Capstone Project Inc.

Addressing Safety at Seattle Intersections

A Machine Learning Approach

The Opportunity

- Seattle has captured extensive collision data for several decades
- Using this data and machine learning models, it is possible to identify the most important factors in predicting collision severity
- These factors can then be used by traffic planners and engineers to increase safety at intersections

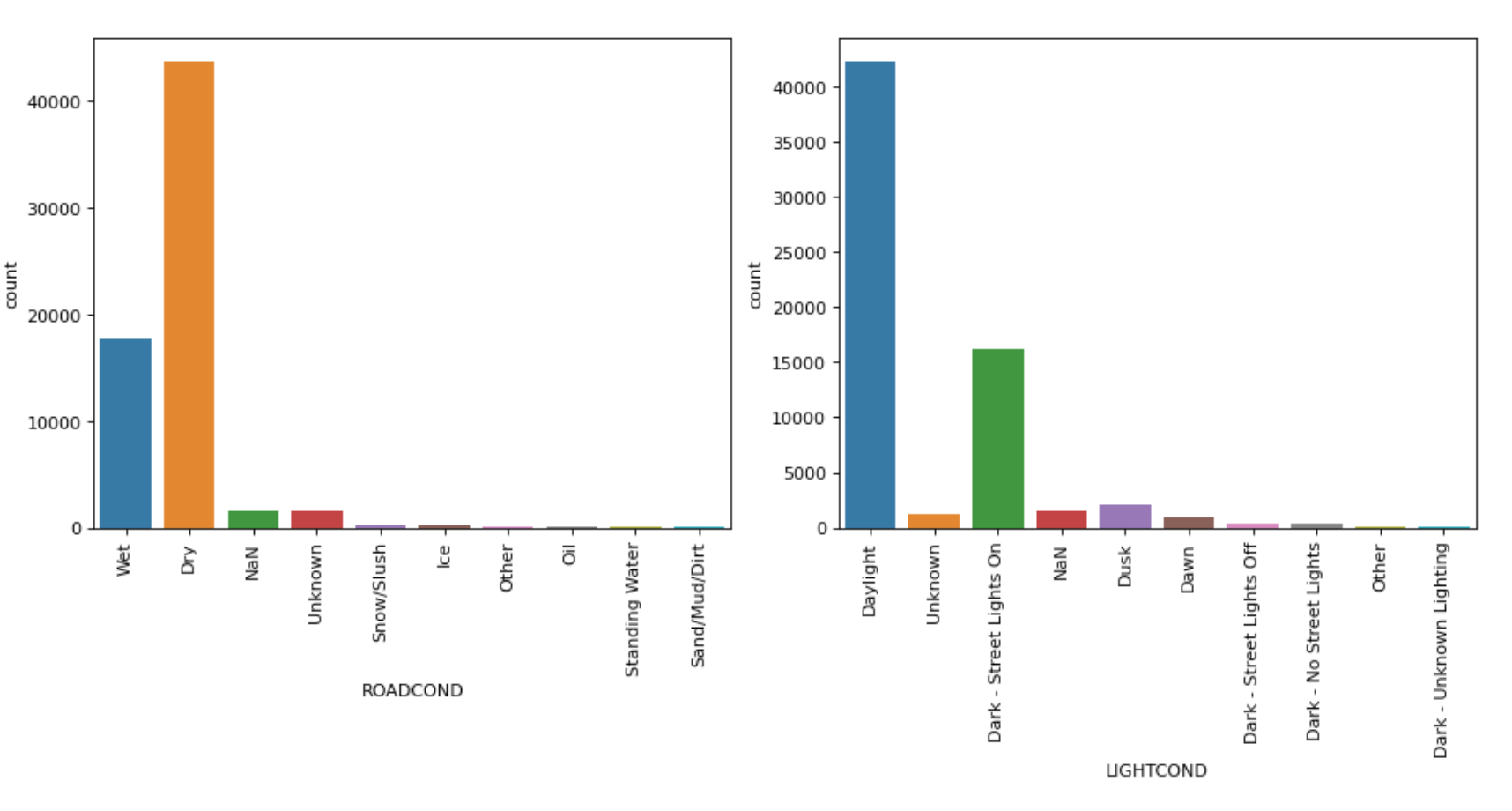
The project plan

1) Assess the state and completeness of the data
2) Identify key attributes potentially influencing collision severity
3) Clean and transform the data as necessary
4) Build machine learning models and assess for accuracy in predicting collision severity
5) Use attribute importance within the chosen model to understand factors that most strongly influence collision severity
6) Make preliminary investigative recommendations on possible safety modifications

Data summary

- 194673 collision records from 2004 to 2020
- 65070 collision records at intersections
- Collision severity measured from least to most severe
- ~35 attributes worth of data
- 12 were manually identified for inclusion in the model as being potentially useful for predicting collision severity

Example: road & light conditions



Attributes used in models

- PEDCOUNT – pedestrian count
- PEDCYLCOUNT – bicycle count
- VEHCOUNT – vehicle count
- INATTENTIONIND – was collision caused by inattention (y/n)
- UNDERINFL – was driver under the influence (y/n)
- COLLISIONTYPE – type of collision e.g. head on, parked car
- WEATHER – weather conditions e.g. overcast, snowing
- ROADCOND – road conditions e.g dry, wet, oil
- LIGHTCOND – light conditions e.g. daylight, dark with no street lights
- PEDROWNOTGRNT – if pedestrian right of way was granted (y/n)
- SPEEDING – if speeding was involved (y/n)
- HITPARKEDCAR – if a parked car was hit (y/n)

Machine learning models: goal

- Use data to train multiple machine learning models
- Determine suitable model based on accuracy and ability to determine feature importance
- Identify most important attributes contributing to collision severity

Machine learning models: choices

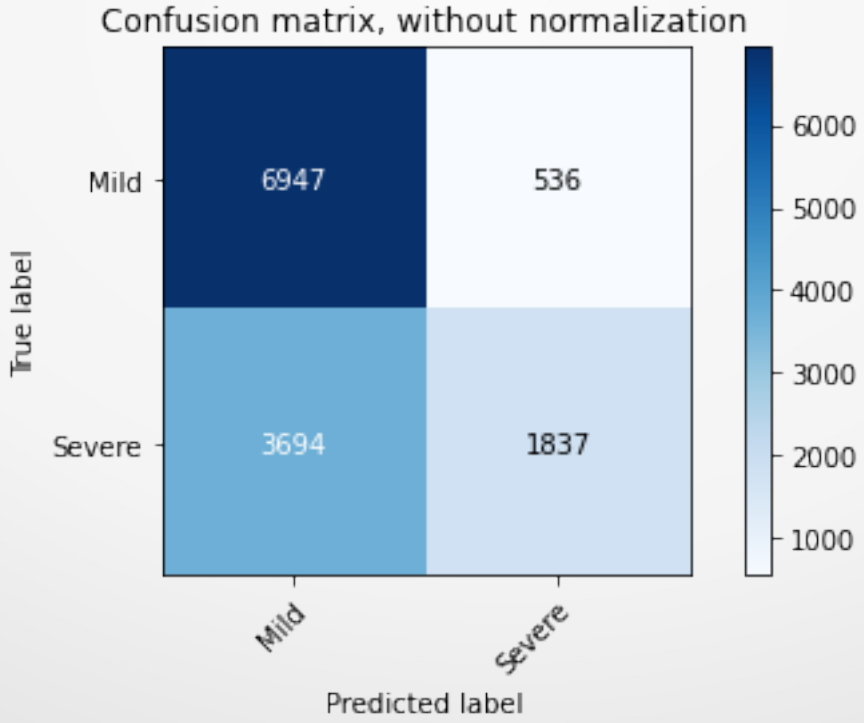
- This is a *classification* problem
- As a result, three models with fundamentally differing approaches to learning were used
- Decision Tree
- Random Forest
- Logistic Regression

Machine learning models: eval

- To assess effectiveness of the model, 20% of the data was reserved for testing
- Scores used to assess:
  - Jaccard score
  - Accuracy score
  - Weighted F1

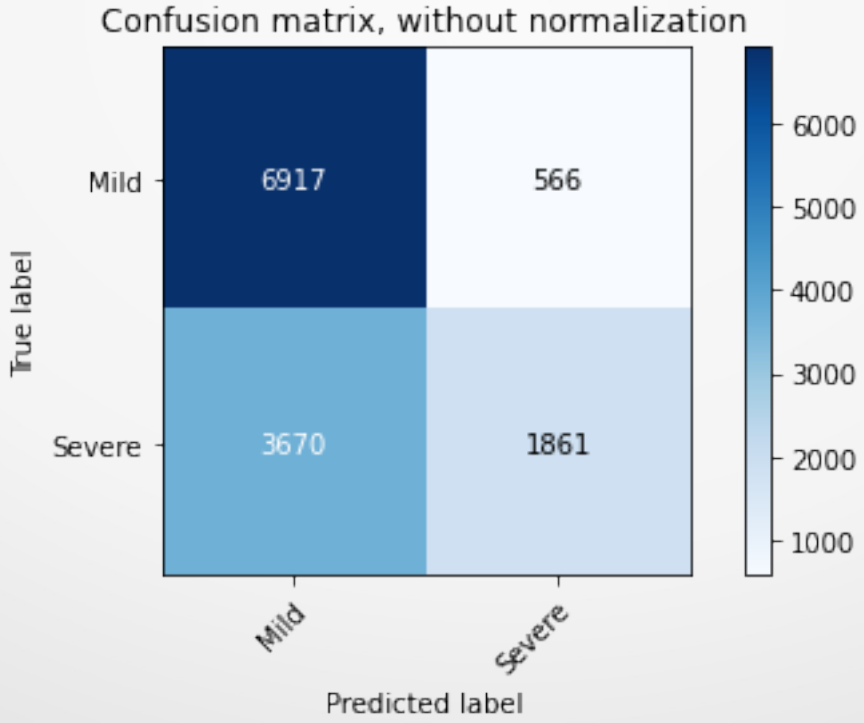
Results: Decision Tree

- Jaccard score 0.3
- Accuracy score 67.5%
- Weighted average F1 0.64



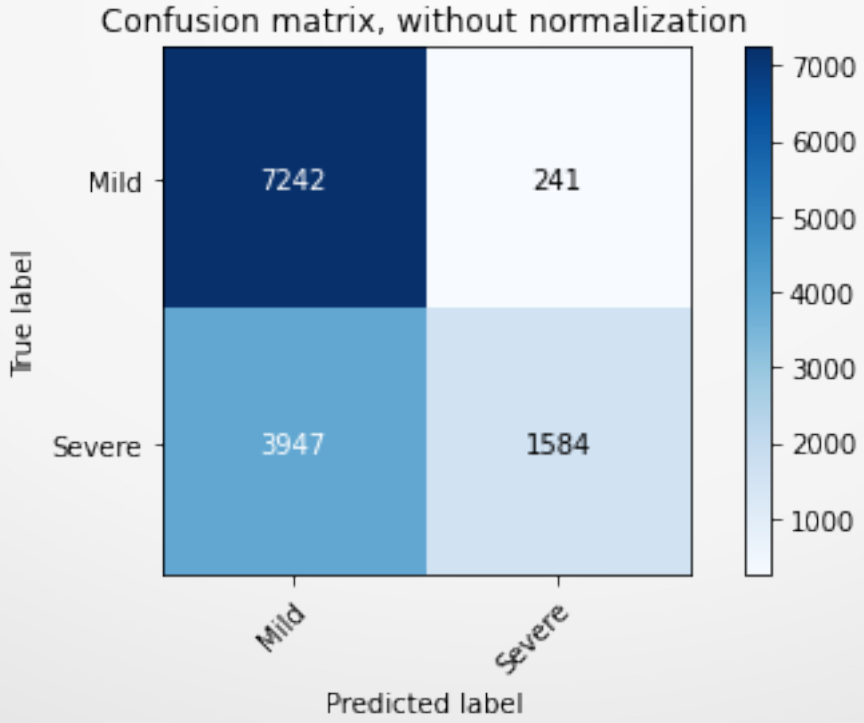
Results: Random Forest

- Jaccard score 0.31
- Accuracy score 67.4%
- Weighted average F1 0.64



Results: Logistic Regression

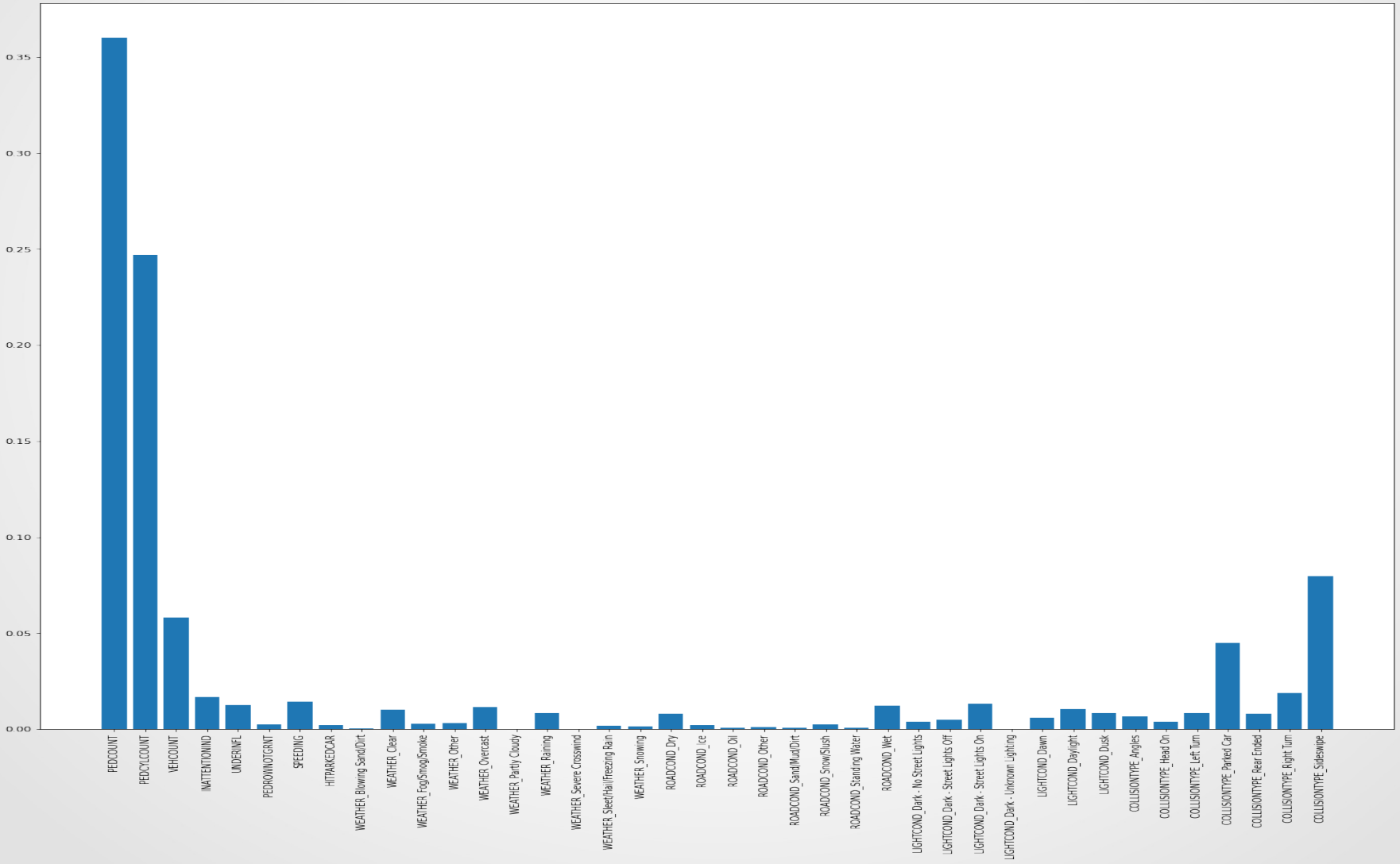
- Jaccard score 0.26
- Accuracy score 67.8%
- Weighted average F1 0.63



Machine Learning Summary

- All models performed modestly well, with accuracy scores at ~67%
- Trivial differences between models, with slightly different strengths and weaknesses for each
- As such, the chosen model was based not on accuracy but on *ease of attribute importance* assessment
- **Winner: Decision Tree takes the decision**

Decision Tree Attribute Importance

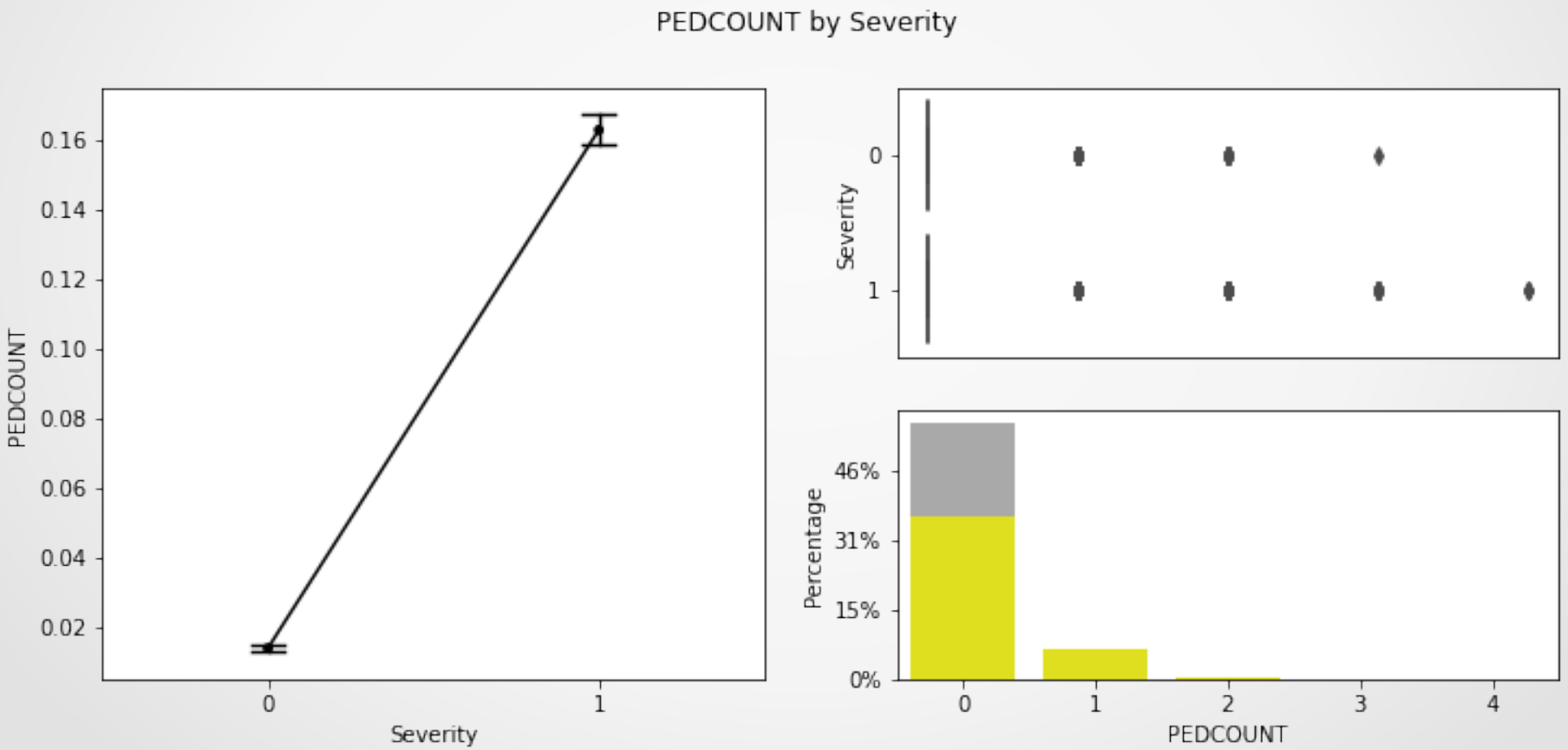


Attributes (in order)

- Pedestrian(s) present
- Bicyclist(s) present
- Sideswipe style collisions
- Number of vehicles involved
- Presence of parked cars

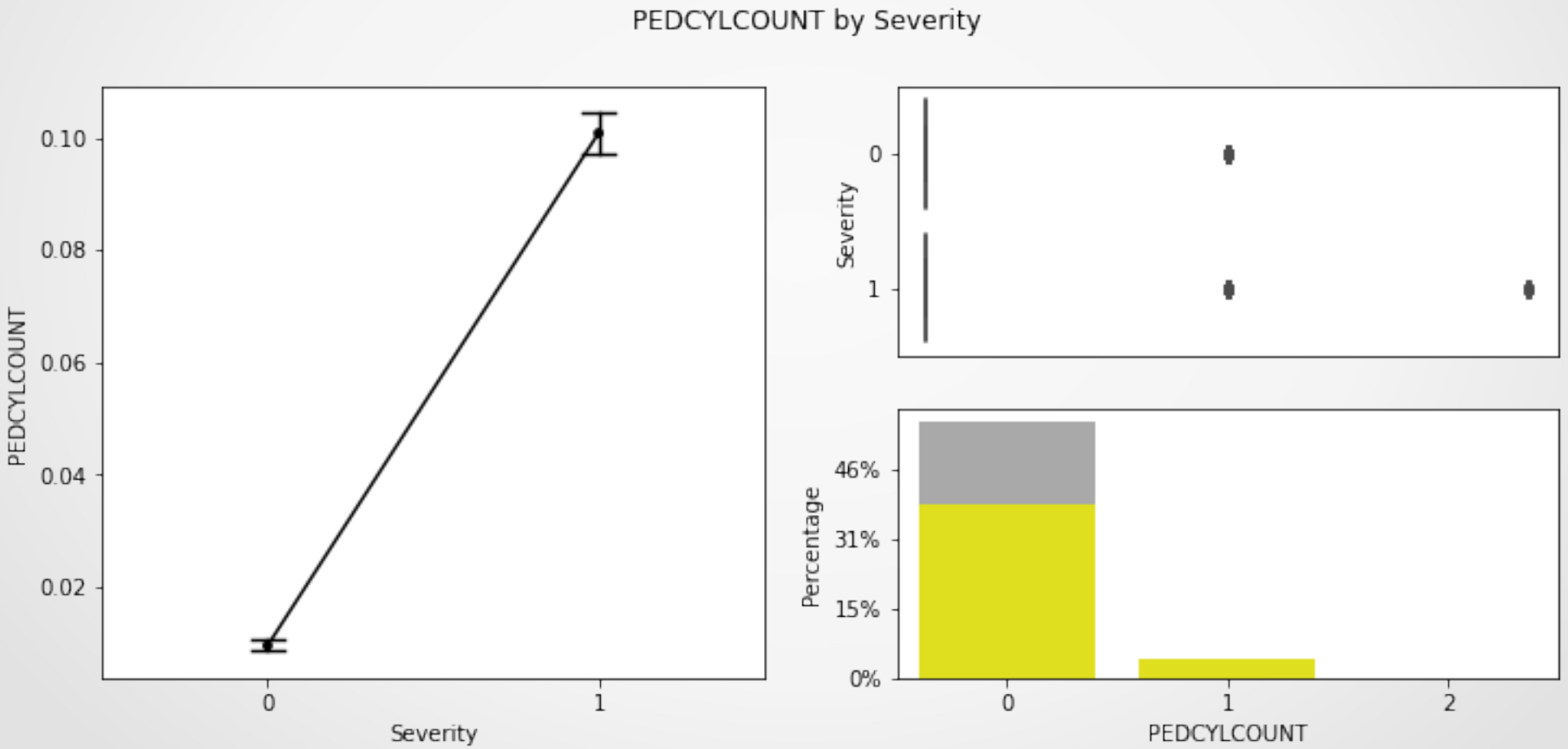
Pedestrians

- Pedestrians present tends to increase collision severity



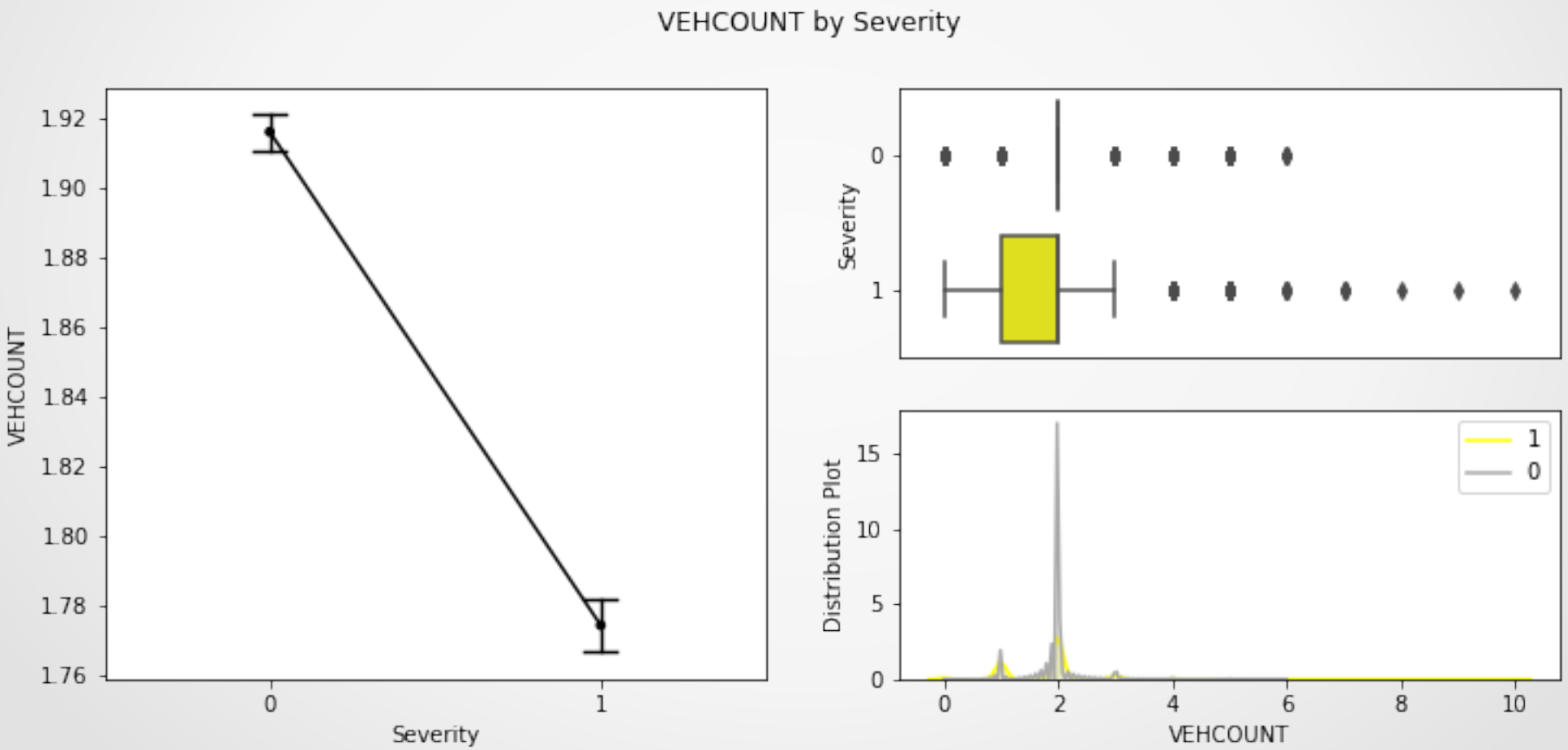
Bicycles

- Bicycles present tends to increase collision severity



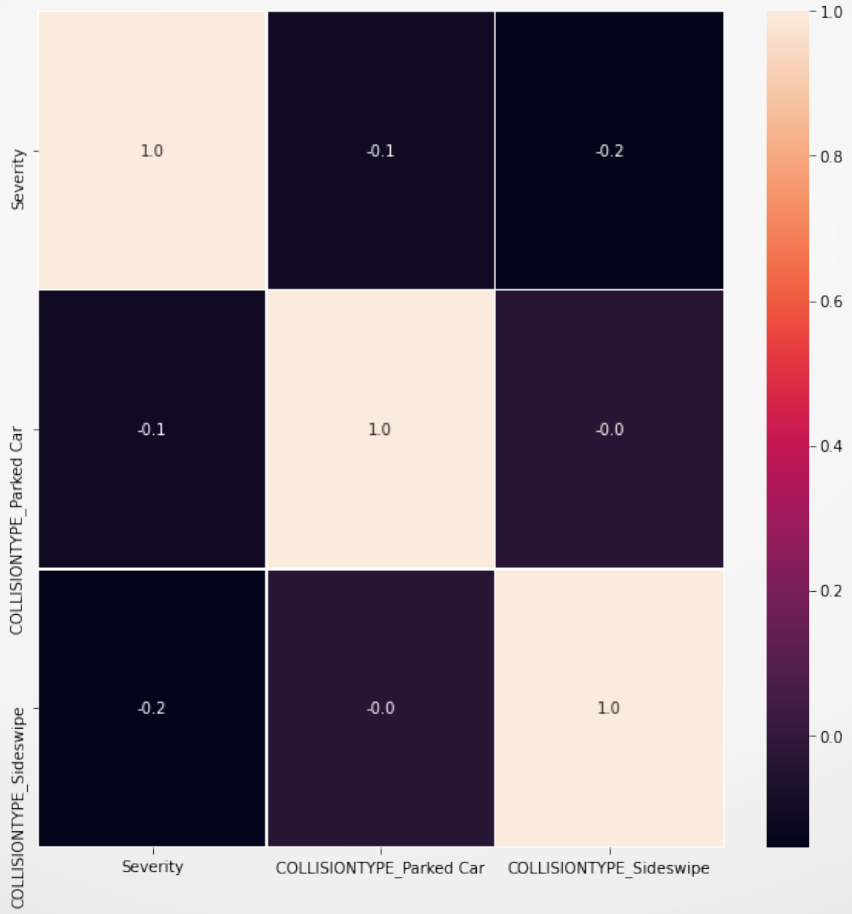
Vehicles

- >Vehicle count tends to decrease collision severity



Parked Cars / Sideswipes

- Parked cars & sideswipes tend to *slightly* decrease collision severity



Assessing Results

- Severity of collision **is most significantly impacted by pedestrian and bicycle involvement**. These factors are by far the largest and most important, likely overwhelming all other factors.

- Vehicle involvement, parked cars, and sideswipes all decrease collision severity. However, this effect may not be independent of the pedestrian effect.

- As such, the number one priority for reducing collision severity at intersections is to increase protections for pedestrians and bicycles.

Recommendations

- Consider how to protect pedestrians and bicyclists:
  - Change signage
  - Increase visibility (visual restrictions)
  - Speed restrictions
  - Time of day vehicle traffic restrictions

  Much depends on the specific characteristics and purpose of each intersection.

Other considerations

- To specifically reduce vehicle-vehicle collisions, more analysis is required
  - Remove data on collisions that involve pedestrians and bicycles and focusing strictly on situations involving vehicles. May produce significantly different results in terms of determining safety requirements.
- Additional future work could also include more sophisticated modelling to increase predictive accuracy.
- Enormous blind spot for analysis: data only looks at collisions. More data on traffic at intersections with no collisions would be very helpful, as it would potentially be just as helpful to look at what makes a safe intersection as what makes an unsafe intersection.

THANK YOU!