



IBM Data Science

CAR ACCIDENT SEVERITY

CAPSTONE PROJECT

ARUN CHAUHAN

19th September 2020

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

To achieve our goals we will be using Data Collision data (Data-collisions.csv)
All the data is recorded by Traffic control and SPD. This includes all types of collisions.
Collisions will display at the intersection or mid-block of a segment.

Out[5]:

	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
...
194668	2	-122.290826	47.565408	219543	309534	310814	E871089	Matched	Block	NaN	...	Dry	Daylight
194669	1	-122.344626	47.690924	219544	309085	310365	E875731	Matched	Block	NaN	...	Wet	Daylight
194670	2	-122.306689	47.683047	219545	311280	312640	3809984	Matched	Intersection	24760.0	...	Dry	Daylight
194671	2	-122.355317	47.678734	219546	309514	310794	3810083	Matched	Intersection	24349.0	...	Dry	Dusk
194672	1	-122.289360	47.611017	219547	308220	309500	E868008	Matched	Block	NaN	...	Wet	Daylight

194673 rows × 38 columns

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.