**Busines Understanding**

As a hospital director that plans resources you want to predict risk and severity of accidents on streets to help people that were victims of car accidents. As a head of police department you want to know how to protect citizens and patrol the streets so to limit dangerous behaviour on the streets. Aim of this project is to predict the severity of road accident so to limit the number of incidences on the way and optimize resources to protect victims. To understand this problem I will use Seatle incident data that uses metrics such as weather and road condition from 2004 to present.

**Hypothesis**

Some features may be more important than others to predict severity. The questions that I would like to ask are:

*Primary*:

**H0**: If LIGHCOND is related to SEVERITYCODE then as the more light there is the less severe the accidents are

**H0**: If SPEEDING cause more severe incidents then number of incidents recorded as fatality will be bigger when speeding

**H0**: If SPEEDING UNDERINTL increases the number of incidents then collisions under the substance are more severe

*Secondary questions*:

1. Are there particular seasons that increase the severity?
2. What was the biggest accident recorded so far?
3. What percentage of collision involve bicycles?
4. What percentage of collisions involved pedestrians?
5. What kind of collisions do we have the most?
6. One locations more dangerous than the other?
7. Is is more likely to have incident on the intersection of the roads?
8. What is distribution of severity?

**INFO about the dataset**

In the raw dataset there are 38 columns, for full list please refer to metadata. The aim is to predict severity of the accident using SeverityCode.

We have 3 unique identifiers, that can be removed: OBJECTID, INCKEY, COLDETKEY. We don't need columns: X, Y, SDOTCOLNUM and the SEVERITYCODE.1 is the same as SEVERITY CODE

In the columns: EXCEPTRSNDESC, INATTENTIONIND, PEDROWNOTGRNT, SPEEDING is more than 80% of missing values so we want to remove those. The columns that I will use for the project are:

