Team members have drafted their project, including the following:

✓ Selected topic

✓ Reason why they selected their topic

✓ Description of their source of data

✓ Questions they hope to answer with the data

Team members present a provisional machine learning model that stands in for the final machine learning model and accomplishes the following:

✓ Takes in data in from the provisional database

✓ Outputs label(s) for input data

----

The Rise of Traffic Fatalities During the Pandemic and its Relation to the Price of Gas

According to multiple reports from U.S. Department of Transportation’s National Highway Traffic Safety Administration, there was a rise in traffic fatalities during the pandemic. While the NHTSA attributes this rise as positively correlated to the rise of DUI/DWI’s, lack of seatbelt usage, and speeding, our team is curious to find out how the price of gasoline, which notably fell to $1.841 a gallon during May 2020 of the pandemic, may also be correlated. We also hope to see if the subsequent historic rise of gasoline prices shows a negative relationship with traffic fatalities.

We are sourcing our data related to the traffic fatalities directly National Highway Traffic Safety Administration. We are sourcing our data related to the average U.S. price of gasoline from the U.S. Energy Information Administration.

Dependent Variable:

Whether a traffic incident resulted in a fatality.

Independent Variables:

* U.S. Average price of gasoline (dollars per gallon)
* Failure to require restraint use by self or passengers (0 or 1)
* Whether or not there was impaired driving due to the use of alcohol, drugs, or prescription medication (0 or 1)
* Whether or not there was involvement of a vehicle travelling at an illegal high speed (0 or 1)
* Weather
* Day of the Week
* Light level

<https://www.nhtsa.gov/press-releases/2020-fatality-data-show-increased-traffic-fatalities-during-pandemic>

<https://www.nhtsa.gov/press-releases/early-estimate-2021-traffic-fatalities>

Data exploration phase:

While pondering the viability of our chosen topic, the group worked together using data bases such as INSERT\_GOOGLE\_DATA\_SITE\_HERE and Statistica to find look for legitimate sources. We knew theoretically that we needed data regarding the price of gas and traffic accidents that ended in fatalities, as well as the variables mentioned in the original National Highway Traffic Safety Administration (NHTSA) press release (violations regarding seat belt usage, DUI/DWI, and excessive speeding. We ultimately sourced our data directly from the NHTSA and the U.S. Energy Information Administration (EIA).

The XLS file provided by the EIA was simple and straightforward, while the zip file of CSV files had our data scattered about across different sheets. Dissecting the files together, we managed to find the variables we looked work with together and highlighted which columns of data we want to merge later in the project.

Data analysis:

Our machine learning model is a logarithmic regression to estimate the relationship between our chosen variables (month, day of the week, gas price, light level, weather condition, rural/urban, age of driver, alcohol violation, seat belt violation, speeding violation, and whether an accident resulted in a fatality). Using the logarithmic regression, R can generate summary statistics that will highlight which of our variables were the most significant.

Ultimately, we want to see if gas prices give any explanatory value in our model. We can also measure the probability that each variable may result in a fatal car accident given the presence of a certain violation and/or circumstance.

=IF(U2="January",2.548,IF(U2="February",2.442,IF(U2="March",2.234,IF(U2="April",1.841,IF(U2="May",1.87,IF(U2="June",2.082,IF(U2="July",2.183,IF(U2="August",2.182,IF(U2="September",2.183,IF(U2="October",2.158,IF(U2="November",2.108,IF(U2="December",2.195,0))))))))))))

E column = VIOLATION DESC

Q Column = DUI\_DWI = =IF(OR(E2="Alcohol, drug, or impairment violations generally", E2="Driving under influence of substance not intended to intoxicate", E2="Driving while intoxicated (alcohol or drugs) or BAC above limit (any detectable BAC for CDLs)", E2="Driving with detectable alcohol"), 1, 0)

T column = Seatbelt\_Vio = =IF(E2="Failure to require restraint use (by self or passengers)",1,0)

U Column = Speeding\_Vio = =IF(OR(E2="Speed greater than reasonable & prudent (not necessarily over the limit)",E2="Speed related violations, generally", E2="Speeding (above the speed limit)"),1,0)