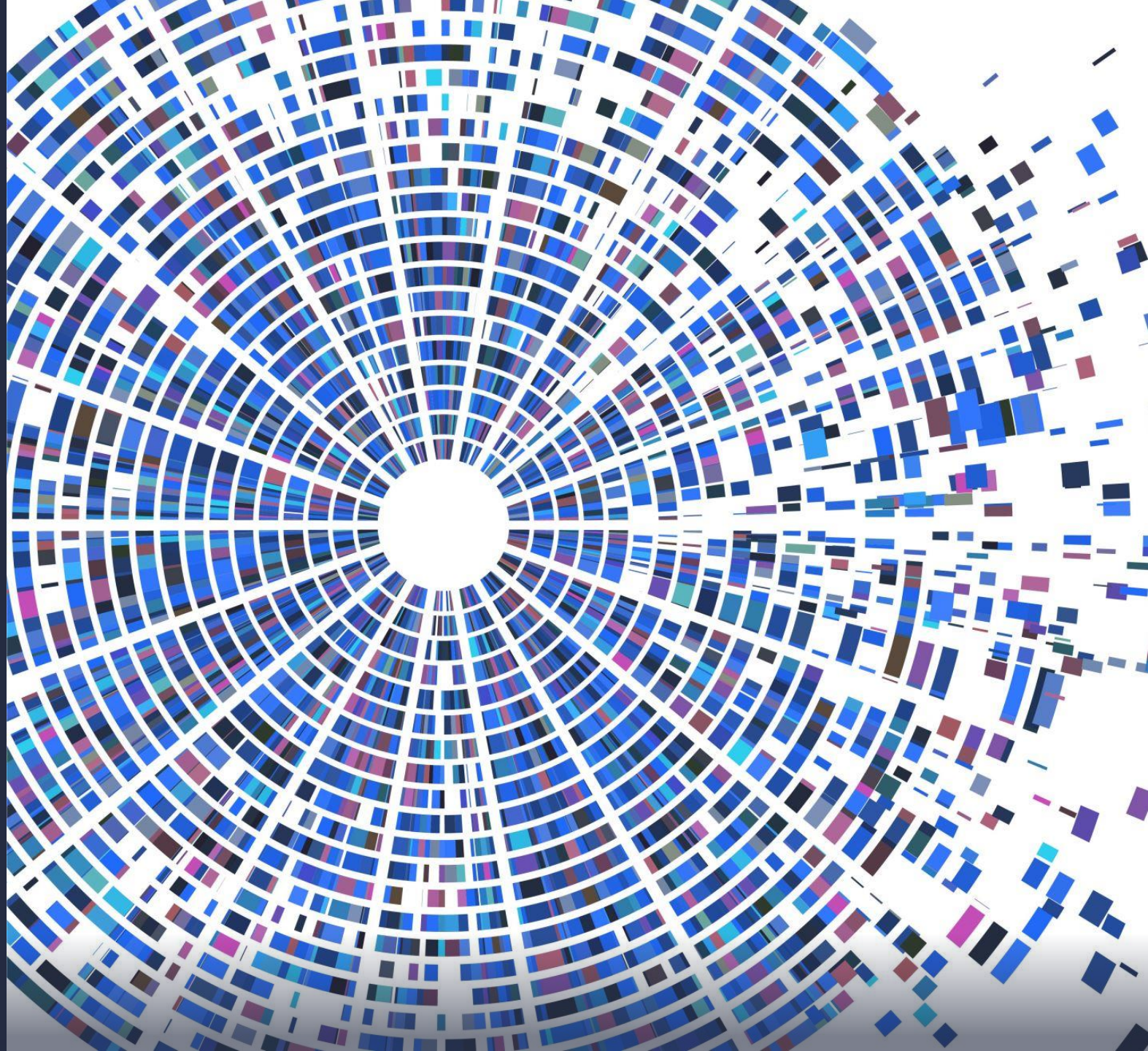


Predicting Car Accident Severity



Predicting car accidents severity is valuable for Seattle Drivers

- This report will target those driving around Seattle
- Therefore drivers will drive more carefully or even change travel if they able to.
- Will try to detect collisions that occur around the city
 - Possibility of getting into a car accident
 - Severity of the accidents
- Insurance companies have interest as well



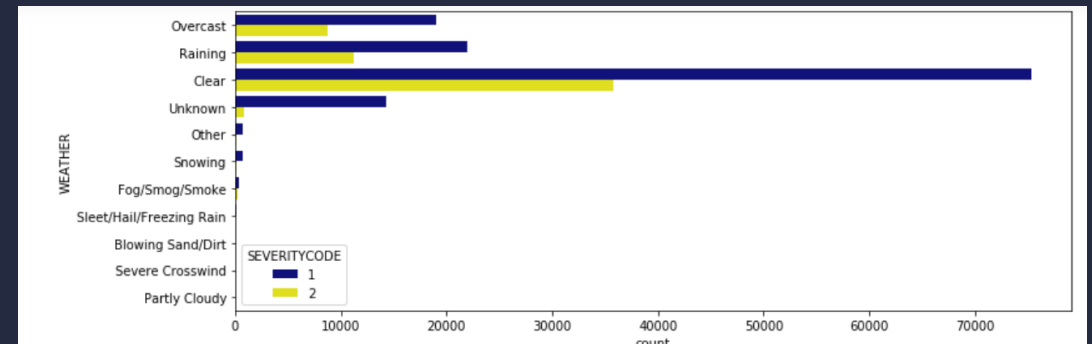
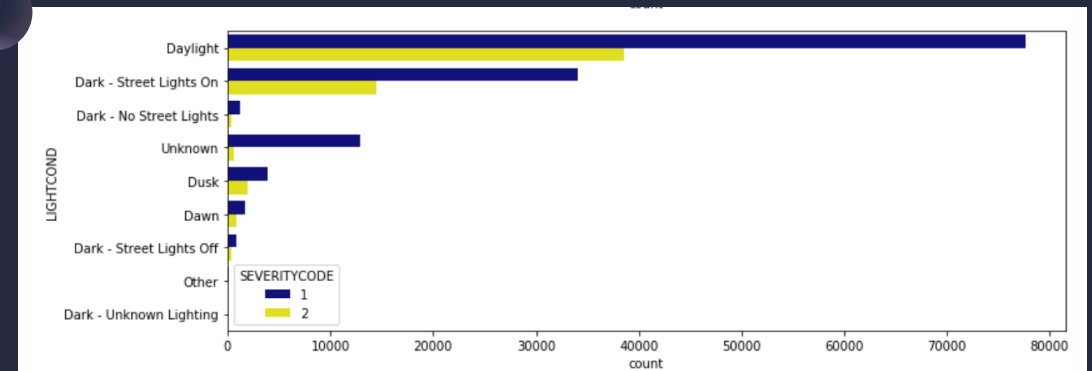
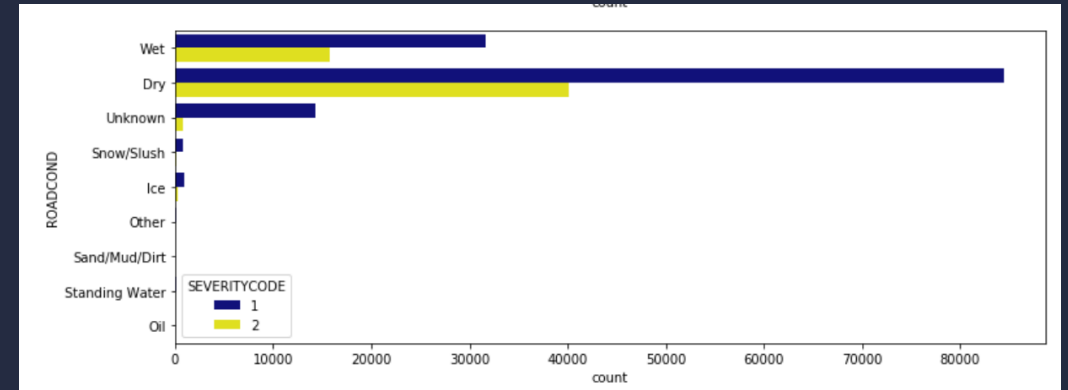
Data acquisition and cleaning

- Seattle Collision data downloaded from Coursera Applied Data Science Capstone
- The data is also found at provided by Seattle Police Department and recorded by Traffic Records <https://data-seattlecitygis.opendata.arcgis.com/>
- In total, 19,4673 rows and 38 features in the raw dataset
- Irrelevant features, rows containing irrelevant or missing data were dropped
- Cleaned data contains 4 features



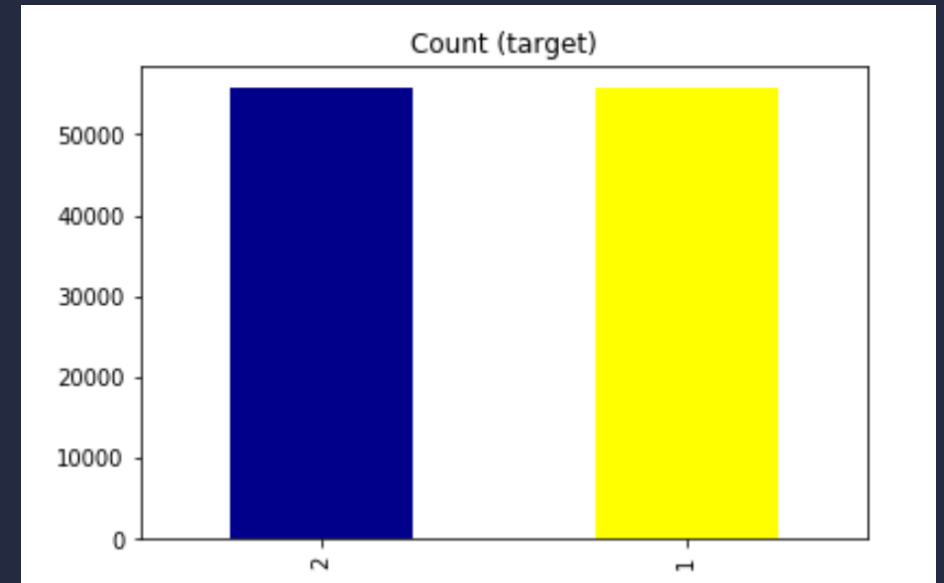
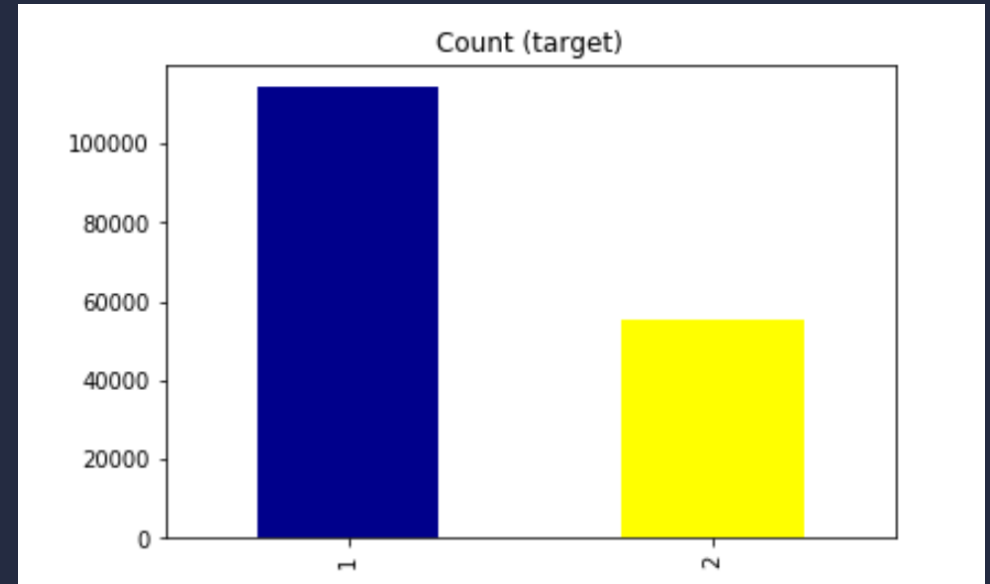
Using SEVERITYCODE as severity measure

- Most collisions occur during the clear weather, daylight and dry roads
- Most of the collisions are classified as property damages



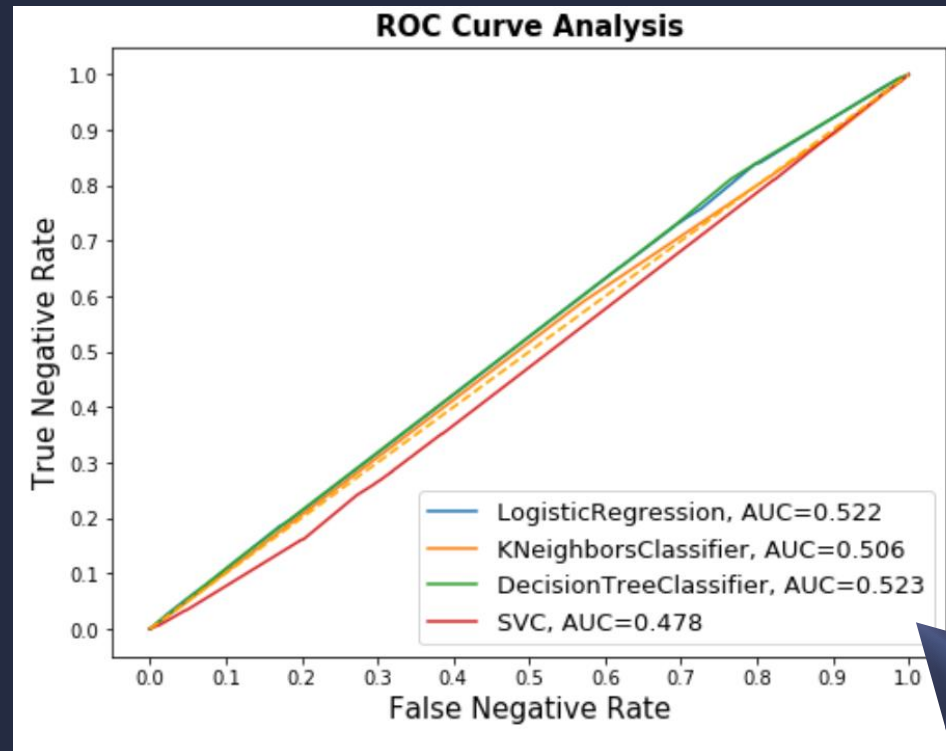
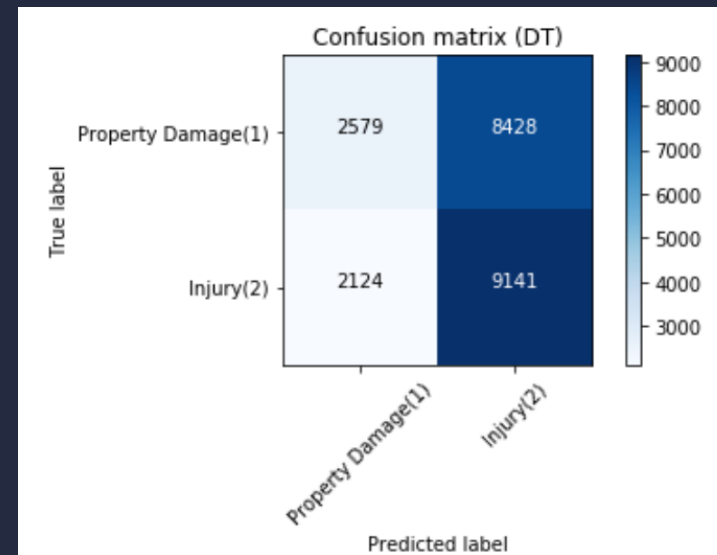
Dealing with unbalanced dataset

- The data was unbalanced by 43%, before cleaning and 49% after cleaning
- Down sampling was used to balance the data



Classification models performance

- F1-score:
 - 0.4830-0.5075 between 4 models
- Jaccard:
 - 0.5106-0.5262 between 4 models
- Decision Tree performed best among single algorithms



Conclusion and future directions

- Built useful models to predict whether a driver will get into an accident and the severity of that accident.
- Accuracy of the models has room for improvement.
- Ideas include:
 - JUNCTIONTYPE
 - ADDTYPE
 - Displaying the location of the accidents on map (longitude, latitude)
 - A description of the collision (whether it was a vehicle to vehicle or vehicle to bike accident)