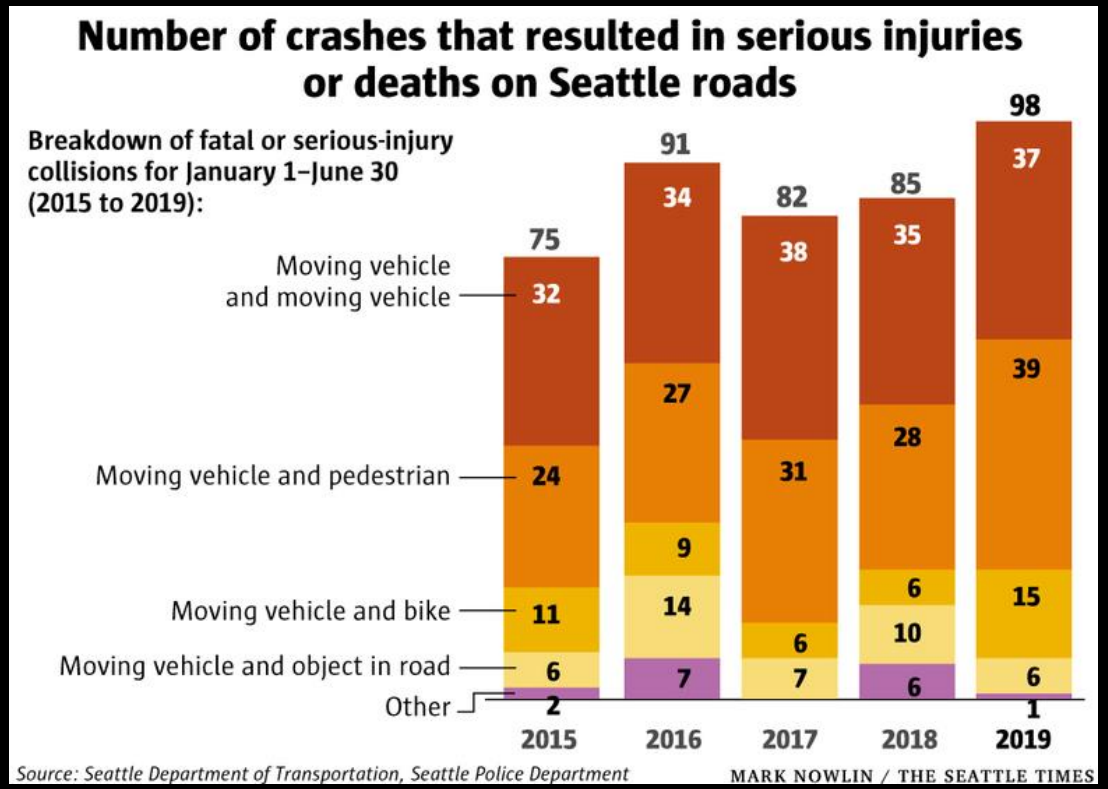
Predicting Car accident severity

Moving towards ‘Vision Zero’ by 2030

Background

The Seattle Department of Transportation (SDOT) has a mission to eliminate all traffic-related deaths and injuries by 2030, also known as the ‘Vision Zero’ program.

On average, the city faces 85 cases related to serious-injury or collisions and these numbers are far from achieving SDOT’s target



Source: https://www.seattletimes.com/seattle-news/transportation/sdot-data-show-nearly-100-serious-injury-or-fatal-collisions-on-seattle-streets-in-first-half-of-2019/

To better predict severity and improve road conditions

Business objective and goals

To help SDOT understand which attributes in their data set are the best predictors (i.e weather, location, road conditions) of accident severity

The predictive model will support their investment decisions, thereby reducing the cases of injury and collisions

Source: xxxx

Severitycode will be assigned as the dependent variable

Data source

Timeframe:

2004-2020

194673 rows of data

37 attributes, including:

- Severitycode (Target field)
  - 3—fatality
  - 2b—serious injury
  - 2—injury
  - 1—prop damage
  - 0—unknown
- Location
- Injuries
- Fatalities
- JunctionType
- Weather
- RoadCond(itions)
- Speeding

Source: Dataset provided by SPD and recorded by Traffic records

Data cleaning

Data set has been reduced to 184,541 entries and 31 attributes

Summary of Changes

Attribute	Changes	Rationale
All	Replace empty cells with NaN	Ensure that numpy library can detect missing data
SEVERITYCODE	2→1, 1→0	Only 2 outcomes (1&2), transform it to binary values
INTKEY, EXCEPTRSNCODE, EXCEPTRSNDESC, INATTTENTIONID, PEDROWNOTGRNT, SPEEDING, SDOTCOLNUM	Drop column	Either because there were more than 180,000 unfilled cells or if it does not add any value to the analysis (eg. SDOTCOLNUM)
X, Y, ST_COLCODE, ST_COLDESC, UNDERINFL	Remove rows with NaN	Values are too specific to be replaced with frequency or mean data
UNDERINFL	“Y”→1, “N”→0	Alignment with the rest of the data
ADDRTYPE, COLLISIONTYPE, JUNCITIONTYPE,WEATHER,	Replace NaN with mode data	Reasonable to replace with the mode of the data set

Week 3 (TBC)

xxx

xx

Source: xxxx