

# Predicting Incident Severity

IBM DATA SCIENCE - CAPSTONE

# Predicting Incident Severity based on outside and driver-related conditions

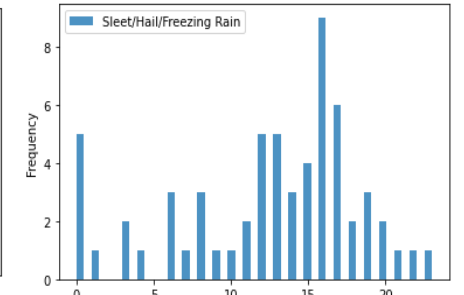
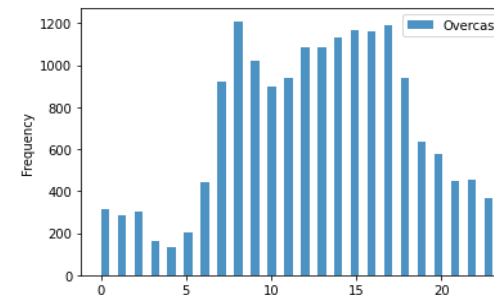
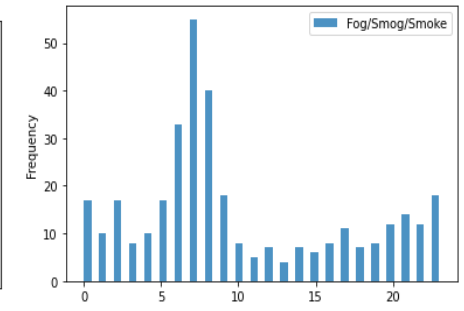
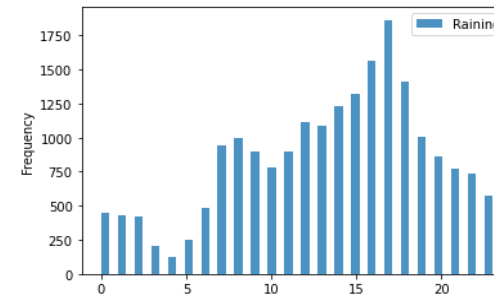
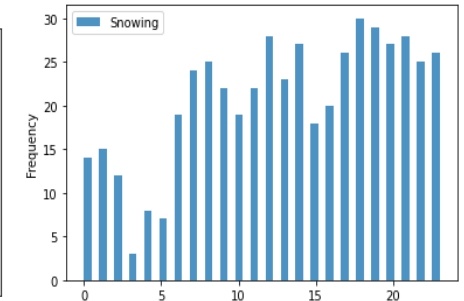
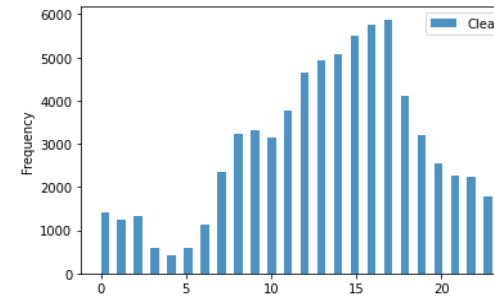
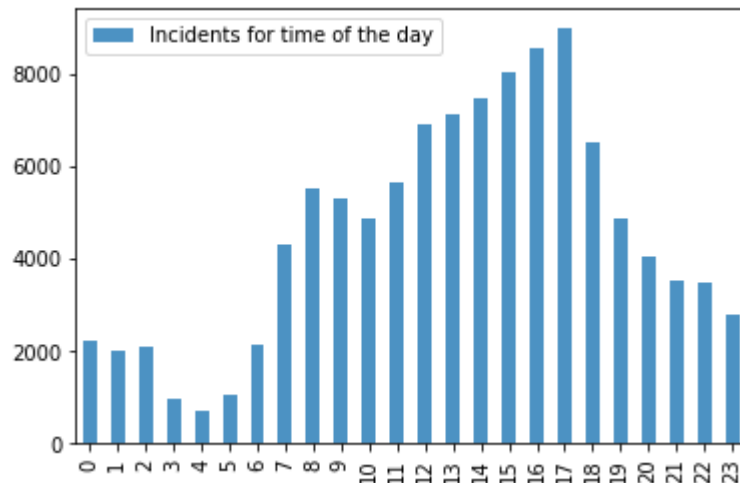
- ▶ In 2015, a crash occurred in Washington every 4.5 minutes, while in 2017 – every 4 minutes.
- ▶ To reduce number of incidents by changing some of road conditions or traffic management, this study could be interested for Seattle Road department or even for Washington State Department of Transportation (SDOT) to know that changing road or light conditions could be helpful or not.
- ▶ Seattle police department could use this data for improving safety transportation while outside conditions are not clear

# Data acquisition and cleaning

- ▶ Original dataset was provided by Washington State Department of Transportation (SDOT) and could be found on Kaggle [here](#). The limited dataset with reduced number of rows for only few severity types could be found on IBM shared folder [here](#). The both datasets contain details of car incidents from 2004 till present.
- ▶ The metadata for the dataset with explanatory details of each column could be also found on IBM shared folder [here](#).
- ▶ All cases not related to at least one active vehicle (Bicyclists, Trains, Pedestrians, ...) were removed.
- ▶ All features except outside and driver-related condition were removed.
- ▶ Total number of cleaned records: 127,273 rows and 8 columns

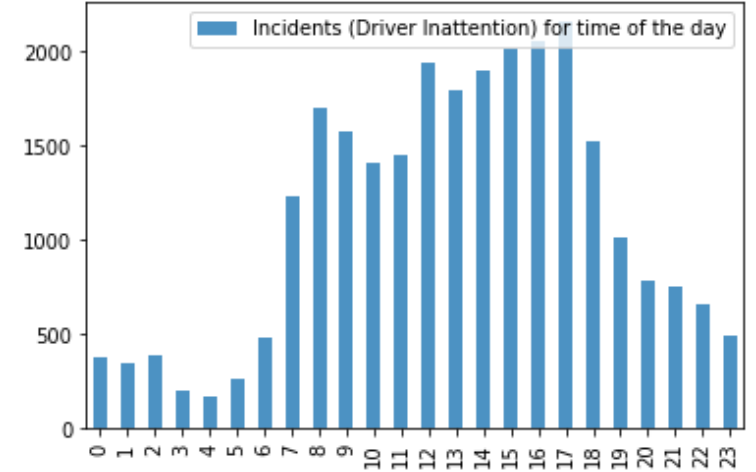
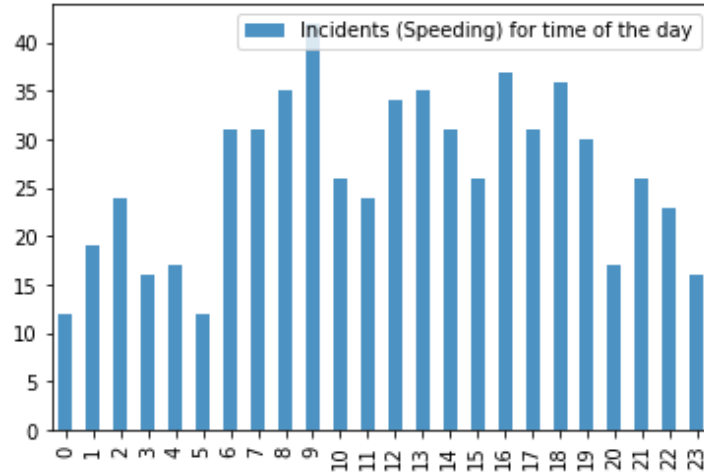
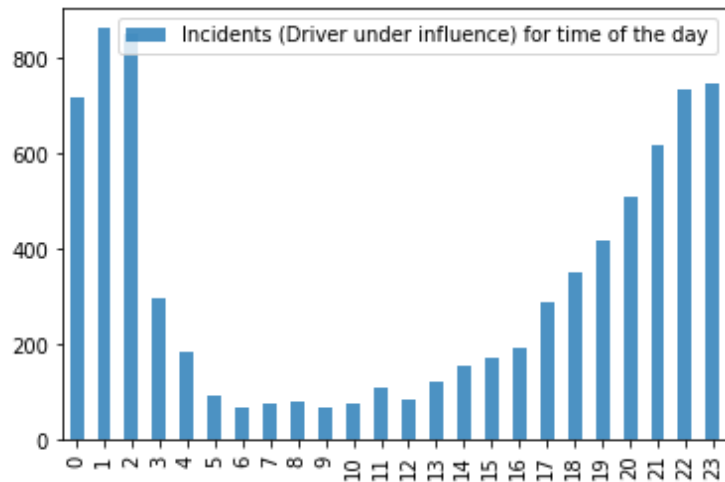
# Incidents based outside conditions

- ▶ While snowing, the incidents have similar probability during the day
- ▶ Overcast weather conditions have very similar to snow probability
- ▶ While Raining, most incidents are happened during evening time
- ▶ Fog / Smog / Smoke weather could be a reason of incidents for morning time
- ▶ Incidents in clear weather more likely happening on evening time (*may be when most of the people are commuting from work back to home*)



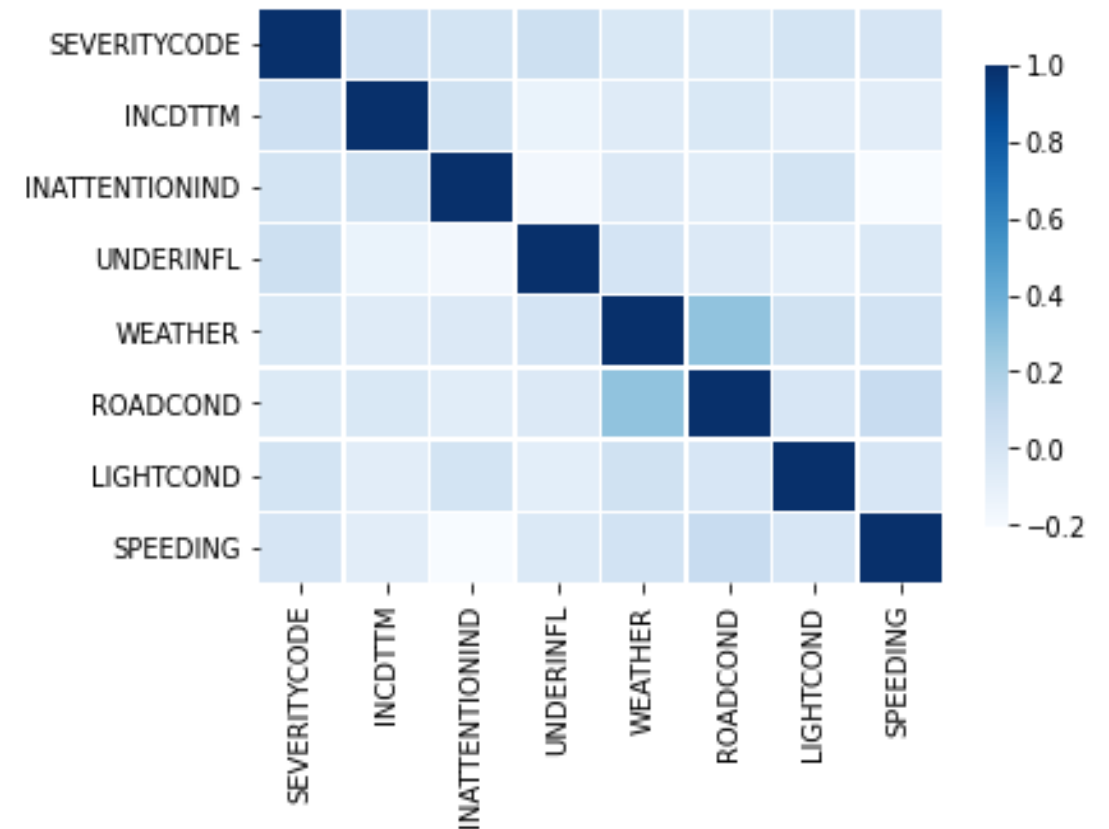
# Incidents based on driver-related conditions

- ▶ The most cases related to driver under influence were taken place at night time
- ▶ Speeding related cases are distributed during the whole day
- ▶ Driver inattention cases very close to standard view (*weather related*)



100

- ▶ Weather and road conditions have good correlation
- ▶ Severity somehow correlated with time driver under influence and speeding
- ▶ Speeding is not correlated with inattention
- ▶ Light condition somehow correlated with Time, weather and Road conditions



# Results - Predictive Modeling

- ▶ Generally, there are two types of predictive models – classification and regression. The difference is that while classification models could be helpful with prediction of output based on some inputs, they are not providing any additional information on the delta or amount of the changes of target feature.
- ▶ In our case, we have only binary output feature, so we could use only classification models.
- ▶ To understand which model would be most relevant for our study, we could use the following models with evaluation:
  - ▶ K-Nearest Neighbors (KNN)
  - ▶ Decision Tree
  - ▶ Logistic Regression (LR)
  - ▶ Support Vector Machine (SVM)

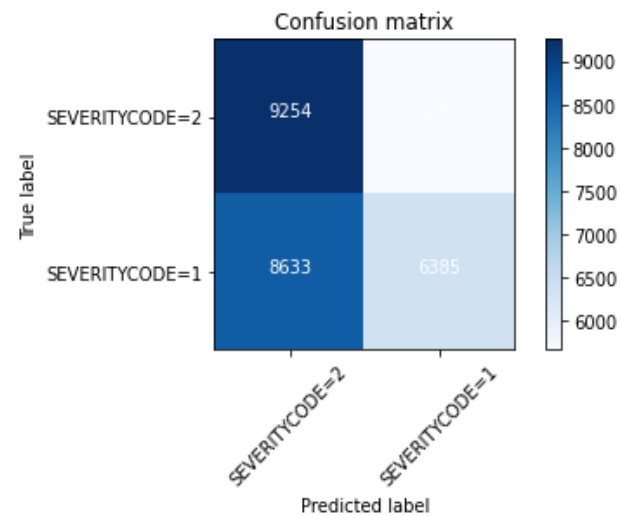
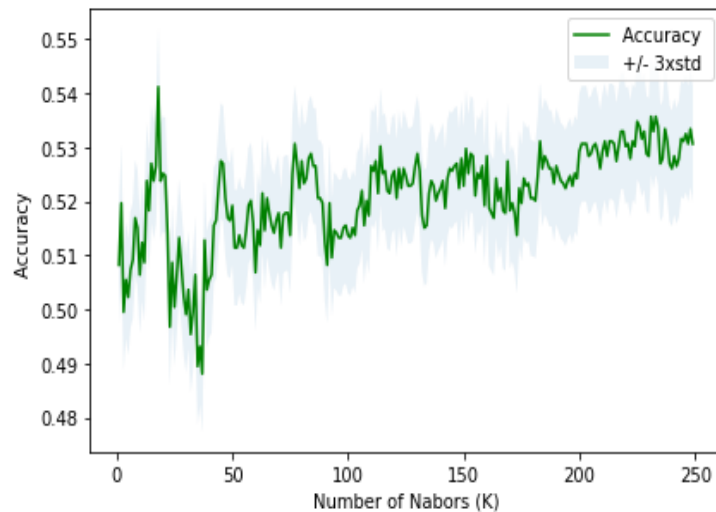
# Classification models – Best Parameters

MODEL	BEST PARAMETERS	ACCURACY
KNN	k = 21	0.534
DECISION TREE	criterion = 'entropy' max_depth = 50 max_features = 'auto' splitter = 'random'	0.532
LOGISTIC REGRESSION	C = 1 max_iter = 100 solver = 'lbfgs'	0.521
SVM	decision_function_shape = 'ovo' gamma = 'scale' kernel = 'linear'	0.527



# Classification Models - Performance

MODEL	F1 SCORE	ACCURACY	JACCARD SCORE	LOG LOSS
KNN	0.533	0.534	0.429	N/A
DECISION TREE	0.525	0.532	0.305	N/A
LOGISTIC REGRESSION	0.518	0.521	0.284	0.695
SVM	0.521	0.527	0.302	N/A



# Conclusion and future directions

- ▶ Reviewing the details of the results in that particular study, we have analyzed relationship between severity of incidents (property damage vs injuries) based on both – outside conditions as well as human-related (driver condition) conditions.
- ▶ We were able to find out that more injuries rather than property damage cases related to driver conditions, while outside conditions (weather / road / light) is more related to property damage (higher probability). Also, some cases were related to Dark conditions, while Street Lights were Off, and they possibly could be mitigated if Street Lights would be turned On. In addition to that, on snowing roads the most of the cases were happened during morning time, so may be using some salt in the nights/evenings before snowing could help to reduce such cases.
- ▶ Finally, we have created classification prediction models that based on outside conditions as well as driver conditions could predict (with some level of probability) the severity of incident, if it would happen.
- ▶ That information could help drivers, road workers as well as police to take some measures for reducing incidents overall and injuries in particular case.
- ▶ Place for improvement: The current study was based on SDOT dataset data, limited by only few severity codes as well as only outside and driver conditions were taken in consideration. Both of these types of conditions could be also related to area / district of the city or age of the driver or season of the year (i.e. winter vs summer) and models, based on additional features could have better result of prediction probability of the incident severity.