

# Reducing Accidents

Pradeep Saravana

September 15 2020

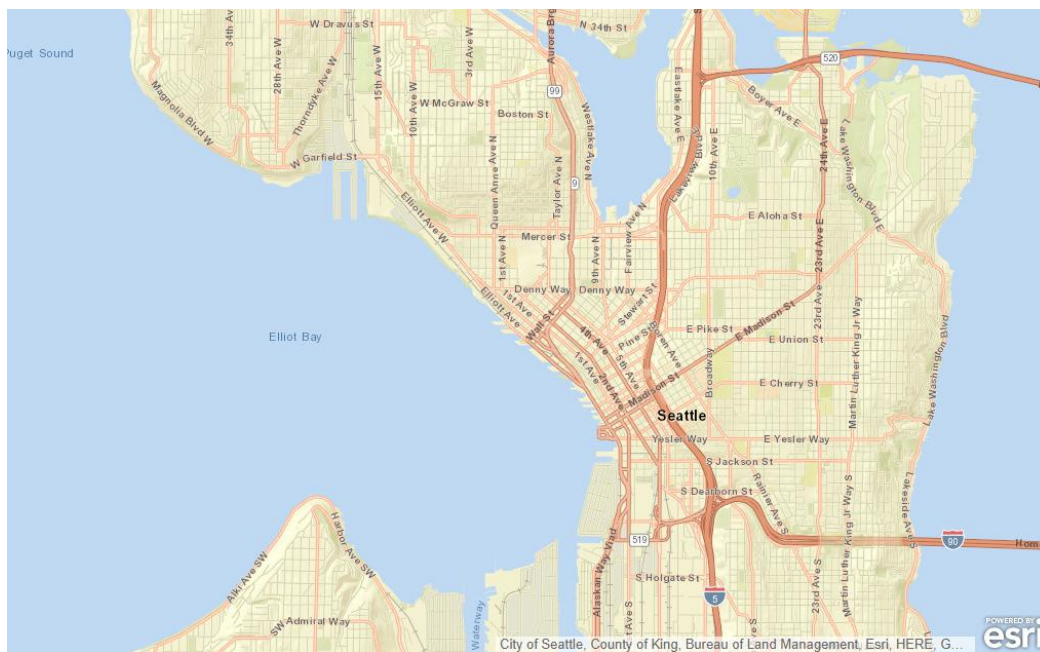
## 1. Introduction

### 1.1 Background

With the increase of population comes the need for increased transportation facilities. Fast paced lifestyle of public demands for shorter commute time. Governments however have constraints and may never be able to satisfy each and every one of its citizen on their travel needs each and every time a need arises. Having a personal vehicle is a straightforward and a quick fix to a flexible transport solution. Despite being expensive compared to public transport, it gives its owner round the clock service on any public road.

Developments in engineering and other technology along with increased affordability have exponentially increased the number of vehicles on road.

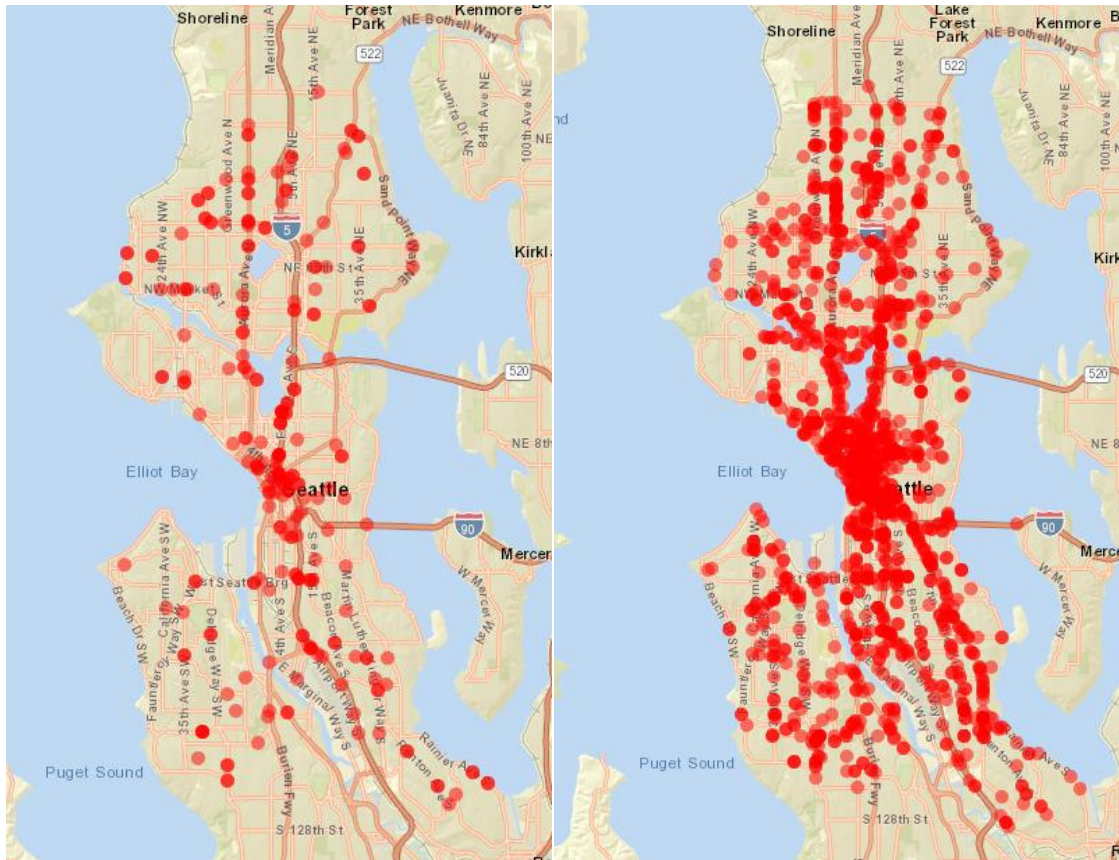
In this report we will be discussing about transport related issues in and around Seattle. We tried to see if we could identify underlying causes for increased number of accidents in certain areas compared to other areas.



Courtesy SDOT

### 1.2 Problem

Making roads safer for everybody has been a priority for the government and local authorities. We are trying to identify causes of accidents, and to reduce them continuously. We aim to make Seattle an accident free city in Five years time.



The first map shows the 244 fatal accidents that have taken place since January 2004. The second map shows the areas in which serious accidents have occurred during the same period. This clearly shows that certain areas are prone to more accidents and certain areas are relatively safer. There could be firm underlying explanations behind these patterns. Which is what we are trying to identify and address.

### 1.3 Interest

Seattle Department of Transport, Seattle Police department, Government, drivers, passengers and pedestrians will all want to have an accident free city. Findings from this research upon verification shall be shared with all concerned parties.

## 2. Data

### 2.1 Data sources

There are a number of publicly available datasets containing traffic and accident data. This analysis is based on a sample dataset provided by Coursera, and the dataset can be found [here](#) and the metadata can be found [here](#). The data contains details of collisions that took place in Seattle between 1<sup>st</sup> January 2004 and 20<sup>th</sup> May 2020.

	SEVERITYCODE	X	...	CROSSWALKKEY	HITPARKEDCAR
0	2	-122.323148	...	0	N
1	1	-122.347294	...	0	N
2	1	-122.334540	...	0	N
3	1	-122.334803	...	0	N
4	2	-122.306426	...	0	N
...	...	...	...	...	...
194668	2	-122.290826	...	0	N
194669	1	-122.344526	...	0	N
194670	2	-122.306689	...	0	N
194671	2	-122.355317	...	0	N
194672	1	-122.289360	...	0	N

[194673 rows x 38 columns]

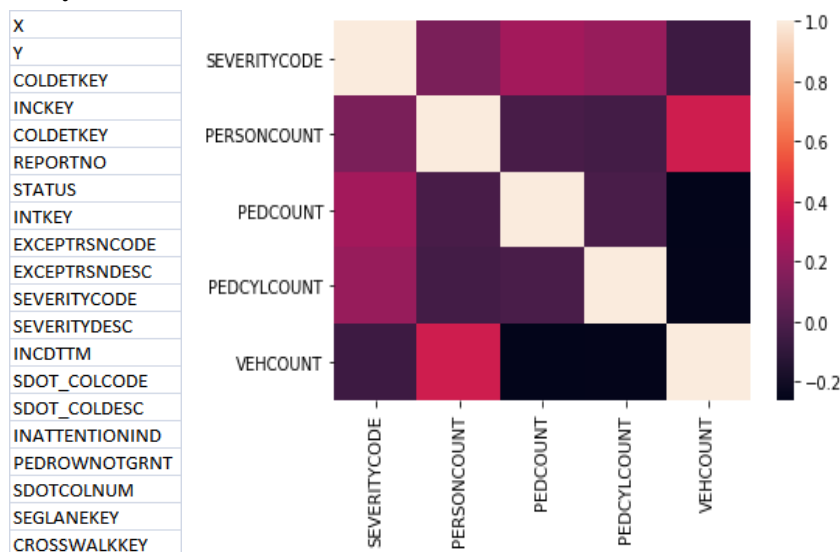
	SEVERITYCODE	X	...	SEGLANEKEY	CROSSWALKKEY
count	194673.000000	189339.000000	...	194673.000000	1.946730e+05
mean	1.298901	-122.330518	...	269.401114	9.782452e+03
std	0.457778	0.029976	...	3315.776055	7.226926e+04
min	1.000000	-122.419091	...	0.000000	0.000000e+00
25%	1.000000	-122.348673	...	0.000000	0.000000e+00
50%	1.000000	-122.330224	...	0.000000	0.000000e+00
75%	2.000000	-122.311937	...	0.000000	0.000000e+00
max	2.000000	-122.238949	...	525241.000000	5.239700e+06

## 2.2 Data cleaning

The number of fields without data, were very high but this was expected as the data set contains incidents far back from 2004. Where means of data collection was very limited compared to what we have today. We can expect to reduce null values and have complete datasets in future to make better decisions.

ADDRTYPE	1926
INTKEY	129603
LOCATION	2677
EXCEPTRSNCODE	109862
EXCEPTRSNDESC	189035
SEVERITYCODE.1	0
SEVERITYDESC	0
COLLISIONTYPE	4904
PERSONCOUNT	0
PEDCOUNT	0
PEDCYLCOUNT	0
VEHCOUNT	0
INCDATE	0
INCDTTM	0
JUNCTIONTYPE	6329
SDOT_COLCODE	0
SDOT_COLDESC	0
INATTENTIONIND	164868
UNDERINFL	4884
WEATHER	5081
ROADCOND	5012
LIGHTCOND	5170
PEDROWNOTGRNT	190006
SDOTCOLNUM	79737
SPEEDING	185340
ST_COLCODE	18
ST_COLDESC	4904

All reference identity columns and lesser important fields have been excluded for ease of analysis. Such as



## 2.3 Feature selection

17 Features were retained for analysis of the dataset. The features are given below.

```
Index(['SEVERITYCODE', 'ADDRTYPE', 'LOCATION', 'COLLISIONTYPE', 'PERSONCOUNT',  
      'PEDCOUNT', 'PEDCYLCOUNT', 'VEHCOUNT', 'INCDATE', 'JUNCTIONTYPE',  
      'UNDERINFL', 'WEATHER', 'ROADCOND', 'LIGHTCOND', 'SPEEDING',  
      'ST_COLDESC', 'HITPARKEDCAR'],  
      dtype='object')
```

These fields contain the environmental variables such as location, timing of the day, weather, road condition. It also has details about the severity of those accidents. We will have to find a direct relationship between one or more environmental variables (Causation variables) and number of accidents (Result).

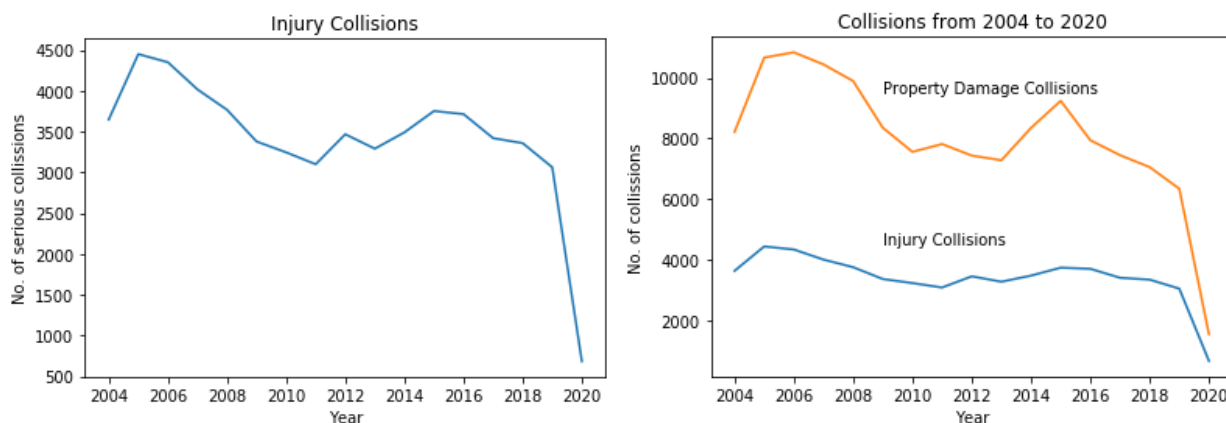
## 3. METHOD

### 3.1 Basic Exploration

	SEVERITYCODE	PERSONCOUNT	PEDCOUNT	PEDCYLCOUNT	VEHCOUNT
SEVERITYCODE	1.000000	0.130949	0.246338	0.214218	-0.054686
PERSONCOUNT	0.130949	1.000000	-0.023464	-0.038809	0.380523
PEDCOUNT	0.246338	-0.023464	1.000000	-0.016920	-0.261285
PEDCYLCOUNT	0.214218	-0.038809	-0.016920	1.000000	-0.253773
VEHCOUNT	-0.054686	0.380523	-0.261285	-0.253773	1.000000

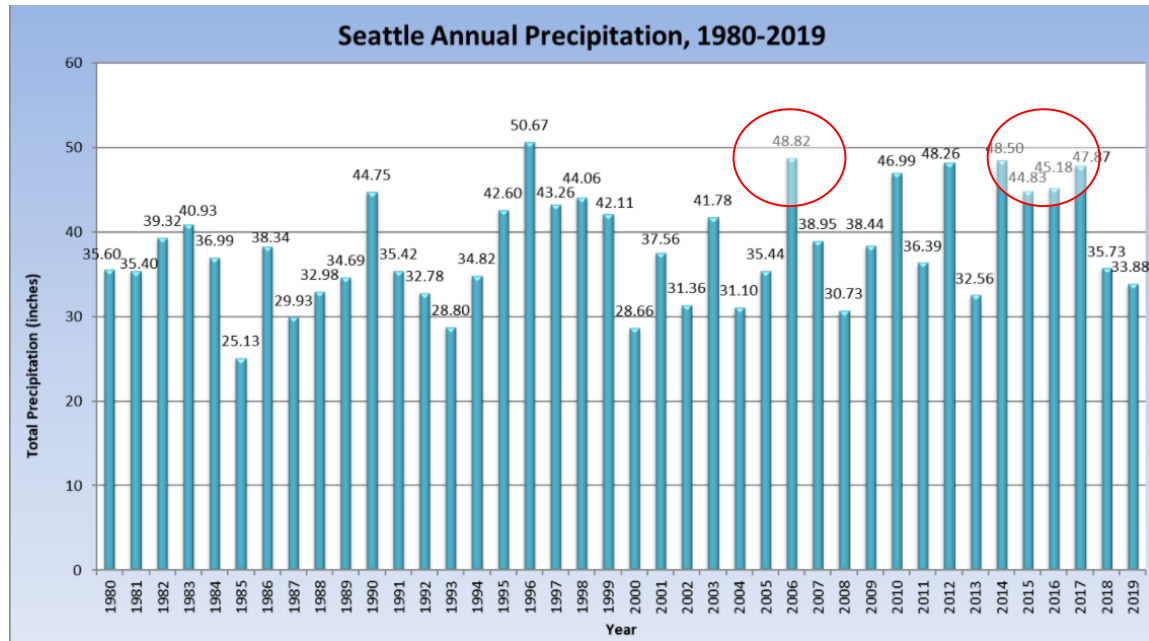
From the outset we could not see any linear relationships with the given dataset. We had to do a more detailed and in depth analysis to find what we were looking for.

### 3.2 The Trend



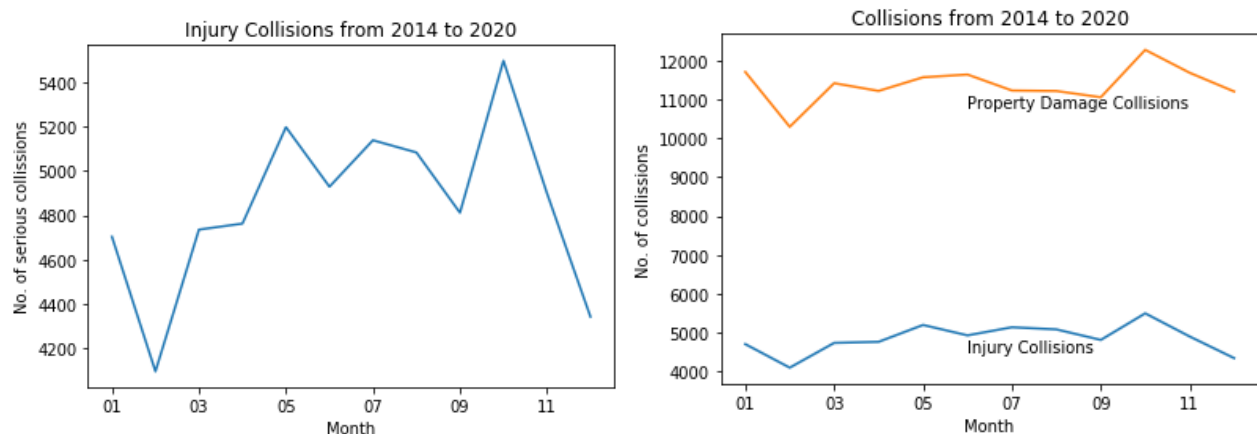
The number of incidents have on averaged been on the decline compared to 2005/2006 levels. Despite seeing a downward trend there was a spike during 2015 & 2016 compared previous years. This phenomenon needs to be identified and addressed in order to avoid such an unexpected spike happening again in the near future or in 2025/26!..

### 3.3 Possible Causes



Both 2006 & 2016 had seen above 45 inches of rain fall. The spike in number of collisions in 2012 also could be seen as result of high rain fall. However we need more data to verify this.

We looked at the number of collisions on a monthly basis to see if our initial theory of “Rain causes Accidents” holds.

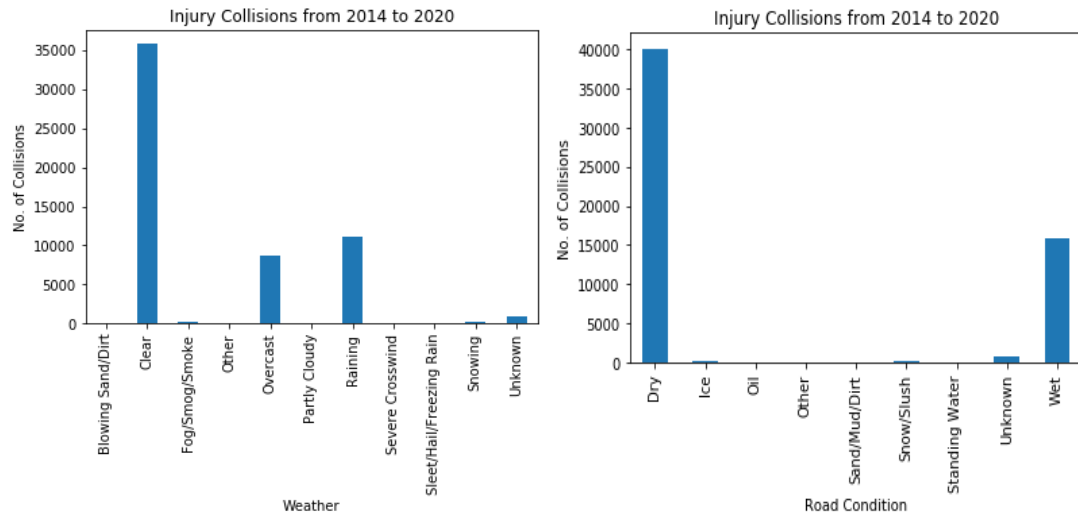


The highest number of collision were recorded in the month of October, this coincides perfectly with the month of highest rainfall!.. We can also see higher levels accidents in May, which is a rainy month in the first half of the year.

However the trend does not explain the reason for having lowest number of accidents in February, which normally receives higher rainfall compared to June to September. There for had to further investigate.



Seattle gets 150 days of rain a year on average, which is around 41% of the days.

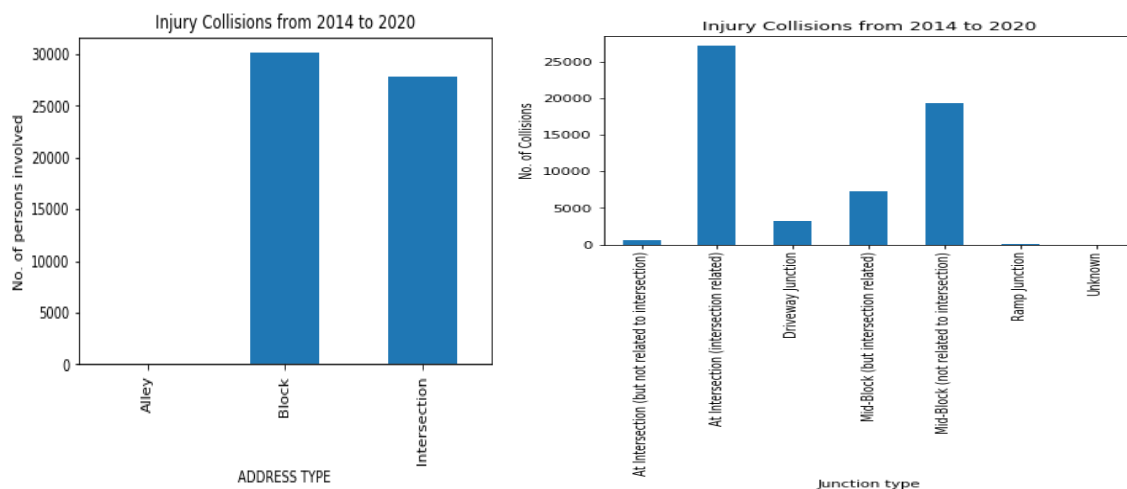


Only 37% of collisions have taken place during overcast or rainy weather.

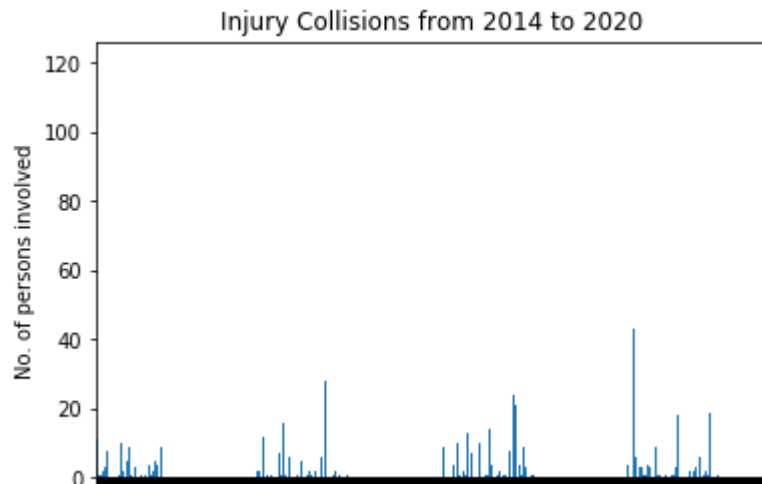
Only 31% of collisions had taken place on wet roads.

If rain was to be considered the prime cause of accidents then at least 42% of accidents should taken place on rainy, overcast or slippery roads. We can now say even though the weather may have played a part in accidents, we cannot single it out as the only reason behind accidents.

Looking at the type of roads and addresses it was clear that 99% collisions took place in blocks or intersections. More variables started emerging. We were working with a number of other unknowns now.



Further analyzing the data brought up a handful of areas were home to most of the accidents.

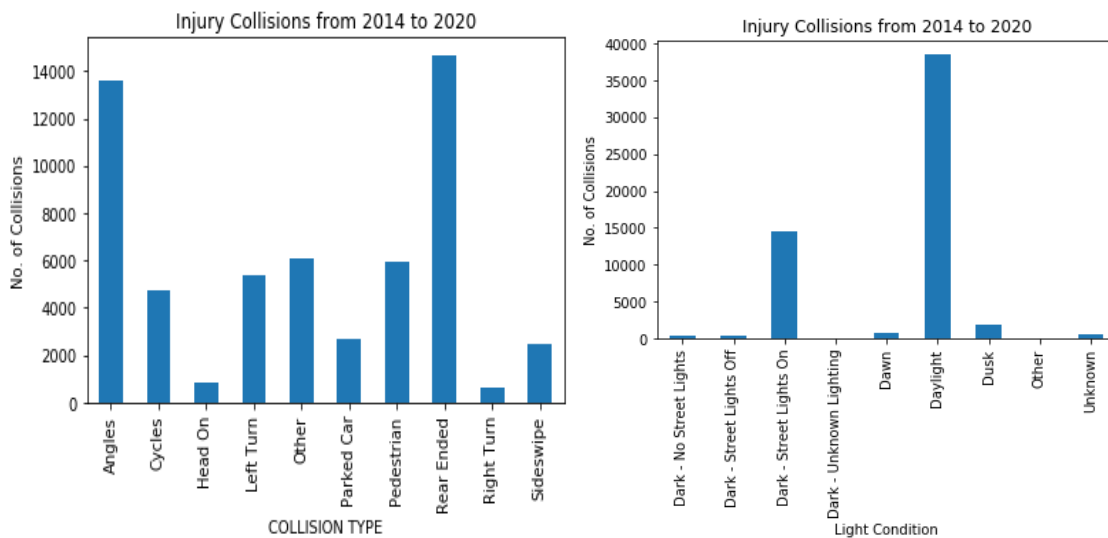


The top ten areas are listed below

	LOCATION	PERSONCOUNT
10802	BATTERY ST TUNNEL NB BETWEEN ALASKAN WY VI NB ...	276
10803	BATTERY ST TUNNEL SB BETWEEN AURORA AVE N AND ...	271
16841	N NORTHGATE WAY BETWEEN MERIDIAN AVE N AND COR...	265
10636	AURORA AVE N BETWEEN N 117TH PL AND N 125TH ST	254
9152	6TH AVE AND JAMES ST	252
10640	AURORA AVE N BETWEEN N 130TH ST AND N 135TH ST	239
10345	ALASKAN WY VI NB BETWEEN S ROYAL BROUGHAM WAY ...	238
19351	RAINIER AVE S BETWEEN S BAYVIEW ST AND S MCCLE...	231
10357	ALASKAN WY VI SB BETWEEN COLUMBIA ST ON RP AND...	212
23757	WEST SEATTLE BR EB BETWEEN ALASKAN WY VI NB ON...	212

## 4. Results

Also looking at the data it was evident that different types of accidents had occurred with contradicting levels. Which may point to a human error as the reason for accidents.



## 5. Conclusion

We can say that accidents are caused by a combination of environmental and human factors



## **6. Future**

We will need more information to further analyze causes of accidents with certainty and to predict risk areas with highest levels of certainty to make roads of Seattle safe for everyone