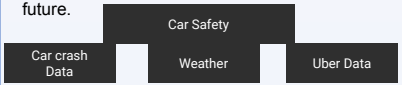
Safety of Ride-sharing Industry could be Predicted Using Car Crash Analytics

Danting Huang, Jinyan Yu, Yanyue Fu, Kai Sun, Aparna Raman

UBER

Abstract

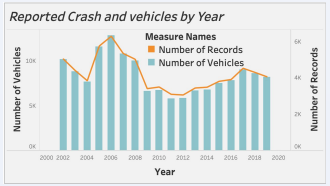
Our project yearns to solve the problem of road safety and mapping some of the characteristics most common with crashes and road fatalities. Our project will set the stage for present and future year car sharing industry to make its car rides safer for everyone and increase the consumer base in the future.



Our Dataset

Uber dataset consists of 330568 rows and 57 variables with data divided into 5 categories: time, location, car type, price and weather. It was collected in 2018 for three weeks in Nov - Dec'18 and at a regular interval of 5 mins.

Boston car crash data is from 2002 to 2020 Feb, and consists of 101694 rows and 25 variables can be divided between 4 categories: crash time, crash location, type of car and severity of crash.



Objectives

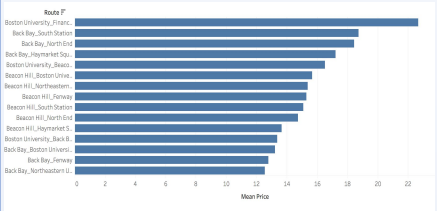
- Determine the factors that most affect the way crashes occur and their intensity
- Predict the number of crashes for a 60 day period
- The effect of weather on uber rides and car crashes
- In the end we would want to give recommendations to ridesharing industry (Uber) of how to make their rides safer

Exploratory Analysis

Route analysis

By checking Boston downtown famous places, we make 12 starting points and 12 destinations connected to get in total 72 recorded routines.

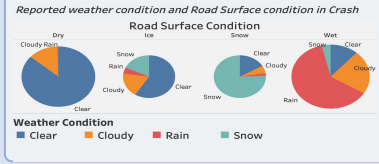
- We assume students are frugal and call Uber pools most while there is no significant difference among various type called by university-route
- Mean price of university-route is even higher
- BU to financial district is the most costly routes



Weather analysis

Weather condition broken down by road surface condition. Most Car crash happened during four kinds of road surface: Dry, Ice, Snow and Wet, and different weather conditions cause those road surface. In the following pie chart, color shows details about weather condition. Size shows percentile of number of records. The results show:

- Most car accidents happened on wet and dry road surface
- Rain was the main cause of the car crash



In order to explore the specific of rain on car crash, different rainy level days were defined by subdividing the variable of humidity. We find:

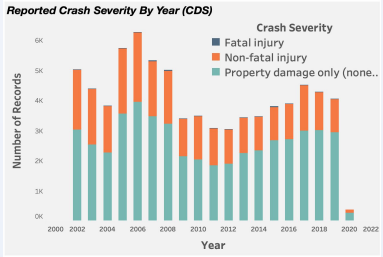
- Most car crash happened on light rain, and most of them caused non-fatal injury and property damage only.
- 8 am, 1 pm and 4 pm are the most likely time for a car accident on light rain.

Except humidity, we want to figure out what other variables may affect car crash by combining Uber and Crash dataset. We finally find that these three important variables affect car accidents:

- Apparent Temperature
- Dew Point
- Wind Gust

Collision Analysis

In brief, over half of the collision were mild and 35% of them would cause non-fatal injury and only 0.4% cause fatal injury. The amount of non-fatal injury is decreasing by year.



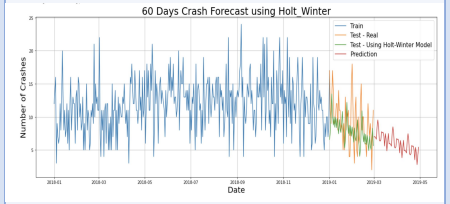
- 83.6% of people died because of 'single vehicle crash' and 'angle crash'.
- Over 40% of people got hurt due to rear-end crash and 25% caused by angle crash.
- For property damage only, the top three collision manners are read-end, angle and sideswipe (with same direction).

Results

The regression results shows that the following factors has significant impact on crash severity:

- Number of vehicles
- Average temperature
- Average humidity
- Average Visibility
- Manner of collision
- Road Surface Condition

60 Days Crash Prediction



Conclusion

Below measures could be implemented in Uber and other ride-sharing companies:

- Drivers will receive notifications when passing through areas with high crash risk.
- Uber will reroute upon drivers' and passengers' request based on the volume and price.
- As weather could affect road safety, Uber could increase price accordingly based on weather and road surface conditions.

Please take a look at our github repository for more info