



PREDICTING COLLISION SEVERITY IN SEATTLE

FACTORS LEADING TO INJURIES IN COLLISIONS

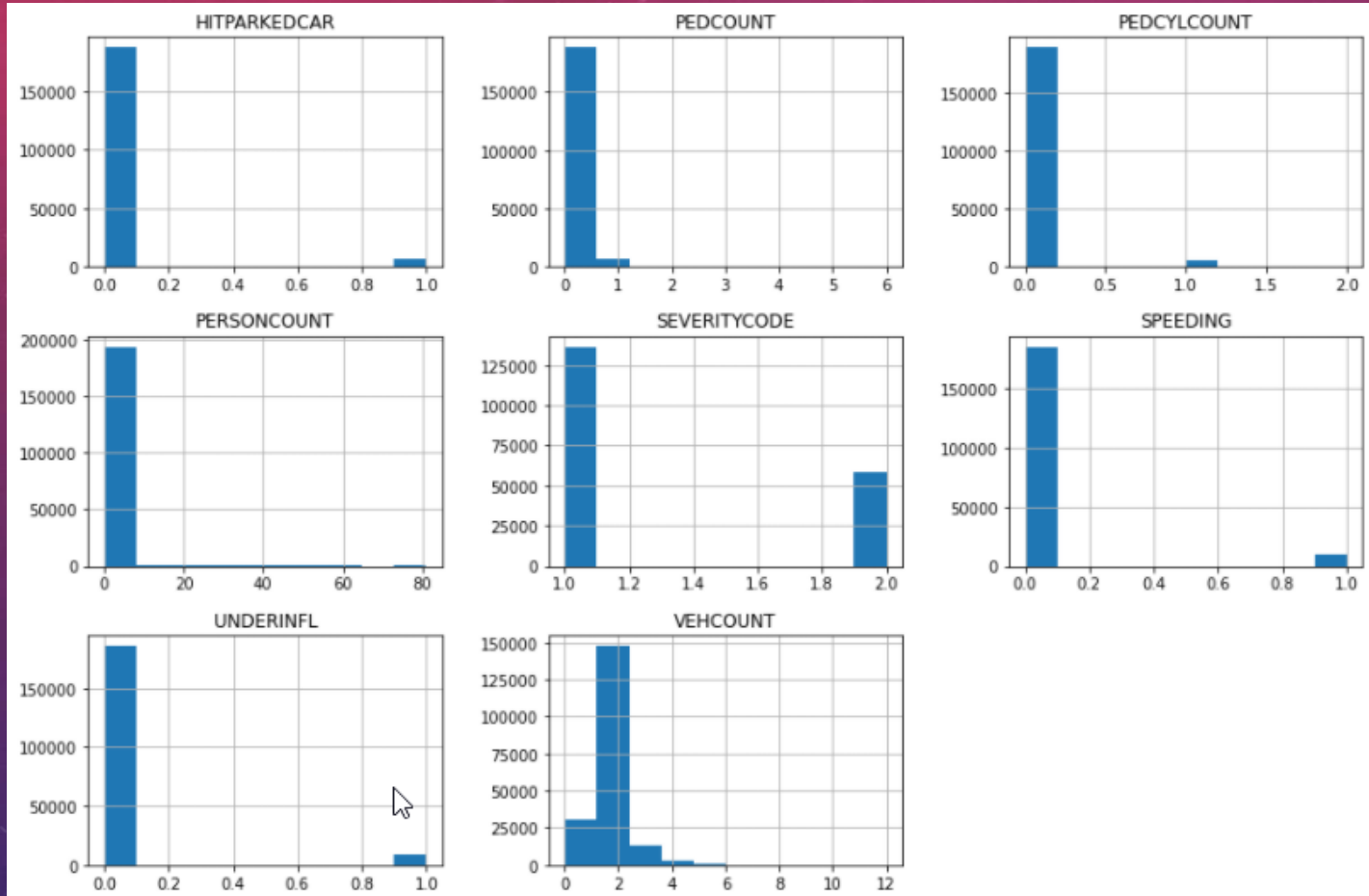
DATA EXPLORATION AND HYGIENE

- Example data came from the Coursera capstone project assignment
- Dataset contained 194,643 records with 38 fields.
 - Selected data for analysis contained all rows for 15 fields; 12 of these fields were analyzed for relevance.
 - Replaced all missing values with zeros and normalized true/false, yes/no, and mixed bool indicators to 1s or 0s
 - Replaced all missing values in text fields with 'Unknown'
 - Duplicate severity code was dropped.
 - One-hot encoding was done for weather and collision type.

```
OBJECTID
[ 1 2 3 ... 219545 219546 219547]
SEVERITYCODE
[2 1]
SEVERITYDESC
['Injury Collision' 'Property Damage Only Collision']
ADDRTYPE
['Intersection' 'Block' 'Alley' nan]
COLLISIONTYPE
['Angles' 'Sideswipe' 'Parked Car' 'Other' 'Cycles' 'Rear Ended' 'Head On'
'Unknown' 'Left Turn' 'Pedestrian' 'Right Turn']
PERSONCOUNT
[ 2 4 3 0 1 5 6 16 8 7 11 9 12 17 26 22 10 37 13 36 28 14 53 19
30 29 23 44 15 32 21 41 27 20 35 43 81 18 25 48 24 34 57 39 47 54 31]
PEDCOUNT
[0 1 2 3 4 5 6]
PEDCYLCOUNT
[0 1 2]
VEHCOUNT
[ 2 3 1 0 4 7 5 6 8 11 9 10 12]
UNDERINFL
[0 1]
WEATHER
['Overcast' 'Raining' 'Clear' 'Unknown' 'Other' 'Snowing' 'Fog/Smog/Smoke'
'Sleet/Hail/Freezing Rain' 'Blowing Sand/Dirt' 'Severe Crosswind'
'Partly Cloudy']
ROADCOND
['Wet' 'Dry' 'Unknown' 'Snow/Slush' 'Ice' 'Other' 'Sand/Mud/Dirt'
'Standing Water' 'Oil']
LIGHTCOND
['Daylight' 'Dark - Street Lights On' 'Dark - No Street Lights' 'Unknown'
'twilight' 'Dark - Street Lights Off' 'Other' 'Dark - Unknown Lighting']
SPEEDING
[0 1]
HITPARKEDCAR
[0 1]
```

Normalized data for each field (post processing)

HISTOGRAMS OF NUMERIC FIELDS

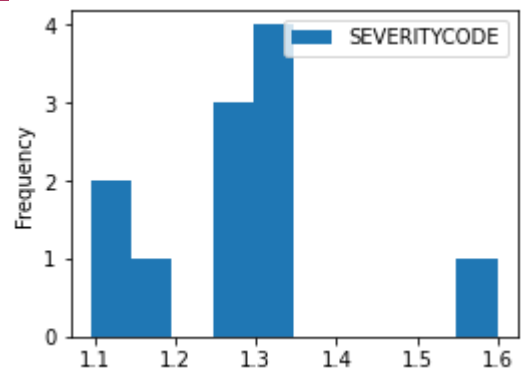


Numeric fields contained largely the same values and were therefore generally not useful.

SEVERITY FOR WEATHER AND COLLISION TYPE

normalized severity code for weather

| WEATHER | SEVERITYCODE |
|--------------------------|--------------|
| Blowing Sand/Dirt | 1.267857 |
| Clear | 1.322491 |
| Fog/Smog/Smoke | 1.328647 |
| Other | 1.139423 |
| Overcast | 1.315544 |
| Partly Cloudy | 1.600000 |
| Raining | 1.337185 |
| Severe Crosswind | 1.280000 |
| Sleet/Hail/Freezing Rain | 1.247788 |
| Snowing | 1.188534 |
| Unknown | 1.094190 |



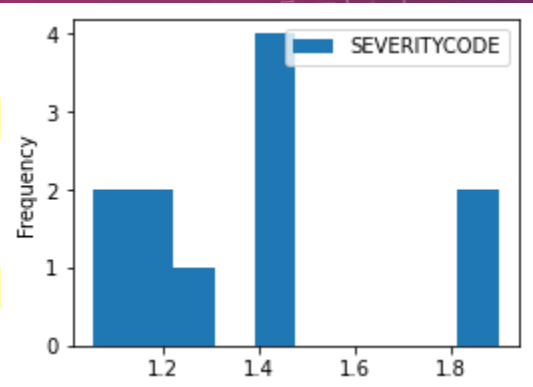
Normalized severity by weather type

| WEATHER | SEVERITYCODE | |
|--------------------------|--------------|----------|
| Blowing Sand/Dirt | 1 | 0.732143 |
| | 2 | 0.267857 |
| Clear | 1 | 0.677509 |
| | 2 | 0.322491 |
| Fog/Smog/Smoke | 1 | 0.671353 |
| | 2 | 0.328647 |
| Other | 1 | 0.860577 |
| | 2 | 0.139423 |
| Overcast | 1 | 0.684456 |
| | 2 | 0.315544 |
| Partly Cloudy | 2 | 0.600000 |
| | 1 | 0.400000 |
| Raining | 1 | 0.662815 |
| | 2 | 0.337185 |
| Severe Crosswind | 1 | 0.720000 |
| | 2 | 0.280000 |
| Sleet/Hail/Freezing Rain | 1 | 0.752212 |
| | 2 | 0.247788 |
| Snowing | 1 | 0.811466 |
| | 2 | 0.188534 |
| Unknown | 1 | 0.945928 |
| | 2 | 0.054072 |

- Note that the average of 1.6 indicates that there are more injuries than property damage (50/50 = 1.5)
- Injuries on partly cloudy days are 60%.
- Correlation is not causation as there are more often partially cloudy days in Seattle.

SEVERITYCODE

| COLLISIONTYPE | SEVERITYCODE |
|---------------|--------------|
| Angles | 1.392917 |
| Cycles | 1.876085 |
| Head On | 1.430830 |
| Left Turn | 1.394877 |
| Other | 1.257858 |
| Parked Car | 1.055473 |
| Pedestrian | 1.898305 |
| Rear Ended | 1.430361 |
| Right Turn | 1.206022 |
| Sideswipe | 1.134666 |
| Unknown | 1.212276 |



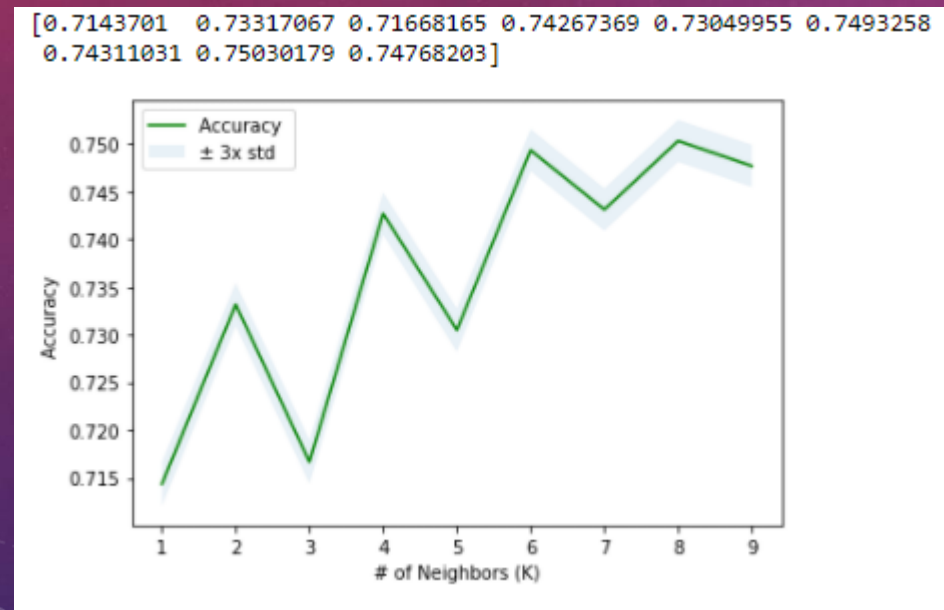
| COLLISIONTYPE | SEVERITYCODE | |
|---------------|--------------|----------|
| Angles | 1 | 0.607083 |
| | 2 | 0.392917 |
| Cycles | 2 | 0.876085 |
| | 1 | 0.123915 |
| Head On | 1 | 0.569170 |
| | 2 | 0.430830 |
| Left Turn | 1 | 0.605123 |
| | 2 | 0.394877 |
| Other | 1 | 0.742142 |
| | 2 | 0.257858 |
| Parked Car | 1 | 0.944527 |
| | 2 | 0.055473 |
| Pedestrian | 2 | 0.898305 |
| | 1 | 0.101695 |
| Rear Ended | 1 | 0.569639 |
| | 2 | 0.430361 |
| Right Turn | 1 | 0.793978 |
| | 2 | 0.206022 |
| Sideswipe | 1 | 0.865334 |
| | 2 | 0.134666 |

- Note that the averages of more than 1.5 indicate more injuries than property damage (50/50 = 1.5)
- Injuries are far more likely (almost 90%) when a pedestrian or cyclist is involved.

KNN RESULTS

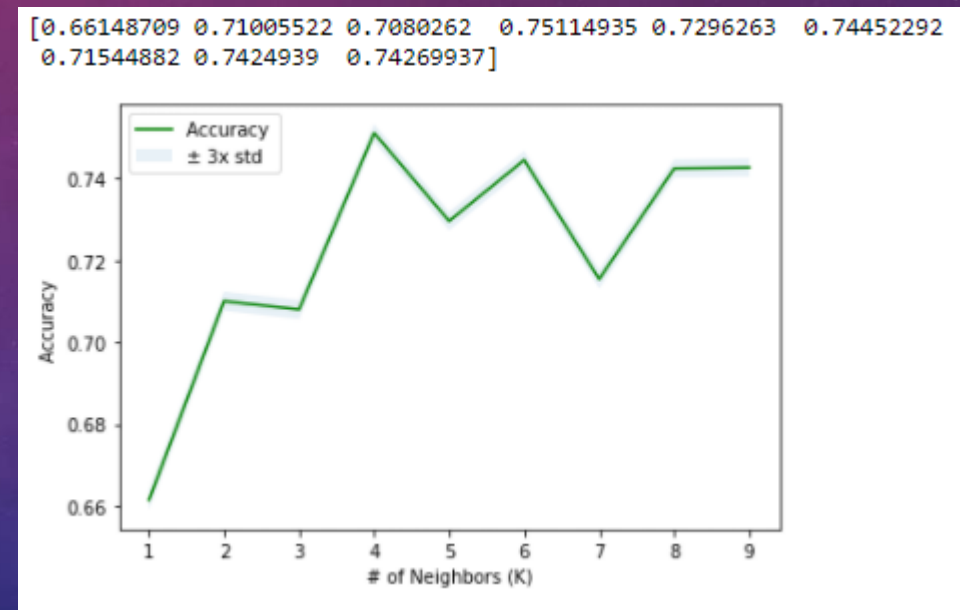
Models were created using KNN test/train data using dataframes that were one hot encoded for either weather or collision. They were also tested to determine the best k for testing

Weather



```
Train set acc: 0.7472999524843006  
Test set acc: 0.7503017850263259
```

Collision Type



```
KNeighborsClassifier(n_neighbors=4)  
Train set acc: 0.7489758440457692  
Test set acc: 0.7511493514832412
```


CONCLUSION AND FUTURE DIRECTION

- The greatest indicator of whether or not there will be an injury is whether or not a pedestrian or cyclist is involved.
- Additional data including severity greater than 2 would be necessary to make a better analysis.
- More widespread (including other cities, states, or countries) would be beneficial to determine if something like address type or road conditions would have more or less of an effect.
- In the future if I were to redo this I would probably drop the pedestrian and cyclist data and attempt to determine which features had an effect solely on the remaining incidents.

The background is a gradient of deep purple and blue, filled with numerous out-of-focus circular light spots (bokeh) in various shades. Overlaid on the left side are several faint, white geometric patterns, including concentric circles, arcs, and a large circular scale with numerical markings from 140 to 260. Some of these patterns have small arrows indicating direction.

THANK YOU!

NATHAN