Can Ride Sharing Industry Make Rides Safer If They Knew More About Crash Analytics?

Team1: Aparna Raman, Danting Huang, Kai Sun, Jinyan Yu, Yanyue Fu

UBER

Our Schedule

09/2019	Set project goals to predict WTP and demand for Uber and Lyft
10/2019 - 12/2019	Gathered data from Uber directly and also from various resources
01/2020	Performed exploratory analysis, and realized the research limitation for demand forecast
02/2020	Adjusted our research direction to focus on ride safety by combining Uber and Car crash datasets
02/2020 - now	Performed exploratory analysis, regression and crash prediction

Business Problem

Our service provider is faced with a major problem pertaining to keeping their customer base. A current series of road accidents have left the Boston customer base perplexed and unsure about trusting strangers with their lives on the road.

Dataset

1. Uber Dataset : Kaggle
2. 2002-2020 Boston vehicle crash data : Mass Gov

Dataset Overview

Uber Data (Kaggle):

- 330568 data points for 3 weeks of data- November 2018 to December 2018
- 57 Variables divided into 5 categories: time, location, car type, price and weather situation
- Half of the variables are associated with weather, combined with date and route information to analyze price and location

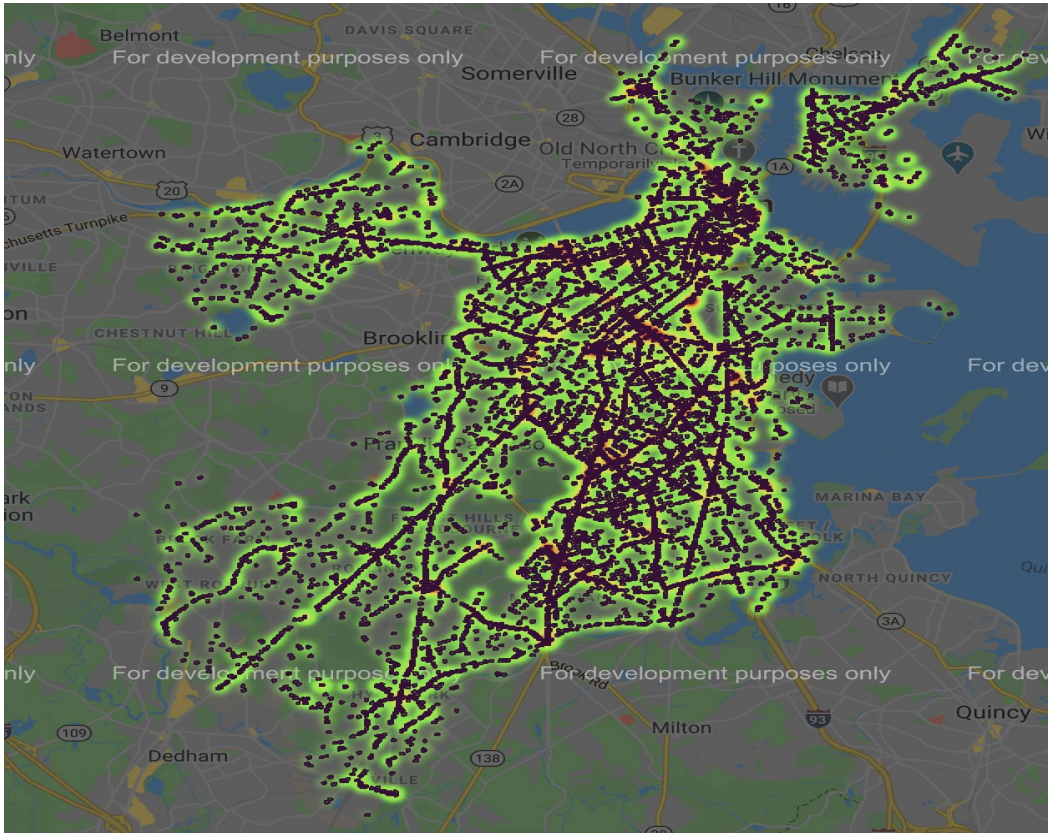
Boston Crash Data (Mass gov):

- 101694 data points for Boston area car crash data for the period 2002-2020
- 25 variables divided into 4 categories: crash time, crash location, types and severity of car crashes.
- Data being used to assess the number of crashes in Boston downtown and the different factors affecting crash such as weather, type of car, fatality based on manner of collisions etc.

Hypotheses Pertaining to our Business Problem

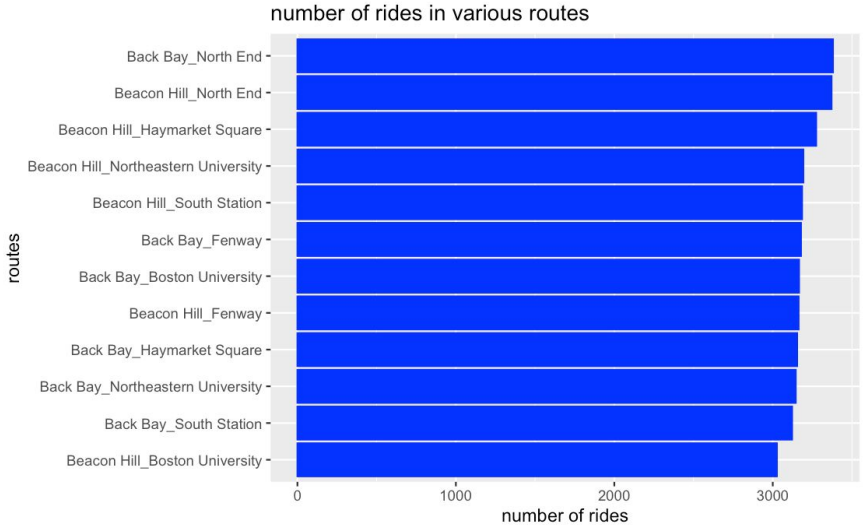
- Determine the factors that most affect the way crashes occur and their intensity.
- Perform exploratory data analysis on different routes in Uber rides
- Analyse weather and road surface conditions which might affect crash rate
- Predict the number of crashes for a 60 day period

An Overview of car crash locations in Boston area



- Centralized on the city center (most)
- Centralized near the airport (few)
- Scattered around the city

Uber Route Analysis: Busiest Routes



school <dbl>	price <dbl>
0	15.03809
1	17.31145

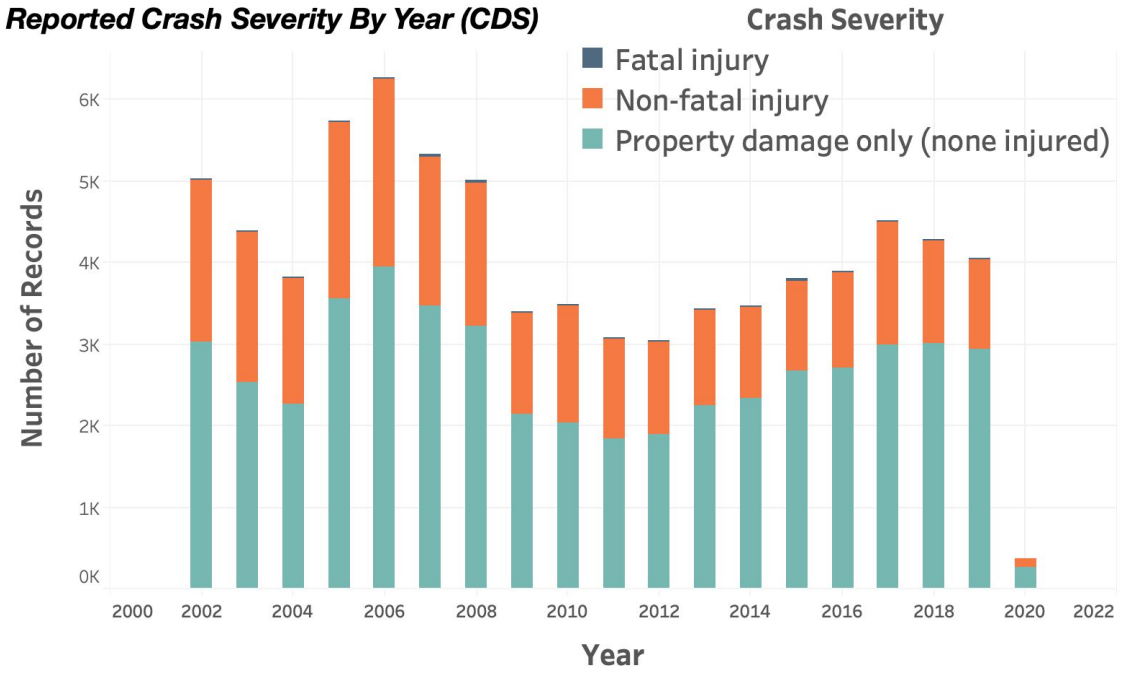
1）12 famous starting points and 12 destinations >>72 routines recorded in Boston downtown (Show the top ones）

2）check university-related route containing BU, NEU and calculate average price per ride and number of rides based on university routes and other routes

Uber Route Analysis: Priciest Routes

check the mean price based on various routes >> BU to the financial district most costly

Collision Analysis: Overview of Crash



- Over past 20 years, 64.3% of the crash was mild, 35.3% of them caused non-fatal injury and only 0.4% leaded to fatal injury.
- Before 2009, the number of vehicle accidents was really high.
- After 2010, the crashes caused property damage only have increased significantly, whereas the crashes caused non-fatal injuries were on a decline.

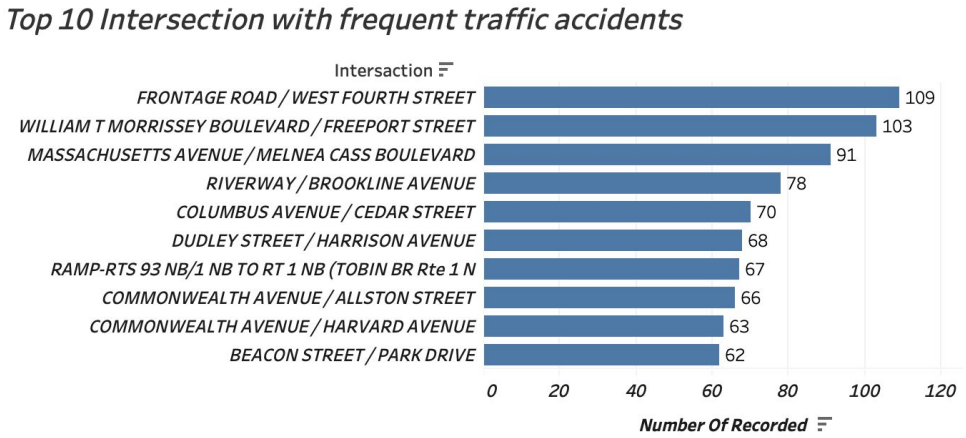
Collision Analysis: Find the most dangerous collision

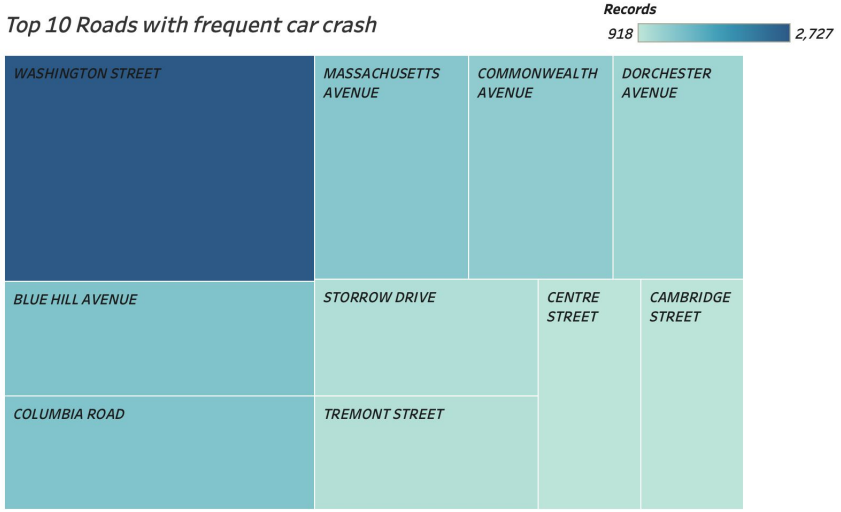
Average Severity Score for Manner of Collision

Manner of Collision	
Head-on	2.5800
Single vehicle crash	2.5120
Angle	2.4110
Rear-end	2.4020
Front to Front	2.1500
Sideswipe, opposite direc..	2.0910
Front to Rear	2.0730
Rear-to-rear	1.9620
Sideswipe, same direction	1.9540
Rear to Side	1.9230

- Designed an ***Average Severity Score System*** based on “KABCO Injury Classification Scale and Definitions”.
- Concluded that **Head-on Collision** is the most dangerous type of car accident.
- Common causes of Head-on Collision: ***Distracted driving, Fatigue, Reckless driving, Speeding***

Collision Analysis: Find the most dangerous road





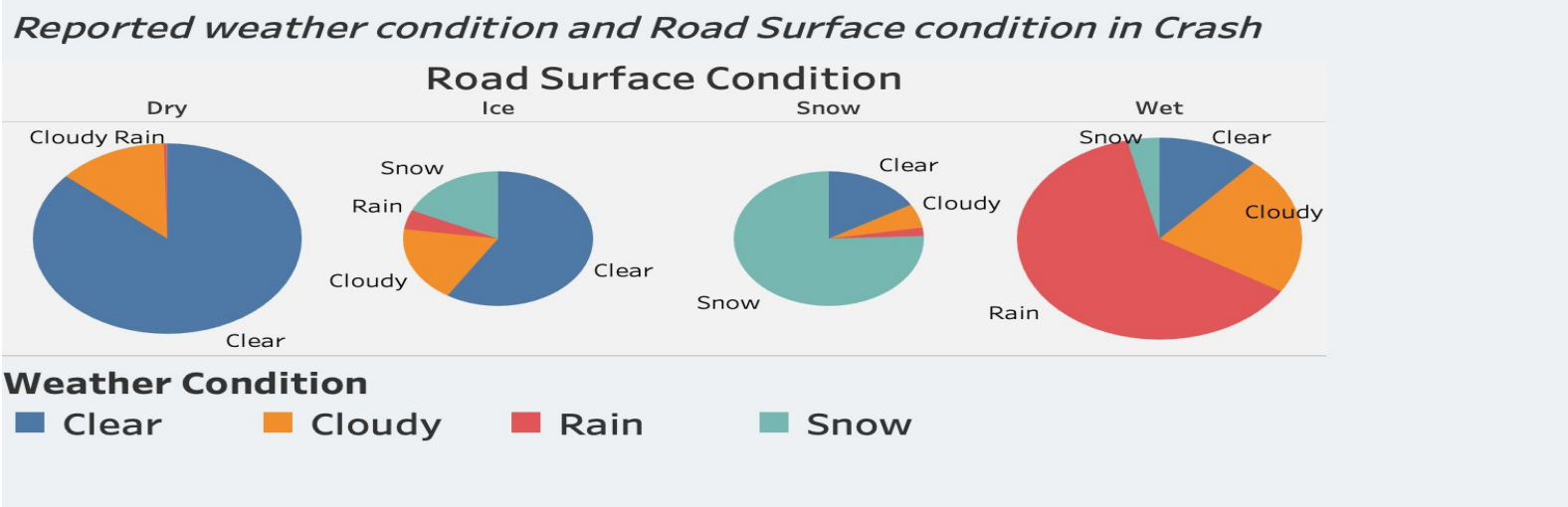
Top three roads:

- Washington St (2727 crashes)
- Blue Hill Ave (1375 crashes)
- Columbia Rd. (1368 crashes)

Tips for Uber Driver:

- Know the speed limit
- Focus on the road
- Don’t to be overconfident
- Stay stone cold sober

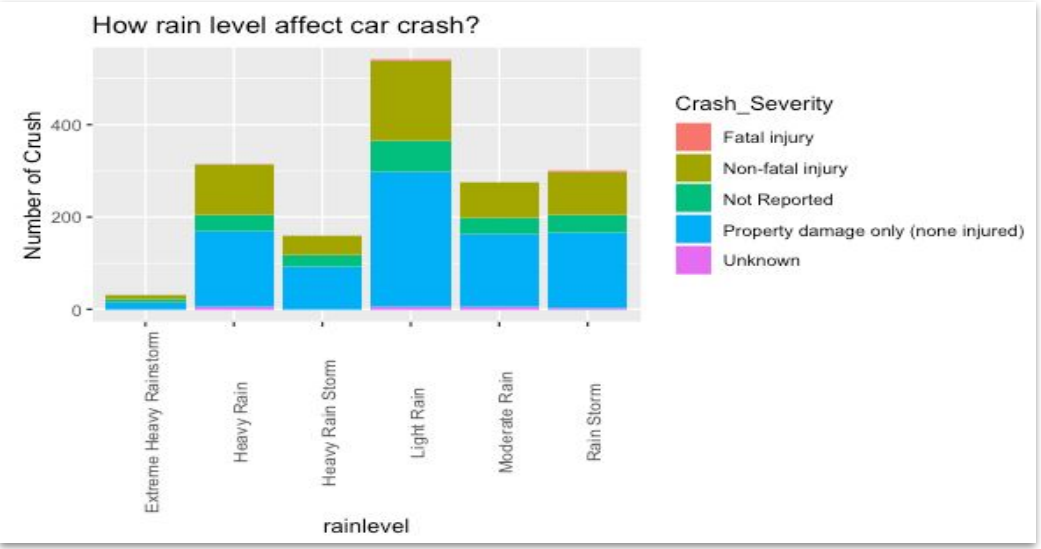
Weather Affecting Car Crash

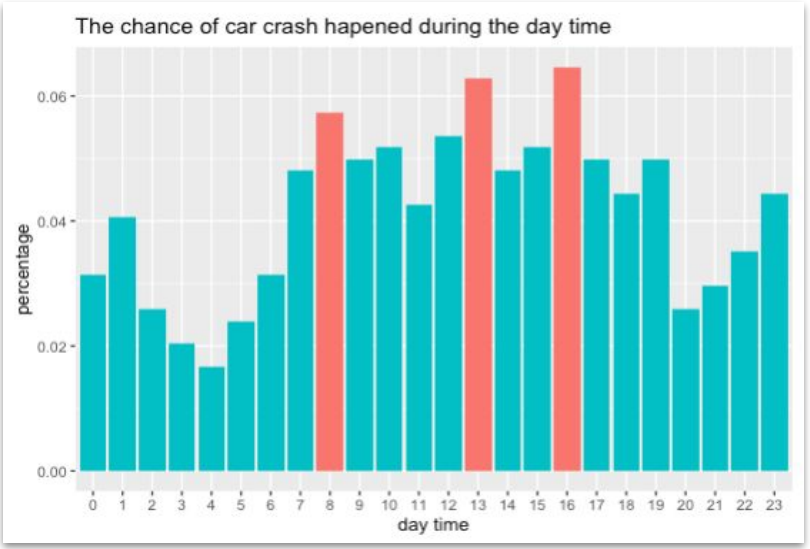


- Most Car accidents happened on the dry and wet road surface
- Plan to choose one important weather variable precipitation to specifically define rain level for 7 situations: No Rain, Drizzle, Light Rain, Moderate Rain, Heavy Rain, Rain Storm, Heavy Rain Storm and Extreme Heavy Rainstorm

Weather Affecting Car Crash

- Most car accidents happened during the light rain caused property damage only.
- 8am, 1pm and 4pm during the light rain has more chance to cause car crash.





- However, we got different important variables which affect car crash after combing Uber and Crash dataset: **Apparent Temperature, Dew Point and Wind Gust.**

Regression Result

OLS Regression Results

Dep. Variable:	Crash_Severity	**R-squared (uncentered):**	0.933
Model:	OLS	**Adj. R-squared (uncentered):**	0.932
Method:	Least Squares	**F-statistic:**	3718.
Date:	Wed, 15 Apr 2020	**Prob (F-statistic):**	0.00
Time:	15:22:37	**Log-Likelihood:**	-4374.7
No. Observations:	4582	**AIC:**	8783.
Df Residuals:	4565	**BIC:**	8893.
Df Model:	17		
Covariance Type:	nonrobust		

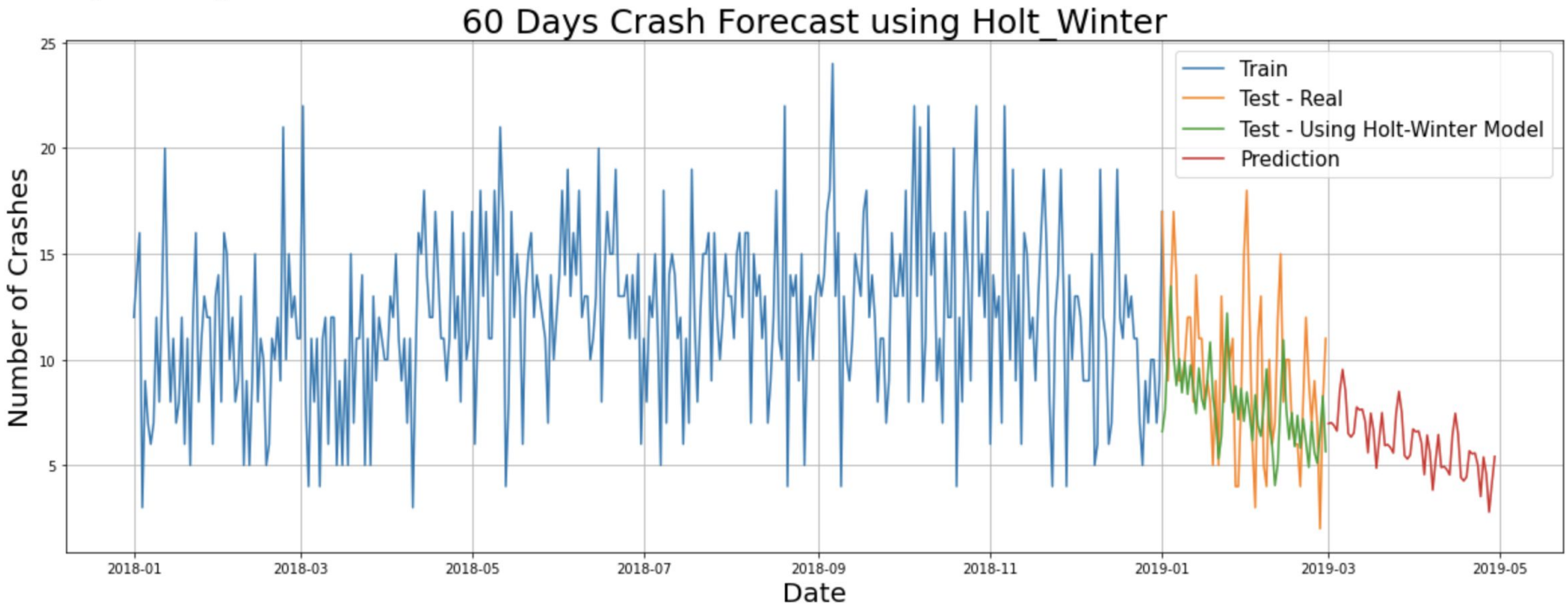
	coef	std err	t	P>|t|	[0.025	0.975]
Manner_of_Collision	-0.0019	0.003	-0.679	0.497	-0.007	0.004
Most_Harmful_Events	-0.0018	0.000	-8.073	0.000	-0.002	-0.001
Vehicle_Configuration	0.0001	0.000	0.680	0.496	-0.000	0.001
Road_Surface_Condition	0.0055	0.004	1.502	0.133	-0.002	0.013
Ambient_Light	-0.0054	0.005	-1.109	0.267	-0.015	0.004
Weather_Condition	-0.0022	0.001	-1.735	0.083	-0.005	0.000
At_Roadway_Intersection	7.286e-05	2.52e-05	2.894	0.004	2.35e-05	0.000
Distance_From_Nearest_Exit	6.448e-05	0.000	0.285	0.775	-0.000	0.001
Distance_From_Nearest_Landmark	-4.879e-06	4.67e-05	-0.105	0.917	-9.64e-05	8.66e-05
Number_of_Vehicles	0.1185	0.014	8.670	0.000	0.092	0.145
Total_Nonfatal_Injuries	-0.8423	0.012	-72.865	0.000	-0.865	-0.820
avg_temp_(f)	0.0357	0.003	13.773	0.000	0.031	0.041
avg_dew_point_(f)	-0.0394	0.003	-14.275	0.000	-0.045	-0.034
avg_humidity_(%)	0.0245	0.001	23.725	0.000	0.023	0.027
avg_visibility_(mi)	0.0461	0.006	8.205	0.000	0.035	0.057
avg_wind_(mph)	0.0065	0.004	1.686	0.092	-0.001	0.014
high_wind_gust_(mph)	0.0016	0.002	0.993	0.321	-0.002	0.005

Omnibus:	673.249	**Durbin-Watson:**	1.925
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	4999.168
Skew:	0.484	**Prob(JB):**	0.00
Kurtosis:	8.025	**Cond. No.**	2.04e+03

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

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

60 Days Crash Forecast



Conclusion

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


Q&A