

IBM Coursera Capstone

INTRODUCTION/BUSINESS PROBLEM

It is extremely essential for proper understanding of public safety when it comes to constructing roads by the municipal corporation or any other private entity. If the officials have access to various meaningful insights regarding these road accidents, the future construction of roads could be done in a manner that would ultimately lead to a safer and more seamless experience for the public. After studying and fabricating trends from the previous years, these entities will be at a much better position to make decisions which greatly benefit the public and the corresponding stakeholders. For the existing roads, these organisations can also put up various signs at strategic areas to further alert the civilians. Various infographics related to these car accidents could be issued in public interest to further alert the civilians regarding this and hence spread awareness.

We know how effective Machine Learning is when it comes to predicting/classifying based on some previous trends. By using some of the machine learning models I plan to contribute towards the safety of the civilians and elaborate various factors that go into a road accident.

Using inferences from the previous trends we can alert the public with some key findings and thus make them more careful towards car accidents thereby reducing it.

So my business problem aims to aid the road-building organisations to be more aware and educated about the car accidents before constructing newer roads in the city of Seattle so that these accidents don't repeat as often. It will compel the officials to strategically come up with various junction types to reduce the number of accidents accordingly.

This is no way is restricted only to Seattle, since all the cities that have roads similar to Seattle can take some inferences from this work as well (with a few modifications of course).

DATA DESCRIPTION

The dataset that I am going to be working with is a Collision dataset that records various factors when an accident takes place at different locations in the city of Seattle. These accidents have been recorded since the year 2004. The data for analysis was retrieved from the Road Accident Severity Data from the Seattle State Department of Transport from Data-Collisions. It is a CSV(Comma Separated Value) file that contains 194673 rows and 38 columns.

There are 37 other attributes that are a mixture of text and numbers, i.e., both categorical and numerical data types are present. The label is chosen to be the "accident severity" and is encoded as follows:

- 3—fatality
- 2b—serious injury
- 2—injury
- 1—prop damage
- 0—unknown

Here is a more detailed version description of the dataset as a whole:

```
In [11]: df.shape
Out[11]: (194673, 38)

In [16]: df.describe()
Out[16]:
```

	SEVERITYCODE	X	Y	OBJECTID	INCKEY	COLDKETKEY	INTKEY	SEVERITYCODE.1	PERSONCOUNT	PEI
count	194673.000000	189339.000000	189339.000000	194673.000000	194673.000000	194673.000000	65070.000000	194673.000000	194673.000000	194673.000000
mean	1.298901	-122.330518	47.619543	108479.364930	141091.456350	141298.811381	37558.450576	1.298901	2.444427	1.946730e+05
std	0.457778	0.029976	0.056157	62649.722558	86634.402737	86986.542110	51745.990273	0.457778	1.345929	1.946730e+05
min	1.000000	-122.419091	47.495573	1.000000	1001.000000	1001.000000	23807.000000	1.000000	0.000000	0.000000e+00
25%	1.000000	-122.348673	47.575956	54267.000000	70383.000000	70383.000000	28667.000000	1.000000	2.000000	2.000000e+00
50%	1.000000	-122.330224	47.615369	106912.000000	123363.000000	123363.000000	29973.000000	1.000000	2.000000	2.000000e+00
75%	2.000000	-122.311937	47.663664	162272.000000	203319.000000	203459.000000	33973.000000	2.000000	3.000000	3.000000e+00
max	2.000000	-122.238949	47.734142	219547.000000	331454.000000	332954.000000	757580.000000	2.000000	81.000000	8.100000e+01

Other attributes:

```
In [11]: df.shape
Out[11]: (194673, 38)

In [16]: df.describe()
Out[16]:
```

INTKEY	SEVERITYCODE.1	PERSONCOUNT	PEDCOUNT	PEDCYLCOUNT	VEHCOUNT	SDOT_COLCODE	SDOTCOLNUM	SEGLANEKEY	CROSSWALKKEY
0.000000	194673.000000	194673.000000	194673.000000	194673.000000	194673.000000	194673.000000	1.149360e+05	194673.000000	1.946730e+05
8.450576	1.298901	2.444427	0.037139	0.028391	1.920780	13.867768	7.972521e+06	269.401114	9.782452e+03
5.990273	0.457778	1.345929	0.198150	0.167413	0.631047	6.868755	2.553533e+06	3315.776055	7.226926e+04
7.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.007024e+06	0.000000	0.000000e+00
7.000000	1.000000	2.000000	0.000000	0.000000	2.000000	11.000000	6.040015e+06	0.000000	0.000000e+00
3.000000	1.000000	2.000000	0.000000	0.000000	2.000000	13.000000	8.023022e+06	0.000000	0.000000e+00
3.000000	2.000000	3.000000	0.000000	0.000000	2.000000	14.000000	1.015501e+07	0.000000	0.000000e+00
0.000000	2.000000	81.000000	6.000000	2.000000	12.000000	69.000000	1.307202e+07	525241.000000	5.239700e+06