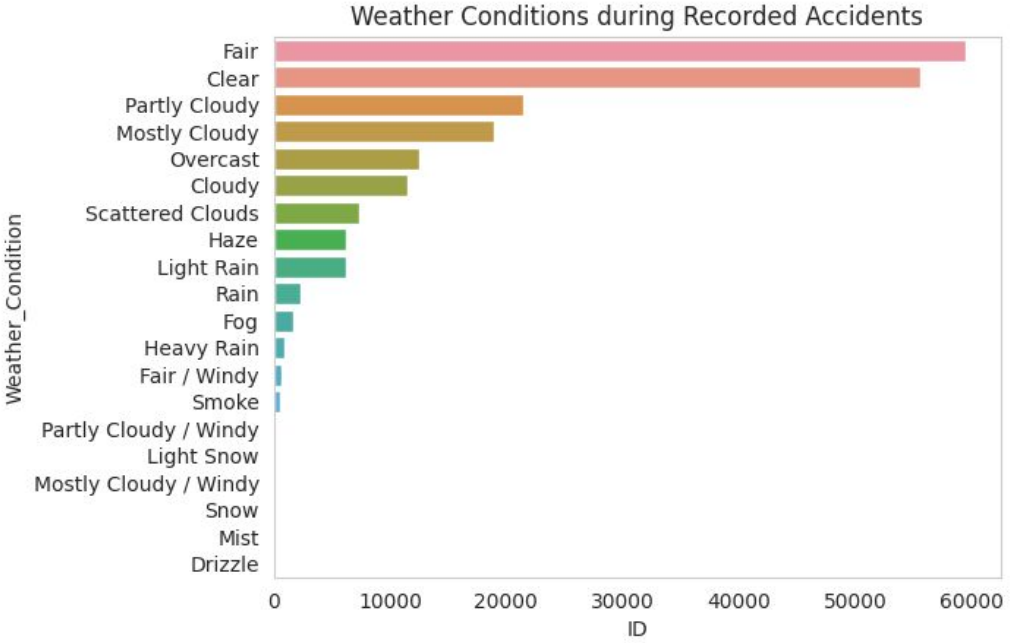
Severity is a number between 1 and 4, where 1 indicates the least impact on traffic (i.e., short delay as a result of the accident) and 4 indicates a significant impact on traffic (i.e., long delay).

Sample selection is biased as there is no evidence that the dataset is complete.

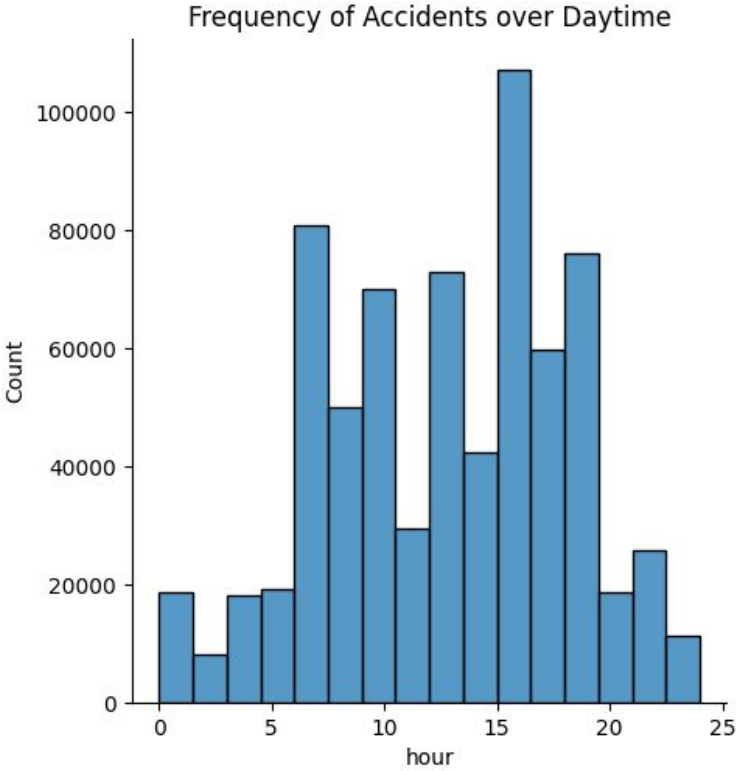
Google Colab: https://colab.research.google.com/drive/1NICl3jKb0Vzqjbsv7TrvcPCFomgNfrdL?usp=sharing

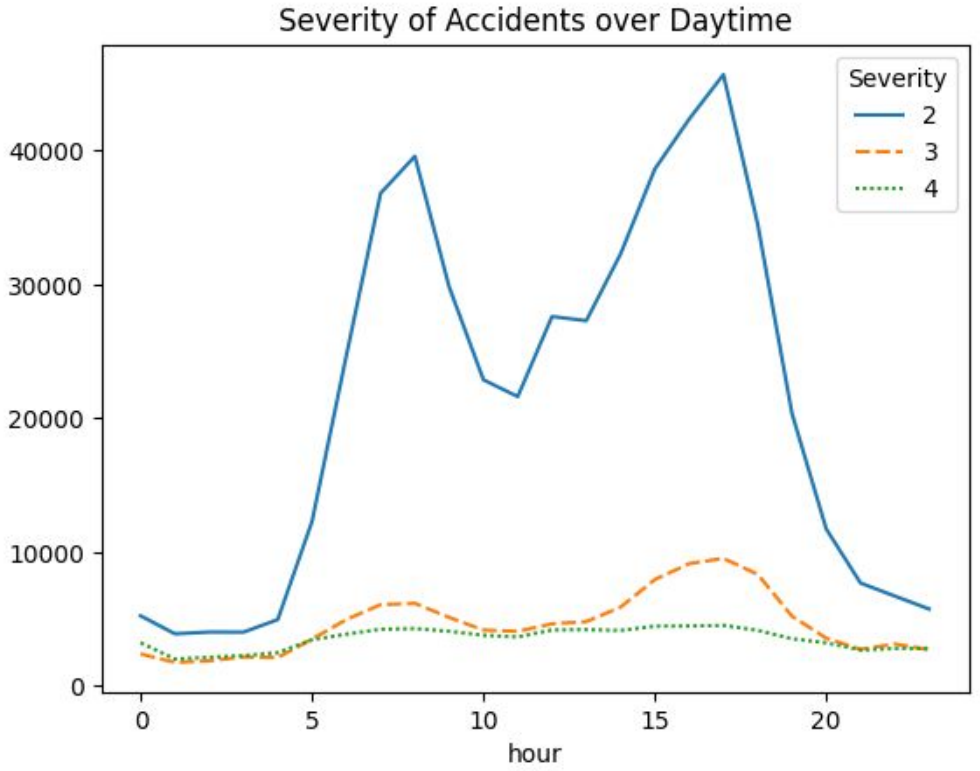
Severity & Frequency Analysis in the US



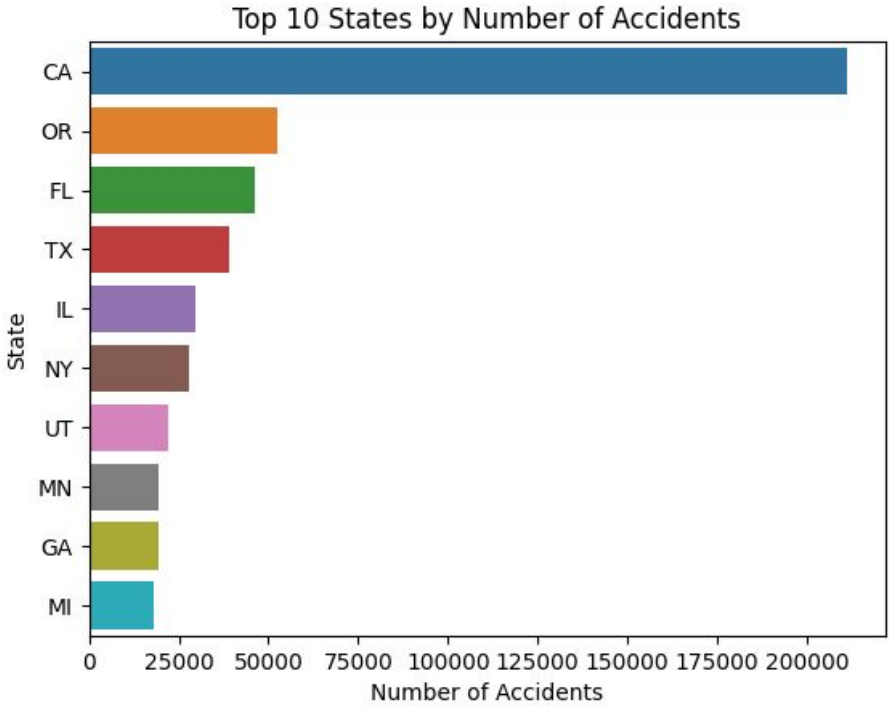


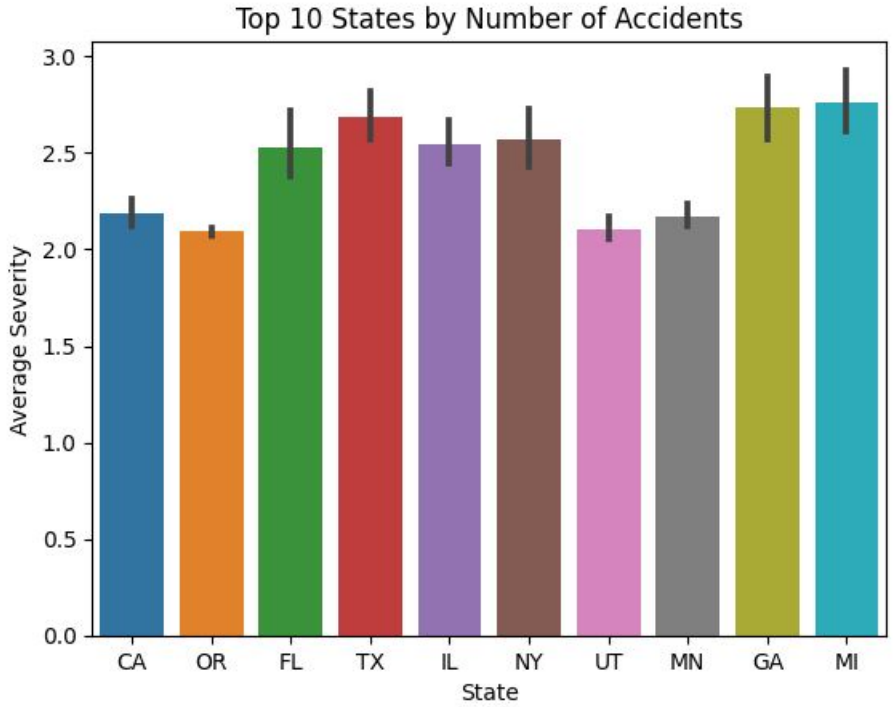
Severity & Frequency Analysis in the US



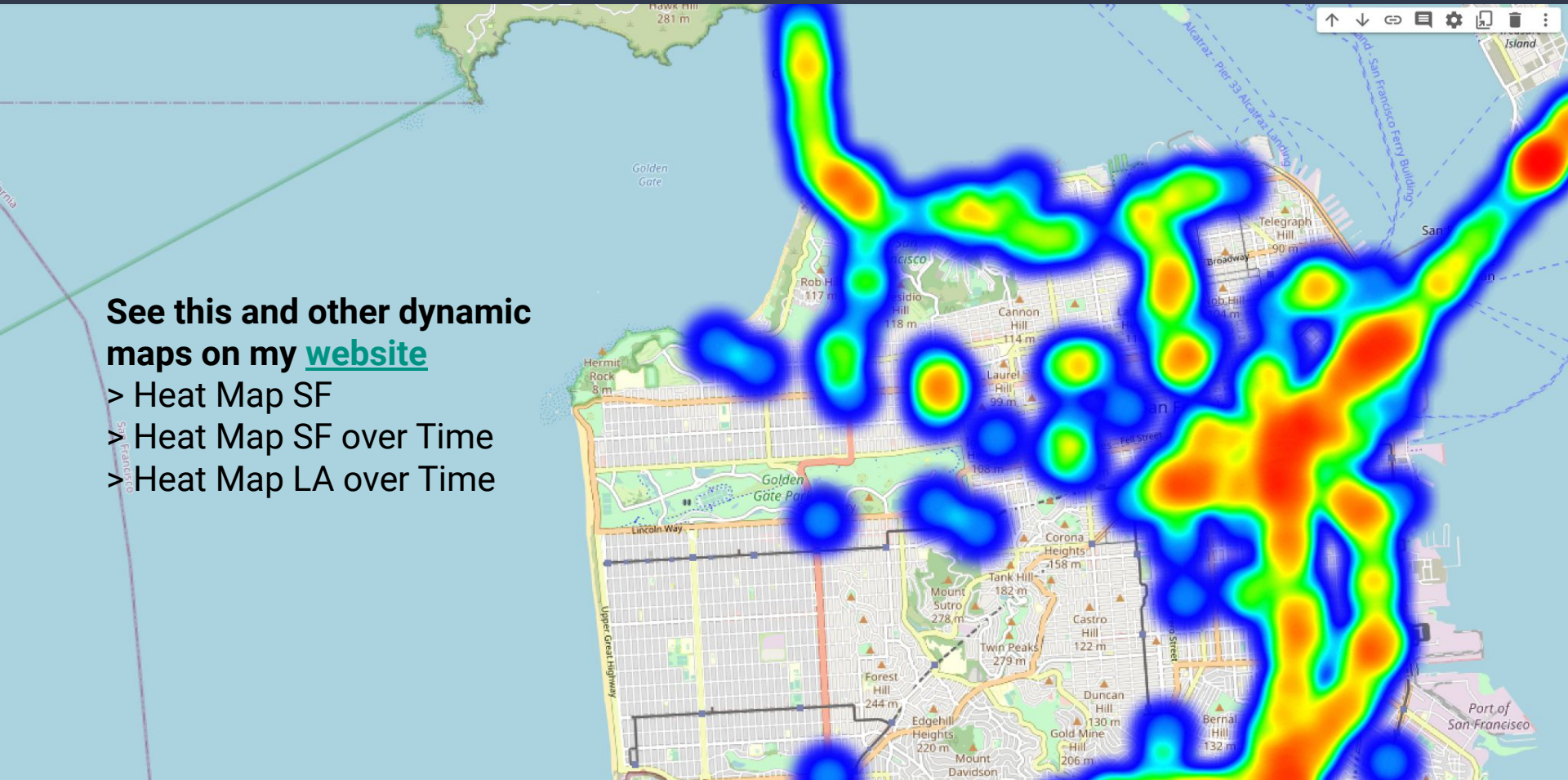


Analysis of the States





GEO Analysis of the Accidents' Frequency



See this and other dynamic maps on my website

> Heat Map SF

> Heat Map SF over Time

> Heat Map LA over Time