

## **DUI Testing Location and Police Deployment to Reduce Fatality in Road Accidents**

### **1. Project Description**

This is a sub-project of a big project. The primary goal of this sub-project is to use past fatal accident data and real time traffic data to estimate the traffic police force that needs to be deployed in certain areas to prevent/improve fatal accidents. The final form of this project will be using past accident data and real time data to investigate issues on certain locations and try to predict fatal accidents in certain areas at a certain time.(I know this is very ideal, but it is practical and applicable).

### **2. How to run my Code**

The first part of data is stored in the form of csv files under the Data folder.

The input year will only take '2015', '2016', '2017', '2018', '2019', '2020' because only these six years of data are saved in the Data folder.

`get_accident_data(year: str)`

This function takes a string argument year.

it will return a dataframe from accident.csv of that year.

`count_accident_by_state(year: str)`

This function will take the year and return a dictionary of the number of accidents in each state.

`count_accident_by_weather(year: str)`

This function will take the year of wanted data and return two dictionaries by state, one dictionaries for counting rainy days accidents and the other counting non-raining days accidents.

`count_DUI_accident(year: str)`

This function will take the year of wanted data and return a dictionary by state for the number of DUI accidents.

`rate_of_DUI(year: str)`

This function will take the year of wanted data and return the DUI rate among all fatal accidents.

`histogram_plot_DUI_ratio(year: str)`

This function will take the year as an input plot of the Percent of DUI Fatal Accidents in All Fatal Accidents by State in that year.

`line_plot_DUI_ratio_2015_to_2020(states: list)`

This function will take a list of states as input and plot a line graph to show the change to DUI rate by state from 2015 to 2020.

It will also return a list of dictionaries for each year.

`get_realtime_accident_data(response_limit: str)`

This function makes an API request to the server, getting responses from the server and returning responses in json format.

the argument takes only integers and will return that amount of most recent data.

`search_city_location(search_city: str)`

This function will take the name of the city as a string and return a list of strings containing the longitude and latitude of the bottom left corner of the city and upper right corner of the city.

`count_alert(city: str)`

This function will take the name of the city as a string and return the number of real time alerts within 24 hours where the alert type is 'ACCIDENT' or 'HAZARD'. We eliminate the alert of 'JAM' and 'ROAD\_CLOSED' in this function.

link to github repository

<https://github.com/pyro157/DSCI-510-Final-Project.git>

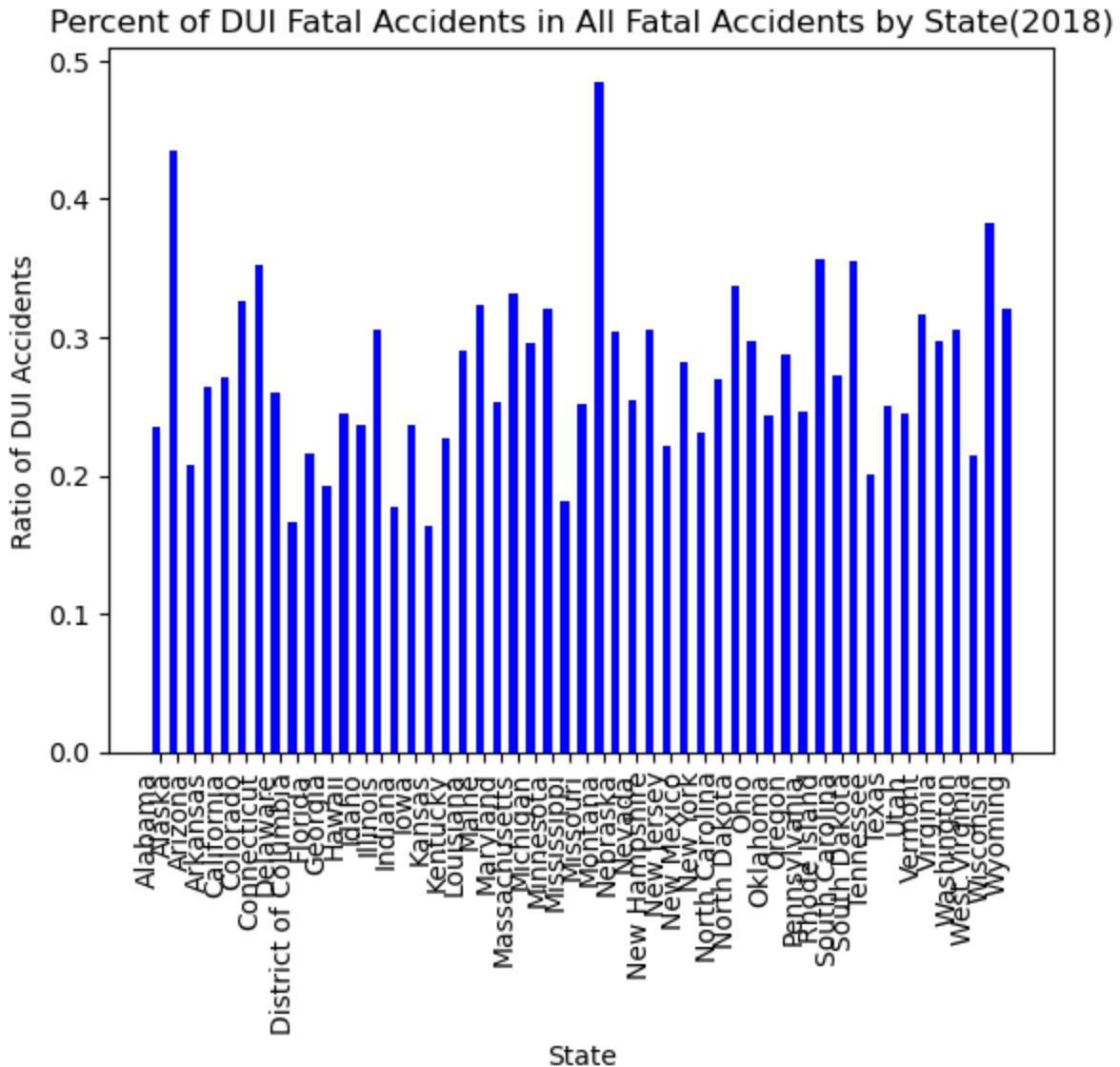
### **3. About My Data**

I have collected the Data about road accidents in the United States. I have used 2 different data sources where they are csv files that can be downloaded through <https://cdan.nhtsa.gov/> .

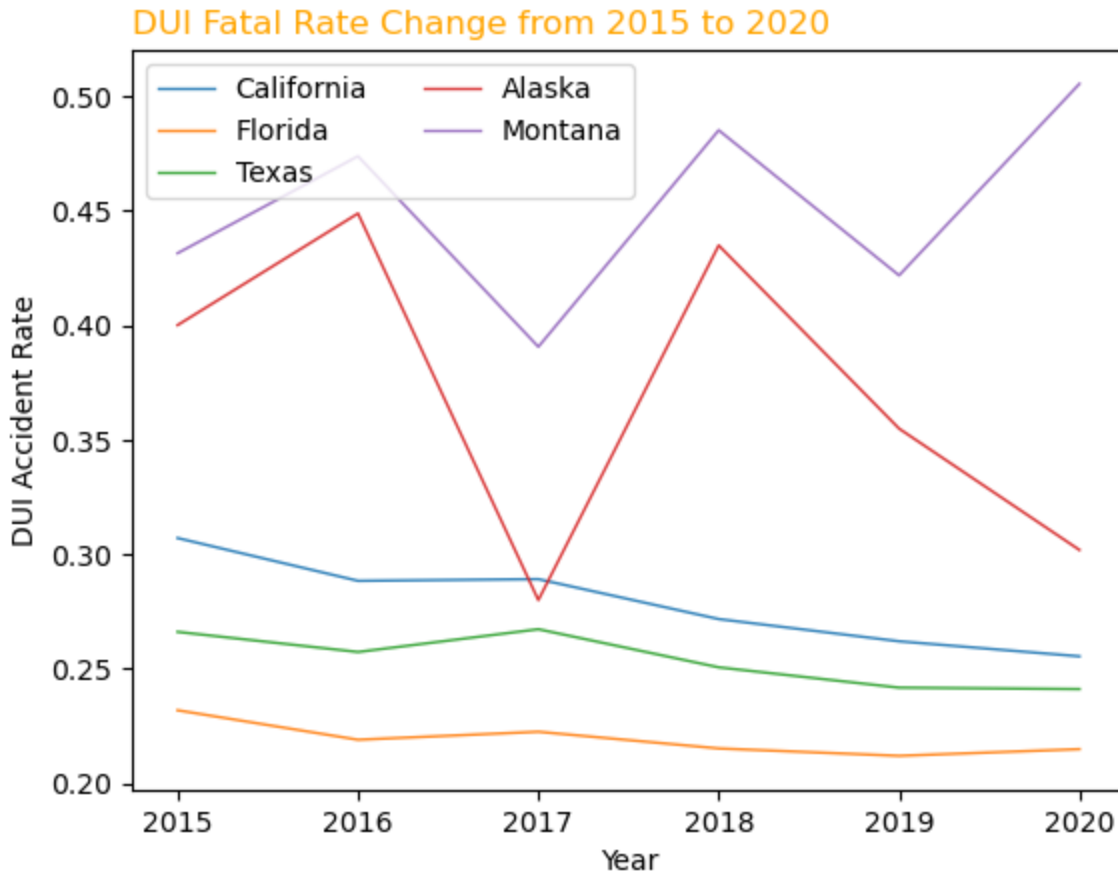
Another one is real time data from rapid API, where we can use the get method to request data from <https://realtime-road-accidents-in-usa.p.rapidapi.com> . There are more than 90000 data samples from 2018-2020, and for real time data, the user can enter the size of samples from most recent. My original plan was to use most of the factors in the data to create a complete program/application to predict fatal accidents. I realize that it requires much more time and effort rather than one month's development, so I decided to fulfill the requirements of the final project and do a section of the project I have planned in the beginning.

### **4. Analyses and Visualizations**

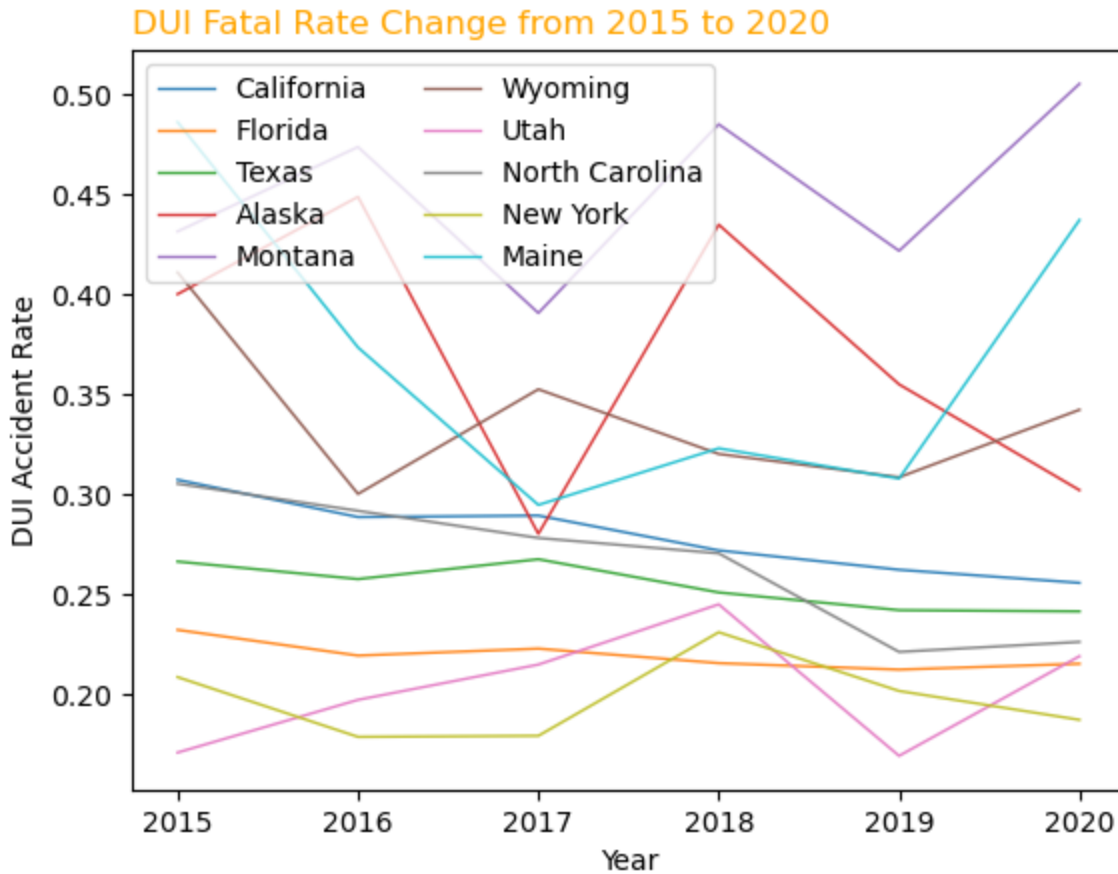
I have done analysis on the DUI fatal accident rate among all fatal accidents. I have done a histogram to get the result of the DUI rate by state in a certain year(where you can input the year from 2015 to 2020). And I have also done a line graph to show the change to DUI rate by state from 2015 to 2020.



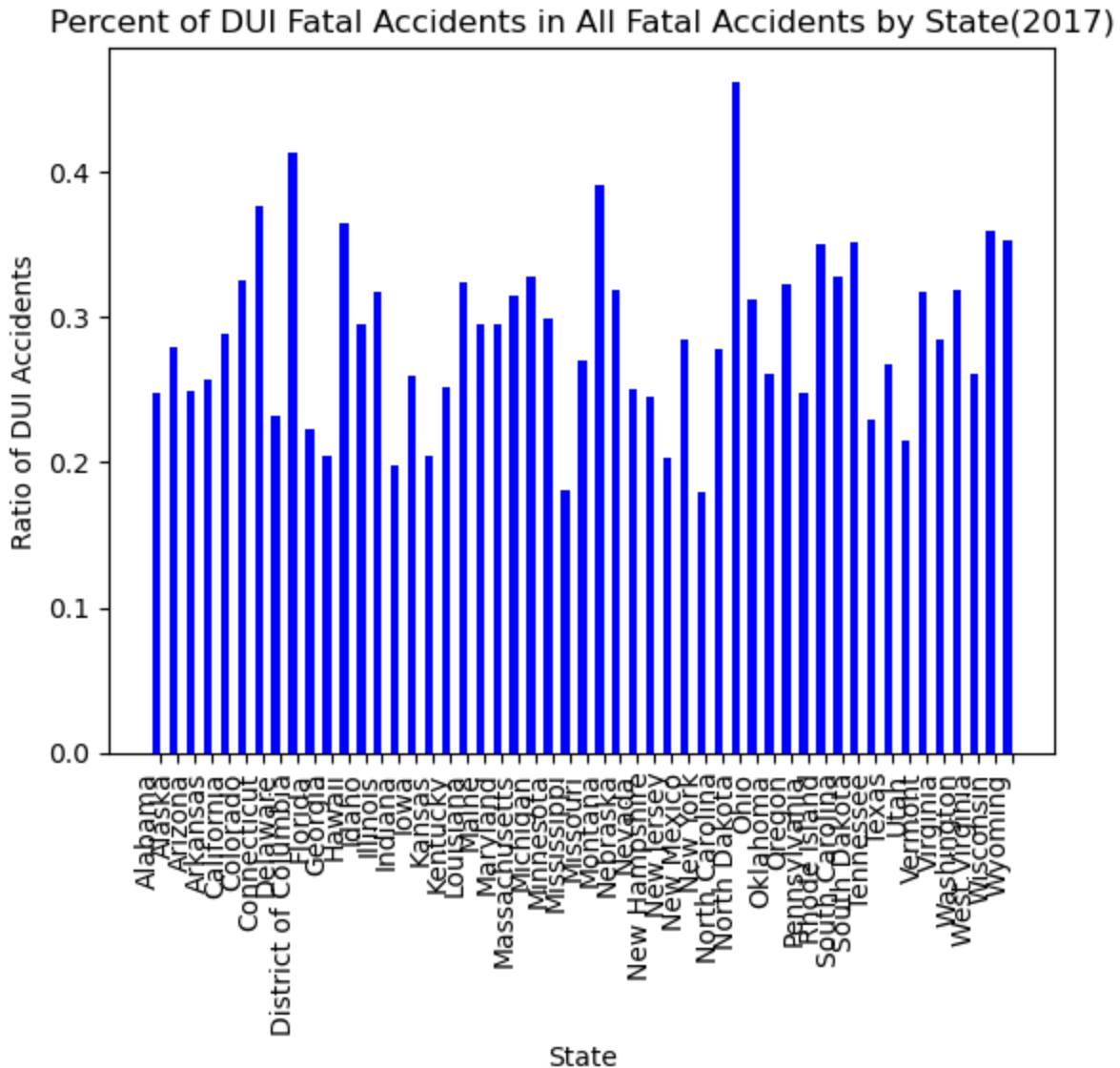
The first figure shows the DUI rate in fatal accidents by state in the year 2018. We can see this result is independent from the number of people we have in that state because it is the ratio to the state's total population. The second figure is the DUI rate change from 2015 to 2020 in 5 selected states, we can see a clear pattern for each state and we can see a significant difference between the southern and northern states. We do not know why we get this result, this is what I will investigate by analyzing other variables in my data set.



The second figure is the DUI rate change from 2015 to 2020 in 5 selected states, we can see a clear pattern for each state and we can see a significant difference between the southern and northern states. In order to investigate more about this result I have done DUI rate change for 10 cites.



We can conclude that northern states have much higher dui fatal rates than southern states. This finding is very important because we can ask why northern states have much higher DUI fatalities than southern states, what can be changed to improve accidents? Asking the right question/making the right hypothesis will save thousands of peoples lives each year. In the beginning I tried to use t-test to show that there is a significant difference between the rate of northern and southern states. I did not use hypothesis testing for it because the rate data set is very small and the concept from northern to southern is geographically continuous. We can see a trend of decreasing from all states in 2017, so I decided to observe this data again by counting the in 2017:



Each state is much lower than the other years of the same state. We can investigate what policy was published that year or what factors of the traffic environment have changed in that year.

## 5. Future Work

From the result and visualizations, I made several bold assumptions where needs to be testified using more data. I would like to use weather data.

1. The Northern state has a higher DUI fatal rate because the temperature is very low and people waiting for rescue freeze to death.
2. The sudden drop of DUI rate in 2017 is because the temperature is much warmer than other years.

If given more time, I will just follow the original goal and analyze the cause of fatal accidents from a huge number of factors, and give each factor a weight of influence. At the end of day we

want to create a prediction model where we know how to deploy a police force and notify city planning about the traffic issues of certain areas.