# Analysis of U.S. Car Accident

Team Members: Tianchen Wang, Zitao Cheng, Yu Li

# Questions needed to study

1. Whether is the car accident related to day or night
2. Whether time period(day, month, year) would affect the car accident rate?
3. What range of visibility would cause car accidents more frequently
4. Where is the most likely place to have a car accident(county-level)?
5. The relationship is between road types and car accident types
6. What are the influence factors on the U.S. car accident rate?

# List of Tools

- Pandas
- Numpy
- Plotly
- Jupyter Notebook
- Microsoft Excel



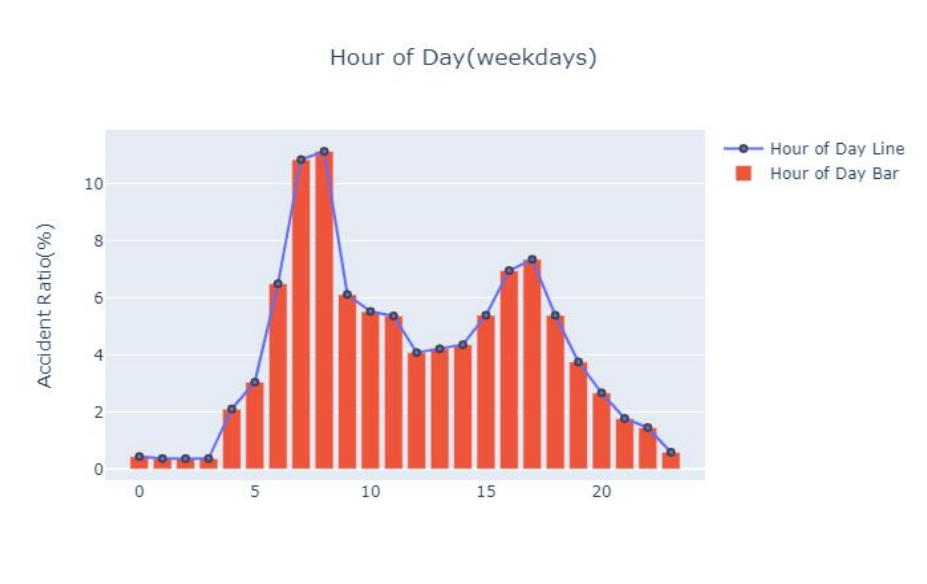


# Main Tasks - Data Cleaning and Selection

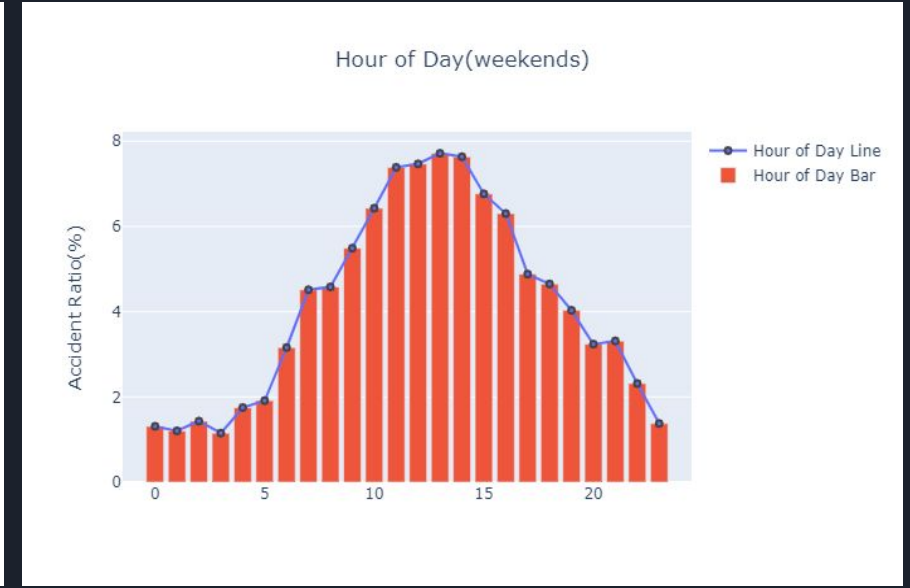
- Data Cleaning
  - Remove useless attributes
    - Example: “End_Lat”, “End_Lng”
  - Fill a mean or mode value in missing place intuitively
- Data Selection
  - Select useful attributes when generating specific graphs
    - Example: “Start_Time”

# Main Tasks - Graph Analysis

- Temporal




The U.S. accident rate is happening each hour of the day on the weekday.

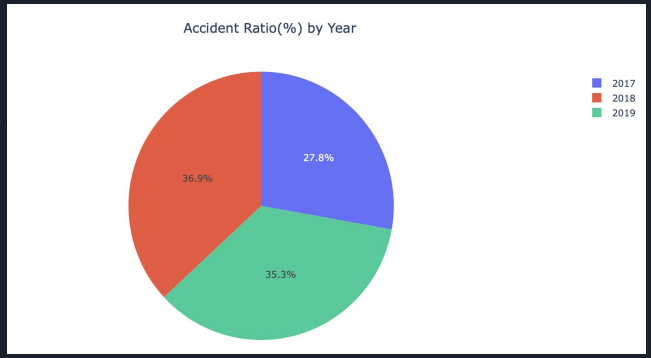



The U.S. accident rate is happening each hour of the day at the weekend.

# Main Tasks - Graph Analysis

- Temporal


The U.S. accident rate is happening each day of the week from Monday to Sunday.

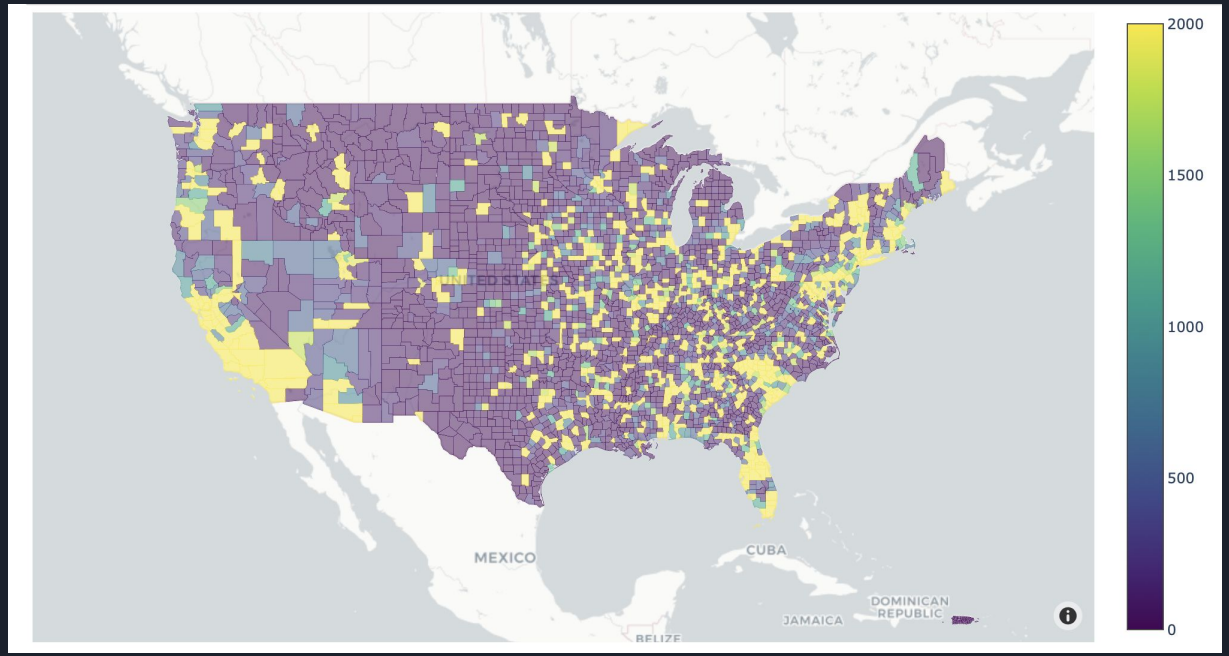



Total U.S. accident counts in 2017, 2018, and 2019. All of these years have approximately the same U.S. car accident proportion


Monthly U.S. accidents that happened in 2017, 2018, and 2019.

# Main Tasks - Graph Analysis

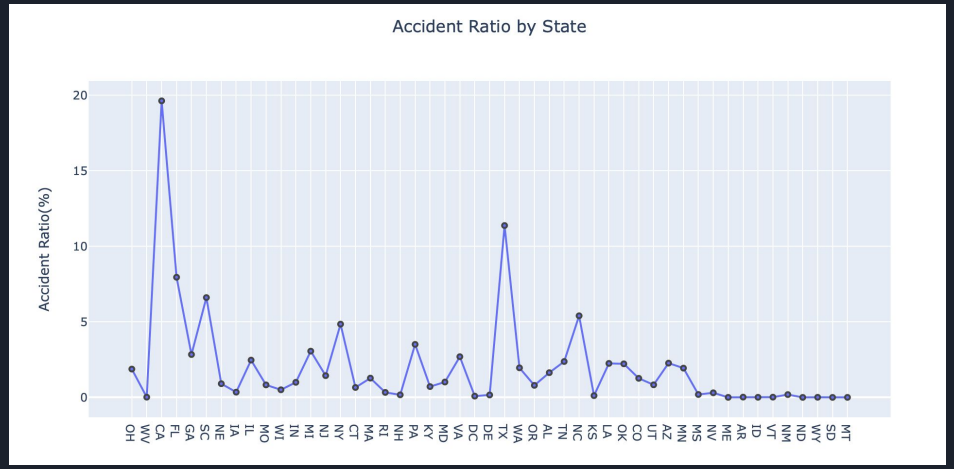
- Geospatial - map chart



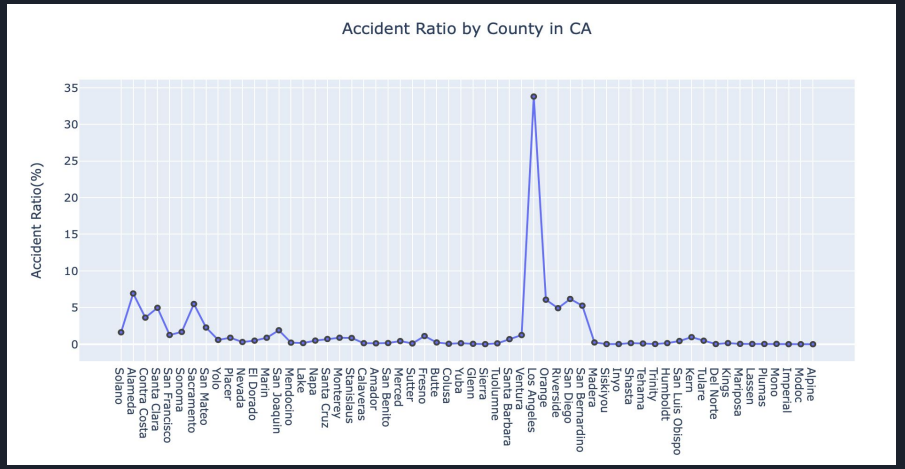

The U.S. accidents count in different counties. (Alaska, Hawaii, and Puerto Rico are not included in the map)

# Main Tasks - Graph Analysis

- Geospatial - line chart



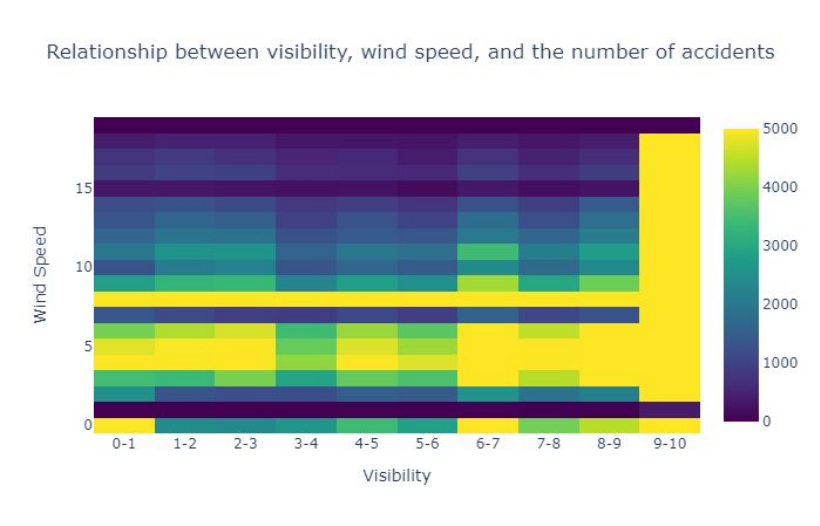

The U.S. accident rate in different states in a line chart (The state of California has the most severe car accidents ratio)




The accident rate in California as a county Level which Los Angeles has the highest accident ratio

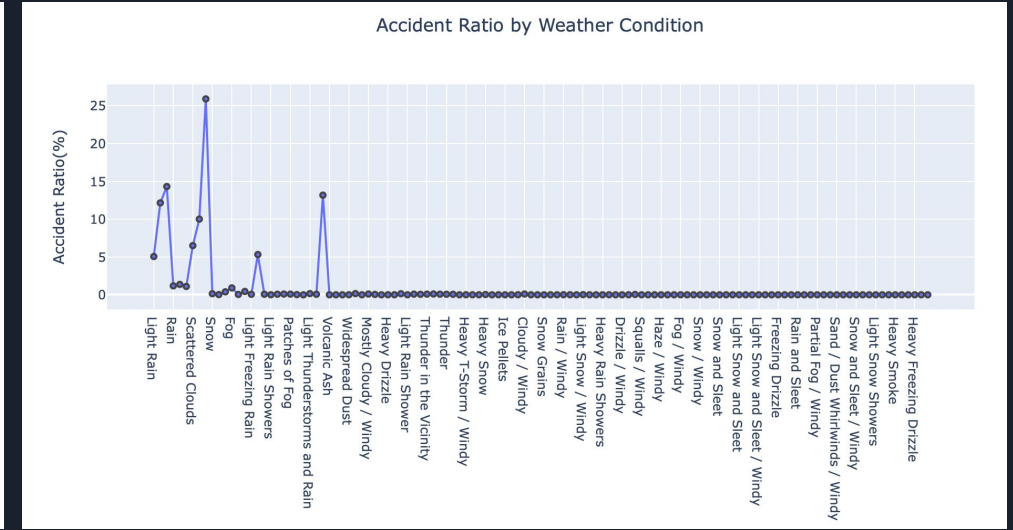
# Main Tasks - Graph Analysis

- Visibility and Wind Speed




The accident rate happening in different wind speed and visibility conditions.

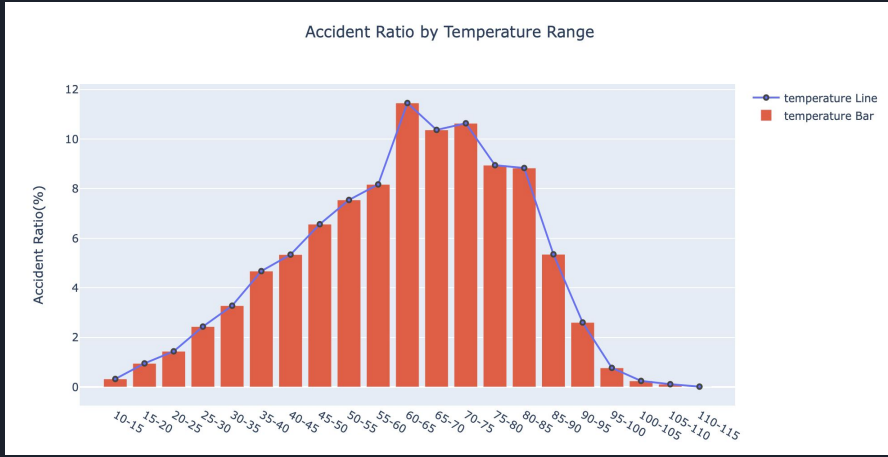
- Weather Conditions



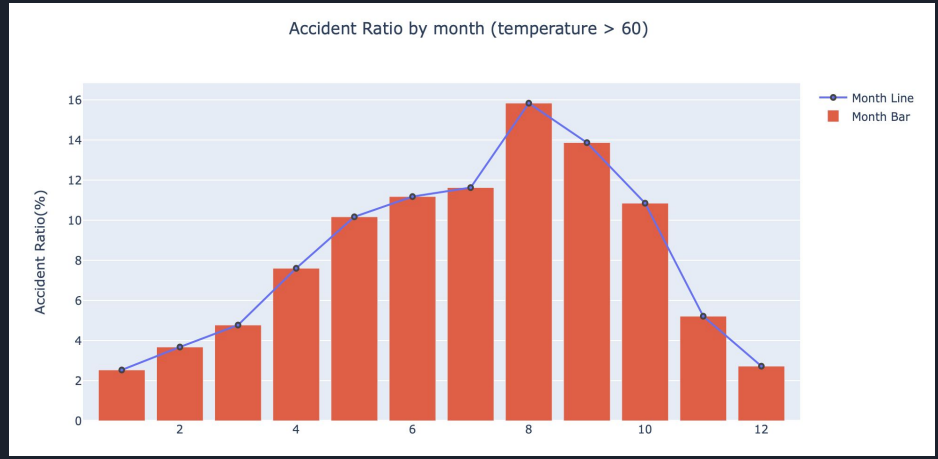

The U.S. car accident rate depends on different weather conditions. Good weather conditions have a higher accident ratio.

# Main Tasks - Graph Analysis

- Temperature Range




The U.S. car accident rate based on temperature range, which approximately appears as a normal distribution




The U.S. car accident rate distributed only the temperature is above 60℉

# Discussion Points

Good Point

- Normalization
- Question-based research

Bad Point

- Evaluation

What we do in the future ?

- Produce a mathematical model to try to predict the U.S. car accident ratio in the future

# Key Results

1. Drivers should drive more carefully during weekdays' rush hour and weekends' noon.
2. Since there is a lot of traffic in metropolises, drivers should pay more attention to cars' distance to avoid unnecessary traffic accidents.
3. In densely populated cities,  drivers should focus more on road conditions to reduce car accidents.
4. Traffic lights are the most critical parts in the traffic, pay close attention to the traffic lights, especially at the junctions and crossings.
5. Los Angeles tend to be the highest car accident ratio as county level in the United States.
6. During fair days and clear days, people tend to have more outdoor activities. Hence there will be more traffic jams, and drivers should drive more carefully.
7. In summer, high temperatures will cause drivers' inattention; people should avoid driving in such conditions.