

A blue ribbon graphic with a 3D effect, featuring a lighter blue top surface and a darker blue bottom surface, framing the text.

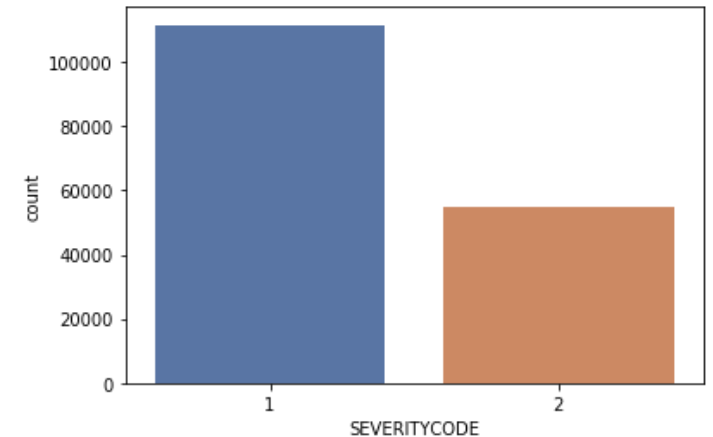
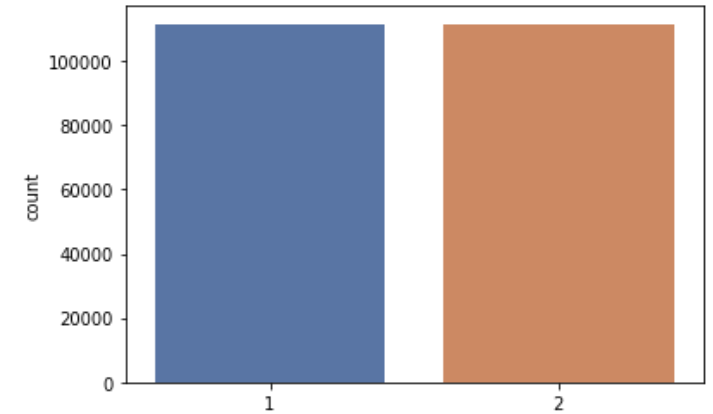
# Applied Data Science Capstone Project

# Business Understanding:

- Imagine the ability to anticipate whether an auto accident is likely and its severity that would allow the ability to save lives. The objective of the project is to predict the likelihood and severity of an accident so that the driver can adjust driving behavior and/or journey plan.

# Data understanding:

- The data is both categorical and numerical.
- Data used: accident location, event description, weather, and road conditions.
- Data is missing for several attributes and unbalanced. Oversampling and under sampling compared.



# Modeling:

- The data was then encoded using sklearn's preprocessing LabelEncoder as some of the attributes were categorical.
- The oversampled/undersampled data was then split into training and testing sets
- K Nearest Neighbors, Decision Tree, Logistic Regression, Naïve Bayes, and Random Forest algorithms. For the KNN model, the optimal K was determined to be K=6.



# Evaluation:

- Comparison of the ML models

	Jaccard	F1-score	Recall	Precision
K Neighbors	0.689362	0.694387	0.724108	0.711193
DecisionTree	0.706268	0.659339	0.72509	0.768958
LogisticRegression	0.704081	0.673087	0.727056	0.744422
Naive Bayes	0.69427	0.663835	0.717369	0.72101
Random Forest	0.692533	0.705986	0.730525	0.718708

