# ACCIDENT SEVERITY DUE TO EXTERNAL FACTORS

### BACKGROUND

In 2013, 54 million people worldwide sustained injuries from traffic collisions. This resulted in 1.4 million deaths in 2013, up from 1.1 million deaths in 1990. About 68,000 of these occurred in children less than five years old.

According to Florida Safety. Org, the five environmental forces that influence/effect driver performance and thereby leading to the impact of the collision are-

1. Darkness.
2. Weather
3. Unfamiliar Terrain
4. Special Traffic Controls

According to road safety organization, Victoria- When road crashes are investigated, in the majority of cases it is the combination of a number of factors that have caused the crash. It is suggested that in about 95 per cent of motor vehicle crashes human factors alone, or in combination with one or more other factors, are the major contributors.

Factors that can contribute to a road crash:

**Human factors**– the behaviour of the people involved – such as: deliberate risk taking; distraction (mobile phones, music, friends); tiredness (resulting in poor concentration); driving while affected by alcohol or drugs; lateness (resulting in speeding); disregard for road rules; choosing the unsafe option (such as a pedestrian crossing between moving vehicles)

**Vehicle factors** – features of the vehicle involved – such as poorly maintained vehicle; lack of modern safety features (such as lane assist, brake assist, electronic stability control (ESC), ABS brakes)

**Environmental factors** – features of the road and surrounding area – such as the surface and condition of the road, and the roadside; poor visibility as a result of heavy rain or bright sunshine; wet roads needing people to travel slower so they have more time to come to a stop; speed limits that make pedestrians and cyclists more vulnerable.

Here, our attempt would be to understand how influential an impact does environmental factors play a role in the accident severity.

## Business Problem

Accident Severity prediction is essential as nations, Governments and individuals each want to achieve efficiency in their own domains. Collisions, accidents and damage to property/assets cumulatively are a huge loss for the economy- physically, financially affecting the economy, which strives to achieve and sustain an efficient approach of its resources.

External (which include environmental ) factors, reduce the governments/ corporations efforts to achieve efficiency.

Through the real-time governmental MIS on accident severity, we aim to identify how influential are external/ environmental factors in determining the severity of a collision.

## Data Source

As stated, the reliable government data on accident severity is well documented and maintained in a database which records the traffic information based in Seattle. The sample dataset on collisions can be found out through the link <https://data-seattlecitygis.opendata.arcgis.com/datasets/collisions/data>

### Data Parameters

The data-set comprises of the following fields which gives us a brief on the entire collision brief for the records captured.

|  |  |
| --- | --- |
| Attribute | Description |
| OBJECTID | ESRI unique identifier |
| SHAPE | ESRI geometry field |
| INCKEY | A unique key for the incident |
| COLDETKEY | Secondary key for the incident |
| ADDRTYPE | Collision address type:   * Alley * Block * Intersection |
| INTKEY | Key that corresponds to the intersection  associated with a collision |
| LOCATION | Description of the general location of the  collision |
| SEVERITYCODE | A code that corresponds to the severity of the collision:   * 3—fatality * 2b—serious injury * 2—injury * 1—prop damage * 0—unknown |
| SEVERITYDESC | A detailed description of the severity of the  collision |
| COLLISIONTYPE | Collision type |
| PERSONCOUNT | The total number of people involved in the  collision |
| PEDCOUNT | The number of pedestrians involved in the  collision. This is entered by the state. |
| PEDCYLCOUNT | The number of bicycles involved in the collision.  This is entered by the state. |
| VEHCOUNT | The number of vehicles involved in the collision.  This is entered by the state. |
| INJURIES | The number of total injuries in the collision. This  is entered by the state. |
| SERIOUSINJURIES | The number of serious injuries in the collision.  This is entered by the state. |
| FATALITIES | The number of fatalities in the collision. This is  entered by the state. |
| INCDATE | The date of the incident. |
| INCDTTM | The date and time of the incident. |
| JUNCTIONTYPE | Category of junction at which collision took  place |
| SDOT\_COLCODE | A code given to the collision by SDOT. |
| SDOT\_COLDESC | A description of the collision corresponding to  the collision code. |
| INATTENTIONIND | Whether or not collision was due to inattention.  (Y/N) |
| UNDERINFL | Whether or not a driver involved was under the  influence of drugs or alcohol. |
| WEATHER | A description of the weather conditions during  the time of the collision. |
| ROADCOND | The condition of the road during the collision. |
| LIGHTCOND | The light conditions during the collision. |
| PEDROWNOTGRNT | Whether or not the pedestrian right of way was  not granted. (Y/N) |
| SDOTCOLNUM | A number given to the collision by SDOT. |
| SPEEDING | Whether or not speeding was a factor in the  collision. (Y/N) |
| ST\_COLCODE | A code provided by the state that describes the collision. For more information about these codes, please see the State Collision Code  Dictionary. |
| ST\_COLDESC | A description that corresponds to the state’s  coding designation. |
| SEGLANEKEY | A key for the lane segment in which the collision  occurred. |
| CROSSWALKKEY | A key for the crosswalk at which the collision  occurred. |
| HITPARKEDCAR | Whether or not the collision involved hitting a  parked car. (Y/N) |

#### 4.2 Data Wrangling

As stated, we aim to understand the influence of external factors in determining the severity of a collision. Hence, the following information has been considered keeping the above perspective to hold true.

1. Data inputs attributed **status** as “matched” by the Seattle MIS have only been considered and “unmatched” criterias have been removed.
2. The following attributes have been considered-

|  |  |  |
| --- | --- | --- |
| **Attribute** | **Description** | **Reasoning** |
| ADDRTYPE | Collision address type:  Alley  Block  Intersection | Alleys, blocks and intersections are not under the control of the driver |
| SEVERITYCODE | A code that corresponds to the severity of the collision:  3—fatality  2b—serious injury  2—injury  1—prop damage  0—unknown | The attribute which will be finally predicted using our model |
| COLLISIONTYPE | Collision type | External factors not in control of the driver |
| WEATHER | A description of the weather conditions during  the time of the collision. | External factors |
| ROADCOND | The condition of the road during the collision. | External factors |
| LIGHTCOND | The light conditions during the collision. | External factors |
| SPEEDING | Whether or not speeding was a factor in the  collision. (Y/N) | Speeding depends upon the road, environment which is an external factor |

1. Rows with unknown values in severity code, WEATHER, LIGHTCOND, WEATHERCOND, ROADCOND were dropped
2. Rows with null values in severity code, addrtype, collision type, weather, roadcond, lightcond have been removed/dropped.
3. Speeding blank values have been converted to “0” and SPEEDING ‘Y’ values have been converted to 1
4. Severity Code has been converted to the following numeric codes to convert the dataset into int type.

|  |  |
| --- | --- |
| Severity Code | Description |
| 1 | Property damage |
| 2 | Injury |
| 3 | Fatality |
| 4 | Serious Injury |

1. The categorical values of certain parameters have been converted to the following numeric codes-

|  |  |
| --- | --- |
| ADDRTYPE | ADDRTYPE CODE |
| Block | 1 |
| Intersection | 2 |
| Alley | 0 |

|  |  |
| --- | --- |
| WEATHER | WEATHERCODE |
| Clear | 1 |
| Raining | 6 |
| Overcast | 4 |
| Snowing | 9 |
| Other | 3 |
| Fog/Smog/Smoke | 2 |
| Sleet/Hail/Freezing/Rain | 8 |
| Blowing Sand/Dirt | 0 |
| Severe Crosswind | 7 |
| Partly cloudy | 5 |

|  |  |
| --- | --- |
| LIGHTCOND | LIGHTCONDCODE |
| Daylight | 5 |
| Dark-Street Lights On | 2 |
| Dusk | 6 |
| Dawn | 4 |
| Dark-No street lights | 0 |
| Dark-Street lights off | 1 |
| Other | 7 |
| Dark-Unknown Lighting | 3 |

|  |  |
| --- | --- |
| ROADCOND | ROADCONDCODE |
| Dry | 0 |
| Wet | 7 |
| Ice | 1 |
| Snow/Slush | 5 |
| Other | 3 |
| Standing water | 6 |
| Sand/Mud/Dirt | 4 |
| Oil | 2 |

### Exploratory Data Analysis

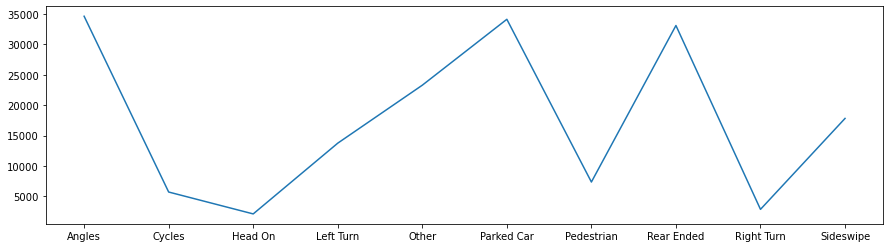
On observing our datasets, we see the above mentioned column values in the Collision dataset

.

Most frequent and recorded collision type are those in

* Angles
* Parked Cars
* Rear Ended Collisions

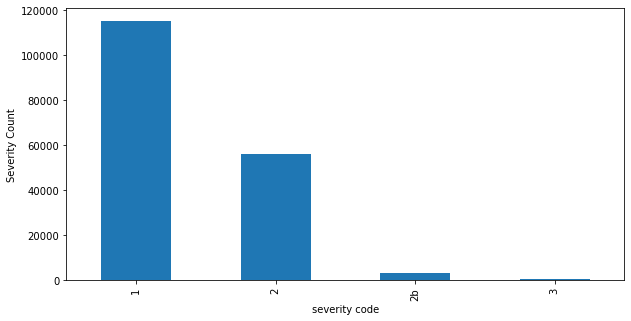
Head-on Collisions are very less



We also understand that from the given dataset, the collisions are more when the roads are dry.



When considering severity, property damage is considered the most common form of the result of collision.



##### Model Development & Evaluation

For approaching the model development phase, the Support Vector Mechanism of Machine Learning model has been used, alongside with K-Nearest Neighbour

The reasons for choosing them are as follows-

1. Clustering of parameters help us in providing a better scope of understanding and prediction
2. We want to negate outliers as much as possible
3. Large Datasets are more accurate while using KNN.
4. Hence, a clustering based approach is used.



| **Algorithm** | **Jaccard** | **F1-score** |
| --- | --- | --- |
| **KNN** | 0.52 | 0.65 |
| **SVM** | 0.47 | 0.58 |

We find the KNN model to be the more responsive to understand the problem, when compared to Support Vector Mechanism.

### Conclusion

In this study, we analyzed the role of external , environmental effects in influencing the probability of a traffic collision. While as already stated, there are many factors playing a role in identification of collision, external environmental influence is also relevant.

The purpose of this study was to understand how high the influence is, amd whether can we predict the collision by only considering environmental factors.

The answer we get is NO.

Even if government, corporations, individuals ensure that the highways and related traffic is streamlined, efficient. They upgrade their systems and adopt the best of practices. Still, there will be collisions as a probability of 50% is dependant on external factors viz. The Spot of collision, The entities involved in collision, the road, weather, lighting conditions, the clarity, ease of roads which leads to speeding.