EMSAS

Emergency Medical Services Alert System

Ruben Goncalves

November 2020

The problem

Road accidents happen.

Road accidents can lead to injuries so it is vital that a sufficient number of paramedics and emergency support staff is present when needed.

The business needs

There is a need for a model that can predict if a serious incident is likely to occur to alert the Emergency Medical Services (EMS) which can prepare paramedics in advance and improve readiness.

Dataset

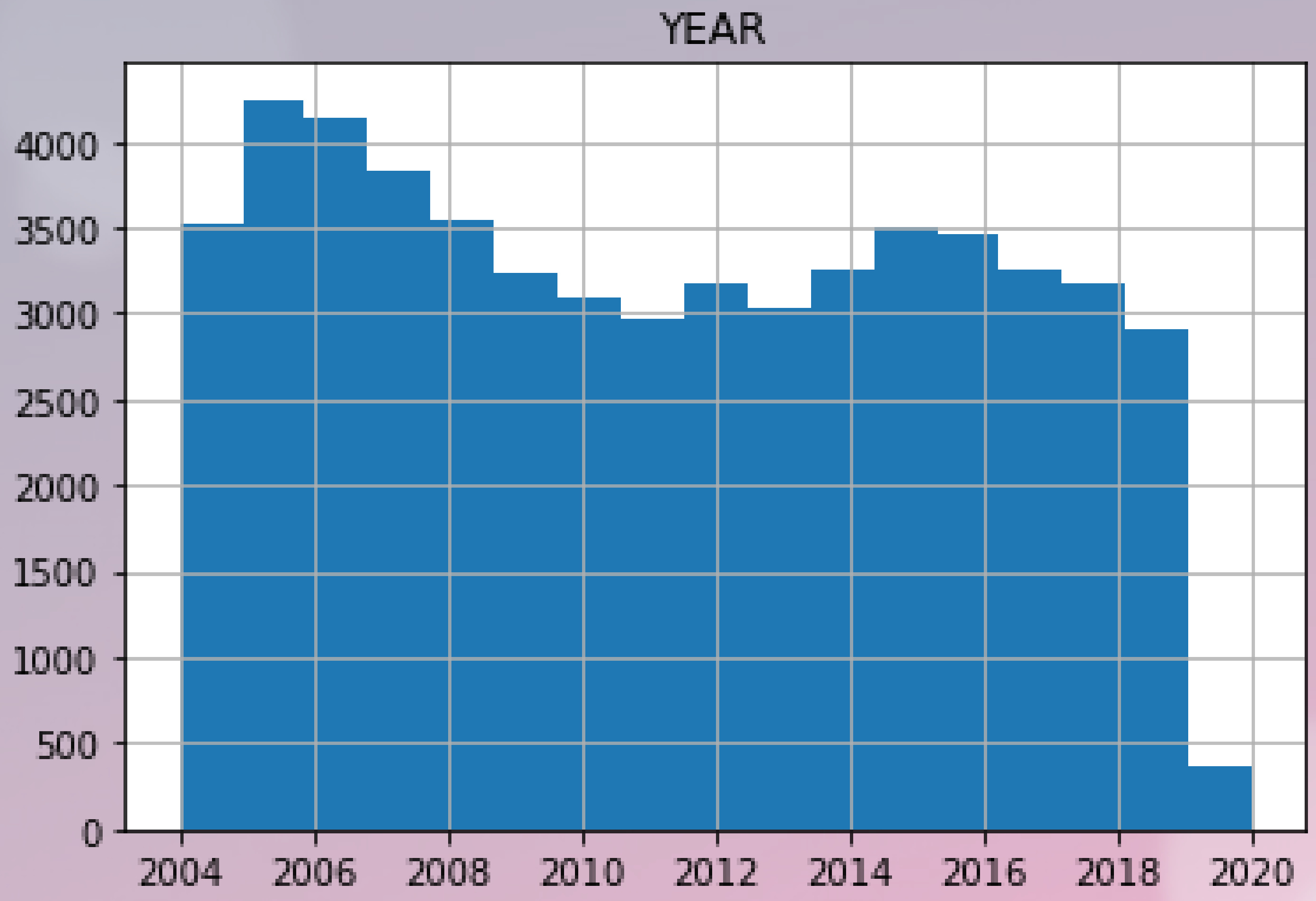
The accident data collected is from the city of Seattle.

To allow the model development only certain variables were kept for the purpose of analysis.

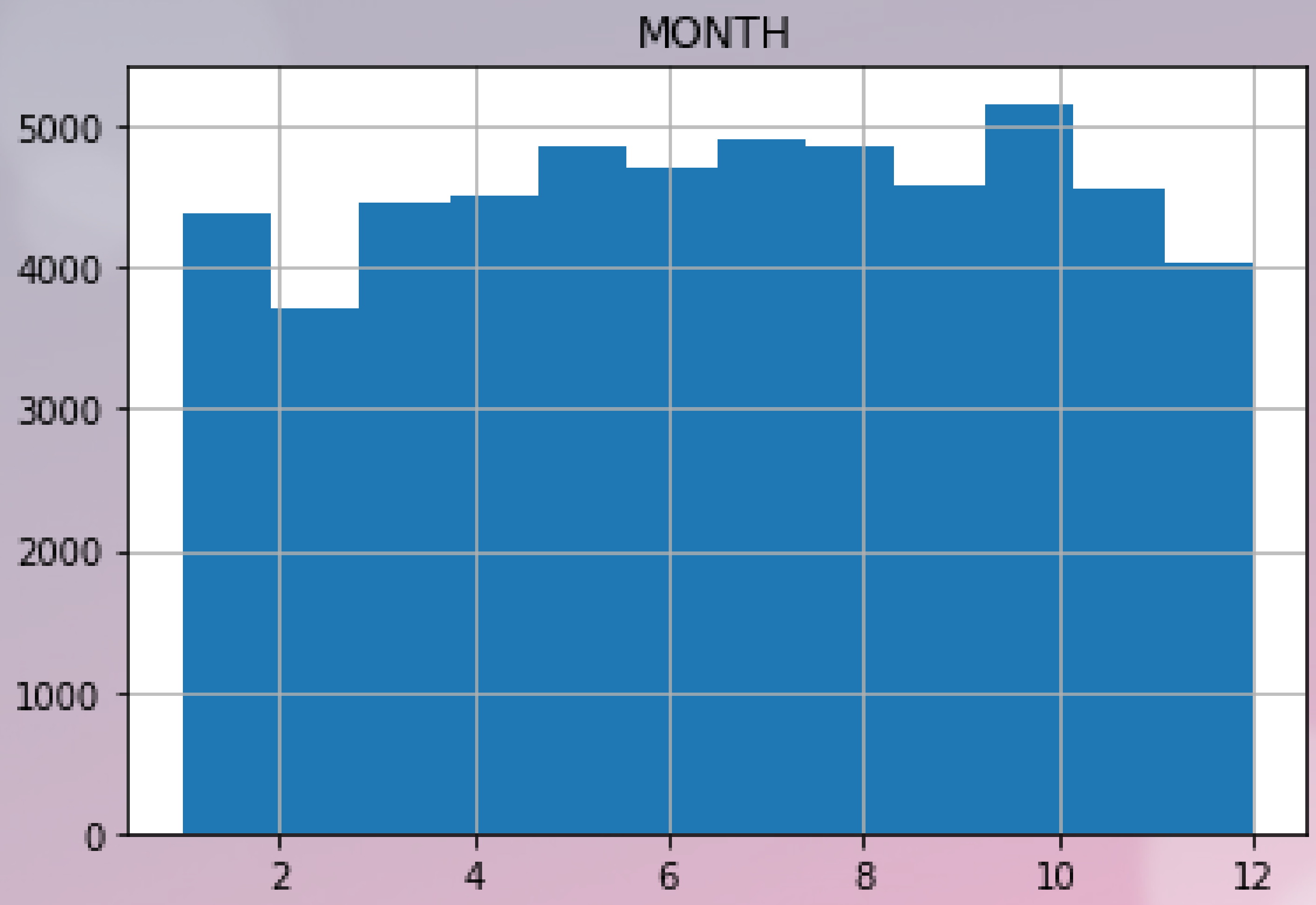
Table 1

Variable removed from the dataset	Reason for the removal from the dataset
OBJECTID, INCKEY, COLDETKEY, REPORTNO and SDOTCOLNUM	Are for internal use of the city of Seattle and are unique identifiers that convey no information.
EXCEPTRSNCODE, EXCEPTRSNDESC	Have no metadata to describe the features and STATUS is similarly undocumented.
SEVERITYCODE.1 and SEVERITYDESC	Are duplicates of the target variable and convey the same information.
PERSONCOUNT, PEDCOUNT, PEDCYLCOUNT, VEHCOUNT	While may be useful for analysis post incident, they are descriptive in nature and can not be used for prediction as one can not in advance know many people, pedestrians, bicycles or vehicles will be involved in an accident.
COLLISIONTYPE, JUNCTIONTYPE, INATTENTIONIND, UNDERINFL, PEDROWNOTGRNT, SPEEDING, SEGLANEKEY, HITPARKEDCAR	Have the same limitations as the features above since one can not know if people will be inattentive, under the influence, if they grant a pedestrian right of way, are speeding, on which lane they are, the type of junction or collision when accident happens or if they hit a parked car until after the accident.
SDOT_COLCODE, SDOT_COLDESC, ST_COLCODE and ST_COLDESC	Are highly descriptive features that while possessing plenty of useful data, they can not be used as a basis for an alert of the Emergency Medical Services.
CROSSWALKKEY, INTKEY and ADDRTYPE	Duplicates - Information is already present in the coordinate variables X and Y.

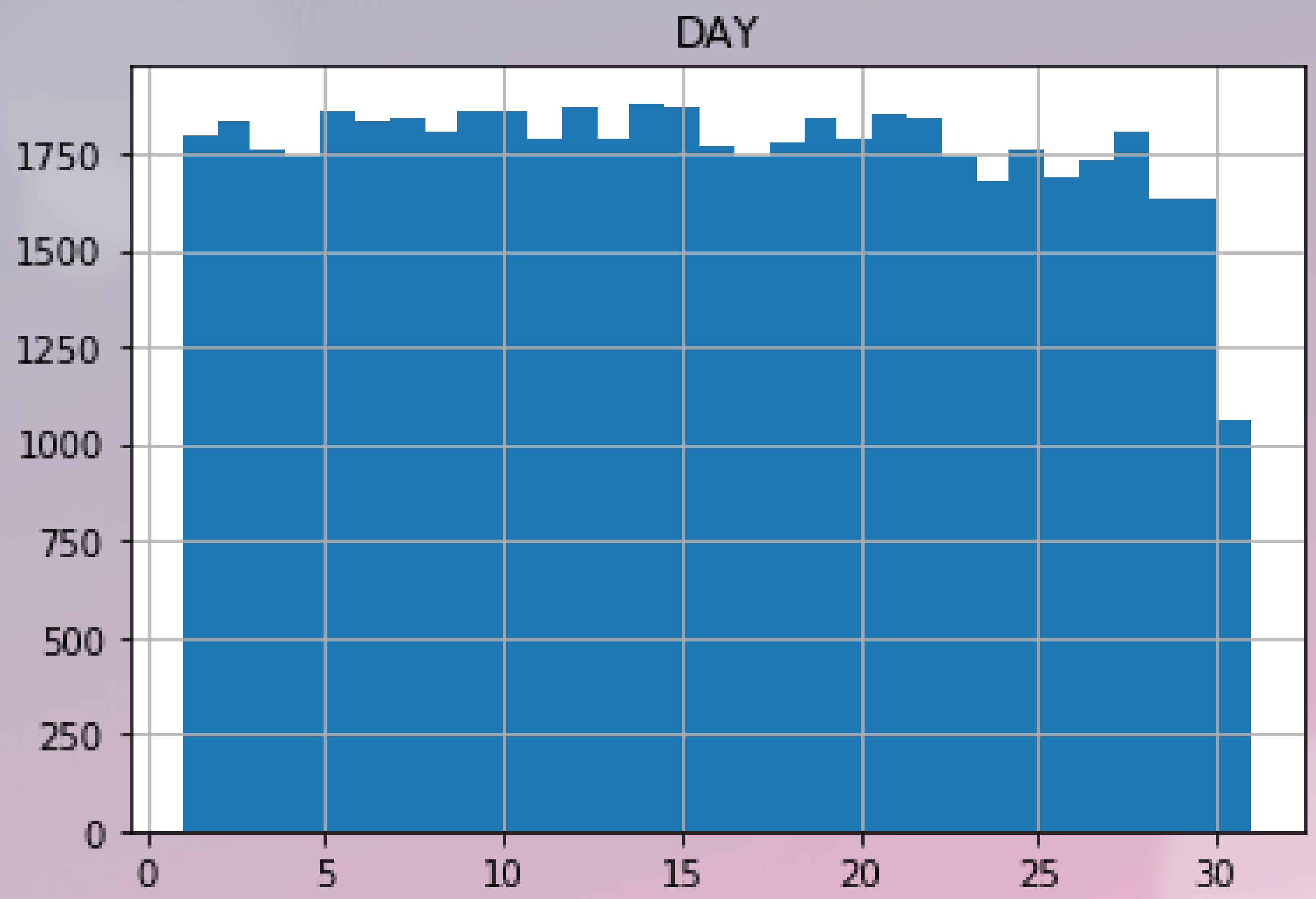
Analysis of dataset of Total number of Injuries (Class 2) based on the Year the accidents occurred.



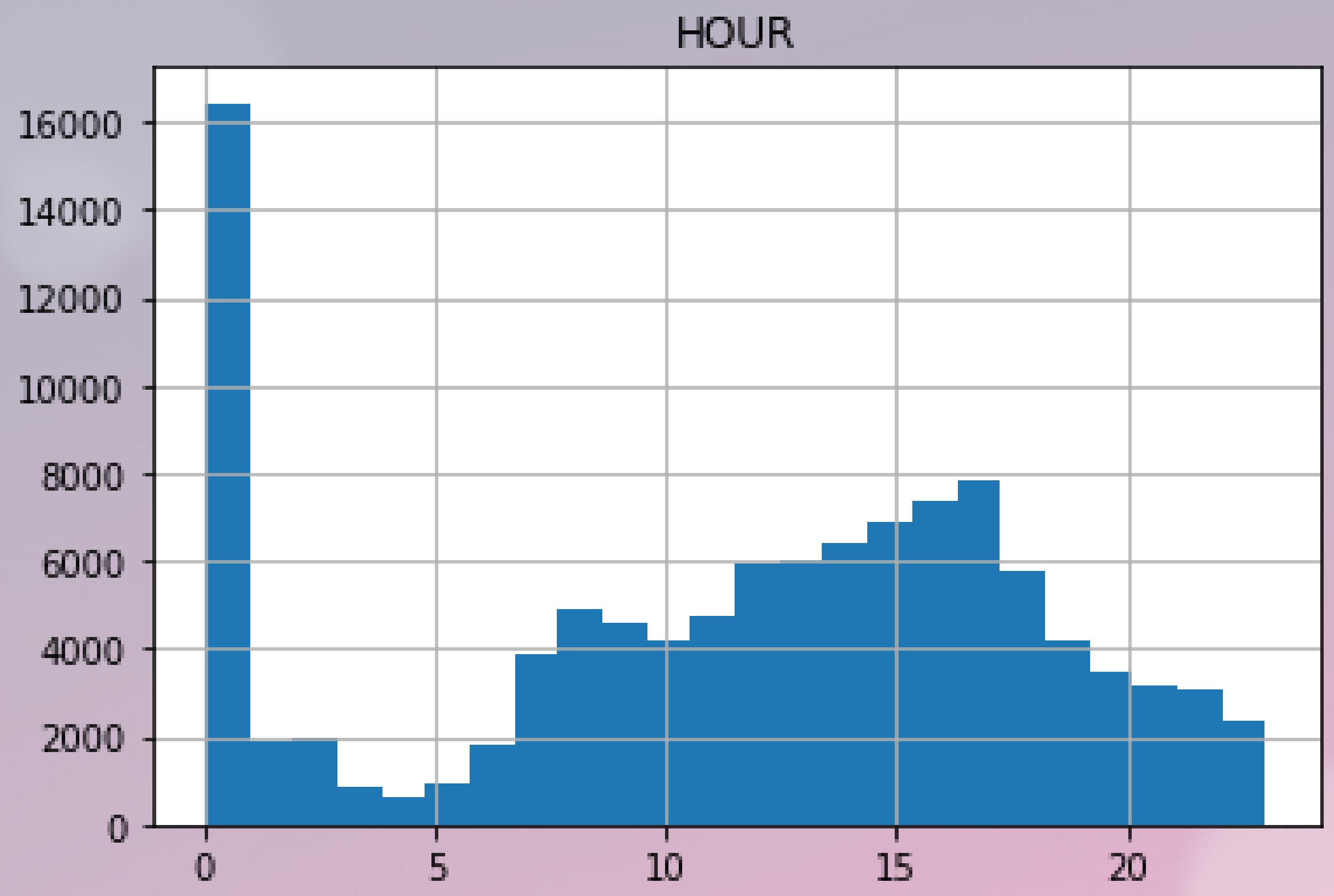
Analysis of dataset of Total number of Injuries (Class 2) based on the Month the accidents occurred.



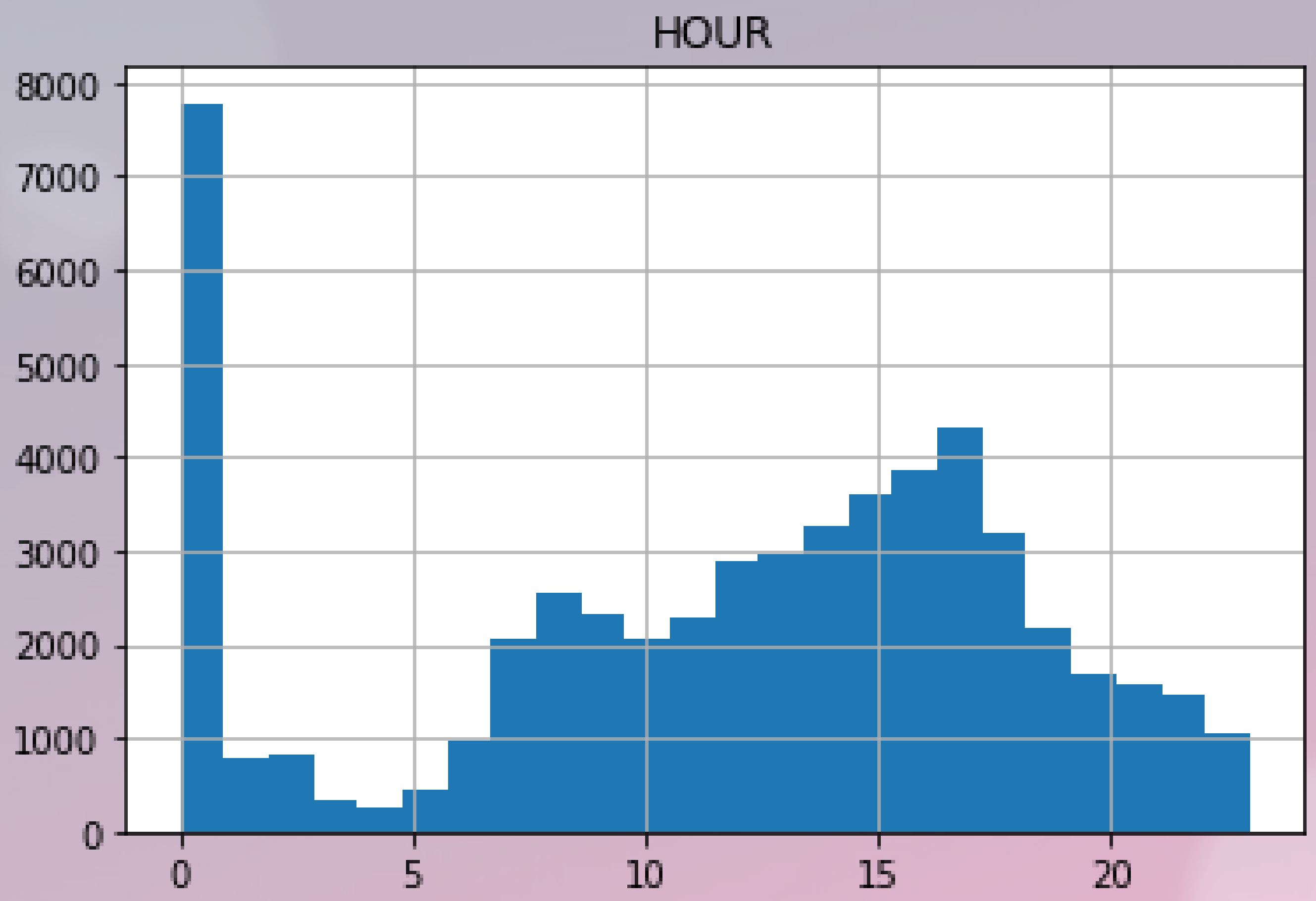
Analysis of dataset of Total number of Injuries (Class 2) based on the Day the accidents occurred.



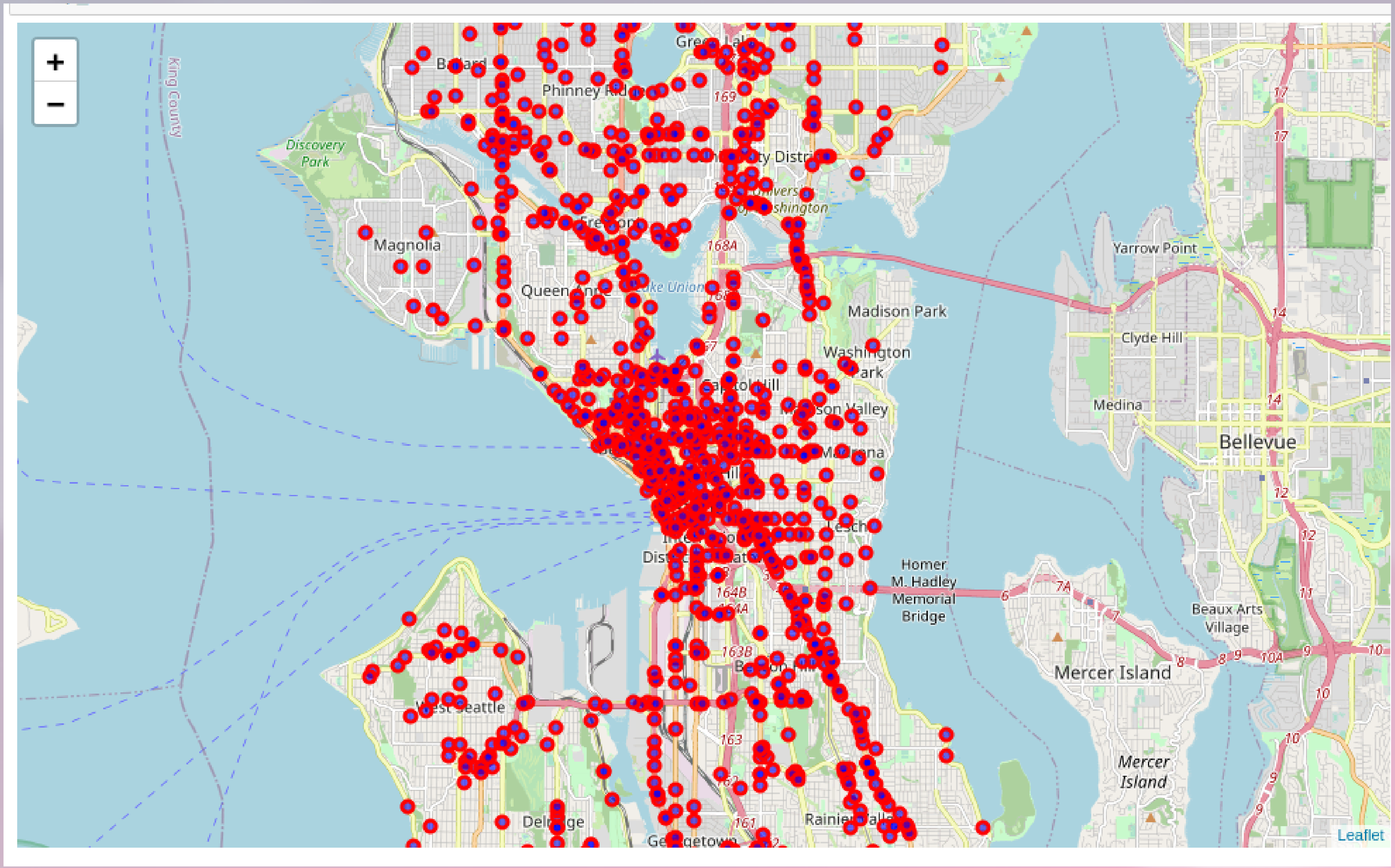
Analysis of dataset of Total number of Accidents (Class 1 + Class 2) based on the Hour the accidents occurred.



Analysis of dataset of Total number of Injuries (Class 2) based on the Hour the accidents occurred.

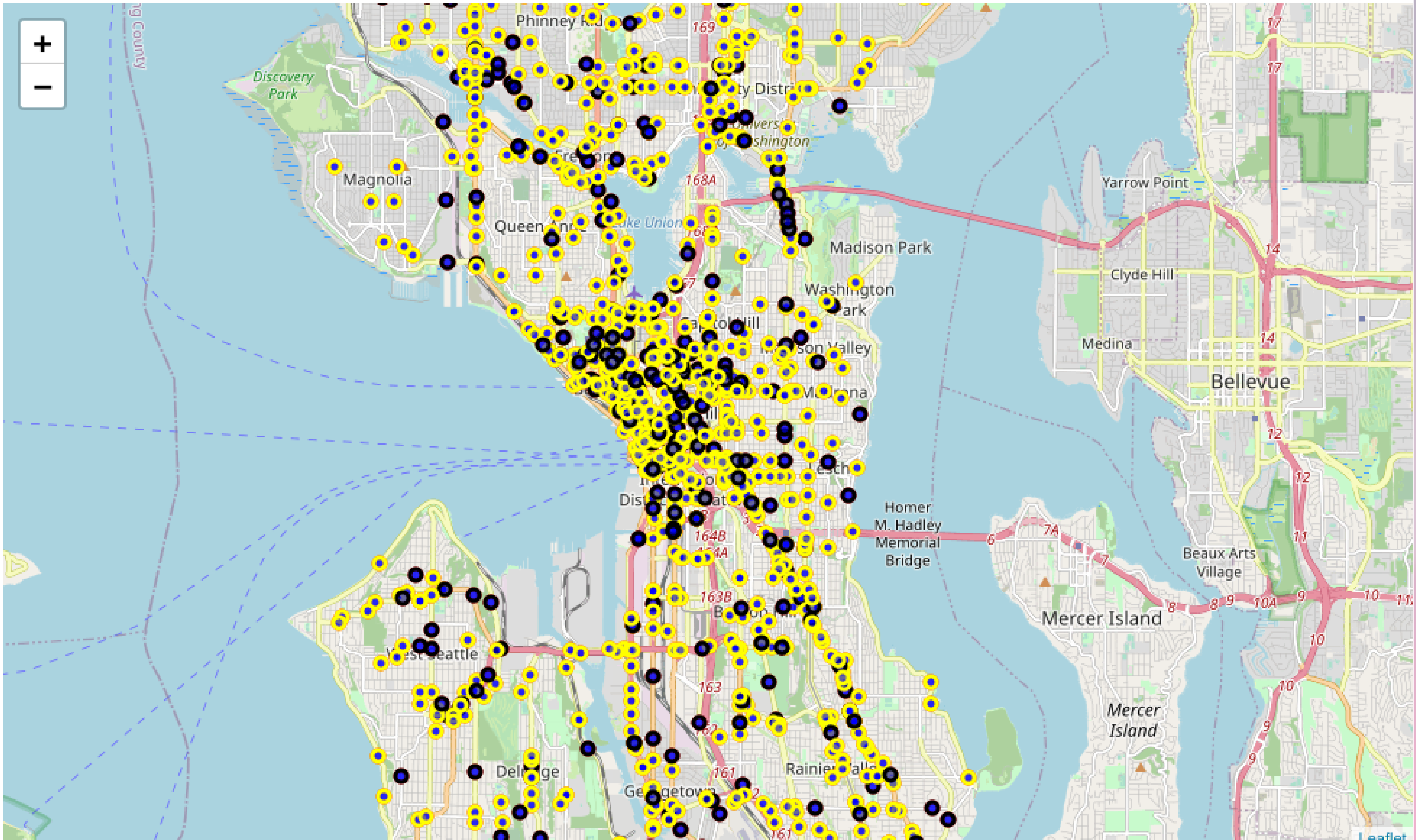


Visual representation of the injuries (Class 2 entries) on a Seattle map. Number of entries limited to 1500.

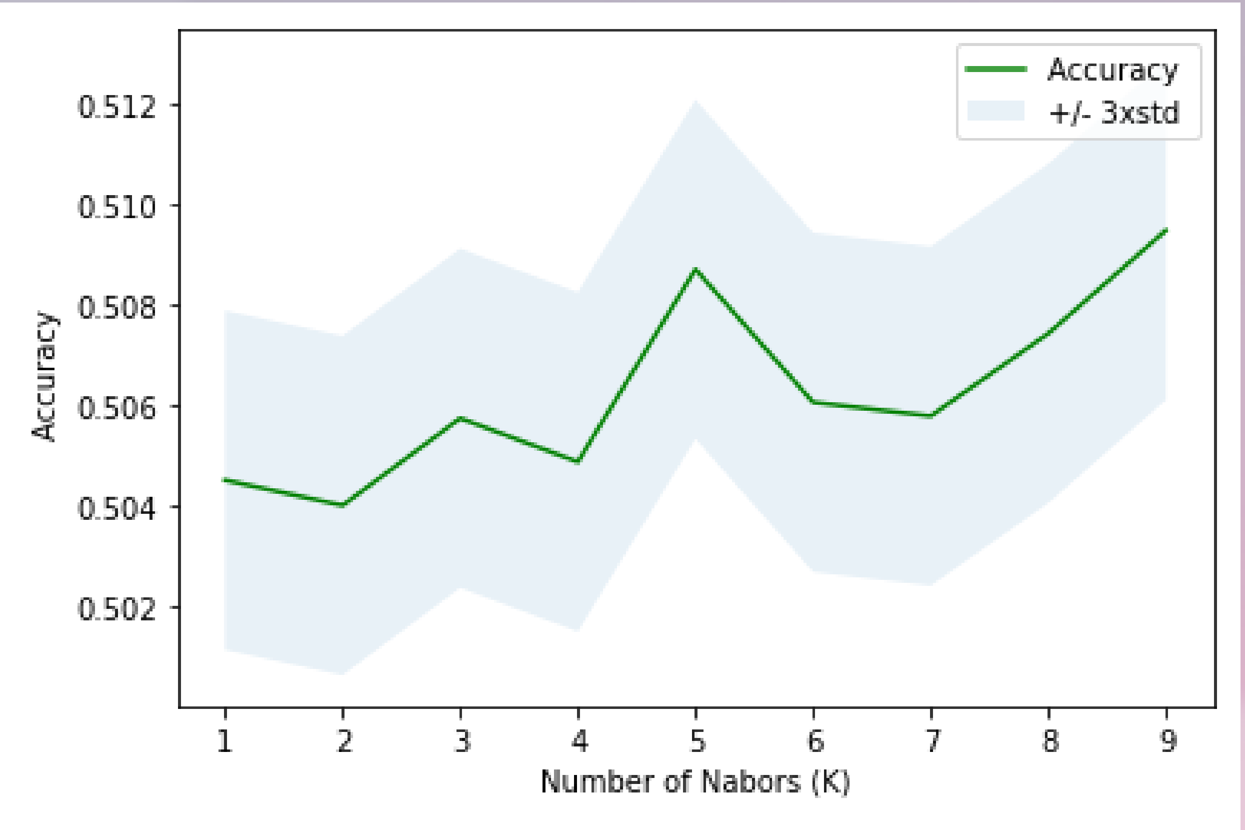


Visual representation of the injuries (Class 2 entries) on a Seattle map by period of the day.

Number of entries limited to 1500. Blue dots – accidents that happen between midnight and 5 am. Yellow dots – accidents that happen between 6 am and 11 pm.



Graphic representation of the KNN model accuracy for K between 1 and 9.



Classification report for the KNN model. K = 9.

	precision	recall	f1-score	support
1	0.51	0.50	0.51	10895
2	0.51	0.51	0.51	10990
micro avg	0.51	0.51	0.51	21885
macro avg	0.51	0.51	0.51	21885
weighted avg	0.51	0.51	0.51	21885

Classification report for the Decision Tree model. Max depth =10.

	precision	recall	f1-score	support
1	0.51	0.57	0.54	10895
2	0.52	0.46	0.49	10990
micro avg	0.52	0.52	0.52	21885
macro avg	0.52	0.52	0.52	21885
weighted avg	0.52	0.52	0.52	21885

Classification report for the Logistic Regression model.

	precision	recall	f1-score	support
1	0.52	0.43	0.47	10895
2	0.52	0.60	0.56	10990
micro avg	0.52	0.52	0.52	21885
macro avg	0.52	0.52	0.51	21885
weighted avg	0.52	0.52	0.51	21885

Classification report for the SVM model.

	precision	recall	f1-score	support
1	0.52	0.35	0.42	10895
2	0.52	0.68	0.59	10990
micro avg	0.52	0.52	0.52	21885
macro avg	0.52	0.52	0.50	21885
weighted avg	0.52	0.52	0.50	21885

Frequency Analysis

Incidence was determined using the formula:

Incidence = (number of class 2 entries / (total number of entries))
= (58188/(136485+58188))
= 0.29890123437764865

Number of daily accidents was determined using the data from 2019 as it was most recent complete year available. The formula used was:

Accperday = (8246/365)
= 22.59178082191781

Number of injuries per day was determined using the formula:

Injperday = Accperday*Incidence
= 6.752711174460523

Expected number of injuries per hour at any given day.

	TOTALINJURIESPERHOURPERDAY
0	0.130206
1	0.105639
2	0.110553
3	0.077796
4	0.059780
5	0.085985
6	0.137576
7	0.328382
8	0.404540
9	0.369327
10	0.340666
11	0.380792
12	0.432383
13	0.416824
14	0.484793
15	0.438116
16	0.444667
17	0.513455
18	0.381611
19	0.285799
20	0.242397
21	0.222743
22	0.205546
23	0.153136

Conclusion

An estimate of the number of injuries per hour was found as well as that accidents are clustered near the city centre which can be used to plan and inform alertness levels.

No significant relation was found between road condition, light conditions and weather in determining if an accident will result in injury or not.