FUNDAMENTALS OF DATA ANALYTICS PROJECT

TOPIC: - ANALYZE TRANSFER STUDENT DATA AND RECOMMEND A POLICY CHANGE

INTRODUCTION

The Leeds School of Business at the University of Colorado is thinking about, boosting the GPA requirements for students transferring from other CU schools and institutions. The aim is to reduce the number of transfer students by 25-50% to make course scheduling a little easier for the administration. After looking at the last five semesters worth of data (Spring 2021 through Spring 2023) for admitted transfer students, our team has come up with some thoughts on how to tweak the GPA and grades criteria for key courses like ECON and MATH to reach the enrollment reduction goals

METHODS AND FINDINGS:

"Method 1 was conducted by Sai Arvind Atluri, Method 2 by Sandeep Reddy Modugu, and Method 3 by Amandeep Bhardwaj."

Using GPA Thresholds to Model Enrollment Reductions (Method 1):-
Analysis of overall GPAs showed minimums of 3.14 and 3.30 were needed to reduce enrollment by 25% and 50% respectively. Visualizing the GPA data illustrated the impact of raising GPA cut-offs on student eligibility.

Leveraging Grade Requirements for ECON Courses to Control Transfers (Method 2):-
Prescribing minimum grades of B in ECON 2010 and 2020 reduced eligible transfers by 25%, from 908 to 679 students. Raising the bar to B+ achieved a 50% reduction. Focusing on these critical prerequisite courses allowed strategic control of transfers.

Utilizing Math Course Grades to Limit Enrollment (Method 3):-
Requiring minimum B grades in MATH 1112 and 2510 decreased eligible transfers by 22.3% to 706 students. Further raising the grade standard to B+ yielded a 50% reduction. Targeting these challenging math courses helped dial in the desired enrollment impacts.

DATA CLEANING AND MERGING

Data from multiple semesters was consolidated into a master data frame called all_data to enable analysis across terms.

all_data <- rbind (F21, S21, F22, S22, S23)

This part of the code was used to reassign all instances of 'TA' and 'TA-' grades to 'A' and 'A-' respectively. The B's and C's got reassigned based on the same logic.

all_data [all_data == 'TA'] <- 'A', all_data [all_data == 'TA-'] <- 'A-'

The raw grade data included numerous permutations of the same letter grade spanning across various semesters (e.g., "B," "B+," "P+ (B")," "B (P+")," "B- (P+")," and so on). The summary grade tables were cleaned by combining all variations of a grade into one.

all_data [all_data == "P+ (B)"] <- "B", all_data [all_data == "P+ (B-)"] <- "B-"
all_data [all_data == "B (P+)"] <- "B", all_data [all_data == "B- (P+)"] <- "B-"
all_data [all_data == "P (B-)"] <- "B-", all_data [all_data == "B+ (SU20)"] <- "B+"

A few rarely occurring grades like AP, Exempt Exam, and T were removed during data cleaning due to insufficient sample sizes for analysis.

METHOD 1:- "Analyzing Student GPA Data to Identify Minimum Cut-offs for Transfer Eligibility"

The key elements I would aim to convey in the heading are:
- The goal of using GPA thresholds to reduce eligibility
- The data-driven approach of analyzing actual student GPA distributions
- The modelling of different percentage reductions (25% and 50%)
- The focus on the transfer student population specifically

The code calculates the number of transfer students that would be ineligible to achieve enrollment reductions of 25% and 50% of the students and transfer_students has a total students of 908.

transfer_students_25 <- transfer_students * 0.25, transfer_students_50 <- transfer_students * 0.50

In other words, the minimum GPA threshold has to be set such that 227 & 454 fewer students
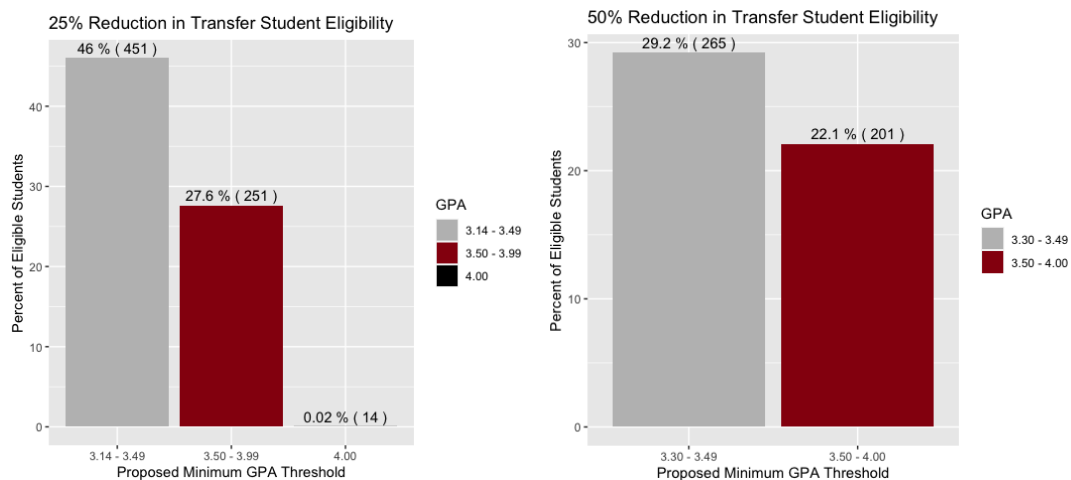
The aim was to restrict transfer students by raising the bar on the GPA required to get into the program. Tougher standards mean fewer students make the cut. The magic number was finding the GPA - high enough to trim enrollment by 25% and 50%, but not so high it shuts out tons of qualified applicants.

For a 25% enrollment cut, the magic minimum GPA was 3.14 - any lower would slash enrollment too drastically. To cut enrollment to a 50% reduction, the minimum GPA needed to be set at 3.30. The sample code for the 25% reduction is below and focuses on the ranges between 3.14 - 3.50 and 3.50 - 4.0 where large clusters of students fell.

sum (all_data$`Cumulative GPA` < 4.00 & all_data$`Cumulative GPA` > 3.50)
sum (all_data$`Cumulative GPA` <= 3.50 & all_data$`Cumulative GPA` >= 3.14)
mean (all_data$`Cumulative GPA` < 4.00 & all_data$`Cumulative GPA` > 3.50)
mean (all_data$`Cumulative GPA` <= 3.50 & all_data$`Cumulative GPA` >= 3.14)

Visualizing the different GPAs with stacked bars made it crystal clear what the impact would be. Labels show how many students and what percentages could still make it over each higher GPA hurdle. In plain terms, the data revealed the minimum GPAs needed to achieve the desired downsizing in the transfer size.

data_GPA <- data.frame(GPA = c("3.14", "3.50", "4.00"), Percent_Eligible = c(46.0, 27.6, 0.02), Count_Eligible = c(451,251,14))

ggplot(data_GPA, aes(x=GPA, y=Percent_Eligible, fill = GPA)) + geom_col() +
        geom_text(aes(label=paste(Percent_Eligible,"% (", Count_Eligible,")")), vjust=-0.5)

METHOD 2:- "Modelling a 25% and 50% Reduction with Mandatory ECON 2010 and 2020"

The enclosed R code comprises grades from ECON 2010 and ECON 2020 courses to create data frames prescribing the reduction in the number of transfer students by approximately 25% and close to 50%. The sample code for the 25% reduction is below.

Creating a subset data frame 'ECON_25' to model an approximately 25% reduction in the number of transfer students.

ECON_25 <- subset (all_data, all_data$`ECON 2010` != "B-" & all_data$`ECON 2020` != "B-" & all_data$`ECON 2020` != "C+")

Generating frequency tables for 'ECON 2010' and 'ECON 2020' column values within the 'ECON_25' data frame

table (ECON_25$`ECON 2010`), table (ECON_25$`ECON 2020`)

Counting the number of rows in the 'ECON_25' data frame by using the command nrow(ECON_25)

By adjusting the required grade from "B-" or higher to a "B" or higher for the ECON 2010 and ECON 2020 courses, the ECON_25 data frame reveals a 25% decrease in the number of transfer students. This change led to a total reduction in transfer students, going down from 908 to 679 students.

Although the process and code for filtering remained unchanged, there was a shift in grading criteria. By elevating the required grade benchmark from "B-" to "B+" or higher for ECON 2010 and ECON 2020 courses, the ECON_50 data frame illustrates a nearly 50% decrease in transfer students. Consequently, the count of transfer students decreased significantly from 908 to 436, marking a substantial 52% reduction.

We proposed these adjustments to the ECON grade standards because there was a significant range in the grades of transfer students taking ECON 2010 and ECON 2020 courses. By setting specific standards for these courses, we aimed to achieve the targeted reduction in the number of transfer students, all while keeping their GPA and other factors intact.

METHOD 3:- "Modelling a 25% and 50% Reduction with Mandatory MATH 1112 and MATH 2510"

After observing significant variations in grades among transfer students taking MATH 1112 and MATH 2510 courses, we put forth these changes to the MATH grading criteria. The sample code for the 25% reduction is below.

Creating a subset data frame 'MATHS_25' to model an approximately 25% reduction in the number of transfer students.

MATHS_25 <- subset(all_data, all_data$`MATH 1112` != "B-" & all_data$`MATH 1112` != "C+" & all_data$`MATH 2510` != "B-")

Generating frequency tables for 'MATH 1112 and 'MATH 2510' column values within the 'ECON_25' data frame

table (MATHS_25$`MATH 1112`), table (MATHS_25$`MATH 2510`)

Counting the number of rows in the 'MATHS_25' data frame by using the command nrow(MATHS_25)

With the help of this R subset function, we sorted through the data to create a subgroup known as 'MATHS_25'. This subgroup, 'MATHS_25', contains 706 entries, which is about 22.34% of the original 908 students. To achieve the intended 25% reduction in transfer students, the associate dean of undergraduate programs could consider raising the acceptable grade to a 'B' for both math courses MATH 1112 & MATH 2510.

The associate dean of undergraduate programs might consider raising the bar for acceptable grades in both math classes MATH 1112 & MATH 2510 to at least a "B+" if they aim for a significant 50% reduction in transfer students. The subset will exclusively consist of students who scored "B+" or higher in either MATH 1112 or MATH 2510. What's intriguing is that the MATHS_50 group is made up of 442 students, which makes up 51% of the initial 908 students. It's worth noting that this strategy relies on the same code used to achieve a 25% reduction initially.

CONCLUSION:-
In simple terms, our analysis provides a plan backed by historical data to address the challenge of reducing the number of transfer students at the Leeds School of Business, aiming for a decrease of 25% and 50%. We dug into GPA requirements, course grades, and enrollment info to find the best strategies. Our findings suggest that tweaking the minimum GPA needed for specific courses like ECON 2010, ECON 2020, MATH 1112, and MATH 2510 can really make a difference in hitting our enrollment goals. These changes offer a well-rounded solution to make course schedules better while still keeping our educational quality intact. And all of this was done through R.

CONTRIBUTIONS:-
Sai Arvind Atluri:- Consolidated data across semesters into one dataset to enable term-wide analysis. After standardizing grades and removing outliers, we calculated overall GPAs for all transfers together. The analysis identified the minimum GPA thresholds needed to reduce transfer eligibility by 25% and 50%. Stacked bar charts clearly illustrated how raising the GPA bar would limit transfer students.

Sandeep Reddy Modugu:- Consolidated data across semesters into one dataset to enable term-wide analysis. The subset tool filtered out students not meeting the minimum grades set for ECON 2010 and 2020. Raising the grade bar resulted in a 25% and 50% reduction in eligible transfers.

Amandeep Bhardwaj:- Consolidated data across semesters into one dataset to enable term-wide analysis. Using the subset tool, we filtered out students who didn't make the minimum grade we set for MATH 1112 and 2510 resulting in a 25% and 50% reduction in eligible transfers.

Team Contribution:- Our team worked together, but each focused on their own method. To keep things simple, we used the same data frame name "all_data" for consistency. As a group, we brainstormed the introduction, data prep, and conclusion sections. The methods and findings were individual contributions tied together through the shared report.