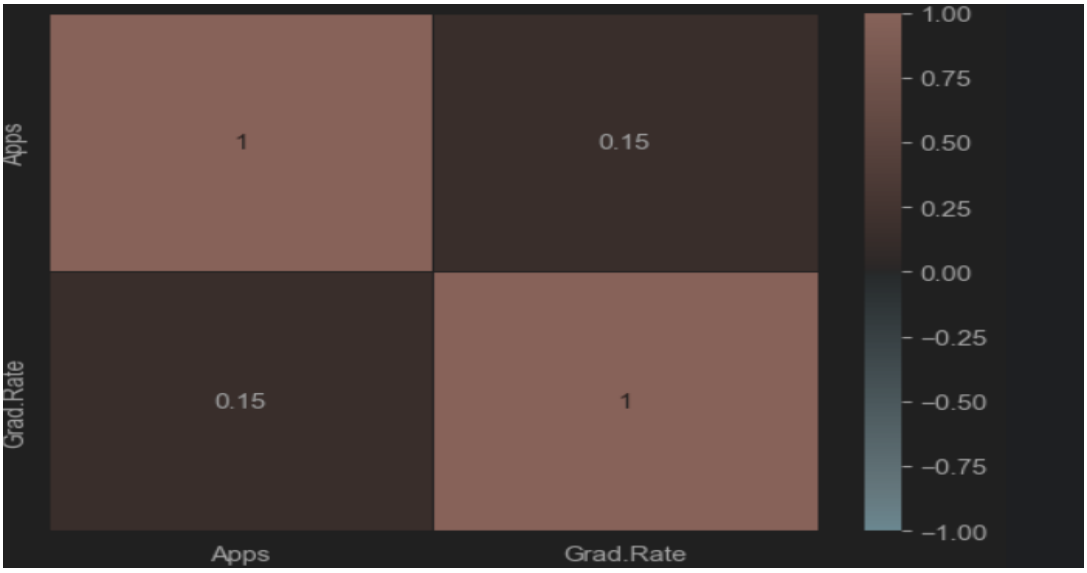
UMKC SPRING SEMSETER

Course: Statistical Learning

Professor: Adu

1.Perform data analysis on the dataset with Graduation rate as the response. Comment on the output. Did you find any patterns?

Graduation rate vs Number of Applications Received

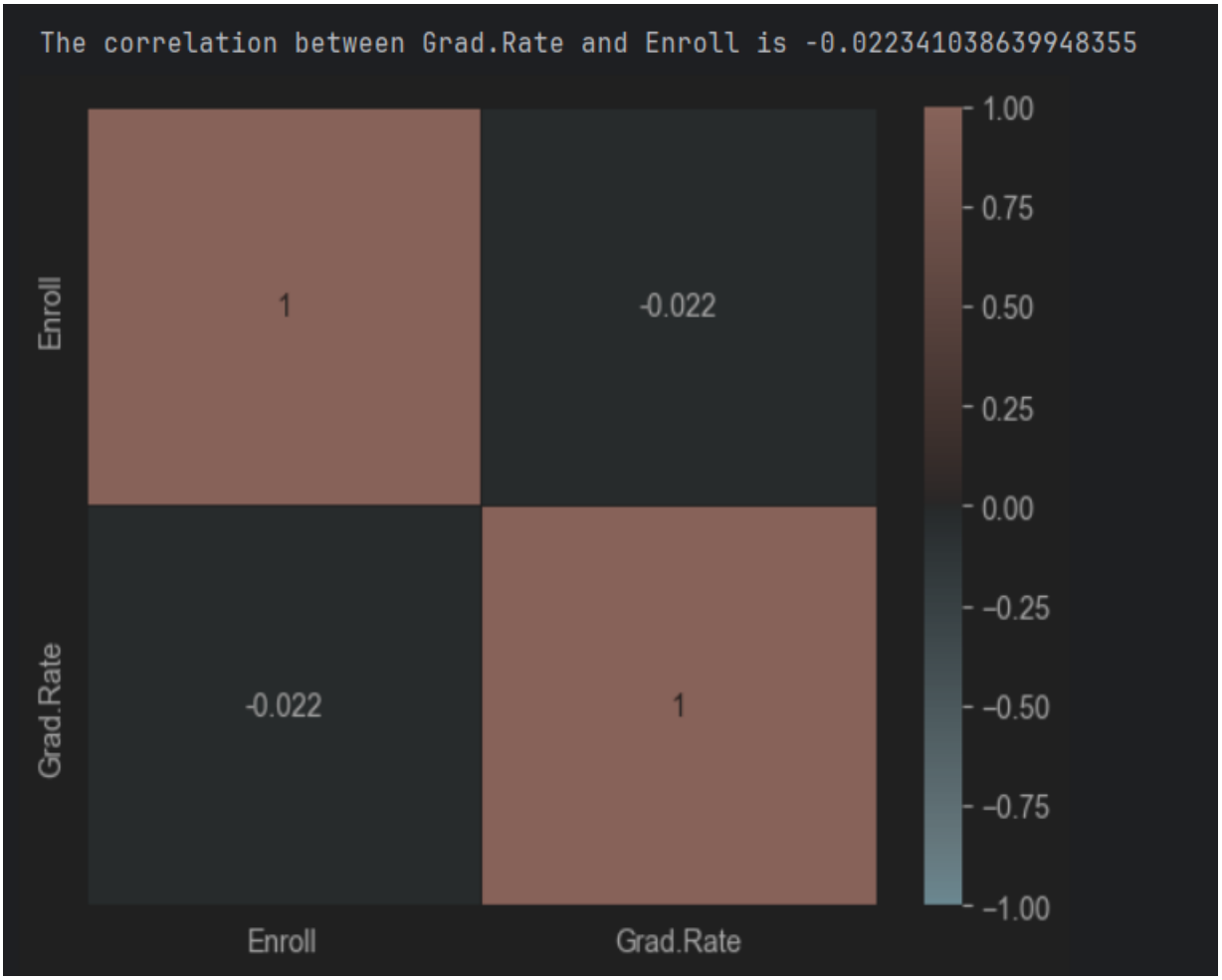


The correlation between the graduation rate and the number of applications received is 0.15, indicating a weak relationship. This means that as the graduation rate increases, the number of applicants slightly increases; however, the correlation does not strongly support predictive conclusions.

Graduation Rate Vs Number of Applications Accepted

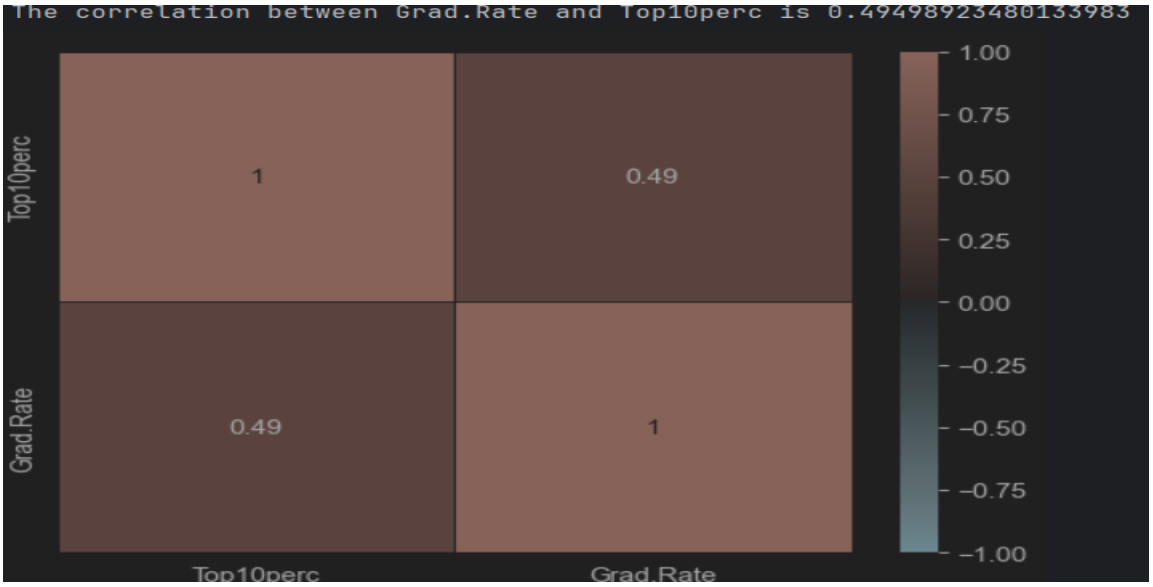
The correlation coefficient between graduation rates and the number of applications accepted is 0.067. Since this value is close to zero, it indicates that graduation rates have little to no impact on the number of applications accepted.

Graduation Rates vs Number of Students Enrolled.



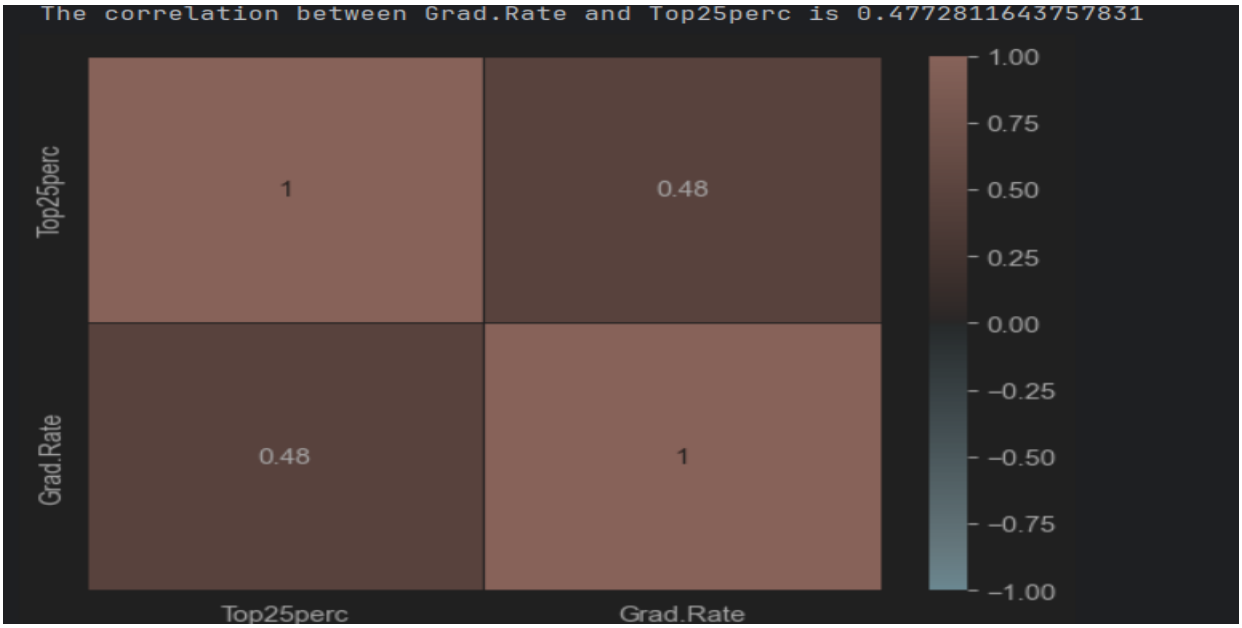
The correlation coefficient between graduation rates and the number of students enrolled is -0.027. This is a weak correlation, and since the value is close to zero, it indicates that graduation rates have little to no impact on the number of students enrolled.

Graduation Rates vs Top 10Percent of New Students from top 10% of H.S.class.



The correlation between the graduation rate and the percentage of new students from the top 10% of their high school class is 0.49, indicating a moderate positive relationship. This means that as the number of students from top high schools increases in colleges, graduation rates tend to increase. However, the correlation only moderately supports predictive conclusions.

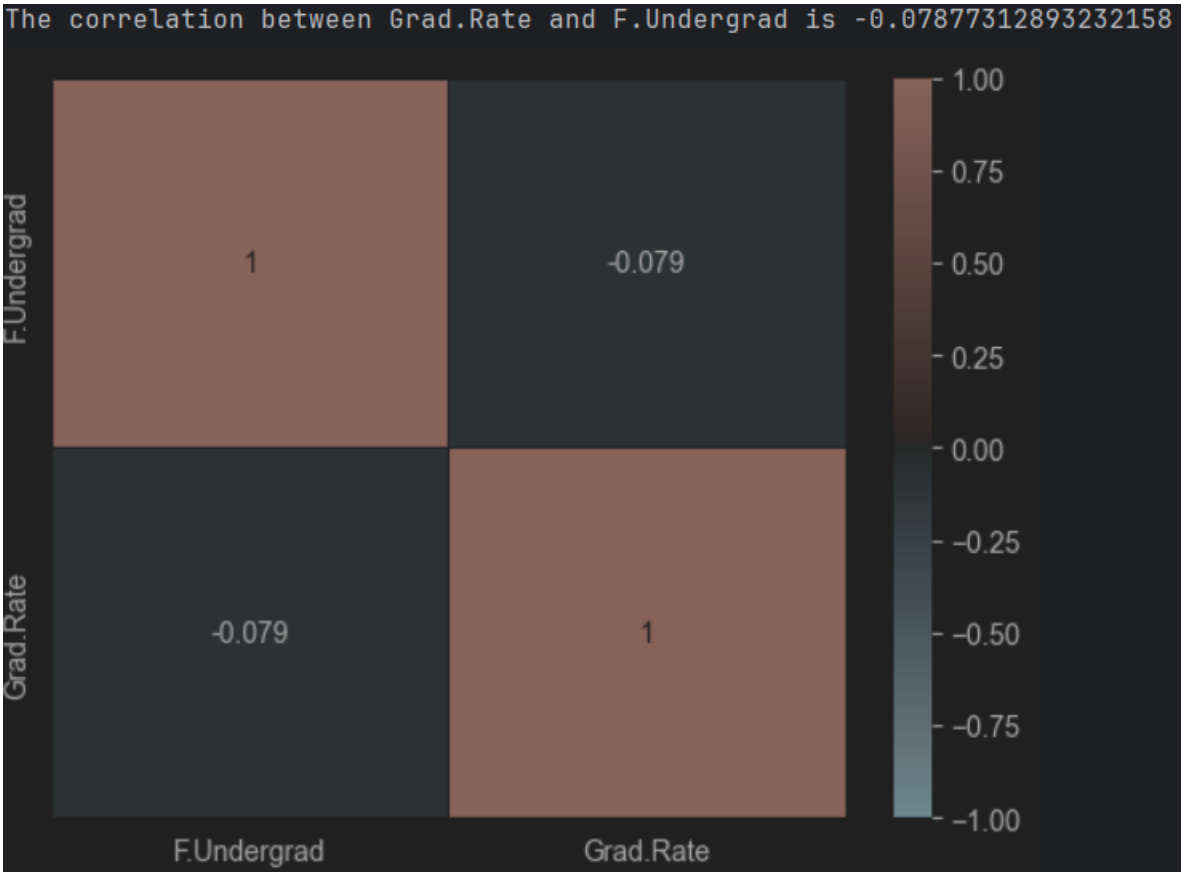
Graduation Rate vs Pct. new students from top 25% of H.S. class



The correlation between the graduation rate and the percentage of new students from the top 25% of high school class is 0.48, indicating a moderate positive relationship. This means that as

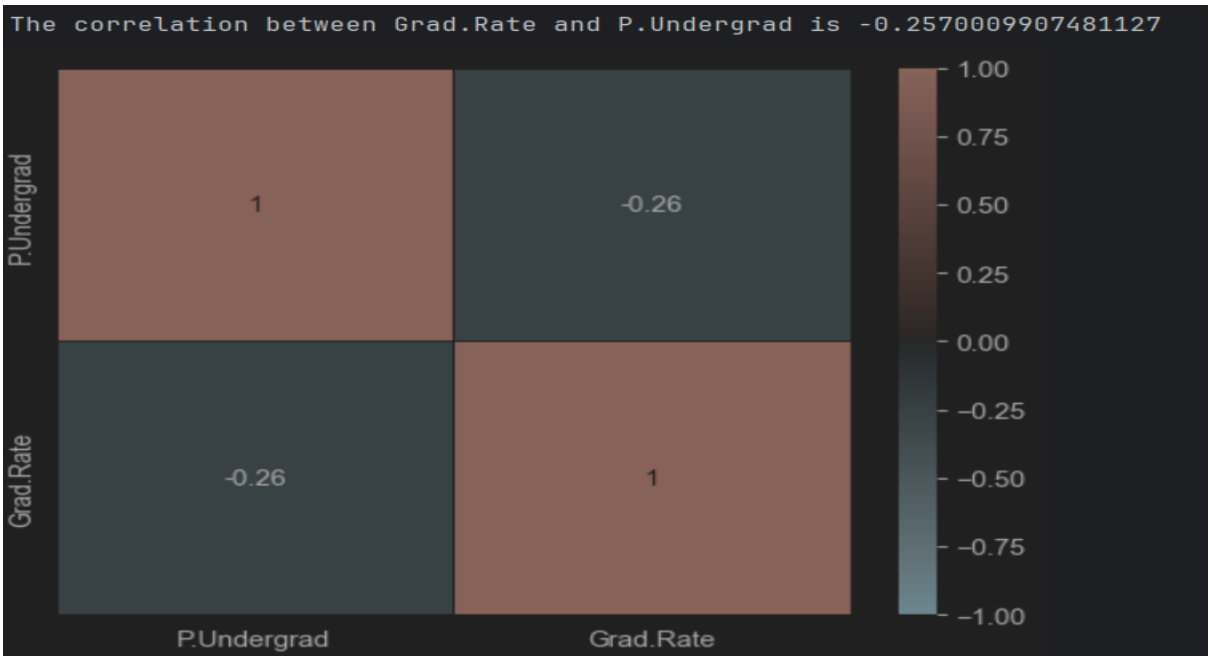
the number of students from top high schools increases in colleges, graduation rates tend to increase. However, the correlation only moderately supports predictive conclusions.

Graduation Rate vs Number of full-time undergraduates



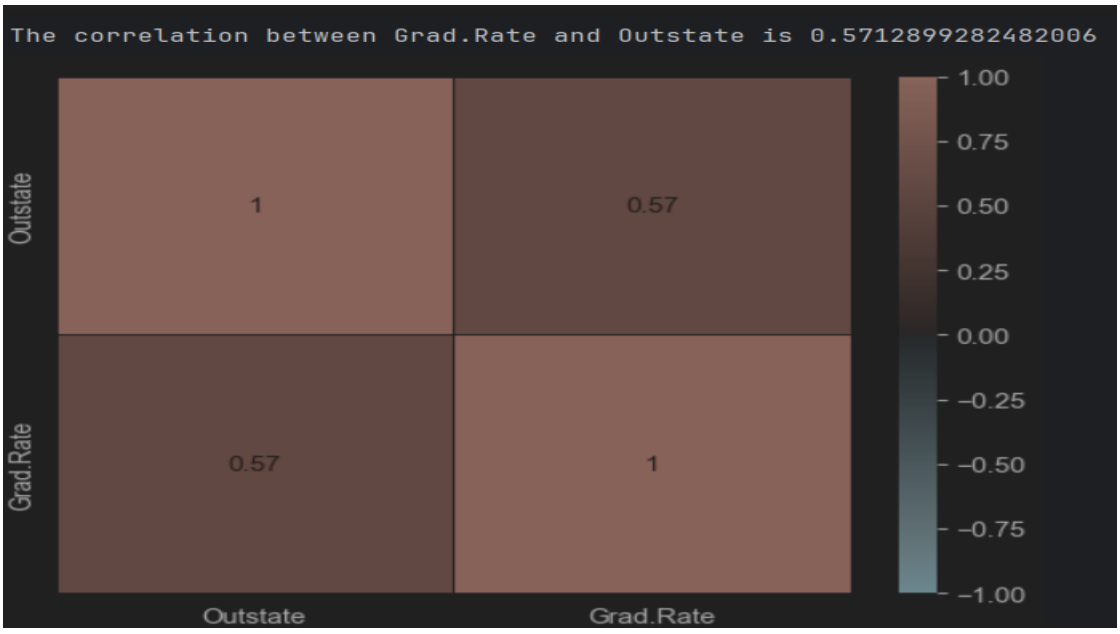
The correlation coefficient between graduation rates and the number of full-time undergraduates is -0.079. This is a weak correlation, and since the value is close to zero, it indicates that graduation rates have little to no impact on the number of full-time undergraduates enrolled in colleges, and vice versa.

Graduation Rates and Number of Parttime Undergraduates.



The correlation coefficient between graduation rates and the number of part-time undergraduates is -0.26. This is a weak negative correlation. It indicates that the graduation rates tend to decrease as the number of part-time undergraduates enrolled in college increases. However, the relationship is not strong to draw the conclusion.

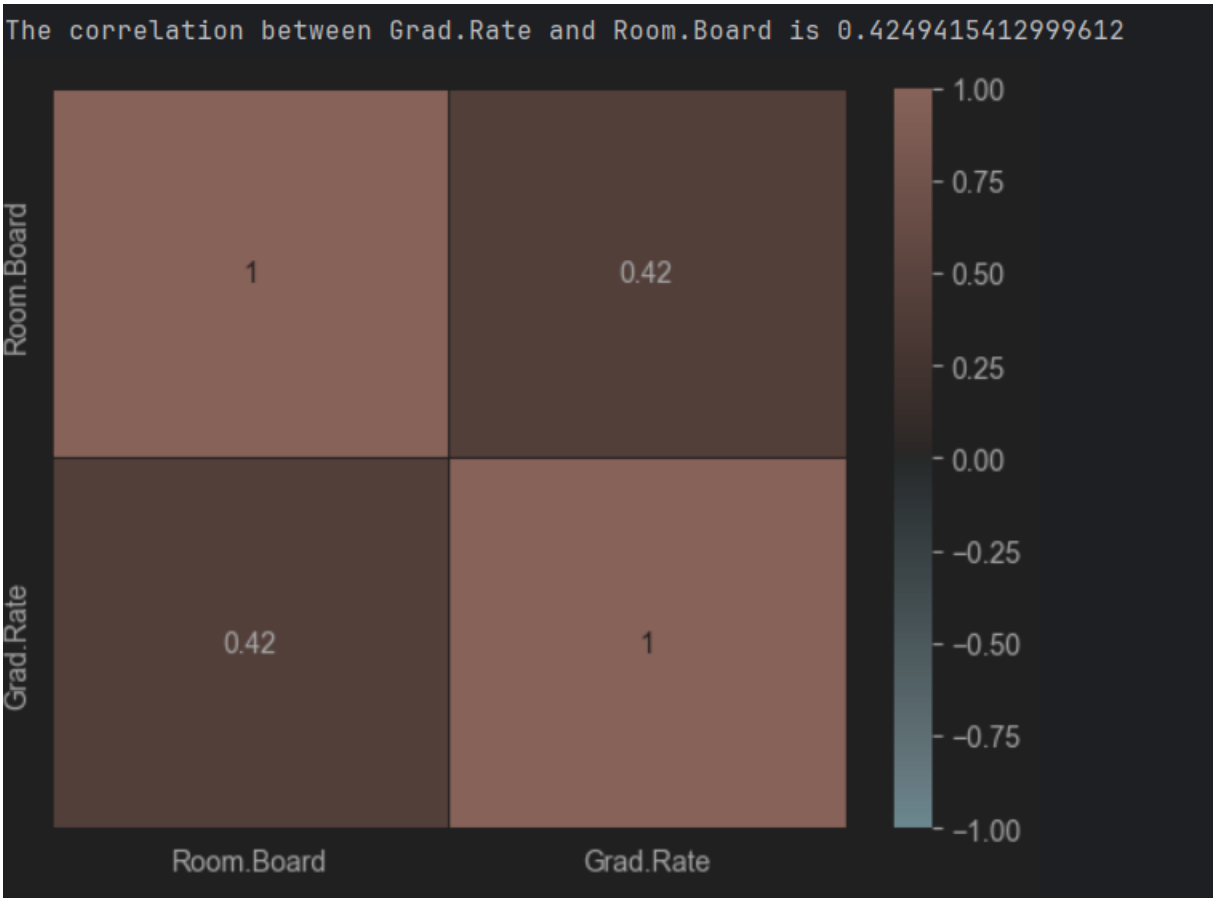
Graduation Rates and Out-of-state tuition



The correlation coefficient between graduation rates and out-of-state tuition is 0.57, this is a moderate positive relationship between out of state tuition and graduation rates. This means as

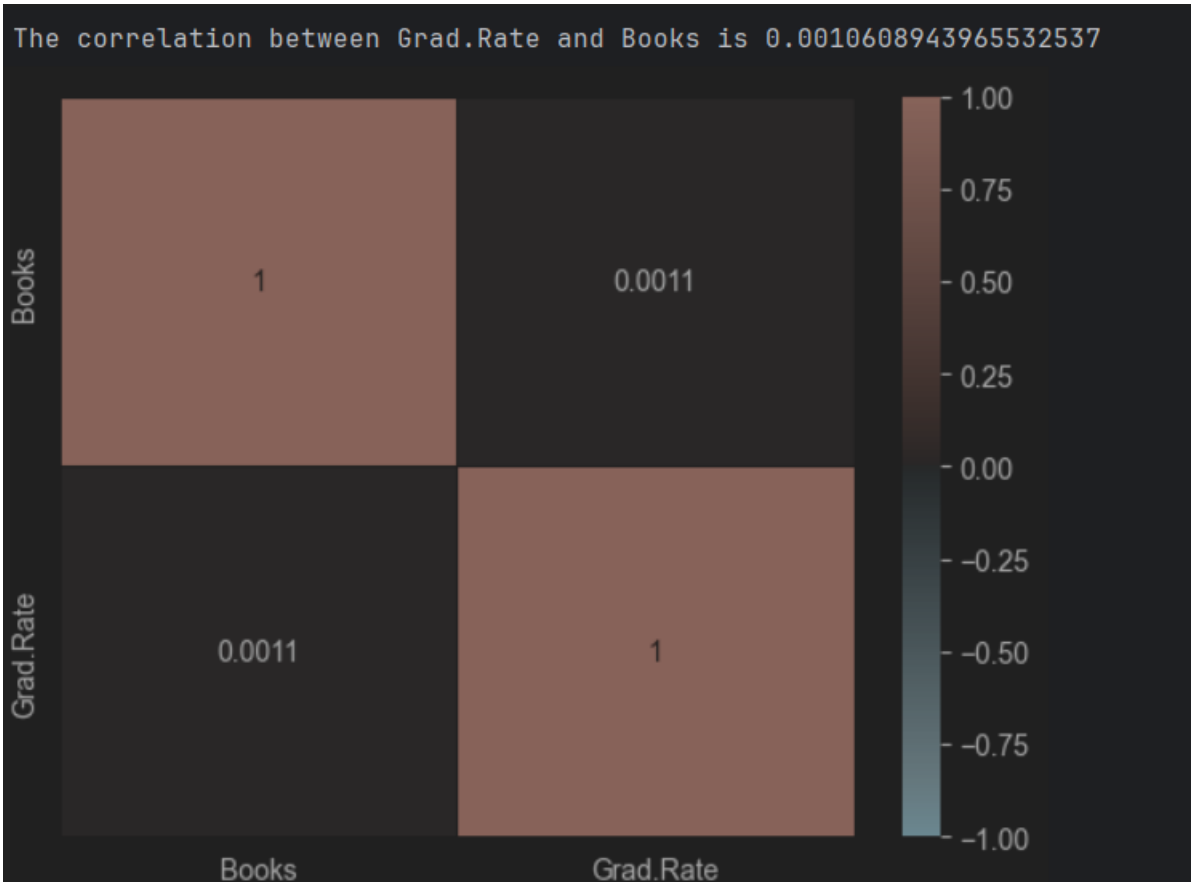
out-of-state tuition increases the graduation rates tends to increase, and vice versa. However, the relationship is not strong.

Graduation Rates vs Room and Board Costs.



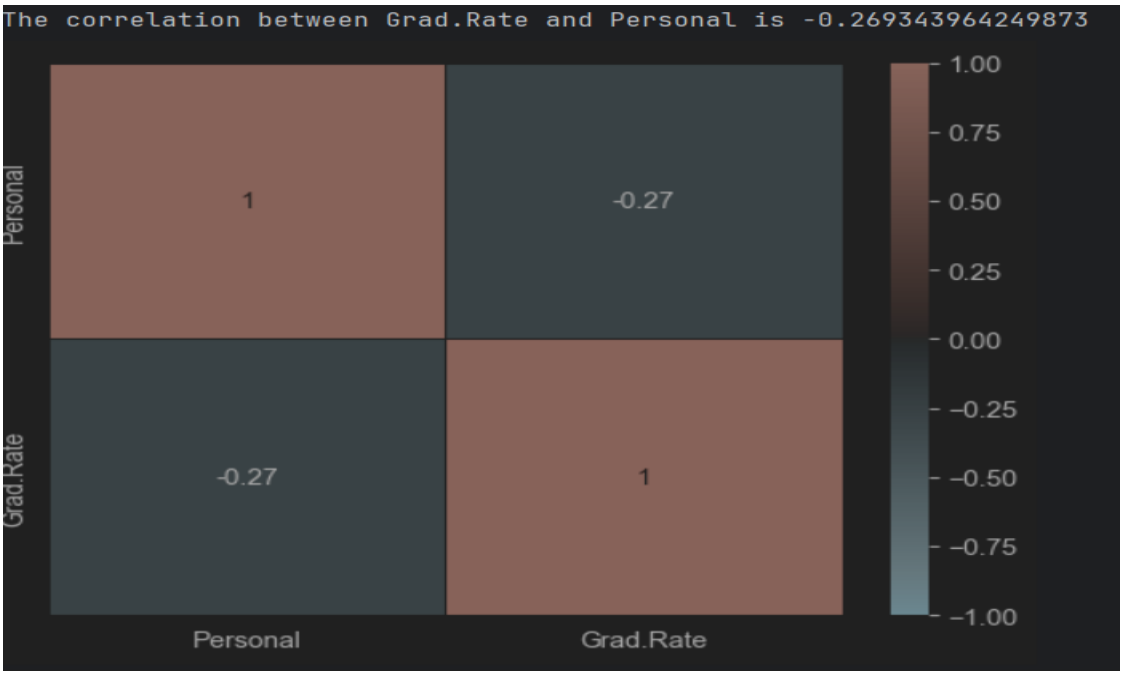
The correlation coefficient between graduation rates vs room and board cost is 0.42. This is a moderate positive relationship between room and board cost and graduation rates. This means as room and board costs increases the graduation rates tend to increase, and vice versa. However, the relationship is not strong

Graduation VS Estimated Book Cost



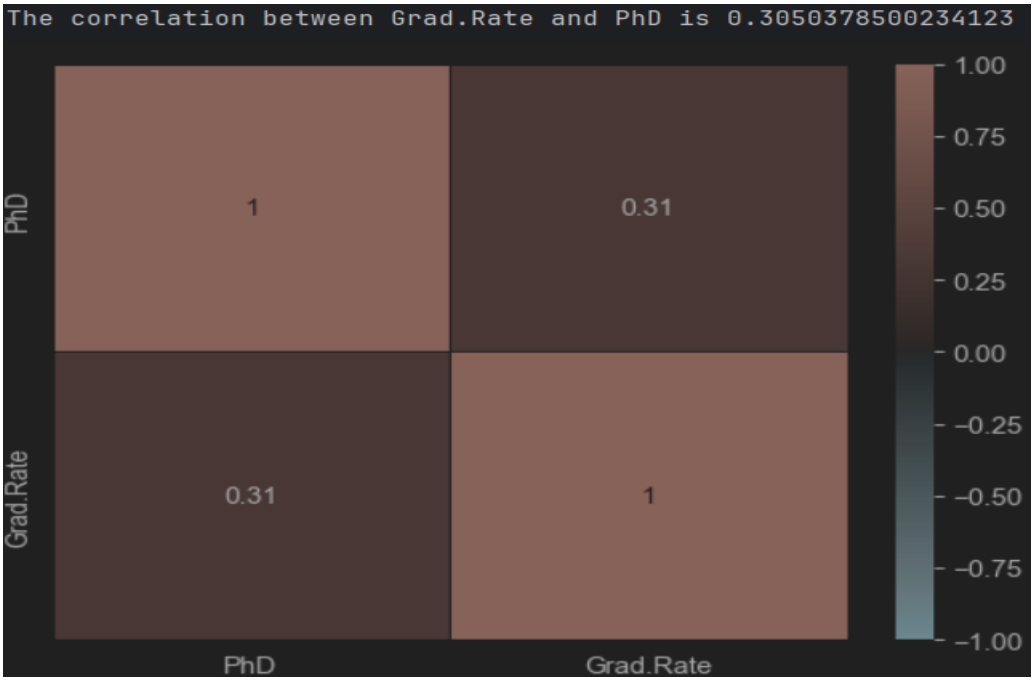
The correlation coefficient between graduation rates vs room and estimated cost of books is 0.001. This is an extremely weak positive relationship between the cost of books and graduation rates. This means there's no relationship to conclude that the cost of books affects the graduation rates.

Graduation Rates Vs Estimated Personal Spending.



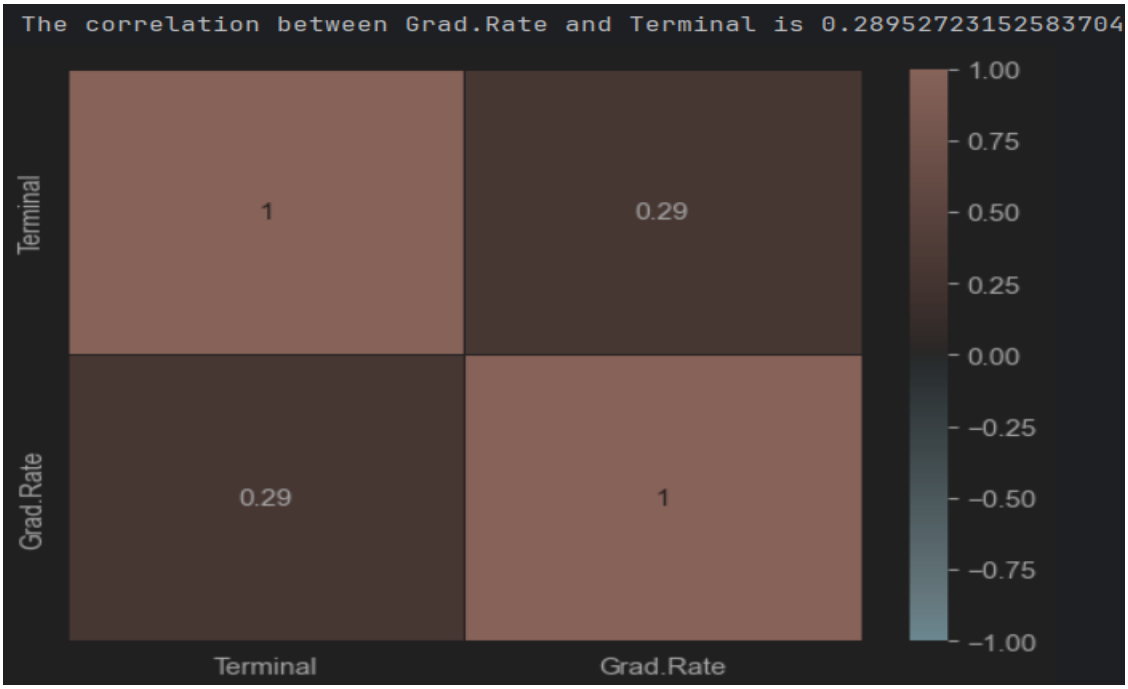
The correlation coefficient between graduation rates and estimated personal spending is -0.27. This is a weak negative correlation. It indicates that as personal spending increases among students, graduation rate tend to decrease. However, the relationship is not strong to draw the conclusion

Graduation Rate vs Percentage of faculty with PH.D's



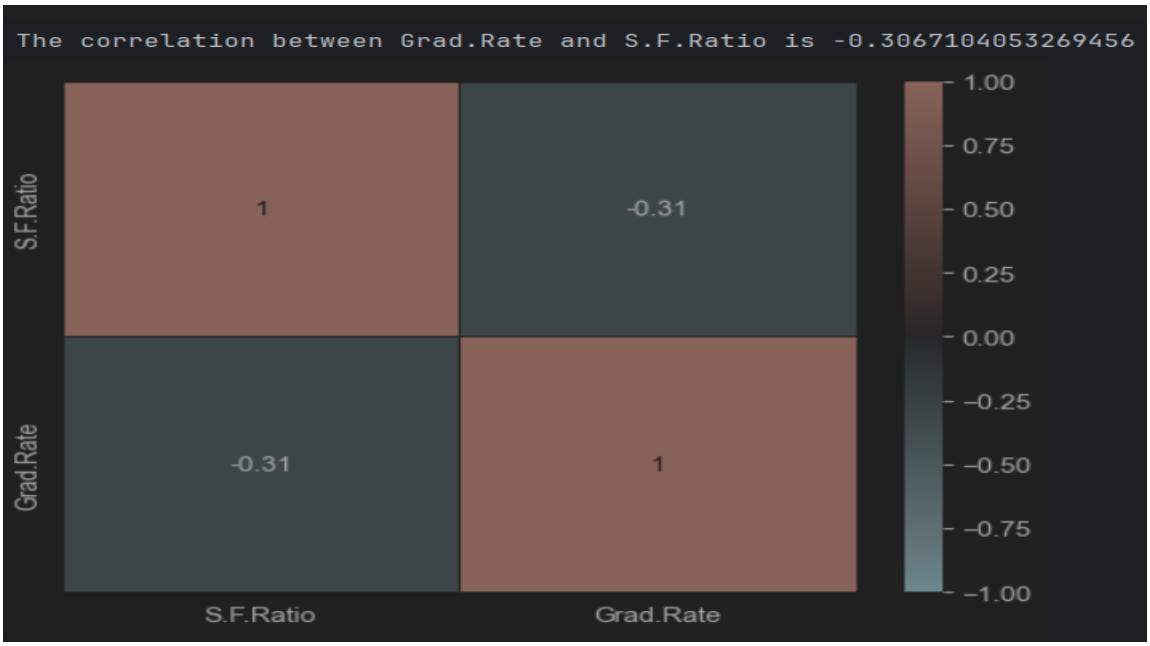
The correlation between the graduation rate and the percentage of faculty with a PHD degree is 0.31. This represents a weak positive correlation, meaning that graduation rates tend to slightly increase as the percentage of faculty with a Ph.D. increases. However, we cannot conclude that one causes the other.

GRADUATION RATE VS PERCENT OF FACULTY WITH TERMINAL DEGREE



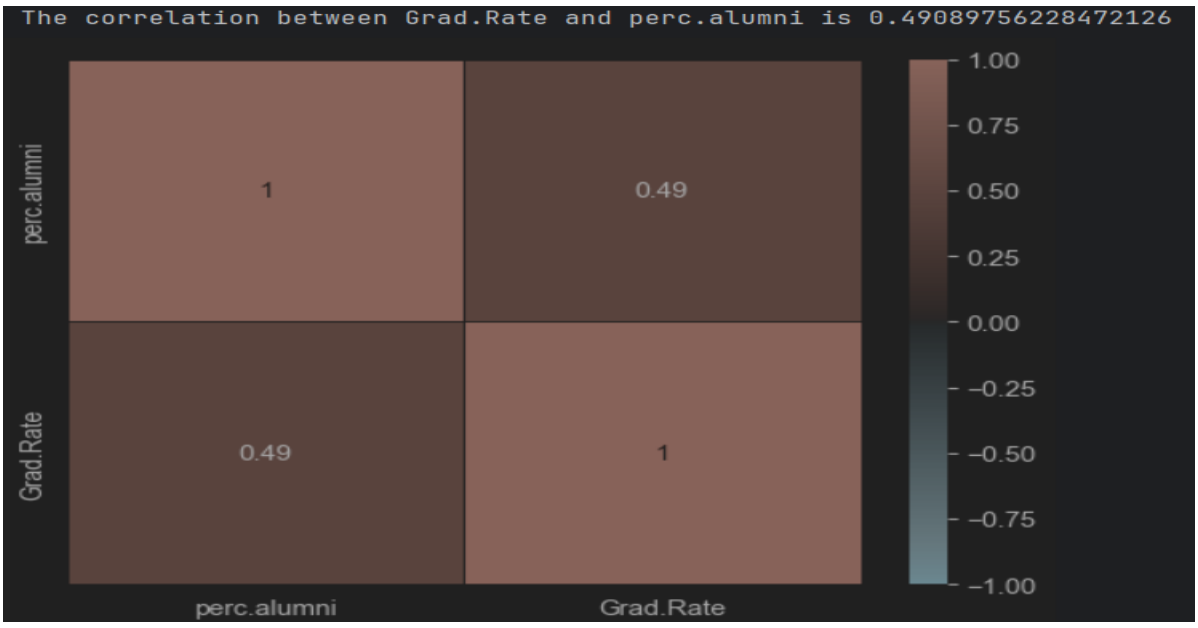
The correlation between the graduation rate and the percentage of faculty with a terminal degree is 0.29. This represents a weak positive correlation, meaning that graduation rates tend to slightly increase as the percentage of faculty with a terminal degree increases. However, we cannot conclude that one causes the other

GRADUATION RATE VS STUDENT/FACULTY RATIO



The correlation between the graduation rate and student/faculty ratio is -0.31. This represents a weak negative correlation, meaning that graduation rates tend to slightly decrease as the ratio of student to faculty increases. However, we cannot conclude that one causes the other.

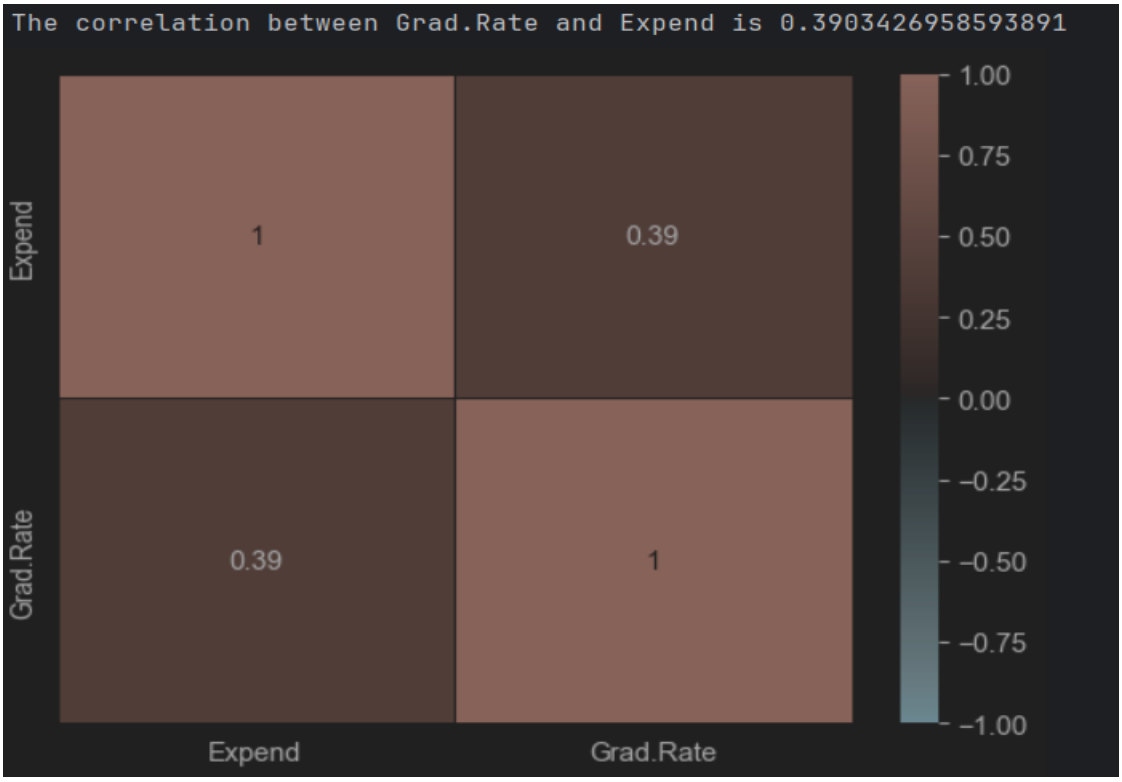
Graduation Rates vs Percent of Alumni who donate



The correlation coefficient between graduation rates and the number of alumni who donate is 0.49, this is a moderate positive relationship between the number of alumni who donate and

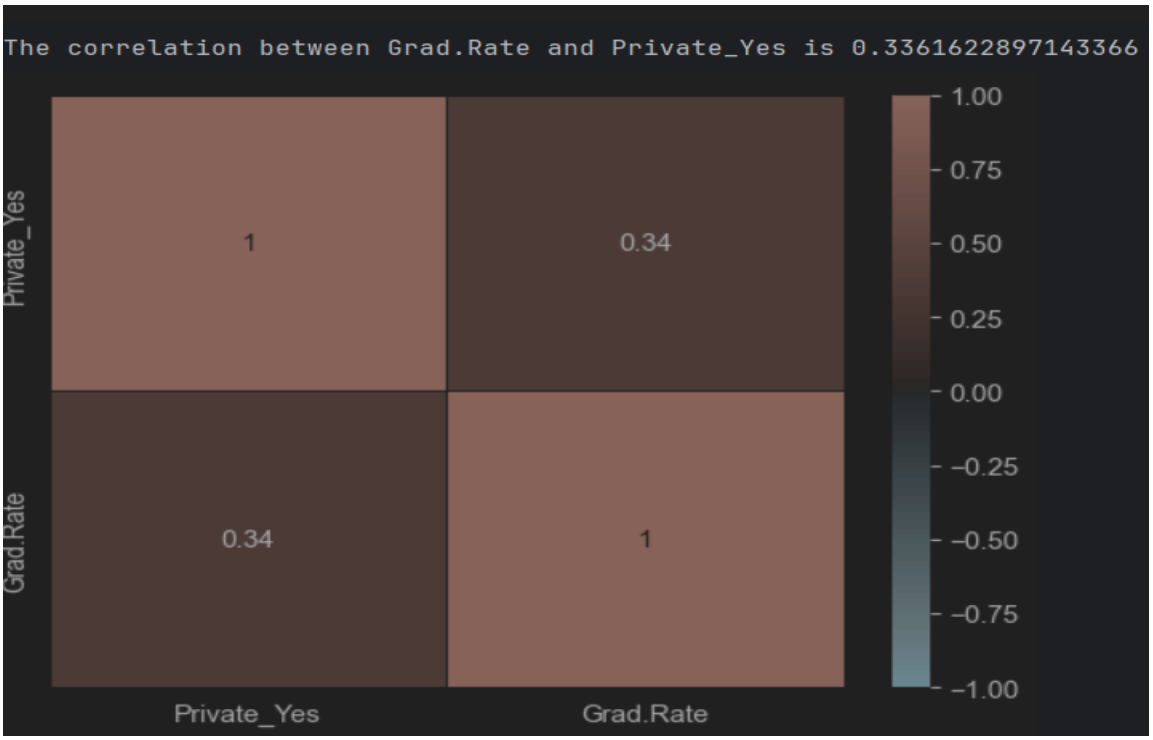
graduation rates. This means as the number of alumni who donate increases the graduation rates tend to increase, and vice versa. However, the relationship is not strong

Graduation Rates vs Instruction Expenditure Per Student:



The correlation coefficient between graduation rates and instructional expenditure per student is 0.39. This indicates a weak to moderate positive relationship between instructional expenditure per student and graduation rates. This means that as a school spends more on instructional materials for students, graduation rates tend to increase, and vice versa. However, the relationship is not strong.

Graduation vs Private University



The correlation coefficient between graduation rates and private school is 0.34. This indicates a weak to moderate positive relationship between private university and graduation rates. This means, on average private schools tends to have high graduation rates. However, the relationship is not strong.

#3. A Is there a relationship between the predictors and the response? Use the ***p-values*** or ***t-statistic*** .

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0064	0.034	-0.190	0.849	-0.073	0.060
Apps	0.1443	0.057	2.530	0.012	0.032	0.256
Accept	-0.0109	0.072	-0.152	0.879	-0.152	0.130
Enroll	0.0229	0.077	0.298	0.766	-0.128	0.174
Top10perc	0.0853	0.065	1.317	0.188	-0.042	0.213
Top25perc	0.1243	0.069	1.811	0.071	-0.011	0.259

F.Undergrad	-0.0532	0.066	-0.810	0.418	-0.182	0.076
P.Undergrad	-0.0505	0.016	-3.146	0.002	-0.082	-0.019
Outstate	0.2168	0.060	3.634	0.000	0.100	0.334
Room.Board	0.1253	0.040	3.167	0.002	0.048	0.203
Books	-0.0102	0.017	-0.605	0.546	-0.043	0.023
Personal	-0.0427	0.029	-1.464	0.144	-0.100	0.015
PhD	0.1151	0.060	1.913	0.056	-0.003	0.233
Terminal	-0.1034	0.058	-1.768	0.078	-0.218	0.011
S.F.Ratio	0.0203	0.035	0.574	0.566	-0.049	0.090
perc.alumni	0.1977	0.040	4.920	0.000	0.119	0.277
Expend	-0.0778	0.028	-2.794	0.005	-0.132	-0.023
Private_Yes	0.1267	0.076	1.660	0.097	-0.023	0.277

The p-value for number of applications, number of part-time undergraduates, out-of-state tuition, room and board costs, expenditure on instructional services per student, and the percentage of alumni who donate is less than 0.05. This indicates that all these independent variables are statistically significant in predicting the graduation rate of universities.

The t-statistics for the number of applicants, out-of-state tuition, room and board costs, and the percentage of alumni who donate are greater than 2. This indicates that as these independent variables increase, graduation rates also increase. On the other hand, the t-statistics for instructional expenditure and the number of part-time undergraduate students are less than -2. This suggests that as instructional expenditure per student increases and the number of part-time undergraduates rises, graduation rates decrease. All predictors with t-statistics greater than 2 or less than -2 are statistically significant in predicting the graduation rate of universities.

#3. B **What does the coefficient of each predictor suggest**?

The coefficient of the number of applications received is 0.1443. A one-unit increase in the number of applications leads to a 0.1443 increase in the graduation rate.

The coefficient of applications accepted is -0.0109. A one -unit increase in the number of applications accepted leads to 0.0109 decrease in the graduation rate.

The coefficient number of new students enrolled is 0.0229. This means a one unit increase in the number of new students enrolled leads to a 0.0229 increase in the graduation rate.

The coefficient of the percentage of new students from the top 10% of their high school class is 0.0853. This means that a one-unit increase leads to a 0.0853 increase in the graduation rate.

The coefficient of the percentage of new students from the top 25% of their high school class is 0.1243. This means that a one-unit increase leads to a 0.1243 increase in the graduation rate

The coefficient of the number of full-time undergraduates is -0.0532. This means that a one-unit increase leads to a 0.0532 decrease in the graduation rate.

The coefficient of the number of part-time undergraduates is -0.0505. This means that a one-unit increase leads to a 0.0505 decrease in the graduation rate

The coefficient of out of state tuition is 0.2168. This means that a one-unit increase leads to a 0.2168 increase in the graduation rate

The coefficient of room and board costs is 0.1253. This means that a one-unit increase leads to a 0.1253 increase in the graduation rate

The coefficient of estimated book costs is -0.0102. This means that a one-unit increase leads to a 0.0102 decrease in the graduation rate

The coefficient of estimated personal spending is -0.0427. This means that a one-unit increase leads to a 0.0427 decrease in the graduation rate

The coefficient of percentage of faculty with PH.D.s is 0.1151. This means that a one-unit increase leads to a 0.1151 increase in the graduation rate

The coefficient of percentage of faculty with terminal degree is -0.1034. This means that a one-unit increase leads to a 0.1034 decrease in the graduation rate

The coefficient of student/faculty ratio is 0.0203. This means that a one-unit increase leads to a 0.0203 increase in the graduation rate

The coefficient percentage of alumni who donate is 0.1977. This means that a one-unit increase leads to a 0.1977 increase in the graduation rate

The coefficient of instructional expenditure per student is -0.0778. This means that a one-unit increase leads to a 0.0778 decrease in the graduation rate

The coefficient of a college being private is 0.1267. This means that a one-unit increase leads to a 0.1267 increase in the graduation rate

#3. C **How well can the model explain the relationship between the predictors and the response variable**?

Out of 17 predictors, 6 were statistically significant ($p < 0.05$), and 4 were marginally significant ($0.05 < p < 0.1$). This suggests that the model captured meaningful relationships for more than 50% of the predictors.

#4. B **Fit Ridge regression model comment on the coefficient of each predictor**

```
Apps           0.129786
Accept         0.068968
Enroll        -0.003465
Top10perc      0.111925
Top25perc      0.120724
F.Undergrad   -0.041170
P.Undergrad   -0.117412
Outstate       0.168236
Room.Board     0.125845
Books         -0.018280
Personal      -0.058227
```

```
PhD             0.062160
Terminal       -0.037157
S.F.Ratio       0.016328
perc.alumni     0.167846
Expend         -0.057157
Private_Yes     0.086754
```

The coefficient of the number of applications received is 0.129786. A one-unit increase in the number of applications leads to a 0.129786 increase in the graduation rate.

The coefficient of applications accepted is 0.068968. A one -unit increase in the number of applications accepted leads to 0.068968 increase in the graduation rate.

The coefficient number of new students enrolled is - 0.003465. This means a one unit increase in the number of new students enrolled leads to a 0.003465 decrease in the graduation rate.

The coefficient of the percentage of new students from the top 10% of their high school class is 0.112. This means that a one-unit increase leads to a 0.112 increase in the graduation rate.

The coefficient of the percentage of new students from the top 25% of their high school class is 0.1243. This means that a one-unit increase leads to a 0.1243 increase in the graduation rate

The coefficient of the number of full-time undergraduates is -0.04112. This means that a one-unit increase leads to a 0.04112decrease in the graduation rate.

The coefficient of the number of part-time undergraduates is -0.117. This means that a one-unit increase leads to a 0.117 decrease in the graduation rate

The coefficient of out of state tuition is 0.168. This means that a one-unit increase leads to a 0.168 increase in the graduation rate

The coefficient of room and board costs is 0.126. This means that a one-unit increase leads to a 0.126 increase in the graduation rate

The coefficient of estimated book costs is -0.018. This means that a one-unit increase leads to a 0.018 decrease in the graduation rate

The coefficient of estimated personal spending is -0.058. This means that a one-unit increase leads to a 0.058 decrease in the graduation rate

The coefficient of percentage of faculty with PH.D.s is 0.062. This means that a one-unit increase leads to a 0.062 increase in the graduation rate

The coefficient of percentage of faculty with terminal degree is -0.037. This means that a one-unit increase leads to a 0.037 decrease in the graduation rate

The coefficient of student/faculty ratio is 0.0163. This means that a one-unit increase leads to a 0.0163 increase in the graduation rate

The coefficient percentage of alumni who donate is 0.168. This means that a one-unit increase leads to a 0.168 increase in the graduation rate

The coefficient of instructional expenditure per student is -0.057. This means that a one-unit increase leads to a 0.057 decrease in the graduation rate

The coefficient of a college being private is 0.086. This means that a one-unit increase leads to a 0.086 increase in the graduation rate

No 4 part C. **How well can the model explain the relationship between the predictors and the response variable**?

```
115.5064850041579
R squared is 0.488
```

The r-squared of Ridge model is 0.488. This shows the model explains 48.8 % the variation of the dependent variable (graduation rate) based on the features in the model.

5 B **What does the coefficient of each predictor suggest**?

Apps	0.17068942552882832
Accept	0.0
Enroll	-0.0
Top10perc	0.07918037031922491
Top25perc	0.1327034839967485
F.Undergrad	-0.0
P.Undergrad	-0.12943072395622177
Outstate	0.20817348123374543
Room.Board	0.11672940650482748
Books	-0.0053172084049236705
Personal	-0.05003398588082524
PhD	0.008111432656013218
Terminal	-0.0
S.F.Ratio	0.0
perc.alumni	0.1782982754856687
Expend	-0.05952105895015908
Private_Yes	0.062110884776370584

The coefficient of the number of applications received is 0.17068. A one-unit increase in the number of applications leads to a 0.17068 increase in the graduation rate.

The coefficient of applications accepted is 0.0. A one -unit increase in the number of applications accepted leads to 0.0increase in the graduation rate.

The coefficient number of new students enrolled is - 0.0. This means a one unit increase in the number of new students enrolled leads to a 0.00 decrease in the graduation rate.

The coefficient of the percentage of new students from the top 10% of their high school class is 0.0792. This means that a one-unit increase leads to a 0.0792increase in the graduation rate.

The coefficient of the percentage of new students from the top 25% of their high school class is 0.133. This means that a one-unit increase leads to a 0.133 increase in the graduation rate

The coefficient of the number of full-time undergraduates is -0.0. This means that a one-unit increase leads to a 0.0decrease in the graduation rate.

The coefficient of the number of part-time undergraduates is -0.129. This means that a one-unit increase leads to a 0.129 decrease in the graduation rate

The coefficient of out-of-state tuition is 0.208. This means that a one-unit increase leads to a 0.208 increase in the graduation rate

The coefficient of room and board costs is 0.1167. This means that a one-unit increase leads to a 0.1167 increase in the graduation rate

The coefficient of estimated book costs is -0.0053. This means that a one-unit increase leads to a 0.0053 decrease in the graduation rate

The coefficient of estimated personal spending is -0.050. This means that a one-unit increase leads to a 0.050 decrease in the graduation rate

The coefficient of percentage of faculty with PH.D.s is 0.0081. This means that a one-unit increase leads to a 0.0081 increase in the graduation rate

The coefficient percentage of faculty with terminal degrees is -0.00. This means that a one-unit increase leads to a 0.00 decrease in the graduation rate

The coefficient of student/faculty ratio is 0.00. This means that a one-unit increase leads to a 0.00 increase in the graduation rate

The coefficient percentage of alumni who donate is 0.178. This means that a one-unit increase leads to a 0.178 increase in the graduation rate

The coefficient in instructional expenditure per student is -0.059. This means that a one-unit increase leads to a 0.059 decrease in the graduation rate

The coefficient of a college being private is 0.086. This means that a one-unit increase leads to a 0.062 increase in the graduation rate

#5 C **How well can the model explain the relationship between the predictors and the response variable**

```
0.020185
Best R_squared  for lasso model is 0.493
```

The r-squared of lasso model is 0.493. This shows the model explains 49.83% the variation of the dependent variable (graduation rate) based on the features in the model

No 6 A : Which predictors were significant across all three models

The number of applications received, the percent of new students from the top 10% of their high school class, the percent of new students from the top 25% of their high school class, out-of-state

tuition, room and board, the percentage of alumni who donate, and whether the school is private were the most significant predictors across the three models, based on their coefficient results.

No 6B: **Which predictors shrank the most**?

"The number of applicants accepted, number of new students enrolled, number of full-time undergraduates, number of faculty with terminal degrees, and student-to-faculty ratio shrank the most because their coefficients were pulled to zero in the Lasso model.

No 6C**: Which model shrank the predictors the most**?.

The Lasso model shrank the most predictors; out of 17, the coefficients of 5 predictors were pulled to zero.

No 7**: Conclusion**:

The multiple regression had an R-squared of 0.454, Ridge regression had an R-squared of 0.488, and Least Absolute Shrinkage and Selection Operator (Lasso) had an R-squared of 0.493. All three models had an R-squared of less than 50%, meaning they explained less than half of the variation in the dependent variable based on the available predictors. Therefore, none of the models fully captured the variation in college graduation rates. The most significant predictors included the percentage of students from the top 10% of high school class, the percentage from the top 25% of high school class, out-of-state tuition, and the percentage of alumni who donate.

When compared to existing research on college graduation rates, the models' results generally align with findings in higher education studies. For example, according to **Travis Thornton**, factors that greatly contribute to a student's graduation from college include a high GPA from high school rather than ACT scores, cost and fees, location, safety, academic support, and career services**. Ronald Stolberg, PhD**, also argues that academic performance, financial stability, social support, academic support centers, and campus engagement are important factors**. Michael T. Nietzel**, in his article published on Forbes, states that factors such as attending private colleges, race, gender, and geographical location contribute to higher education completion rates.

The models identified factors such as a college being private, the percentage of students from the top 25% and top 10% of high school classes, out-of-state tuition, and the percentage of alumni who donate as strongly influencing high graduation rates. However, the models did not find academic support from institutions or high GPA as significant predictors of high graduation rates, which could be due to limitations in the available data. More focus should be placed on incorporating additional data when training our models to make better predictions and account for other potential predictors

Work citied

Regression Coefficients - GeeksforGeeks

R Squared | Definition, Formula & How to Calculate

What Factors Affect College Graduation Rates? - CLJ

Factors That Lead to Higher Graduation Rates and Less Drop-Out - Psi Chi, The International Honor Society in Psychology

Latest College Completion Rate Remains Flat At 62%