EXPLORATORY DATA ANALYSIS & VISUALIZATION OF UNITED STATES YEARLY ACCIDENTS WITH A DASHBOARD

Amir Khoeilar

Abstract

We live in a country that it is estimated to have at least 115 million cars and trucks on its streets everyday. Roads full of cars and trucks operated by humans, unfortunately will cause in many accidents. The main causes of accidents have be recognized as reckless driving, distracted driving , vehicle defects and intoxicated driving.
I'm going to be analyzing the data yearly collected from all of the unites states accidents and perform an exploratory data analysis with visualizations and provide an online dashboard for educational purposes and also providing a tool for authorities in order recognize and better understanding of the accidents happening locally.

Design

I began with downloading the data that from Kaggle. The data itself was collected yearly from 2016 till 2020 yearly with APIs from different websites. I put the data into Sqlite database and with the use of Sqlalchemy, I separately got the data for 2019 and 2020. After converting both two tables into separate CSV file, I performed simple EDA in order to have a better understanding of the data. I used visualization tools like Matplotlib, Seaborn, Folium and Tableau in order to show different graphs from the data. Lastly, using visual studio, I wrote a script for streamline where I wanted to make local application where the visualizations were displayed.

Data

The data was provided by the Kaggle website where the owner has used APIs to collect the data. The accident data are collected from February 2016 to Dec 2020, using multiple APIs that provide streaming traffic incident (or event) data. These APIs broadcast traffic data captured by a variety of entities, such as the US and state departments of transportation, law enforcement agencies, traffic cameras, and traffic sensors within the road-networks.The data that I used was only for the year 2019 and 2020 which each one has about 271000 and 780000 accident data.

Algorithms

Feature Engineering

1. Exploratory Data Analysis

Model Evaluation and Selection

In the conclusion, the dashboard provided will display different visualizations with regards of the data provided early and eventually would bring extensive sight towards accidents happening around US.

Tools

- SQLite, Sqlalchemy
- numpy, pandas
- Seaborn, matplotlib, Tableau , Folium
- Streamlit

Communication

There is going to be the code for the data operations and also PDF slides of powerpoint presentation available on my GitHub account.

https://github.com/rezxkoi/Data_ENG_USA_Accidents_2019vs2020-covid-_app