IBM DATA SCIENCE CAPSTONE PROJECT

Philip P.

October 2020

BUSINESS CASE

- What set of **stochastic** and **non-stochastic** variables most strongly correlate with car accident severity?

- What urban design principles can be gleaned from car accident data to reduce accident severity?

KEY VARIABLES

- **Stochastic**: weather, road condition, light condition
- **Non-stochastic**: crosswalks, sidewalks, junction type

DATA DESCRIPTION

- Approx. 20,000 car accidents in Seattle from 2004 to present, collected and managed by the Seattle Department of Transportation
  - Sufficiently large, diverse; numerical and categorical
  - Official source and curation obviates issues with ground truthing
  - Publicly available

DATA PREPARATION

- Several fields may be dropped as irrelevant to study:
  - X and Y coordinates, street address, date of accident, etc
- Relevant categorical variables are assigned floating integer values to accommodate statistical inference:
  - Road condition, light condition, weather

METHODOLOGY

- Multiple linear regressions to establish causality and correlations between multiple variables and target
  - Data segregated into train and test sets with appropriate degree of randomness
  - Variance score and confusion matrix computed to ascertain model fit and explanability
- Logistic regression performed to classify categorical variables for predictive analysis of accident severity based
  - Similar train and test sets, split anew to avoid overfit
  - F1 score used to determine model accuracy

RESULTS

- Low correlations for stochastic variables as measured by accuracy tests
  - Conversion from categorical variables to floats introduced degree of noise into the model
  - No one stochastic variable offers in-depth explanability or correlation to overall accident severity

- Moderately higher correlation for non-stochastic variables as measured by accuracy tests
  - Junction type and address type show greater correlation to accident severity

DISCUSSION

- Categorical nature of data made hypothesis testing more difficult, prone to errors, and undermine correlative potential

- Follow-on research might focus more narrowly on street design elements and car accident severity
  - Street design elements appeared to have greater correlation to accident severity, despite drawbacks of experimental approach
  - More generally, one can control for street design elements…one cannot control the weather

CONCLUSION

- Street design considerations may reduce the severity of car accidents within a metropolitan setting. Principal elements include junction type and address type (block or intersection, for example)

- Additional data and analysis is necessary to fully explicate causal and correlative relationships between urban design elements and accident severity, given limitations of initial data pool