Identifying Drunk Driving in Fatal Car Crashes

By Skyler Lehto

Motivation

- 32,000 fatal car crashes in 2015
- 27 % involved drunk driving
- Department has limited resources to investigate
- Which crashes most likely involved drunk driving?

Process

Data

- Used 2015 data from NHTSA
- Merged compatible variables from datasets

Model

- Trained and tested several model types
- Selected logistic regression as best model

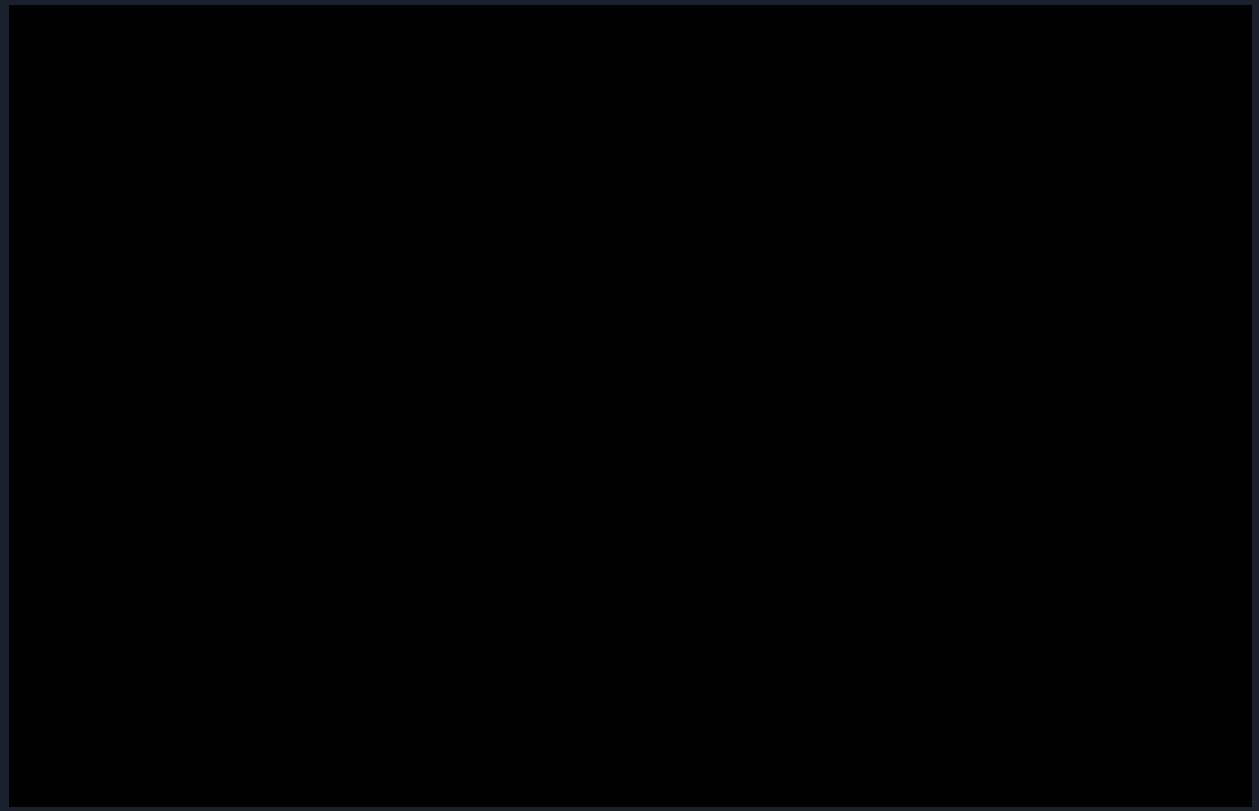
Analysis

- Identified most useful variables
- Visualized model probabilities in Tableau

Key Findings

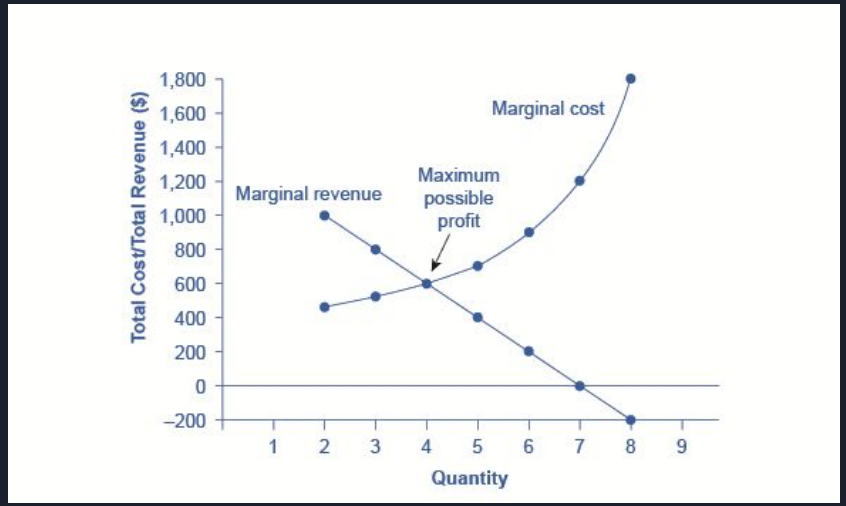
- Time of day is highly predictive
  - 10 pm - 4 am more likely
- Day of week also predictive
  - Saturday and Sunday more likely
- Other variables are less important

Model Predictions (Visualized)



What Would a Business Do?

- Marginal cost is increasing
- Marginal revenue is decreasing
- Where does MR=MC ?



Recommendations

- Use a 20% probability threshold for investigation
  - Heuristic: Social benefit of identifying drunk driving is ~5x cost of investigation
  - Contingent on resources in budget