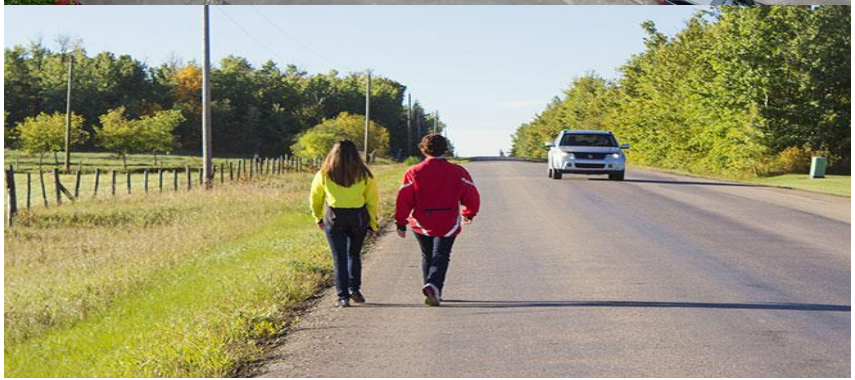
Road Fatalities Evaluation and Prediction

HSO

Outline

- Contextualization
- Problem
- Objectives
- Dataset
- Exploratory Analysis
- Association Rules
- Accident Forecasting
- Accident Patterns
- Discussion
- Conclusions

Contextualization



Statistical Release 8 November 2018

Department for Transport

Reported road casualties in Great Britain: quarterly provisional estimates year ending June 2018

About this release

This publication provides an estimate of the number of personal-injury road traffic accidents in Great Britain that were reported to the police for the year ending June 2018. It also includes the number of people killed or injured in these accidents and which road user group they were in. These figures are provisional as they might change following the end of year validation process.

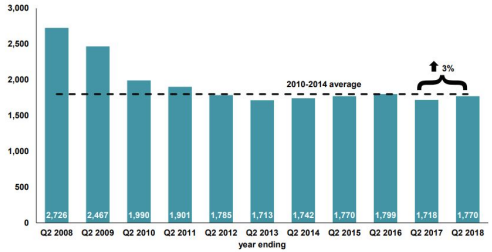
Definition

Casualty: A person killed or injured in an accident. Casualties are sub-divided into killed, seriously injured and slightly injured.

Rolling year: a period of 12 months that begins and ends on a set day. In this publication the rolling year ending

There were 1,770 road deaths in the year ending June 2018. This is a similar level to that seen since 2012.

Road deaths: GB, rolling years ending June, 2008-2018



Problem

- Accident Causes are not Immediately Evident.
  - While fatal accidents can occur in many places, some location are more prominent to occur
  - However identify the causes the lead to the occurrence of fatal accidents is important
  - Several categories of accidents occur daily, but hidden pattern may exist and are hard to identify in useful time.
- Huge Demand in Data Analysis
  - Mining road accident to establish relations of causality requires a fair knowledge of the cases
  - Immediate conclusions while can be obtained in loco, are sometimes hard to determine offline.
- Incorporate side information
  - Side information can be useful to sustain the conclusions.
- Fatal Accidents Prediction
  - Traffic has own dynamics, keep track of modifications and details that lead to fatal accident are hard

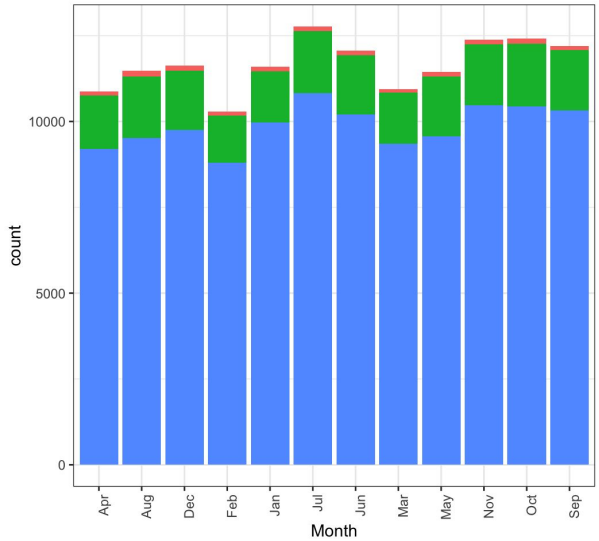
Objectives

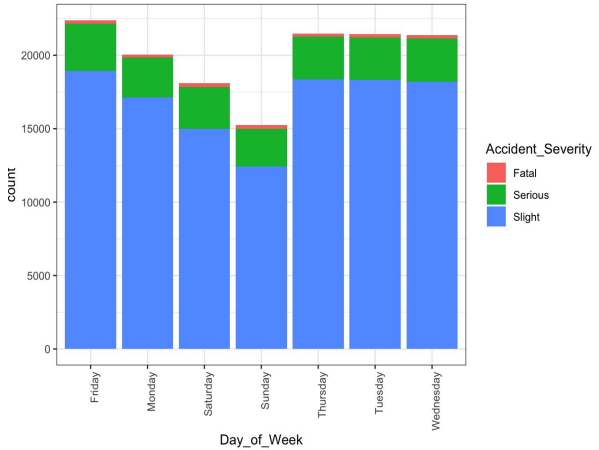
- Road Accident Analysis and Forecasting of Fatal Crashes.
  - Accidents are due to a variate number of factors.
  - Identify the prominent factors that lead to the occurrence of fatal accidents
  - Forecast the number of accident given a set of features
  - Identify patterns of accident and provide to authorities relevant processed information for accident mitigation
- Handle Extreme Class Imbalance and Noisy Information
  - Accident reports frequently contain incomplete information that poses difficulties for a deep analysis
  - While common patterns are easily find, abnormal ones are more rare and sometimes relevant
- Provide an Deep Analysis to Support Our Developments
  - Exploratory analysis by means of visualization
  - Mining Strategies by exploring association rules to find relations of causality
  - ML modeling for forecast number of
  - Clustering to detect emerging behaviours

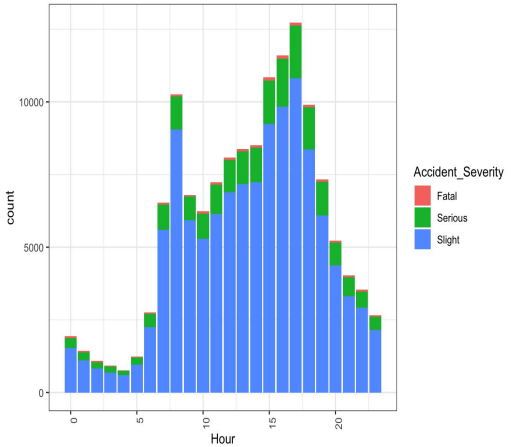
Dataset

- Dataset Encompasses Several Features.
  - Road Safety Data provided by the UK Road Administration department,
  - Containing road safety data about the circumstances of personal injury road accidents in GB, the types of vehicles involved, and the consequential casualties combined with relevant discriminative variables.
  - Additionally, other auxiliary files contain the code mapping of the several variables in the dataset and some supplemental information. 140k entries
- A Brief Glimpse of the variables.
  - Accident Index & char\\
  - Location Easting OSGR & int \\
  - Location Northing OSGR & int\\
  - Longitude & num \\
  - Latitude & num \\
  - Police Force & int\\
  - Accident Severity & int \\
  - Number of Vehicles & int \\
  - Number of Casualties & int \\
  - Date & char\\
  - Day of Week & int\\
  - Time & char\\
  - Local Authority Highway & char\\
  - X1st Road Class & int\\
  - X1st Road Number & int \\
  - Road Type & int\\
  - Speed limit & int\\
  - Junction Detail & int \\
  - Junction Control & int\\
  - Pedestrian Crossing Human Control & int\\
  - Pedestrian Crossing Physical Facilities & int\\
  - Light Conditions & int\\
  - Weather Conditions & int\\
  - Road Surface Conditions & int\\
  - Special Conditions at Site & int\\
  - Carriageway Hazards & int\\
  - Urban or Rural Area & int \\
  - Did Police Officer Attend Scene of Accident & int \\
  - LSOA of Accident Location & char \\

Exploratory Analysis



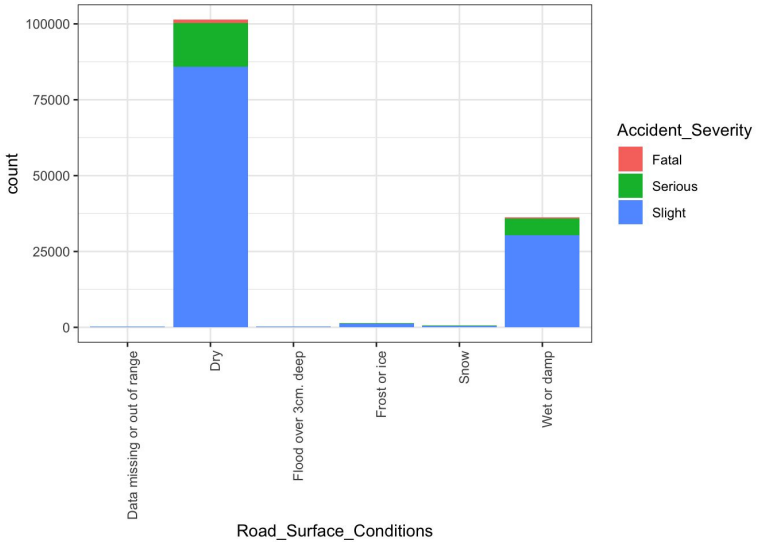


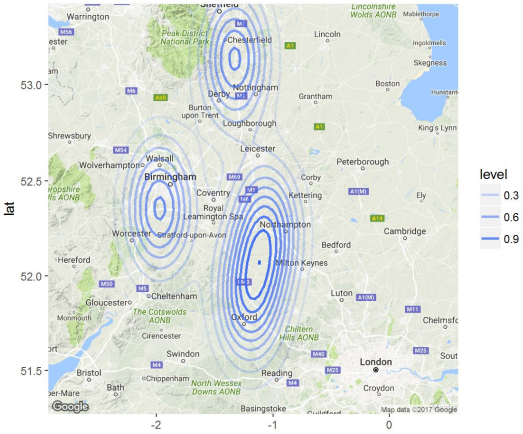


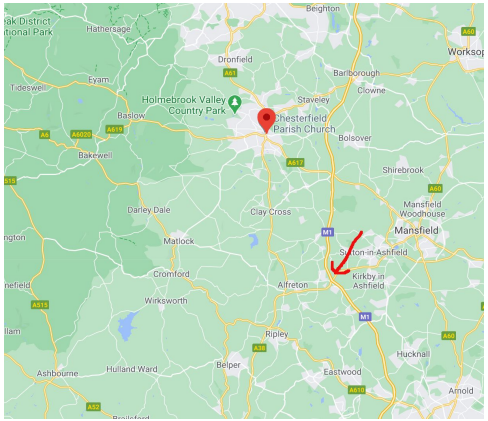
Observations:

* Sunday account for a reduced number of accident and fatalities (common).
* Afternoon accounts for many accidents, congestions and reduced visibility.
* No clear evidence for monthly abnormal patterns.

Exploratory Analysis



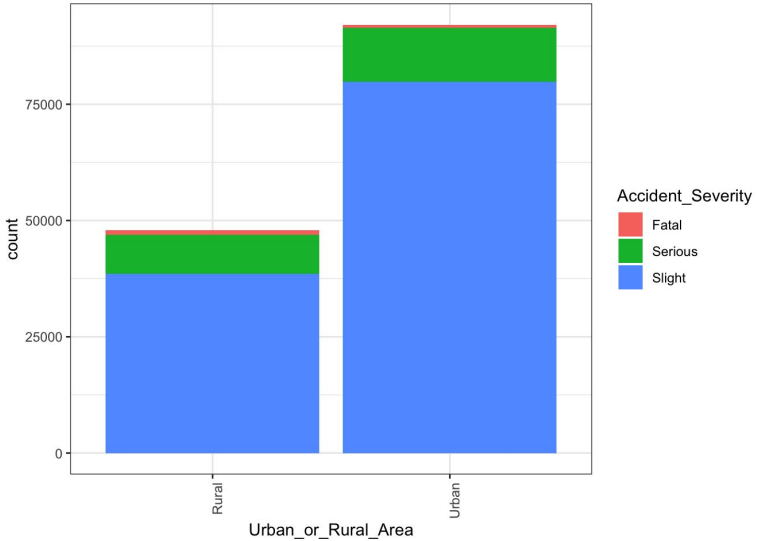


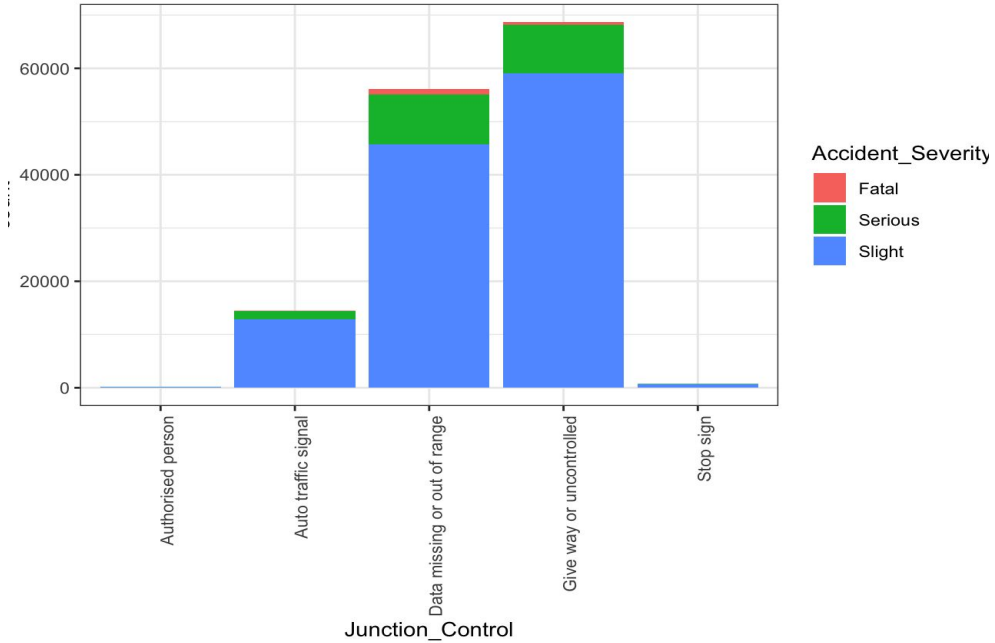


Observations:

* Most accidents occur in dry condition.
* Some locations account for large number of vehicles involved.
* Busy highway that connect major cities entry and exit points.

Exploratory Analysis

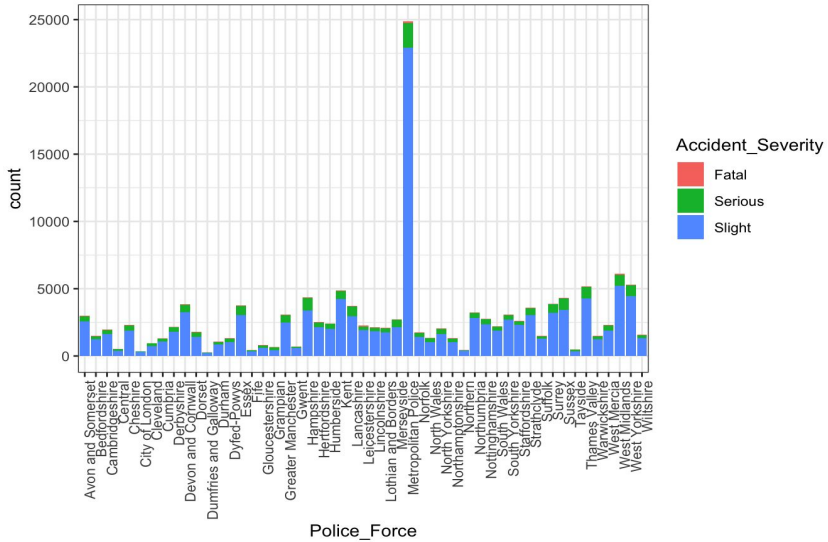


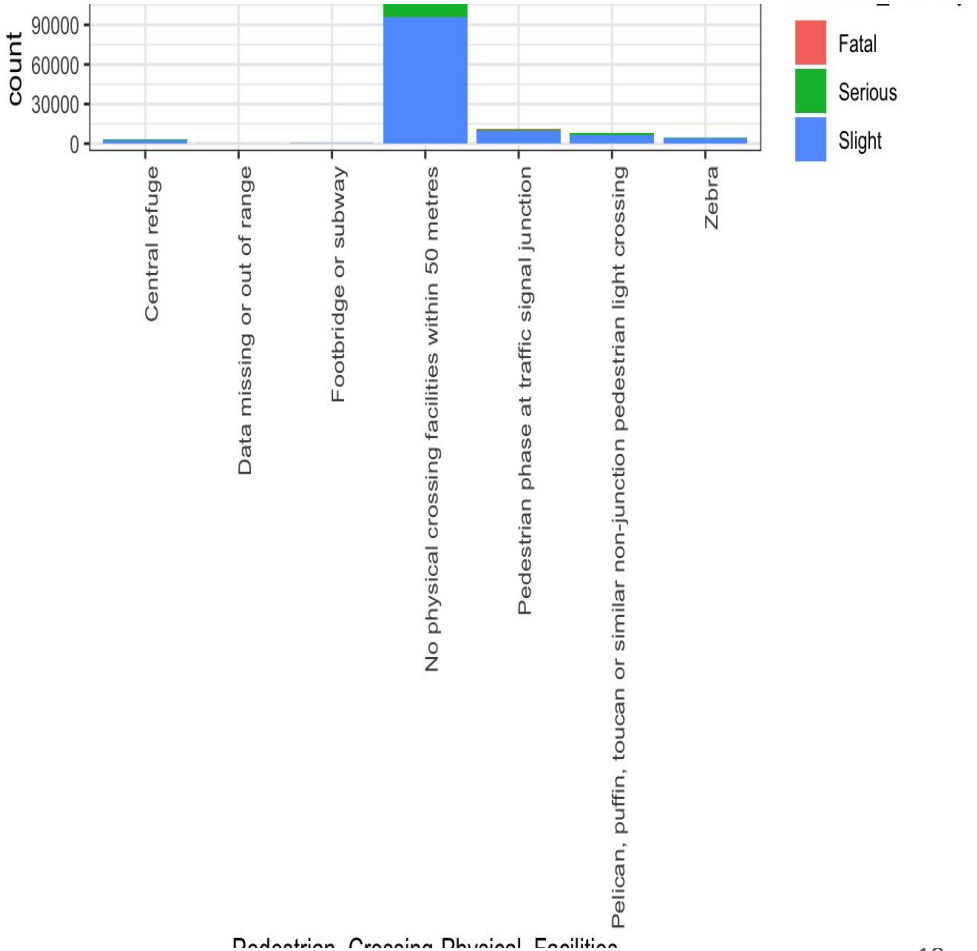


Observations:

* Most accidents occur urban scenarios, but rural exhibit a higher incidence in fatal cases.
* Most accidents occurs near crossing where precedence must be given and no semaphores.
* The majority respect the STOP sign

Exploratory Analysis





Observations:

* Police is present in major urban areas
* No physical crossing for pedestrians have a huge impact

Association Rules

```
## [2] {Pedestrian_Crossing-Physical_Facilities=No physical crossing facilities within 50 metres} => {Casualties_
Class=Large} 0.8605183          1    1
```


```
## [2] {Pedestrian_Crossing-Physical_Facilities=No physical crossing facilities within 50 metres} => {Casualties_
Class=Large} 0.8605183          1    1
```


```
## [5] {Road_Type=Single carriageway}                                                => {Accident_Se
verity=Fatal} 0.7493812          1    1
## [6] {Pedestrian_Crossing-Physical_Facilities=No physical crossing facilities within 50 metres,
##      Weather_Conditions=Fine no high winds}                                       => {Accident_Se
verity=Fatal} 0.7054455          1    1
```

Observations:

* Lack of crossing facilities have a huge impact
* Apparently rough weather don't have a impact on the fatalities occurrence.

Association Rules

```
##     lhs                                                                                                  rhs
support confidence lift
## [1] {}                                                                                                => {Accident_Se
verity=Fatal} 1.0000000          1    1
## [2] {Pedestrian_Crossing-Physical_Facilities=No physical crossing facilities within 50 metres} => {Accident_Se
veritv=Fatal} 0.8620050          1    1
```

```
##     lhs                                                                                                    rhs
support confidence lift
## [1] {}                                                                                                  => {Casualties_
Class=Large} 1.0000000          1    1
## [2] {Pedestrian_Crossing-Physical_Facilities=No physical crossing facilities within 50 metres} => {Casualties_
Class=Large} 0.8605183          1    1
## [3] {Weather_Conditions=Fine no high winds}                                                             => {Casualties_
Class=Large} 0.8170732          1    1
```

Observations:

* Lack of crossing facilities have a huge impact
* Apparently rough weather don't have a impact on the fatalities occurrence.

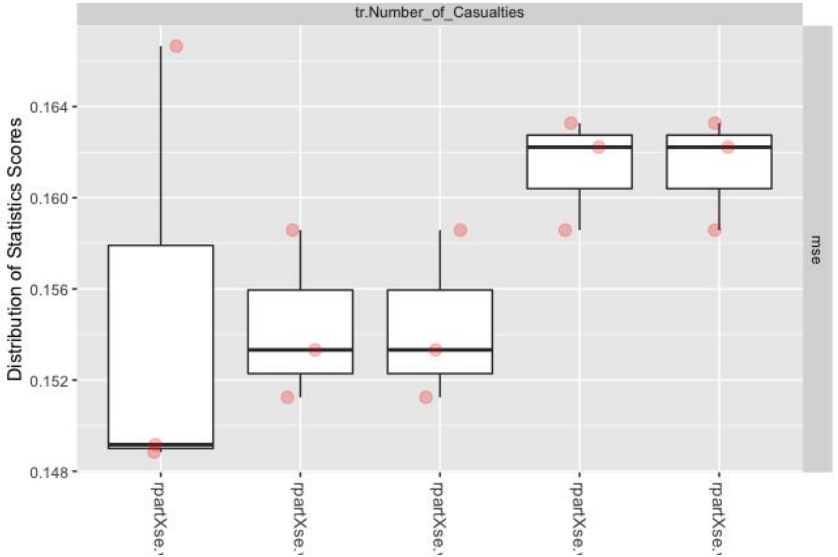
Predictive Modeling

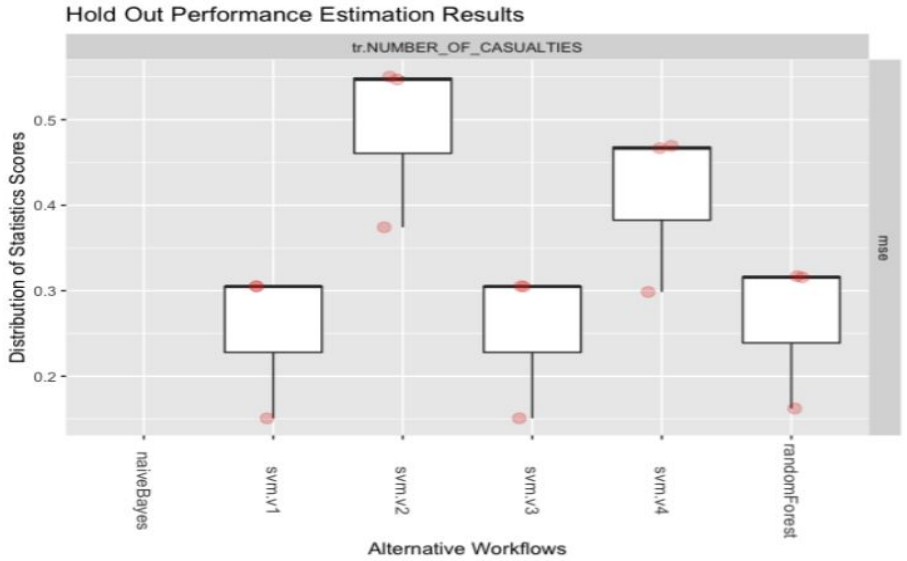
- Longitude \\
- Latitude \\
- Police Force \\
- Accident Severity \\
- Number of Vehicles \\
- Number of Casualties \\
- Police Force \\
- Number of Vehicles \\
- Number of Casualties (target)\\
- Local Authority District \\
- First Road Class \\
- Road Type \\
- Speed limit \\
- Junction Detail \\
- Junction Control \\
- Pedestrian Crossing-Human Control \\
- Pedestrian Crossing-Physical Facilities \\
- Light Conditions \\
- Weather Conditions \\
- Road Surface Conditions \\
- Special Conditions at Site \\
- Carriageway Hazards \\
- Urban or Rural Area\\
- Did Police Officer Attend Scene of Accident \\
- Season \\
- Month \\
- Hour \\
- Day Period \\
- Casualties Class \\
- Number Vehicles Class\\

Observations:

* Longitude and latitude were summarized
* Columns with too many NA were removed
* correlation filter selection, to obtain uncorrelated features
* Variables with too many categories were aggregated

Predictive Modeling

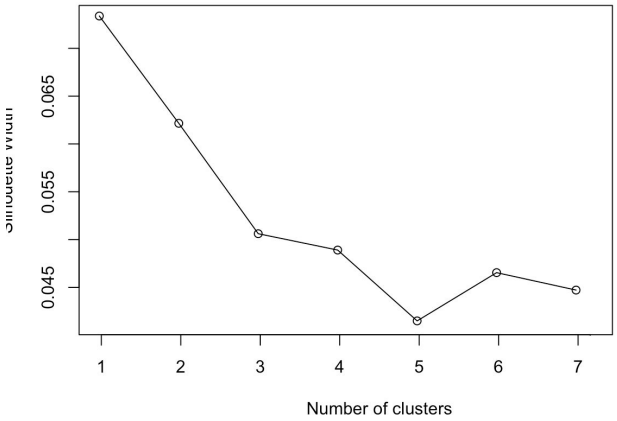




Observations:

* 70/30 split in 140K entries
* Several models and hyperparameters evaluated.
* RF obtained best with a MSE of 0.2

Unsupervised Analysis



Observations:

* Elbow Suggest the existence of 5 Clusters
* While some cluster are evident, (urbar, pedestrian), other are bit misleading (large center hard to separate cluster)

Conclusions

- *Many other experiments we conducted, but in this presentation only the most prominent were showed*
- *Association rules found many relation, but most of then are obvious and don't convey relevant information that lead to occurrence of fatal accidents*
- *Exploratory analysis corroborated the findings of association rules*
- *Predictive modeling showed some promising results, but more evaluations should be made*
- *Unsupervised cluster has potential, using spectral of hierarchical clustering can be useful.*