VISUAL SIX
COVID-19 Impact on US Traffic Accidents

Team 036 – Bavithra Radhakrishna, Chandra Sekhar Nookarapu, Lohitha Rajasekhar, Sravya Karamched, Swaminathan Murugesan, Uday Bag

Georgia Tech

COVID-19 CORONAVIRUS DISEASE 2019

Motivation

Every year 6 million accidents occur on an average in the USA, more than 90 people die in car accidents every day, and 3 million people are injured in car accidents every year. Because of their frequency, traffic accidents are a major cause of death globally, cutting short millions of lives per year. Therefore, a system that can predict the occurrence of traffic accidents or accident-prone areas can potentially save lives. With these facts, we aim to build an Interactive User Interface that provides a thorough visual analysis of US Traffic Accidents.

Data

US Traffic Accident
Source: Kaggle
Size: 1.5 GB
Records : 4.2 million
Timeline: Feb 2016 to Dec 2020

COVID
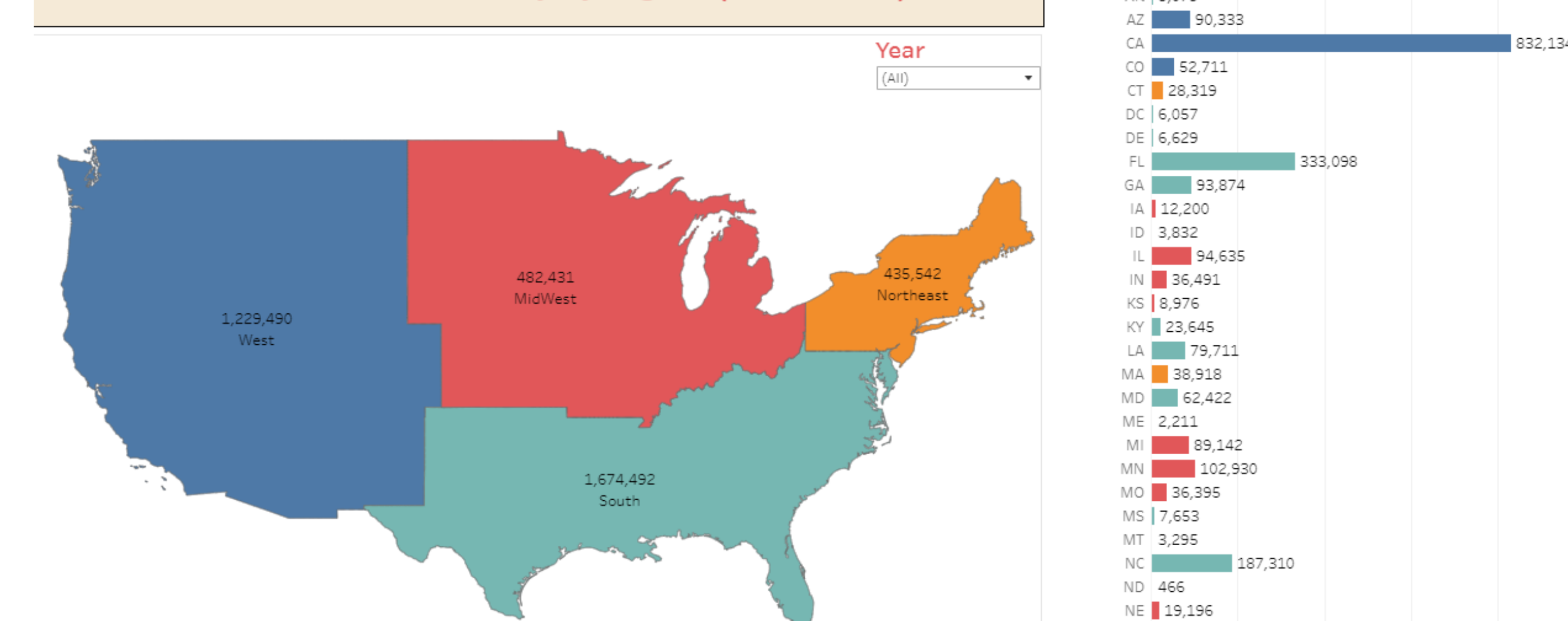**Source:** USA Facts
Size: 100 MB
Records: 1.2 million
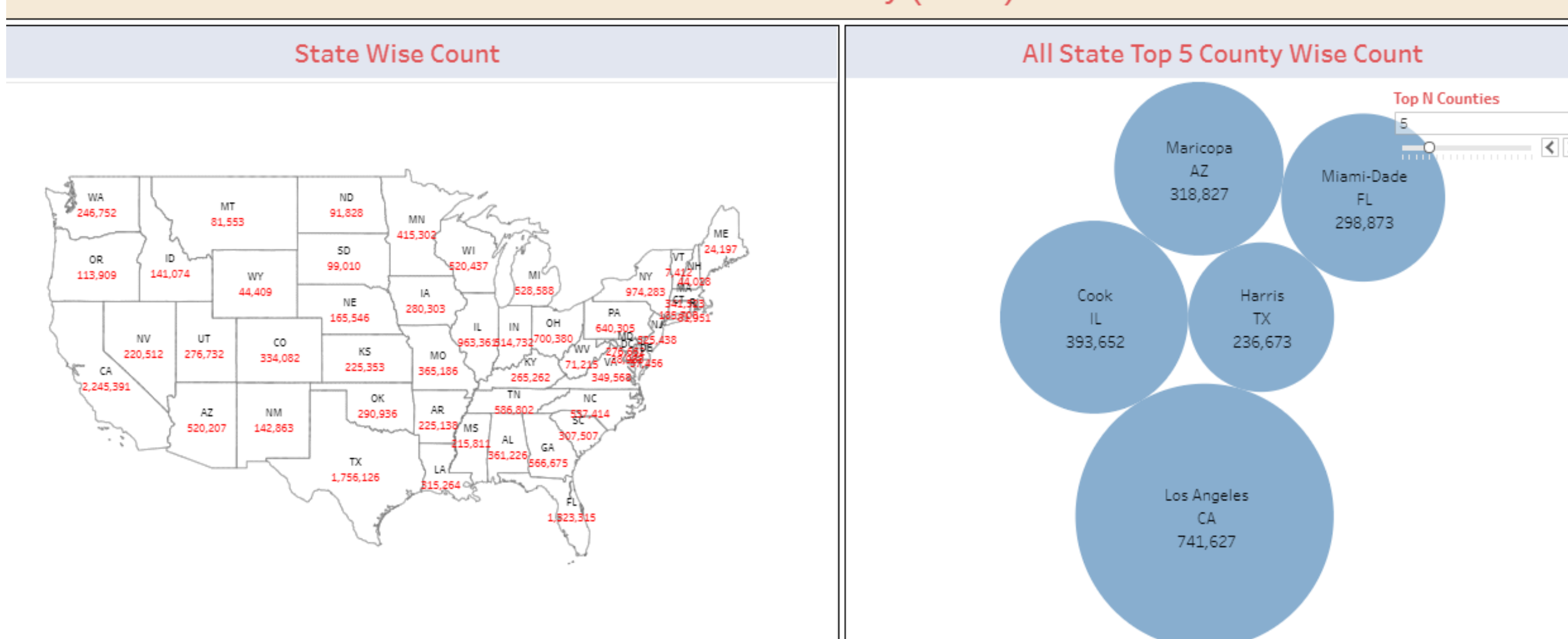**Timeline:** Jan 2020 to Dec 2020

Visualization

Traffic accident summary by region (2017-2020)

COVID cases summary (2020) — State Wise Count / All State Top 5 County Wise Count
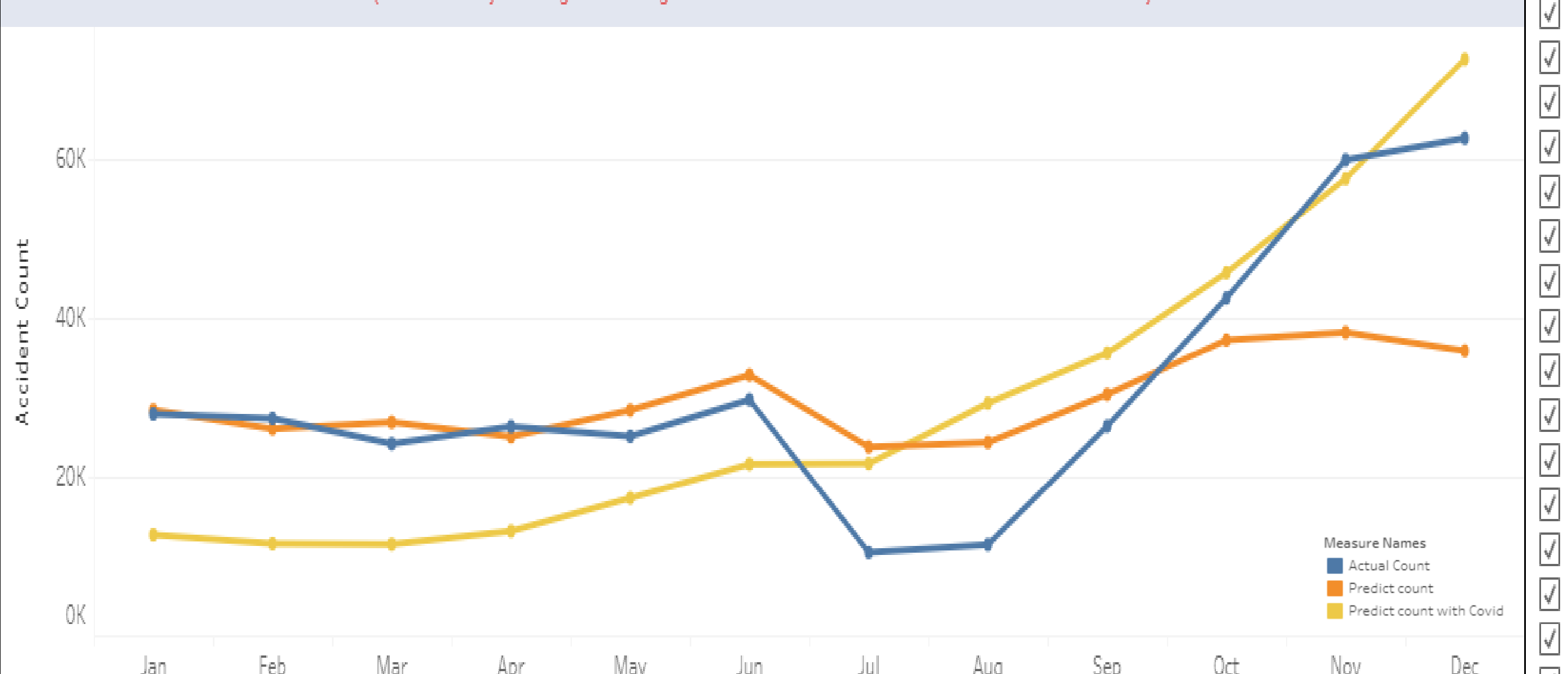

Actual vs Predicted accident count (2020)
(Model built by training and testing on 2020 data - With and without Covid counts as a factor)

Methodology

- Datasets hosted on AWS S3 bucket for storage
- Perform data manipulations using various tools(EMR, PySpark, SageMaker) on both datasets
- Analyze and locate the most concentrated traffic locations, having numerous accident criteria and apply ML model to find a comparative analysis on COVID data.
- Study accident impacts due to COVID cases and find similar/outlier trends.
- Develop Tableau dashboards that show the analysis of US traffic accidents

COVID-19 Impact on US Traffic Accidents — System Design Arch/Flow
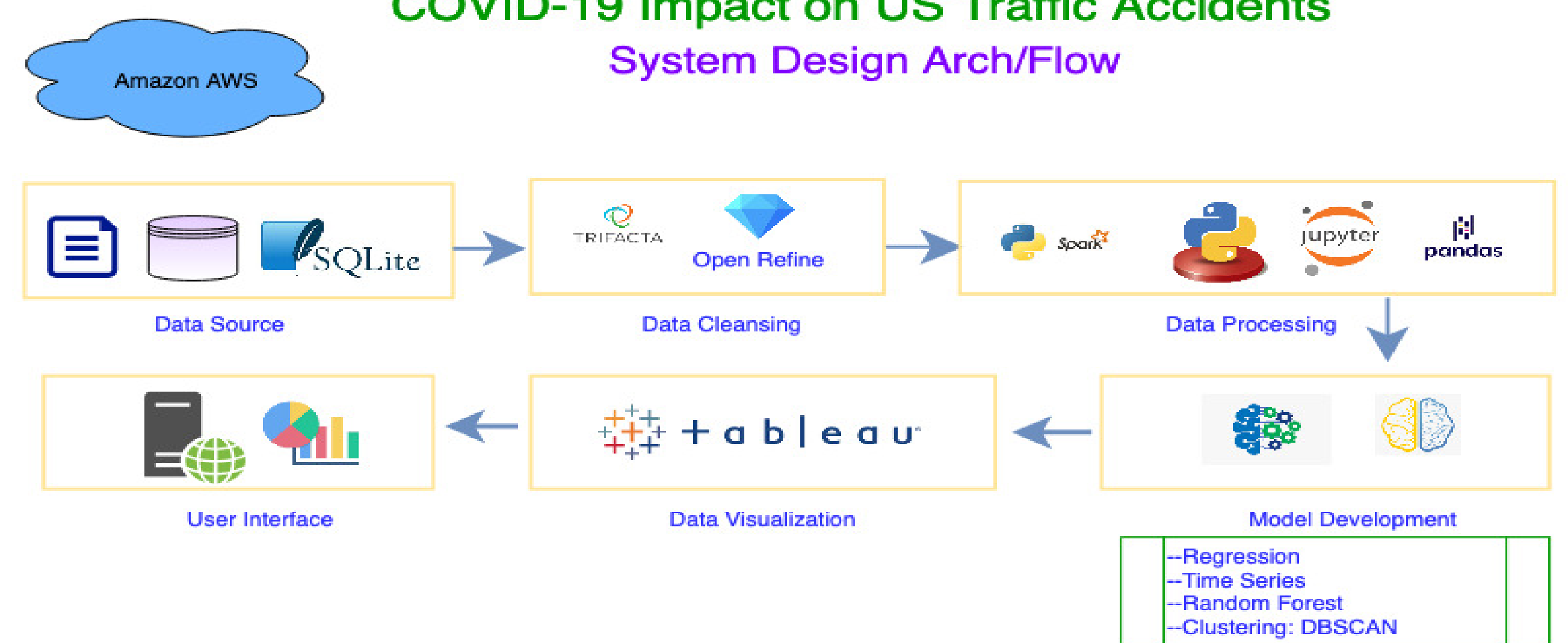
Experiments/Results

Model-1: Prediction of accident counts for 2020 using 2020 data as training/test sets.
Algorithms: k-means/DBSCAN, PCA, Linear/Logistic Regression
Results: The results show that densely populated locations had better metric values compared to sparse populated locations.

Model-2: Prediction of accident counts for 2020 using 2020 data as training/test sets with COVID as additional feature to the above model.
Algorithms: k-means/DBSCAN, PCA, Linear/Logistic Regression
Results: The results show that there is no significant correlation between the counts.

Model-3: Prediction/Trend analysis of accident counts for 2020 using 2016-2019 data as a training set.
Algorithms: k-means/DBSCAN, PCA, Linear Regression, Time series
Results: The results show that past trends are still valid during COVID-19.

Conclusion/Future Work

Our qualitative results show that there is necessarily no significant correlation between the traffic accidents and COVID-19. However, they provide a considerable starting point for future guidance in implementing new traffic rules and regulations. We observed that the past trend of increase in the number of accidents continued during the COVID-19 era. We noticed a dip in the accident count during mid of 2020(July, Aug) in majority of the states and this might be due to the lockdowns and stay-at-home orders enforced by the government. Although, the level of detail needed to interpret this fact is not always easily available to the public.

While analyzing the predictions of our model, the features available in the datasets is ranked as 5th factor of cause of accidents. To improve the accuracy of the model, identifying the right data sources and considering the additional causes as features for model building is suggested in the future.