Students' Early Attrition Modelling for Clearwater State University

CAPSTONE PROJECT REPORT

BY- **SAMYAK NAYAK**

OUTLINE

- INTRODUCTION
- GOAL
- DATASET DESCRIPTION
- DATA VISUALIZATION
- MODELS
- RESULTS
- FEATURE IMPORTANCE PLOT
- PARTIAL DEPENDENCY PLOT
- OBSERVATIONS
- CONCLUSION

INTRODUCTION

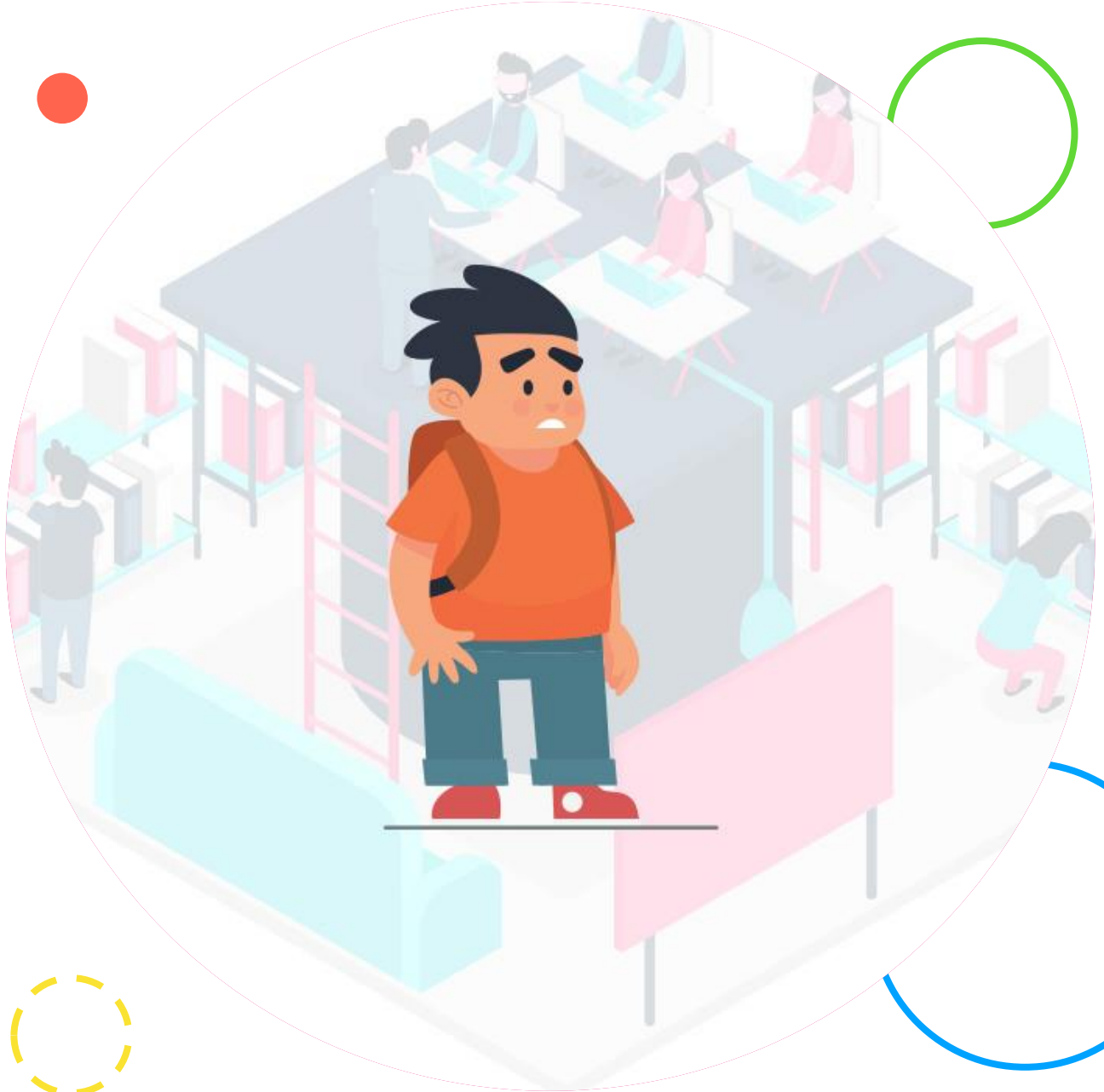
Clearwater State University offers a wide variety of degree programs, from online degrees to a doctorate in education.

Some key strategic goals of the University are:

- Increase enrolment of students
- Improve retention, progression and graduation rates
- Recruit better academically qualified undergraduate and graduate students
- Increase external funding and recognition

One of the strategic goal of the university is to reduce the early attrition rate, among students, which would help university to improve retention, progression as well as graduation rates.

Early attrition is defined as student drop out within one year of successfully joining the program (completed admission process).



GOAL

The aim of this project is to identify the key drivers of early attrition and build a predictive model to identify students with higher early attrition risk, based on the data related to student’s demographic profile, course preferences, performance record, grades, financial background, financial aid and other application information.

DATASET DESCRIPTION

The dataset is about student's demographic profile, course preferences, performance record, grades, financial background, financial aid and other application information .

There are 3400 data points, with 56 features, out of which 36 are categorical variables, and 20 are numerical ones.

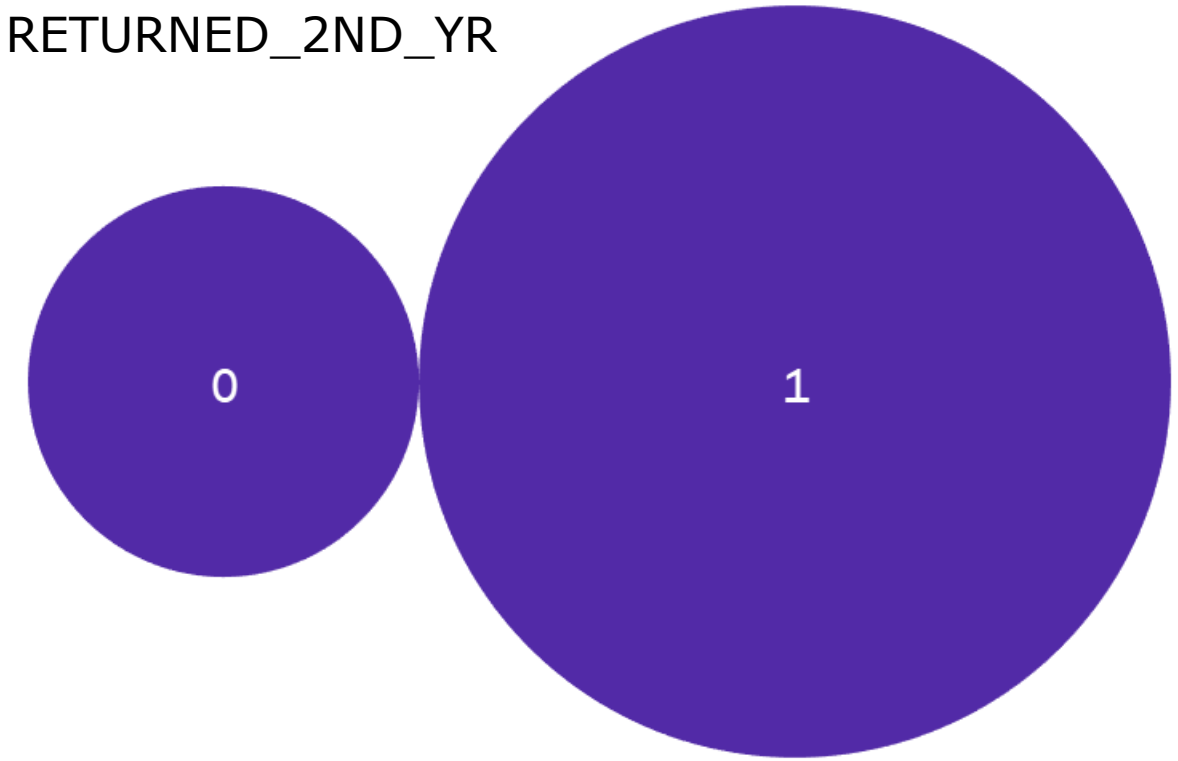
The target variable is –RETURNED_2ND_YR, having binary values (0 for "NO" and 1 for "YES").All other variables are independent ones.

VARIABLE DICTIONARY

Variables	Description
CORE_COURSE_GRADE_2_S	Grade in Core course 2 opted in Second semester
CORE_COURSE_NAME_3_S	Core course 3 opted in Second semester
CORE_COURSE_GRADE_3_S	Grade in Core course 3 opted in Second semester
CORE_COURSE_NAME_4_S	Core course 4 opted in Second semester
CORE_COURSE_GRADE_4_S	Grade in Core course 4 opted in Second semester
CORE_COURSE_NAME_5_S	Core course 5 opted in Second semester
CORE_COURSE_GRADE_5_S	Grade in Core course 5 opted in Second semester
CORE_COURSE_NAME_6_S	Core course 6 opted in Second semester
CORE_COURSE_GRADE_6_S	Grade in Core course 6 opted in Second semester
HOUSING_STS	Indicator of whether the student is staying in campus or outside
RETURNED_2ND_YR	Indicates whether the student came back to First semester in 2nd year
DISTANCE_FROM_HOME	Distance from the university to student's home
HIGH_SCHL_GPA	Student's High School GPA score
HIGH_SCHL_NAME	High School from where the student graduated
FATHER_HI_EDU_CD	Father's educational status code
FATHER_HI_EDU_DESC	Father's educational status
MOTHER_HI_EDU_CD	Mother's educational status code
MOTHER_HI_EDU_DESC	Mother's educational status
DEGREE_GROUP_CD	Degree code for which student has enrolled in university
DEGREE_GROUP_DESC	Degree for which student has enrolled in university
FIRST_TERM_ATTEMPT_HRS	# Hours attempted by student(Or # Grade points attempted by Student in First semester)
FIRST_TERM_EARNED_HRS	# Hours earned by student(Or # Grade points earned by Student in First semester)
SECOND_TERM_ATTEMPT_HRS	# Hours attempted by student(Or # Grade points attempted by Student in Second semester)
SECOND_TERM_EARNED_HRS	# Hours earned by student(Or # Grade points earned by Student in Second semester)
GROSS_FIN_NEED	Financial need of Student
COST_OF_ATTEND	Course Fees
EST_FAM_CONTRIBUTION	Estimated Family contribution towards course fees

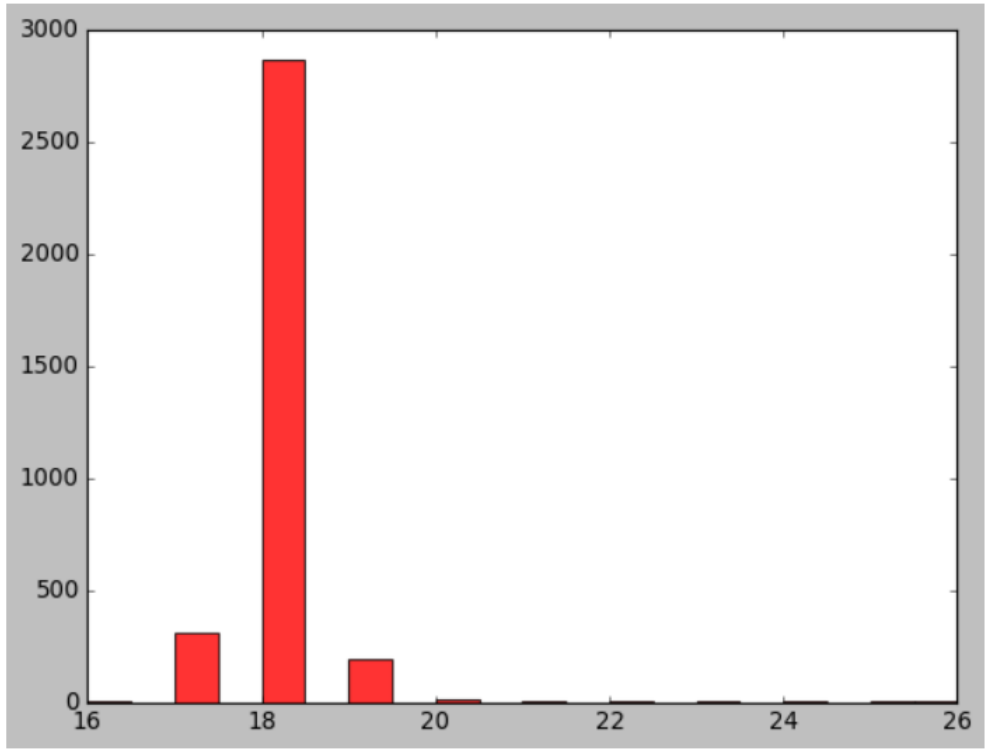
Variables	Description
STUDENT IDENTIFIER	Student Identifier
STDNT_AGE	Age of the Student Enrolled
STDNT_GENDER	Gender of the student
STDNT_BACKGROUND	Backgroud of Student
IN_STATE_FLAG	Indicator of whether Student is in same state as university
INTERNATIONAL_STS	Indicator of whether Student is an International Student
STDNT_MAJOR	Student's Major course in University
STDNT_MINOR	Student's Minor course in University
STDNT_TEST_ENTRANCE1	Student's Entrance 1 score
STDNT_TEST_ENTRANCE2	Student's Entrance 2 score
STDNT_TEST_ENTRANCE_COMB	Student's score calculated both on Entrance1 & Entrance2 score
FIRST_TERM	First semester year
CORE_COURSE_NAME_1_F	Core course 1 opted in First semester
CORE_COURSE_GRADE_1_F	Grade in Core course 1 opted in First semester
CORE_COURSE_NAME_2_F	Core course 2 opted in First semester
CORE_COURSE_GRADE_2_F	Grade in Core course 2 opted in First semester
CORE_COURSE_NAME_3_F	Core course 3 opted in First semester
CORE_COURSE_GRADE_3_F	Grade in Core course 3 opted in First semester
CORE_COURSE_NAME_4_F	Core course 4 opted in First semester
CORE_COURSE_GRADE_4_F	Grade in Core course 4 opted in First semester
CORE_COURSE_NAME_5_F	Core course 5 opted in First semester
CORE_COURSE_GRADE_5_F	Grade in Core course 5 opted in First semester
CORE_COURSE_NAME_6_F	Core course 6 opted in First semester
CORE_COURSE_GRADE_6_F	Grade in Core course 6 opted in First semester
SECOND_TERM	Second semester year
UNMET_NEED	Unmet financial need of the student
CORE_COURSE_NAME_1_S	Core course 1 opted in Second semester
CORE_COURSE_GRADE_1_S	Grade in Core course 1 opted in Second semester
CORE_COURSE_NAME_2_S	Core course 2 opted in Second semester

DATA VISUALISATIONS

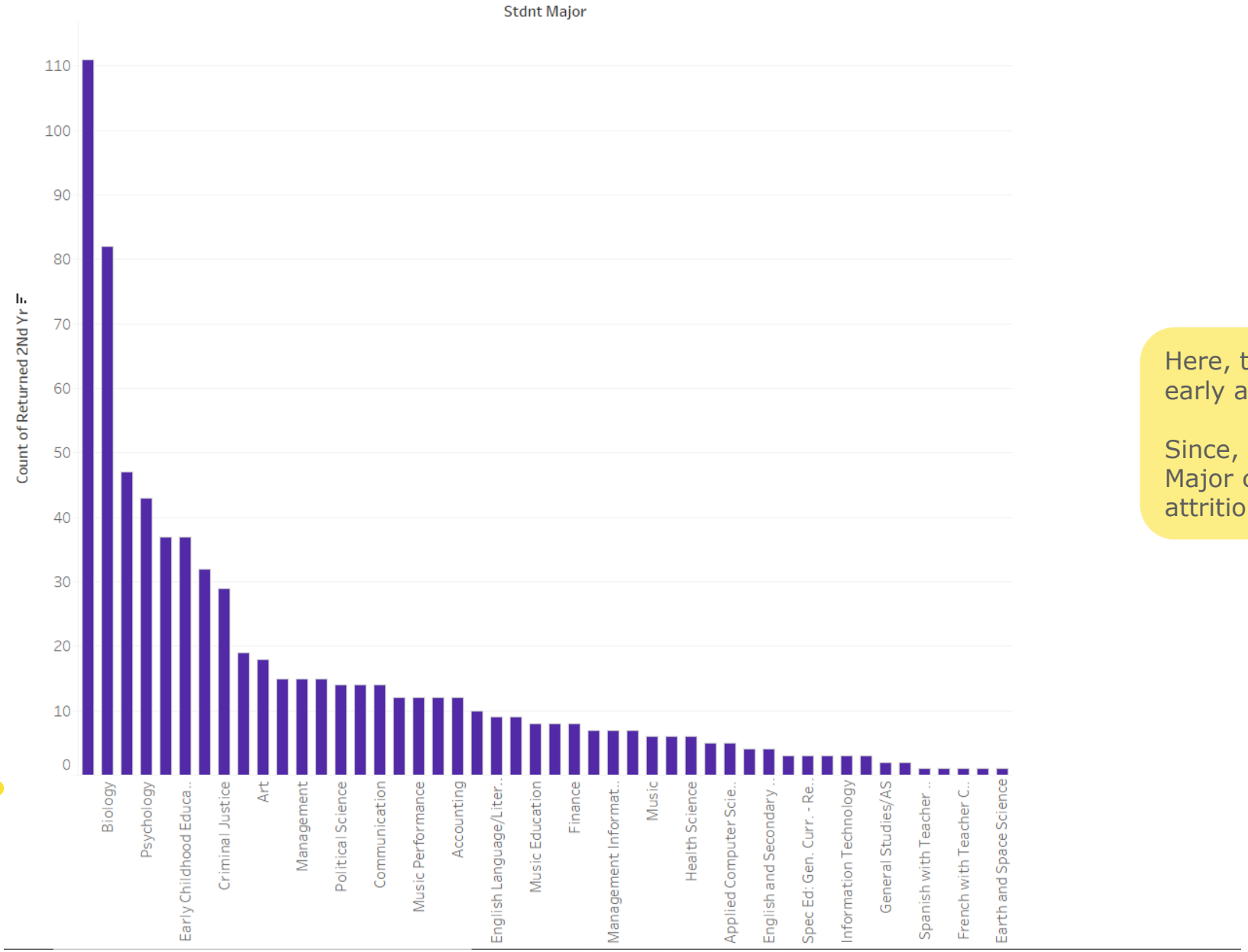


The target variable-RETURNED_2ND_YR contains only binary values.

STDNT_AGE

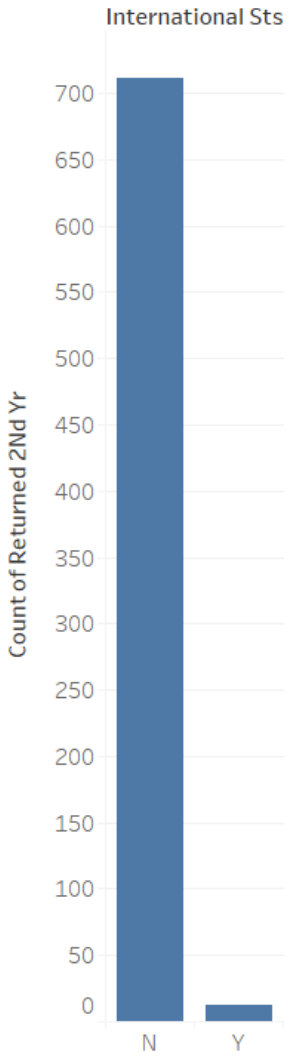


From this plot, it can be concluded that the STDNT_AGE is right skewed.

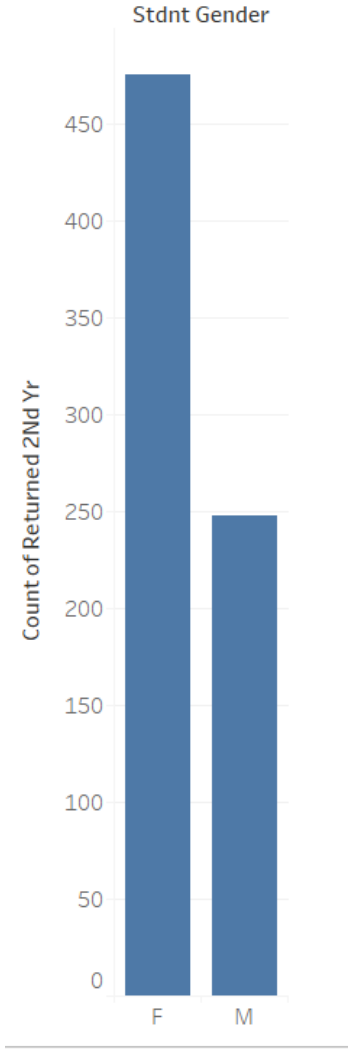


Here, the plot shows the count of number of early attritions from each major course.

Since, the count differs for each category, Major course might be a factor affecting early attrition rate.



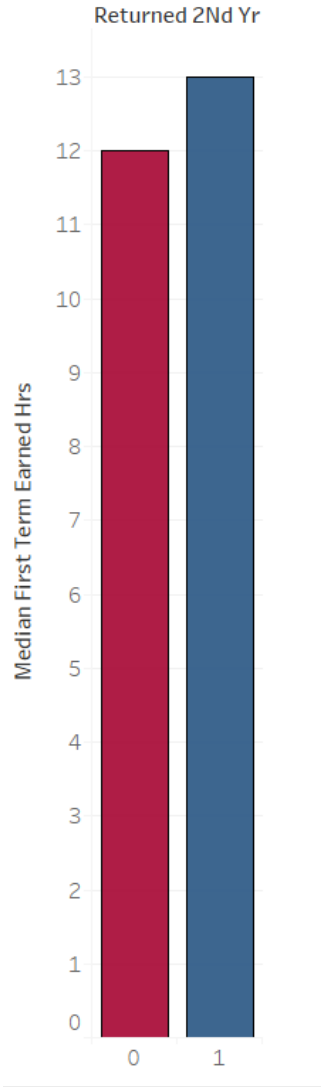
This plot shows, the number of early attritions for each category – whether the student is international or not.

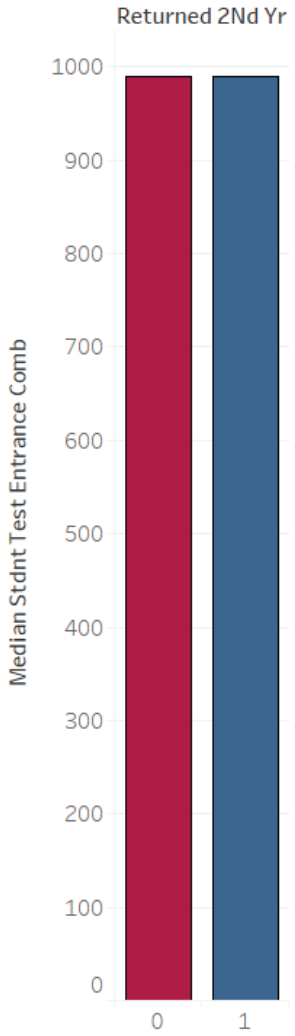


This plot shows, the number of early attritions for each category – whether the student is male or female.

This plot shows, the median of First Term Earned Hrs, for each category of Returned_2nd_Year.

There seems a little association between these two variables.

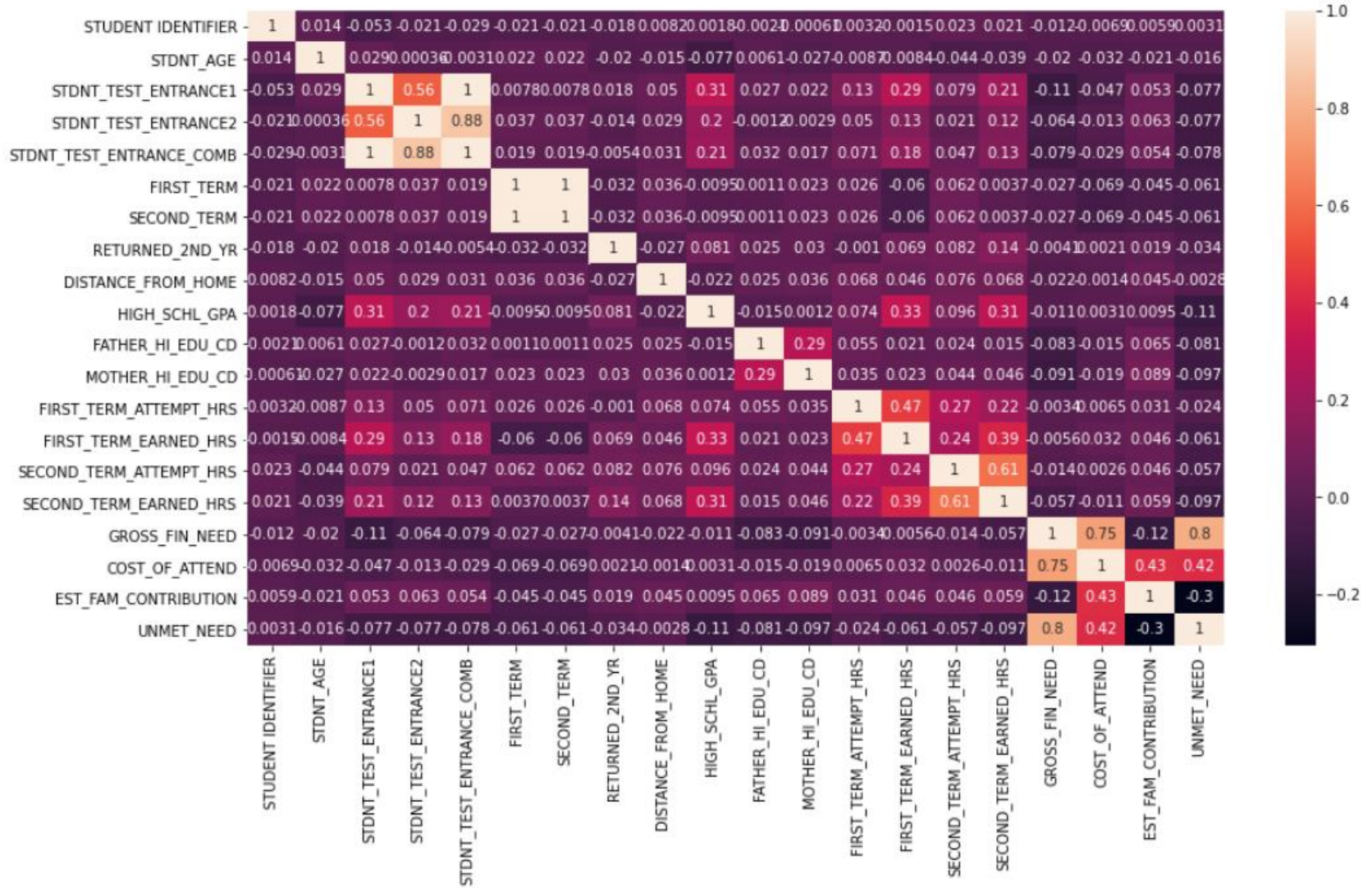




This plot shows, the median of Student Test Entrance Test score, for each category of Returned_2nd_Year.

There doesn't seem much relation between these two variables.

The Heat-Map (or Correlation Matrix) for all the variables in our dataset.

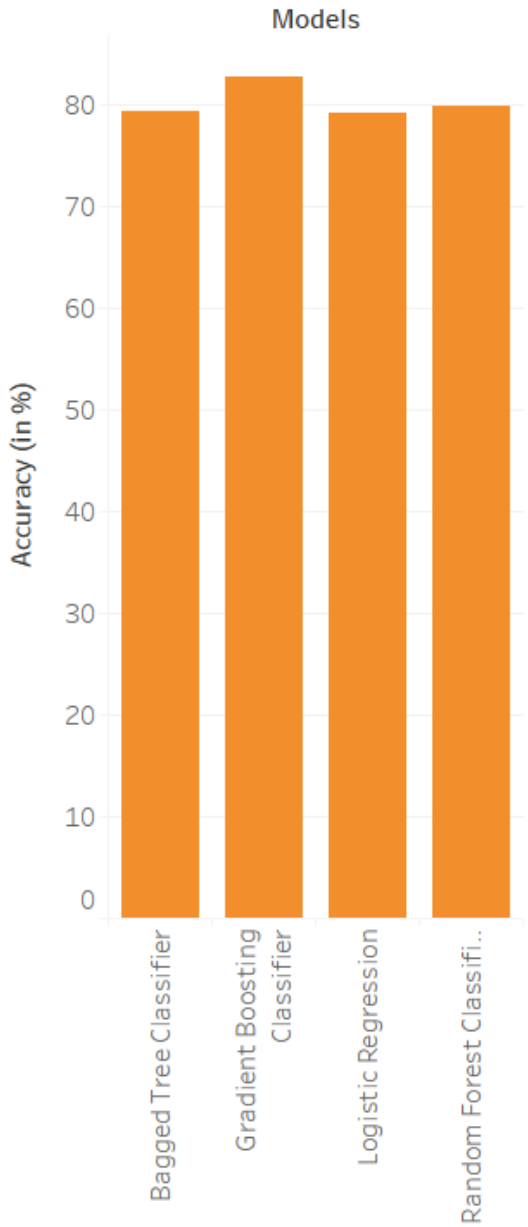


MODELS USED

- LOGISTIC REGRESSION
- RANDOM FOREST CLASSIFIER
- BAGGING TREE CLASSIFIER
- GRADIENT BOOSTING CLASSIFIER

RESULT

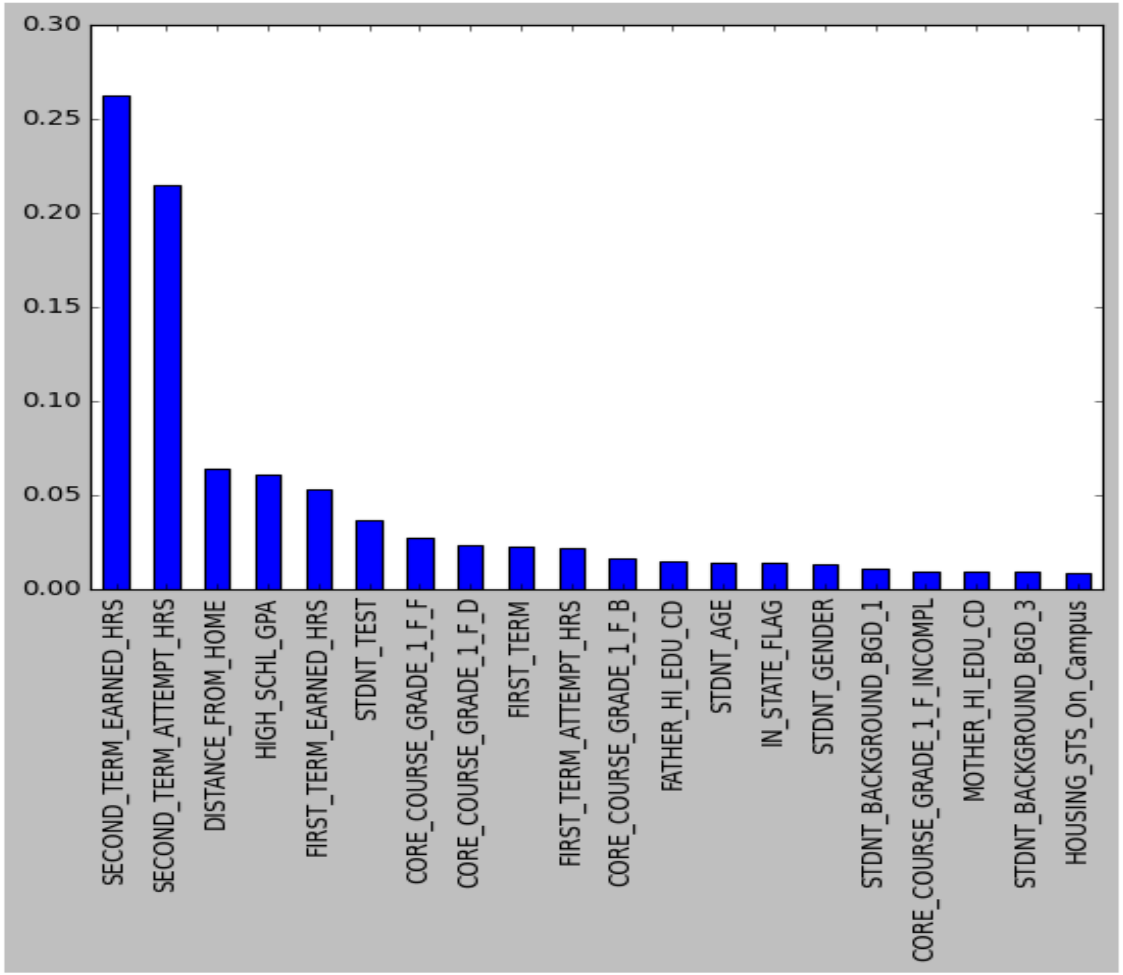
The best working model, on the basis of Accuracy is Gradient Boosting Classifier, with an accuracy of 82.68%.



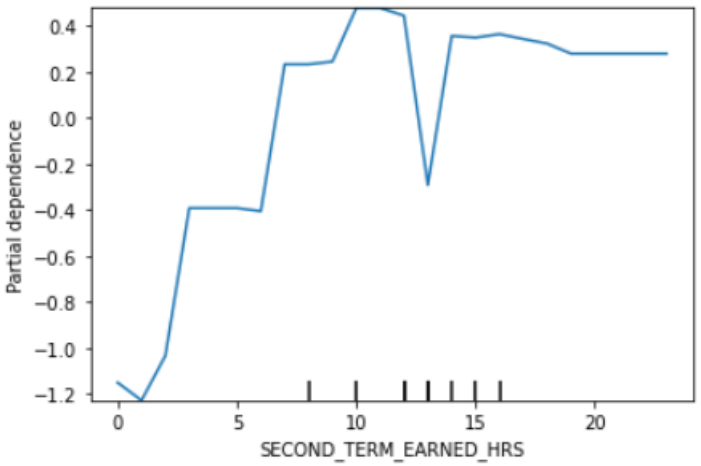
FEATURE IMPORTANCE PLOT FOR GRADIENT BOOSTING CLASSIFIER

The variables, affecting the early attrition rate is-

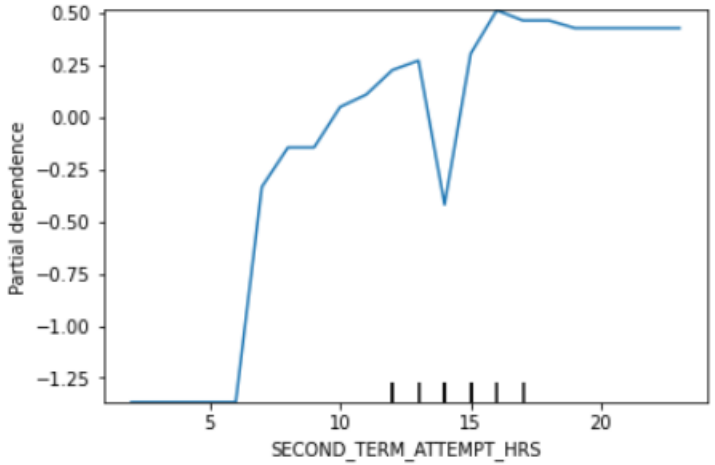
SECOND_TERM_EARNED_HRS, SECOND_TERM_ATTEMPT_HOURS, DISTANCE_FROM_HOME, HIGH_SCHOOL_GPA, FIRST_TERM_EARNED_HRS, STDNT_TEST, CORE_COURSE_GRADE, FIRST_TERM, FIRST_TERM_ATTEMPT_HOURS, FATHER_HI_EDU_CD, STDNT_AGE, STDNT_GENDER, STDNT_BACKGROUND



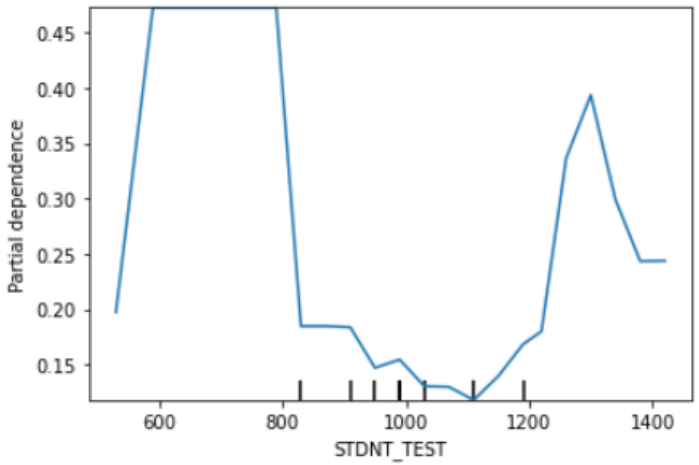
PARTIAL DEPENDENCY PLOTS



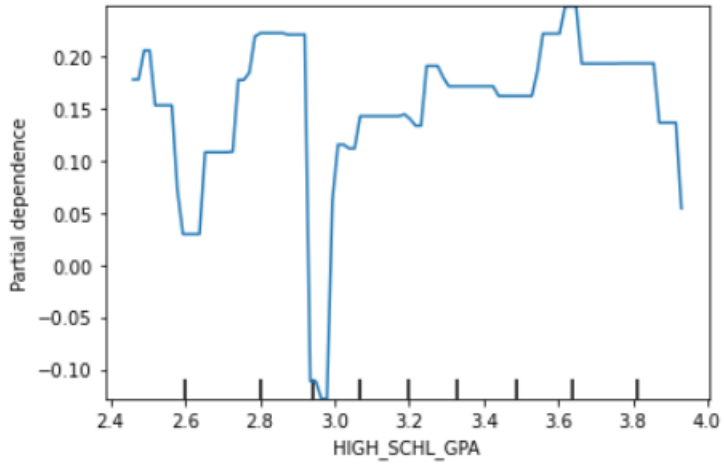
This partial dependency plot shows that, as students' SECOND_TERM_EARNED_HRS increases, their chances of early attrition decreases



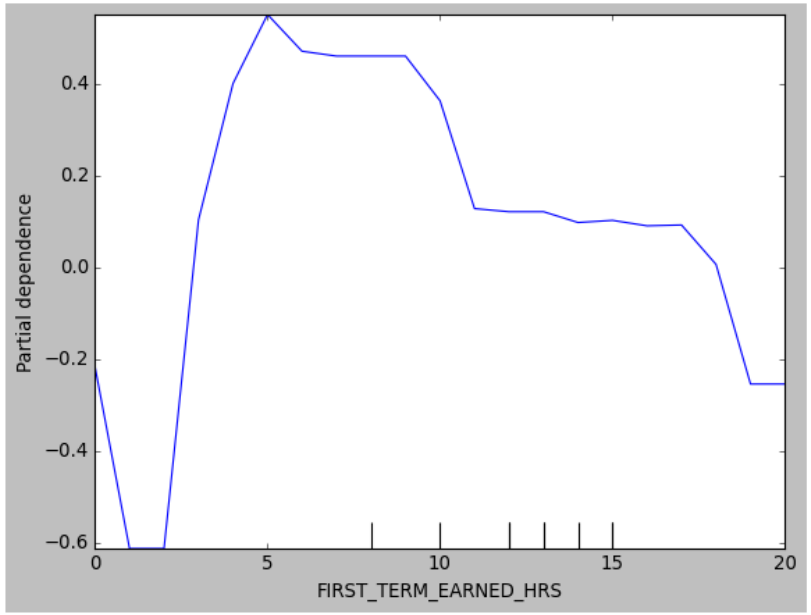
This partial dependency plot shows that, as students' SECOND_TERM_ATTEMPT_HRS increases, their chances of early attrition decreases



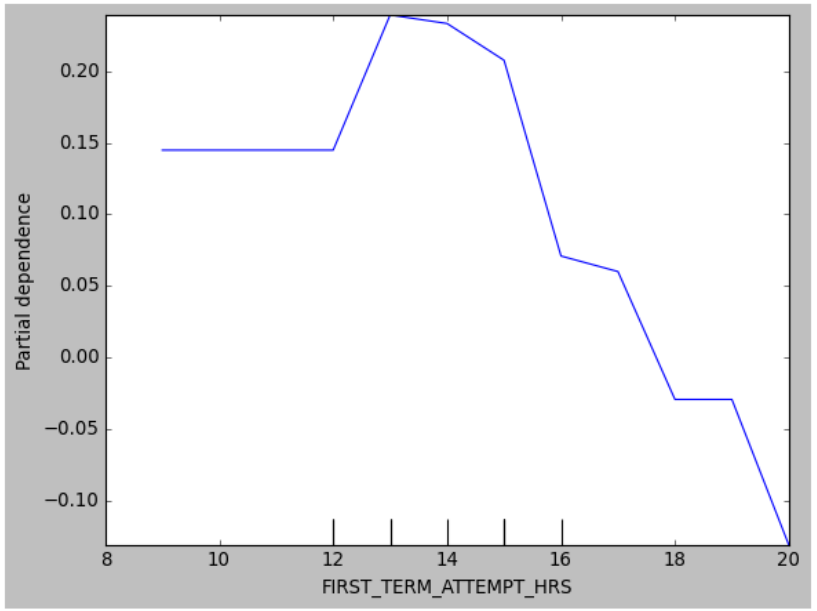
This partial dependency plot shows that- students having Combined test marks between 800-1200 are more likely to leave the college after first year.



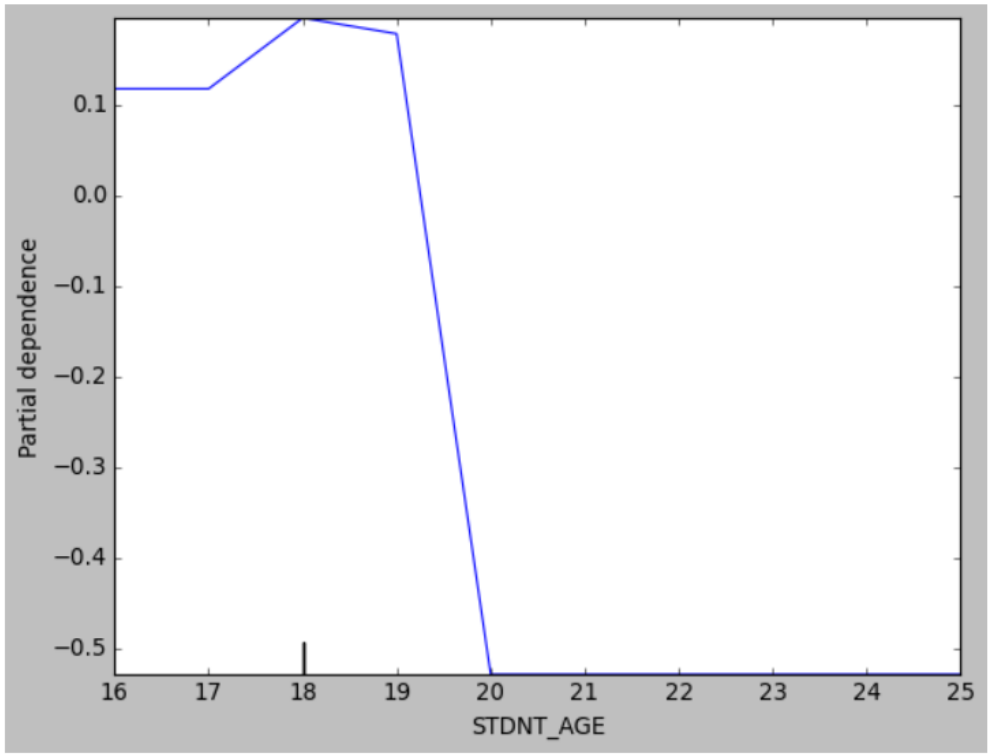
This partial dependency plot shows some interesting facts- students having HIGH_SCHL_GRADES around 3.0 are more likely to leave school after first year



This partial dependency plot shows that- students who have earned between 15-20 hours in their first term are more likely to go for early attrition, than the ones having earned hours between 5-10.



This partial dependency plot shows that, as students' FIRST_TERM_ATTEMPT_HRS increases, their chances of early attrition increases.



This partial dependency plot shows that, as students' AGE increases, their chances of early attrition increases as well.

OBSERVATIONS

1. As the students' SECOND_TERM_EARNED_HRS increases, their chances of early attrition decreases.
2. As students' SECOND_TERM_ATTEMPT_HRS increases, their chances of early attrition decreases
3. The students having Combined test marks between 800-1200 are more likely to leave the college after first year
4. The students having HIGH_SCHL_GRADES around 3.0 are more likely to leave school after first year
5. The students who have earned between 15-20 hours in their first term are more likely to go for early attrition, than the ones having earned hours between 5-10.
6. The as the students' FIRST_TERM_ATTEMPT_HRS increases, their chances of early attrition increases.

CONCLUSION

Based on the observations, that have been made, the university should take following steps in order to reduce the early attrition rate among students-

1. Students should be encouraged to attempt less hours in their first term, whereas attempt more hours in their second term.
2. University should conduct special counselling sessions for the students having age above 20.
3. University should have special counselling sessions for students who have Combined test scores between 800-1200, since they are more likely to go for early attrition.
4. The students who have HIGH_SCHL_GRADES around 3.0, should be given more focus, so that they are less likely to leave the university.