IBM Data Science Capstone Project – Car Accident Severity

Week 1

Introduction & Busines Understanding:

Car collision are quite common around the world. They cost countries billions of dollars in healthcare cost and property damage and result on thousands of deaths and injuries. To reduce the number of car collision and its severity, a model will be developed to predict the severity of an accident given attributes like weather conditions, car speed, light conditions, and road conditions and several other attributes.

For the data source of this project, I will be using the data from the City of Seattle's police department showing all the car collisions from 2004 to 2020. This model can be used in an application to warn drivers to be more cautious when driving in bad weather and bad road conditions. This model can also be used by city planners to better design roads and to install warning signs and speed bumps in dangerous road intersections.

Data Understanding:

The dataset provided by the City of Seattle's police department had a total of 194,673 collisions from 2004 to 2020. The dataset has 38 attributes describing the details of each car collision including the severity, coordinates, the number of vehicles, date & time, weather & road conditions for each collision. The figures below highlight the initial analysis made on the used dataset.

In [18]:
```
df.head()
```
Out[18]:

	SEVERITYCODE	X	Y	OBJECTID	INCKEY	COLDETKEY	REPORTNO	STATUS	ADDRTYPE	INTKEY	...	ROADCOND	LIGHTCOND
0	2	-122.323148	47.703140	1	1307	1307	3502005	Matched	Intersection	37475.0	...	Wet	Daylight
1	1	-122.347294	47.647172	2	52200	52200	2607959	Matched	Block	NaN	...	Wet	Dark - Street Lights On
2	1	-122.334540	47.607871	3	26700	26700	1482393	Matched	Block	NaN	...	Dry	Daylight
3	1	-122.334803	47.604803	4	1144	1144	3503937	Matched	Block	NaN	...	Dry	Daylight
4	2	-122.306426	47.545739	5	17700	17700	1807429	Matched	Intersection	34387.0	...	Wet	Daylight

5 rows × 38 columns

In [21]:

```
df.describe()
```

Out[21]:

	SEVERITYCODE	X	Y	OBJECTID	INCKEY	COLDETKEY	INTKEY	SEVERITYCODE.1	PERSONCOUNT
count	194673.000000	189339.000000	189339.000000	194673.000000	194673.000000	194673.000000	65070.000000	194673.000000	194673.000000
mean	1.298901	-122.330518	47.619543	108479.364930	141091.456350	141298.811381	37558.450576	1.298901	2.444427
std	0.457778	0.029976	0.056157	62649.722558	86634.402737	86986.542110	51745.990273	0.457778	1.345929
min	1.000000	-122.419091	47.495573	1.000000	1001.000000	1001.000000	23807.000000	1.000000	0.000000
25%	1.000000	-122.348673	47.575956	54267.000000	70383.000000	70383.000000	28667.000000	1.000000	2.000000
50%	1.000000	-122.330224	47.615369	106912.000000	123363.000000	123363.000000	29973.000000	1.000000	2.000000
75%	2.000000	-122.311937	47.663664	162272.000000	203319.000000	203459.000000	33973.000000	2.000000	3.000000
max	2.000000	-122.238949	47.734142	219547.000000	331454.000000	332954.000000	757580.000000	2.000000	81.000000

In [29]:

```
df.dtypes
```

Out[29]:

```
SEVERITYCODE        int64
X                 float64
Y                 float64
OBJECTID            int64
INCKEY              int64
COLDETKEY           int64
REPORTNO           object
STATUS             object
ADDRTYPE           object
INTKEY            float64
LOCATION           object
EXCEPTRSNCODE      object
EXCEPTRSNDESC      object
SEVERITYCODE.1      int64
SEVERITYDESC       object
COLLISIONTYPE      object
PERSONCOUNT         int64
PEDCOUNT            int64
PEDCYLCOUNT         int64
VEHCOUNT            int64
INCDATE            object
INCDTTM            object
JUNCTIONTYPE       object
SDOT_COLCODE        int64
SDOT_COLDESC       object
INATTENTIONIND     object
UNDERINFL          object
WEATHER            object
ROADCOND           object
LIGHTCOND          object
PEDROWNOTGRNT      object
SDOTCOLNUM        float64
SPEEDING           object
ST_COLCODE         object
ST_COLDESC         object
SEGLANEKEY          int64
CROSSWALKKEY        int64
HITPARKEDCAR       object
dtype: object
```

In our model, I will be using (SEVERITYCODE: a code that corresponds to the severity of the collision) as our dependent variable (Y), which will be the target attribute to be predicted**.** As for the independent variables (X), I will be using the following attributes:

- ADDTYPE: Collision address type (Alley, Block, Intersection)
- WEATHER: A description of the weather conditions during the time of the collision
- ROADCOND: The condition of the road during the collision.
- VEHCOUNT: The number of vehicles involved in the collision. This is entered by the state.
- PERSONCOUNT: The total number of people involved in the collision
- COLLISIONTYPE: Collision type
- LIGHTCOND: The light conditions during the collision.