Leverage the severity of accidents from environmental conditions

- A case study in Seattle

Yunya Gao

2020-09-30

Content

Under what environmental conditions, property collisions can happen more easily?

Under what environmental conditions, injury collisions can happen more easily?

1 Introduction

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.

source

2 Data description

Traffic collisions data from

Seattle Department of Transportation

There are 37 attributes in the given table, which include location (X and Y), timestamps, severity types, weather, light, road conditions, etc.

- SEVERITYCODE: dependent variables to indicate the severity of the traffic collisions;
- WEATHER, ROADCOND, LIGHTCOND, ADDRTYPE: independent variables
- INCDATE: temporal variables

3 Methodology

Data preparation

Original Data

	SEVERITYDESC	INCDATE	WEATHER	ROADCOND	LIGHTCOND	COLLISIONTYPE	ADDRTYPE
0	Injury Collision	2004/01/01 00:00:00+00	Overcast	Dry	Daylight	Other	Block
1	Property Damage Only Collision	2004/01/01 00:00:00+00	Unknown	Ice	Dark - Street Lights On	Parked Car	Block
2	Property Damage Only Collision	2004/01/01 00:00:00+00	Raining	Wet	Daylight	Rear Ended	Block
3	Property Damage Only Collision	2004/01/01 00:00:00+00	Raining	Wet	Dark - Street Lights On	Angles	Intersection
4	Injury Collision	2004/01/01 00:00:00+00	Overcast	Wet	Dark - Street Lights On	Parked Car	Block

	False	True
SEVERITYDESC	194673.0	NaN
INCDATE	194673.0	NaN
WEATHER	173669.0	21004.0
ROADCOND	174451.0	20222.0
LIGHTCOND	175784.0	18889.0
COLLISIONTYPE	166066.0	28607.0
ADDRTYPE	192747.0	1926.0

	False
SEVERITYDESC	146877
INCDATE	146877
WEATHER	146877
ROADCOND	146877
LIGHTCOND	146877
COLLISIONTYPE	146877
ADDRTYPE	146877

Dealing with missing data

	Injury Collision	Property Damage Only Collision	Clear	Overcast	Raining	Snowing	Dry	Ice	Snow/Slush	Wet	...	Cycles	Head On	Left Turn	Parked Car	Pedestrian	Rear Ended	Right Turn	Sideswipe
0	0	1	0	0	1	0	0	0	0	1	...	0	0	0	0	0	1	0	0
1	0	1	0	0	1	0	0	0	0	1	...	0	0	0	0	0	0	0	0
2	1	0	0	1	0	0	0	0	0	1	...	0	0	0	1	0	0	0	0
3	0	1	0	1	0	0	0	0	0	1	...	0	0	0	0	0	1	0	0
4	0	1	0	0	1	0	0	0	0	1	...	0	0	0	0	0	0	0	0

Ddependent variables: Injury Collision, Property Damage Only Collision.

Independent variables related to environmental conditions: 25 (weather, road, light, address type)

The total number of data is 146876.

One-hot encoding categorical data

3 Methodology

Analysis

Exploratory analysis:

- Time series analysis (Are there any patterns for collisions in terms of time?)

- Regression analysis (What are the relations between different variables?)

Two types of models:

- KNN for modelling categorical data

- Multiple linear regression for modelling monthly aggregated number of traffic collisions

3 Methodology

Analysis

Exploratory analysis:

- Time series analysis (Are there any patterns for collisions in terms of time?)

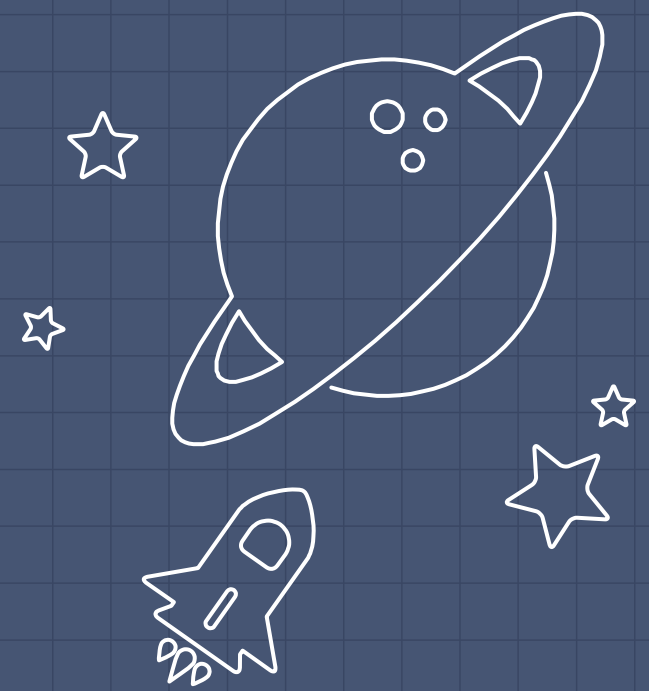
- Regression analysis (What are the relations between different variables?)

Two types of models:

- KNN for modelling categorical data

- Multiple linear regression for modelling monthly aggregated number of traffic collisions

Results and Discussion



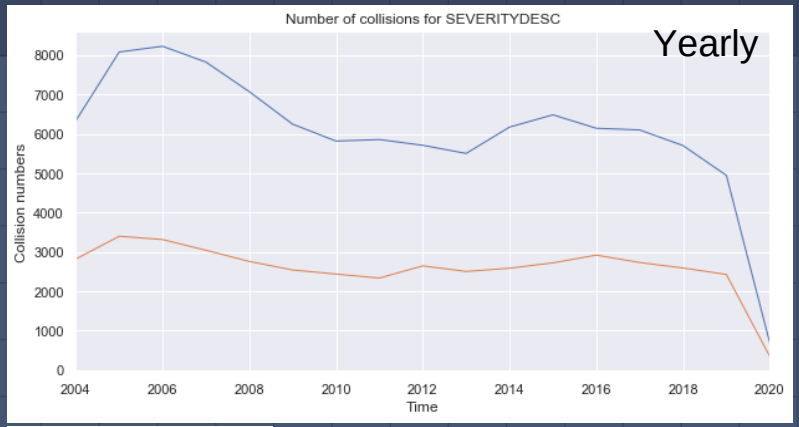
Highlight:

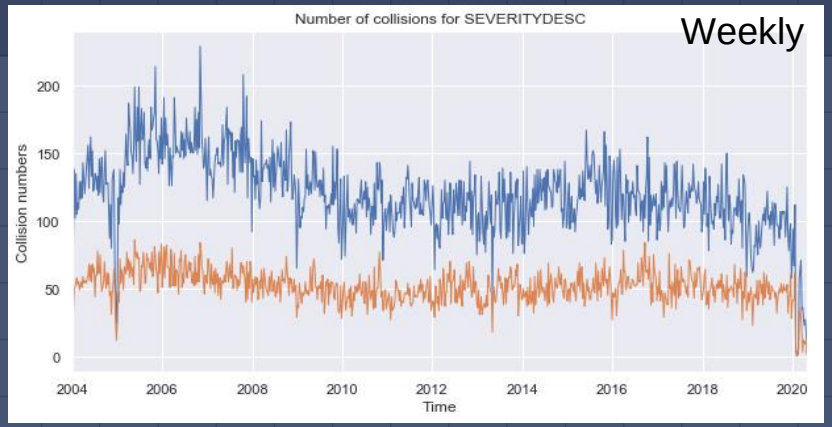
- Property collisions tend to happen when environmental conditions are “GOOD”. This may be related to over-confidence.
- Injury collisions tend to happen when environmental conditions are “BAD”.

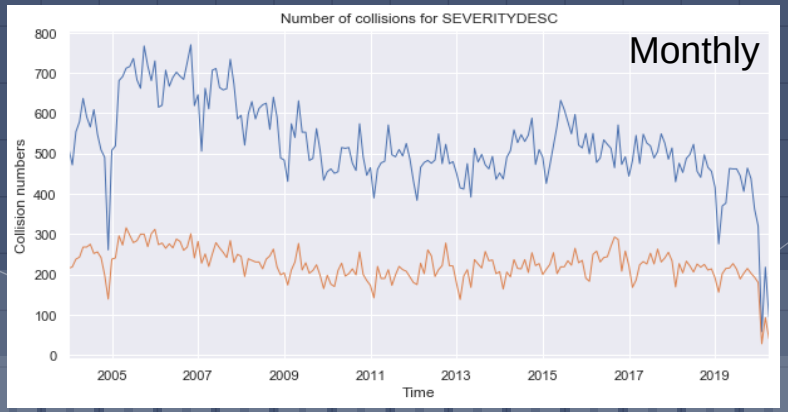
- Linear regression is not suitable for this case.

4 Results and Discussion

Time series analysis



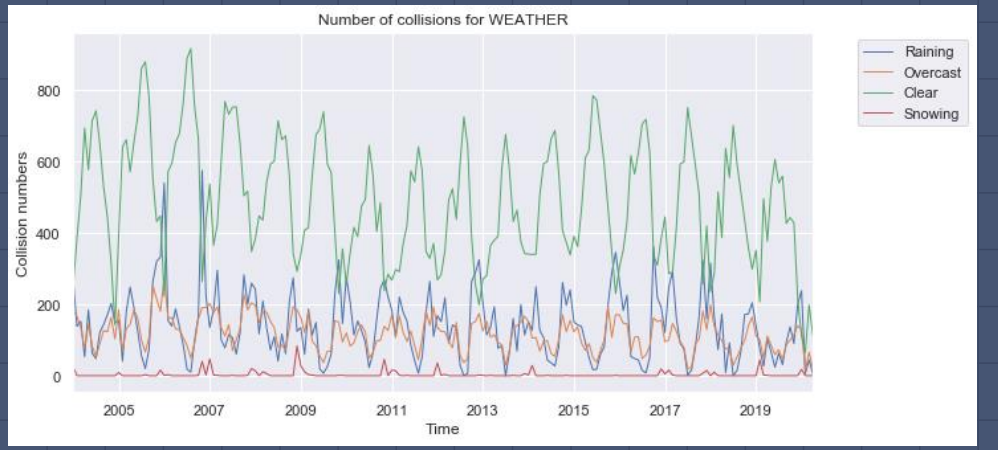




- The number of property collisions are much higher - than that of injury collisions.
- It can be seen some "periodic cycles" like sin or cos curves on both two types.
- The peaks and valleys are mostly matched in terms of time.

4 Results and Discussion

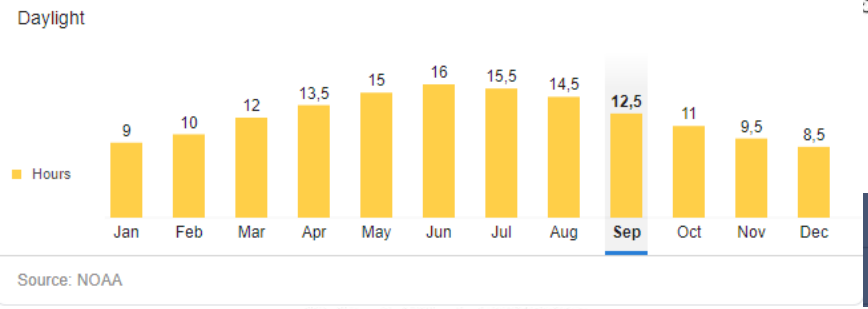
- Time series analysis

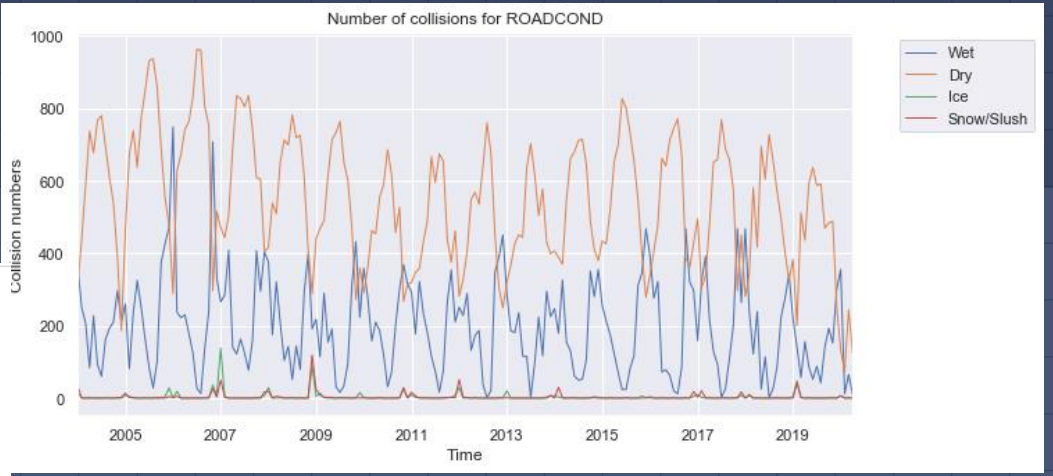


Weather conditions

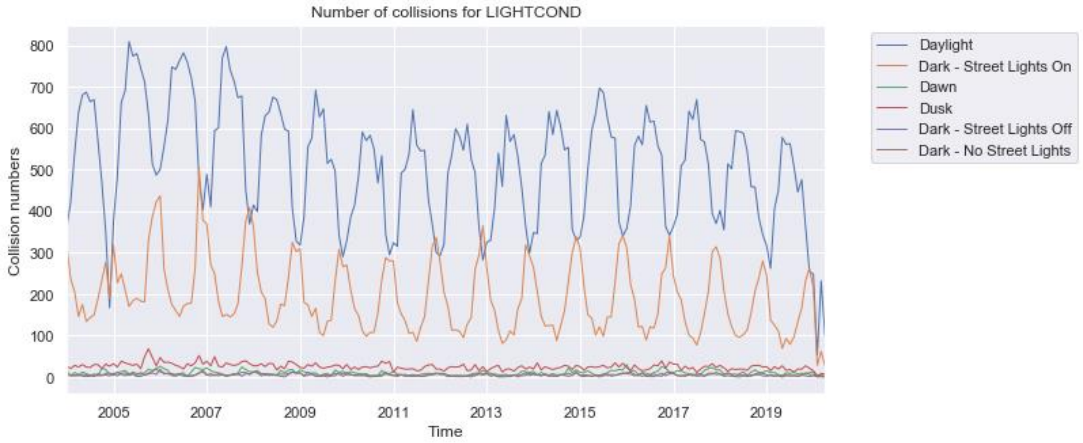
4 Results and Discussion

- Time series analysis





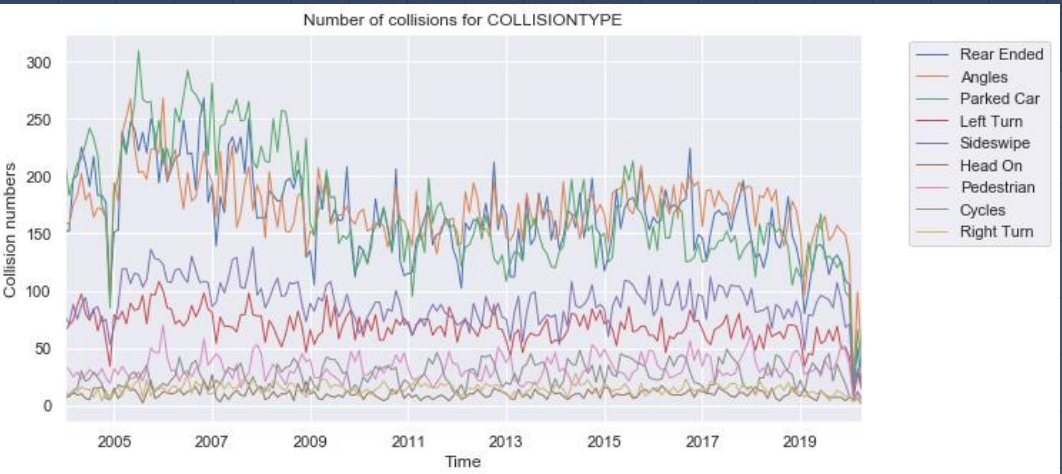
Road conditions



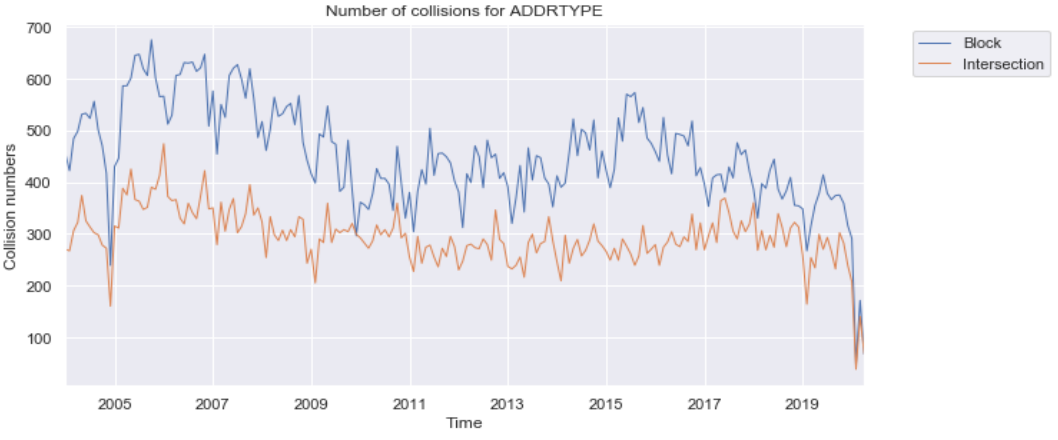
Light conditions

4 Results and Discussion

Time series analysis



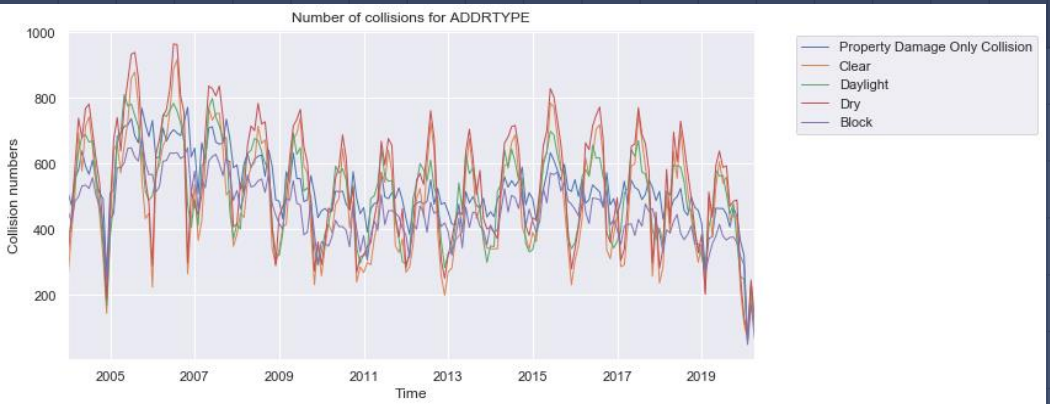
Collision types

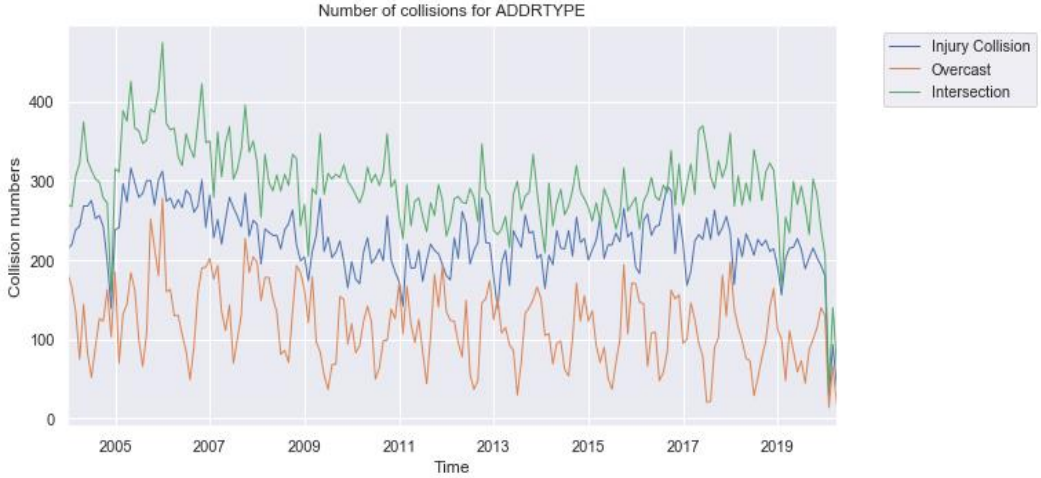


Address types

4 Results and Discussion

Time series analysis



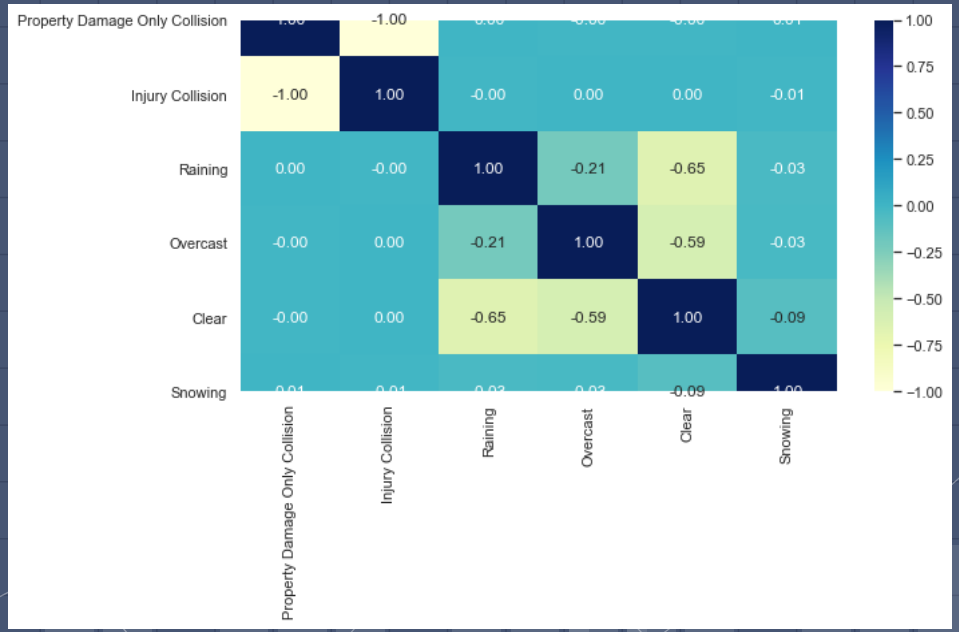


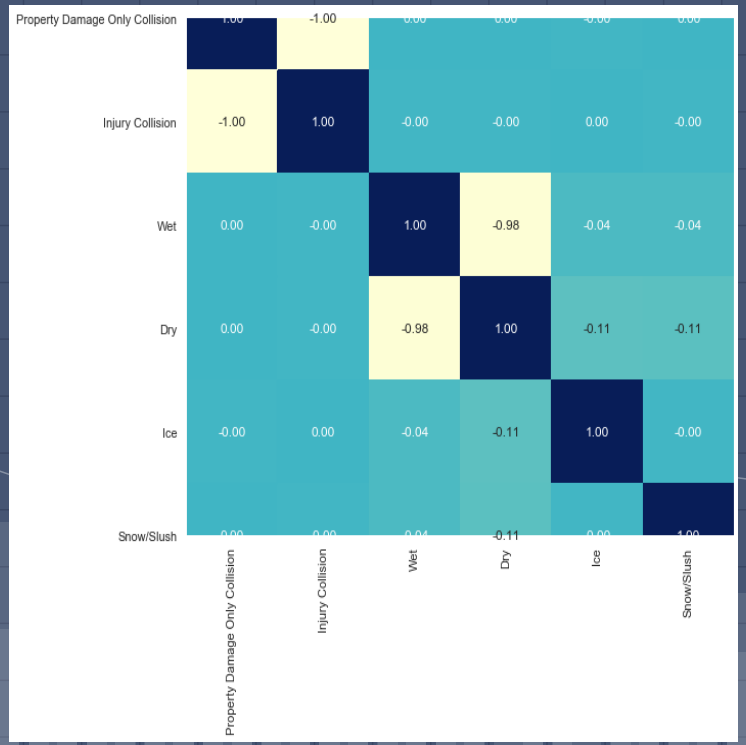
Property collisions and variables with similar shapes of curves

Injury collisions and variables with similar shapes of curves

4 Results and Discussion

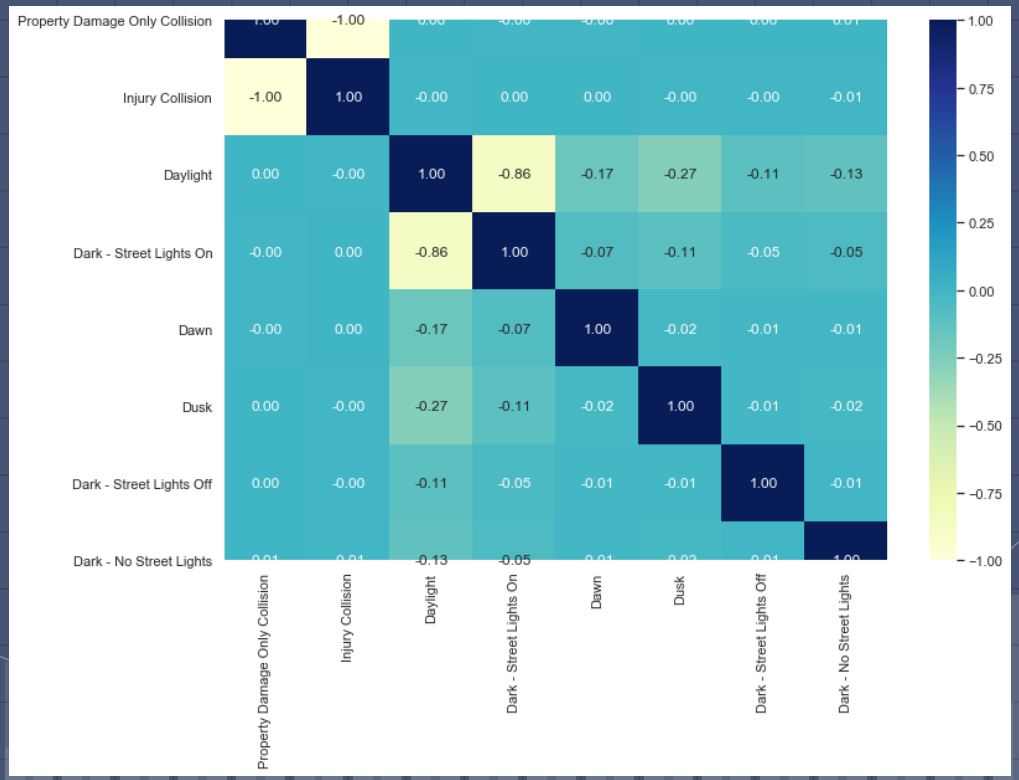
Regression analysis

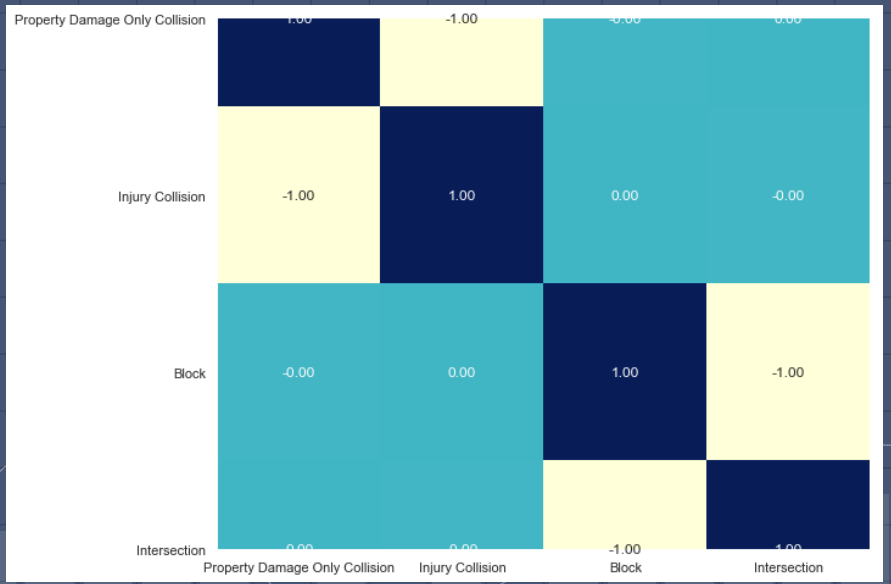




4 Results and Discussion

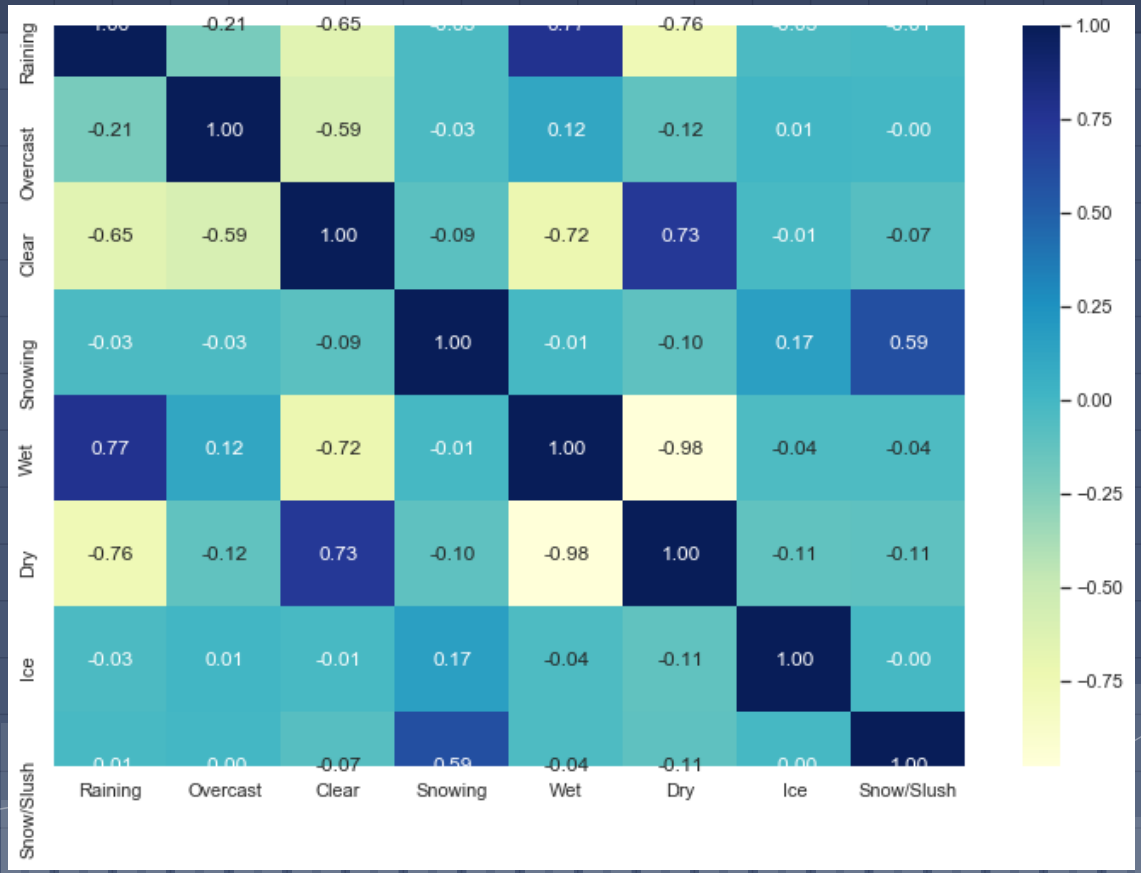
Regression analysis





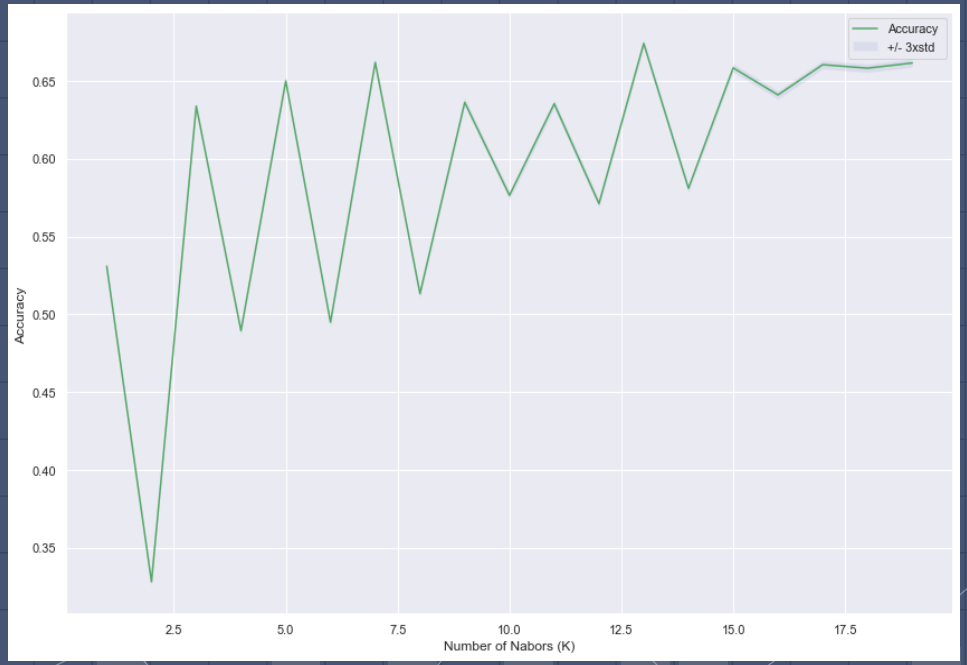
4 Results and Discussion

Regression analysis



4 Results and Discussion

KNN model



80% of data were selected for training while 20% for testing

When K=13, the model performs best with an accuracy score 0.67.

Help analyse the possibility of the occurrence of traffic collisions, such as predicting how traffic collisions can change under the big background of climate change

4 Results and Discussion

Multiple Linear Regression

	Clear	Overcast	Raining	Snowing
Injury Collision	-14.457162	-3.877285	-9.955556	3.494033
Property Damage Only Collision	14.457162	3.877285	9.955556	-3.494033

	Dry	Ice	Snow/Slush	Wet
Injury Collision	360.545568	27.968318	20.977232	264.202792
Property Damage Only Collision	-360.545568	-27.968318	-20.977232	-264.202792

	Dark - No Street Lights
Injury Collision	-3.416603
Property Damage Only Collision	3.949881

	Dark - Street Lights Off
Injury Collision	-4.401069
Property Damage Only Collision	4.947450

	Dark - Street Lights On	Dawn	Daylight
Injury Collision	-31.072708	-1.501542	-59.998332
Property Damage Only Collision	45.868516	2.453250	84.486315

	Dusk	Angles	Cycles	Head On
Injury Collision	-2.653365	-51.134256	-16.673632	-7.136045
Property Damage Only Collision	4.058756	75.590338	25.425142	10.431165

	Left Turn	Parked Car	Pedestrian	Rear Ended
Injury Collision	-26.242889	-109.314552	-19.750121	-68.971935
Property Damage Only Collision	37.895440	151.221451	28.336279	101.785917

	Right Turn	Sideswipe	Block	Intersection
Injury Collision	-10.627652	-39.670071	66.611518	37.669887
Property Damage Only Collision	14.437407	55.279158	-66.611518	-37.669887

80% of data were selected for training while 20% for testing

R2 score: 0.9607

Initially, building multiple regression models for monthly aggregated data is for urban planners to see which factors play an important role in resulting traffic collisions.
But as mentioned above, the linear regression may not be suitable for this case.

5 Conclusions

- Most of the curves of selected variables related to environmental conditions are periodic. The basic reason for is related to the location and climate of Seattle.
- Property collisions tend to happen when environmental conditions are "good" for driving. Good environmental conditions may let people be over-confident on driving.
- The trained KNN model can be helpful for urban planners or traffic department to check the possibility of occurrence of traffic collisions for different environmental conditions so as to make some preparations in advance.
- Linear regression models may not work well for this case.

6 Recommendations

- As for property collisions, increase the number of announcements in car radios to remind them or other ways especially when the weather conditions are good for driving
- As for injury collisions, make more mirrors in the places like intersection and encourage people to install some equipment for their vehicles when snowing.

Thanks for reading.
Yunya Gao