Mont St. Michel

Date: November 5, 2021

To: Professor Callaghan, President

From: Muktha Mukkanna Swamy, Poorvi Suhane, Venkat Naren Pagadala, Vijeta Mallesh Kodam

Subject: Effect of Seminars on Student Retention

**Problem statement:** Mont St. Michel college had a high dropout rate of students. The college wanted to know whether the voluntary one-credit seminars will have a positive impact on students’ retention in the college.

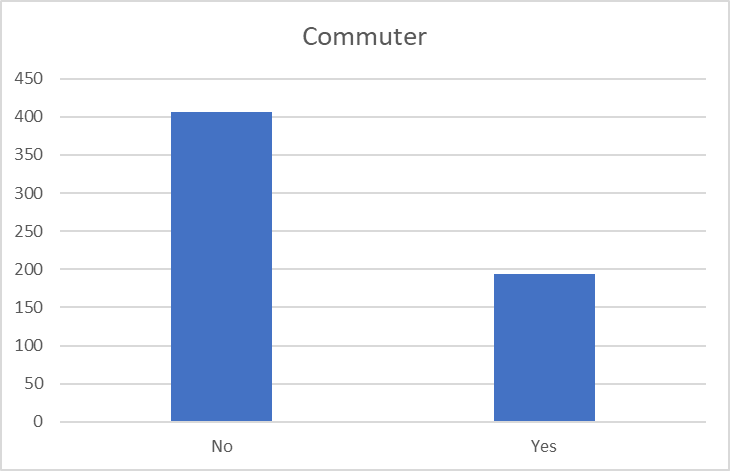
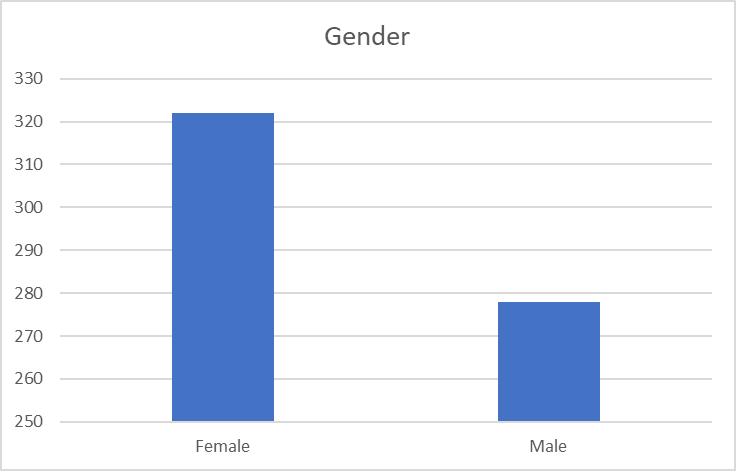
**Background:** The student retention percentage of Mont St. Michel college students in their first year was increasing. Consequently, the college had organized a one hour long seminar with the faculty to encourage the students to build connections on campus. The college also wanted to know if the seminar had a positive impact on student retention. They also intended to utilize this opportunity to persuade the college administrators to not resume the funding for this initiative.

The college administration felt that the probability of first year college students leaving at the end of the year had lower high-school GPAs. Along with this, The Student Support Administrator doubted that the students who enrolled in a high number of units had immense schedules which hence, discouraged them from staying enrolled in the course.

**Demographic variables:** Students retention is dependent on a wide variety of factors such as gender, age of the high school students, and whether those students live off the campus or not.

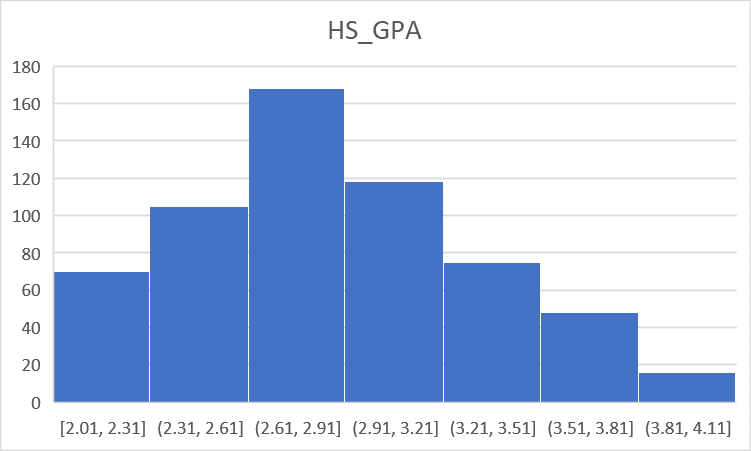
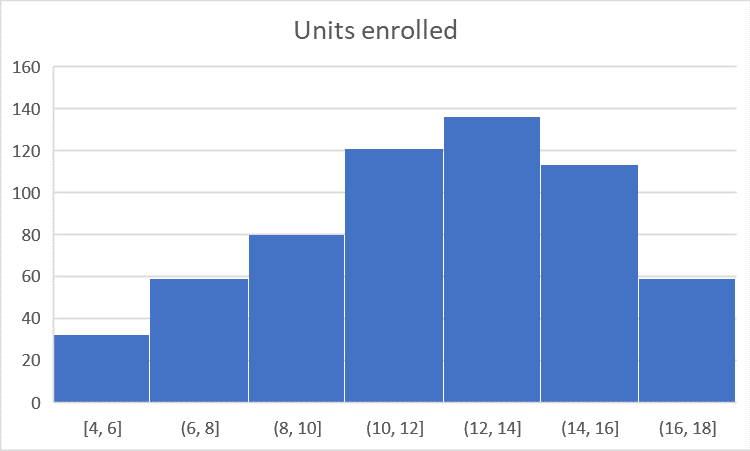
**Age:** The age of majority of the high school students were between 18-20 years which means more younger people were high school graduates.Out of all the students, 46.33 % were male and 53.67 % were female students. This reveals that there were 7.34 % more female students than male students who attended high school.

**Gender:** There were more female students than male students. Out of all the students, 46.33 % were male and 53.67% were female students. This reveals that there were 7.34 % more female students than the males who attended high school.

****

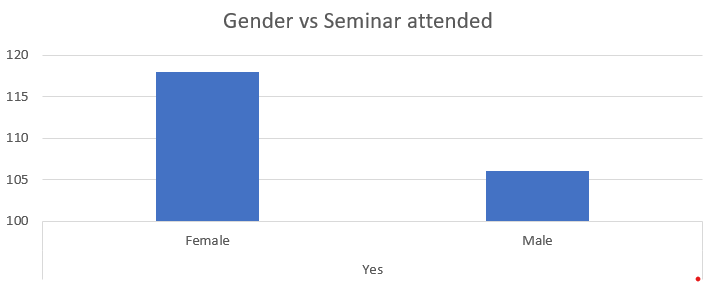
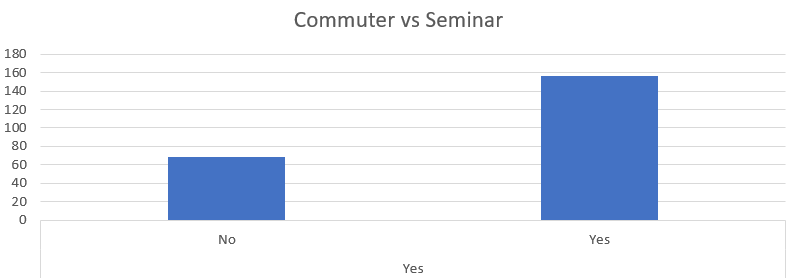
**Commuter:** 194 students commuted to the campus whereas 406 students lived near the campus. This implies that 67.67 % of the high school students had to travel a long distance to attend the classes.

**Units Enrolled:** The majority of students were enrolled in 12-14 units in their high school. The average number of units enrolled by the students was 12, whereas the median was 19, which indicates that the distribution is skewed to the left, i.e negatively skewed. This implies that the data is not normally distributed and the majority of the students had enrolled in more than 12 units. The maximum number of units taken by a student is 18 and the minimum was 4. This data is unimodal.



**High School GPA:** The mean and the median high school GPA of the students was 2.8. This signifies that the data were normally distributed and the majority of the students had GPAs around 2.8. The minimum GPA was 2.01 and the maximum GPA was 4. This data is unimodal.

**Seminar Attended:** Around 37.33 % of the students attended the seminar whereas 62.67 % of the students were not interested in attending the seminar. There were 118 female students and 106 male students who attended the seminar.

**Attending Seminar:** For student retention, the model can predict for about 82.3 %. We think that gender and GPA do not have any major effect on first year seminars for retention. Whether the student is local or not, makes a difference in the student’s decision of attending the seminar. This means there is approximately an 82.3 % chance that a local student will attend the seminar. Some fewer local people attended the seminar.

**Benefits of the seminar:** The first take on the proposed plan was to evaluate the number of students who attended the seminar.

**Reasons for considering these variables for regression:**

**Age:** Older people will prefer to be financially stable rather than attending high school and getting an education. Older people seem less interested in attending seminars as they have other commitments such as jobs, family. So, they are less likely to attend the seminar. Moreover, younger people have more curiosity to learn as compared to older people.

If attendance is a factor in determining the high school grades of students, then there is a possibility that some students who are even attending the seminar, are more concerned about their GPA rather than their enrollment in college.

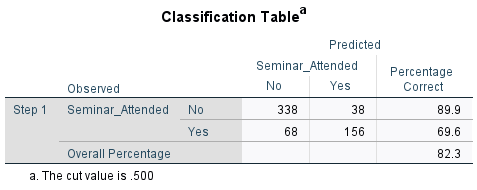
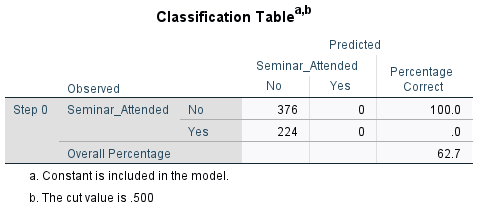
**Commuters:** Commuters are less likely to attend the seminar as they have to move their schedules and make time for the seminar. Commuting is an additional cost for students. Along with this, the time taken by students to travel via private/public transportation to attend the seminar will affect the total population of the seminar.

**Gender:** In some cultures, women are not encouraged to pursue quality education.

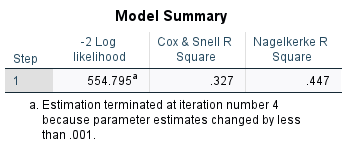
**GPA:** We think that students with a high GPA are more inclined towards learning and are more likely to attend the seminar.

**Units enrolled:** If the students are enrolled in more units, they would rather spend their time improving their academic performance and would not bother to attend the seminar.

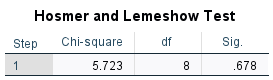
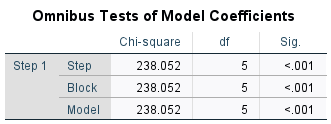
**Predictive model for attending the seminar :**

****

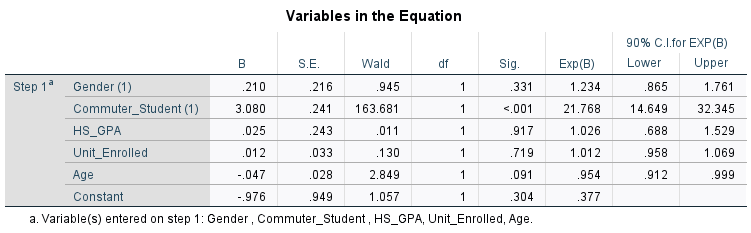
The accuracy of the model has been increased from 62.7 % to 82.3%. That means 82.3% values are correctly predicted.



The data we have, suggested to differ between 0.327 and 0.447 from the average of predicted values.

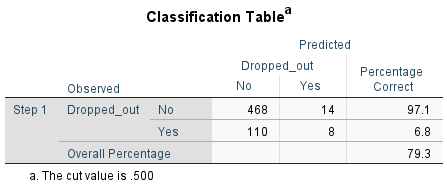
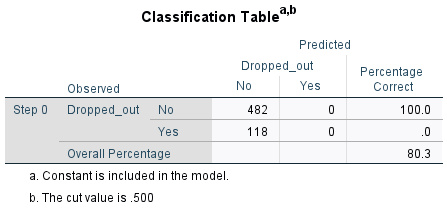


The Omnibus test is significant as the significance value is less than 0.001. And the Hosmer and Lemeshow tests are insignificant. Hence, the model is a good fit.

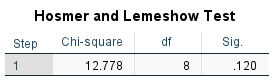
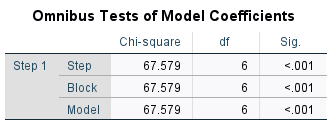


Whether the student will attend the seminar or not depends upon their age and the student's ability to commute as they are significant. However, the odds ratio of 21.768 for commuters indicates that the commuters will be less likely to attend the seminar and the Odds ratio of 0.954 for age also indicates that younger people are more likely to attend the seminar as compared to older people.

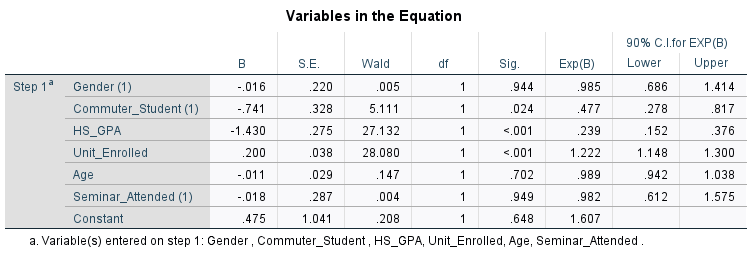
**Predictive model for Dropping out:**



The accuracy of the model has been decreased from 80.3% to 79.3% and the variance is suggested between 0.107 and 0.169.

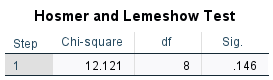
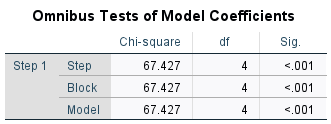


The Omnibus test is significant and the Hosmer and Lemeshow test is insignificant. Hence, the model is a good fit.

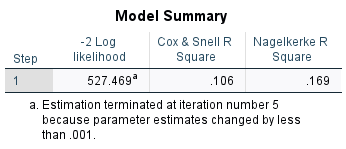
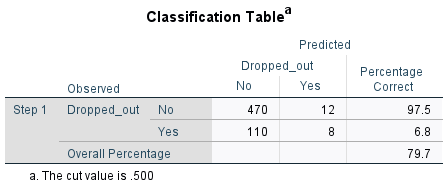


The student’s dropout decision depends on the student's GPA, the number of units they are enrolled in, and whether the student is a commuter or not. The odds ratio for a student's GPA implies that students with lower GPAs are more likely to drop out of college. The odds ratio for commuters indicates that students who are commuters are less likely to drop out of college and the odds ratio for units enrolled implies that students who are enrolled in more units are more likely to drop out after their first year. As per our analysis, the effectiveness of a first year seminar on retention is insignificant. We think that the seminar has a neutral effect on student retention. If there will be multiple modes of communication for seminars, then the students who are commuting will be more interested in attending the seminar.

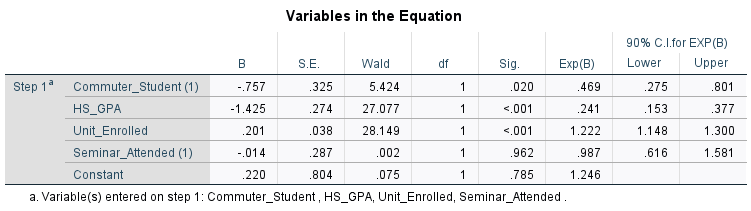
To enhance the accuracy of the model, we think it’s best to eliminate age and gender from the regression analysis. These two variables contribute the least in the model to estimate the student dropout rate.



The Omnibus test is significant for this model and the poor-fit test i.e., Hosmer and Lemeshow test for this model is insignificant having the significance value greater than alpha. Therefore, it is proved that this model is good.



After eliminating the age and gender in our analysis, we acquired an accuracy of 79.7% in our model. The variance suggested for this model is from 0.106 to 0.169.



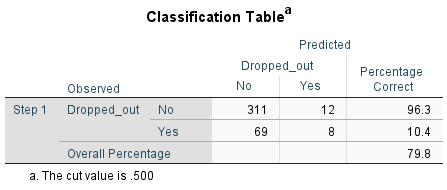
The odds ratio of 0.469 for commuters indicates that students who commute to the campus are more likely to drop out of college. Similarly, the odds ratio of 1.222 for units enrolled implies that students who enrolled in more units are more likely to drop out of college. Additionally, the odds ratio of 0.241 for GPA indicates that students having lower GPAs are more likely to drop out of college.

Logistic regression equation for the best model is:

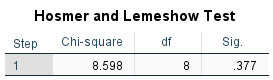
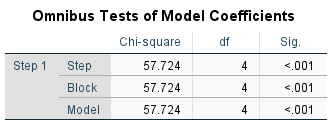
**1/(1 + exp(- (0.220 - 0.014 x1 - 0.757 x2 - 1.425 x3 + 0.201 x4)))**

where x1 represents seminar attended, x2 represents commuters, x3 represents high school GPA, x4 represents units enrolled.

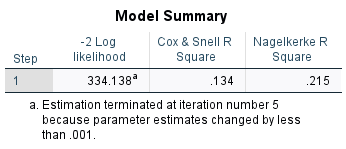
**Simple split- Training dataset:** For the training dataset, we have randomly chosen 400 different cases out of 600, which is 66.67% of the total data.

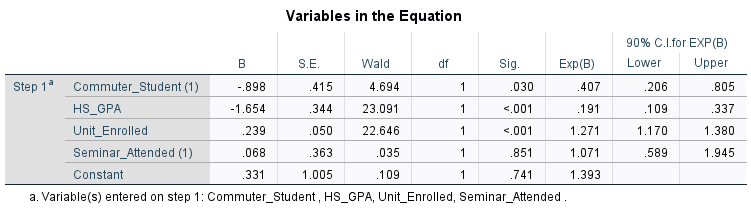


As per the training dataset, the best accuracy we obtained is 79.8%.



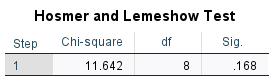
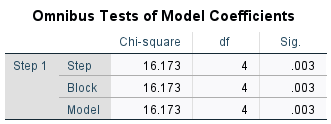
The Omnibus test for this model is significant. And the Hosmer and Lemeshow tests are insignificant which means the model is a good fit model. The variance suggested for this model is from 0.134 to 0.215.

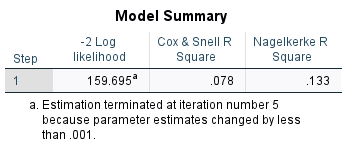




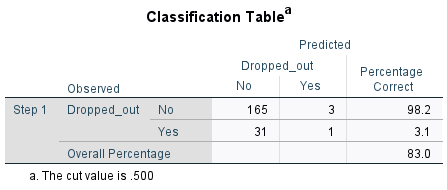
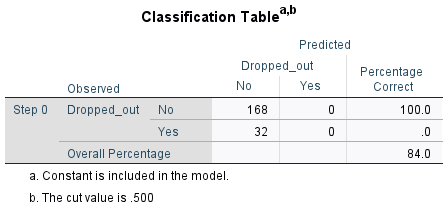
Being the odds ratio of 0.407 for commuters implies that students who commute more are less likely to drop out of college. Added to this, the more units the students will be enrolled in, the more likely will be the students to decide to drop out of the college. GPA having a 0.191 odds ratio indicates that students having low GPA are more likely to decide to drop out of college. And Units Enrolled has the Odds Ratio of 1.271 which indicates that more will be the more units enrolled by the students more will be the dropout rate.

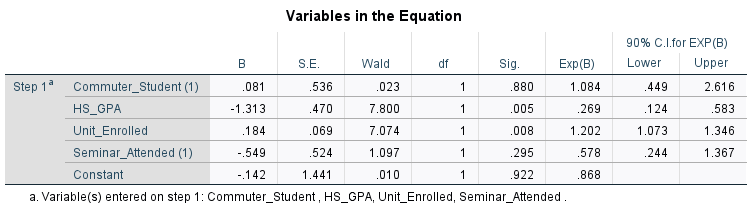
**Simple split- Validation Set:** For the testing dataset, we have randomly chosen 200 cases out of 600, which makes 33.33% of the total data.





The Omnibus test for this model is significant and the Hosmer and Lemeshow test i.e the poor-fit test being insignificant. Hence we concluded that the model is a good-fit model, with the variance suggested between 0.078 and 0.133. Here, the accuracy has decreased from 84% to 83%.



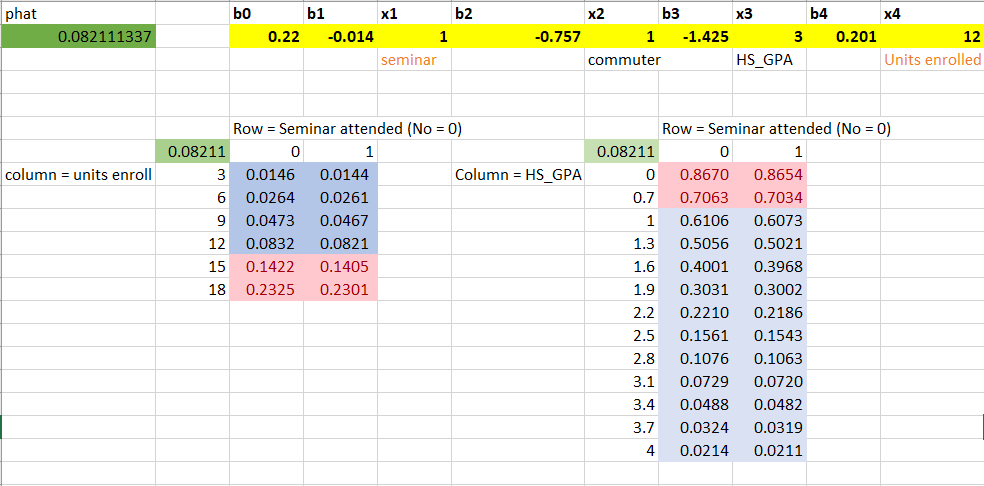


According to our testing dataset, we obtained the odds ratio for units enrolled is 1.202. This signifies that students who enrolled in more units are more likely to drop out of college. 0.269 is the odds ratio for GPA reveals that students having lower GPA are more likely to drop out of college. Hence, we consider the model without age and gender as the best model with the best results.

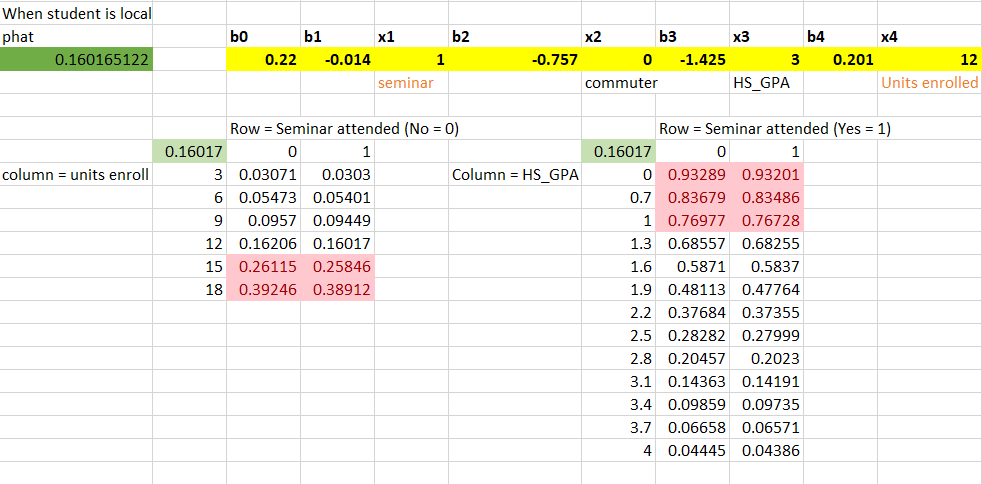
**Testing the model using a two-input data table:**

We are now testing whether the seminar will have a positive effect or not on student retention.

In this two input data table, we are considering students who commute. In the first table, we conclude that students who are not attending the seminar and who are enrolled in more units are more likely to drop out. Added to this, students who didn’t attend the seminar and have low high school GPA have a high probability of more than 70% drop out.



In the second table, we are considering local students. We are also considering the same three parameters as commuters. Students who don’t commute to the campus have more probability of dropping out than commuters.



**Conclusion and Recommendations:**

As per the analysis, students who are enrolled in more units are more likely to attend the seminar irrespective of whether the student is a commuter or not. The same would apply to those who are enrolled in fewer units.

By having the seminar online, the commuting cost for students will be reduced as well as the cost of conducting the seminar will be reduced. Students will be more likely to attend the seminar, hence, increasing the strength of the attendees for the seminar. It recommended other variables to the regression model such as Finance, Marital status, Kids, Language, College ranking, College tuition fees, Race, and Parents' highest education.

Keeping an attendance record of students for seminars will motivate students to attend. The college should make sure that students can enroll in limited units unless they have good grades.

Also, the college could provide transportation to the students who are commuting to the campus which will, in turn, reduce student retention. We recommend the school continue its funding for conducting the seminars.