

CUNY School of Professional Studies
MS DATA ANALYTICS
Advanced Programming Techniques DATA 602
Spring 2017 Project Part 1

Ahsanul Choudhury

Project proposal

Data source: <https://github.com/fivethirtyeight/data/blob/master/bad-drivers/bad-drivers.csv>

The dataset contains data on number of drivers involved in fatal collisions per billion miles, different conditions under those collisions happened, car insurance premium in U.S. dollars and losses incurred by insurance companies for collisions per insured driver in U.S. dollars for all 50 states of United States of America and District of Columbia. I have used this dataset for DATA 606 final project in Fall 2016. I picked two variables from the dataset and tried to predict average car insurance premium using those two variables and compare them with actual data using *R* which included analysis and visualization. For the DATA 602 Spring 2017 project I am proposing to do the same using *Python*.