PROBLEM STATEMENT

TECHNOLOGY: DATA SCIENCE

Students from different cities from the state of Maharashtra had applied for the Cloud Counselage Internship Program. We have the dataset of consisting information of all the students. Using this data we want to get more insights and draw out more meaningful conclusions. Interns are expected to build a data visualization model and find the best data segmentation model using the student's dataset. Following are the tasks interns need to perform :

1. Interns need to preprocess the data for missing values, unknown values, encoding categorical values.
2. Create a data visualization model to build graphs from the dataset answering the following questions:

a. The number of students applied to different technologies.
b. The number of students applied for Data Science who knew ''Python" and who didn't.
c. The different ways students learned about this program.
d. Students who are in the fourth year and have a CGPA greater than 8.0.
e. Students who applied for Digital Marketing with verbal and written communication score greater than 8.
f. Year-wise and area of study wise classification of students.
g. City and college wise classification of students.
h. Plot the relationship between the CGPA and the target variable.
i. Plot the relationship between the Area of Interest and the target variable.
j. Plot the relationship between the year of study, major, and the target variable.

3. Identify the best binary classifier to classify data into "eligible/1" and "not eligible/0".

Input file

https://cloudcounselage24.bitrix24.com/disk/showFile/36677/?&ncc=1&ts=1590841589&filename=DS_DATESET.csv

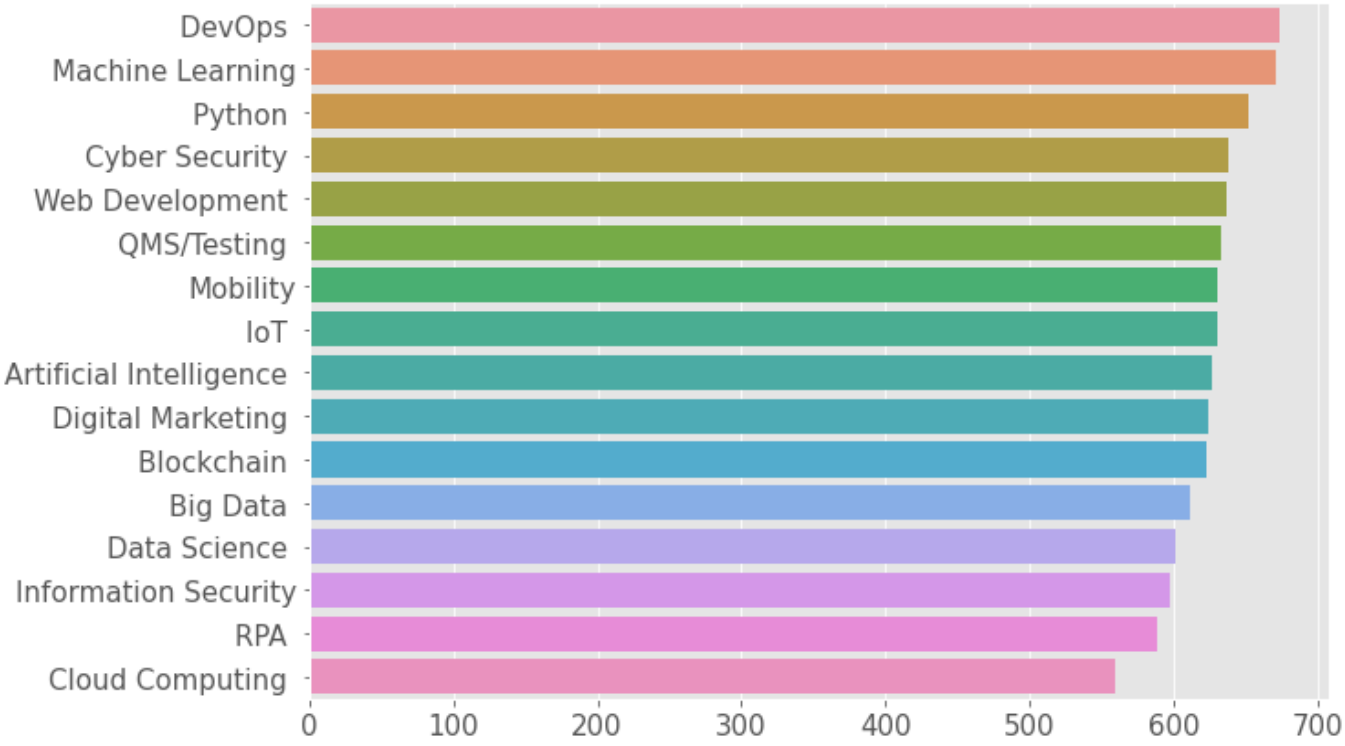
Input

Path of the input dataset file

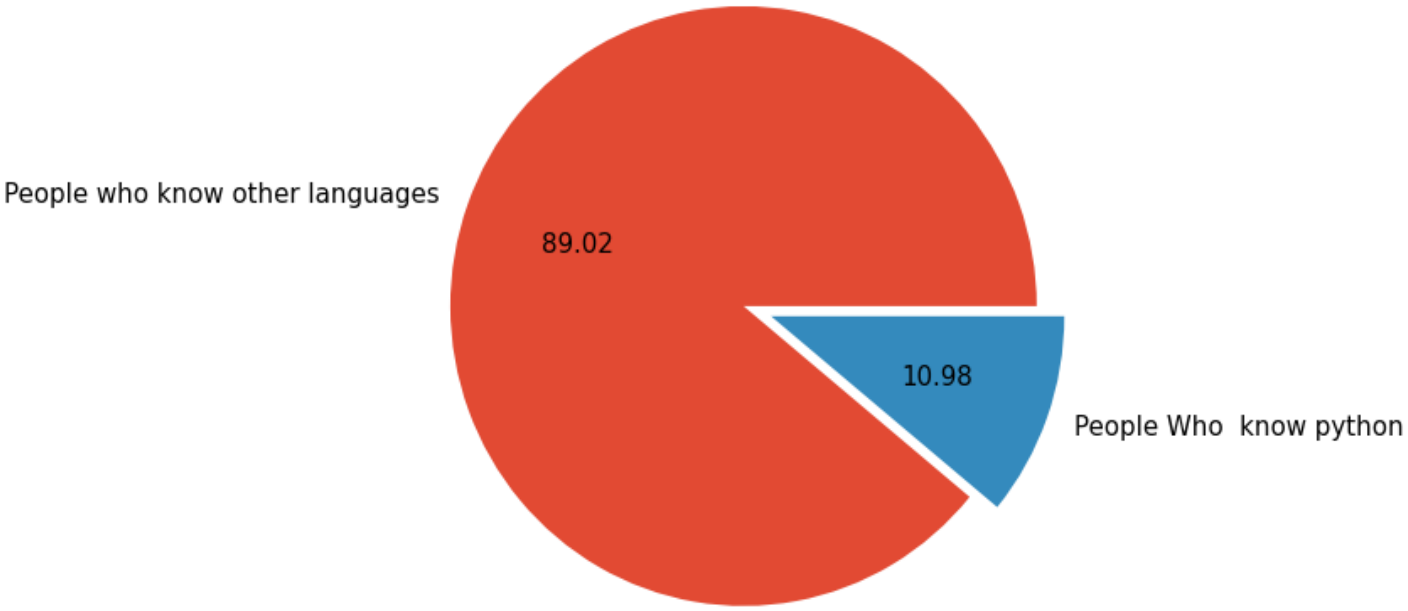
Output

For the data classification, In output print accuracy of the best model identified.
For the data visualization model, in output code need to create a pdf file of all the graphs generated by the model.

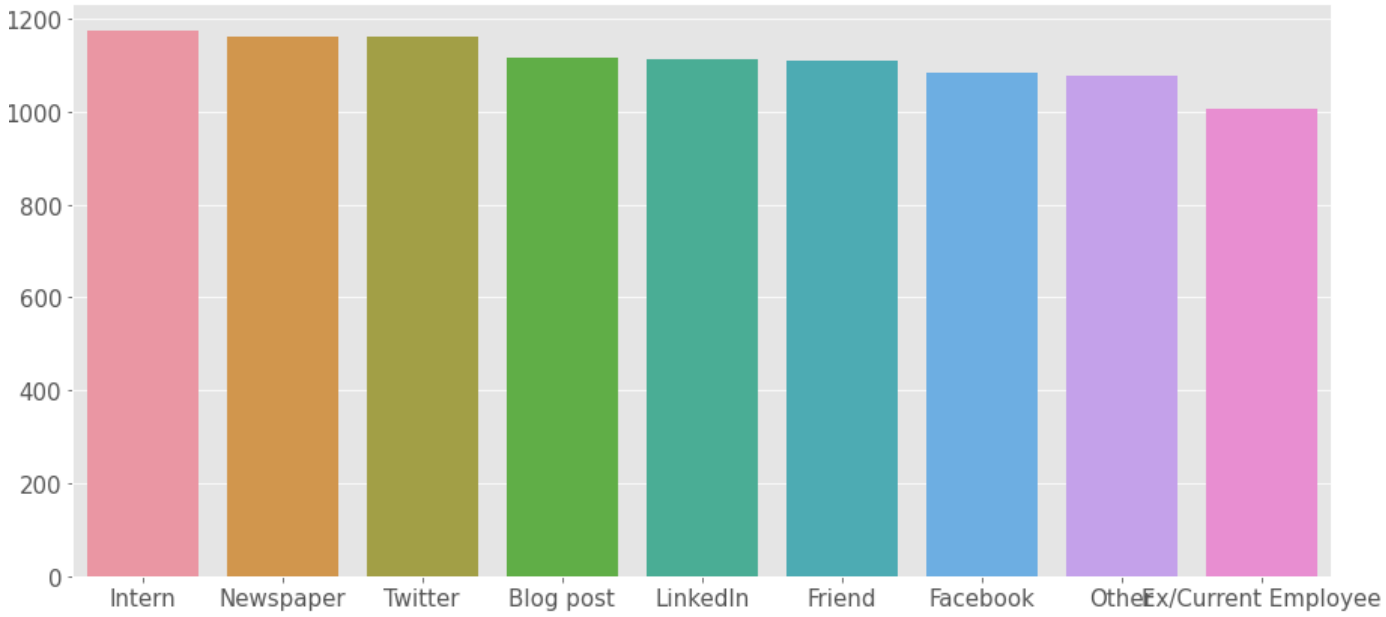
a. The number of students applied to different technologies.

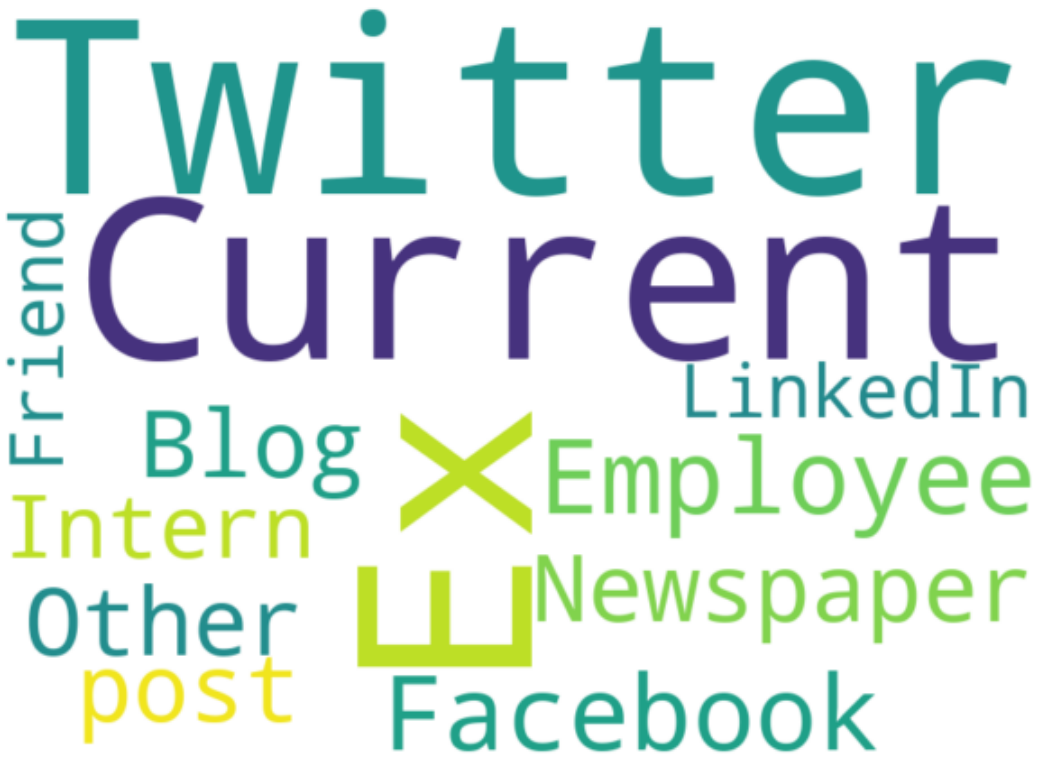


b. The number of students applied for Data Science who knew ‘’Python” and who didn’t.

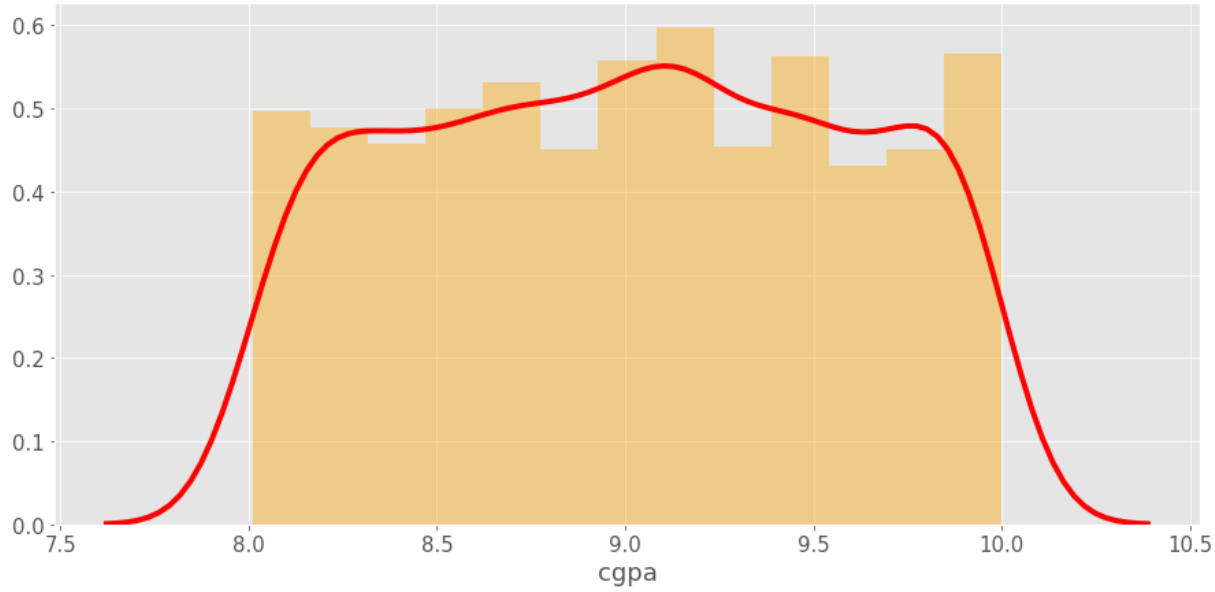


c. The different ways students learned about this program.

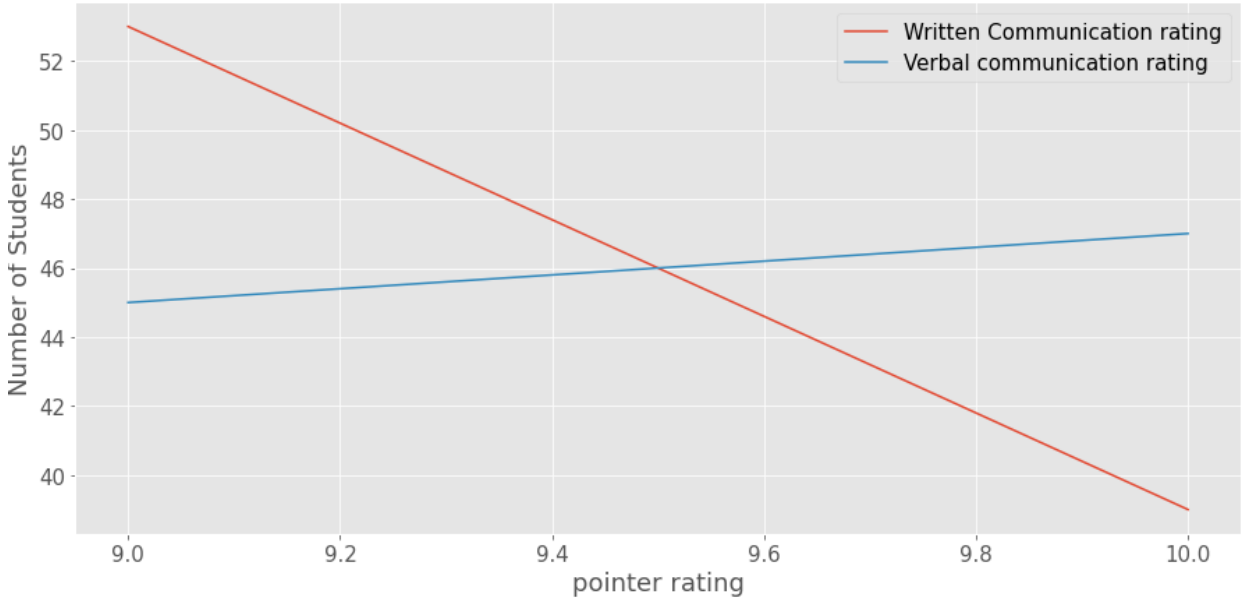




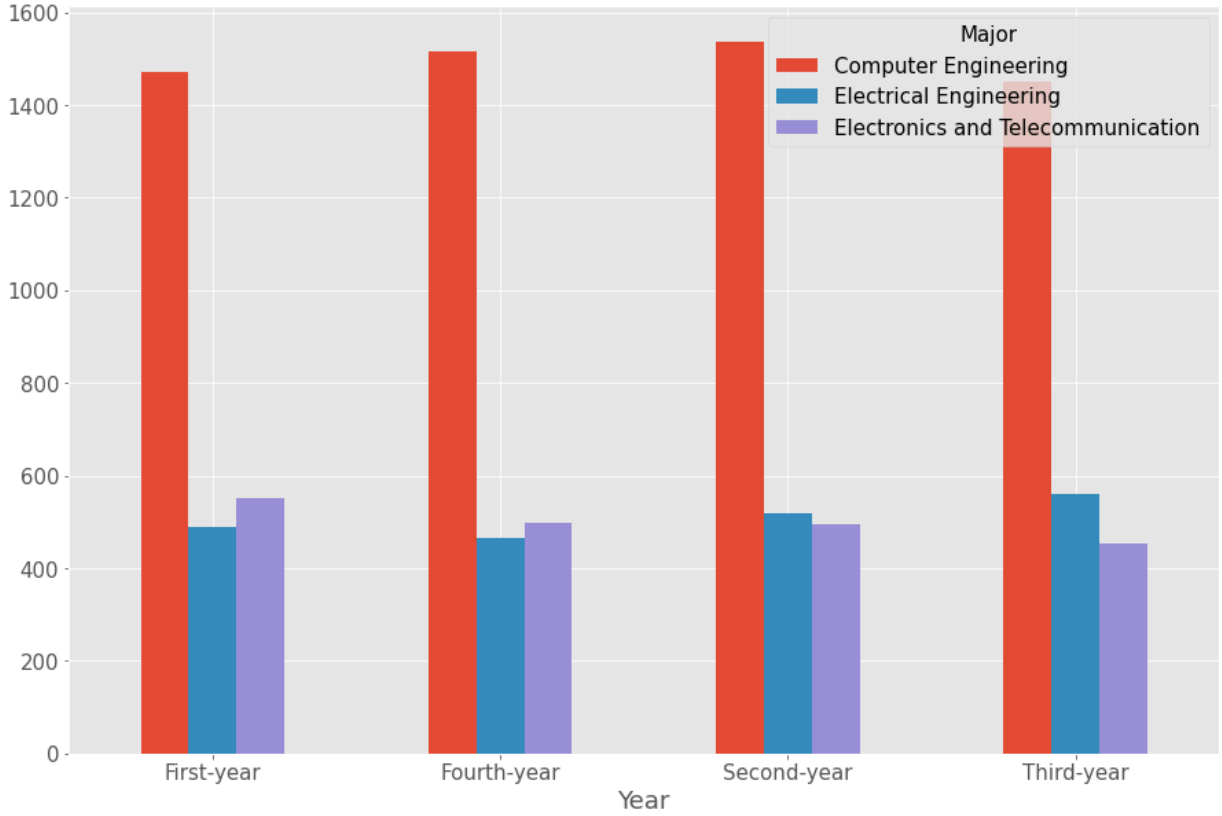
d. Students who are in the fourth year and have a CGPA greater than 8.0



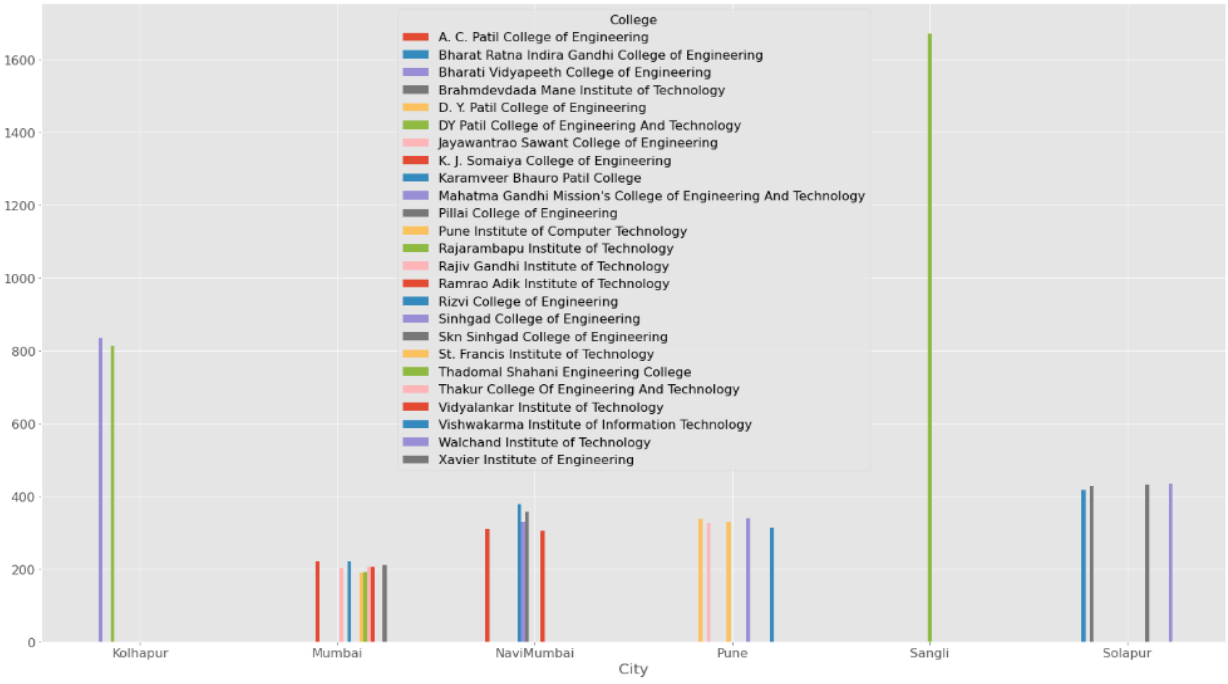
e. Students who applied for Digital Marketing with verbal and written communication score greater than 8



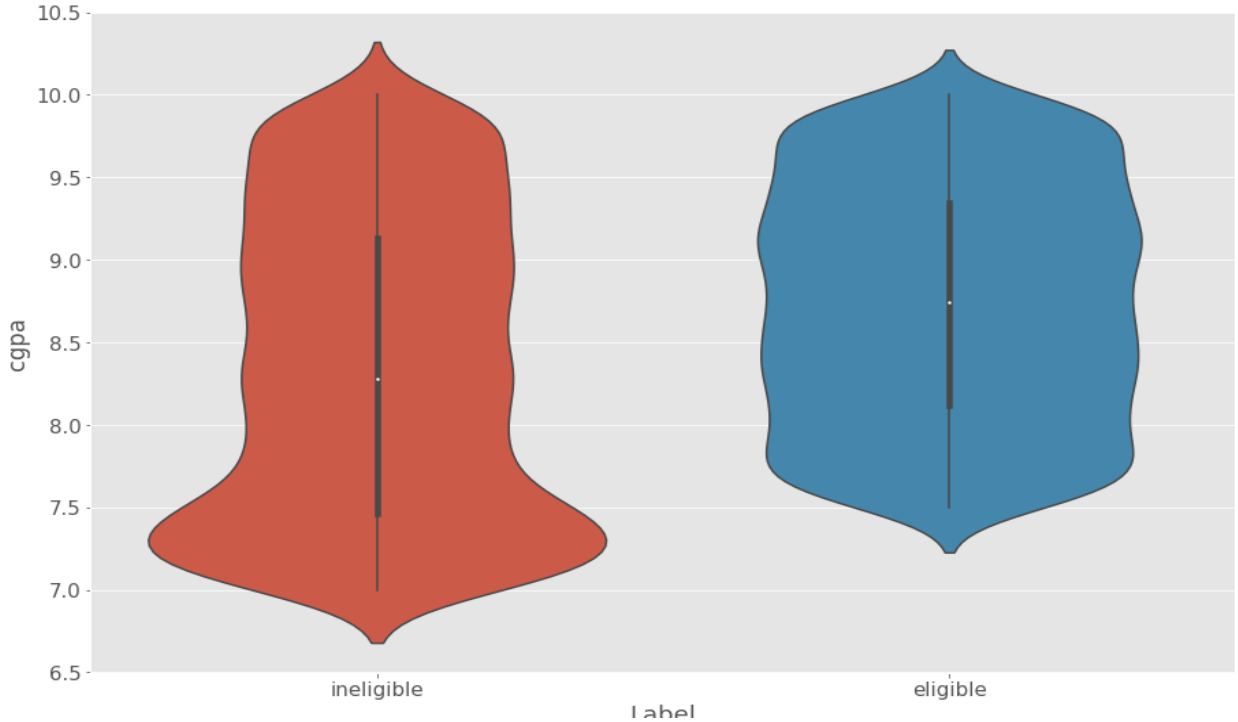
f. Year-wise and area of study wise classification of students.



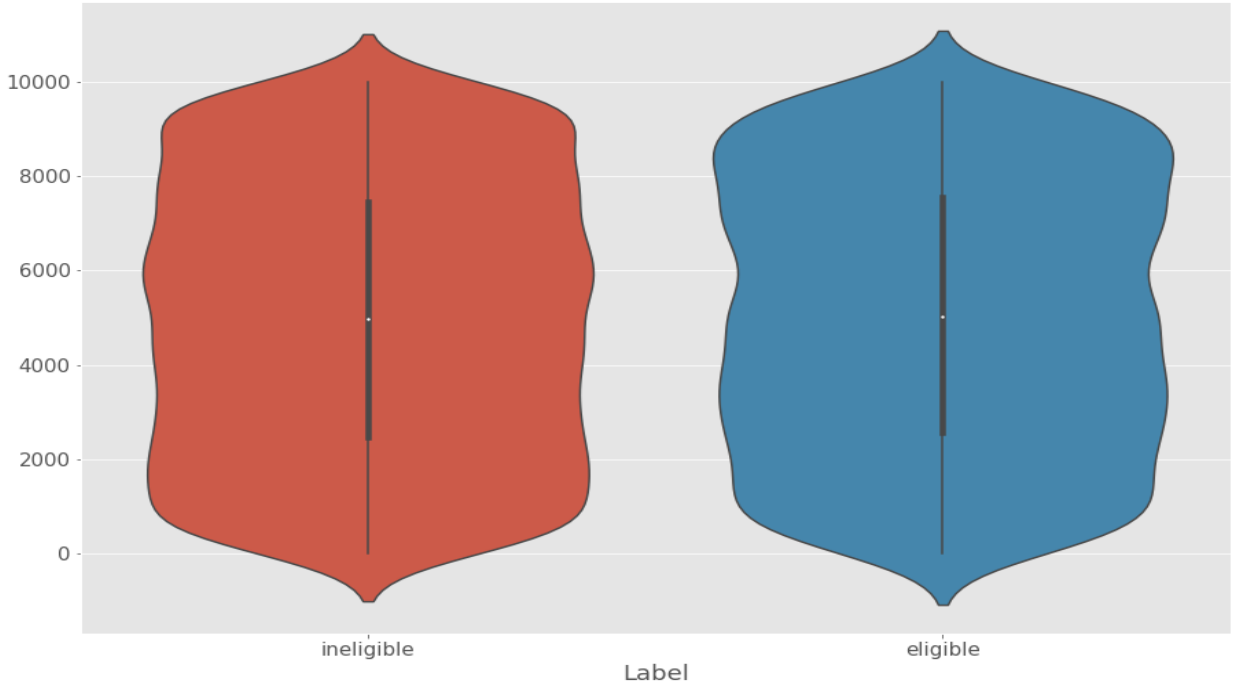
g. City and college wise classification of students



h. Plot the relationship between the CGPA and the target variable.



i. Plot the relationship between the Area of Interest and the target variable.



j. Plot the relationship between the year of study, major, and the target variable.