Introduction / Business Problem

Business Problem:
What is the probability that an road accident will happen?
And how severe will it be?
Are certain conditions (weather, daytime etc) more likely to result in an accident?
How can i improve the forecast to guarantee the neccessary precautions?
A better forecast will help to save lives, because the ambulance will know how to react to a certain type of accident and how many employees they will need in their different departments.
Departments near a location where many severe accidents will happen need more people and other kinds of equipment than a department near a location where only minor accidents happen.
Help should arrive as quickly as possible.

Target Audience:

Ambulance
Fire Departments
Police

Aim:

Reduce the time the ambulance will need to reach a severe accident with the right equipment available
Improve the efficiency of the emergency services
We need to know where most of the severe accidents happen and what causes this accidents

Data

The Dataset contains all types of collisions in Seattle ranging from 2004 to present.
It is provided by The SDOT Traffic Management Division and the Traffic Records Group and they update this data weekly.

Data location: Data
Data Description: Data Description

The CSV File has 38 columns.

```
['SEVERITYCODE', 'X', 'Y', 'OBJECTID', 'INCKEY', 'COLDETKEY', 'REPORTNO',
 'STATUS', 'ADDRTYPE', 'INTKEY', 'LOCATION', 'EXCEPTRSNCODE',
 'EXCEPTRSNDESC', 'SEVERITYCODE.1', 'SEVERITYDESC', 'COLLISIONTYPE',
 'PERSONCOUNT', 'PEDCOUNT', 'PEDCYLCOUNT', 'VEHCOUNT', 'INCDATE',
 'INCDTTM', 'JUNCTIONTYPE', 'SDOT_COLCODE', 'SDOT_COLDESC',
 'INATTENTIONIND', 'UNDERINFL', 'WEATHER', 'ROADCOND', 'LIGHTCOND',
 'PEDROWNOTGRNT', 'SDOTCOLNUM', 'SPEEDING', 'ST_COLCODE', 'ST_COLDESC',
 'SEGLANEKEY', 'CROSSWALKKEY', 'HITPARKEDCAR'],
```

The File has 194673 entries.

```
RangeIndex: 194673 entries, 0 to 194672
```

Missing values:

SEVERITYCODE	0
X	5334
Y	5334
OBJECTID	0
INCKEY	0
COLDETKEY	0
REPORTNO	0
STATUS	0
ADDRTYPE	1926
INTKEY	129603
LOCATION	2677
EXCEPTRSNCODE	109862
EXCEPTRSNDESC	189035
SEVERITYCODE.1	0
SEVERITYDESC	0
COLLISIONTYPE	4904
PERSONCOUNT	0
PEDCOUNT	0
PEDCYLCOUNT	0
VEHCOUNT	0
INCDATE	0
INCDTTM	0
JUNCTIONTYPE	6329
SDOT_COLCODE	0
SDOT_COLDESC	0
INATTENTIONIND	164868
UNDERINFL	4884
WEATHER	5081
ROADCOND	5012
LIGHTCOND	5170
PEDROWNOTGRNT	190006
SDOTCOLNUM	79737
SPEEDING	185340
ST_COLCODE	18
ST_COLDESC	4904
SEGLANEKEY	0
CROSSWALKKEY	0
HITPARKEDCAR	0

As we can see there are multiple columns with missing values.
I will clean or drop this columns.
I first thought that I could implement the SPEEDING column into my model but with 185340 missing values it is not a good fit.

Important columns i will use to build my machine learning algorithm:

INCKEY: Unique Key for the Incident
SEVERITY CODE: 1 or 2 dependent on the type oft he severity
WEATHER: Weather Conditions
ROADCOND: Road Conditions
LIGHTCOND: Light Conditions
INCDTTM: Date and Time of the incident
X and Y: Latitude and longitude