**Car Accidents Severity:**

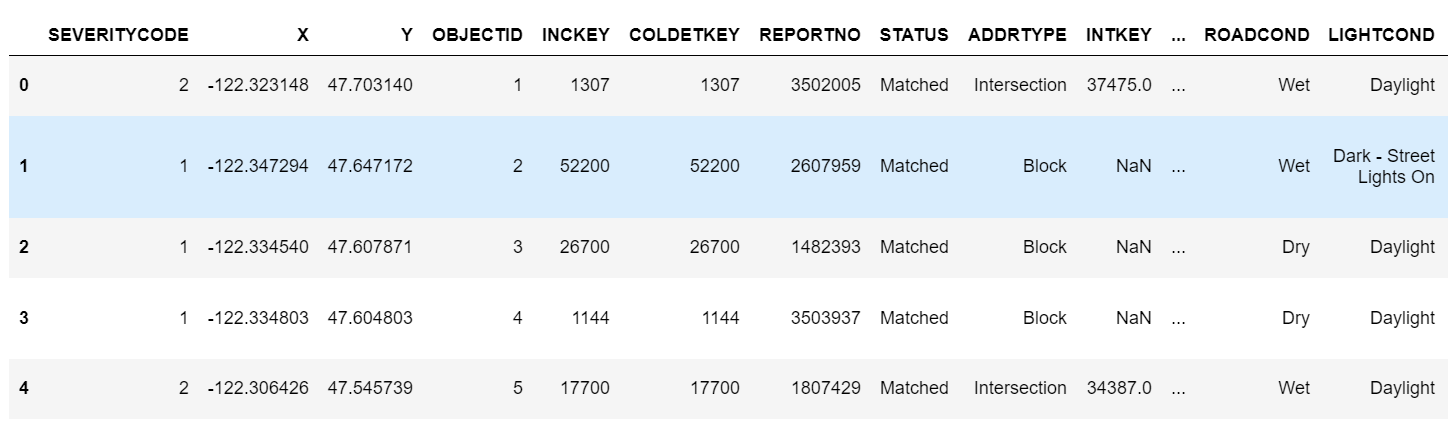
**Introduction:**

**The Car Accidents Severity data set is provided as part of Capstone project. It contains totally 37 attributes and 194673 instances. The business problem chosen is predicting the accident severity using vehicle count, person count and collision description as main features. Severity, class label for the data set is classified into Injury collision and Property Damage Only Collision.**

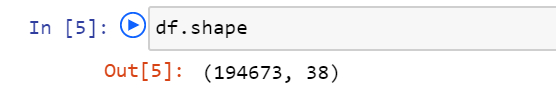
**Business Problem:**

**Are we able to predict accident severity using vehicle count, person count and collision description?**

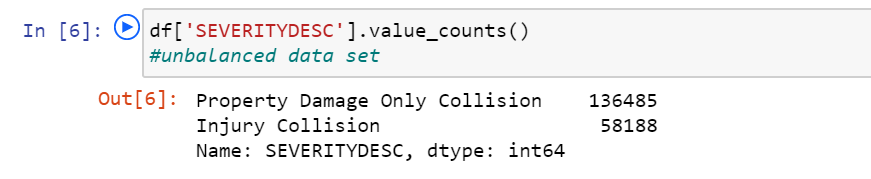
**Data Set: Car Accident Severity**

****

**Data Set Size:**

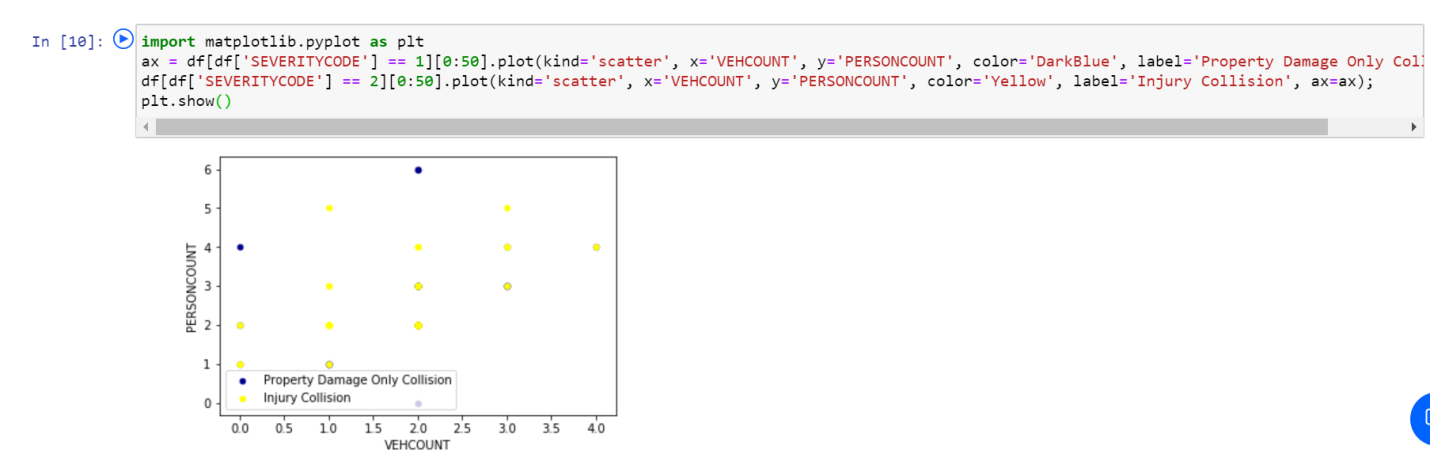
****

**Data Set Type = Unbalanced labeled data set**

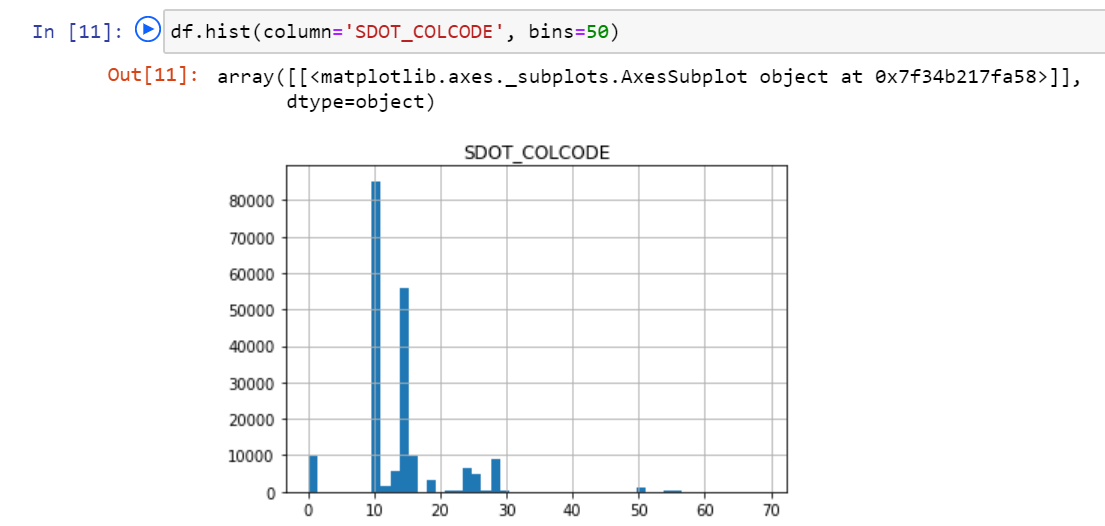
****

**Initial Data Visualization:**

**Plotting scatter plot to find the relation between attributes with accident severity**

****

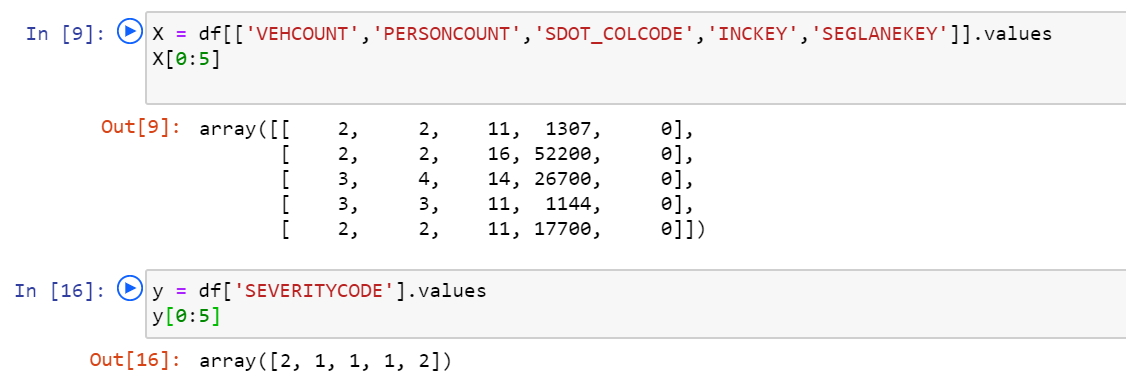
**Plotting histogram for collision description and found “MOTOR VEHICLE STRUCK MOTOR VEHICLE, REAR END” causes more accidents.**

****

**Model Selection:K-nearest neighbors algorithm:**

**We chose K-nearest neighbors algorithm as it is optimized model and easy to find the best accuracy for our dataset.**

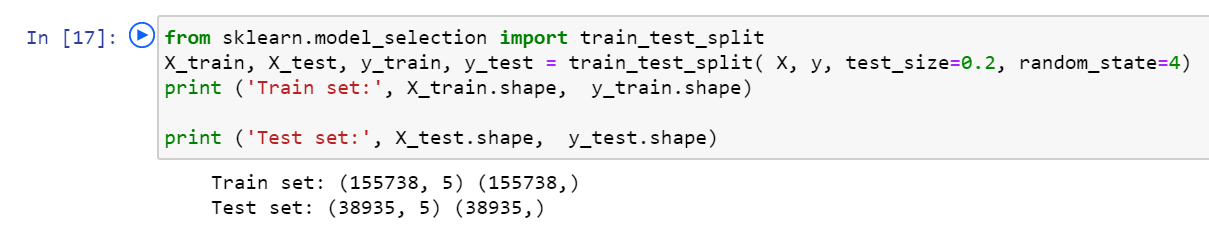
**Feature Attributes and Target Variable:**

****

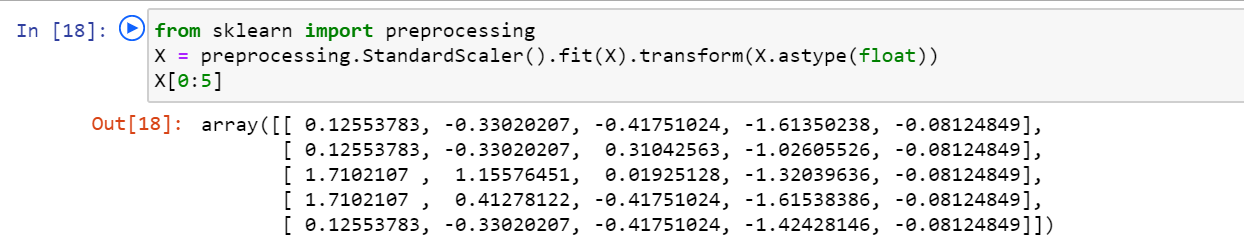
**Here,**

|  |  |
| --- | --- |
| **SEVERITYCODE** | **SEVERITYDESC** |
| 1 | Property Damage Only Collision |
| 2 | Injury Collision |

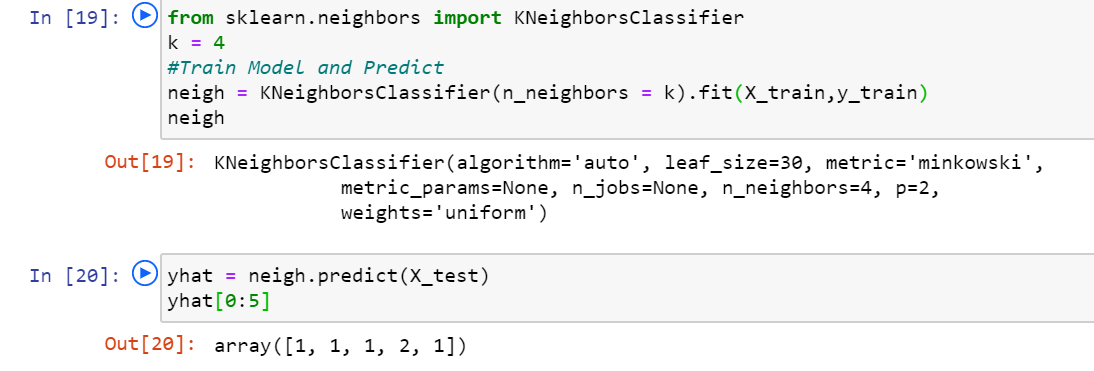
**Splitting Data Set for training and testing**

****

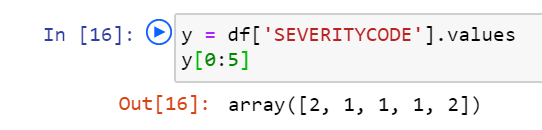
**Building the model:**

****

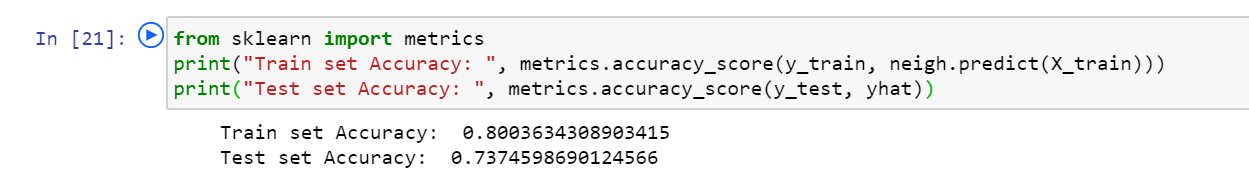
**Initially K set to 4:**

****

**After built the model with k=4, only two accidents are predicted correctly for 5 instances.**

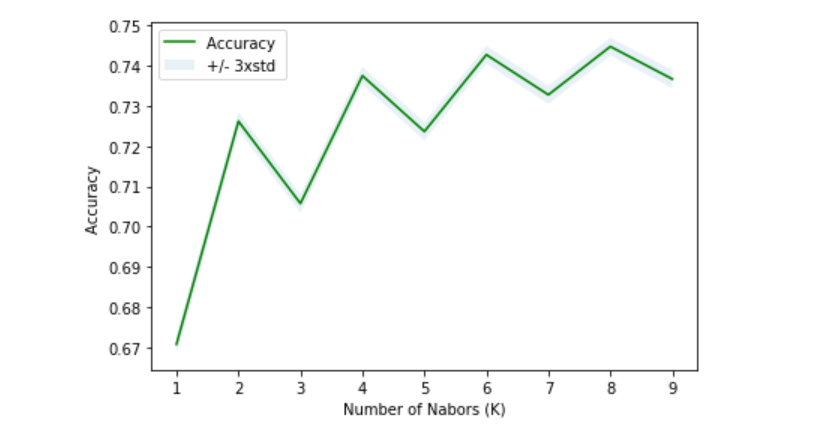
****

**Model Evaluation:**

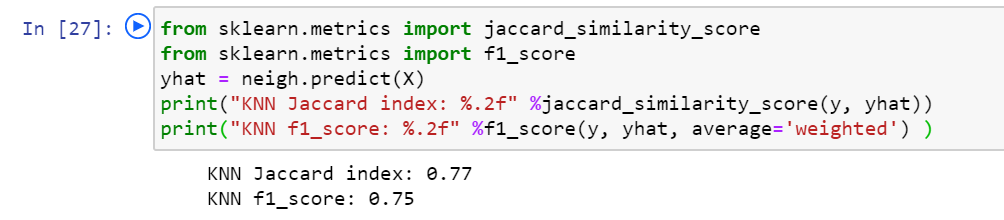
****

**Best Accuracy:**

**Best Accuracy=74% when k=8**

****

**Model evaluation using Jaccard Index and F1\_Score:**

****

**Conclusion:**

**We got approximately 75% accuracy for the car accident severity data set with 194673 instances using k-nearest neighbors algorithm.**