Capstone Project

Car accident severity

BART FISCHER 09-10-2020

Introduction

- Worldwide goal is a reductions of traffic accidents.
  - In 2013 => 1.25 million traffic accident deaths.
- There’s a need to predict accidents and accident severity for
  - Safer routes
  - Optimize public transport
  - Improvement on infrastructure costs
- Relevant for drivers, public transport officials and first responders

-

Goal

- Investigate a road accident dataset with a binary severity class
- Try to predict the severity of an accident,
  - Severity being property damage or injury.

Data provided

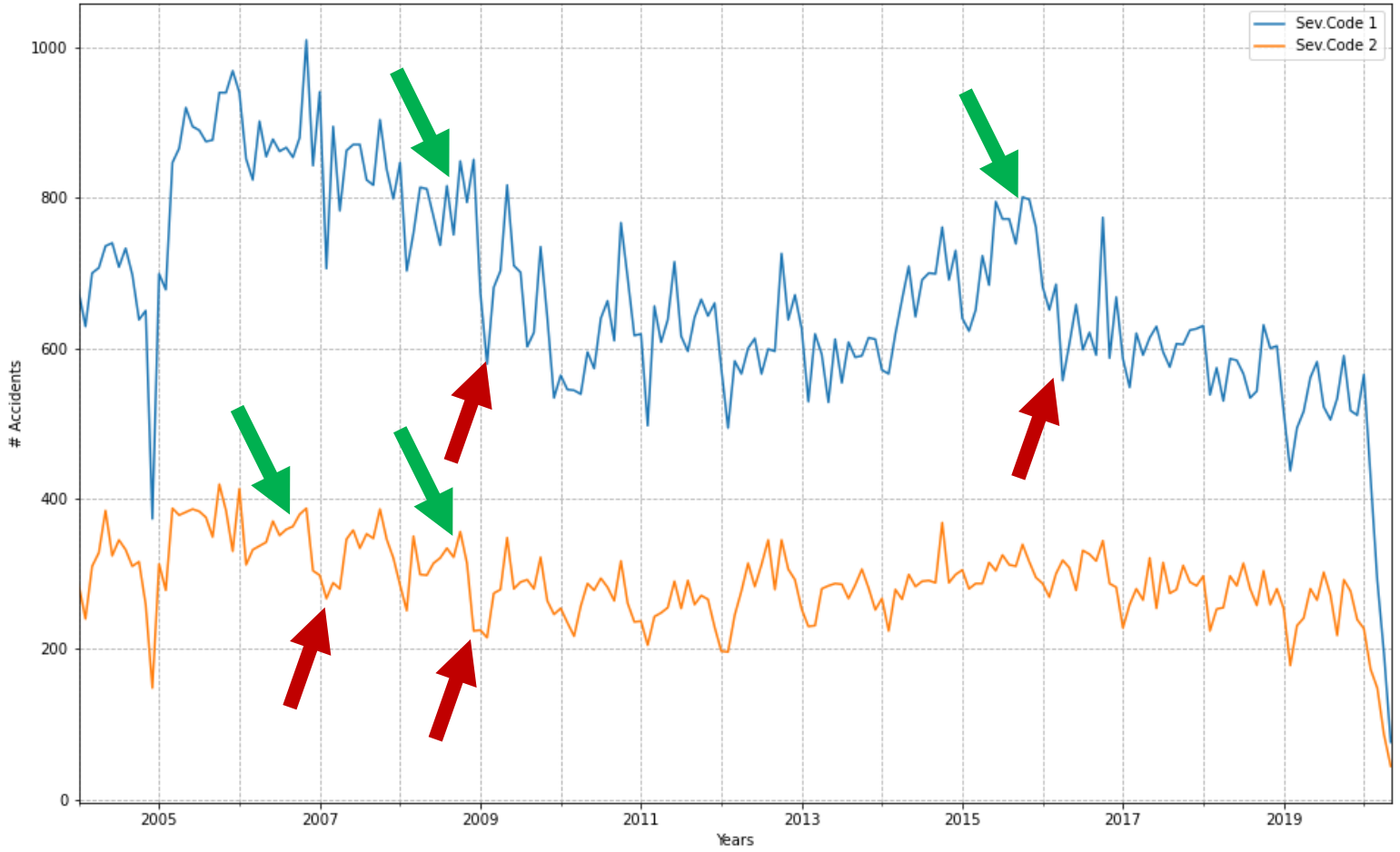
- Accident from Jan-2004 to May-2020 in the Seattle (USA) area.
- Total of 194673 collisions
- Total of 38 attributes
  - Binary severity classification
  - Determine: Relevant and Non-relevant attributes

Data cleaned

- All non-relevant attributes dropped
- Final data set
  - 189339 collisions
    - 132221 Sev.Class 1 & 57118 Sev.Class 2.
  - Attributes remaining
    - 1 Binary severity classification
    - Locations attributes
    - Time related attributes
    - Accident specific attributes
  - Additional attributes created
    - [‘Ratio’] = # Sev.Code 1/ # Sev.Code 2

```
Int64Index: 189339 entries, 0 to 194672
Data columns (total 18 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   SEVERITYCODE    189339 non-null  int64
 1   X               189339 non-null  float64
 2   Y               189339 non-null  float64
 3   LOCATION        189339 non-null  object
 4   COLLISIONTYPE   184582 non-null  object
 5   PERSONCOUNT     189339 non-null  int64
 6   PEDCOUNT        189339 non-null  int64
 7   PEDCYLCOUNT     189339 non-null  int64
 8   VEHCOUNT        189339 non-null  int64
 9   INCDATE         189339 non-null  object
 10  INCDTTM         189339 non-null  object
 11  JUNCTIONTYPE    185146 non-null  object
 12  SDOT_COLDESC    189339 non-null  object
 13  UNDERINFL       184602 non-null  object
 14  WEATHER         184414 non-null  object
 15  ROADCOND        184481 non-null  object
 16  LIGHTCOND       184327 non-null  object
 17  HITPARKEDCAR    189339 non-null  object
dtypes: float64(2), int64(5), object(11)
```

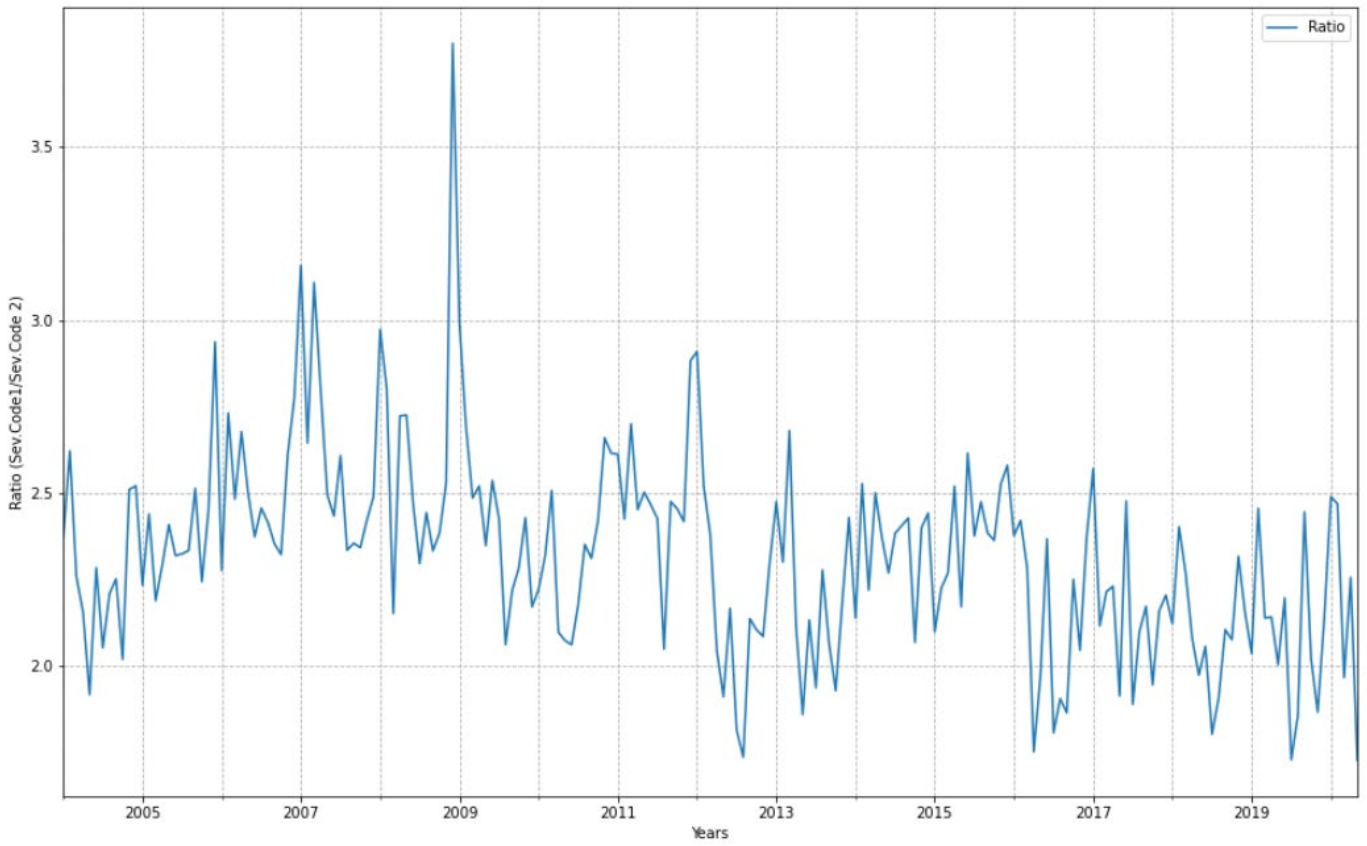
Methodology



Both types of severity over time.

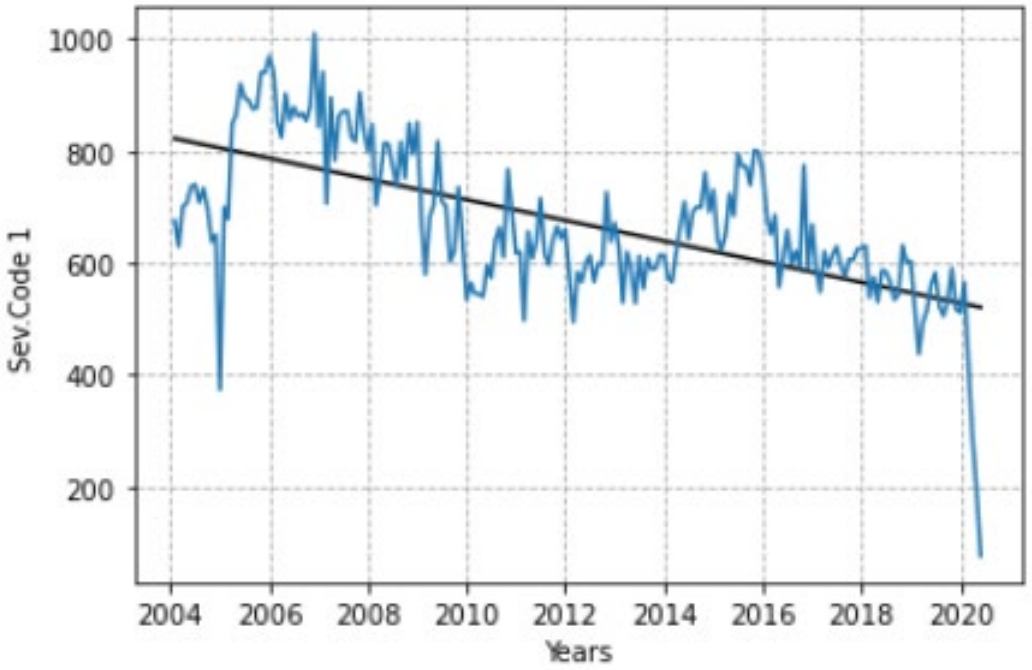
Note the yearly dip in winter compared to summer high of the years

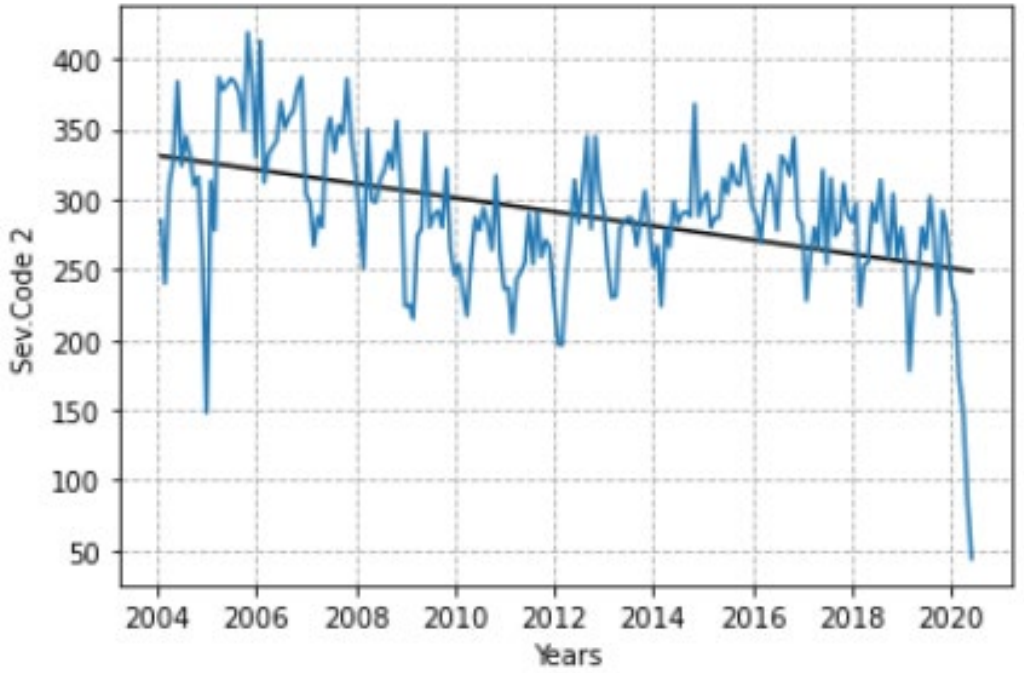
Methodology



Collision ratio of Sev. Code 1 / Sev. Code 2 over time.

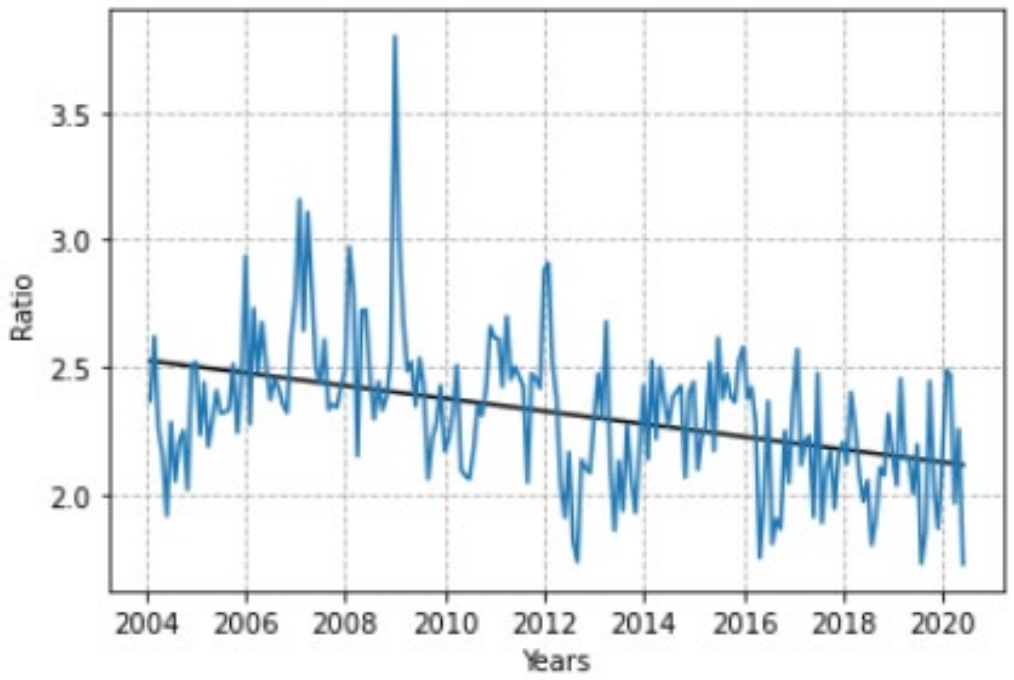
Analysis





Linear Regression of collision cases of Sev. Code 1 (left) and Sev. Code 2 (right) over time.

Analysis

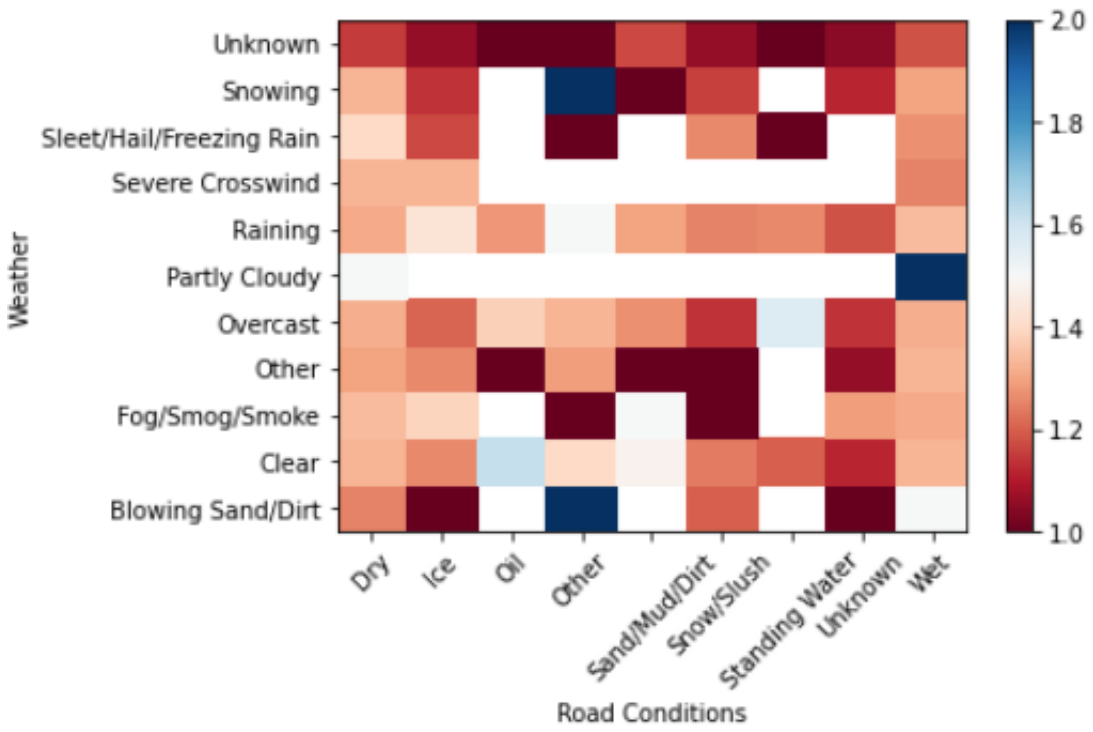


Linear Regression of collision cases ratio of Sev. Code 1 / Sev. Code 2 over time.

Discussion

- There is a decrease of total amount of cases over time
  - Injuries didn’t decrease as much collisions, as the ratio shows.
- A pattern of summer/winter highs and lows needs to further investigated.
- The OBJECT parameters can be further explored with grouping.
  - Example with “Weather” & “Road Conditions”

Discussion



Heat map of the grouping with 2 OBJECT attributes as example for further exploration.

‘Partly Clouded’ & ‘Wet’ has a blue value of 2, indication that with these conditions only severity code 2 accidents occur. Whilst ‘Partly Clouded’ and all other Road Condition parameters show a white value of 1.5 indicating a 1/1 ratio between code 1 and code 2 accidents.

Conclusion

- Much more can be investigated, but wasn't possible due to time limits.
- The creation of additional parameters, eg ['Ratio'], is useful,
- From this parameter an indicator is the focus should shift more to injuries response than collateral damage response
- The OBJECT parameters can be further explored with grouping.
- The location parameters should be further explored with FOLIUM.