

Seattle Vehicle Collision Analysis

*Project Report in partial fulfilment of
IBM Data Science Professional Certificate*

Anmol Tripathi

August 26, 2020

1. Introduction

1.1. Background

According to the annual United States road crash statistics by ASIRT, more than 38,000 people die every year in crashes on U.S. roadways. The U.S. traffic fatality rate is 12.4 deaths per 100,000 inhabitants. It is evident that with the increasing number of vehicles on urban and suburban roads, the cases of vehicle accidents are also increasing. This project aims to analyse vehicle collision data available in public domain provided by Seattle Geo Data (SDOT) and derive some useful insights.

1.2. Problem

Data provided by the Seattle Department of Transportation (SDOT) on vehicle collisions along with its severity might be useful to derive insights and may show some pattern with the environmental factors like weather, road conditions etc.

Main objective of this analysis is as follows:

- Derive some useful insights from the data
- Develop a model to predict the severity of an accident

1.3. Interest

This project might interest the local government, SDOT and all the stakeholders of Seattle Traffic Management.

2. Data acquisition and cleaning

2.1. Data Sources

Data used for this analysis was originally fetched from Seattle Geodata. [Click here](#) to visit the page. The dataset has been modified as per requirement of the Capstone Project of IBM Data Science Professional Certificate. The modified dataset can be downloaded in csv format from [here](#). The timeframe of the data starts from 2004. *For this analysis, original data is being used.*

2.2. About Dataset

The dataset consists of 40 columns having different kinds of data like, collision severity, road conditions, number of people involved, location of collision, weather etc. Meta-data of the dataset can be viewed [here](#).

2.3.Dataset Summary

Data Set Basics	
Title	Collisions—All Years
Abstract	All collisions provided by SPD and recorded by Traffic Records.
Description	This includes all types of collisions. Collisions will display at the intersection or mid-block of a segment. Timeframe: 2004 to Present
Supplemental info	-
Update Frequency	Weekly
Keyword(s)	SDOT, Seattle, Transportation, Accidents, Bicycle, Car, Collisions, Pedestrian, Traffic, Vehicle
Contact Information	
Contact Organization	SDOT Traffic Management Division, Traffic Records Group
Contact Person	SDOT GIS Analyst
Contact Email	DOT_IT_GIS@seattle.gov

2.4 Data Cleaning

Data had many columns with ambiguous values such as column UNDERINFL had values “Y”, “N”, “0”, “1”. Since the column describes whether a driver involved was under the influence of drug or alcohol, these either must be “Y” or “N”. Converted all the 0,1 to “Y” and “N” respectively. INCKEY and COLDETKEY are not useful for our analysis so it is better to drop them. SEVERITYCODE has a code 0 corresponding to Unknown category, it is better to classify all the missing SEVERITYCODE to 0.

3. Feature Engineering

70% of data has time associated with each collision. It is easier to analyse with respect to time if we segregate the data in certain categories. I have segregated the data in 6 categories as following.

TIME	
Late Night	00:00 to 04:00
Early Morning	04:00 to 08:00
Morning	08:00 to 12:00
Noon	12:00 to 16:00
Eve	16:00 to 20:00
Night	20:00 to 00:00

4. Analysis

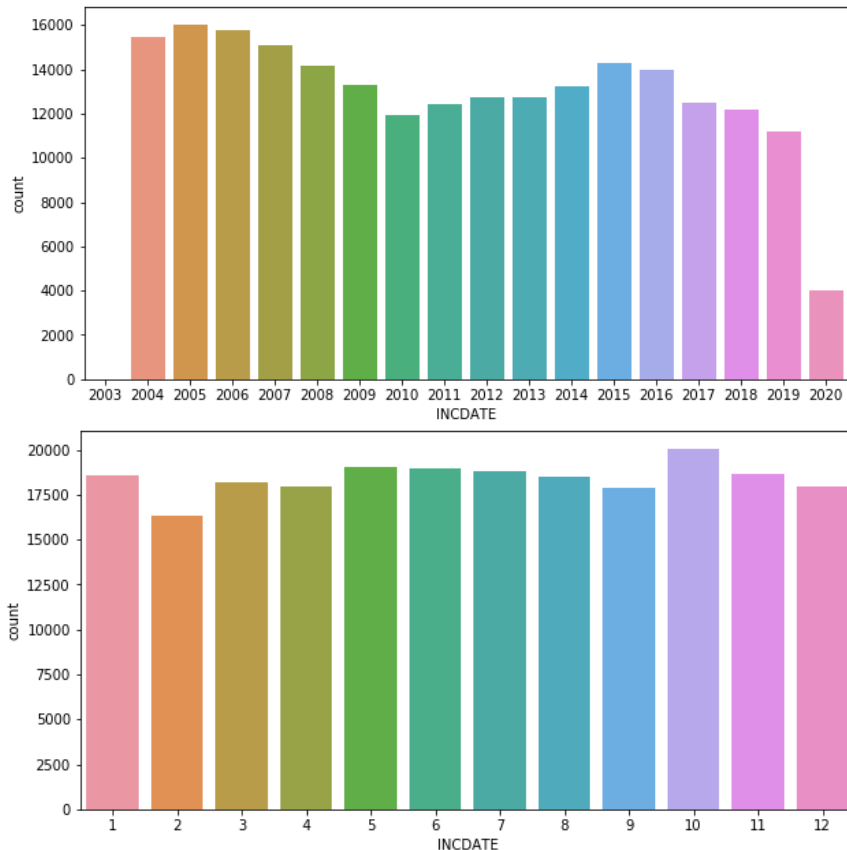
4.1. Analysing the severity

62% of accidents belong to category 1 which is property damage only collision

and 26% belong to category 2 which is injury collision. It shows that majority of accidents happening are just causing property damage. Fatal collisions are ~0.15% of dataset. Whereas, we do not have any data about 10% of collisions.

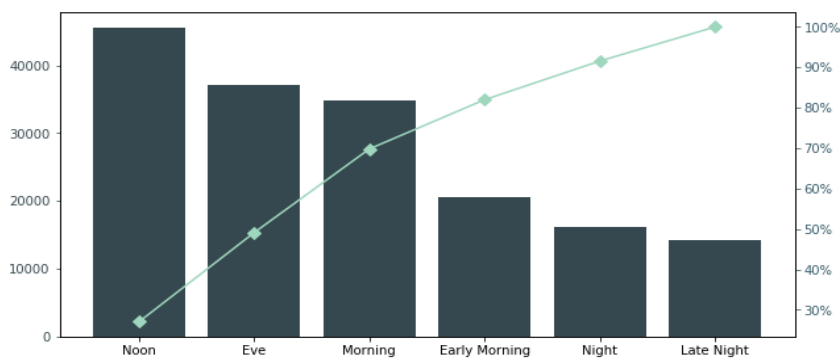
4.2. Analysis by Month and Year

Although is no clear trend visible, but it seems that cases are declining yearly. There is no pattern with month visible.

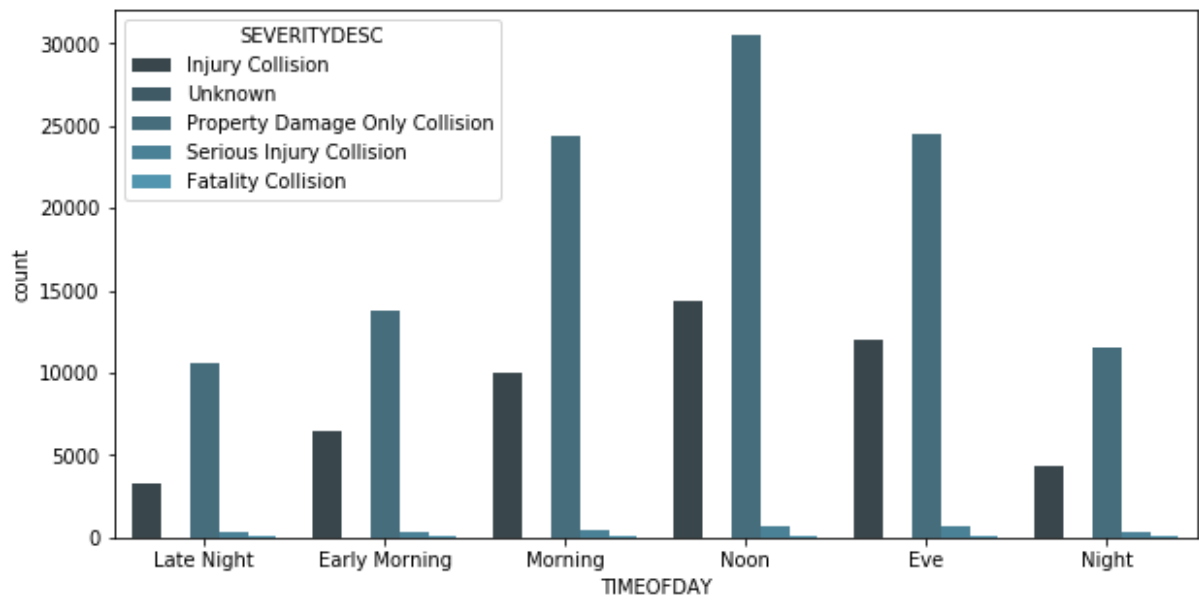
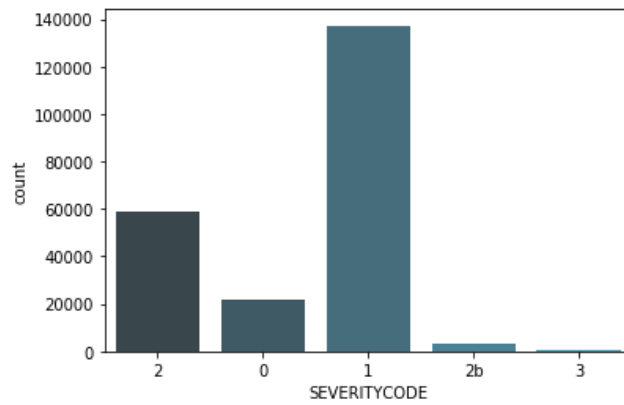


4.3. Analysis by Time

We can see that almost 50% of cases happen during the daytime i.e. between 12:00 P.M. and 08:00 P.M.



By splitting the data on severity shows that cases of injury collision are almost always half of that of property damage only collision. In general, cases of property damage is maximum followed by injury collision.



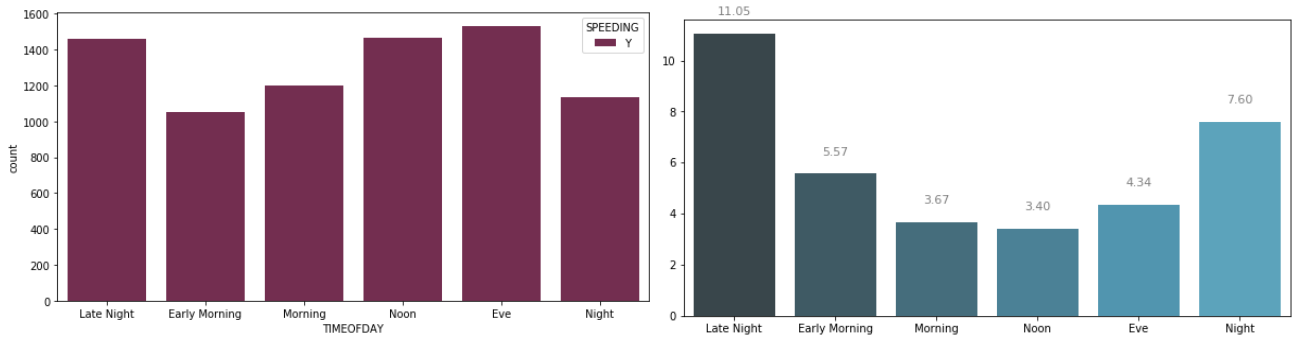
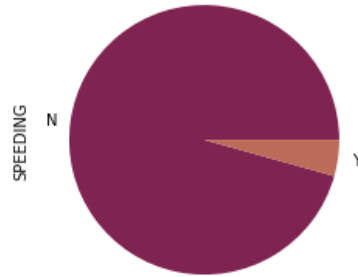
SEVERITYCODE	0	1	2	2b	3
TIMEOFDAY					
Late Night	0	10639	3232	314	52
Early Morning	1	13748	6407	326	37
Morning	4	24403	10039	432	41
Noon	1	30534	14354	647	63
Eve	0	24509	11998	650	58
Night	0	11500	4295	288	41

SEVERITYCODE	SEVERITYDESC
2	Injury Collision
0	Unknown
1	Property Damage Only Collision
2b	Serious Injury Collision
3	Fatality Collision

There is no clear relation between severity of accidents and time of the day. During late night, cases of severity 1 (“Property Damage only Collision”) incidents are higher than rest of the day percentage wise. 3 of 4 accidents happening during late night are just doing property damage.

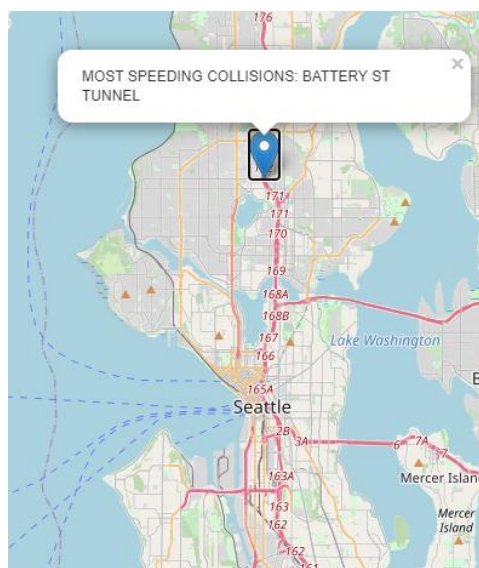
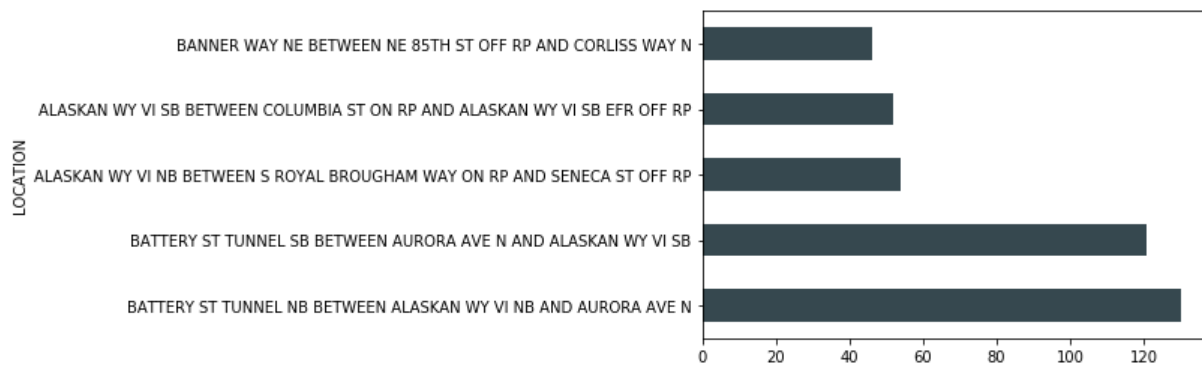
4.4 Analysis on Speeding

5% of all the collision cases are because of speeding. Distribution with time shows that there are lot many cases of speeding during evening and late night.



Percentage wise tells some other story. There are a lot many (11.05%) speeding accidents in late night as compared to any other time.

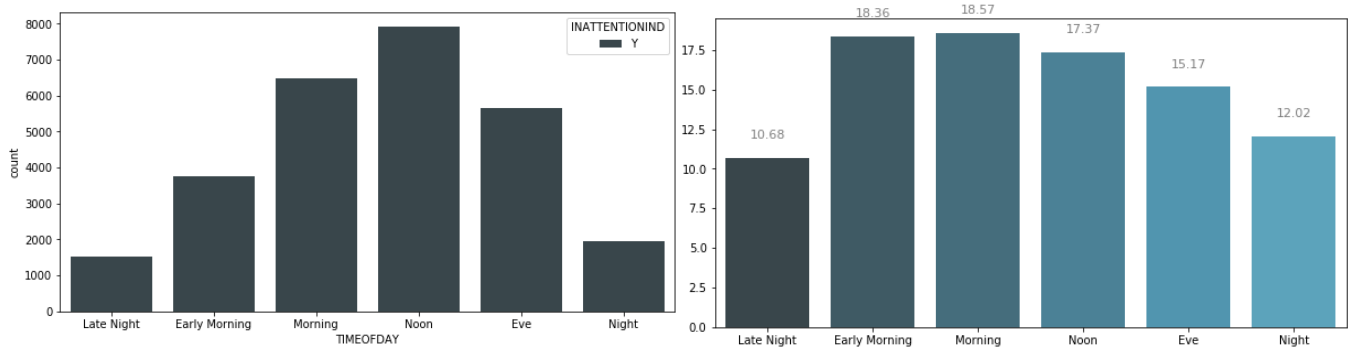
If we see location wise, we notice that there are a lot many cases of speeding collisions at “Battery St. Tunnel”.



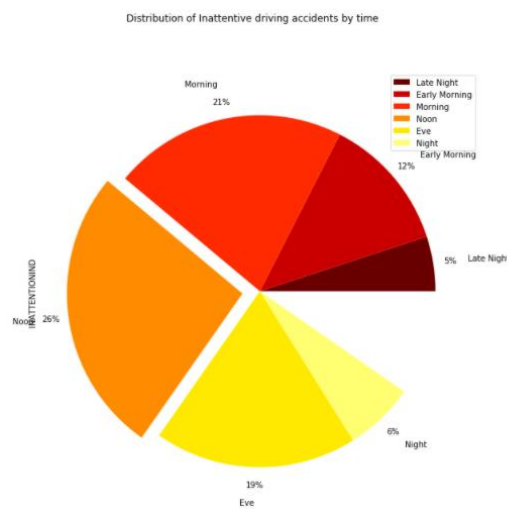
SEVERITYDESC	
Property Damage Only Collision	58.873665
Injury Collision	36.303808
Serious Injury Collision	3.927069
Fatality Collision	0.895458

It is worth noting that percentage of
Injury collision **increases** to 36.3% from 26.52%
Serious injury collision **increases** to 3.92% from 1.39%
Fatality injury collision **increases** to 0.89% from 0.15% because of speeding.

4.4 Analysis on collision due to inattention

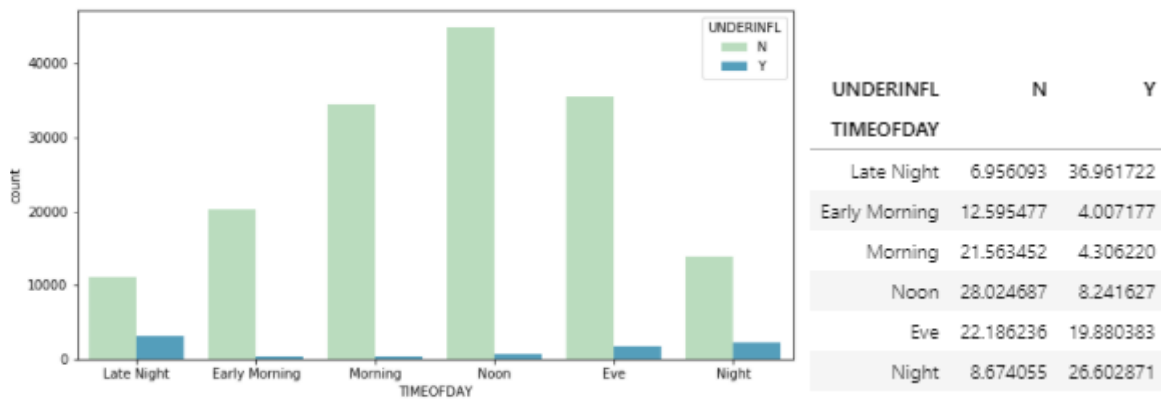


We can see that ~18% of collisions happening in morning and early morning are due to inattention of drivers.



4.5 Analysis on collision due under influence of drug/alcohol

Alcohol/drug influenced accidents are becoming a huge problem for metropolitans. It is quite evident from the data that majority of accidents under influence are happening in late night and in evening.



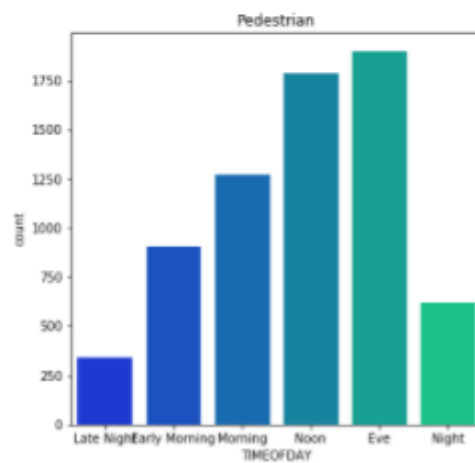
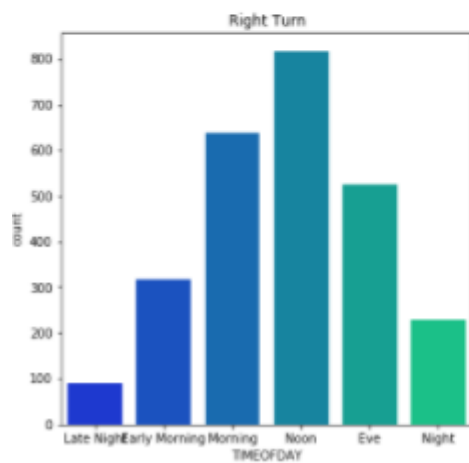
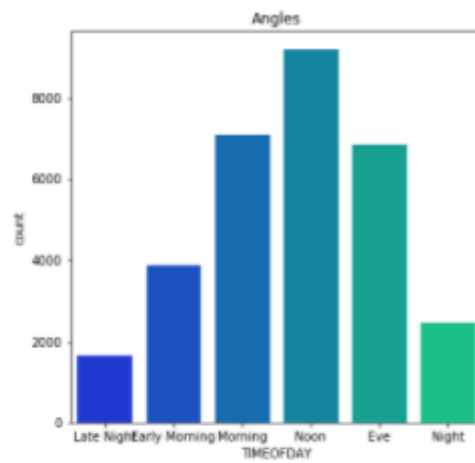
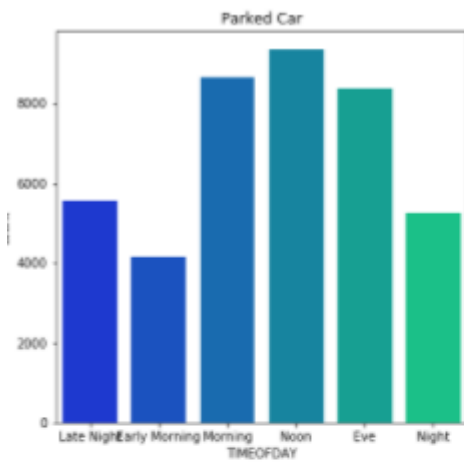
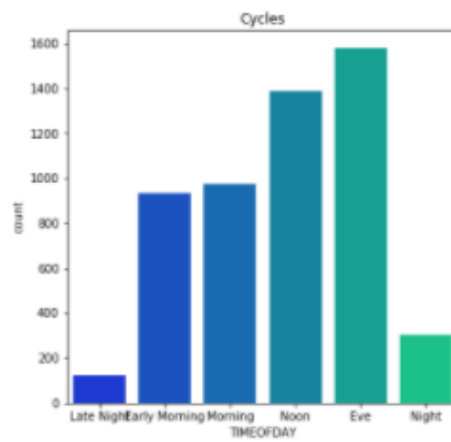
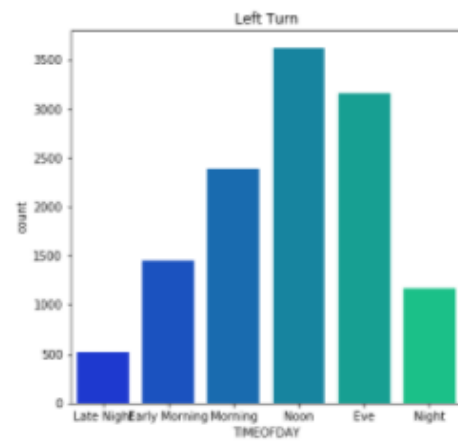
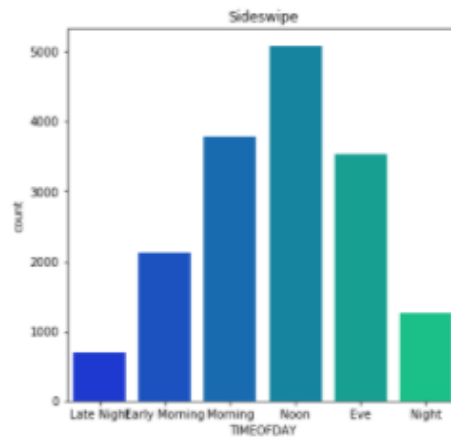
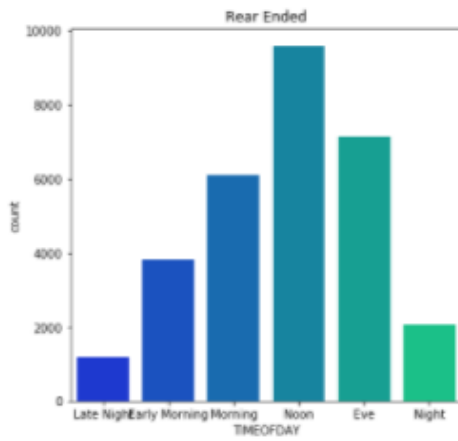
We can see there are relatively more cases of alcoholic and drug influenced collision during December (Probably because of holiday season)

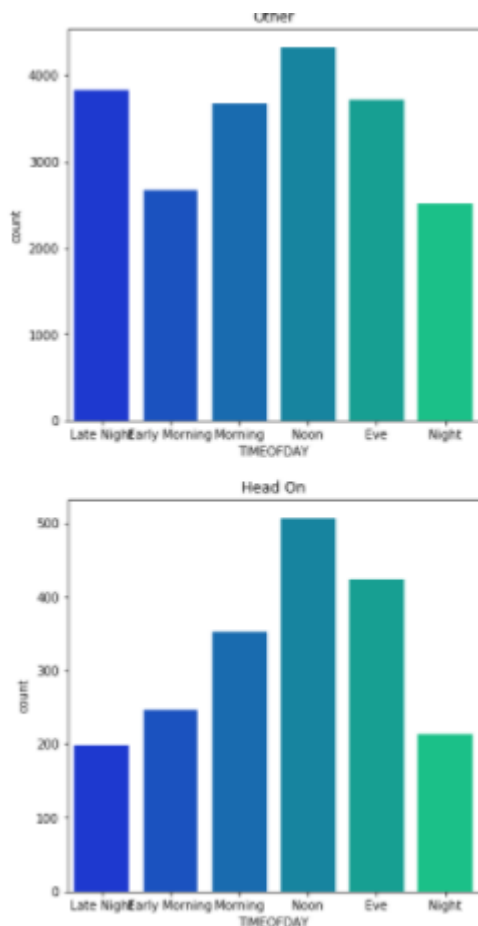


"ALASKAN VI SB BETWEEN COLUMBIA ST ON RP AND ALASKAN WY VI SB EFR OFF RP" has the greatest number of cases of under influence collisions.

4.6 Analysis by collision type

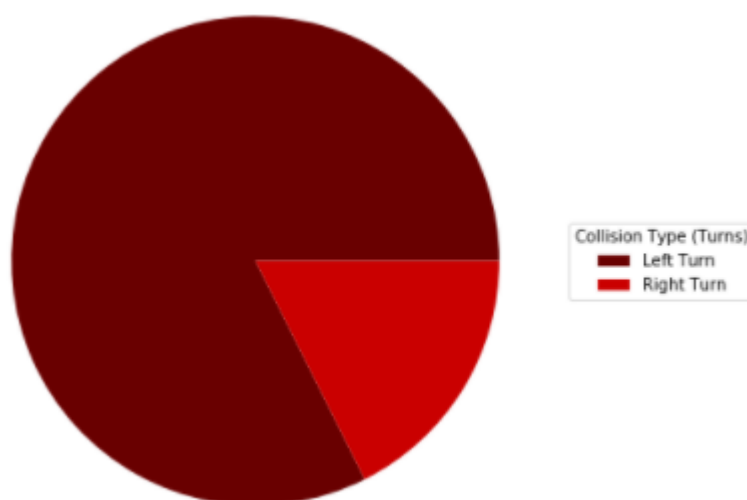
We can analyse the distribution of collisions by the collision type. Collision type is further bifurcated by SDOT collision code. Since there are a lot many categories, it is better to plot them and have an overview.





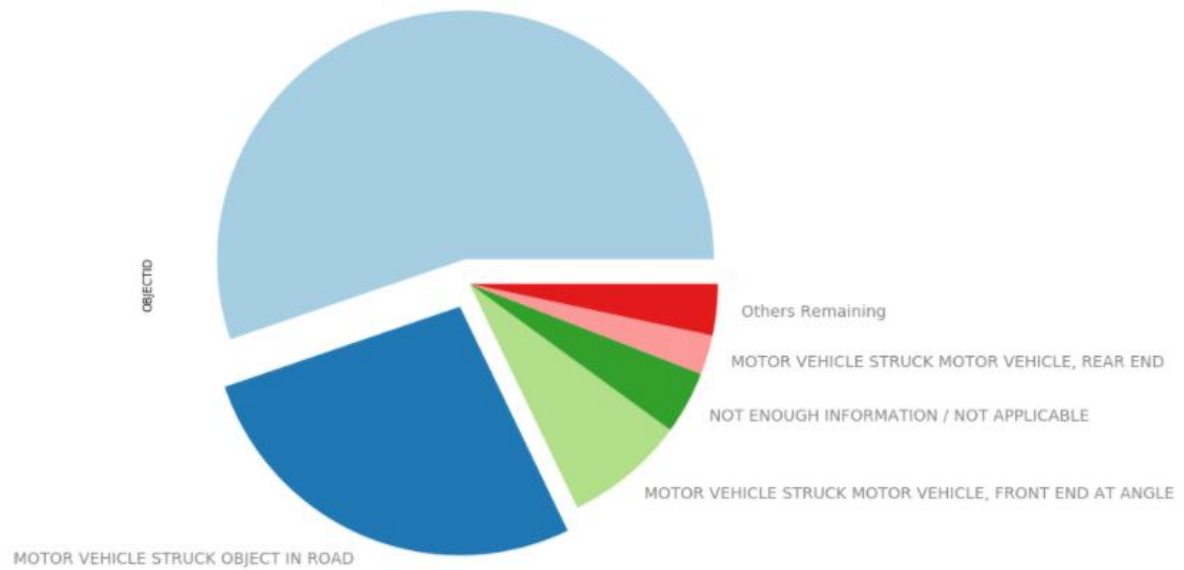
TIMEOFDAY	Late Night	Early Morning	Morning	Noon	Eve	Night
COLLISIONTYPE						
Angles	11.694880	18.938286	20.287032	20.135553	18.432226	15.387955
Cycles	0.877994	4.543239	2.792976	3.040074	4.245942	1.891707
Head On	1.397766	1.204056	1.011201	1.112062	1.139417	1.327296
Left Turn	3.610311	7.087842	6.854966	7.946744	8.502634	7.225702
Other	26.873639	13.030126	10.530236	9.471168	9.977964	15.592632
Parked Car	39.060195	20.249586	24.712825	20.460179	22.468559	32.531167
Pedestrian	2.395168	4.421371	3.646624	3.917440	5.105880	3.857843
Rear Ended	8.470886	18.636053	17.514108	20.997565	19.200795	12.882218
Right Turn	0.632156	1.550161	1.824744	1.792020	1.410835	1.420331
Sideswipe	4.987006	10.339280	10.825289	11.127196	9.515748	7.883148

People are more likely to hit Parked car during Night. We can also see there are a lot of "Other" collision type of cases in late night.



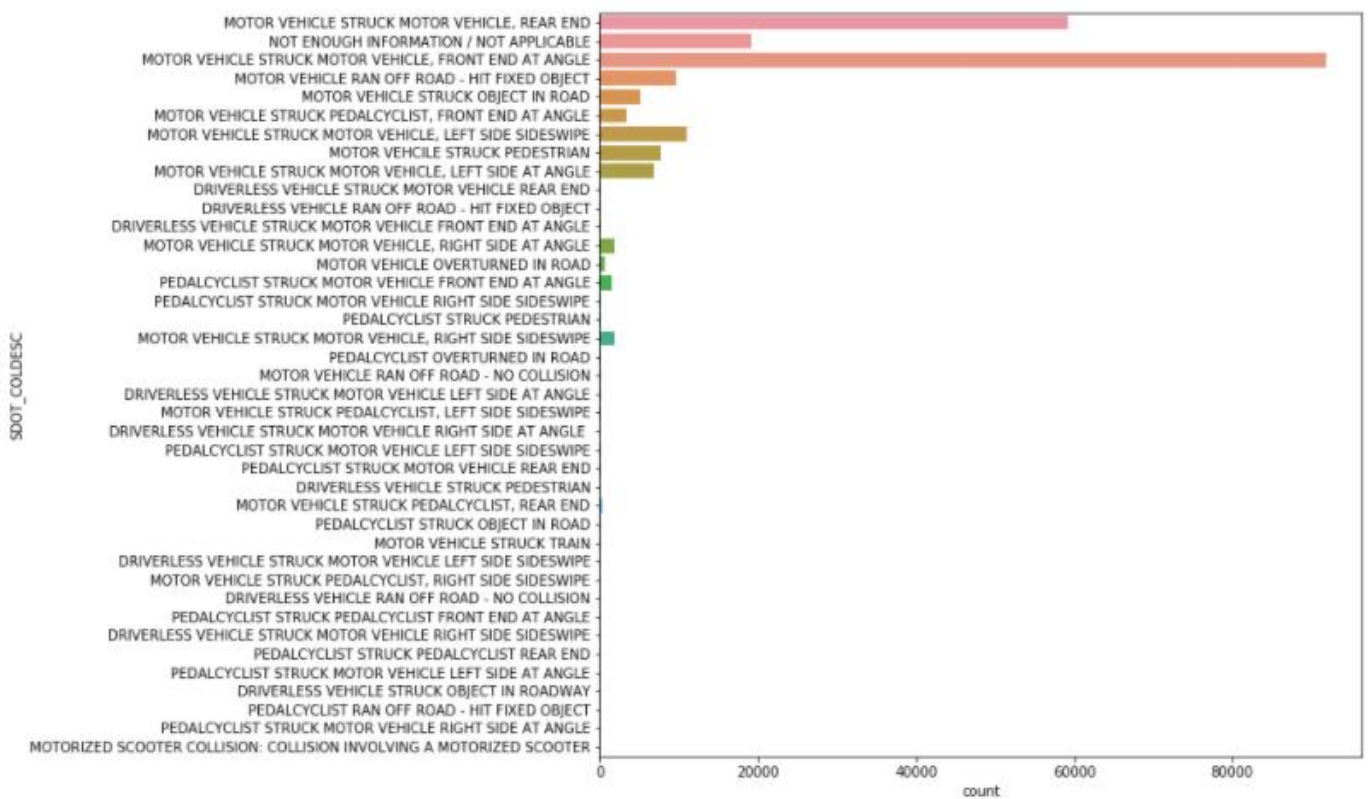
There is ~65% more chance to encounter a collision while taking a Left turn than taking a Right turn.

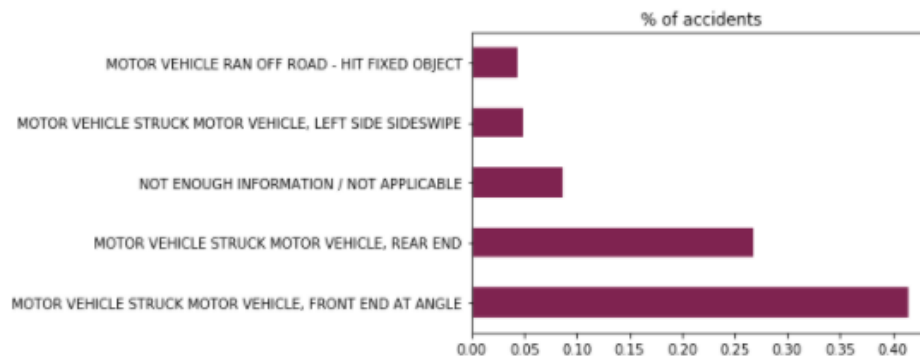
MOTOR VEHICLE RAN OFF ROAD - HIT FIXED OBJECT



Motor Vehicles tend to hit fixed objects more during late night-time.

4.7 Analysis by collision type specified



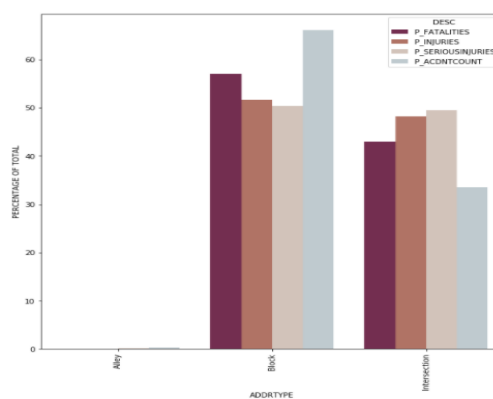


We can see ~40% of collisions are of motor vehicle striking another motor vehicle's front end at an angle.

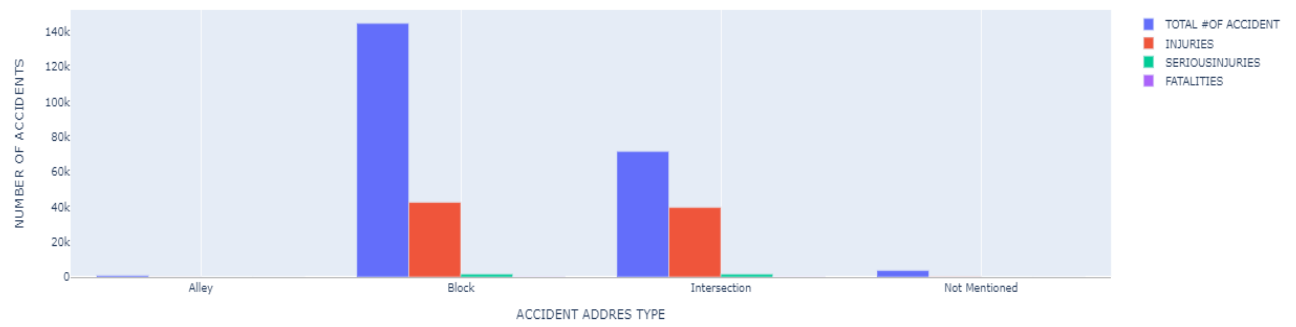
TIMEOFDAY	Late Night	Early Morning	Morning	Noon	Eve	Night
COLLISIONTYPE						
Angles	5.344761	12.471109	22.733693	29.468413	22.017848	7.964176
Cycles	2.357156	17.574958	18.385819	26.136149	29.794456	5.751461
Head On	10.236626	12.705761	18.158436	26.080247	21.810700	11.008230
Left Turn	4.174450	11.808658	19.434744	29.424186	25.696418	9.461545
Other	18.465251	12.900579	17.741313	20.839768	17.919884	12.133205
Parked Car	13.472720	10.063960	20.900766	22.599089	20.256323	12.707142
Pedestrian	4.993410	13.281593	18.641089	26.153170	27.822522	9.108215
Rear Ended	4.028325	12.769724	20.422206	31.976084	23.865990	6.937671
Right Turn	3.440367	12.155963	24.350153	31.230887	20.068807	8.753823
Sideswipe	4.304335	12.858442	22.909973	30.754774	21.467111	7.705365

4.8 Analysis on location

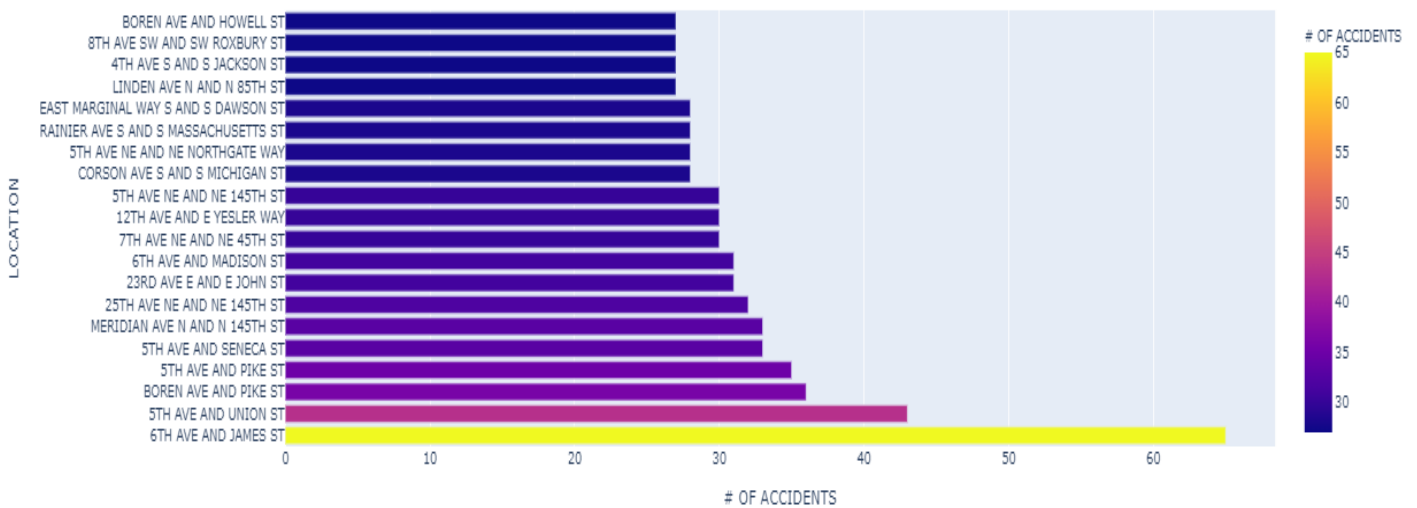
	ADDRTYPE	FATALITIES	INJURIES	SERIOUSINJURIES	ACDNTCOUNT
0	Alley	0	96	5	872
1	Block	212	42580	1687	144661
2	Intersection	160	39680	1656	71693
3	Not Mentioned	0	239	3	3711



Location wise Type of Injuries



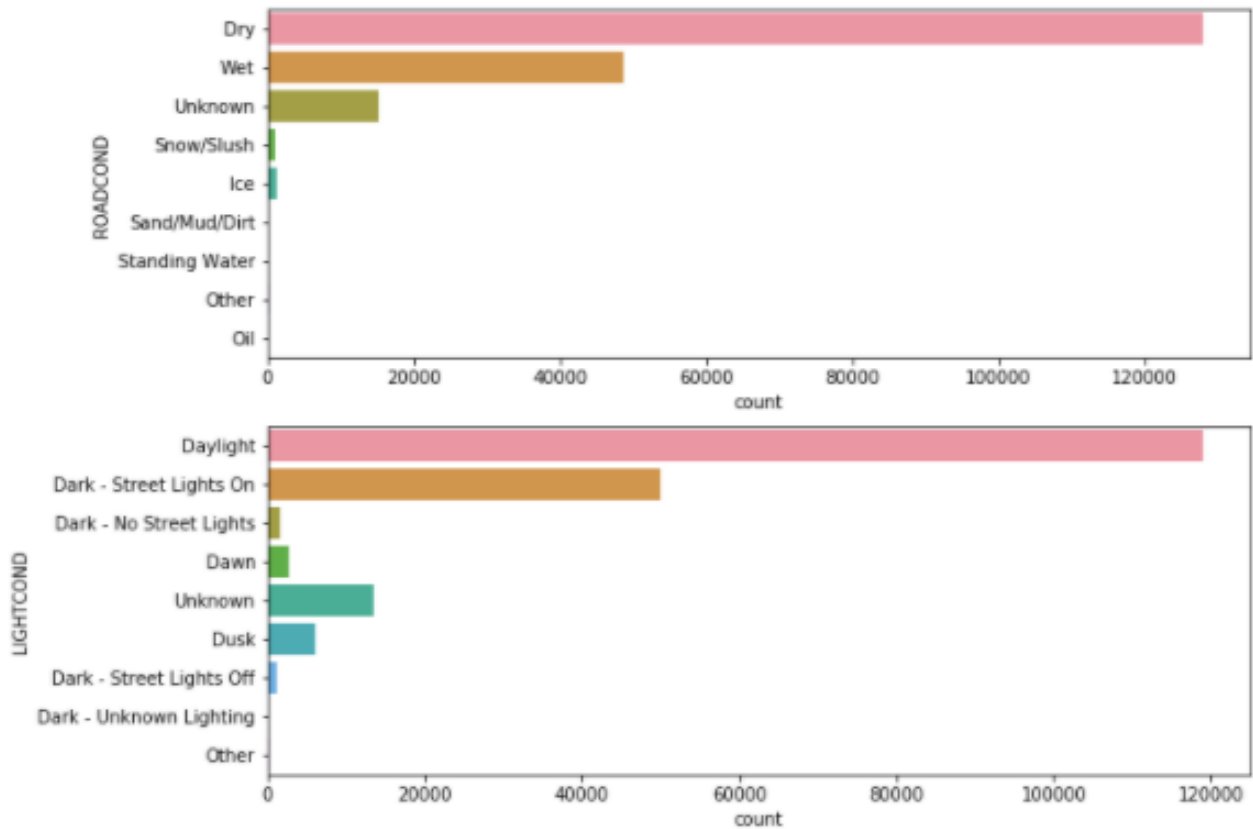
COLLISIONTYPE	Alley	Block	Intersection	Not Mentioned
Angles	58	5771	29545	66
Cycles	8	2500	3393	4
Head On	5	1697	472	6
Left Turn	0	2167	11878	30
Other	294	20153	3890	174
Parked Car	326	45661	1316	1142
Pedestrian	39	2275	5331	5
Rear Ended	11	30083	4267	244
Right Turn	0	1255	1735	16
Sideswipe	16	14757	3939	137



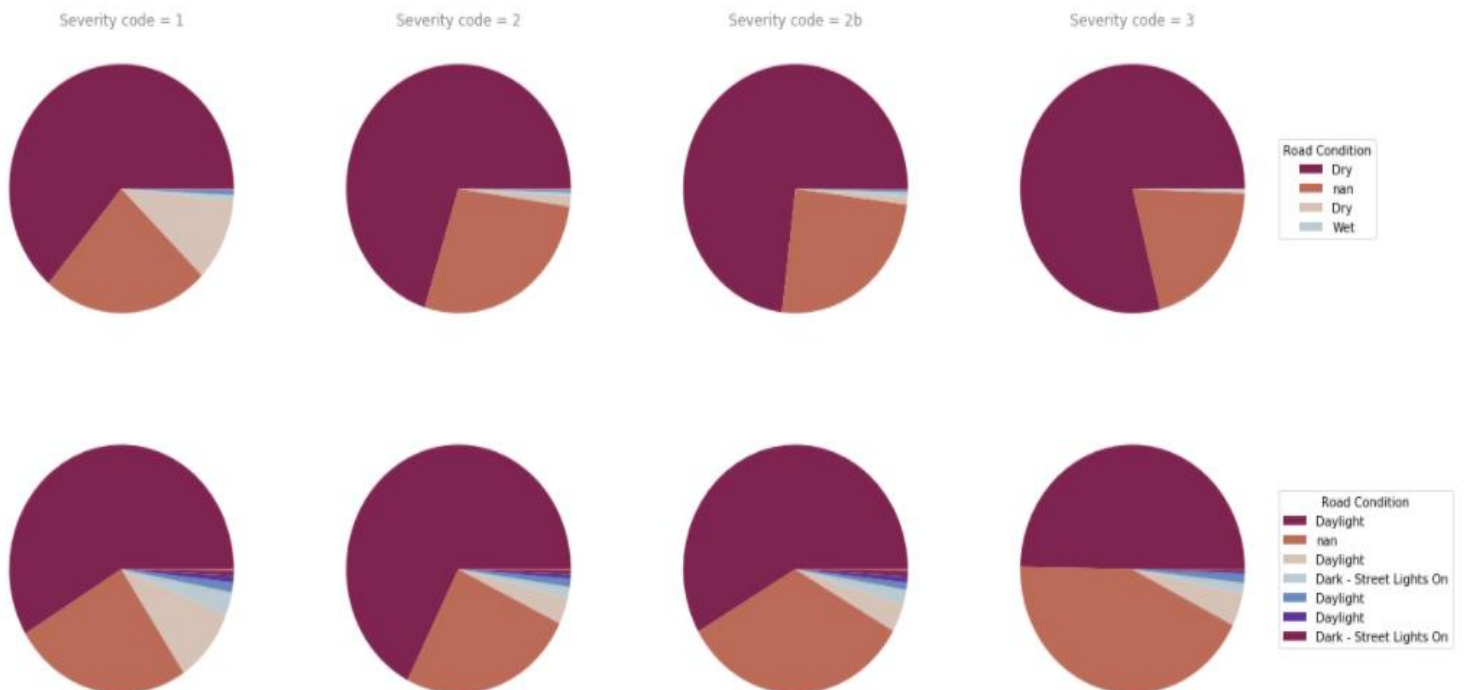
Location "6th AVE AND JAMES ST" has abnormally high number of collisions since 2016

4.9 Analysis of road and light conditions

We have data of road and light conditions where the collision happened.



We can see that majority of collisions happened at dry roads and in daylight. We can check for severity on these conditions.

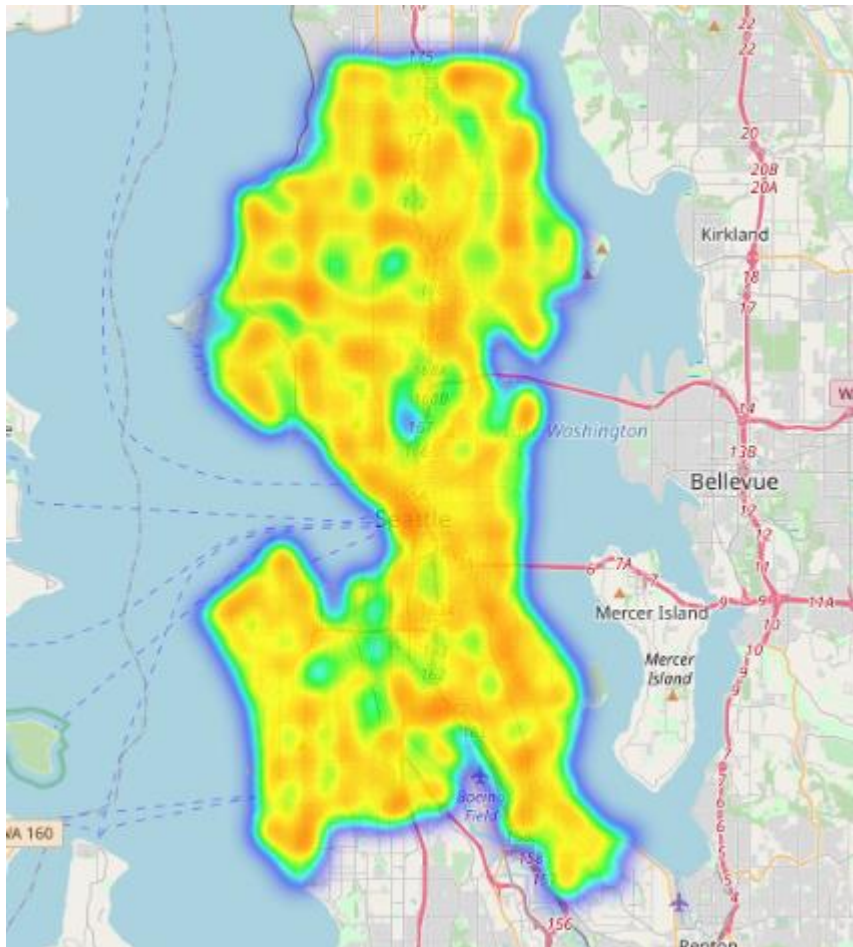
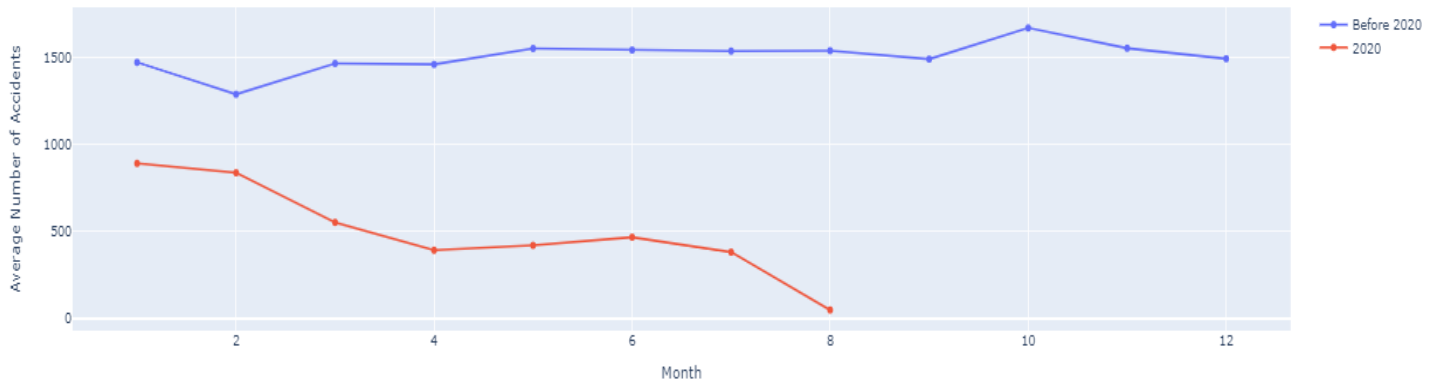


Severe accidents are relatively less in daylight

4.10 Effect of COVID-19 pandemic on vehicle collisions

In 2020, whole world faced a pandemic of the COVID-19 outbreak. Its effect on travel behaviour have also been sudden and wide ranging. We can see that there is a sudden drop in accident cases in 2020, due to COVID-19 pandemic.

Impact of COVID-19 on Road Accidents



5. Conclusion

Road safety is very important nowadays as there has been too much of increase in vehicles on road and it becomes crucial that we follow the traffic rules and laws and save ourselves and our families from road accidents. Also making other people aware of road safety rules could make our roads safe.