

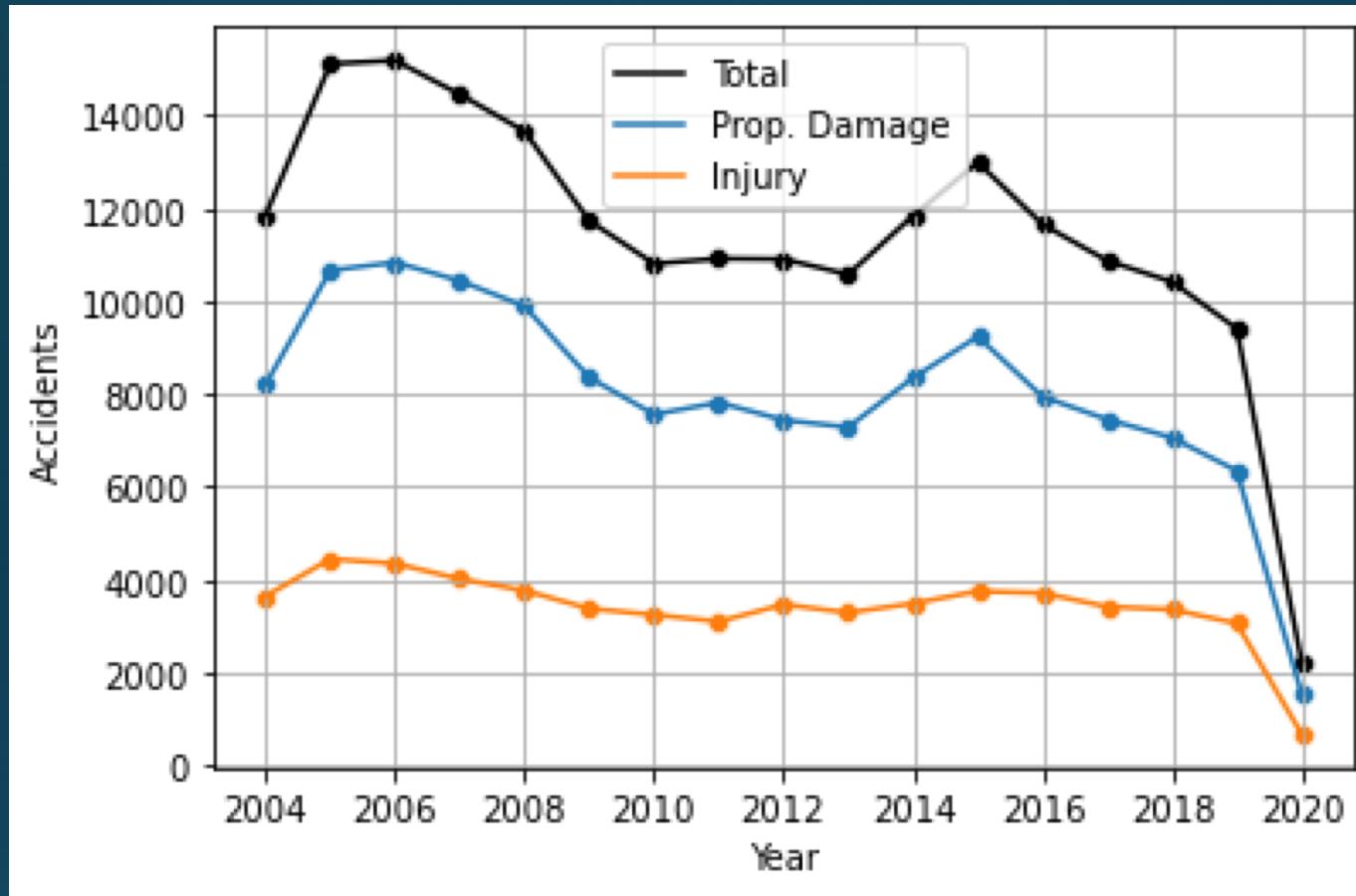
# Predicting Vehicle Accident Severity

- Traffic accidents are a **major concern** to individual safety and financial costs.
- Accidents range in severity from:
  1. Limited to property damage
  2. Injury
  3. Major Injury
  4. Fatality

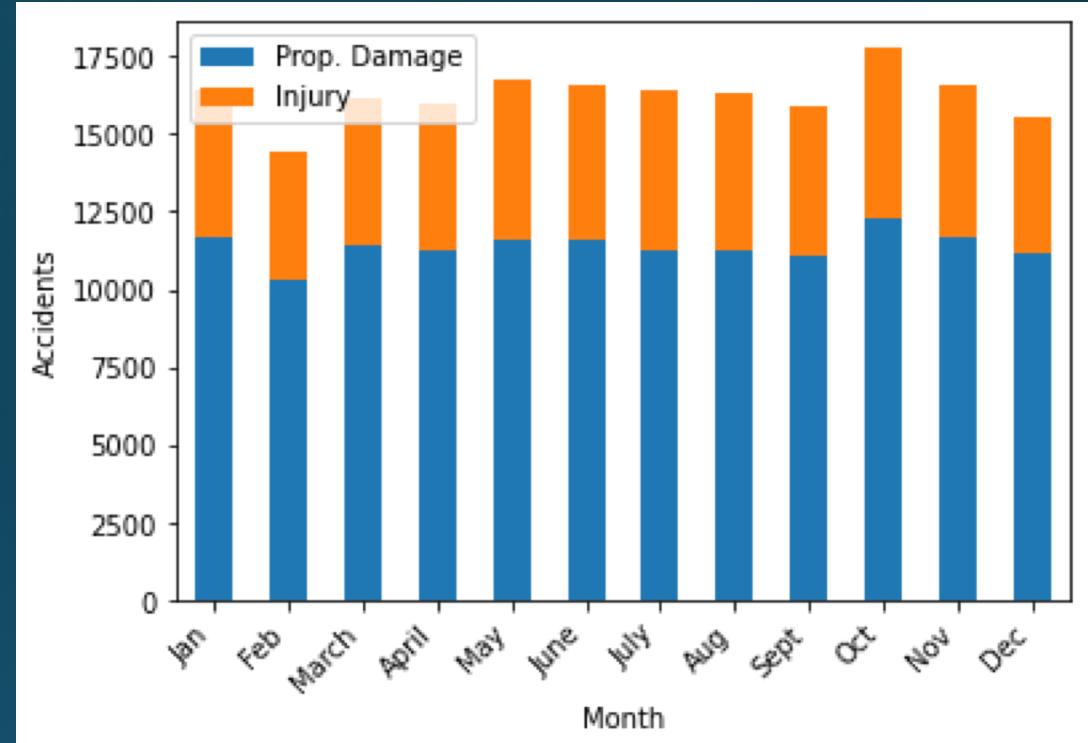
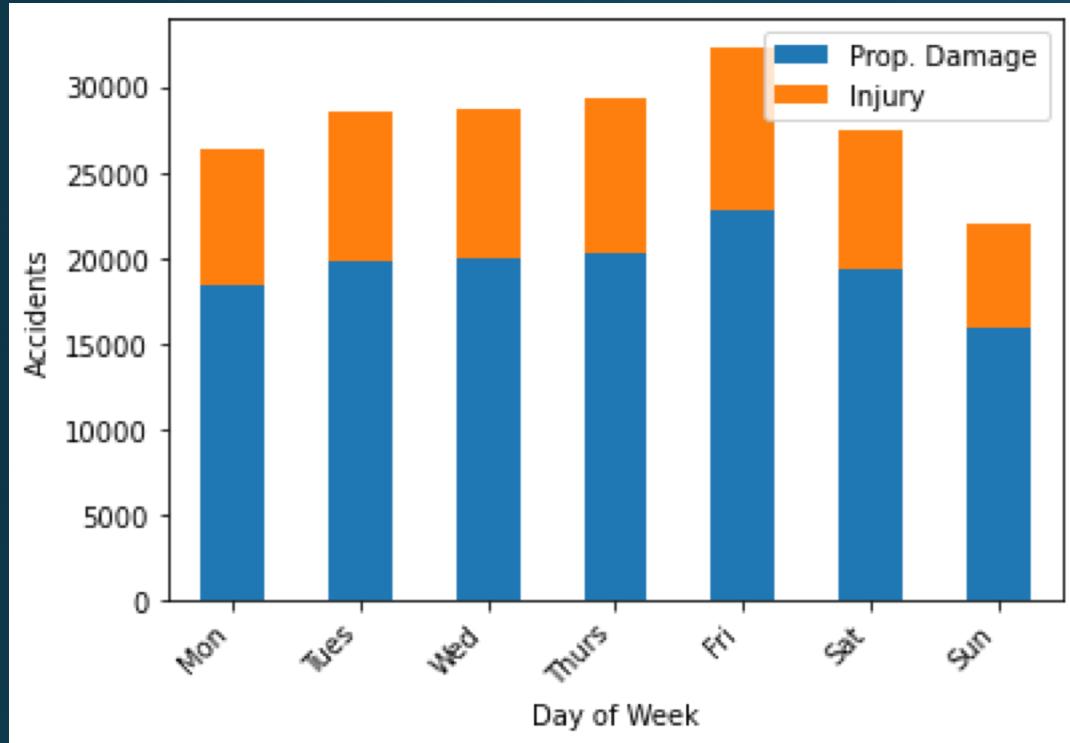
**Question:** Can the factors that increase the accident severity be isolated?

# Data Acquisition and Cleansing

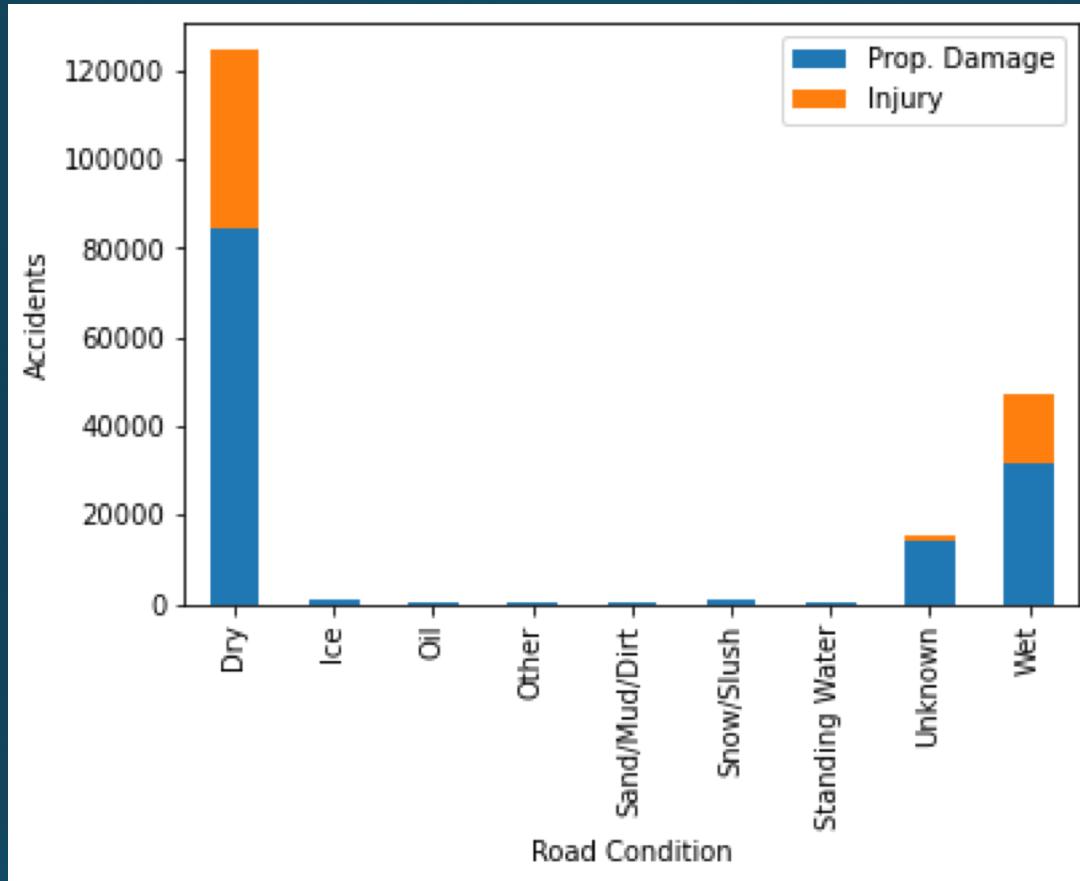
- Traffic collisions reported within the Seattle-area from 2004-present are provided by the SPD and recorded by Traffic Records.
- This dataset provides a number of details to determine these hazardous conditions. Particular factors include:
  - Date
  - Driver distraction (Y/N)
  - Driver under the influence (Y/N)
  - Road Conditions (Dry, Wet, etc)
  - Light Conditions (Daylight, Dark, etc)
  - Speeding involved (Y/N)
  - Parked Car (Y/N)
- This dataset was cleaned to include these factors and organized for analysis.



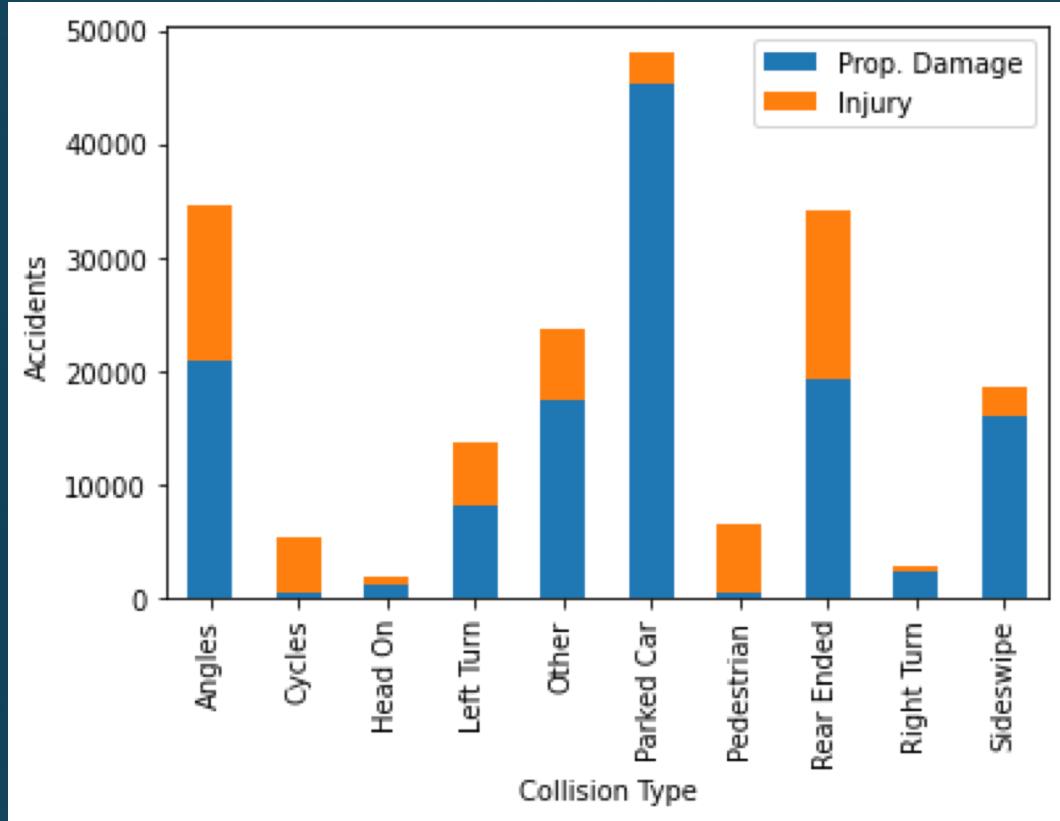
- The number of accidents generally reduces with increasing year.
- Accidents resulting in property damage > Accidents resulting in injury
- Similar trend, except slight peak of 'Prop Damage' in 2015



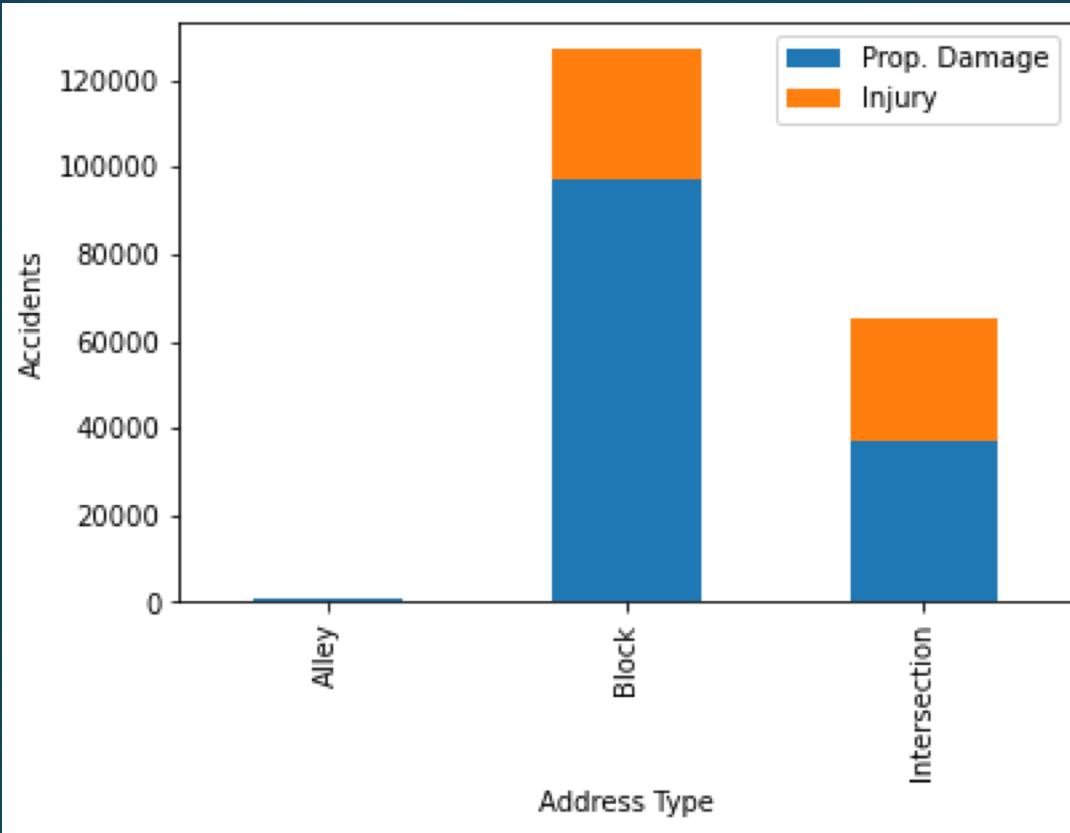
- Day of week and Month had little influence on number of accidents and accident severity.
- Slight increase in number of accidents on Fridays, and reduction on Sundays.
- Slight increase in October, reduction in Feb.



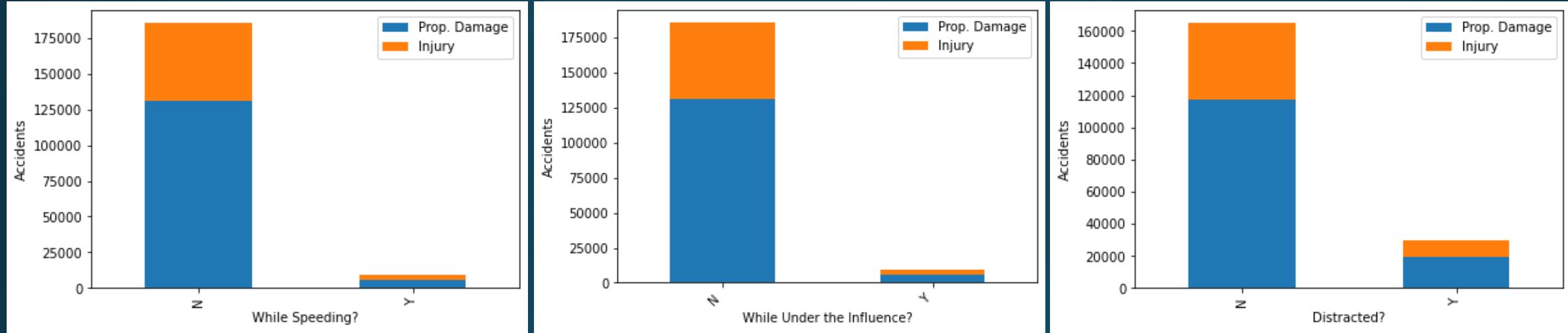
- Accidents most frequent on 'Dry' days.
  - Must analyze weather data for additional insight.
- Similar proportions of Prop. Damage/Injury between 'Dry' and 'Wet'



- Most accidents were reported with parked cars, which also have relatively low severity.
- Collisions at 'angles' and 'rear ends' are fairly frequent, with increased severity.
- Collisions with cycles and pedestrians are infrequent, but have the highest severity.



- Accidents occur more frequently on 'Blocks.'
- Slight inclination of 'Intersections' resulting in proportionally more injuries than 'Blocks.'



- Low number of reported accidents involving speeding or driving under the influence.
- Slightly increased number of reported accidents for distracted drivers.

# Modeling

The following models were used to predict accident severity with moderate success:

- Decision Tree
- Logistic Regression

The following models were attempted, but did could not compute in a practical amount of time:

- K-Nearest Neighbors
- SVM

Algorithm	Jaccard	F1-score	LogLoss
KNN	ERR	ERR	NA
Decision Tree	0.7323	0.8455	NA
SVM	ERR	ERR	NA
LogisticRegression	0.6913	0.8175	0.5815

- Decision tree shows slightly improved accuracy relative to logistic regression.

# Modeling

The following models were used to predict accident severity with moderate success:

- Decision Tree
- Logistic Regression

The following models were attempted, but did could not compute in a practical amount of time:

- K-Nearest Neighbors
- SVM

Algorithm	Jaccard	F1-score	LogLoss
KNN	ERR	ERR	NA
Decision Tree	0.7323	0.8455	NA
SVM	ERR	ERR	NA
LogisticRegression	0.6913	0.8175	0.5815

- Decision tree shows slightly improved accuracy relative to logistic regression.

# Conclusion

**Question:** Can the factors that increase the accident severity be isolated?

- Collision type appeared most significant.
  - Accidents with parked cars are most frequent but least severe.
  - Accidents at angles and rear-ends are still quite frequent and notably severe
  - While infrequent, accidents with cycles and pedestrians are most severe.
- A decision tree model was developed that could predict accident severity with adequate accuracy.