Predicting Severity of Seattle Collisions

Miguel Enrique Játiva Jiménez

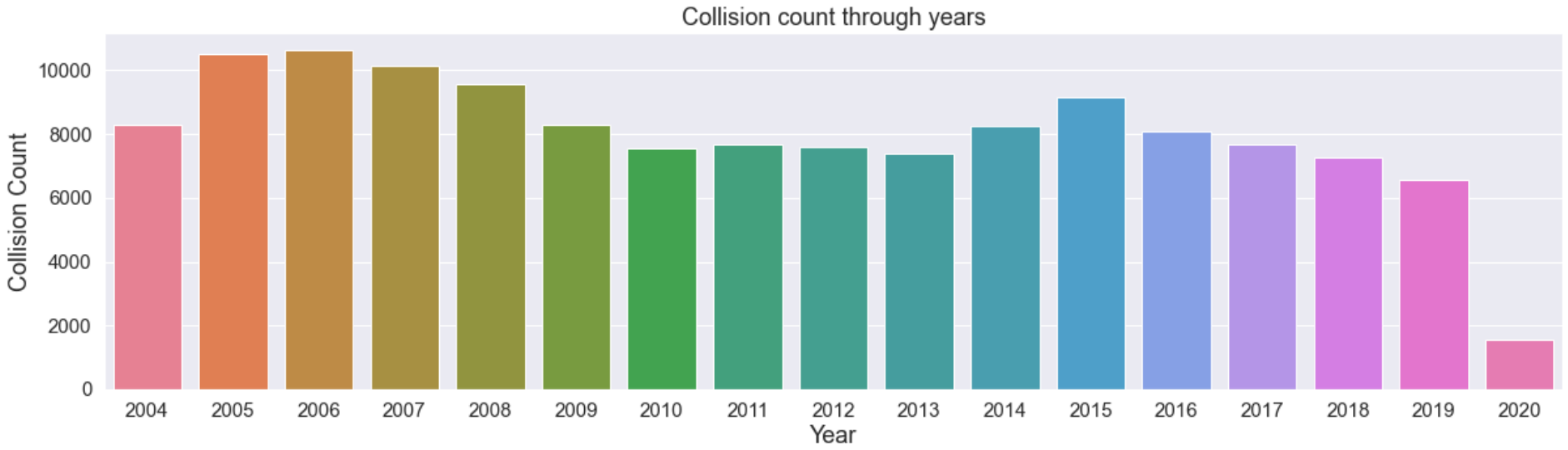
Predicting severity on collisions is important

- It would be useful to predict the severity of the collision given the context, this could be of interest to the **Health Care System** since it would allow it to assign the right amount of resources to each collision depending on the severity predicted.

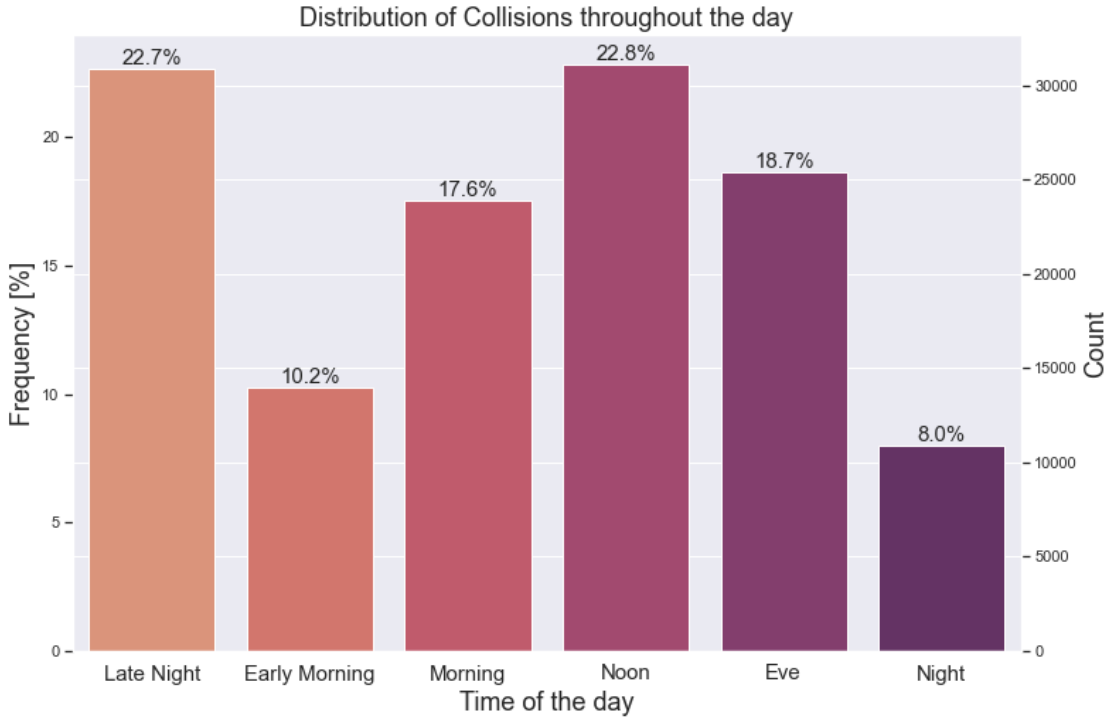
Data acquisition and cleaning

- The data is obtained from the Seattle Government more specifically from SDOT Traffic Management Division: https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv
- In total, 194673 rows and 38 features in the raw dataset.
- Duplicate, id or features with most values missing were dropped.
- Cleaned data contains 18 features.

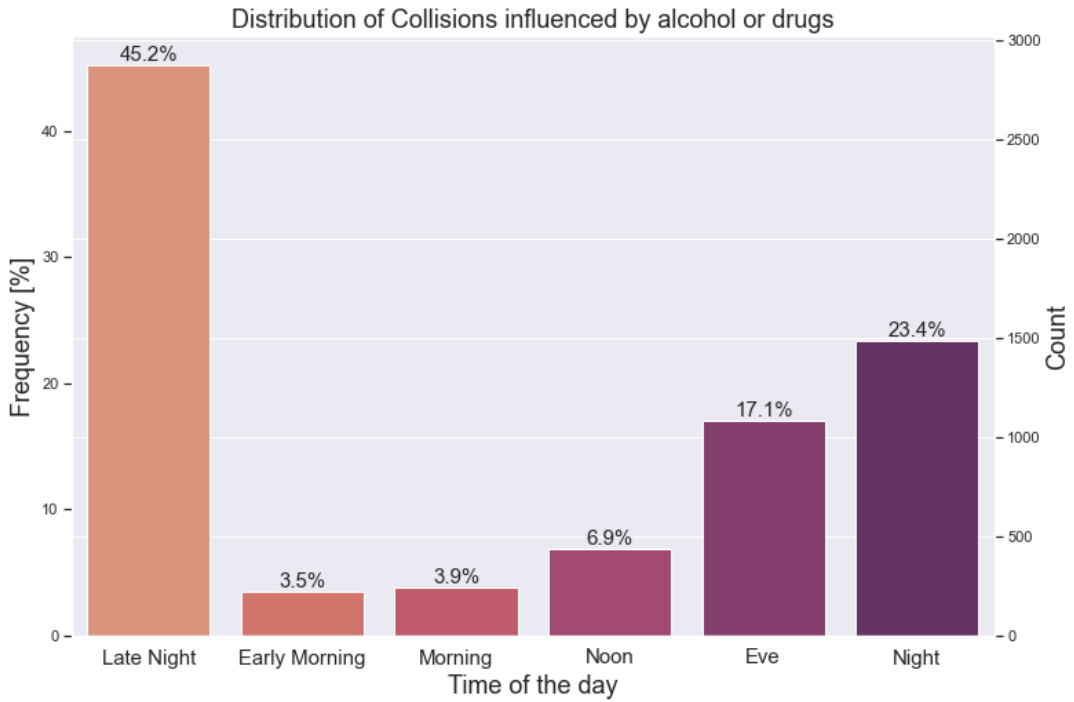
The amount of collisions is decreasing



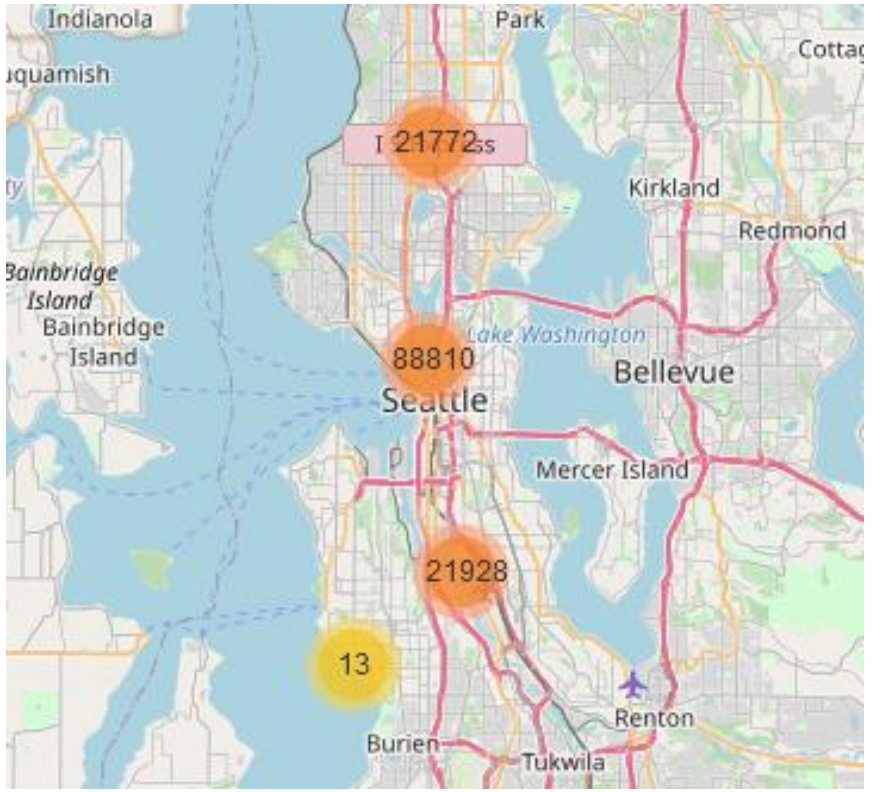
Most collisions happen at Late Nigh and Noon



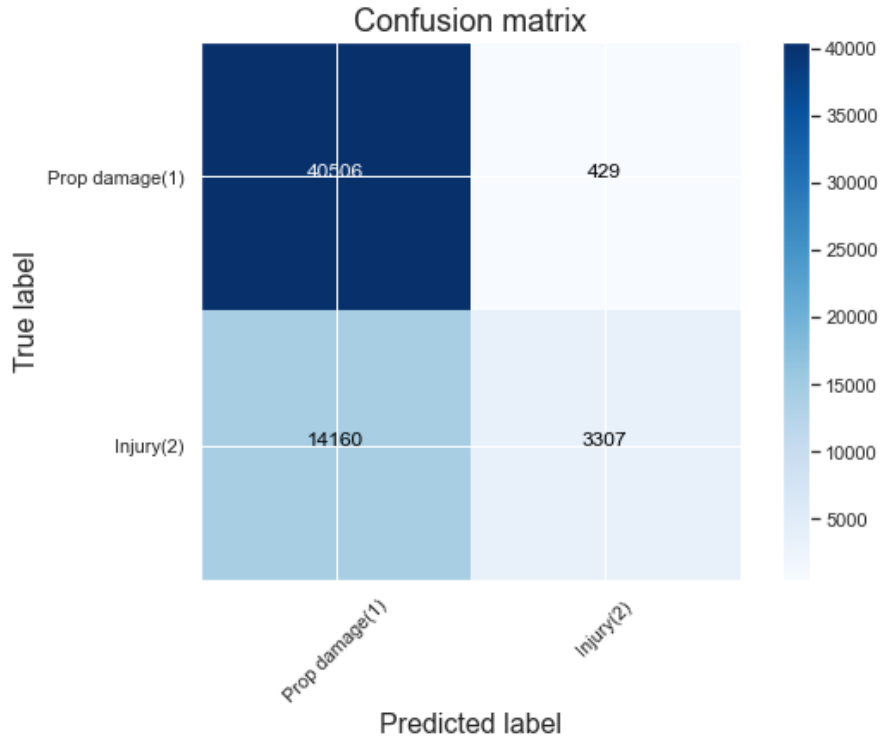
People influenced by alcohol/drugs have more collisions at Late Night



Most collisions happen in the center of Seattle

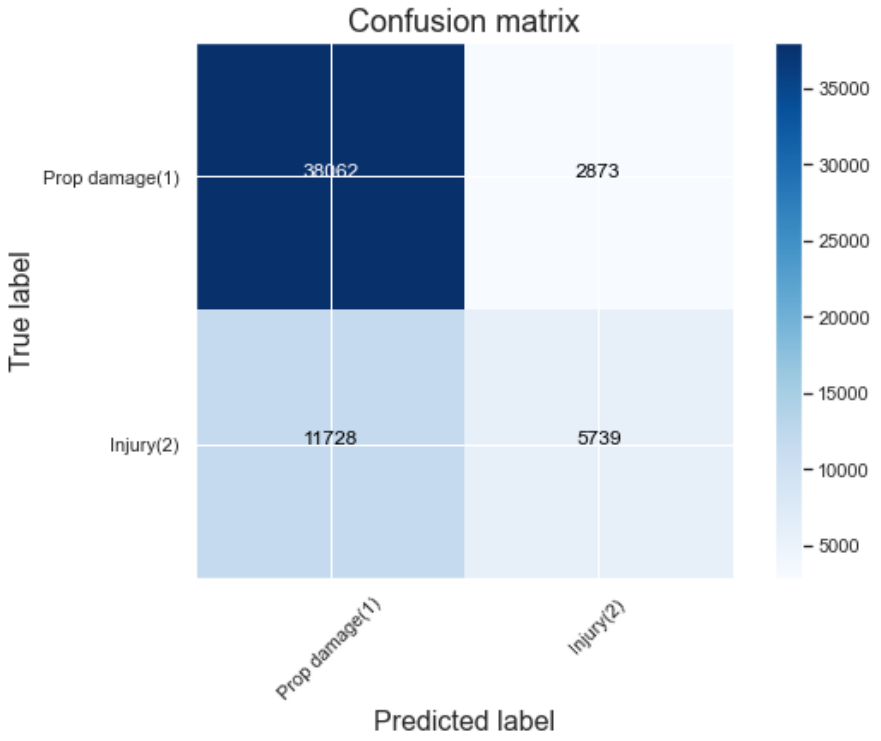


Decision Tree model performance



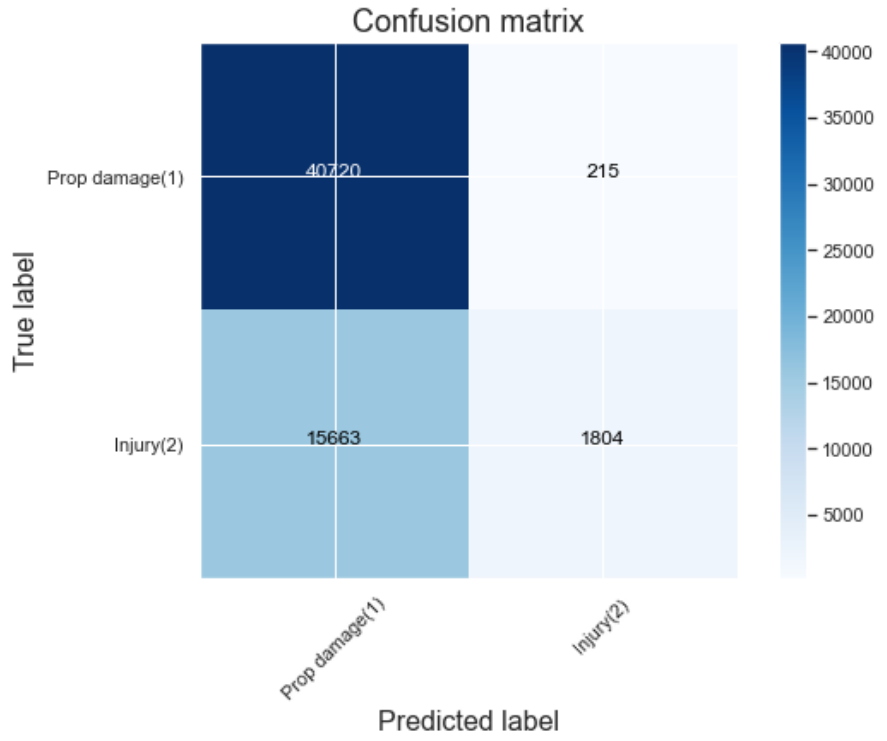
- Accuracy: 0.750197
- Jaccard: 0.735203
- F1-score: 0.687250

K Nearest Neighbors model performance



- Accuracy: 0.749991
- Jaccard: 0.722747
- F1-score: 0.719748

Linear Regression model performance



- Accuracy: 0.728126
- Jaccard: 0.719460
- F1-score: 0.641937
- LogLoss: 0.672446

Conclusion and future directions

- Built models to predict the severity of collisions.
- Accuracy of the models has room for improvement.
- Include more data sources to have a more balanced data set.