

Zachary Picard

Project 2

4/11/2024

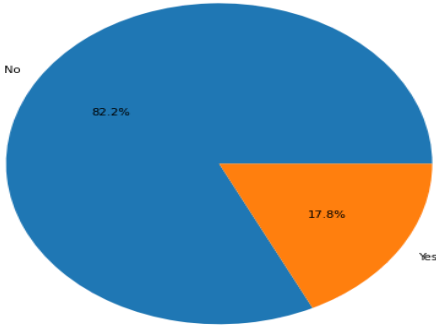
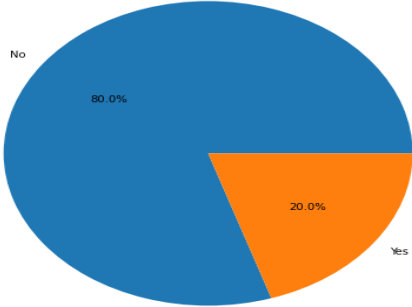
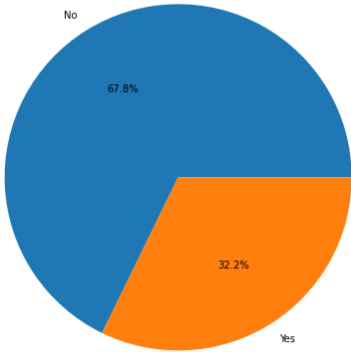
Purpose: The purpose of this analysis was to conduct an investigation on students' ethical behavior at Bayview University. Some faculty members were led to believe that cheating was becoming more prevalent among students, while other faculty members did not see it to be a problem. This analysis stems from an anonymous survey commissioned by the dean of the business school, and the corresponding findings uncover the truth behind the ethical behavior of the students.

Background: There has been recent speculation revolving around the wall street brokers in 2008 and 2009, and their poor ethical behavior. This demographic of people consisted of many positions such as executives, financial managers and other corporate officers. There was a speculation that the cause of this unethical behavior stemmed from habits developed in schooling. This is what led to a study administered at the Bayview University, to conduct a survey on business students to investigate the ethical behavior and tendencies of students.

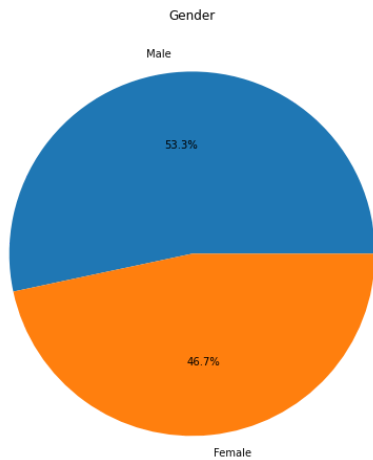
Scope: In this analysis we looked at descriptive statistics and frequency distributions, found disparities between genders and students who had previously cheated. We created confidence intervals for the proportions of cheaters at the school for all students, male students and female students. Additionally, the analysis dove into the correlation values between the discrete variables for both men and women and compared. We then visualized them using heat maps. Lastly we conducted hypothesis tests to compare proportions of students at other universities.

Methodology: This analysis was created using Python in collaboration with Jupyter Notebooks, specifically the package Pandas, to analyze and visualize trends within the dataset.

Descriptive Statistics

<p>Copied From Internet Frequency Distribution ALL Percentage of Copied from Internet</p>  <table border="1"><thead><tr><th>Response</th><th>Percentage</th></tr></thead><tbody><tr><td>No</td><td>82.2%</td></tr><tr><td>Yes</td><td>17.8%</td></tr></tbody></table>	Response	Percentage	No	82.2%	Yes	17.8%	<p>As we can see from the pie chart, 82.2% of the students from the survey had concluded that they had not copied from the internet, as compared to the 17.8% of the students who had copied from the internet.</p>
Response	Percentage						
No	82.2%						
Yes	17.8%						
<p>Copied On Exam Frequency Distribution ALL Copied on Exam</p>  <table border="1"><thead><tr><th>Response</th><th>Percentage</th></tr></thead><tbody><tr><td>No</td><td>80.0%</td></tr><tr><td>Yes</td><td>20.0%</td></tr></tbody></table>	Response	Percentage	No	80.0%	Yes	20.0%	<p>Here we can conclude that 80% of the students in the survey have not admitted to cheating on an exam, while 20% of the students had copied from an exam.</p> <p>This means that out of the whole study, students are more likely to cheat by copying on an exam than copying from the internet. This is most likely because, there is a larger chance of getting caught modernly by cheating on the internet than copying on another students exam</p>
Response	Percentage						
No	80.0%						
Yes	20.0%						
<p>Collaborated on Individual Project Frequency Distribution ALL Collaborated on Individual Project</p>  <table border="1"><thead><tr><th>Response</th><th>Percentage</th></tr></thead><tbody><tr><td>No</td><td>67.8%</td></tr><tr><td>Yes</td><td>32.2%</td></tr></tbody></table>	Response	Percentage	No	67.8%	Yes	32.2%	<p>Here we can see that the largest portion of cheating from the survey clearly comes from people who collaborate on individual projects. 67.8% of students did not admit to cheating like this, while 32.2% did.</p>
Response	Percentage						
No	67.8%						
Yes	32.2%						

Gender Frequency Distribution ALL

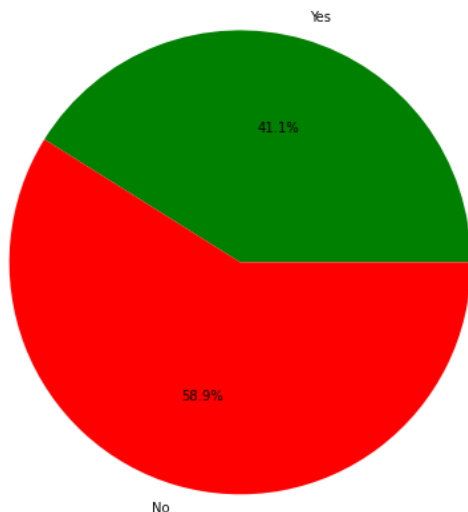


Here we can see that 53.3% of the students from the survey are male while 46.7% of the students are female.

I believe there are slightly more males in the study because on average women are outnumbered by men in most business schools.

Next I would like to create a pie chart to get a grasp of what percentage of students had cheated in any of the three categories. This will give us a visual grasp as to what the ethical behavior of the students looks like as a whole.

Cheating Frequency Distribution ALL



Here we can see that in total 41.1% of students had answered yes to cheating in one form in the survey, while 58.9% of the students who responded to the survey answered no.

This leads me to believe that a student is more likely to be unethical and cheat in more than one way if they had already answered yes to cheating. This means that a student who has already made an unethical decision could be more likely to make another one.

To confirm this hypothesis, I will perform a further analysis below.

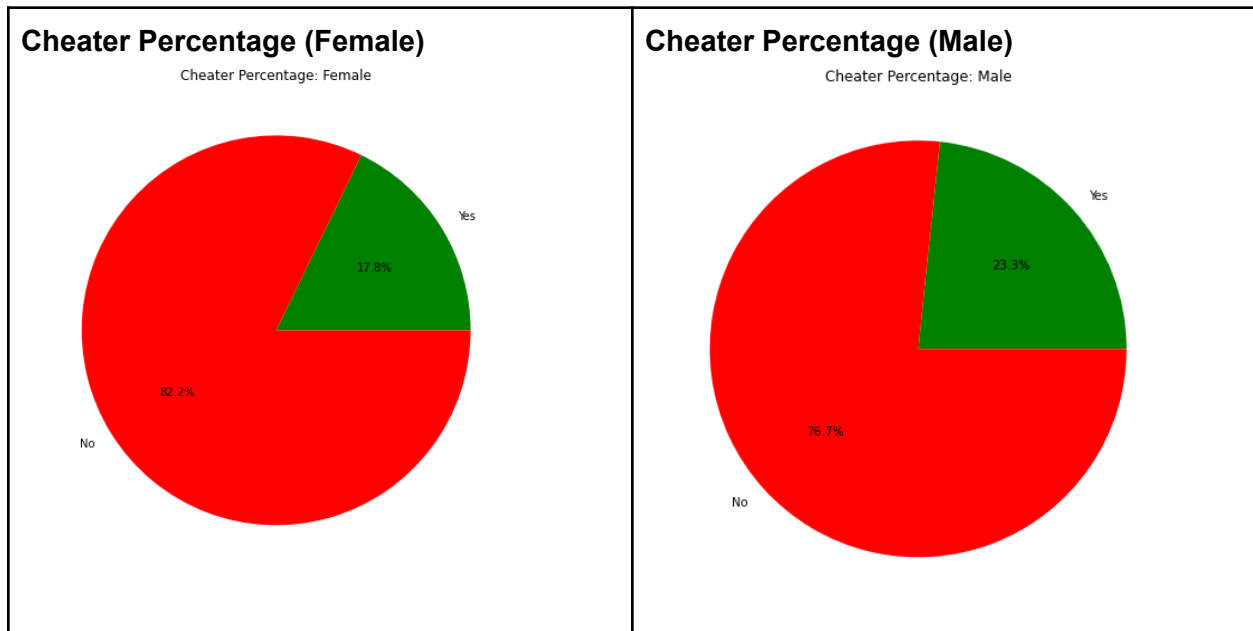
If a student has already answered yes to one of the columns, are they more likely to have cheated in multiple different ways?

This was the original hypothesis, based on the frequency distributions for the different columns. After performing some analysis. We find that 54% of cheaters only cheated in 1 category, while 46% of cheaters, cheated in more than one category, while this isn't significant evidence to say

that you are MORE likely to cheat in multiple columns if you are a cheater, we can say that you have a large chance of being someone who cheats in multiple ways, if you had already answered 'Yes' to cheating in one of the 3 survey questions.

What about Males Vs. Females ethical habits?

It would also be interesting to have a look at what proportion of the cheaters were males vs. females. Additionally, it would be interesting to expand on the last question we answered and see if males who cheat once are more likely to cheat in a different method as well then females who cheat once and how likely they are to cheat in a different method again.



Here we can see that there is a significantly higher percentage of cheaters who are male vs cheaters who are female. The difference is about 5%. Additionally, we want to take a look at the likelihood of the two subsets of the data cheating in more than one way, given they are in at least one way a cheater. This can be seen Below

The students (Female) that only cheated in one category makes up 50%
The students (Female) that cheated in MORE THAN one category makes up 50%
The students (Male) that only cheated in one category makes up 57%
The students (Male) that cheated in MORE THAN one category makes up 43%

So, we conclude that male students are more likely to only cheat in one way then women, and that women are more likely to be a multi-type cheater academically, given they are a cheater in at least one category. On the contrary women however, are overall less likely to cheat based on the demographic. This leads me to believe that many more men are willing to cheat in one way,

possibly a few times, whereas women are less likely to cheat at all, but if they do, become more chronic cheaters, and will cheat in multiple ways.

95% Confidence Interval for Proportion of cheaters(All)

<p>Students Who Cheat - (0.4, 0.42)</p> <p>If we were to repeat this process/survey, we can say that 95% of the time the proportion 0.4 - 0.42 would contain the population proportion of saying yes to having cheated in any category</p>	<p>Students Who Don't Cheat - (0.57, 0.60)</p> <p>If we were to repeat this process/survey, we can say that 95% of the time the proportion 0.57 - 0.6 would contain the population proportion of saying no to having cheated in any category</p>
--	--

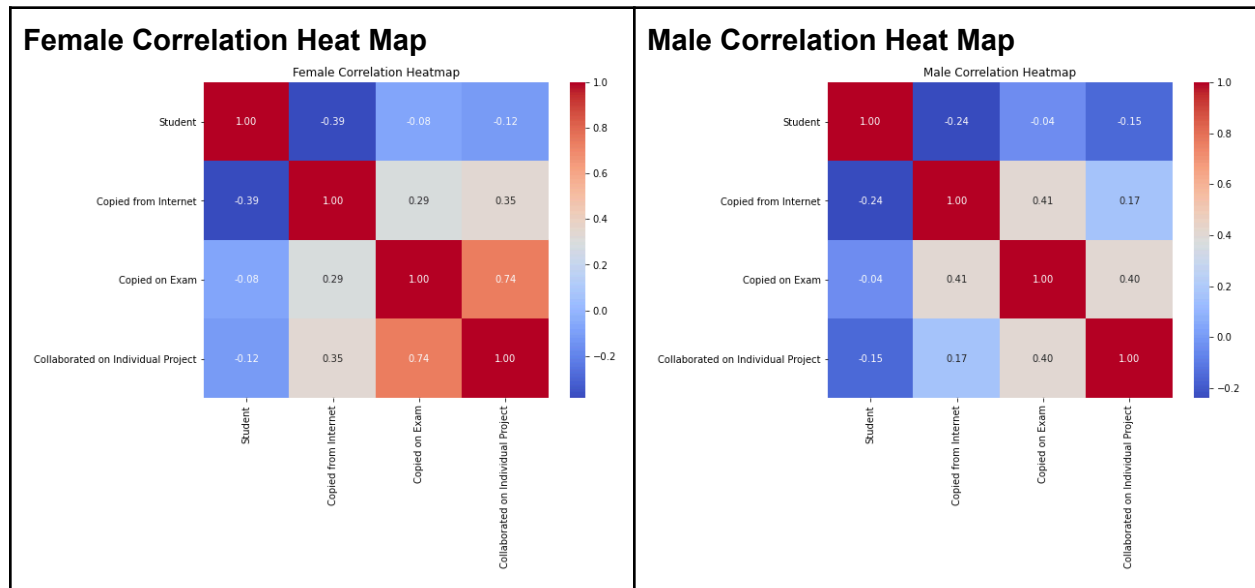
95% Confidence Interval for Proportion of cheaters(Female)

<p>Female Students Who Cheat - (0.169, 0.186)</p> <p>If we were to repeat this process/survey, we can say that 95% of the time the proportion 0.169 - 0.186 would contain the population proportion of saying yes to having cheated being a female</p>	<p>Students Who Don't Cheat - (0.279, 0.298)</p> <p>If we were to repeat this process/survey, we can say that 95% of the time the proportion 0.279 - 0.298 would contain the population proportion of saying no to having cheated in any category as a female</p>
--	---

95% Confidence Interval for Proportion of cheaters(Male)

<p>Students Who Cheat - (0.224, 0.242)</p> <p>If we were to repeat this process/survey, we can say that 95% of the time the proportion 0.224 - 0.242 would contain the population proportion of saying yes to having cheated being a male</p>	<p>Students Who Don't Cheat - (0.290, 0.309)</p> <p>If we were to repeat this process/survey, we can say that 95% of the time the proportion 0.290 - 0.309 would contain the population proportion of saying no to having cheated in any category as a male</p>
---	---

Correlation Between Types of Cheating in Males vs. Females



Here we are trying to investigate if there are significant differences in the correlation values between men and women and the different forms of cheating. We had already previously established some trends behind the general likelihood of a gender to cheat academically more than once, but now we are diving into further specifics. To accomplish this, we must first transform the values of the dataset into 1's and 0's so that we can see a correlation. After plotting the corresponding Heat Maps, we come to the following conclusions:

- Men are 10% more likely to copy on an exam given they copied from the internet
- Women are 18% more likely to collaborate on an individual paper given they copied from the internet
- Women are 35% more likely to collaborate on an individual project given they had copied on an exam

These are all of the significant differences between populations of males vs. females and the correlation values between the discrete variables, and help to show some of the discrepancies behind the demographics of cheaters at the college.

Hypothesis Test - Proportion of Business Students at Bayview University Involved in Cheating Less than proportion of cheaters in business students at other institutions

H0: $\mu = 56\%$

Ha: $\mu < 56\%$

To begin this hypothesis test, we must first create a new column in our dataframe to help see if the students in the survey are in general a cheater or not (if they have responded 'Yes' to any of the survey questions). This will allow us to calculate the sample proportion as well as the sample variance. We then pick a confidence level of 5%, corresponding to a z critical value of -1.645. We then calculate our z_score based on our p hat which was given through python by that extra column we created with our proportion of cheaters from the sample being at .411. We

also use our $H_0 = 0.56$, and find the length of the dataframe for n . We then calculate the z-score based on the previous criteria and obtain the result -2.847 which lies in the RR for the given level of significance. This means we can reject the null hypothesis and have evidence that the population proportion is less than 0.56 at the given confidence level. These values do agree and we can see that the sample proportion which is .411, and the z score lying in the RR both agree in the conclusion of H_a .

Hypothesis Test - Proportion of Business Students at Bayview University Involved in Cheating Less than proportion of cheaters in non-business students at other institutions

$H_0: \mu = 47\%$

$H_a: \mu < 47\%$

We are now going to mimic the same test but change the value of our null proportion (changing .56 to .47). We are doing this so we can see the comparison to the proportion of cheaters in non-business students at other institutions. We are going to use the same RR (at a 5% confidence level). Here we calculate a z-score of -1.12 which DOES NOT lie inside of the rejection region so we do not have enough evidence to suggest that the population proportion is less than 0.47. These results do not agree, since a value of .411 is less than .47, but at the same time we don't have significant evidence to suggest that the population proportion is less than 0.47 at the 5% significance level.

Hypothesis Test Conclusions

We can use this information to conclude that BayView University and their students have good ethics compared to other schools, and that the dean is incorrect in thinking that the students at Bayview university specifically in the business school cheat more than other universities. In fact despite the disparity between business and non-business students and their cheating habits (56% of business students admitted to cheating and 47% of non-business students admitted to cheating), Bayview University's business students survey results showed that there is evidence to suggest that their population proportion of cheaters is less than that of other institutions business students, therefore proving that while cheating is still an issue at the university, they are performing ethically better than other institutions business students.

Advice to Dean based on analysis:

Based on the analysis, I would focus on a few things for the Dean of Bayview University. First I would begin with understanding that while maintaining a high ethical academic standard is of utmost importance, that there is evidence to suggest that Bayview University is doing better than other institutions. However, there are some measures that I believe could be put forward to help eliminate future cheating and predict possible students who are more likely to cheat.

For example we know that if a student is female and has previously cheated by copying on the exam that there is a very high likelihood they will also at some point collaborate on an individual project and vice versa. This is based on a correlation value of 0.74. But by analyzing figures like

these we can maybe predict cheating before it happens. There are other instances of cases like this listed below as well.

Another example could be predicting that if a student is male and has copied on an exam, they are 10% more likely to copy from the internet as well and vice versa. This is another example of how we can use demographic and our analysis to predict people who are more likely to cheat recurrently

Additionally, we know to look for male cheaters more than female cheaters based on the analysis. However, we have to be very careful going down this road, because we don't take proportion into account. This is important because we can predict that there are more males than females in the business school based on the sample proportion.

Conclusion:

Summary of Findings:

Through this analysis we conducted an investigation of students' ethical behavior at Bayview University (The Business School). We found descriptive statistics based on the frequency distributions of the answers to the anonymous findings, we explored if students were more likely to cheat more than once if they were already a cheater, we looked at the disparities between men and women students. Then we calculated confidence intervals for the proportion of cheaters(all, female/male). We also created correlation heat maps for both men and women to view the differences in cheating in multiple categories and their corresponding correlations. Lastly we conducted hypothesis tests to compare the proportion of business students at Bayview University to the proportions of cheaters of both business students and non-business students at other institutions.

Strategic Implications:

The analysis helped us to create a demographic of students that are more likely to cheat, and the analysis also helped us to compare the students at Bayview University and their ethical academic habits to students of other universities. We learned that the students at Bayview University Business school have a good moral compass as compared to business students at other institutions. We also got some information based on correlation heat maps to predict what students could be cheating in the future.

Recommendations:

Further for the college, after the analysis, we recommend taking statistics of all non-business students at the school, so we can better compare the demographics of students of those at other universities. I also recommend looking into trends behind students' grades and seeing if they have correlation within the categories of types of cheating.

Further Steps:

Furthermore, the dean of the college should concern himself less with the comparison of the business schools to other schools, because there is evidence to suggest that the business students at Bayview cheat less than other schools, instead the dean should focus his attention to creating a model to try and predict which students are more likely to cheat in which way.

