CAR ACCIDENT SEVERITY

Introduction

Seattle is a seaport city on the West Coast of the United States. **Seattle** residents get around by car, trolley, streetcar, public bus, bicycle, on foot, and by rail. With such bustling streets, it's **no** surprise that **Seattle** sees car **accidents every day.** Our main objectives of this project are to analyze the accident data, Predict the chances and severity of an accident through various data science techniques that would eventually help the residents plan their travel more carefully.

DATA

The dataset includes all the data from 2004 to present.

The dataset includes 37 attributes and a separate state collision code dictionary. Other datasets will be obtained from Open Government Data portal and open source research groups.

The unbalanced datasets will be inspected first for proper use. The datasets will allow us to train our ML models and predict the severity of accidents and chances of the same.

```
In [6]: df.describe()
```

Out[6]:

	SEVERITYCODE	X	Y	OBJECTID	INCKEY	COLDETKEY	INTKEY	SEVERITYCODE.1	PERSONCOUNT	PE
count	194673.000000	189339.000000	189339.000000	194673.000000	194673.000000	194673.000000	65070.000000	194673.000000	194673.000000	19467
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	

Methodology

- Once the data has been collected and analysed. I Started with inspecting and cleaning through the following ways

- For my model to be unbiased and give accurate results I balanced the dataset and co-related the Severity index with conditions provided such as Weather, Lighting, Road using K-means approach

Results