# Predicting Severity of Car Accidents

Problem

- 280 million cars in the United States
- 12 million car accidents in 2018, causing 36,560 deaths
- Accidents occur for several factors, such as reckless driving
- Predicting the severity of accidents may help to:
  - Warn drivers
  - Help first responders

Data

- 194,673 accident records collected from Seattle Department of Transportation between 2004 until 2020

- The target label is the accident severity code

- Data contains two codes:

  - Code 1: accident with property damage

  - Code 2: accident with injuries
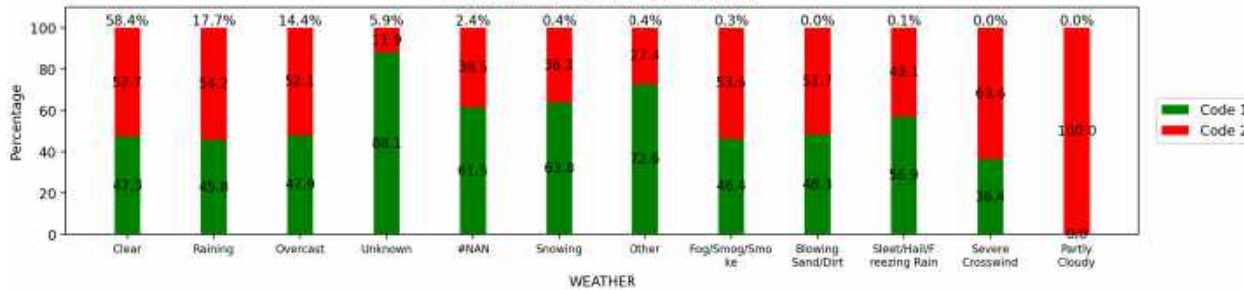

Accidents Grouped by Severity Code
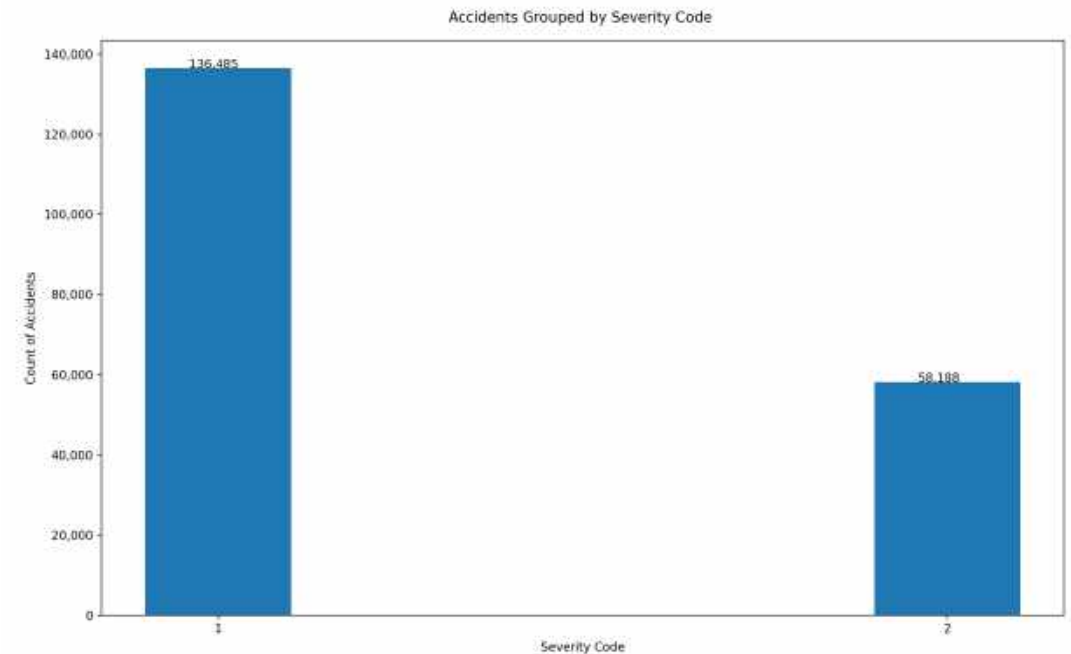
Exploratory Analysis

▶ Analyze on the following features:

Feature
Weather
Accident Day
Accident Month
Accident Hour
Road Condition
Light Condition
Address Type
Collision Type
DUI
Vehicle Count
Persons Count
Pedestrian Count
Bicycle Count

Exploratory Analysis
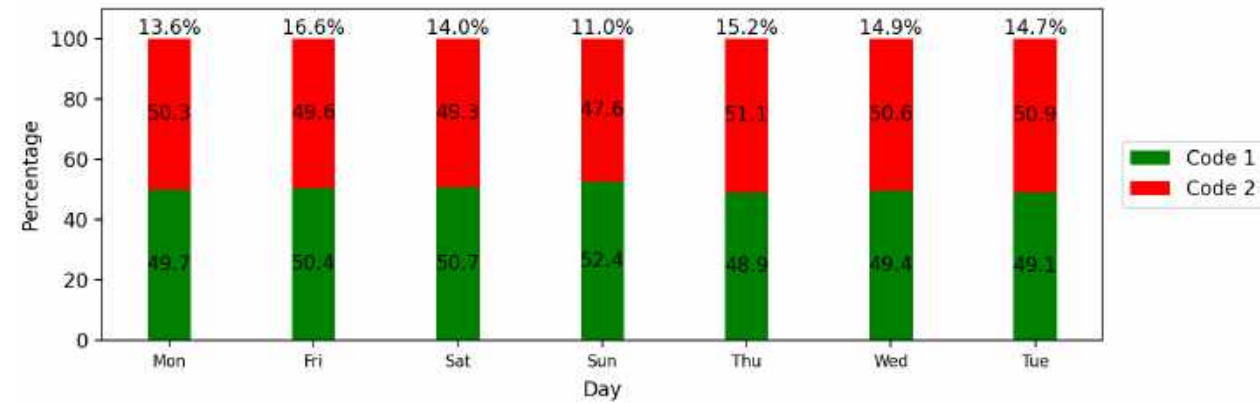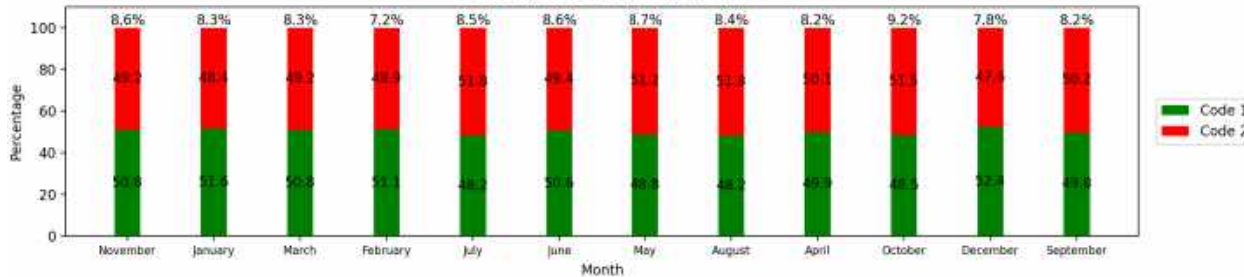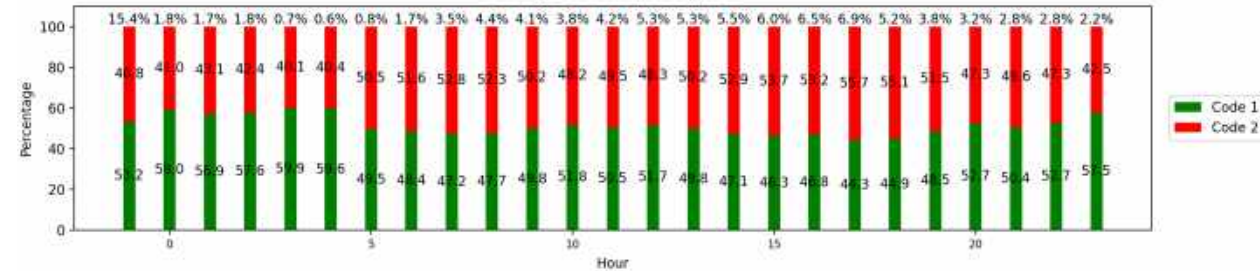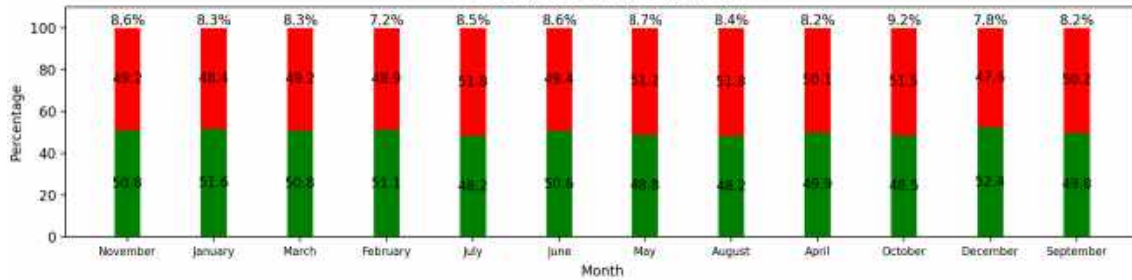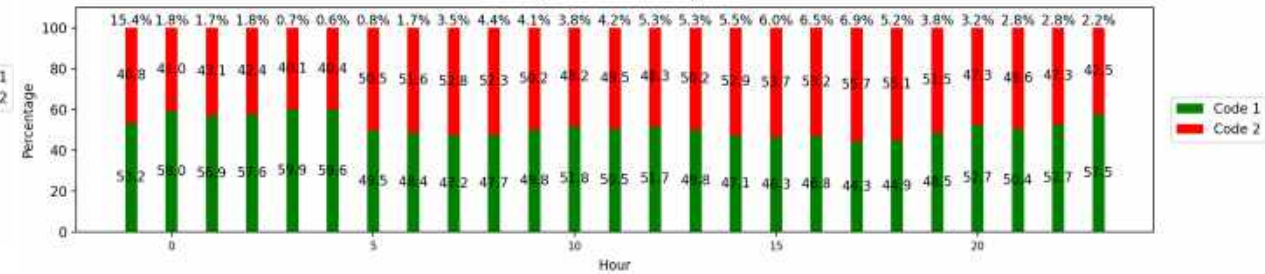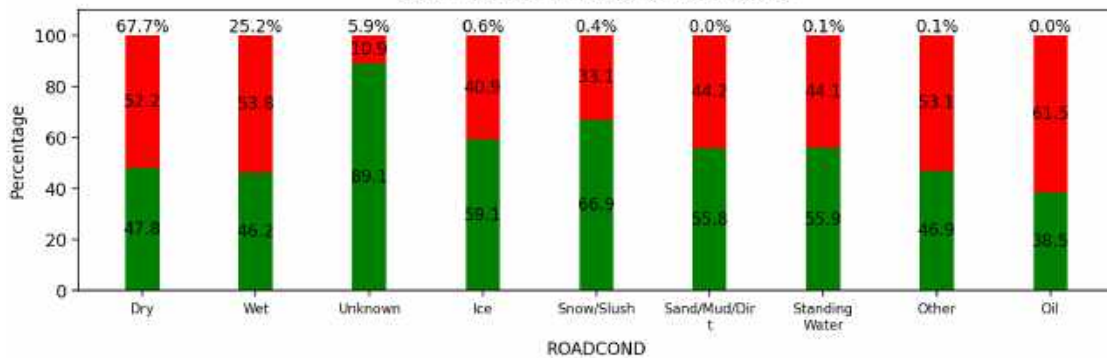
Exploratory Analysis
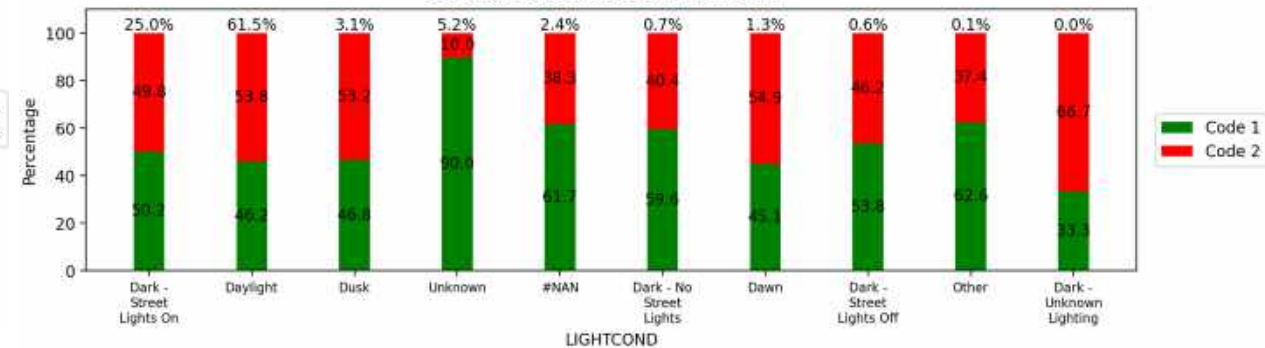
Exploratory Analysis
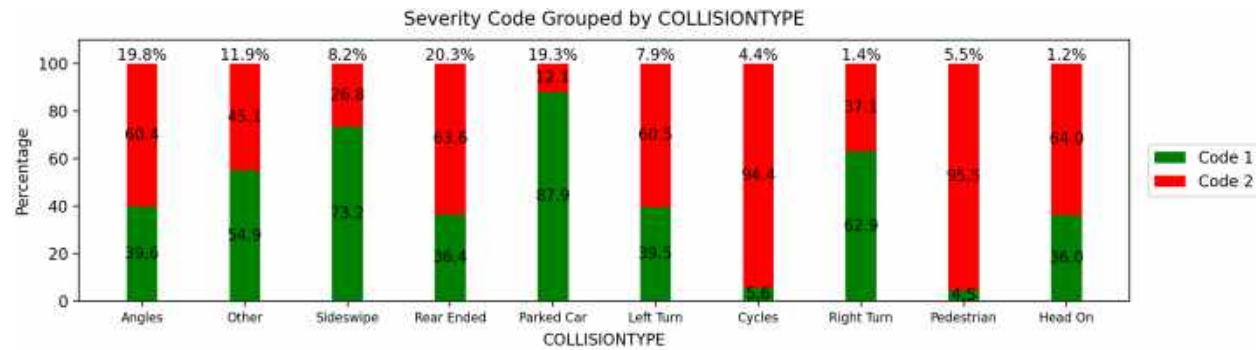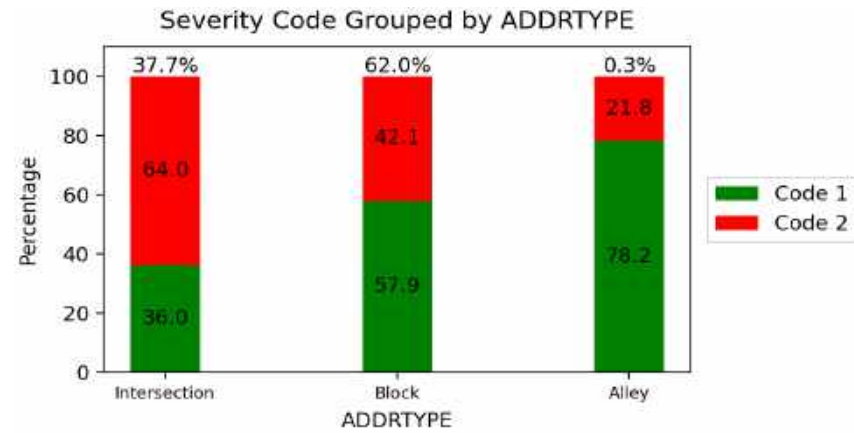
Exploratory Analysis
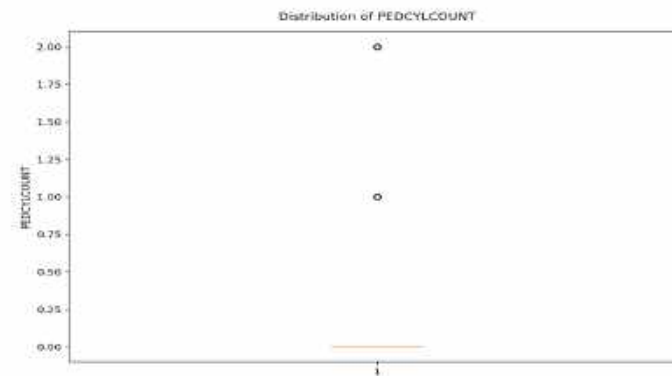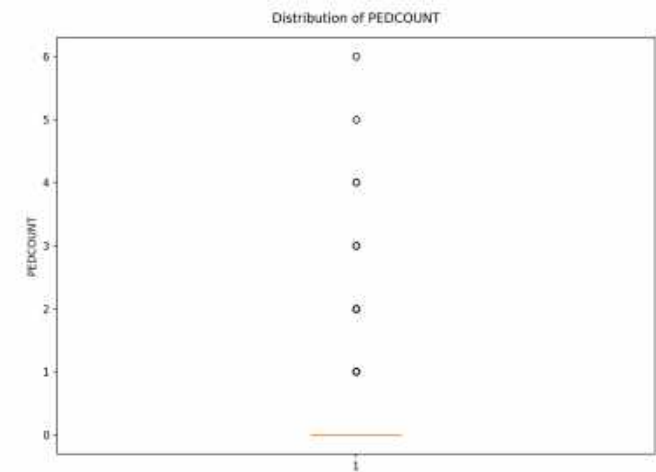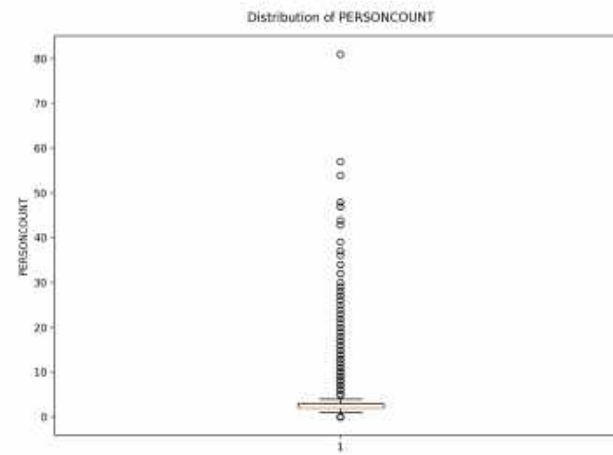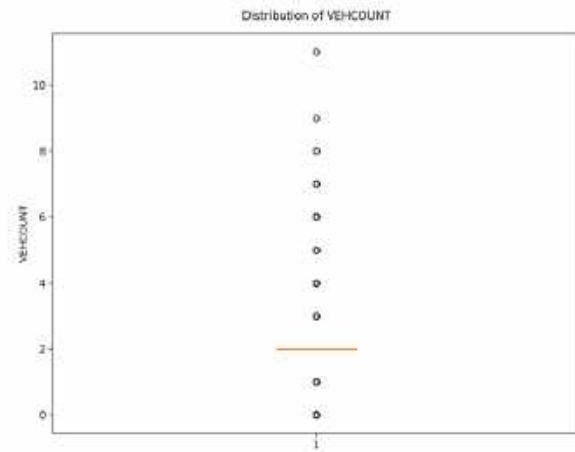
Correlations

▶ Numeric Features

```
                INCKEY  PERSONCOUNT  PEDCOUNT  PEDCYLCOUNT  VEHCOUNT  \
INCKEY         1.000000    -0.066338  0.034381     0.043216 -0.027777
PERSONCOUNT   -0.066338     1.000000 -0.053523    -0.060960  0.430674
PEDCOUNT       0.034381    -0.053523  1.000000    -0.049164 -0.394505
PEDCYLCOUNT    0.043216    -0.060960 -0.049164     1.000000 -0.357746
VEHCOUNT      -0.027777     0.430674 -0.394505    -0.357746  1.000000
SEVERITYCODE   0.030872     0.150489  0.220200     0.187279 -0.096880
```

▶ Categorical Feature

Feature	Cramer's Value
Weather	0.19
Accident Day	0.02
Accident Month	0.03
Road Condition	0.21
Light Condition	0.2
Address Type	0.22
Collision Type	0.49

Feature Selection

▶ Preprocessing of features was performed to remove outliers and handle missing data

Feature
Road Condition
Light Condition
Address Type
Collision Type
Vehicle Count
Persons Count
Pedestrian Count
Bicycle Count

Evaluation

- Random Forest, supervised learning for classification and regression
- Data is split into 90% training and 10% testing
- Training was performed using 10-cross fold validation
- Report accuracy and F1-score

Results

- F1 score for training is 68%
- Testing result:
  - F1 score 67%
  - Accuracy is 70%

Confusion Matrix of the Classifier
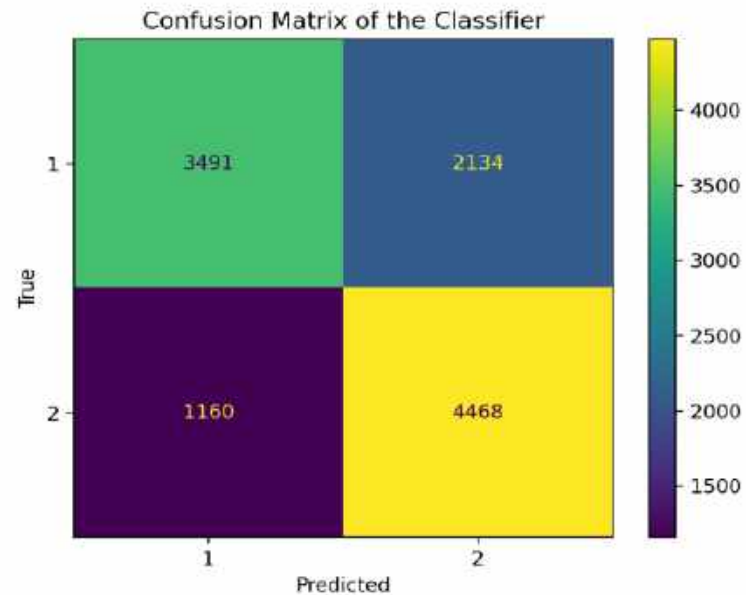
Discussion

- Model achieved low predictability of accident severity
- Features have weak correlation with target variable
- Adding more related features may improve the results

Summary

▶ Predicting the severity of accidents help first responders and to warn drivers

▶ Random Forest model achieved 67% in F1 score and 70% accuracy

▶ Collecting more related features is needed to improve results