SVM Process Documentation

1. Imported the necessary libraries
    a. from msilib.schema import Class
    b. import pandas as pd
    c. import numpy as np
    d. from sklearn.linear_model import LogisticRegressionCV
    e. from sklearn.model_selection import train_test_split
    f. from sklearn.svm import SVC
    g. from sklearn.metrics import confusion_matrix, accuracy_score
    h. from sklearn import preprocessing
    i. from sklearn.model_selection import cross_val_score
    j. from sklearn.metrics import classification_report
2. Read the cars93 csv file using read_csv('cars93.csv') and set it to variable df
3. Cleaning Data
    a. Created a copy of our dataframe using .copy()
    b. Dropped columns not related to our experiment: 'Unnamed: 0', 'Manufacturer', 'Model', 'Type', 'Min.Price', 'Max.Price', 'AirBags', 'DriveTrain', 'Rev.per.mile', 'Passenger', 'Length', 'Wheelbase', 'Width', 'Turn.circle', 'Rear.seat.room', 'Luggage.room', 'Origin', 'Make'
    c. Dropped singular record where 'Cylinders' = 'rotary'
4. Created variables numericData and categoricalData and created copies using copy()
    a. numericData = all columns except 'Mans.trans.avail'
    b. categoricalData = 'Mans.trans.avail'
    c. Set categoricalData for Yes = 1 and No = 0 using map()
5. Used concat() to combine the two together

6. Set x to numericData and y to categorialData
7. Set x and y trains and tests using test size 0.25
8. Used algorithm SVC with default kernel='rbf', c =2, gamma='scaled'
    a. Used **fit** operator on x train and y train
9. Got predictions using x_test
10. Found accuracy by taking in (y_test, y_pred) * 100
11. Confusion matrix using y test and y pred
12. Printed the correlation matrix, confusion matrix, and the accuracy
13. Printed the score
14. Printed the CV (cross-validation) average
15. Printed the classification report
16. Improve accuracy
    a. scaled x using preprocessing attribute of the sklearn module
    b. Kept default kernel = rbf
    c. Changed C from C=1 to C=2
    d. Kept default gamma to 'scale'
17. The Gaussian Naive Bayes model is not as accurate as the SVM model. This could be due to the fact that the columns we are using are not independent of one another in reality. Because Naive Bayes assumes independence it cannot be as accurate when those assumptions are not upheld.