MBA 546 Case Report, Week 7
Topic: Holmes University Retention
Due date: 16 October 2023

Submitted by Samyak Pratap Shah

Holmes University Retention

Executive Summary

At Holmes University, a study was conducted to determine which first-year students might not return for their second year. This understanding is essential, as the plan is to provide these identified students with additional support, potentially increasing retention.

Using data from 500 students, it was found that 80% (or 395 students) returned for their sophomore year, while 20% (or 105 students) did not. Logistic regression, a method suited for yes/no outcomes, was employed for this analysis.

Several variables were assessed for their potential impact on student retention. Here's a snapshot of the findings:

- Athletic Status: Being an athlete or not didn't seem to matter much.
- Gender: Gender also wasn't a significant determinant.
- Financial Aid (FAFSA): Interestingly, students who applied for FAFSA, indicating possible financial need, were more likely to drop out.
- College Affiliation: The type of college a student was enrolled in was crucial. Differences were observed among students from Arts & Sciences, Business, and Education colleges.
- Accommodation: Whether students lived in a dorm or at home didn't significantly sway their decision.
- Academic Performance: Both university and high school GPAs were examined. A higher GPA at the university was associated with higher retention. However, high school GPA didn't make much of a difference.
- Distance from Home: Students who lived further from the university were less likely to return.
- Work Hours: Those who worked more hours weekly were slightly more inclined not to return, although the difference was minimal.

After testing the model's accuracy, it was found that, using a 0.26 probability cutoff, the university can predict with reasonable accuracy which students might not return. Financially, this model's application is a beneficial move for the university. When implemented, the institution stands to gain more than it invests, making this approach both effective and economical. By intervening early, the university can potentially increase its retention rate and provide needed support to its student population.

Introduction

I had a task to identify first-year students at risk of dropping out and not coming back for their second year using information from the school's administration. We want to put these students in a special program where they'll receive extra support and maybe some money to encourage them to stay at the school. I used logistic regression, a type of statistical analysis, to make these predictions.

Data

Model Information		
Data Set	WORK.IMPORT	
Response Variable	Return	Return
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	500
Number of Observations Used	500

Response Profile		
Ordered Value	Return	Total Frequency
1	0	105
2	1	395

Probability modeled is Return='1'.

I used a method called binary logistic regression. This is best for situations where there are only two possible results, like a yes or no answer. Now, diving into the data, a total of 500 student records were analyzed. Here's where it gets interesting: of these students, 105 of them, which is 20%, didn't return for their next year. On the flip side, a significant majority, 395 students or 80%, chose to come back for their sophomore year. It's worth noting that the main goal of this model is to determine the chances of a student not returning, represented by Return=0.

Class Level Information				
Class	**Value**	**Design Variables**		
Athlete	0	1	0	
	1	0	1	
Gender	Female	1	0	
	Male	0	1	
FAFSA	No	1	0	
	Yes	0	1	
College	A&S	1	0	0
	Business	0	1	0
	Education	0	0	1
Accommodations	Dorm	1	0	
	Home	0	1	

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Athlete Classification

The table provides information on how athletes are categorized. When an individual is not classified as an athlete (value is 0), the corresponding design variable for the first column is marked as 1, while the second column remains 0. On the other hand, if an individual is classified as an athlete (value is 1), the first column is marked as 0, and the second is marked as 1.

Gender Distinction

Gender is another category detailed in the table. Specifically, for females, the first design column shows a value of 1 and the second displays a value of 0. In contrast, for males, the first column has a value of 0 while the second column gets a 1.

FAFSA Status

The FAFSA classification indicates a student's financial aid status. When a student hasn't applied for FAFSA (with the value "No"), the design variable is 1 for the first column and 0 for the second. Conversely, if a student has applied (with the value "Yes"), the first column displays a 0 and the second shows a 1.
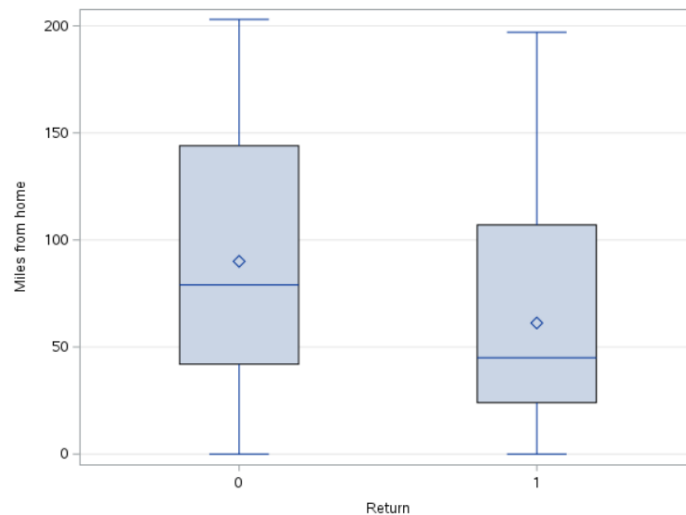
College Affiliation

College categorization is broken down into three parts - Arts & Sciences (A&S), Business, and Education. For students in A&S, the first design column is marked as 1, with the subsequent two columns marked as 0. Business students have the first column as 0, the second as 1, and the third as 0. Lastly, those in Education have the first two columns marked as 0 and the third column marked as 1.

Accommodation Type

The table also gives insights into where students are accommodated. For those living in a dorm, the design variable for the first column is 1 and the second column is 0. However, for students living at home, the first column remains 0 while the second is marked as 1.
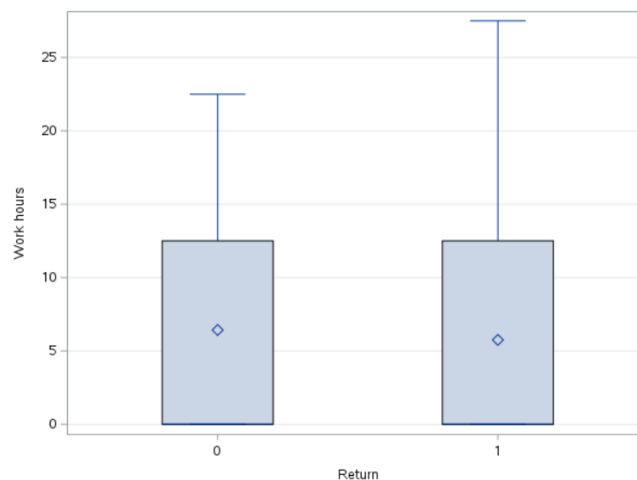
The model convergence status indicates the model's convergence status. With the criterion (GCONV=1E-8) met, it signifies that the statistical model has achieved its desired solution based on the provided threshold.

This above box plot presents the distribution of the "Miles from Home" variable for students who did not return (0) versus those who did return (1).
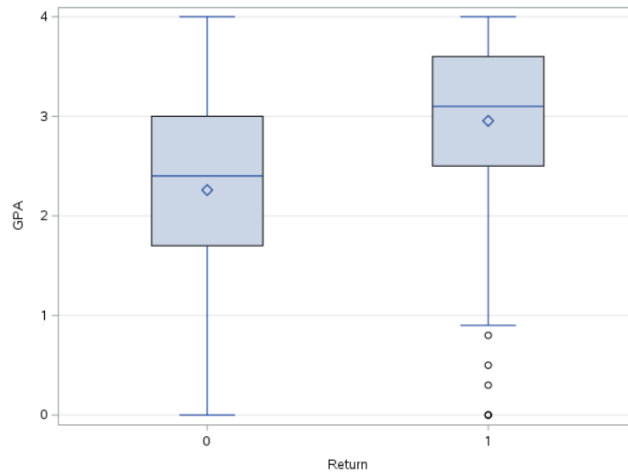
- For students who did not return (0), the median distance from home (represented by the diamond symbol) seems higher than for those who did return.
- The box represents the interquartile range, where the majority of data points lie. The whiskers (vertical lines extending from the box) show the range of the data.

Students who didn't return to Holmes University lived, on average, further away from the university than those who did return.
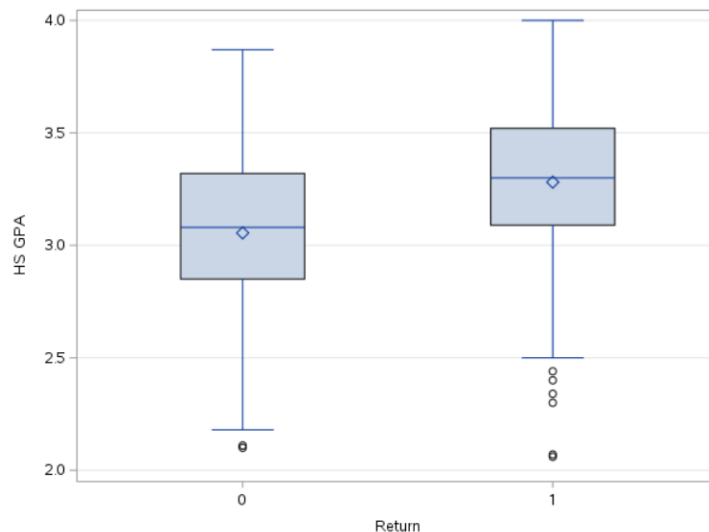

Students who worked more hours were slightly more likely not to return compared to those who worked fewer hours, though the difference isn't very large.

- This plot displays the distribution of the number of work hours per week for students who didn't return (0) versus those who did (1).
- The median work hours for both groups (represented by the diamond symbols) seem quite close, though slightly higher for those who didn't return.

Students with higher GPAs were more likely to return to the university than those with lower GPAs. Some students, however, with exceptionally high or low GPAs are exceptions. The above chart depicts the distribution of the GPA for students who didn't return (0) compared to those who did (1).

- The diamond symbols, representing the median GPA, show that students who returned had a slightly higher GPA compared to those who didn't.
- The circles below and above the boxes are outliers – students whose GPA is notably lower or higher than the majority.

This above plot presents the distribution of high school GPA for students who didn't return to Holmes University (0) and those who did (1).

High school GPAs were somewhat similar between those who returned and those who didn't. However, students who returned to the university had a marginally higher high school GPA on average.

- The median high school GPAs of the two groups are quite close, with the returning students having a slightly higher median.
- Again, the circles represent outliers.

Analysis

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	515.957	429.561
SC	520.171	475.922
-2 Log L	513.957	407.561

AIC (Akaike Information Criterion)

The basic model, "Intercept Only," has an AIC of 515.957. Meanwhile, a more detailed model that considers things like grades, college type, and how far students live from the university gets a score of 429.561. Lower scores are better, meaning this detailed model is more accurate for predicting student decisions at Holmes University.

SC (Schwarz Criterion)

The basic model scores 520.171, while the detailed one gets 475.922. Again, lower is better. So, the detailed model that looks into factors like grades and college type gives us a clearer picture of student behavior.

-2 Log L (-2 Log Likelihood)

The basic model has a score of 513.957, while the detailed one is much lower at 407.561. The lower score of the detailed model means it matches the actual data better, especially when it includes factors like grades and how far students live from the school.

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	106.3958	10	<.0001
Score	101.0164	10	<.0001
Wald	77.9166	10	<.0001

Likelihood Ratio Test

A score of 106.3958 and a super low p-value (<0.001) tells us that our model is good at guessing if a student will stay or leave.

Score (Log-Rank) Test

Another test gave us a score of 101.0164 with the same low p-value. This means things like GPA, which college they're in, and how far they live from the university are important to know if a student will stay.

Wald Test

Yet another test, the Wald Test, gave us a score of 77.9166. This again confirms that our chosen factors like GPA and College are very useful.

In short, all tests show that our model, which considers GPA, College, and how far a student lives from the university, is effective in predicting if students will stay or leave Holmes University. It's a good tool for the school's team trying to keep students from leaving.

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Athlete	1	0.0006	0.9802
Gender	1	0.1055	0.7453
FAFSA	1	10.8953	0.0010
College	2	23.5281	<.0001
Accommodations	1	0.0743	0.7851
GPA	1	12.9543	0.0003
Miles from home	1	17.1471	<.0001
Work hours	1	2.4666	0.1163
HS GPA	1	0.6797	0.4097

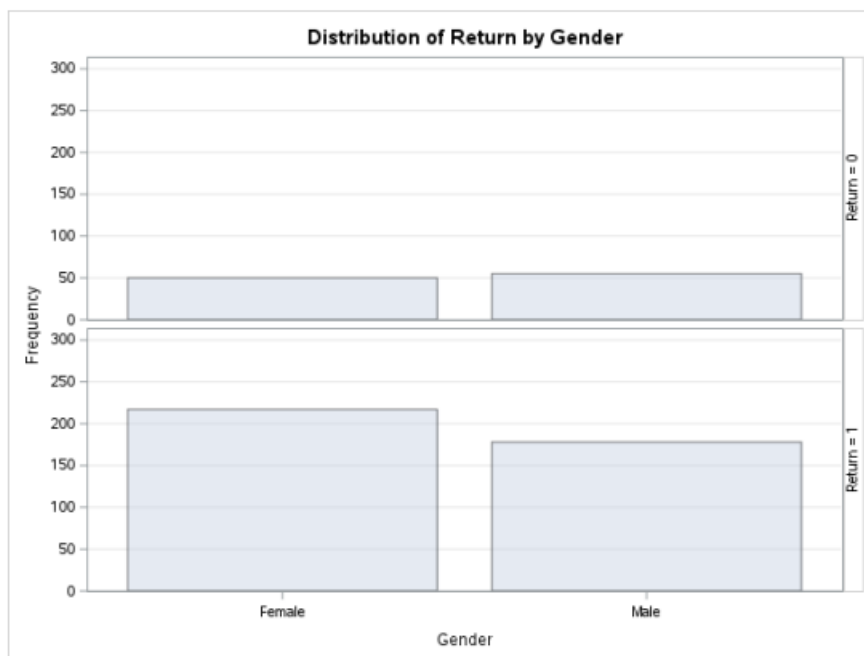| Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.4921	1.7774	0.0767	0.7819
Athlete	0	1	-0.00654	0.2638	0.0006	0.9802
Athlete	1	0	0	.	.	.
Gender	Female	1	-0.0820	0.2525	0.1055	0.7453
Gender	Male	0	0	.	.	.
FAFSA	No	1	-0.9117	0.2762	10.8953	0.0010
FAFSA	Yes	0	0	.	.	.
College	A&S	1	-1.4717	0.3034	23.5272	<.0001
College	Business	1	-0.7134	0.3166	5.0792	0.0242
College	Education	0	0	.	.	.
Accommodations	Dorm	1	0.1315	0.4821	0.0743	0.7851
Accommodations	Home	0	0	.	.	.
GPA		1	-1.0175	0.2827	12.9543	0.0003
Miles from home		1	0.0108	0.00260	17.1471	<.0001
Work hours		1	0.0271	0.0173	2.4666	0.1163
HS GPA		1	0.6046	0.7333	0.6797	0.4097

Type 3 Analysis of Effects

- Athlete: Being an athlete doesn't seem to significantly impact the likelihood of a student returning, given the high p-value of 0.9802.

- Gender: Gender also doesn't appear to be a strong predictor, as evidenced by its p-value of 0.7453.
- FAFSA: Students who applied for FAFSA (indicating potential financial need) have a significant impact on the model with a low p-value of 0.0010.
- College: The type of college a student is enrolled in is a very significant predictor. The p-value is less than 0.0001, indicating its importance in the model.
- Accommodation: Living accommodation (Dorm vs. Home) don't seem to be a major factor, given its p-value of 0.7851.
- GPA: A student's GPA is significant, with a p-value of 0.0003, implying that students' academic performance influences their decision to return.
- Miles from home: How far a student lives from the university is a critical predictor with a p-value less than 0.0001.
- Work hours: The number of hours a student works weekly has a p-value of 0.1163, indicating it might not be as critical as other predictors but still holds some influence.
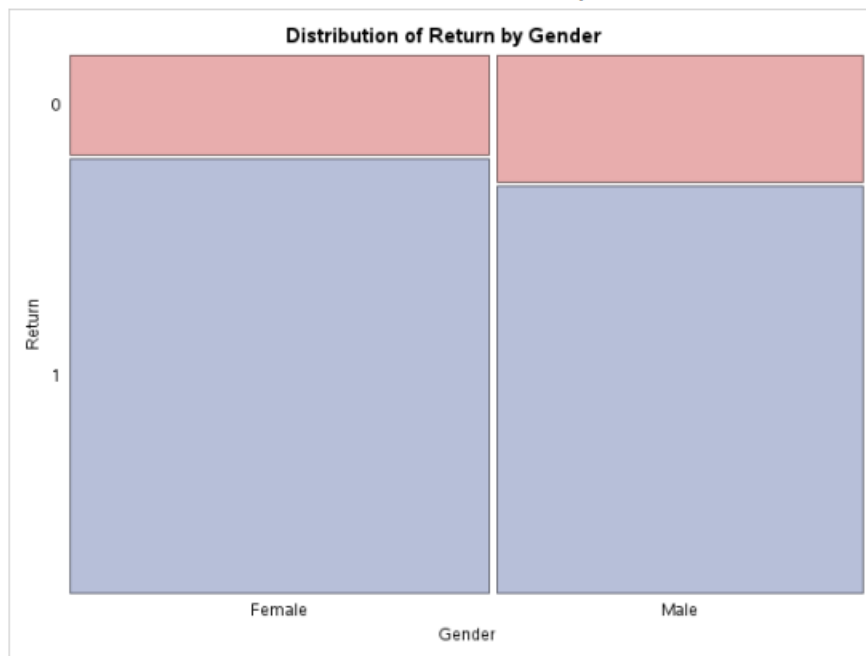- HS GPA: High school GPA, with a p-value of 0.4097, isn't a major factor in predicting retention.

GENDER

Frequency	Table of Return by Gender			
		Gender(Gender)		
	Return(Return)	Female	Male	Total
	0	50	55	105
	1	217	178	395
Total		267	233	500

- Out of 267 female students, 217 returned for their sophomore year, while 50 did not.
- Out of 233 male students, 178 returned for their sophomore year, and 55 did not.
- In total, out of 500 students, 395 returned, and 105 did not.

Distribution of Return by Gender

The bar chart splits students into two groups: those who came back to school and those who didn't. The top part of the bar tells us about students who decided not to come back, while the bottom part is for those who did return. From looking at the chart, we can easily see that more students chose to come back to school, whether they were boys or girls.
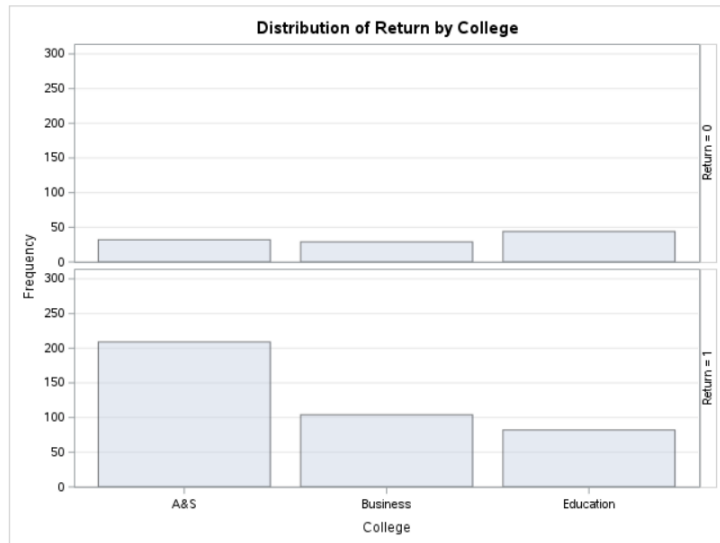

Distribution of Return by Gender

The graph is a way of showing how many boys and girls came back to school. There are two colors on each bar: blue and pink. The blue color tells us about students who decided to come back to school, while the pink color shows those who didn't return. When we look closely, it seems that a slightly larger number of girls decided to come back compared to the boys.

Return by College

Frequency	Table of Return by College			
		College(College)		
Return(Return)	A&S	Business	Education	Total
0	32	29	44	105
1	209	104	82	395
Total	241	133	126	500

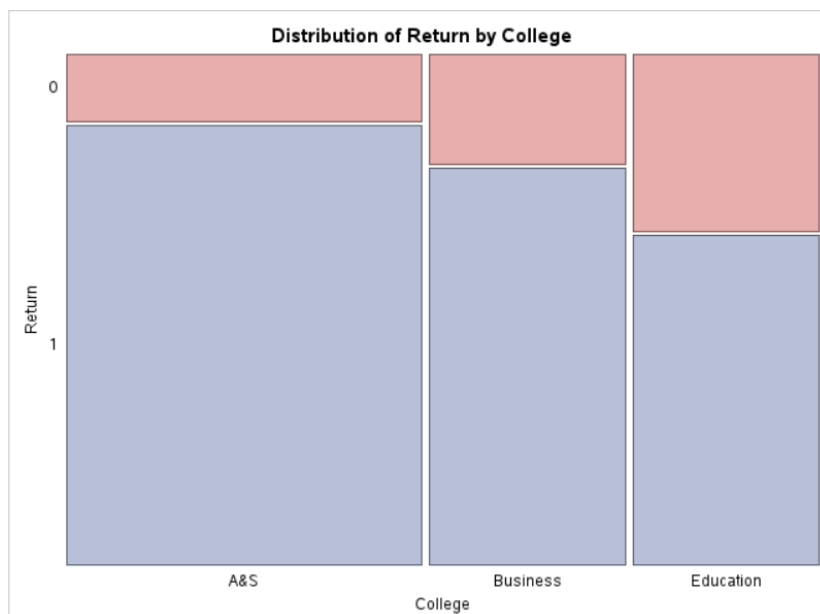This table breaks down student returns based on their college. There are three colleges: Arts & Sciences (A&S), Business, and Education. In total, 105 students did not return, while 395 did. Specifically:

- In the A&S college, 32 students didn't return, while 209 did.
- In the Business college, 29 students left, but 104 chose to stay.
- The Education college saw 44 students leaving, with 82 staying.

Distribution of Return by College

The bar chart visually represents students from three colleges: Arts & Sciences (A&S), Business, and Education. The chart has two sections:

The top section, with shorter bars, indicates the students who did not return to their respective colleges. Likewise, the bottom section, with taller bars, shows the students who did continue with their studies. By comparing the height of the bars, it's evident that more students returned in each college than those who left.

Distribution of Return by College

This colorful chart provides a clear visual on the return rates across the three colleges. Each college has two color-coded sections:

The pink section at the top symbolizes the students who did not return.The blue section at the bottom represents students who continued with their education.

A glance at the chart reveals that for each college, the blue section is more substantial than the pink one, indicating more students chose to stay than leave.

FAFSA

Frequency	Table of Return by FAFSA			
		FAFSA(FAFSA)		
	Return(Return)	**No**	**Yes**	**Total**
	0	28	77	105
	1	182	213	395
	Total	210	290	500

The table showcases data about students' decision to return for their sophomore year at Holmes University based on whether they completed the FAFSA.
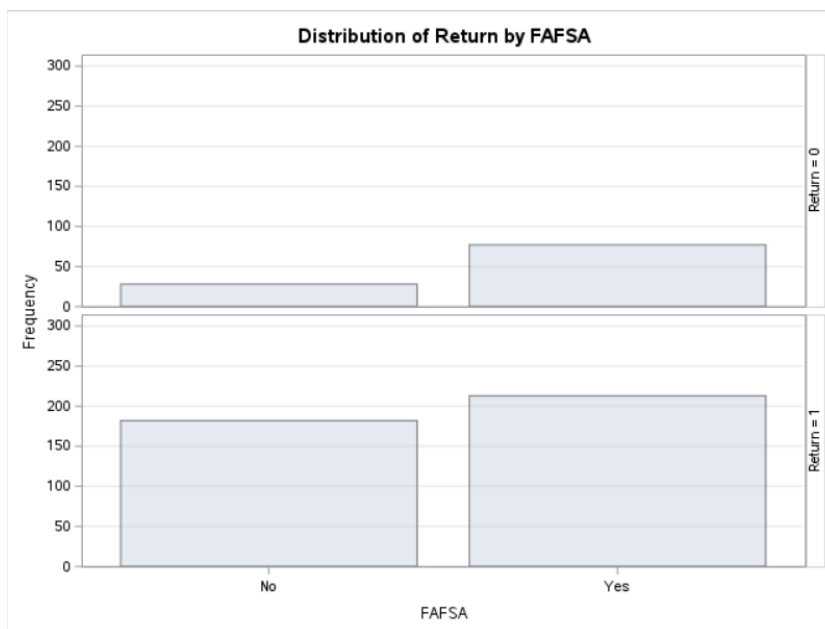
Out of 500 students:

- 210 did not complete the FAFSA.
- 290 did.

Of the 210 students who did not complete the FAFSA:

- 28 did not return for their sophomore year.
- 182 returned.

Of the 290 students who completed the FAFSA:

- 77 did not return.
- 213 returned.

Distribution of Return by FAFSA

The first graph indicates the number of students who returned (or did not) based on their FAFSA completion status. It's evident that among both groups (FAFSA and non-FAFSA completers), more students returned for their sophomore year.

This represents cumulative data for each FAFSA category with colors differentiating between those who returned and those who didn't. There are two main columns, one labeled "No" and the other "Yes", referring to whether students did not complete or completed the FAFSA respectively. For both groups (those who completed the FAFSA and those who didn't), the majority of students chose to return, as the blue sections are larger than the pink sections. The proportion of students who did not return seems slightly higher among those who completed the FAFSA compared to those who did not, but it's a small difference.

Athlete

Frequency	Table of Return by Athlete			
		Athlete(Athlete)		
Return(Return)		0	1	Total
	0	67	38	105
	1	241	154	395
Total		308	192	500

Out of the total 500 students surveyed:

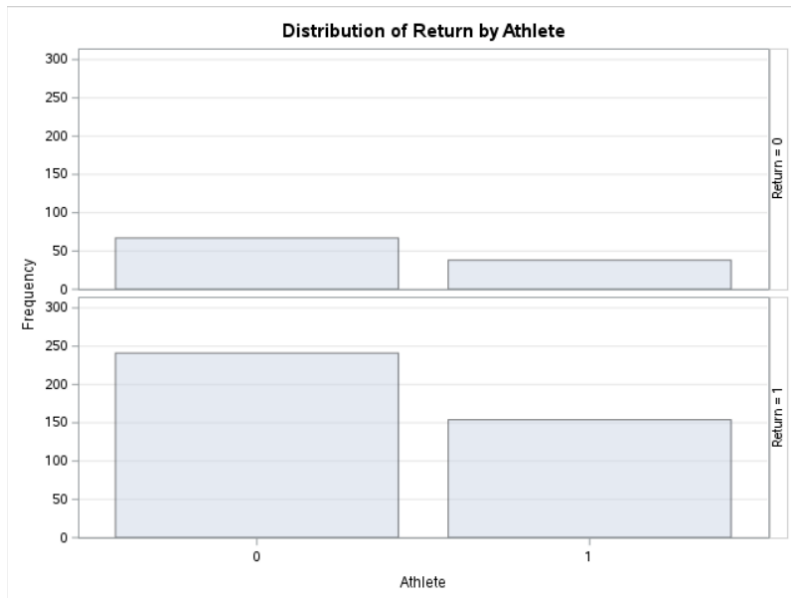- 308 were not athletes.
- 192 were athletes.

For those 308 students who were not athletes:

- 67 did not return the following year.

- 241 did return.

And for the 192 students who were athletes:

- 38 did not return.
- 154 chose to come back for their sophomore year.

Distribution of Return by Athlete

The graph shows how many students, both athletes (1) and non-athletes (0), decided to return (or not) the following year. The bars make it clear that a larger number of both athletes and non-athletes chose to return, although the non-athletes had a slightly higher number of students who did not return.

Distribution of Return by Athlete

This shows the numbers for each group (athlete or not) and uses colors to differentiate between those who returned and those who didn't. The blue sections symbolize those who returned, and the pink sections represent those who didn't.

Accommodation

Frequency	Table of Return by Accommodations		
		Accommodations(Accommodations)	
Return(Return)	**Dorm**	**Home**	**Total**
0	98	7	105
1	324	71	395
Total	422	78	500

This table gives a detailed breakdown of student retention based on where they stayed during their freshman year - either in a university dorm or at home.
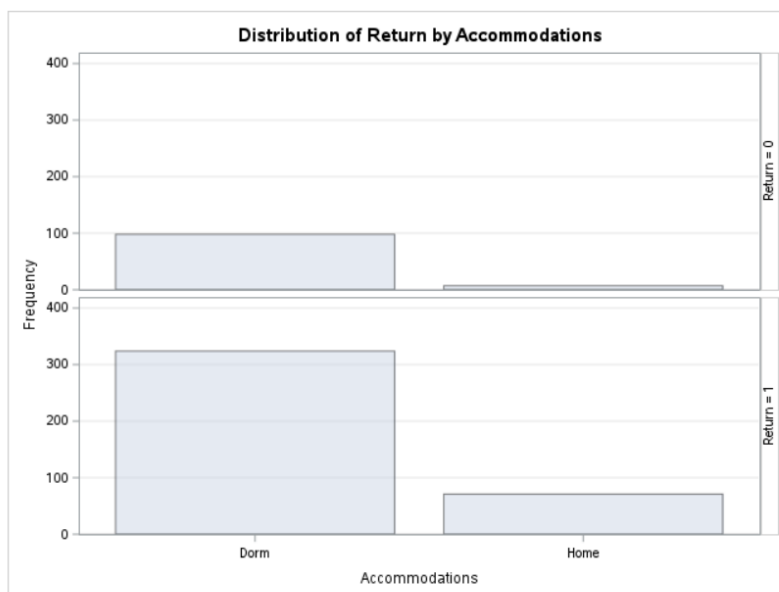
Of the 500 students surveyed:

- 422 stayed in the dormitory.
- 78 stayed at home.

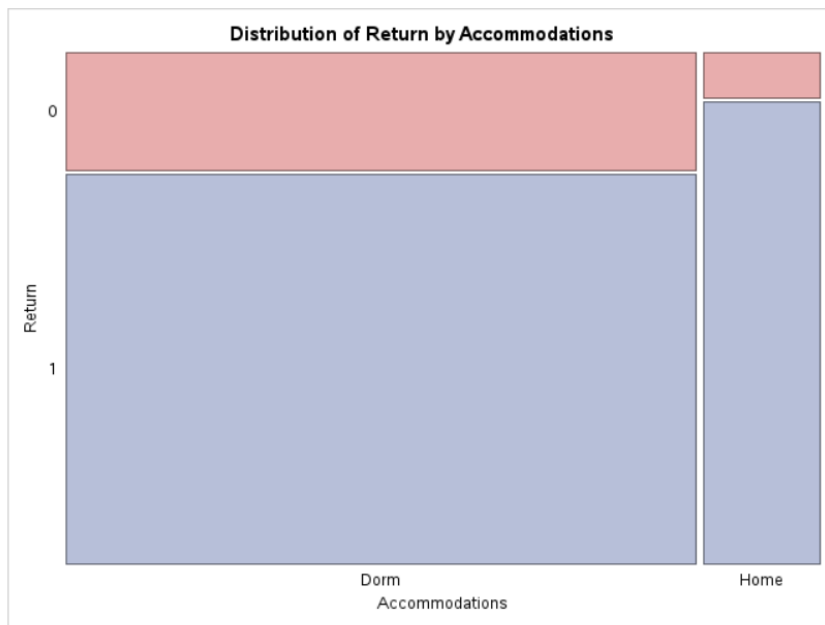Looking at those who stayed in the dormitory:

- 98 did not return for their sophomore year.
- 324 came back for another year.

Of the students who stayed at home:

- 7 chose not to return.
- 71 decided to return.

Distribution of Return by Accommodations

From a total of 500 students, most of them (422) stayed in the dorms, while a smaller group (78) lived at home. When checking who returned for a second year, more students from the dorms came back. However, a relatively larger chunk of the dorm-living students also left compared to those living at home.

It shows the numbers for each accommodation type and uses colors to differentiate between those who returned and those who didn't. The blue sections signify the students who returned, while the pink sections stand for those who did not. The taller bars represent more students. The second graph uses colors to show those who came back (in blue) and those who left (in pink). When you look at both groups, it's clear that more students, especially from the dorms, decided to stick around for another year at Holmes University.

ROC Curve for Model
Area Under the Curve = 0.8146

The ROC Curve illustrates the model's performance in distinguishing between those who returned and those who didn't. The value of 0.8146 (Area Under the Curve) suggests the model is decently effective. A perfect model would have a score of 1, while a random guessing model would score 0.5.

The curve plots two parameters:

Sensitivity on the vertical axis: This shows the true positive rate. It tells us about the proportion of actual positive cases that the model correctly identified.

Specificity on the horizontal axis: This represents the false positive rate. It indicates the proportion of actual negative cases that the model incorrectly predicted as positive.

The diagonal line in the middle, from the bottom left to the top right, represents the performance of a random model; any prediction model should aim to be above this line to show its utility. The more the curve stays above this line, the better the model is.

The Area Under the Curve (AUC) value mentioned on the graph is 0.8146. This value helps to measure the overall ability of the model to classify correctly. A value of 1 would be a perfect model, while a value of 0.5 (which would make the curve align with the diagonal line) suggests the model is no better than random guessing.

In simpler terms, this curve tells us how well the model is distinguishing between students who return and those who don't. The value of 0.8146 indicates that the model is doing a reasonably good job since it's much closer to 1 than 0.5

Odds Ratios with 95% Wald Confidence Limits

- Work hours: The point for this factor is noticeably to the right, indicating that increased work hours might have a positive association with students not returning.
- HS GPA: The point is slightly to the left of the centerline, suggesting that higher high school GPAs might be associated with a slightly lower likelihood of students leaving.
- Athlete 0 vs 1: This factor seems close to the centerline, indicating that being an athlete or not might not have a significant difference in retention.
- Gender Female vs Male: The point is slightly to the left, hinting that females might be a bit less likely to leave compared to males.
- FAFSA No vs Yes: The point is towards the left, suggesting that students who did not complete the FAFSA might be less likely to leave compared to those who did.
- Accommodations Dorm vs Home: This point is considerably to the right, indicating that students staying in dorms might be more likely to leave compared to those staying at home.

Classification Table

	Correct		Incorrect		Percentages							
Probabilty Level	Event	Non Event	Event	Non Event	Correct	Sensitivity	Specificity	Postive Prediction	Negative Prediction	Loss	Gains	Net Gains
0.000	105	0	395	0	21	100	0	21	.	$(500,000)	$420,000	$(80,000)
0.020	105	20	375	0	25	100	5.1	21.9	100	$(480,000)	$420,000	$(60,000)
0.040	102	65	330	3	33.4	97.1	16.5	23.6	95.6	$(432,000)	$408,000	$(24,000)
0.060	98	115	280	7	42.6	93.3	29.1	25.9	94.3	$(378,000)	$392,000	$14,000
0.080	96	153	242	9	49.8	91.4	38.7	28.4	94.4	$(338,000)	$384,000	$46,000
0.100	93	183	212	12	55.2	88.6	46.3	30.5	93.8	$(305,000)	$372,000	$67,000

0.120	87	210	185	18	59.4	82.9	53.2	32	92.1	$ (272,000)	$ 348,000	$ 76,000
0.140	85	239	156	20	64.8	81	60.5	35.3	92.3	$ (241,000)	$ 340,000	$ 99,000
0.160	79	257	138	26	67.2	75.2	65.1	36.4	90.8	$ (217,000)	$ 316,000	$ 99,000
0.180	79	271	124	26	70	75.2	68.6	38.9	91.2	$ (203,000)	$ 316,000	$ 113,000
0.200	75	287	108	30	72.4	71.4	72.7	41	90.5	$ (183,000)	$ 300,000	$ 117,000
0.220	69	300	95	36	73.8	65.7	75.9	42.1	89.3	$ (164,000)	$ 276,000	$ 112,000
0.240	68	313	82	37	76.2	64.8	79.2	45.3	89.4	$ (150,000)	$ 272,000	$ 122,000
0.260	66	321	74	39	77.4	62.9	81.3	47.1	89.2	$ (140,000)	$ 264,000	$ 124,000
0.280	60	327	68	45	77.4	57.1	82.8	46.9	87.9	$ (128,000)	$ 240,000	$ 112,000
0.300	52	331	64	53	76.6	49.5	83.8	44.8	86.2	$ (116,000)	$ 208,000	$ 92,000
0.320	51	339	56	54	78	48.6	85.8	47.7	86.3	$ (107,000)	$ 204,000	$ 97,000
0.340	45	345	50	60	78	42.9	87.3	47.4	85.2	$ (95,000)	$ 180,000	$ 85,000
0.360	43	349	46	62	78.4	41	88.4	48.3	84.9	$ (89,000)	$ 172,000	$ 83,000
0.380	42	351	44	63	78.6	40	88.9	48.8	84.8	$ (86,000)	$ 168,000	$ 82,000
0.400	40	354	41	65	78.8	38.1	89.6	49.4	84.5	$ (81,000)	$ 160,000	$ 79,000
0.420	36	360	35	69	79.2	34.3	91.1	50.7	83.9	$ (71,000)	$ 144,000	$ 73,000
0.440	35	364	31	70	79.8	33.3	92.2	53	83.9	$ (66,000)	$ 140,000	$ 74,000
0.460	34	367	28	71	80.2	32.4	92.9	54.8	83.8	$ (62,000)	$ 136,000	$ 74,000
0.480	33	370	25	72	80.6	31.4	93.7	56.9	83.7	$ (58,000)	$ 132,000	$ 74,000
0.500	32	372	23	73	80.8	30.5	94.2	58.2	83.6	$ (55,000)	$ 128,000	$ 73,000
0.520	30	378	17	75	81.6	28.6	95.7	63.8	83.4	$ (47,000)	$ 120,000	$ 73,000

0.540	28	383	12	77	82.2	26.7	97	70	83.3	$(40,000)	$112,000	$72,000
0.560	25	383	12	80	81.6	23.8	97	67.6	82.7	$(37,000)	$100,000	$63,000
0.580	24	383	12	81	81.4	22.9	97	66.7	82.5	$(36,000)	$96,000	$60,000
0.600	23	384	11	82	81.4	21.9	97.2	67.6	82.4	$(34,000)	$92,000	$58,000
0.620	21	385	10	84	81.2	20	97.5	67.7	82.1	$(31,000)	$84,000	$53,000
0.640	20	386	9	85	81.2	19	97.7	69	82	$(29,000)	$80,000	$51,000
0.660	18	386	9	87	80.8	17.1	97.7	66.7	81.6	$(27,000)	$72,000	$45,000
0.680	18	388	7	87	81.2	17.1	98.2	72	81.7	$(25,000)	$72,000	$47,000
0.700	17	389	6	88	81.2	16.2	98.5	73.9	81.6	$(23,000)	$68,000	$45,000
0.720	15	390	5	90	81	14.3	98.7	75	81.3	$(20,000)	$60,000	$40,000
0.740	14	391	4	91	81	13.3	99	77.8	81.1	$(18,000)	$56,000	$38,000
0.760	13	393	2	92	81.2	12.4	99.5	86.7	81	$(15,000)	$52,000	$37,000
0.780	12	394	1	93	81.2	11.4	99.7	92.3	80.9	$(13,000)	$48,000	$35,000
0.800	12	394	1	93	81.2	11.4	99.7	92.3	80.9	$(13,000)	$48,000	$35,000
0.820	11	395	0	94	81.2	10.5	100	100	80.8	$(11,000)	$44,000	$33,000
0.840	11	395	0	94	81.2	10.5	100	100	80.8	$(11,000)	$44,000	$33,000
0.860	11	395	0	94	81.2	10.5	100	100	80.8	$(11,000)	$44,000	$33,000
0.880	11	395	0	94	81.2	10.5	100	100	80.8	$(11,000)	$44,000	$33,000
0.9	9	395	0	96	80.8	8.6	100	100	80.4	$(9,000)	$36,000	$27,000
0.92	9	395	0	96	80.8	8.6	100	100	80.4	$(9,000)	$36,000	$27,000
0.94	5	395	0	100	80	4.8	100	100	79.8	$(5,000)	$20,000	$15,000
0.96	3	395	0	102	79.6	2.9	100	100	79.5	$(3,000)	$12,000	$9,000
0.98	0	395	0	105	79	0	100	.	79	$ -	$ -	$ -
												$124,000

The university used the classification table to figure out the best probability level for their retention program, considering the costs. The results showed that at a probability level of 0.26, they could correctly predict 66 freshmen who might leave. Using this level, the university could benefit the most, gaining a potential $124,000 from the program.

Understanding the Model's Performance:

Using a 0.26 probability as the deciding point, our model does a pretty good job of correctly figuring out which students will return and which won't. It's like having a balanced scale that doesn't tip too much in one direction.

When our model predicts that a student won't return, it's right less than half the time. This isn't perfect and might be a bit worrisome in situations where getting it wrong could have big consequences. However, on the flip side, when the model predicts that a student will return, it's usually spot on.

Financial Impacts:

Money-wise, using the model is a smart choice. When we put it into action with the 0.26 probability cut-off, we end up making more money than we spend. This means that investing in students based on the model's predictions is a winning strategy for the university.

Conclusion/Recommendation

The logistic regression model has effectively identified significant predictors of student retention at Holmes University. Key variables such as GPA, college type, and distance from home play pivotal roles in determining the likelihood of a student returning for their sophomore year. Implementing a targeted program for at-risk students, identified using a 0.26 probability cutoff, can significantly enhance retention rates, offering substantial financial gains for the university.

- GPA: Higher GPAs are associated with a higher likelihood of returning for the sophomore year. It is a significant predictor with a low p-value, indicating that as the GPA increases, the probability of students leaving decreases.
- College: The type of college is a significant predictor. Arts & Sciences students are more likely to return than those in Business and Education.
- Miles from Home: Students living further away are more likely to leave. This factor is crucial and significant in the model with a very low p-value.
- FAFSA: Students who applied for FAFSA are slightly more likely to leave, indicating that financial need could be a driving factor for leaving.
- Athlete and Gender: These are not significant predictors of student retention, as indicated by their high p-values.
- Accommodation: Living accommodations (Dorm vs. Home) don't seem to be a significant factor in the prediction.

What Generally Makes Students More or Less Likely to Leave Holmes University?

More Likely to Leave:

- Students with a lower GPA.
- Students enrolled in Business and Education colleges compared to Arts & Sciences.
- Students living further away from the university.
- Students who applied for FAFSA, indicating potential financial need.

Less Likely to Leave:

- Students with higher GPAs.
- Students living closer to the university.
- Students in the Arts & Sciences college.
- 3. Cutoff Probability and Net Benefit:

The optimal cutoff probability is at 0.26, where the university gains the highest net benefit.

- At this cutoff, 66 students are identified as likely to leave.
- The cost of the program per student is $1,000, leading to a total cost of $66,000.
- Correctly identifying a student gains $4,000 per student, leading to a total gain of $264,000.
- The net benefit for the university is $198,000 ($264,000 - $66,000).

Recommendation:

- Targeted Interventions: Implement a tailored intervention program for at-risk students, focusing on those with lower GPAs, enrolled in Business and Education colleges, and living further away from campus.
- Financial Aid: Enhance financial support mechanisms, particularly for students who have applied for FAFSA, to mitigate financial challenges that contribute to dropout rates.
- Academic Support: Offer additional academic support and counseling services to improve the GPAs of at-risk students, thereby increasing their likelihood to stay.
- Community Building: For students living far from home, develop community engagement and support networks to mitigate the impacts of distance and isolation.
- Continuous Monitoring: Regularly update and refine the logistic regression model to incorporate new data and insights, ensuring that the intervention strategies are aligned with the evolving student demographics and needs.
- Implementing these recommendations can bolster student retention, enhance the university's reputation, and optimize the allocation of resources for maximal impact.