

# MVP STATISTICAL COMPARISON APPLICATION FOR USA ACCIDENTS BETWEEN THE YEAR 2019 & 2020(COVID YEAR)

The Question:

How has accidents around USA been during a regular year (2019) in comparison with during 2020 which there was a pandemic going on ?

Extra:

- Multi-class classification machine learning model

I have download the USA accidents data from Kaggle which the owner of the data has used different APIs to scraped from different websites. The data includes all the accidents all of over USA from early 2016 till Dec 2020.



I have downloaded the data and stored it with SQL. I have used Pyspark, Pandas and Numpy in order to separate and collect the data for the year 2019 and 2020 implement EDA.

I have used Matplot-lib, Seaborn, Tableau and Folium to show some visualization in regards to my findings with my data.

Lastly, I will be using Visual Studio to write the “.py” file and use Streamlit to show all the visualization as an web application

Next Steps:

- More visualizations
- More EDA
- Tableau