

# SEATTLE CAR ACCIDENT SEVERITY

---

CAPSTONE PROJECT



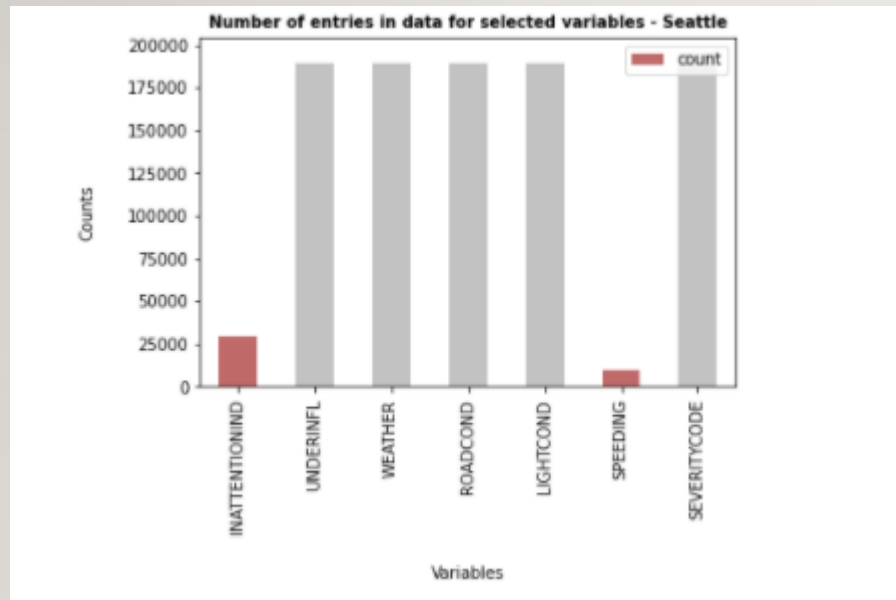
# BUSINESS PROBLEM

---

- The goal of this project is to predict severity of an accident and allow the Seattle government to possibly prevent/reduce car accidents that depend on factors like weather and road condition(wet/dry), traffic situation, light condition etc.
- About 100,000 crashes happen in Seattle every year since 2010 as per WSDOT ten-year summary report([WSDOT 10 Year Summary Report](#)).

# DATA UNDERSTANDING

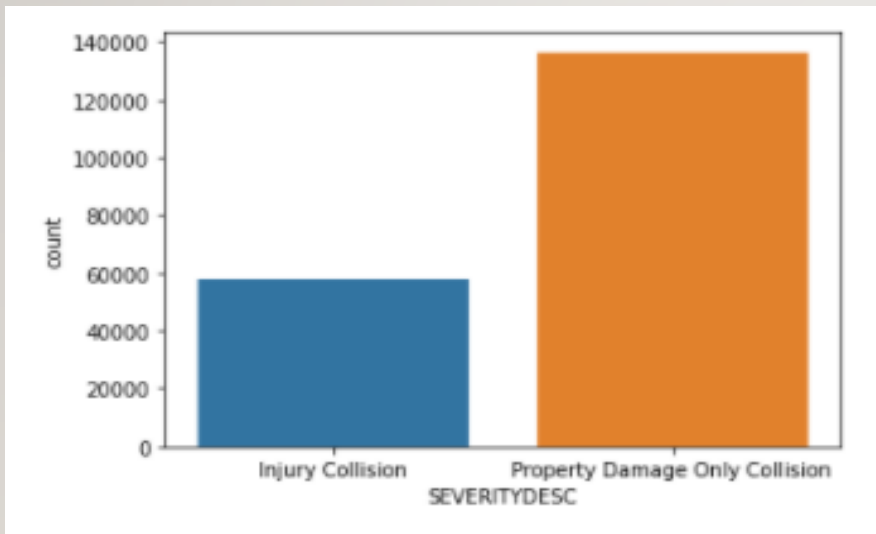
---



- There are 194,673 observations in the data set.
- In total, there are 37 attributes(columns) and 1 dependent variable (labelled data) which is “SEVERITYCODE” in the data and 194,673 rows.
- “SEVERITYCODE” is the code that corresponds to the severity of the collision and is considered as target variable in the analysis.

# COLLISION TYPE

---



- As seen in the figure, data is unbalanced as number of property damage incidents are in ratio of 2:1 with injury collisions
- Data was balanced using SMOTE in this project

# FEATURE SELECTION

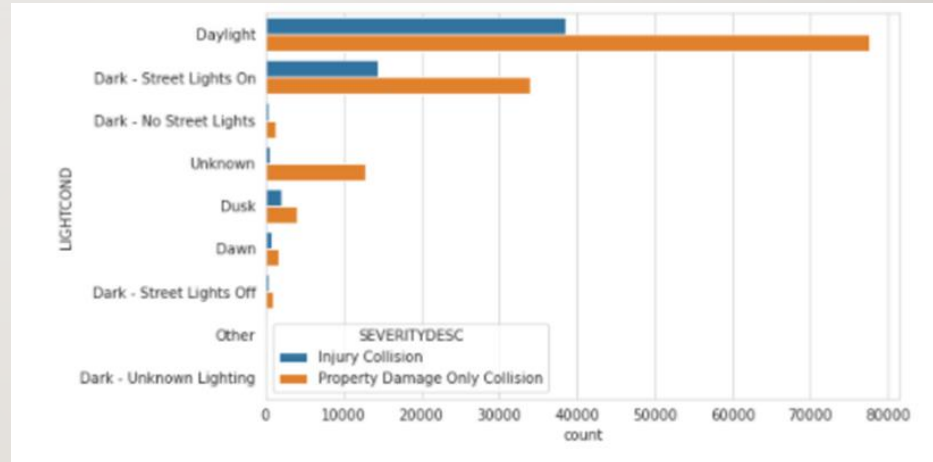
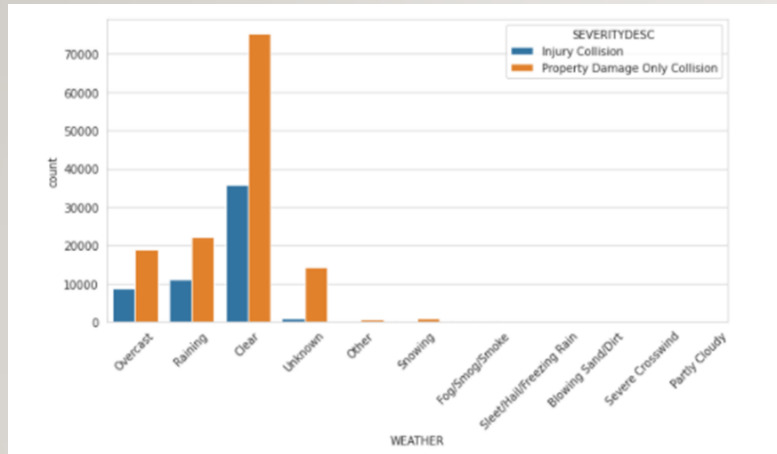
---

Features	Description
WEATHER	A description of the weather conditions during the time of the collision.
ROADCOND	The condition of the road during the collision.
LIGHTCOND	The light conditions during the collision.
SPEEDING	Whether or not speeding was a factor in the collision. (Y/N)
INATTENTIONIND	Whether or not collision was due to inattention. (Y/N)
UNDERINFL	Whether or not a driver involved was under the influence of drugs or alcohol.

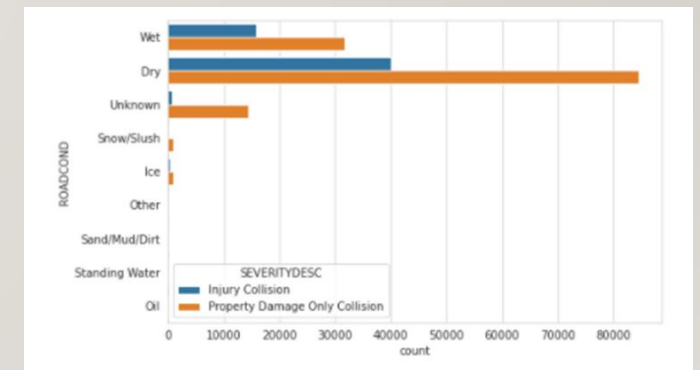
Selected features were fed into the predictive models as these are generally considered to have significant impact on the accident severity



# WEATHER,ROAD,LIGHT CONDITIONS-IMPACT

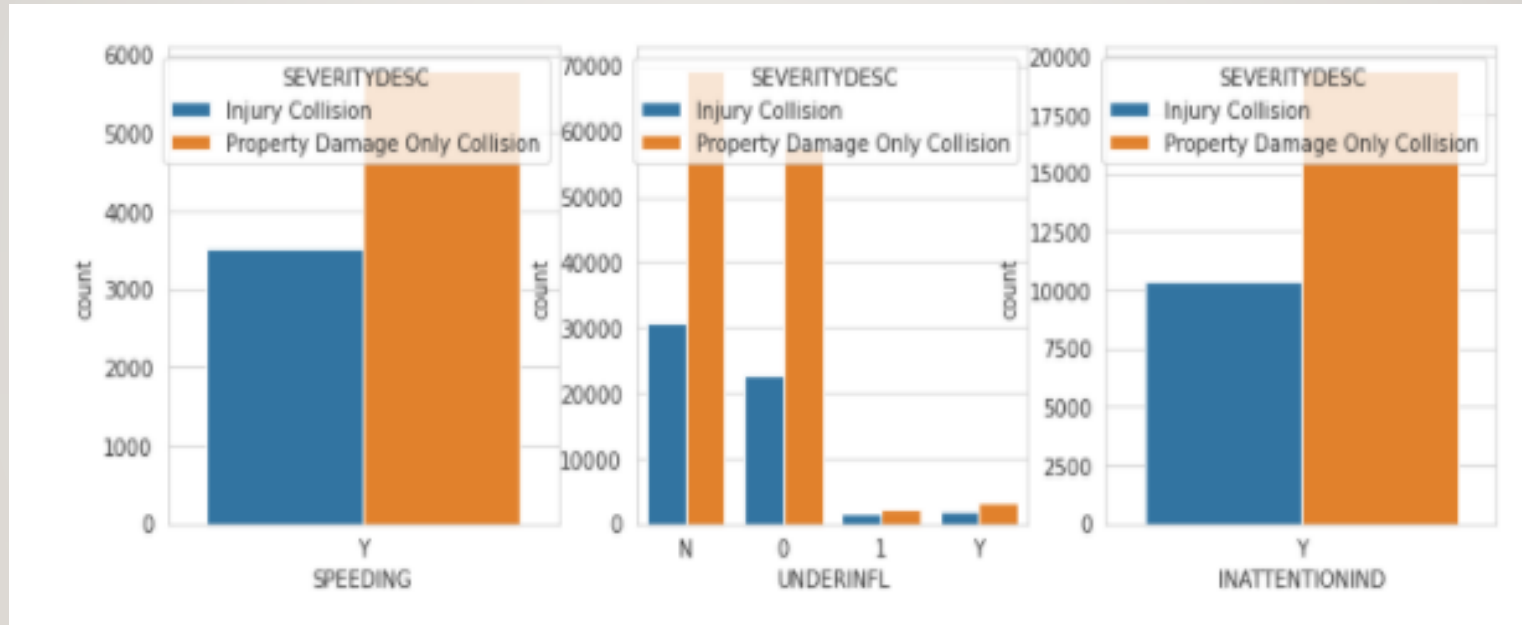


- It is seen that even clear weather has great impact on the collisions
- Many collisions occur un day light and street lights on.
- Both wet and dry road conditions can cause accident



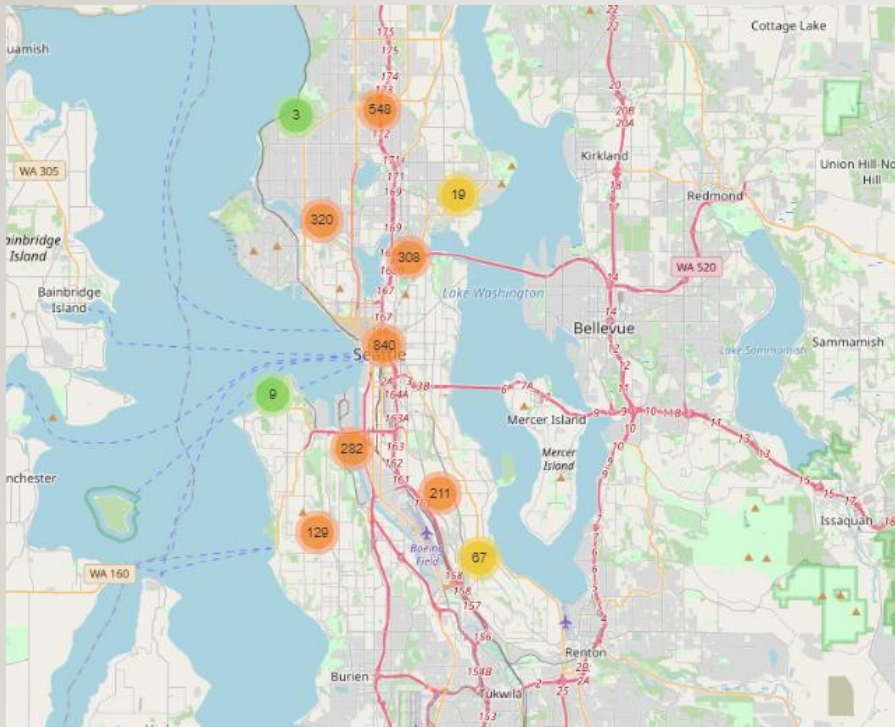
# OTHER ATTRIBUTES(SELECTED)-IMPACT

---



# GEOGRAPHICAL ANALYSIS

---



Majority of injury collisions happen around downtown Seattle where high traffic is expected in busy hours. Clearly drivers should pay extra attention near the accident hot spots.



# ML MODELS COMPARISON

Model	Avg f1-score	Avg Precision	Avg Recall
Decision Tree	0.58	0.58	0.57
LR	0.58	0.58	0.58
Knn	0.56	0.55	0.57

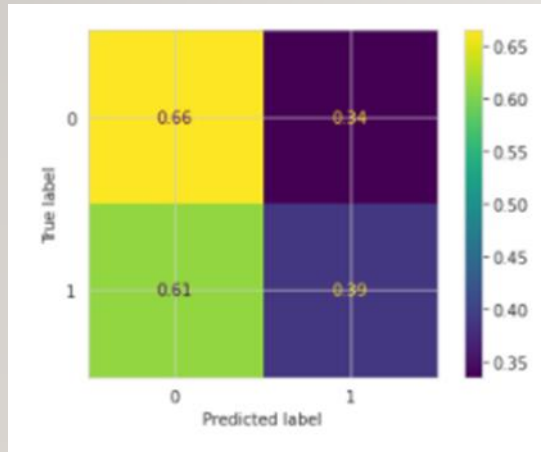
Machine learning models used behave in similar fashion and are biased towards “property damage collisions”.

DT	Precision	Recall	F1-score
0	0.69	0.66	0.68
1	0.36	0.39	0.38
Accuracy	0.57		
Macro avg	0.53	0.53	0.53
Weighted avg	0.58	0.57	0.58

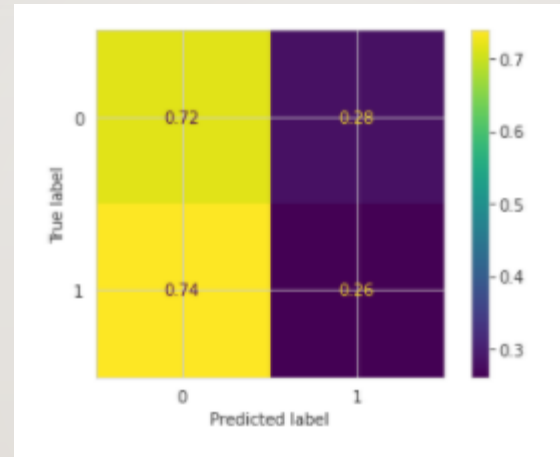
LR	Precision	Recall	F1-score
0	0.69	0.67	0.68
1	0.36	0.37	0.37
Accuracy	0.58		
Macro avg	0.52	0.52	0.52
Weighted avg	0.58	0.58	0.58
Log loss	0.69		

KNN	Precision	Recall	F1-score
0	0.67	0.72	0.69
1	0.31	0.26	0.28
Accuracy	0.57		
Macro avg	0.49	0.49	0.49
Weighted avg	0.55	0.57	0.56

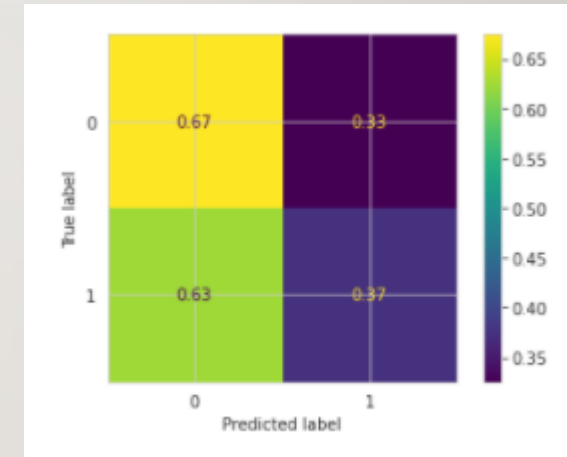
# ML MODELS PERFORMANCE(NORMALIZED)



DT Confusion  
Matrix



LR Confusion  
Matrix



KNN Confusion  
Matrix

# CONCLUSION

---

- Built models to predict severity of accident-injury collision vs property damage
- ML models have room for improvement to get better on accuracy
- Capture missing data like:
  - Weather condition
  - Road condition
  - Light condition