

A decorative graphic on the left side of the slide consisting of white lines and circles on a blue gradient background, resembling a circuit board or neural network structure.

PREDICTING CAR ACCIDENT SEVERITY

DANA ANDREA

PREDICTING ACCIDENT SEVERITY IS KEY FOR FIRST RESPONDERS

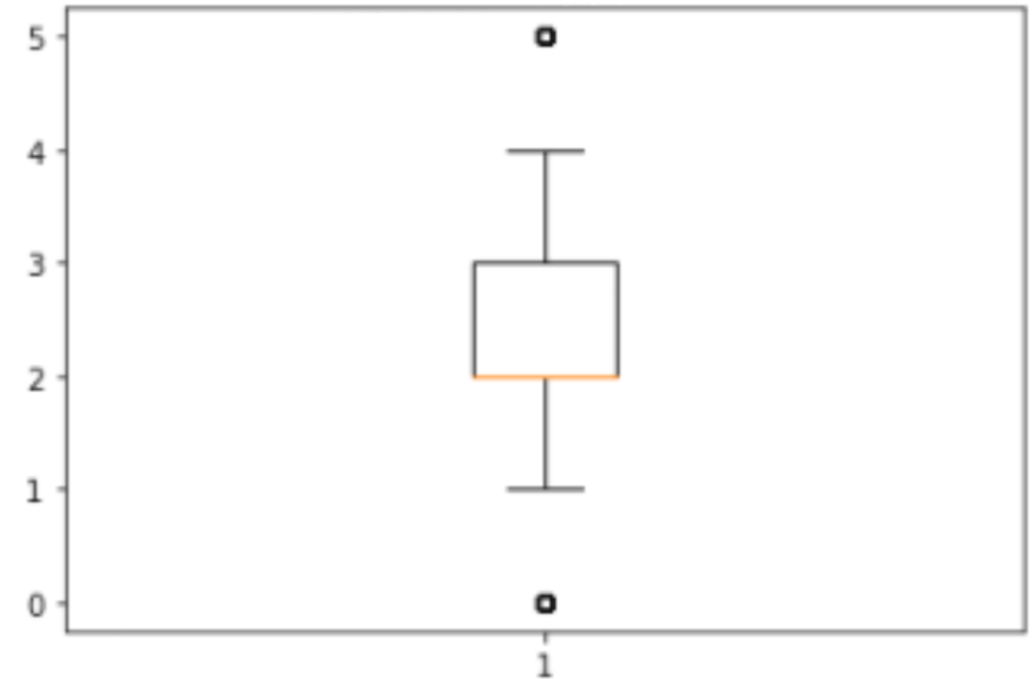
- Firetruck accidents rank as the second leading cause of on-the-job deaths for firefighters.
- Approximately 500 firefighters are involved in fatal firetruck crashes every year; on average, 1 in 100 of those occupants die as a result of the crash.
- Between 2006 and 2016, more than one police officer per week was killed on average from a collision or from being struck directly by another vehicle.
- More than 10,000 ambulance-related collisions occur annually; from 1993 to 2010, approximately 97 EMS technicians were killed in collisions.

DATA ACQUISITION

- The data used in this project contains 194,673 records of motor vehicle accidents and corresponding weather conditions from Seattle, WA.
- The date range for these accidents is from 2004-2020.
- This dataset was provided by Coursera.

DATA CLEANSING

- Remove excess identifier columns
- Convert NaN values to “Unknown”
- Remove Person Count Outliers (limit to accidents involving 5 persons or less)



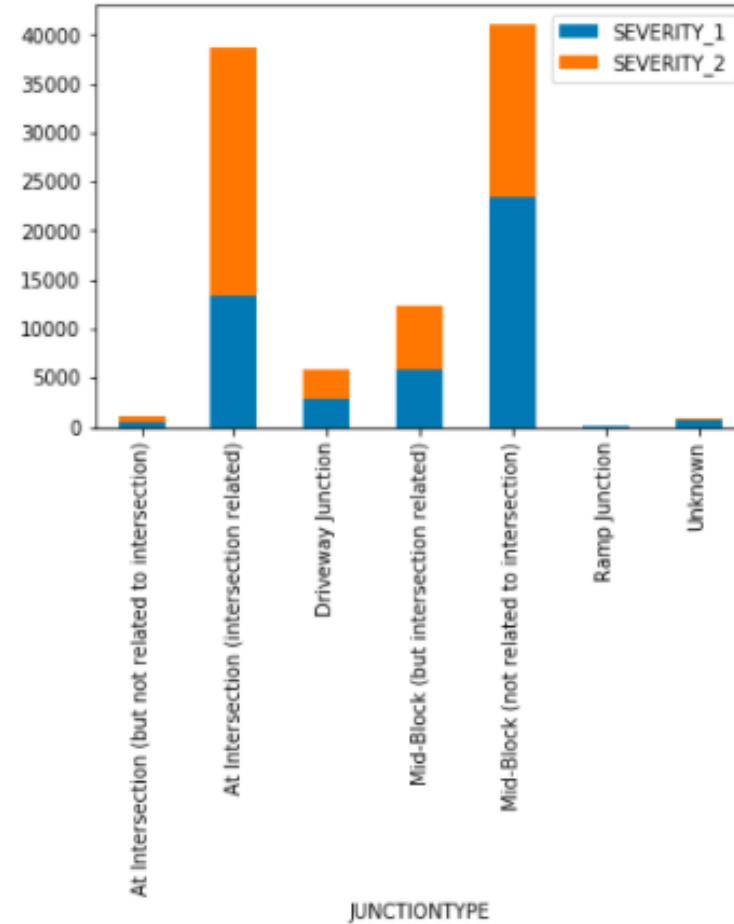
WEATHER CONDITION IMPACT ON ACCIDENT SEVERITY

ROADCOND	Dry	Ice	Oil	Other	Sand/Mud/Dirt	Snow/Slush	Standing Water	Unknown	Wet
SEVERITY_1	47.5%	59.4%	48.7%	46.1%	46.3%	68.9%	50.0%	80.2%	46.3%
SEVERITY_2	52.5%	40.6%	51.3%	53.9%	53.7%	31.1%	50.0%	19.8%	53.7%

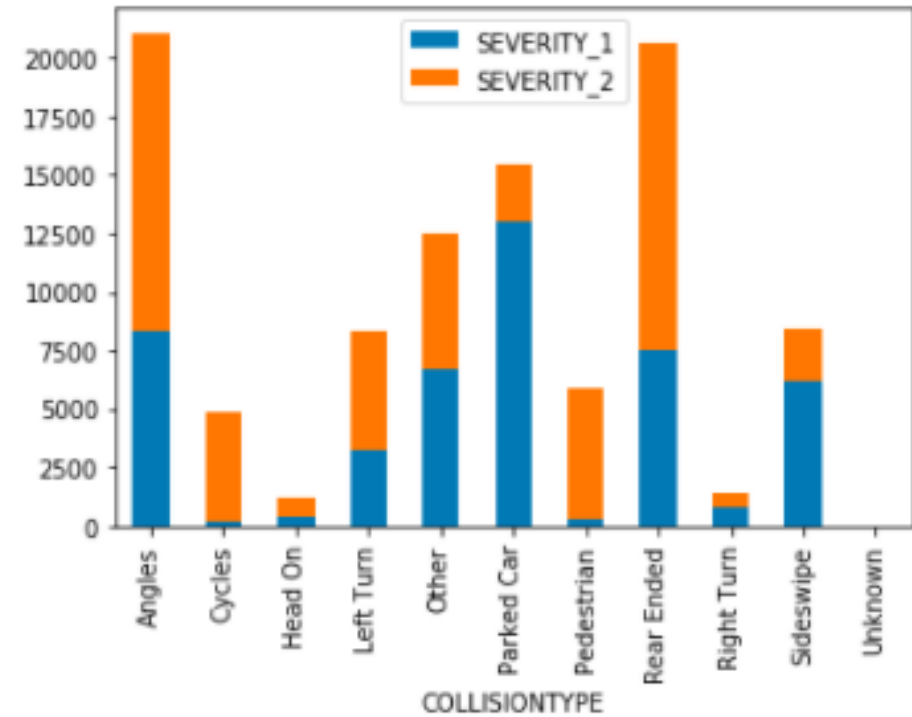
WEATHER	Blowing Sand/Dirt	Clear	Fog/Smog/Smoke	Other	Overcast	Partly Cloudy	Raining	Severe Crosswind	Sleet/Hail/Freezing Rain	Snowing	Unknown
SEVERITY_1	51.9%	47.4%	46.0%	71.2%	48.1%	25.0%	45.7%	53.8%	54.4%	66.0%	79.6%
SEVERITY_2	48.1%	52.6%	54.0%	28.8%	51.9%	75.0%	54.3%	46.2%	45.6%	34.0%	20.4%

LIGHTCOND	Dark - No Street Lights	Dark - Street Lights Off	Dark - Street Lights On	Dark - Unknown Lighting	Dawn	Daylight	Dusk	Other	Unknown
SEVERITY_1	60.2%	55.1%	50.4%	50.0%	45.8%	46.3%	45.7%	60.3%	80.0%
SEVERITY_2	39.8%	44.9%	49.6%	50.0%	54.2%	53.7%	54.3%	39.7%	20.0%

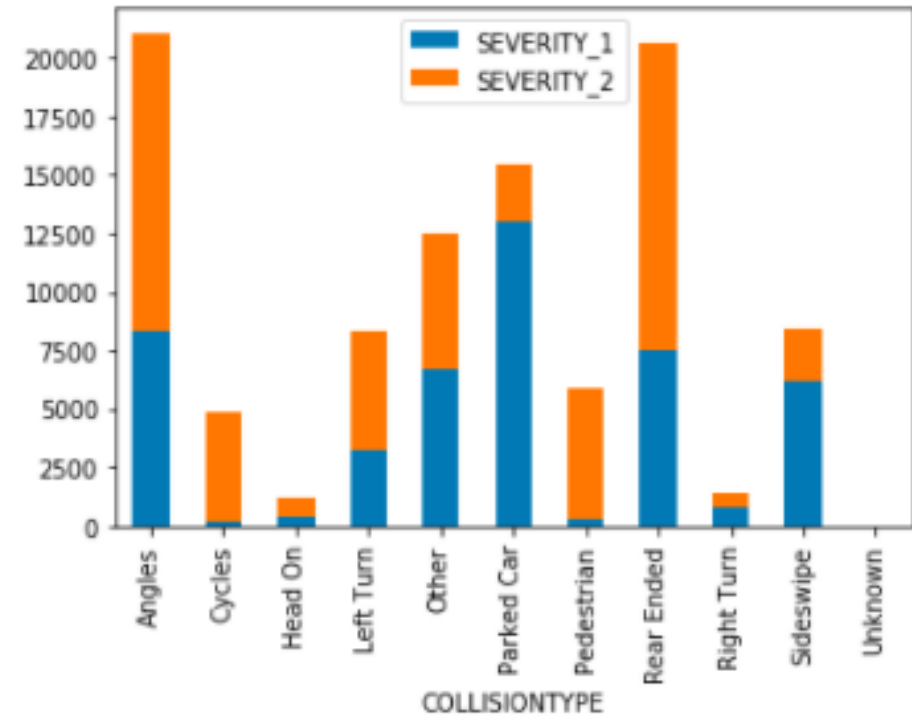
JUNCTION TYPE ANALYSIS



COLLISION TYPE ANALYSIS



COLLISION TYPE ANALYSIS



K-NEAREST NEIGHBOR AND LOGISTIC REGRESSION

```
In [29]: from sklearn import metrics  
print("Train set Accuracy: ", metrics.accuracy_score(y_train, neigh.predict(X_train)))  
print("Test set Accuracy: ", metrics.accuracy_score(y_test, yhat))
```

Train set Accuracy: 0.6619509020127964
Test set Accuracy: 0.6560565870910698

```
In [33]: from sklearn.metrics import jaccard_similarity_score  
jaccard_similarity_score(y_test, yhat)
```

```
Out[33]: 0.6957461440220061
```

CONCLUSION

- K-nearest neighbor yields 66% accuracy
- Logistic Regression yields 70% accuracy
- The insights gained from this model is that accidents involving pedestrians and cyclists at intersections and midblock are the most severe in terms of personal injury.
- This means that in order to reduce accident damage, first responders should be concentrated in locations where pedestrians and cyclists are common - like city centers, school zones, city parks, etc.