

Determining the cause and predicting the severity of traffic related accidents in Michigan.

Team Members:

Ameen Alkhabbaz – MSDSBA 1st Year

Jordan Burton – PhD 4th Year

George Mckenzie – MSDSBA 1st Year

David Emery – MSDSBA 1st year

Mahzad Zavareh – MSDSBA 1st Year

Background, Importance, Data sets, etc.

Traffic accident information is important for insurance companies to price competitive insurance quotes to consumers, for ride-hailing services like Lyft or Uber to avoid dangerous routes, and for emergency response teams to predict the occurrence of an accident. Our group decided to subset the Kaggle datasets US_Accidents_Dec19.csv¹ and 2017_county_data.csv² to specifically analyze traffic accidents throughout Michigan from 2016 – 2019. The accident data contains information about each accident recorded over that time span including accident severity, location, weather conditions, and time. The 2017 county data contains predicted census information for each county in Michigan and includes employment categories, as well as, population and diversity statistics.

How did we come up with and agree on the topic? Are we all in agreement with the topic and scope of the proposal?

Our group explored several different datasets searched by each group member to determine what sparked our interest and what held some commodifiable data. We decided to use the accident and census data as there are many predictors and response terms that we can probe to build visualizations and predict accidents. We are all in agreement with the topic and the scope of the proposal.

5+ questions to be investigated/answered via the project work?

- Is there a correlation between accident severity and other predictors?
- Are there patterns in the accident data?
- Can we determine what causes accidents in a specific location?
- Is prediction of the time of day and location of an accident possible?
- Can we build a visualization tool to allow intuitive interpretations of accident data?
- Are population factors correlated with accidents?
- If changing an existing road, how will it affect the number of accidents?

Keywords of methodologies/tools/frameworks that we are planning to apply to investigate the proposed problems.

- Divide data into training and test data for classification analysis
- Machine learning, multiple regression analysis
- Feature engineering to train model, select variables from data set, create new feature by gathering census data
- Correlation in error terms, as this is timeseries data
- Visual mapping (maps, shinyR)

¹ Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. "A Countrywide Traffic Accident Dataset.", 2019.

² US Census Bureau