

1.The Introduction:

1.1 Background:

Once I was on my way back home from festive shopping when I witnessed a road accident. I was with my sister and it was around 6 o'clock in the evening. In the middle of the road, we saw a crowd surrounding something. We weren't quite sure what was happening as the first thought that came to our mind was that it was probably a quarrel between two men. However, when we reached the spot, we found out an accident had taken place.

Thus, road accidents are bound to happen. You pick up a newspaper and you will find at least one or two news about road accidents daily.

1.2 Problem:

Road accidents have become very common nowadays. As more and more people are buying automobiles, the incidences of road accidents are just increasing day by day. Furthermore, people have also become more careless now. Not many people follow the traffic rules. Especially in big cities, there are various modes of transports. Moreover the roads are becoming narrower and the cities have become more populated.

According to WHO (World Health Organization):

“Every year the lives of approximately 1.35 million people are cut short as a result of a road traffic crash. Between 20 and 50 million more people suffer non-fatal injuries, with many incurring a disability as a result of their injury.

Road traffic injuries cause considerable economic losses to individuals, their families, and to nations as a whole. These losses arise from the cost of treatment as well as lost productivity for those killed or disabled by their injuries, and for family members who need to take time off work or school to care for the injured. Road traffic crashes cost most countries 3% of their gross domestic product.”

Now, It's clear that we are dealing with an important problem. So, This project aims to predict how severity of accidents can be reduced based on some factors. This project will be based on Seattle city (a seaport city on the West Coast of the United States and The largest city in both the state of Washington and the Pacific Northwest region of North America) data.

1.3 Stakeholders:

The reduction of accidents severity can be be benefit to the authority of development in Seattle especially and all governments and authorities of transportation in common to

improves and transportation system and also for individuals to know the factors of accidents and how to prevent them as possible.

2.Data:

2.1 Data source:

Collision data had been fetched from Seattle Department of Transportation Open Data Program in CSV format.

Source:

<https://s3.us.cloud-object-storage.appdomain.cloud/cf-coursesdata/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv>

2.2 Data understanding:

Our dataset contains all the information about a total number of 194673 car accidents. It contains the location, time, date, weather conditions, road conditions etc. The dataset has 37 features and The

dataset has total observations of 194673 with variation in number of observations for every feature.

The dataset also has a lot of empty columns which could have been beneficial had the data been present there. These columns included pedestrian granted way or not, segment lane key, crosswalk key and hit parked car.

2.3 Feature Selecting:

As we are using this dataset to predict the severity of an accident not all the features will be useful for us. So, I will only pick up the features that will help me in this project.

The features that I selected are:

1.SEVERITYCODE:

a code that corresponds to the severity of collision and this column will be the main one as we are predicting the severity.

It has two values:

- 1:prop damage
- 2:injury

2.LIGHTCOND:

The conditions of the light during the collision.

Its values are:

(daylight / dark- street lights on / dark - no street lights / dark- street lights off /dark -

unknown lighting / dusk / dawn /other /
unknown)

3.WEATHER:

A description of weather conditions during the time of the collision.

Its values are:

(overcast - raining - clear - snowing -
severe crosswind - partly cloudy - fog/ smog/
smoke - steel/ hail / freezing rain - blowing
sand/ dirt - other - unknown)

4.ROADCOND:

The conditions of the road during the collision.

Its values are:

(wet - dry- ice - snow/slush -
sand/mud/dirt - standing water - oil - other -
unknown)

2.4 Data cleaning:

First of all, we had to drop the null values from the dataset, which results in about a 10% decrease in the overall dataset. We were able to do this as we realise null values exist for all 3 feature columns normally for each row.

	SEVERITYCODE	WEATHER	LIGHTCOND	ROADCOND
15	1	NaN	NaN	NaN
36	1	NaN	NaN	NaN
53	2	NaN	NaN	NaN
60	1	NaN	NaN	NaN
75	1	NaN	NaN	NaN

So, Number of entries after dropping NA values will be 189337.

Secondly, we will have to down sample our dataset as we have 3 times more rows with a severity code of 1 compared to 2.

```
SEVERITYCODE
1      132285
2       57052
..         ..
```

We will down sample the rows with severity code of 1 to be the same as severity code of 2 for our use case.

```
2       57052
1       57052
```

This an example of our data from the first 5 columns of our dataset:

	SEVERITYCODE	WEATHER	LIGHTCOND	ROADCOND
0	2	Overcast	Daylight	Wet
1	1	Raining	Dark - Street Lights On	Wet
2	1	Overcast	Daylight	Dry
3	1	Clear	Daylight	Dry
4	2	Raining	Daylight	Wet

The downsampled dataset:

	SEVERITYCODE	WEATHER	LIGHTCOND	ROADCOND
82770	1	Clear	Daylight	Dry
122946	1	Clear	Daylight	Dry
102968	1	Clear	Daylight	Dry
13906	1	Overcast	Daylight	Dry
123301	1	Raining	Dark - Street Lights On	Wet

3. Methodology:

I will take every feature alone and start to analyze its values by total as I will make for every feature, except 'SEVERITYCODE', three data frames one for has the total of repeated times for each value in both types of accident 1 and 2. Then I will make another one for the type 1 and one for type 2.

After that, I analyze the results of each data frame of each feature.

Now let's talk in detail:

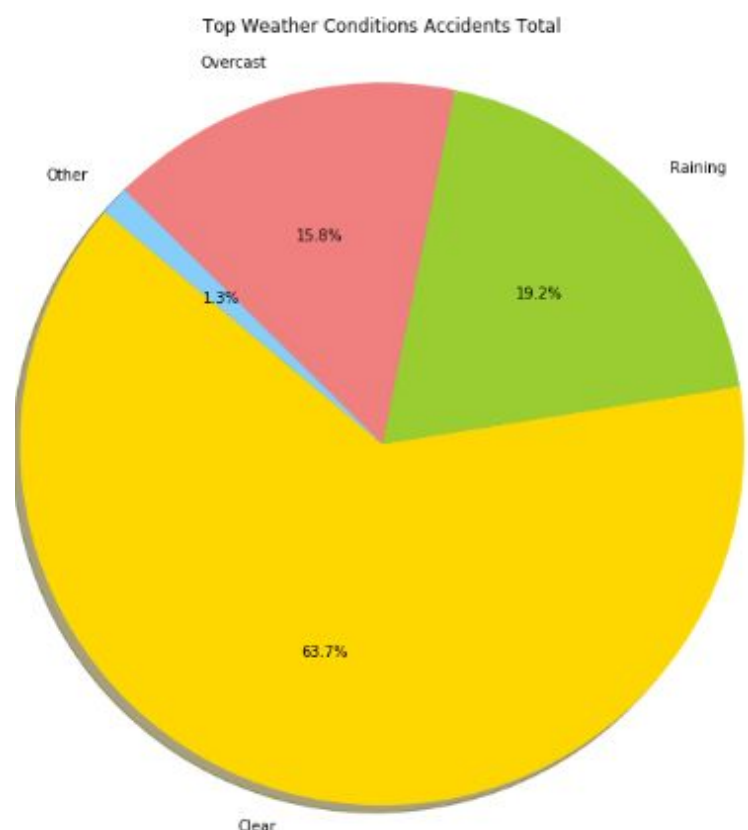
Weather:

The Total:

Below there is a data frame that has the count of every condition accident and a pie chart of its top elements.

The total accidents count

Clear	68248
Raining	20619
Overcast	16951
Unknown	6896
Snowing	500
Other	435
Fog/Smog/Smoke	337
Sleet/Hail/Freezing Rain	71
Blowing Sand/Dirt	30
Severe Crosswind	13
Partly Cloudy	4



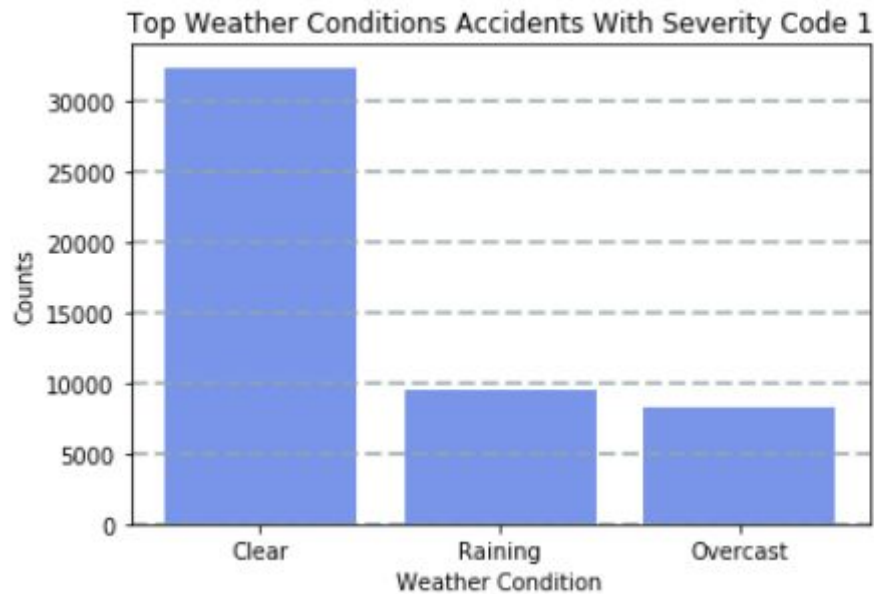
From above it's clear that most of the accidents occur in

Clear weather by 68248 nearly 64% after it is the rainy condition by 20619 a percentage of 20% of the total and the last one is overcast by 16951 is almost 16%.

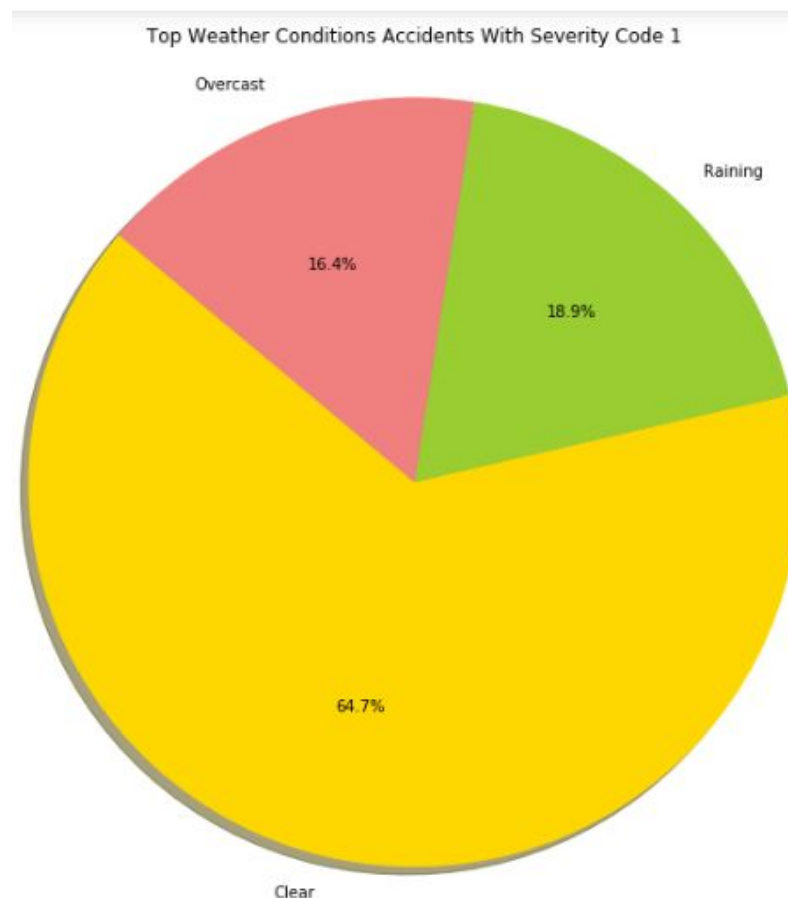
The 1 severity code count:

The total accidents count

Clear	32440
Raining	9451
Overcast	8212
Unknown	6084
Snowing	331
Other	319
Fog/Smog/Smoke	150
Sleet/Hail/Freezing Rain	43
Blowing Sand/Dirt	15
Severe Crosswind	6
Partly Cloudy	1



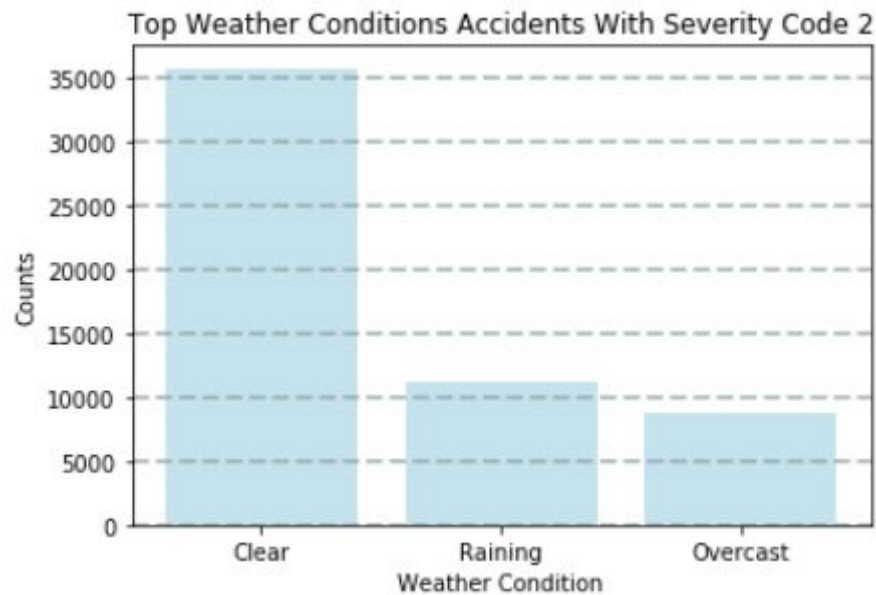
- Clear , Raining ,Overcast are the top three in accidents count.
- Conditions like partly cloudy and Severe crosswind have too few counts that can't be compared to the above.
- The values are near to half the total.
- Two thirds of the accidents happen in Clear weather.



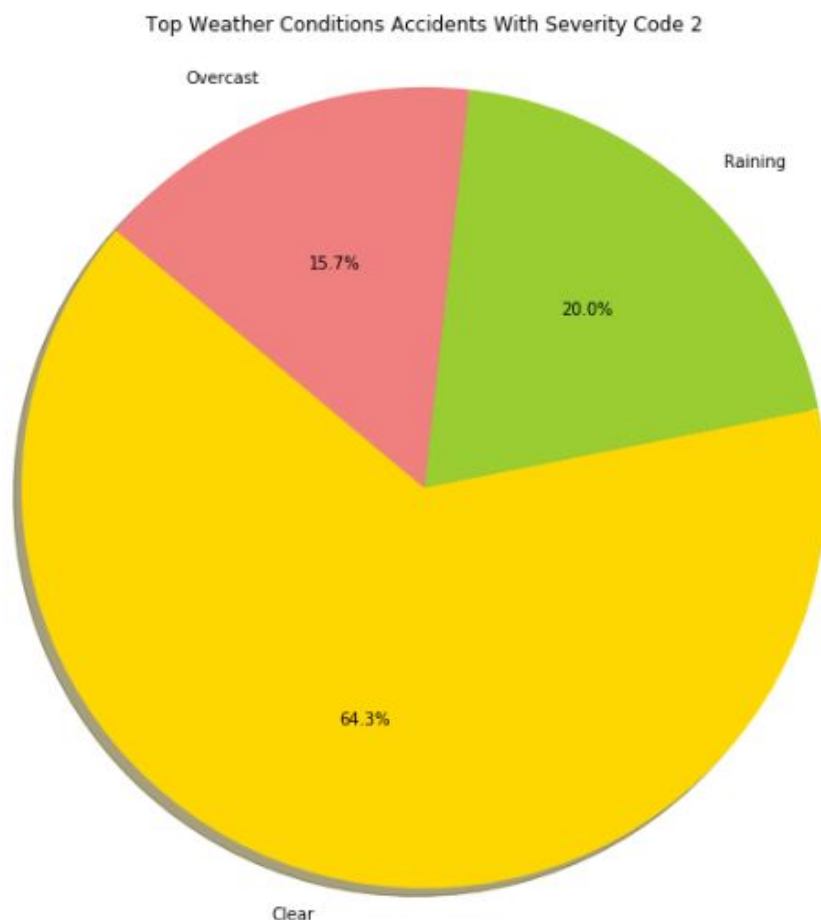
The 2 severity code count:

The total accidents count

Clear	35808
Raining	11168
Overcast	8739
Unknown	812
Fog/Smog/Smoke	187
Snowing	169
Other	116
Sleet/Hail/Freezing Rain	28
Blowing Sand/Dirt	15
Severe Crosswind	7
Partly Cloudy	3



- Clear , Raining ,Overcast still the top three in accidents count.
- The count of Raining and Clear is bigger than it in severity of 1.
- still almost two thirds of the accidents happen in Clear weather.



The Light Conditions:

The total:

Below there are :

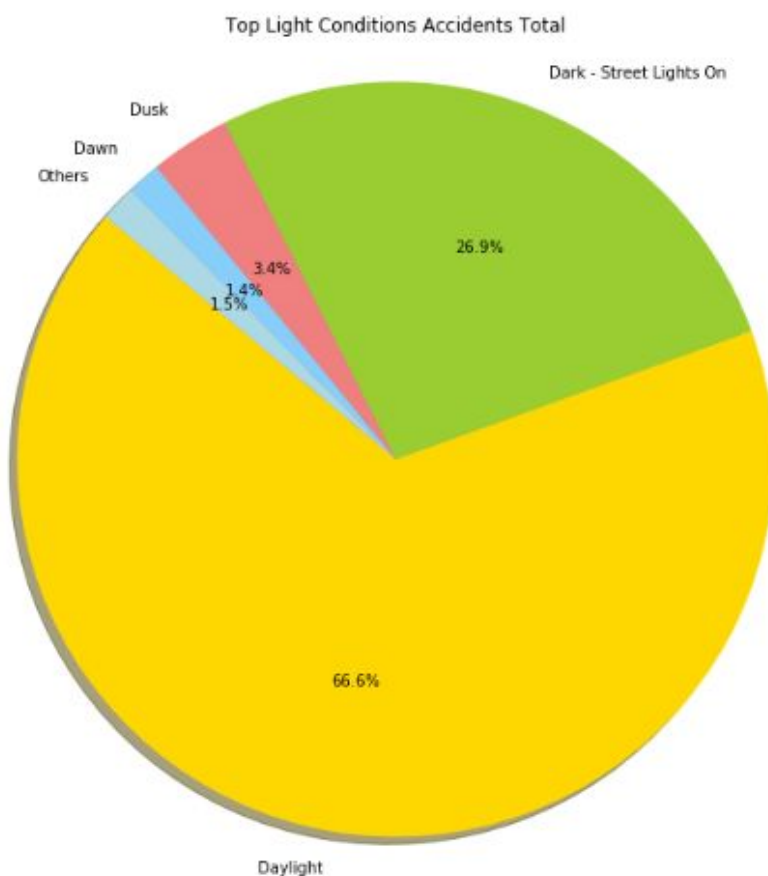
- A data frame has the total number of accidents occurred in every light condition.
- A pie chart shows the percentages of top conditions.

The total accidents count

Daylight	71980
Dark - Street Lights On	29104
Unknown	6087
Dusk	3713
Dawn	1564
Dark - No Street Lights	838
Dark - Street Lights Off	683
Other	128
Dark - Unknown Lighting	7

- From the data frame it's clear that most of the accidents happen in daylight in total of 71980 then the Daylight by 14653.

- The third condition Dusk and the fourth Dawn are more less in count than the first and second .



- The pie chart shows that only 7% of accidents happen in Dusk and Dawn and other conditions.

- 27% of the accidents happen in dark-street lights on

- Top condition is daylight by 66%

The 1 severity code count:

Below there are :

- A data frame has the total number of accidents occurred in every light condition with the severity level of 1.
- A pie chart shows the count percentages of the top 4 light conditions.
- A bar chart shows the total of the top 4 light conditions' counts.

The total accidents count

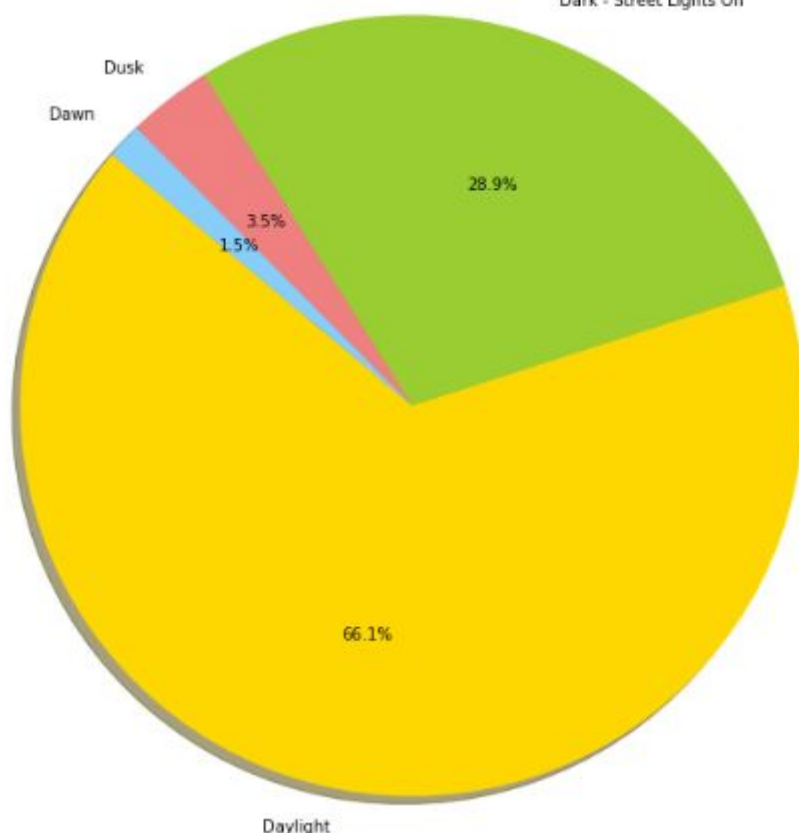
Daylight	33452
Dark - Street Lights On	14653
Unknown	5482
Dusk	1775
Dawn	740
Dark - No Street Lights	504
Dark - Street Lights Off	367
Other	76
Dark - Unknown Lighting	3

- Daylight condition is still the top one in total of 33452 accidents.

- Dark-Street Lights on condition is the second and has a total of 14653 near to the half of Daylight.

Top Light Conditions Accidents With Severity Code 1

Dark - Street Lights On



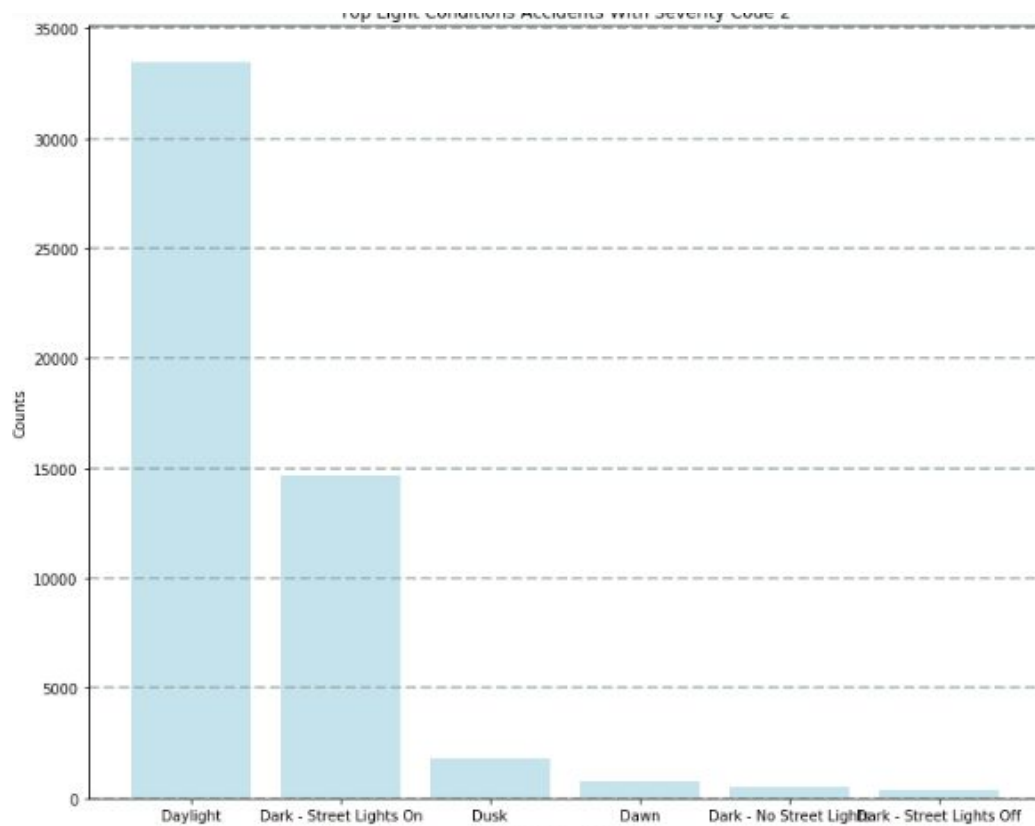
- The pie chart shows the percentages of top 4 conditions : Daylight, Dark-Street Lights on, Dusk and Dawn.

- 66% of the accidents happen in daylight

- Dark-Street Lights on represents the third.

- Dusk and Dawn show only 5%.

This bar chart shows the difference in accidents counts between the Light Conditions with severity code 1.



- There is a huge difference in values between Daylight and other conditions and also between Dark-Street Lights on and the ones after it.

The 2 severity code count:

Below there are :

- A data frame has the total number of accidents occurred in every light condition with the severity level of 2.
- A pie chart shows the count percentages of the top 4 light conditions.
- A bar chart shows the total of the top 4 light conditions' counts.

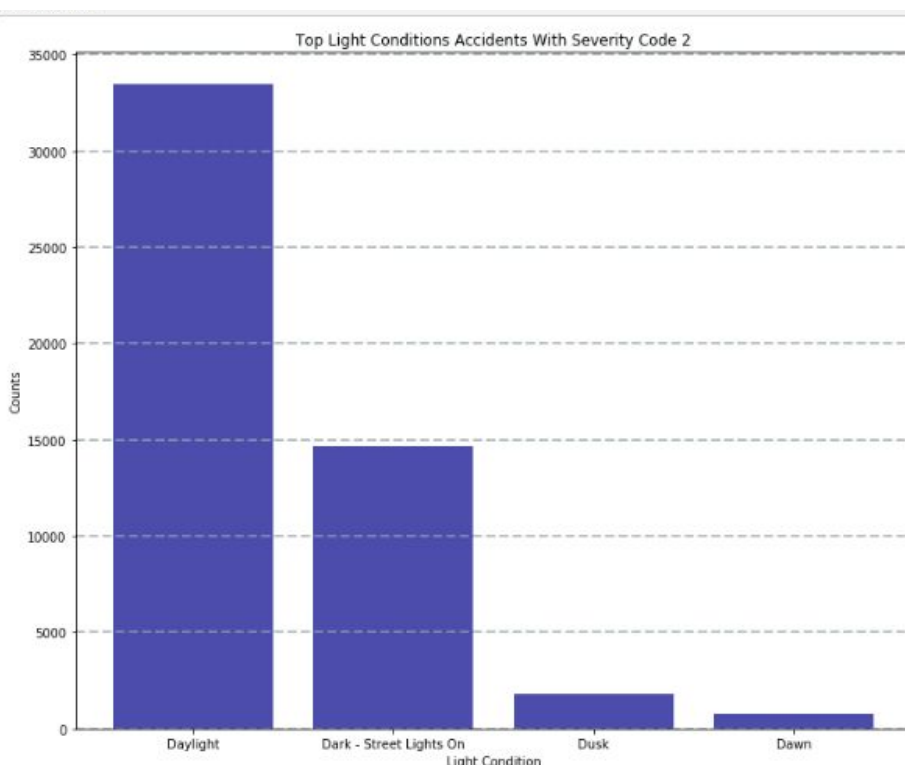
The total accidents count

Daylight	38528
Dark - Street Lights On	14451
Dusk	1938
Dawn	824
Unknown	605
Dark - No Street Lights	334
Dark - Street Lights Off	316
Other	52
Dark - Unknown Lighting	4

- Daylight condition is still the top one in total of 38528 accidents.

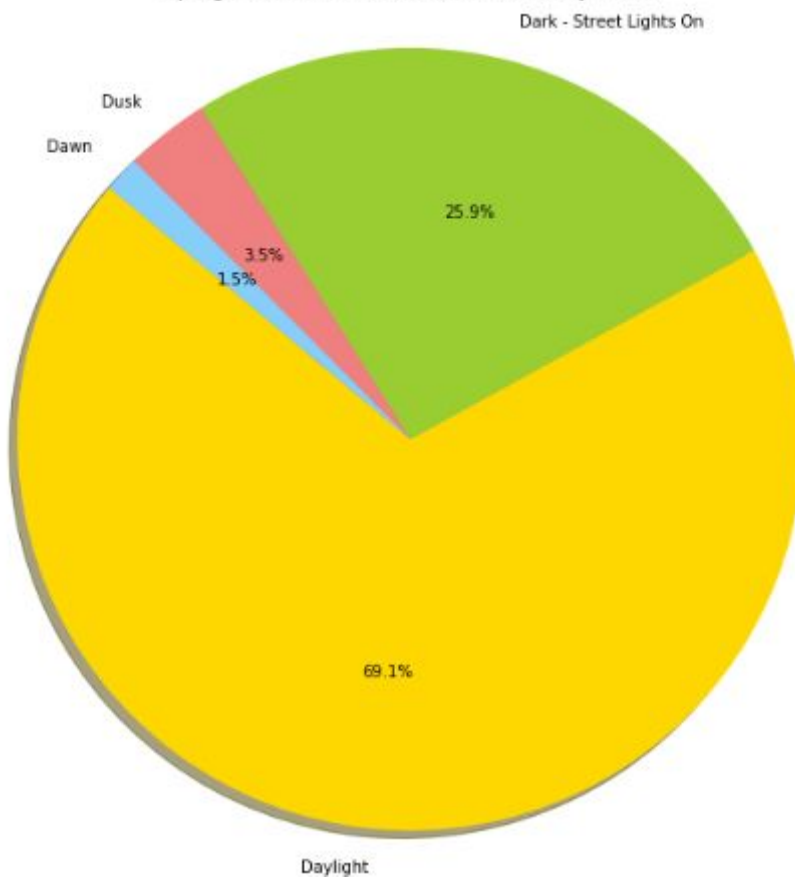
- Dark-Street Lights on condition is the second and has a total of 14451.

- Dusk and Dawn are counts higher than in 1 severity code.



This bar chart shows the difference in accidents counts between the Light Conditions with severity code 2.

Top Light Conditions Accidents With Severity Code 2



- The pie chart shows the percentages of top 4 conditions : Daylight, Dark-Street Lights on, Dusk and Dawn.

- 69.1% of the accidents happen in daylight.

- Dark-Street Lights on represents 26% of the accidents.

- Dusk and Dawn still represent only 5%.

The Road Conditions:

The total:

Below there are :

- A data frame has the total number of accidents occurred in every light condition.
- A pie chart shows the percentages of top conditions.

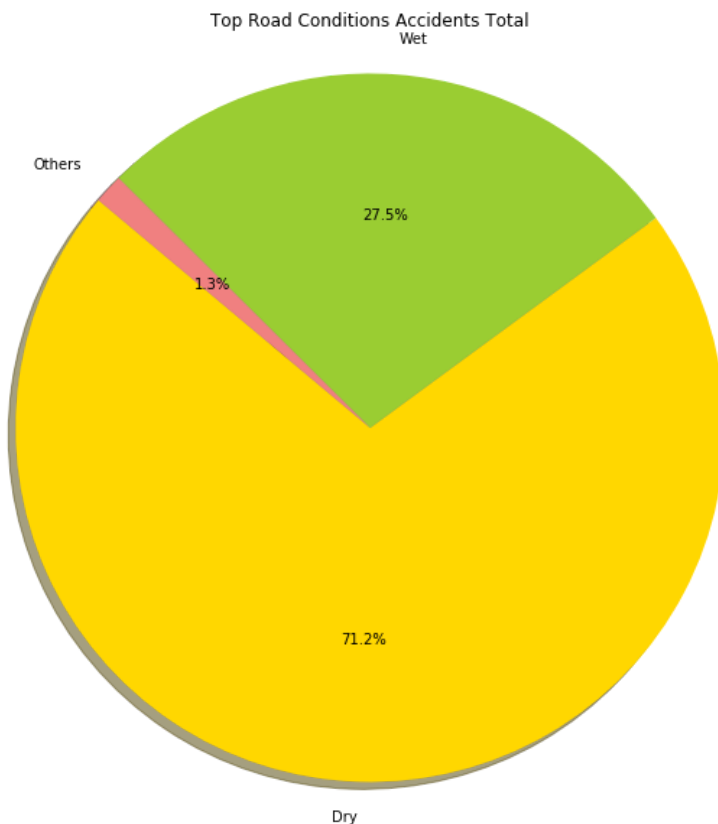
The total accidents count

Dry	76368
Wet	29449
Unknown	6850
Ice	670
Snow/Slush	538
Other	78
Standing Water	67
Oil	43
Sand/Mud/Dirt	41

- Most of the accidents occur in the Dry roads.

- The second one is the Wet roads accidents counts.

- The other conditions counts' are much more fewer than the first two.



The pie chart shows the percentages of top 2 conditions :

- wet
- dry

71.2% of the accidents happen on Dry roads.

27.5% happen on Wet ones.

1.3% happens on other road conditions.

The 1 severity code count:

Below there are :

- A data frame has the total number of accidents occurred in every light condition with the severity level of 1.
- A pie chart shows the count percentages of the top 2 road conditions.
- A bar chart shows the total of the top 2 road conditions' counts.

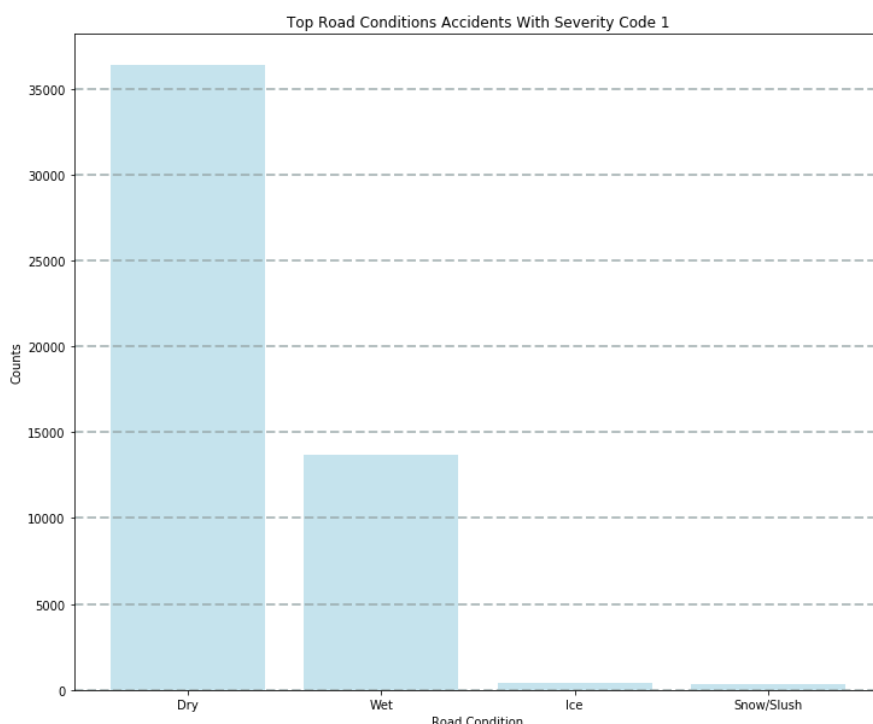
The total accidents count

Dry	36364
Wet	13707
Unknown	6103
Ice	397
Snow/Slush	372
Standing Water	37
Other	35
Oil	19
Sand/Mud/Dirt	18

- It's clear that most accidents with severity code of 1 happen on Dry roads with a count of 36364.

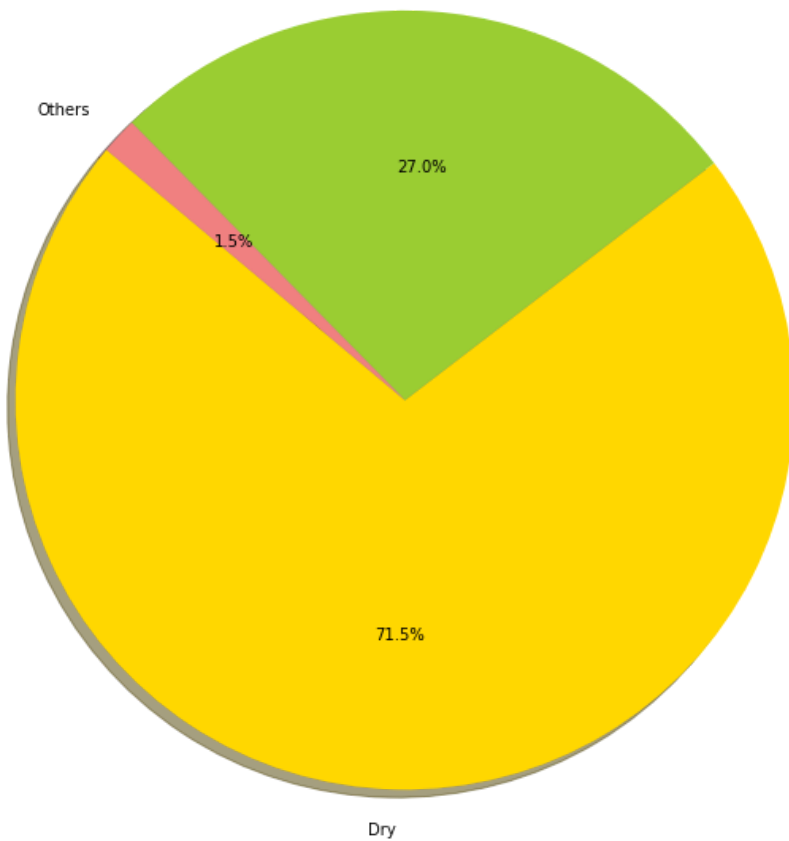
- 13707 accidents happened on Wet roads.

- Other conditions have too few counts like Ice and Oil.



This bar chart shows the difference in accidents counts between the Road Conditions with severity code 1. There is a huge difference between Dry and wet and the others.

Top Road Conditions Accidents With Severity Code 1



The pie chart shows the percentages of top 2 conditions with Severity code 1:
Dry Road 71%
Wet Road 27%
Others 1.5%

The 2 severity code count:

Below there are :

- A data frame has the total number of accidents occurred in every light condition with the severity level of 2.
- A pie chart shows the count percentages of the top 2 road conditions.
- A bar chart shows the total of the top 2 road conditions' counts.

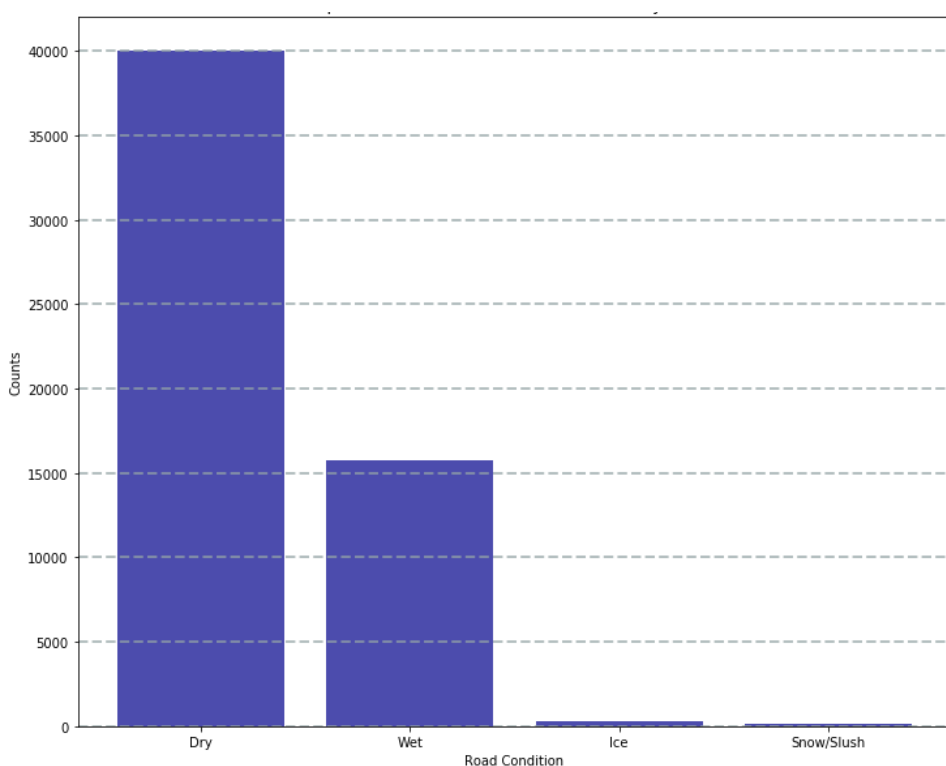
The total accidents count

Dry	40004
Wet	15742
Unknown	747
Ice	273
Snow/Slush	166
Other	43
Standing Water	30
Oil	24
Sand/Mud/Dirt	23

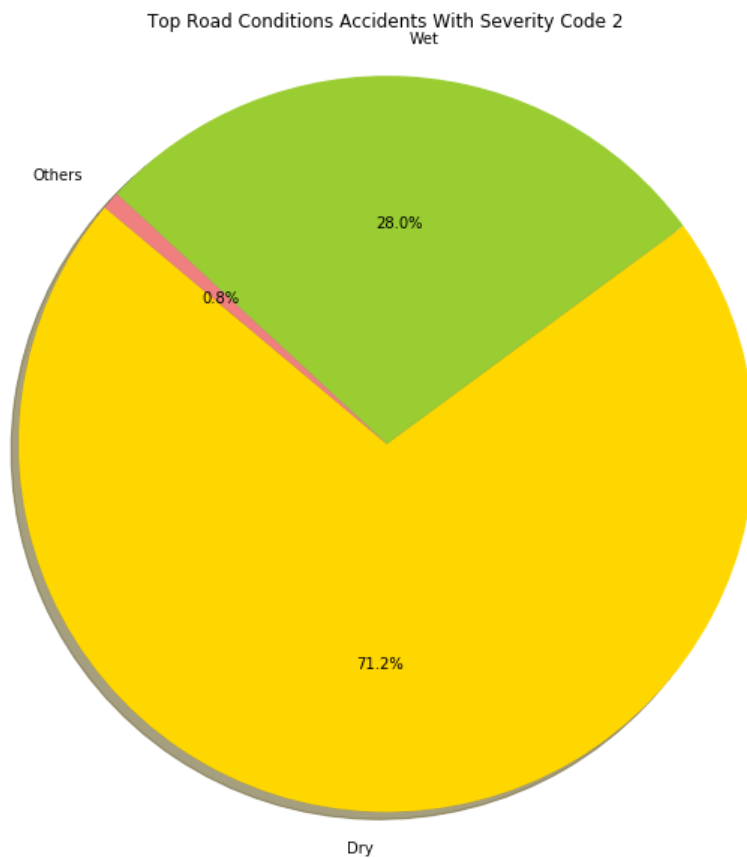
- It's clear that most accidents with severity code of 2 happen on Dry roads with a count of 40004.

- 15742 accidents happened on Wet roads.

- Other conditions have too few counts like Ice and Oil.



This bar chart shows the difference in accidents counts between the Road Conditions with severity code 2. There is a huge difference between Dry and wet and the others.



The pie chart shows the percentages of top 2 conditions with Severity code 1: Dry Road 71.2% Wet Road 28% Others 0.8%

4.Results:

I will compare between The counts of 1 and 2 Severity code for each feature.

Based on the following analysis the skate holders should know the important conditions of severe accidents and make a solution for the problem.

1.Weather:

SEVERITYCODE	1	2
WEATHER		
Blowing Sand/Dirt	15	15
Clear	32440	35808
Fog/Smog/Smoke	150	187
Other	319	116
Overcast	8212	8739
Partly Cloudy	1	3
Raining	9451	11168
Severe Crosswind	6	7
Sleet/Hail/Freezing Rain	43	28
Snowing	331	169

The probability of 2 is higher in:

- Clear: 4.9%
- Raining:8.3%
- Overcast :3.1%
- Fog/Smog/Smoke:10.9%

The probability of 1 is higher in:

- Snowing: 32.4%
- Sleet/Hail/Freezing Rain:21.1%

2.Light:

SEVERITYCODE	1	2
LIGHTCOND		
Dark - No Street Lights	504	334
Dark - Street Lights Off	367	316
Dark - Street Lights On	14653	14451
Dark - Unknown Lighting	3	4
Dawn	740	824
Daylight	33452	38528
Dusk	1775	1938
Other	76	52
Unknown	5482	605

The probability of 2 is higher in:

- Dawn:5.4%
- Daylight:7%
- Dusk: 4.4%

The probability of 1 is higher in:

- Streets-Lights on:0.8%
- Streets-Lights off:7.4%
- No Streets Lights: 20.3%

3.Road:

SEVERITYCODE	1	2
ROADCOND		
Dry	36364	40004
Ice	397	273
Oil	19	24
Other	35	43
Sand/Mud/Dirt	18	23
Snow/Slush	372	166
Standing Water	37	30
Unknown	6103	747
Wet	13707	15742

The probability of 2 is higher in:

- Dry:4.8%
- Oil:11.6%
- Wet:6.9%
- Sand/Mud/Dirt:12.2%

The probability of 1 is higher in:

- Ice: 18.5%
- Snow/Slush:38.3%
- Standing Water: 10.4%

5.Conclusion:

Road accidents, one of the most important problems at our time, have become very

common nowadays. Through the report I tried to make a prediction of accidents severity based on an accidents data set and used the data of lighting on Road and weather conditions to help me find a solution to my problem.