

Anubhav Dubey & Laura Yan  
Professor Reinholtz  
MSBX 5410  
08/12/2023

## **Business Task**

Raise the standards for transferring to Leeds. Reduce the number of transfer students, making the classes easier to schedule. Reduce the number of transfer students by approximately 25% and 50%.

Reduce by 25 % ~ 681 Students  
Reduce by 50% ~ 454 Students

## **Key Stakeholders**

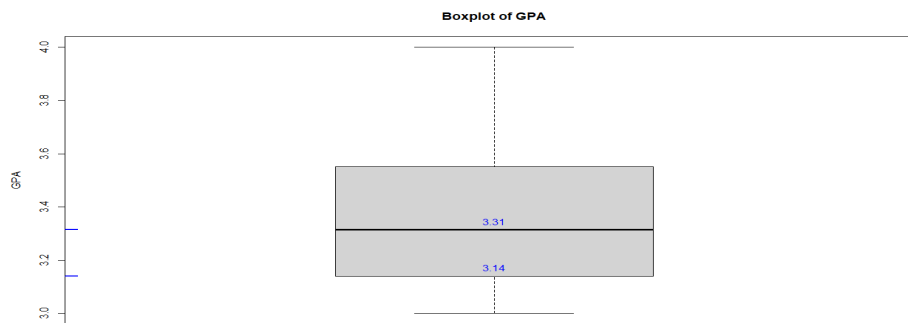
Associate Dean of Undergraduate Programs

## **Phase 1 – Preparing the Data**

- Combined all the semester-wise files containing the student's records, which totaled 908 student records. Used the 'deplyr' R package to bind the files into a single file.
- Cleaned the data by performing the following actions:
  - All the P+ and P grades were converted to B- grades.
  - Changed the grades marked as 'Exempt' and 'Exempt Exam' to AP (Advanced Placement).
  - Changed the grades with a T as a prefix to just the grades.
  - Corrected some grades that were not having proper values e.g., B+ (SU20).
- Added a new column for assigning a range for the Cumulative GPA e.g., '3.0 – 3.1'
- Removed a student record that was having a CGPA less than 3.0 which is the minimum requirement for an admit.

## Phase 2 - Process

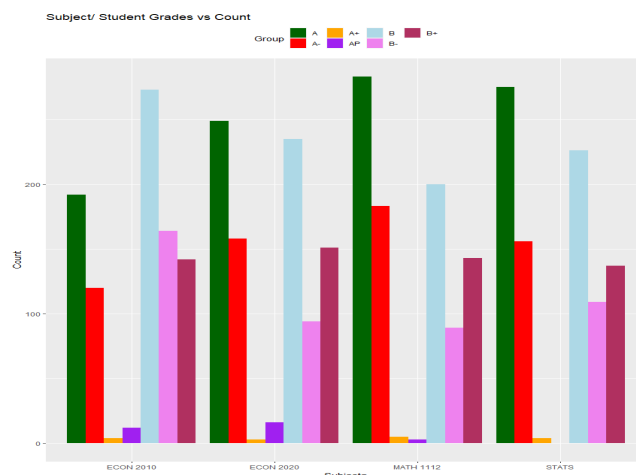
### 1. Method 1 – Student CGPA Quantiles



- Plotted a box plot that shows the quartile values for 0.25 (lower quartile) and 0.50 (middle quartile).
- Filtered the student records to get the number of students above 0.25 and 0.50 quartiles.

```
filtered_df <- allStud[allStud$'Cumulative GPA' > 3.14, ]  
filtered_df  
sum(filtered_df$row_count)
```

### 2. Method 2 – Subject-Wise Grades vs Count



- Used ggplot2 to plot a grouped bar plot to see the subject-wise grade counts.
- Aggregated the GPA and Grades and filtered out the possibilities of the grades the students have.  

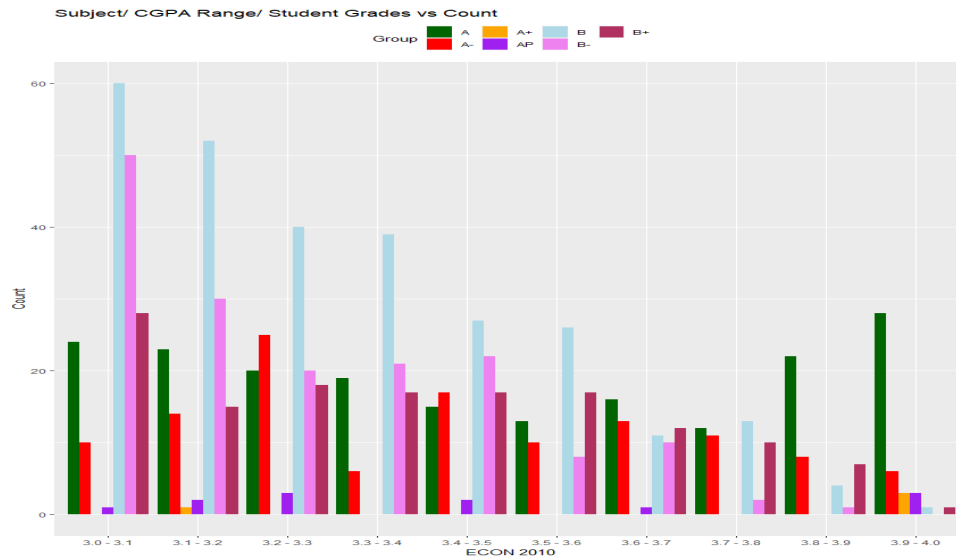
```
allStud$row_count <- 1  
aggregated_data <- aggregate (row_count ~ `ECON 2010` + `ECON 2020` + `MATH 1112` + `MATH 2510`, FUN = length, data = allStud)
```
- Filtered the aggregated data to check for which course can have its minimum grade increased to get the desired standards in admits.

```

filtered_df <- aggregated_data %>%
  filter (`ECON 2010` %in% c("B","B+","A-","A","A+","AP") &
    `ECON 2020` %in% c("B","B+","A-","A","A+","AP") &
    `MATH 1112` %in% c("B-","B","B+","A-","A","A+","AP") &
    `MATH 2510` %in% c("B-","B","B+","A-","A","A+","AP"))

```

### 3. Method 3 – CGPA Range/ Grades vs Count



- Using a grouped bar plot to check the subject-wise/ GPA range grade counts.
- Aggregated the GPA and Grades to filter the possibilities of grade combinations and the GPA of the student.
- Filtered the aggregated data to get the combinations of the grades in each subject and where to increase the minimum requirement along with how to balance it with increasing the minimum CGPA to not have a huge jump in the standards for a particular subject or in the minimum CGPA required.

```

filtered_df <- aggregated_data %>%
  filter (`ECON 2010` %in% c("B","B+","A-","A","A+","AP") & # `ECON 2010` >B-
    `ECON 2020` %in% c("B-","B","B+","A-","A","A+","AP") &
    `MATH 1112` %in% c("B-","B","B+","A-","A","A+","AP") &
    `MATH 2510` %in% c("B-","B","B+","A-","A","A+","AP") &
    `Cumulative GPA` >3.05) # Cumulative GPA` >3.05

```

### Phase 3 – Analyze

1. Method 1
  - Increasing the minimum CGPA required to ~3.14 results in a reduction of admits by 25.13%. If the requirement is increased to ~3.31 it gives a 49.94% reduction in admits.
2. Method 2
  - Increasing the minimum grade required for ECON 2010 and ECON 2020 to B results in 25.35% of the student admits.
  - Changing the minimum grade for MATH 1112 to a B+ and the rest of the courses to a B reduced the admits by 50.6 %.
3. Method 3
  - By increasing the minimum grade requirement for ECON 2010 to a B and a minimum CGPA of ~3.05 we can reduce the admission by 25.24%.
  - We can cut the admits to 50.27% by increasing the minimum grade requirement for all the courses to a B and a minimum CGPA of ~3.15.

### Phase 4 - Recommendations

- Increasing the minimum CGPA to 3.15 may lead to a reduction in admit by 25% or 3.30 to get a reduction of 50% resulting in higher standards of admits.
- Increasing the grade requirement for ECON 2010 and ECON 2020 to a B may result in a 25% reduction of admits or MATH 1112 to a B+ along with B for the rest of the courses for a 50% reduction in admits.
- To raise the standards without having to either only raise the minimum grade or CGPA requirement, we can have a balance between both by increasing the minimum grade for ECON 2010 and CGPA to get a 25% reduction in the admits or increasing minimum grade for all the courses to a B and CGPA to 3.15 to get a 50% reduction.

### Team Contribution

- **Laura:** She worked on method 1 where she created a box plot showing the 0.25 and 0.50 quartile values to get insight on what the minimum CGPA should be in order to reduce the admits by 25% and 50%. Worked on aggregating the data and filtering based on the grades and CGPA that gave the desired numbers of students.
- **Anubhav:** He worked on cleaning and preparing the data using R (deplyr) and deduced methods 2 and 3 where he analyzed the grades and CGPA data on how the minimum requirement can be raised in order to get the desired results. He plotted the grouped bar plots for the methods to show the subject/ grades vs count and subject/ CGPA range/ grades vs count.

## TECHNICAL APPENDIX

### Plots

#### 1. Method 1 – CGPA Box Plot

```
# To get all the CGPA values for all the students
gpa <- allStud$'Cumulative GPA'

# Box Plot
boxplot (gpa, main = "Boxplot of GPA", ylab = "GPA")

# To calculate the quartile values
percentile_values <- quantile (gpa, probs = c (0.25,0.50)) #quantile

# Box Plot Quantile Marking on the Y-axis
rug(percentile_values,side=2,col="blue",lwd=2)

# Box Plot Quantile Values Text on the Plot
text (x = c (1, 1), y = percentile_values, labels = round (percentile_values, 2), pos = 3, col =
'blue')
```

#### 2. Method 2 - Subject/ Student Grades vs Count

```
# Creating a Data Frame having Group, Subgroup, and Value
df <- data.frame(
  Group = c ('A+', 'A+', 'A+', 'A+', 'A', 'A', 'A', 'A', 'A-', 'A-', 'A-', 'A-', 'B+', 'B+', 'B+', 'B+', 'B', 'B', 'B',
'B', 'B-', 'B-', 'B-', 'B-', 'AP', 'AP', 'AP', 'AP'),

  Subgroup = c(names(allStud) [2], names(allStud) [3], names(allStud) [4], 'STATS'),

  Value = c (length (allStud$'ECON 2010'[allStud$'ECON 2010' == 'A+']),
    length (allStud$'ECON 2020'[allStud$'ECON 2020' == 'A+']),
    length (allStud$'MATH 1112'[allStud$'MATH 1112' == 'A+']),
    length (allStud$'MATH 2510'[allStud$'MATH 2510' == 'A+']),

    length (allStud$'ECON 2010'[allStud$'ECON 2010' == 'A']),
    length (allStud$'ECON 2020'[allStud$'ECON 2020' == 'A']),
    length (allStud$'MATH 1112'[allStud$'MATH 1112' == 'A']),
    length (allStud$'MATH 2510'[allStud$'MATH 2510' == 'A']),

    length (allStud$'ECON 2010'[allStud$'ECON 2010' == 'A-']),
    length (allStud$'ECON 2020'[allStud$'ECON 2020' == 'A-']),
    length (allStud$'MATH 1112'[allStud$'MATH 1112' == 'A-']),
```

```

length (allStud$'MATH 2510'[allStud$'MATH 2510' == 'A-']),

length (allStud$'ECON 2010'[allStud$'ECON 2010' == 'B+']),
length (allStud$'ECON 2020'[allStud$'ECON 2020' == 'B+']),
length (allStud$'MATH 1112'[allStud$'MATH 1112' == 'B+']),
length (allStud$'MATH 2510'[allStud$'MATH 2510' == 'B+']),

length (allStud$'ECON 2010'[allStud$'ECON 2010' == 'B']),
length (allStud$'ECON 2020'[allStud$'ECON 2020' == 'B']),
length (allStud$'MATH 1112'[allStud$'MATH 1112' == 'B']),
length (allStud$'MATH 2510'[allStud$'MATH 2510' == 'B']),

length (allStud$'ECON 2010'[allStud$'ECON 2010' == 'B-']),
length (allStud$'ECON 2020'[allStud$'ECON 2020' == 'B-']),
length (allStud$'MATH 1112'[allStud$'MATH 1112' == 'B-']),
length (allStud$'MATH 2510'[allStud$'MATH 2510' == 'B-']),

length (allStud$'ECON 2010'[allStud$'ECON 2010' == 'AP']),
length (allStud$'ECON 2020'[allStud$'ECON 2020' == 'AP']),
length (allStud$'MATH 1112'[allStud$'MATH 1112' == 'AP']),
length (allStud$'MATH 2510'[allStud$'MATH 2510' == 'AP']))
)

# Grouped Bar Plot
ggplot(df, aes(x=Subgroup, y= Value, fill= Group )) +
  geom_bar(stat="identity", position="dodge") +
  scale_fill_manual(values=c("darkgreen", "red", "orange", "purple", "lightblue", "violet", "maroon"))+
  labs (title="Subject/ Student Grades vs Count", x = "Subjects", y = "Count") +
  theme(legend.position="top")

```

### 3. Method 3 – Subject/ CGPA Range/ Grades vs Count

```

# Creating a Data Frame having Group, Subgroup, and Value
df <- data.frame(
  Group = c ('A+', 'A+', 'A+', 'A+', 'A+', 'A+', 'A+', 'A+', 'A+', 'A+', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A', 'A',
    'A', 'A-', 'A-', 'A-', 'A-', 'A-', 'A-', 'A-', 'A-', 'A-', 'B+', 'B+', 'B+', 'B+', 'B+', 'B+', 'B+', 'B+', 'B+', 'B+',
    'B+', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B', 'B-', 'B-', 'B-', 'B-', 'B-', 'B-', 'B-', 'B-', 'B-', 'B-', 'AP', 'AP',
    'AP', 'AP', 'AP', 'AP', 'AP', 'AP', 'AP', 'AP'),
  Subgroup = c(unique(allStud$Category) [2], unique(allStud$Category) [3],
    unique(allStud$Category) [1], unique(allStud$Category) [5],
    unique(allStud$Category) [7], unique(allStud$Category) [9],
    unique(allStud$Category) [6], unique(allStud$Category) [8],
    unique(allStud$Category) [10], unique(allStud$Category) [4]),

```

[illegible]

```

length (allStud$Category[allStud$Category == '3.0 - 3.1' & allStud$'ECON 2010' == 'B']),
length (allStud$Category[allStud$Category == '3.1 - 3.2' & allStud$'ECON 2010' == 'B']),
length (allStud$Category[allStud$Category == '3.2 - 3.3' & allStud$'ECON 2010' == 'B']),
length (allStud$Category[allStud$Category == '3.3 - 3.4' & allStud$'ECON 2010' == 'B']),
length (allStud$Category[allStud$Category == '3.4 - 3.5' & allStud$'ECON 2010' == 'B']),
length (allStud$Category[allStud$Category == '3.5 - 3.6' & allStud$'ECON 2010' == 'B']),
length (allStud$Category[allStud$Category == '3.6 - 3.7' & allStud$'ECON 2010' == 'B']),
length (allStud$Category[allStud$Category == '3.7 - 3.8' & allStud$'ECON 2010' == 'B']),
length (allStud$Category[allStud$Category == '3.8 - 3.9' & allStud$'ECON 2010' == 'B']),
length (allStud$Category[allStud$Category == '3.9 - 4.0' & allStud$'ECON 2010' == 'B']),

length (allStud$Category[allStud$Category == '3.0 - 3.1' & allStud$'ECON 2010' == 'B-']),
length (allStud$Category[allStud$Category == '3.1 - 3.2' & allStud$'ECON 2010' == 'B-']),
length (allStud$Category[allStud$Category == '3.2 - 3.3' & allStud$'ECON 2010' == 'B-']),
length (allStud$Category[allStud$Category == '3.3 - 3.4' & allStud$'ECON 2010' == 'B-']),
length (allStud$Category[allStud$Category == '3.4 - 3.5' & allStud$'ECON 2010' == 'B-']),
length (allStud$Category[allStud$Category == '3.5 - 3.6' & allStud$'ECON 2010' == 'B-']),
length (allStud$Category[allStud$Category == '3.6 - 3.7' & allStud$'ECON 2010' == 'B-']),
length (allStud$Category[allStud$Category == '3.7 - 3.8' & allStud$'ECON 2010' == 'B-']),
length (allStud$Category[allStud$Category == '3.8 - 3.9' & allStud$'ECON 2010' == 'B-']),
length (allStud$Category[allStud$Category == '3.9 - 4.0' & allStud$'ECON 2010' == 'B-']),

length (allStud$Category[allStud$Category == '3.0 - 3.1' & allStud$'ECON 2010' == 'AP']),
length (allStud$Category[allStud$Category == '3.1 - 3.2' & allStud$'ECON 2010' == 'AP']),
length (allStud$Category[allStud$Category == '3.2 - 3.3' & allStud$'ECON 2010' == 'AP']),
length (allStud$Category[allStud$Category == '3.3 - 3.4' & allStud$'ECON 2010' == 'AP']),
length (allStud$Category[allStud$Category == '3.4 - 3.5' & allStud$'ECON 2010' == 'AP']),
length (allStud$Category[allStud$Category == '3.5 - 3.6' & allStud$'ECON 2010' == 'AP']),
length (allStud$Category[allStud$Category == '3.6 - 3.7' & allStud$'ECON 2010' == 'AP']),
length (allStud$Category[allStud$Category == '3.7 - 3.8' & allStud$'ECON 2010' == 'AP']),
length (allStud$Category[allStud$Category == '3.8 - 3.9' & allStud$'ECON 2010' == 'AP']),
length (allStud$Category[allStud$Category == '3.9 - 4.0' & allStud$'ECON 2010' == 'AP'])
)
)

# Grouped Bar Plot
ggplot(df, aes(x=Subgroup, y= Value, fill=Group)) +
  geom_bar(stat="identity", position="dodge") +
  scale_fill_manual(values=c("darkgreen", "red", "orange", 'purple', 'lightblue', 'violet', 'maroon'))+
  labs (title="Subject/ CGPA Range/ Student Grades vs Count ", x = "ECON 2010", y = "Count") +
  theme(legend.position="top")

```