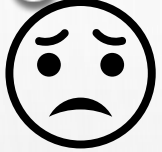


My Cool Capston Project:

# Road collisions in Montreal - Canada

Problem description:



Predicting road accidents is quite uncertain because the relationship between the multiple factors causing an accident is undetermined.



If we could predict traffic accidents, corrective actions could be taken to diminish avoid them.



The good news is that we have data describing the characteristics of more than 190000 accidents since 2012.



The difficulty is that influencing factors are just too many and complex to analyze them by simple observation.



The proposed solution is to build a prototype machine learning model to predict road accidents and its severity, given at least weather and road conditions.

# My Cool Capston Project:

## Road collisions in Montreal - Canada

### Data description

The data-set has 190552 entries, each with 68 attributes

One of the attributes is the severity of the collision, it has the following values,

Value ID	Label	Entries
1	Minor material damages	78300
2	Major material damages	71670
3	Minor injured persons	38831
4	Major injured persons	1543
5	At least one fatality within the next 30 days	208

The number of entries seems to be enough to use the severity attribute to train and test the machine learning model, for further prediction

## As suggested, the road condition attribute is analyzed first

The road condition attribute has the following entries

Condition ID	Description
11	Dry
12	Humid
13	Water accumulation
14	Sand over the road
15	Melt snow
16	Snow
17	Hard snow
18	Icy
19	Muddy
20	Oily
99	Other

## Data assessment:

The correlation indexes to severity are:

Severity	Correlation
1	-0,18
2	-0,18
3	-0,2
4	-0,18
5	-0,85

Meaning that the linear relationship between severity and the road condition is not good.

## Then, the weather condition attribute

The weather condition attribute has the following entries

Condition ID	Description
11	Clear
12	Partially cloudy
13	mist
14	Light Rain
15	Heavy rain
16	Heavy wind

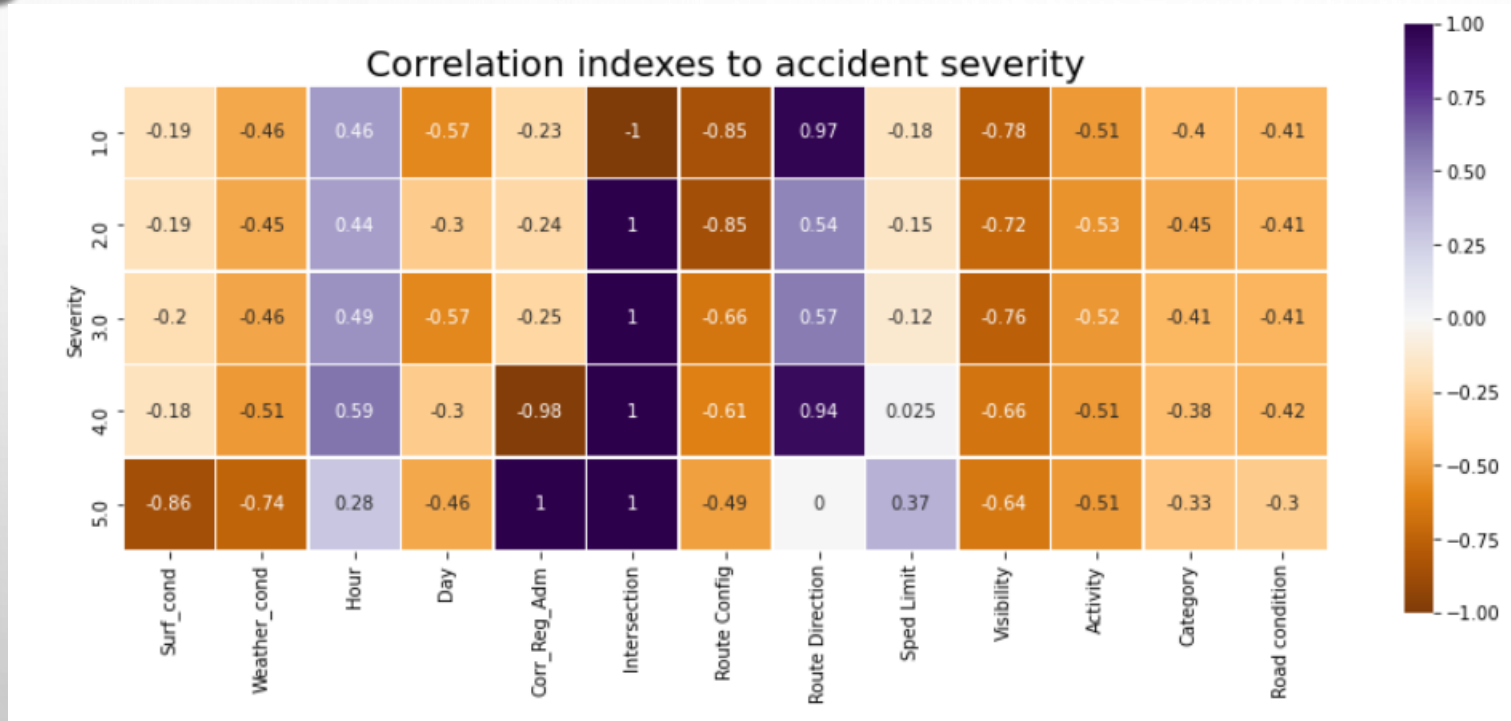
## Data assessment:

The correlation indexes to severity are:

Severity	Correlation
1	-0,46
2	-0,45
3	-0,46
4	-0,50
5	-0,73

Meaning that the linear relationship between severity and the weather condition is fair.

## Finding attributes with better correlation indexes to accident severity



The correlation heat map shows that there are attributes with better linear relationship, these are shown at the center with darker color.

It is worth to make a richer predictive model.