Student Feedback Analysis

Byunghyun Hwang

Introduction

This study investigates the feedback collected from undergraduate and postgraduate students across multiple courses and its relation with several self-reported variables.

It answers the following **research questions**:

1. On average, do students study more than 12 hours/week for each unit?
2. On average, is the satisfaction score above 6.5?
3. Did satisfaction change after course redesign?
4. Analyse satisfaction and gender: a. Does satisfaction vary with gender? b. Is satisfaction different between males and females?
5. Does satisfaction differ by course?
6. What is the effect of gender and level of education on satisfaction? Is there an interaction between gender and level of education?
7. Is the proportion of males equal to that of females?
8. Are course and recommendation in the male population associated?
9. Are gender and perceived improvement associated?
10. What are the key predictors of students' satisfaction score? a. Find the best linear model to explain students' satisfaction score. b. Comment on the effect size of the significant predictor c. Comment on how much of the variation in satisfaction can be explained by the significant predictors.

Data Description

The dataset includes 200 observations of the following variables:

- `student_id`: Participant identifier.
- `course`: One of "Data Science", "Business", "Psychology", "Engineering".
- `level`: One of "Undergraduate", "Postgraduate".
- `gender`: One of "Male", "Female", "Other answer".
- `study_hours`: Weekly hours of study per unit.
- `satisfaction_score`: The satisfaction score by course.
- `recommended`: Would the student recommend this course to other students? Yes/No.
- `improved`: Perceived improvement in student's understanding: is it at the expected level? Yes/No.
- `before_vs_after_score`: The difference in satisfaction score after a redesign.

Exploratory Data Analysis

Summary Statistics

Table 1: Summary of Student Feedback by Gender (Mean)

Characteristic	**Overall** N = 2,000	**Female** N = 981	**Male** N = 929	**Other** N = 90
___course___				
Business	504 (25%)	251 (26%)	229 (25%)	24 (27%)
Data Science	463 (23%)	229 (23%)	217 (23%)	17 (19%)
Engineering	527 (26%)	255 (26%)	240 (26%)	32 (36%)
Psychology	506 (25%)	246 (25%)	243 (26%)	17 (19%)
___level___				
Postgraduate	995 (50%)	467 (48%)	478 (51%)	50 (56%)
Undergraduate	1,005 (50%)	514 (52%)	451 (49%)	40 (44%)
___study_hours___				
Mean (SD)	14.1 (8.8)	14.3 (9.0)	14.0 (8.8)	13.9 (7.4)
___satisfaction_score___				
Mean (SD)	7.0 (1.5)	7.0 (1.5)	7.0 (1.5)	7.1 (1.4)
___recommended___	1,408 (70%)	680 (69%)	667 (72%)	61 (68%)
___improved___	993 (50%)	481 (49%)	469 (50%)	43 (48%)
___before_vs_after_score___				
Mean (SD)	0.0 (1.0)	0.0 (1.0)	0.0 (0.9)	0.1 (1.0)

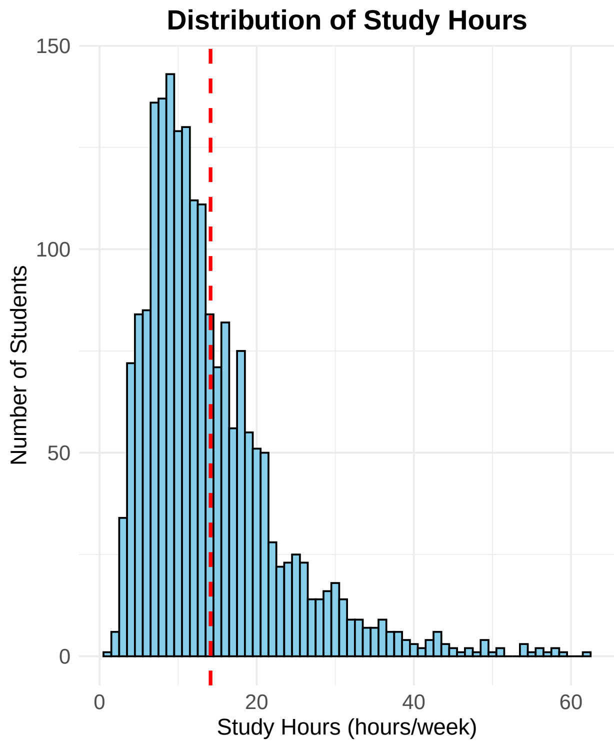
The statistics table by gender shows that the average student satisfaction score was 7. Approximately 70% of students answered that they would recommend the courses. Half of students felt that they had improved. Then, females' satisfaction is slightly higher than males' satisfaction. Furthermore, it seems that there was no significant improvement after course redesign.

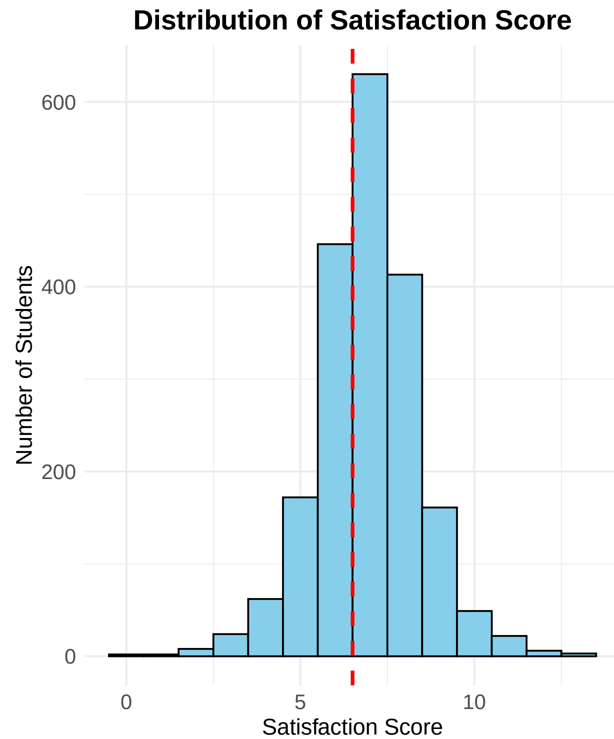
Table 2: Summary of Student Feedback by Course (Mean)

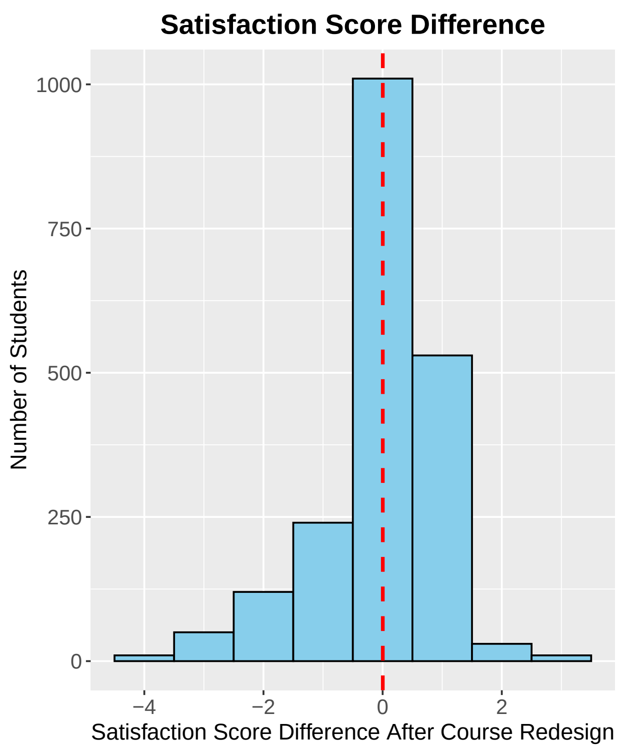
Characteristic	**Overall** N = 2,000	**Business** N = 504	**Data Science** N = 463	**Engineering** N = 527	**Psychology** N = 506
___level___					
Postgraduate	995 (50%)	239 (47%)	244 (53%)	256 (49%)	256 (51%)
Undergraduate	1,005 (50%)	265 (53%)	219 (47%)	271 (51%)	250 (49%)
___gender___					
Female	981 (49%)	251 (50%)	229 (49%)	255 (48%)	246 (49%)
Male	929 (46%)	229 (45%)	217 (47%)	240 (46%)	243 (48%)
Other	90 (4.5%)	24 (4.8%)	17 (3.7%)	32 (6.1%)	17 (3.4%)
___study_hours___					
Mean (SD)	14.1 (8.8)	13.6 (8.7)	14.7 (9.0)	14.3 (8.9)	13.9 (8.8)
___satisfaction_score___					
Mean (SD)	7.0 (1.5)	6.8 (2.2)	7.5 (0.8)	6.5 (0.9)	7.2 (1.4)
___recommended___	1,408 (70%)	370 (73%)	320 (69%)	367 (70%)	351 (69%)
___improved___	993 (50%)	251 (50%)	231 (50%)	275 (52%)	236 (47%)
___before_vs_after_score___					
Mean (SD)	0.0 (1.0)	0.0 (1.0)	0.0 (1.0)	0.0 (0.9)	0.0 (0.9)

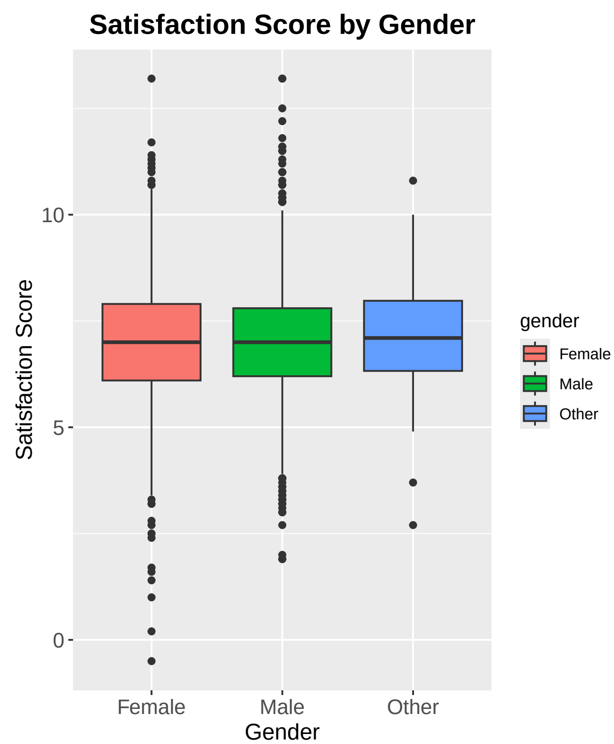
The statistics table by course shows that data science has the highest average satisfaction score of 7.5, while engineering has the lowest average satisfaction score of 6.5. Approximately 70% of students recommend the courses. Particularly, students in business tend to recommend the course relatively more often, compared to students in the other courses. Notably, students in engineering exhibited slightly a higher perceived improvement rate of 52%. In contrast, students in psychology exhibited a slightly lower perceived improvement of 47%.

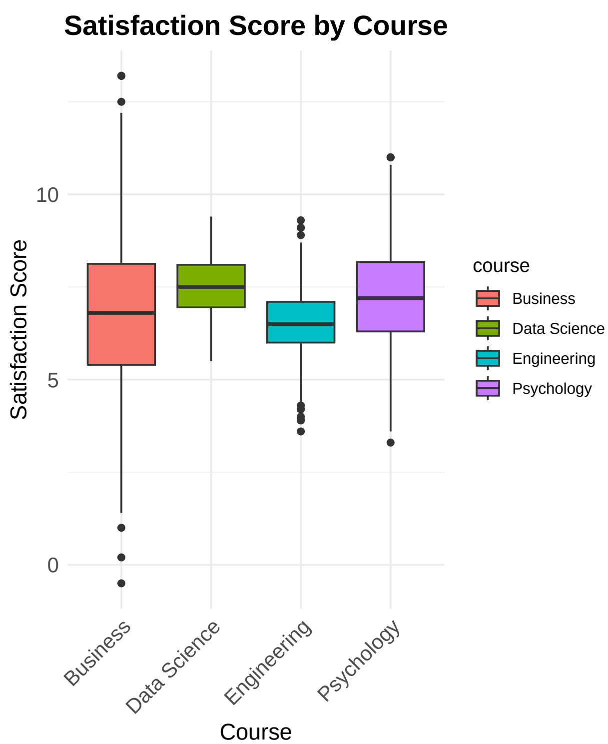
Visual Exploration

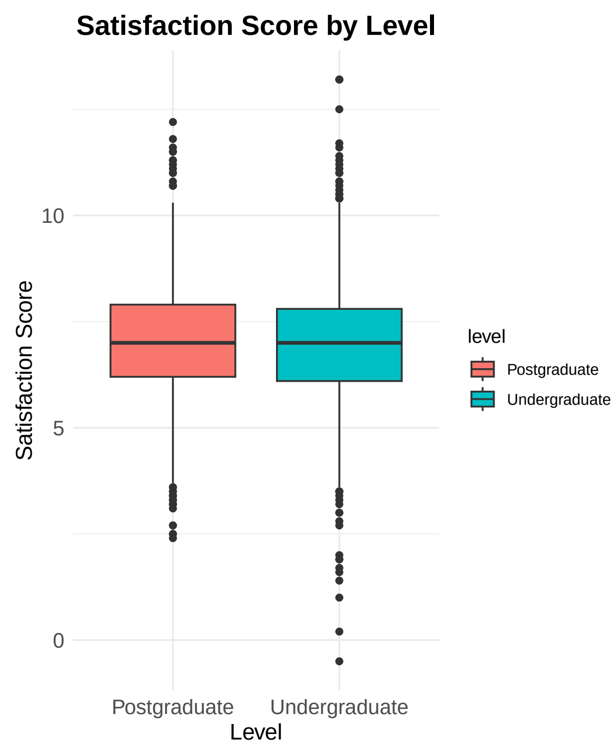


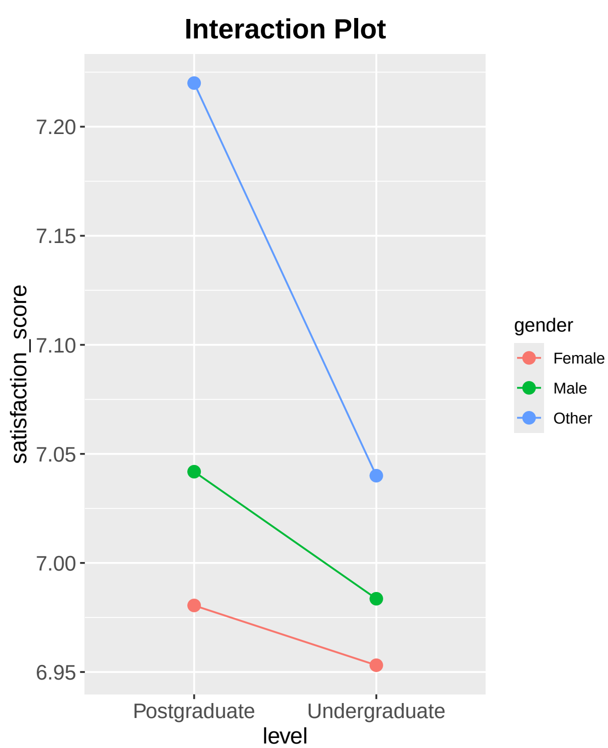


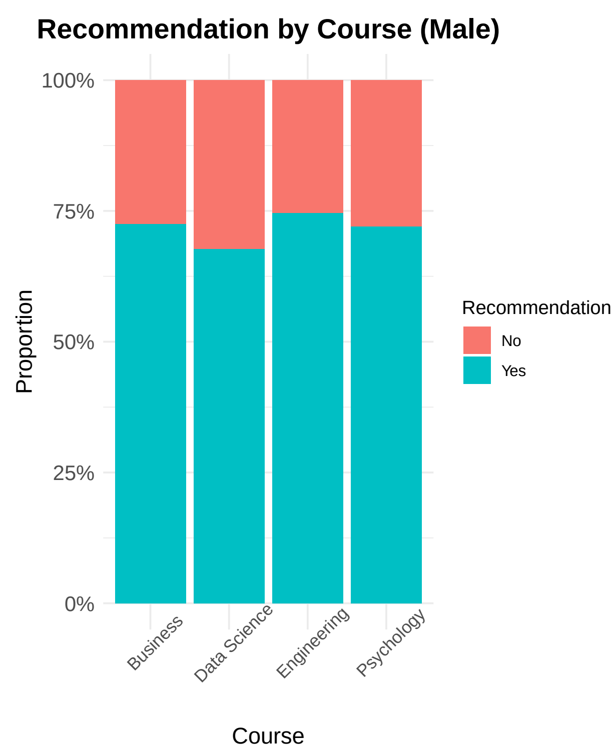


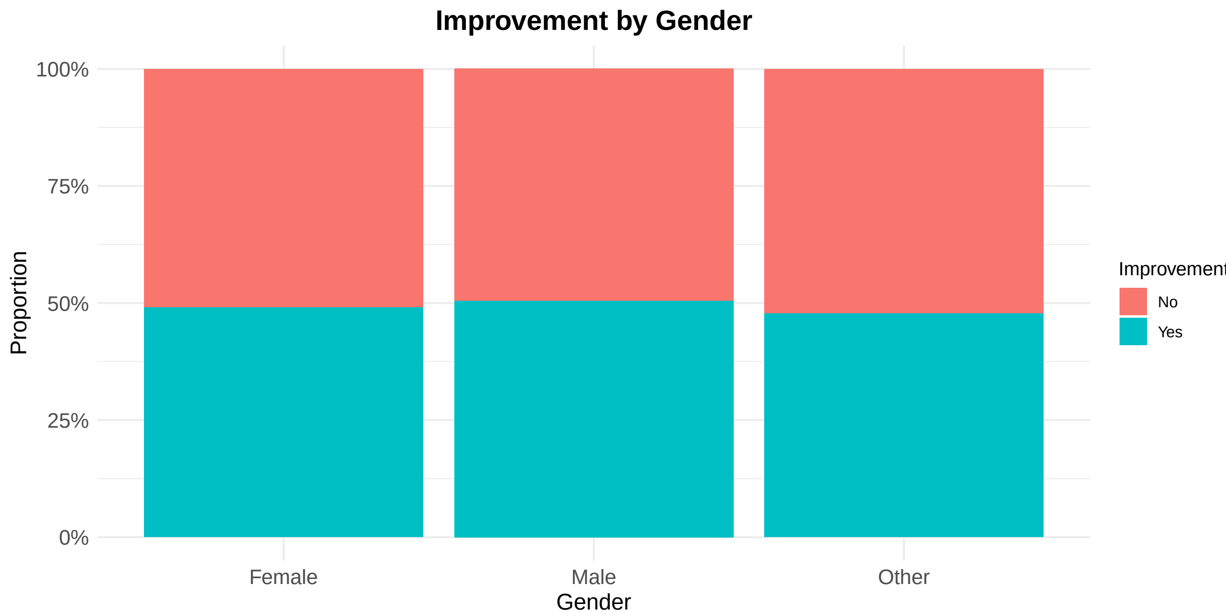



Satisfaction Score by Course
Satisfaction Score
10
5
0
Business
Data Science
Engineering
Psychology
Course
course
Business
Data Science
Engineering
Psychology


Satisfaction Score by Level
Satisfaction Score
10
5
0
Postgraduate
Undergraduate
Level
level
Postgraduate
Undergraduate


Interaction Plot
satisfaction_score
7.20
7.15
7.10
7.05
7.00
6.95
Postgraduate
Undergraduate
level
gender
Female
Male
Other


Recommendation by Course (Male)
Proportion
100%
75%
50%
25%
0%
Business
Data Science
Engineering
Psychology
Course
Recommendation
No
Yes



Results

Do students study more than 12 hours/week/unit on average?

$$H_0 : \mu <= 12 \quad \text{vs.} \quad H_1 : \mu > 12$$

The summary statistics table shows that the average study hours are 14.1, which is greater than 12. A Q-Q plot and a Shapiro-Wilk normality test were conducted to assess if study hours follow a normal distribution (see Appendix 1). As a result ($p < 0.05$), study hours do not follow a normal distribution, making it inappropriate to assess the average study hours. Instead, a Wilcoxon signed rank test was conducted (see Appendix 1). According to the test result ($p < 0.05$), the median study hours are statistically significantly greater than 12.

Is satisfaction score above 6.5 on average?

$$H_0 : \mu <= 6.5 \quad \text{vs.} \quad H_1 : \mu > 6.5$$

The summary statistics table shows that the average satisfaction score is 7, which is greater than 6.5. A Q-Q plot and a Shapiro-Wilk normality test were conducted to assess if satisfaction scores follow a normal distribution (see Appendix 2). As a result ($p < 0.05$), satisfaction scores do not follow a normal distribution, making it inappropriate to assess the average satisfaction score. Instead, a Wilcoxon signed rank test was conducted (see Appendix 2). According to the test result ($p < 0.05$), the median satisfaction score is statistically significantly greater than 6.5.

Did satisfaction change after course redesign?

$$H_0 : \mu_{\text{after}} - \mu_{\text{before}} = 0 \quad \text{vs.} \quad H_1 : \mu_{\text{after}} - \mu_{\text{before}} \neq 0$$

The relative frequencies of student responses regarding satisfaction after course redesign were analyzed (see Appendix 3). 32.5% of students reported decreased satisfaction scores, while 67% of students reported increased satisfaction scores after the course redesign. Then, 0.5% of students reported no changes in satisfaction scores. A Q-Q plot and a Shapiro-Wilk normality test were conducted to assess if the differences in satisfaction scores before and after course redesign follow a normal distribution (see Appendix 3). As a result ($p < 0.05$), the differences in satisfaction scores before and after course redesign do not follow a normal distribution, making it inappropriate to assess the average difference in satisfaction scores before and after course redesign. Instead, a Wilcoxon signed rank test was conducted (see Appendix 3). According to the test result ($p < 0.05$), the median difference in satisfaction scores between before and after course redesign has statistically significantly changed.

Does satisfaction vary with gender? Is it different between males and females?

$$H_0 : \text{median}_{\text{male}} = \text{median}_{\text{female}} = \text{median}_{\text{other}} \text{vs.} \quad H_1 : \text{at least one median}_{\text{i}} \neq \text{median}_{\text{j}}$$

Since students' satisfaction scores do not follow a normal distribution, based on previous analysis, median values were chosen to test the hypothesis. However, normality tests were conducted to confirm the absence of normality. Male and female groups have equal median satisfaction scores of 7, while other group has a satisfaction score of 7.1 (see Appendix 4). Q-Q plots and a Shapiro-Wilk normality test were conducted to assess if the satisfaction scores by gender follow a normal distribution (see Appendix 4). As a result ($p < 0.05$), the satisfaction scores of males and females do not follow a normal distribution. A Kruskal-Wallis rank sum test was conducted (see Appendix 4). According to the result ($p > 0.05$), there are no statistically significant differences in the median satisfaction scores across genders. In addition, a Wilcoxon rank sum test was conducted to assess the difference in the median satisfaction scores between males and females. According to the result ($p > 0.05$), there is no statistically significant difference in the median satisfaction scores between males and females.

Does satisfaction differ by course?

$$H_0 : \text{median}_{\text{male}} = \text{median}_{\text{female}} = \text{median}_{\text{other}} \quad \text{vs.} \quad H_1 : \text{at least one median}_{\text{i}} \neq \text{median}_{\text{j}}$$

Since students' satisfaction scores do not follow a normal distribution based on previous analysis, median values were chosen to test the hypothesis. However, normality tests were conducted to confirm the absence of normality. The median satisfaction scores are 6.8, 7.5, 6.5, and 7.2 for business, data science, engineering, and psychology respectively (see Appendix 5). Q-Q plots and a Shapiro-Wilk normality test were conducted to assess if the satisfaction scores by course follow a normal distribution (see Appendix 5). As a result ($p > 0.05$), the satisfaction scores by course follow a normal distribution, except for the satisfaction scores of data science ($p < 0.05$). Then, a Kruskal-Wallis rank sum test was conducted (see Appendix 5). According to the test result ($p < 0.05$), the median satisfaction scores are statistically significantly different across courses.

What is the effect of gender and level of education on satisfaction? Is there an interaction between gender and level of education?

$$H_0 : \beta_{\text{gender : level}} = 0 \quad \text{vs.} \quad H_1 : \beta_{\text{gender : level}} \neq 0$$

Based on the interaction plot in the EDA section, it seems that there is no interaction between gender and level since the lines do not cross and they are parallel. A Q-Q plot and a Shapiro-Wilk normality test were conducted to assess if the ANOVA residuals follow a normal distribution (see Appendix 6). As a result ($p < 0.05$), the residuals do not follow a normal distribution, suggesting the use of a GLM model. GLM model was selected since residuals do not need to follow a normal distribution and satisfaction scores approximately follow a normal distribution (please refer to the plot of the satisfaction score distribution in the EDM section). According to the GLM result (all p-values > 0.05), gender, level, and their interaction do not have a statistically significant effect on satisfaction. To support the outcome, a two-way ANOVA test was conducted (see Appendix 6). Although its accuracy may be not reliable due to the absence of normality, its result (all p-values > 0.05) also indicates that gender, level, and their interaction do not have a statistically significant effect on satisfaction.

Is the proportion of males equal to that of females?

$$H_0: \; p_{\text{male}} = p_{\text{female}} \quad \text{vs.} \quad H_1: p_{\text{male}} \neq p_{\text{female}}$$

The summary statistics table shows that the total number of females is 981, while the total number of males is 929. A Chi-squared test was conducted (see Appendix 7) and the result ($p > 0.05$) indicates that there is no statistically significant difference in gender proportions.

Are course and recommendation in the male population associated?

$$H_0: \; P(\text{Recommendation} \mid \text{Course}) = P(\text{Recommendation}) \quad \text{vs.}$$
$$H_1: \; P(\text{Recommendation} \mid \text{Course}) \neq P(\text{Recommendation})$$

In the male population, more than half of male students recommend each course (please refer to the plot of recommendation by course in the EDA section). However, a Chi-squared test was conducted (see Appendix 8) and the result ($p > 0.05$) indicates that course and recommendation are not statistically significantly associated.

Are gender and perceived improvement associated?

$$H_0: \; P(\text{Improvement} \mid \text{Gender}) = P(\text{Improvement}) \quad \text{vs.}$$
$$H_1: \; P(\text{Improvement} \mid \text{Gender}) \neq P(\text{Improvement})$$

Approximately half of students perceived improvement across genders (please refer to the plot of improvement by gender in the EDA section). A Chi-squared test was conducted (see Appendix 9) and the result ($p > 0.05$) indicates that there is no statistically significant association between gender and perceived improvement.

What are the key predictors of students' satisfaction score, and how much of the variation in satisfaction can be explained by study habits and demographics?

A linear model was used to identify key predictors for students' satisfaction (see Appendix 10). The key predictor of students' satisfaction is course ($p < 0.05$). Particularly, students in data science and psychology have higher satisfaction on average. The other features do not have statistically significant associations with students' satisfaction. In addition, the interaction of course and gender slightly improved the model performance.

Female students in business showed an average 6.7 satisfaction score ($p < 0.05$). Female students in data science showed a 0.88 points higher average satisfaction score than female students in business ($p < 0.05$). Female students in psychology showed a 0.54 points higher average satisfaction score than female students in business ($p < 0.05$). In addition, male and the interaction of male and data science have marginally significant associations with satisfaction (see Appendix 10). Males in business showed 2.4 points higher average satisfaction score than females in business. Then, the difference in average satisfaction scores between males and females in data science is approximately 0.33 points lower than the difference between males and females in business.

Course and its interaction with gender explain only approximately 6.5% of the variation in satisfaction. For future research, additional predictors that can explain the remaining 93.5% should be identified and incorporated.

Discussion

This study was conducted to find insights into students' satisfaction based on feedback data collected from undergraduate and postgraduate students across multiple courses. Meaningful insights are listed below;

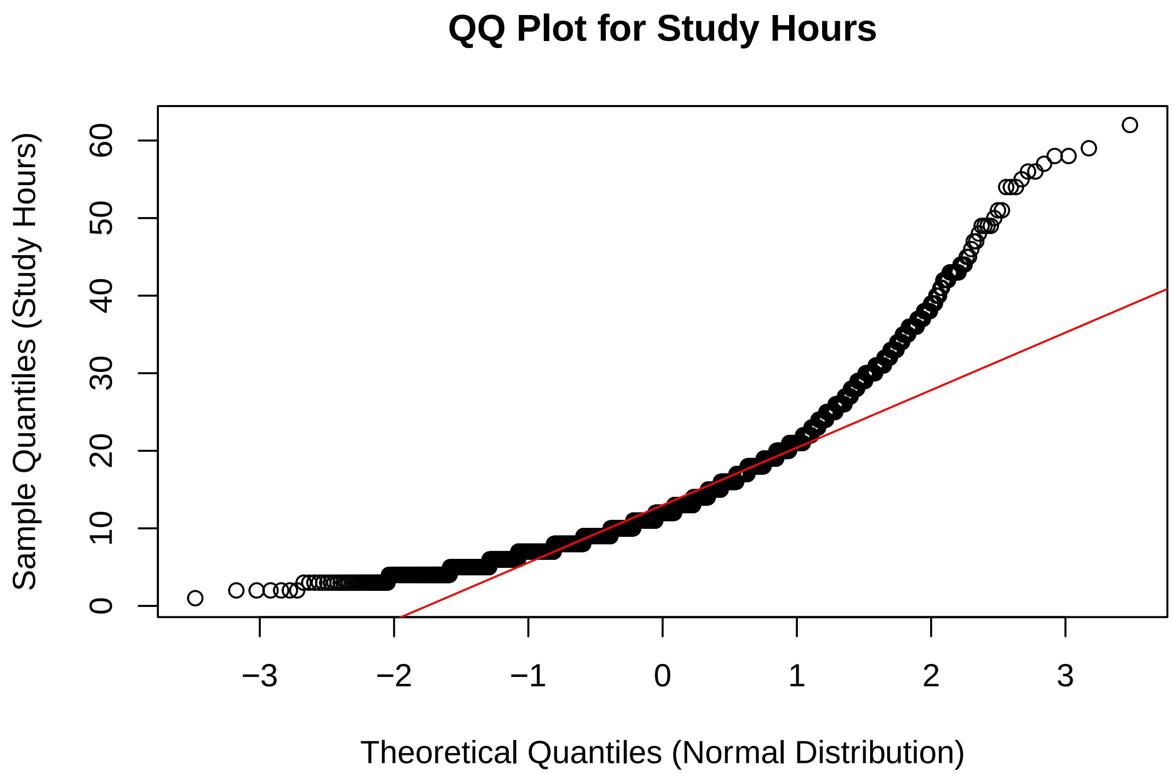
- More than half of students reported satisfaction scores more than 6.5.
- After course redesign, students' satisfaction has changed significantly. 67% of students showed increased satisfaction.
- Students' satisfaction differs across courses. The course with the highest satisfaction is data science and the course with the lowest satisfaction is engineering.
- Students' satisfaction does not differ across genders.
- Gender and education level are not associated with students' satisfaction.
- Gender is not associated with perceived improvement.
- Course is not associated with recommendation.
- The proportions of males and females are equal.
- Course is a key predictor for students' satisfaction.
  - Female students in business showed an average 6.7 satisfaction score.
  - Female students in data science showed a 0.88 points higher average satisfaction score than female students in business.
  - Female students in psychology showed a 0.54 points higher average satisfaction score than female students in business.
  - Males in business showed 2.4 points higher average satisfaction score than females in business.
  - The difference in average satisfaction scores between males and females in data science is approximately 0.33 points lower than the difference between males and females in business.

However, this study has a limitation. Only 6.5% of the variation in satisfaction can be explained by the key predictor and its interaction with gender. The remaining 93.5% can be explained by additional predictors that should be identified and added for future study.

Appendix 1. Do students study more than 12 hours/week on average?

$$H_0 : \mu <= 12 \quad \text{vs.} \quad H_1 : \mu > 12$$

```r
qqnorm(feedback_data$study_hours,
       main = "QQ Plot for Study Hours",
       xlab = "Theoretical Quantiles (Normal Distribution)",
       ylab = "Sample Quantiles (Study Hours)")

qqline(feedback_data$study_hours, col = "red")
```




```r
shapiro.test(feedback_data$study_hours)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  feedback_data$study_hours
## W = 0.86085, p-value < 2.2e-16
```

According to the Q-Q plot and Shapiro-Wilk normality test results, study hours do not follow a normal distribution ($p < 0.05$), indicating the average study hours may be misleading and justifying the use of a Wilcoxon signed rank test.

```r
wilcox.test(feedback_data$study_hours, mu = 12, alternative = "greater")
```

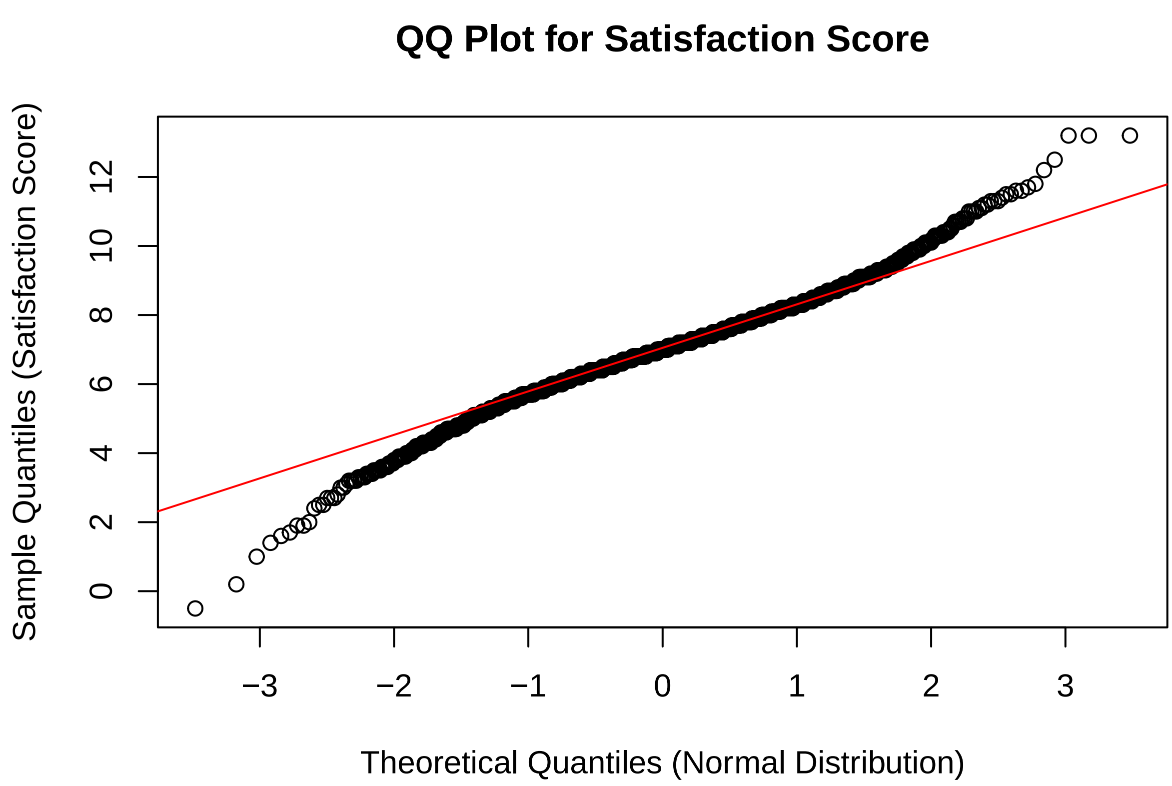
```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  feedback_data$study_hours
## V = 1026039, p-value = 6.708e-09
## alternative hypothesis: true location is greater than 12
```

According to the Wilcoxon signed rank test result, the median study hour is significantly greater than 12 (p < 0.05).

Appendix 2. Is satisfaction score above 6.5 on average?

$$H_0 : \mu <= 6.5 \quad \text{vs.} \quad H_1 : \mu > 6.5$$

```r
qqnorm(feedback_data$satisfaction_score,
       main = "QQ Plot for Satisfaction Score",
       xlab = "Theoretical Quantiles (Normal Distribution)",
       ylab = "Sample Quantiles (Satisfaction Score)")

qqline(feedback_data$satisfaction_score, col = "red")
```



```r
shapiro.test(feedback_data$satisfaction_score)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  feedback_data$satisfaction_score
## W = 0.98237, p-value = 5.208e-15
```

According to the Q-Q plot and Shapiro-Wilk normality test results, satisfaction scores do not follow a normal distribution ($p < 0.05$), indicating that the average satisfaction score may be misleading and justifying the use of a Wilcoxon signed rank test.

```r
wilcox.test(feedback_data$satisfaction_score, mu = 6.5, alternative = "greater")
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  feedback_data$satisfaction_score
## V = 1323507, p-value < 2.2e-16
## alternative hypothesis: true location is greater than 6.5
```

According to the Wilcoxon signed rank test result, the median satisfaction score is statistically significantly greater than 6.5 ($p < 0.05$).

Appendix 3. Did satisfaction change after course redesign?

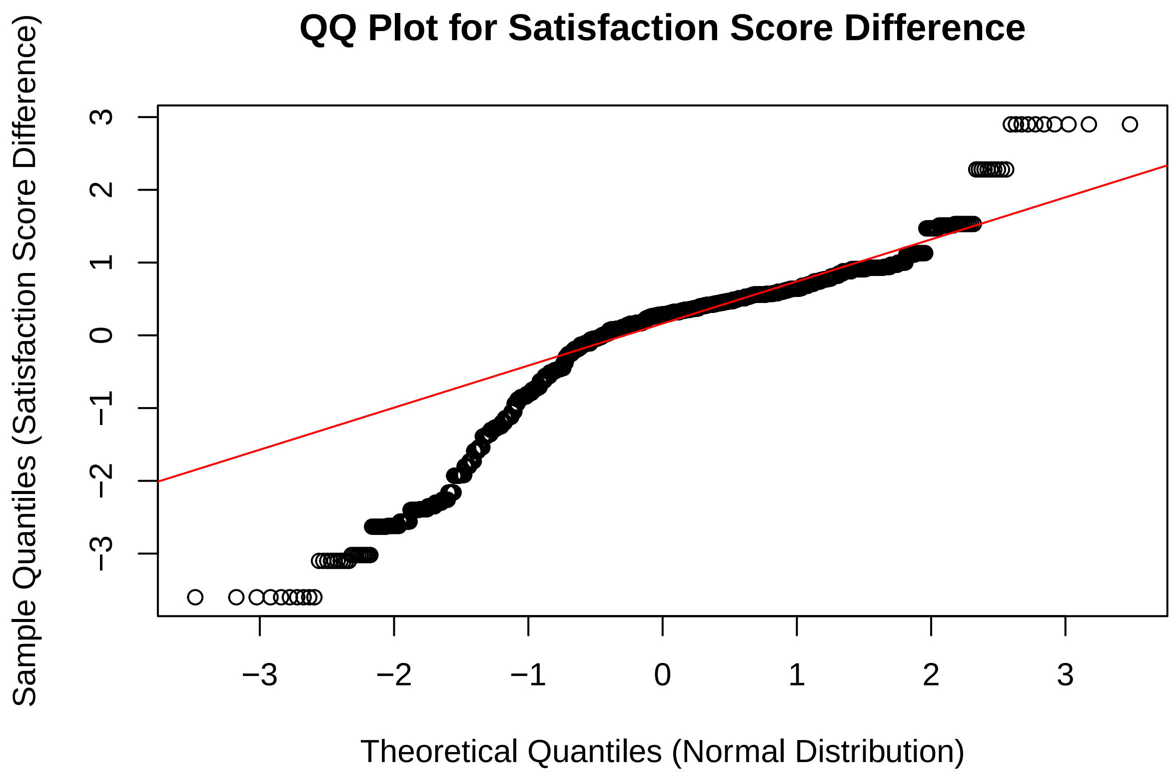
$$H_0 : \mu_{\text{after}} - \mu_{\text{before}} = 0 \quad \text{vs.} \quad H_1 : \mu_{\text{after}} - \mu_{\text{before}} \neq 0$$

```r
satisfaction_change_table <- table(
  Satisfaction_Change_Result =
    ifelse(feedback_data$before_vs_after_score > 0, "Increased",
    ifelse(feedback_data$before_vs_after_score < 0, "Decreased", "No Change"))
)

prop.table(satisfaction_change_table)
```

```
## Satisfaction_Change_Result
## Decreased Increased No Change
##     0.325     0.670     0.005
```

According to the relative frequency table, 32.5% of students showed decreased satisfaction scores and 67% of students showed increased satisfaction scores, while 0.5% of students answered no changes. To assess whether the differences in satisfaction scores between before and after course redesign are statistically significant, a t-test or Wilcoxon signed rank test should be conducted.

```r
qqnorm(feedback_data$before_vs_after_score,
       main = "QQ Plot for Satisfaction Score Difference",
       xlab = "Theoretical Quantiles (Normal Distribution)",
       ylab = "Sample Quantiles (Satisfaction Score Difference)")

qqline(feedback_data$before_vs_after_score,col = "red")
```




```r
shapiro.test(feedback_data$before_vs_after_score)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  feedback_data$before_vs_after_score
## W = 0.86495, p-value < 2.2e-16
```

According to the Q-Q plot and Shapiro-Wilk normality test results, the differences in satisfaction scores between before and after course redesign do not follow a normal distribution ($p < 0.05$), indicating the average difference in satisfaction scores between before and after course redesign may be misleading and justifying the use of a Wilcoxon signed rank test.

```r
wilcox.test(feedback_data$before_vs_after_score, mu = 0, alternative = "two.sided")
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  feedback_data$before_vs_after_score
## V = 1200070, p-value = 2.981e-16
## alternative hypothesis: true location is not equal to 0
```

According to the Wilcoxon signed rank test result, the median difference in satisfaction scores between before and after course redesign is statistically significantly different from zero ($p < 0.05$), suggesting that students' satisfaction scores have changed after course redesign.

Appendix 4. Does satisfaction vary with gender? Is it different between males and females?

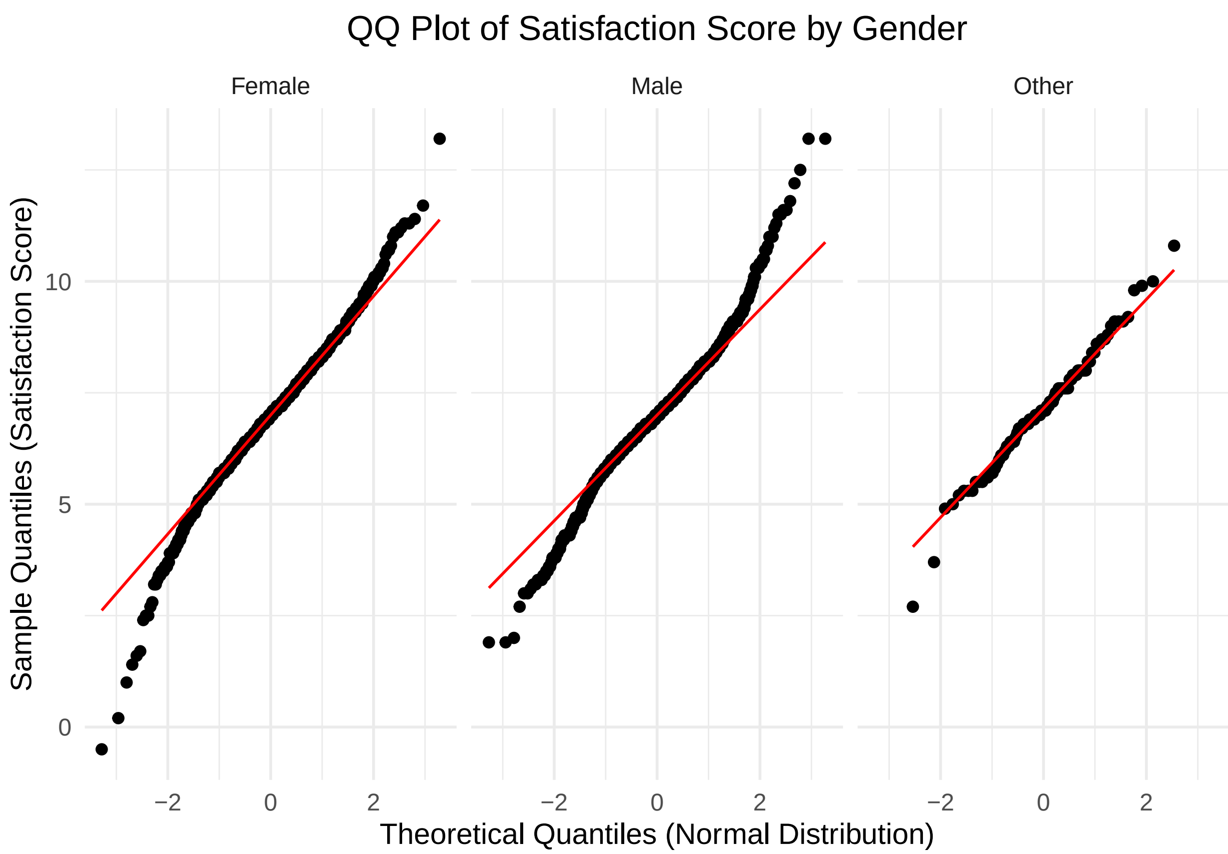
$$H_0 : \text{median}_{\text{male}} = \text{median}_{\text{female}} = \text{median}_{\text{other}} \quad \text{vs.} \quad H_1 : \text{at least one median}_{\text{i}} \neq \text{median}_{\text{j}}$$

```
feedback_data %>%
  group_by(gender) %>%
  summarise(
    count = n(),
    mean_satisfaction_score = mean(satisfaction_score, na.rm = TRUE),
    median_satisfaction_score = median(satisfaction_score, na.rm = TRUE),
  )
```

```
## # A tibble: 3 x 4
##   gender count mean_satisfaction_score median_satisfaction_score
##   <chr>  <int>                   <dbl>                     <dbl>
## 1 Female   981                    6.97                       7
## 2 Male     929                    7.01                       7
## 3 Other     90                    7.14                       7.1
```

Since students' satisfaction scores do not follow a normal distribution based on previous analysis, median values were chosen to test the hypothesis. According to the statistics table, the median satisfaction scores of the three gender categories are 7 or 7.1. To assess whether the median satisfaction scores across genders are statistically significant, a Kruskal-Wallis rank sum test should be conducted.

```
ggplot(feedback_data, aes(sample = satisfaction_score)) +
  stat_qq() +
  stat_qq_line(color = "red") +
  facet_wrap(~ gender) +
  labs(title = "QQ Plot of Satisfaction Score by Gender",
       x = "Theoretical Quantiles (Normal Distribution)",
       y = "Sample Quantiles (Satisfaction Score)") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```




```r
feedback_data %>%
  group_by(gender) %>%
  summarise(p_value = shapiro.test(satisfaction_score)$p.value)
```

```
## # A tibble: 3 x 2
##   gender  p_value
##   <chr>     <dbl>
## 1 Female 1.82e- 9
## 2 Male   5.09e-11
## 3 Other  5.55e- 1
```

Normality tests were conducted to confirm the absence of normality. According to the Q-Q plots and Shapiro-Wilk normality test results, satisfaction scores of females and males do not follow a normal distribution ($p < 0.05$) while satisfaction scores of others follow a normal distribution ($p > 0.05$). This justifies the use of a Kruskal-Wallis rank sum test.

```r
gender_satisfaction_result <- kruskal.test(satisfaction_score ~ gender,
                                           data = feedback_data)

print(gender_satisfaction_result)
```

```
##
##  Kruskal-Wallis rank sum test
##
```

```
## data:  satisfaction_score by gender
## Kruskal-Wallis chi-squared = 1.125, df = 2, p-value = 0.5698
```

According to the Kruskal-Wallis rank sum test result, the null hypothesis is not rejected ($p > 0.05$). The median satisfaction scores are not statistically significantly different across genders.

$$H_0 : \text{median}_{\text{male}} = \text{median}_{\text{female}} \quad \text{vs.} \quad H_1 : \text{median}_{\text{male}} \neq \text{median}_{\text{female}}$$

```r
male_female_df <- subset(feedback_data, gender %in% c("Male", "Female"))

wilcox.test(satisfaction_score ~ gender, data = male_female_df)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  satisfaction_score by gender
## W = 452038, p-value = 0.7627
## alternative hypothesis: true location shift is not equal to 0
```

A Wilcoxon rank sum test result was conducted. According to the result, the null hypothesis is not rejected. The result indicates that the median satisfaction scores of males and females are not statistically significantly different ($p > 0.05$).

Appendix 5. Does satisfaction differ by course?

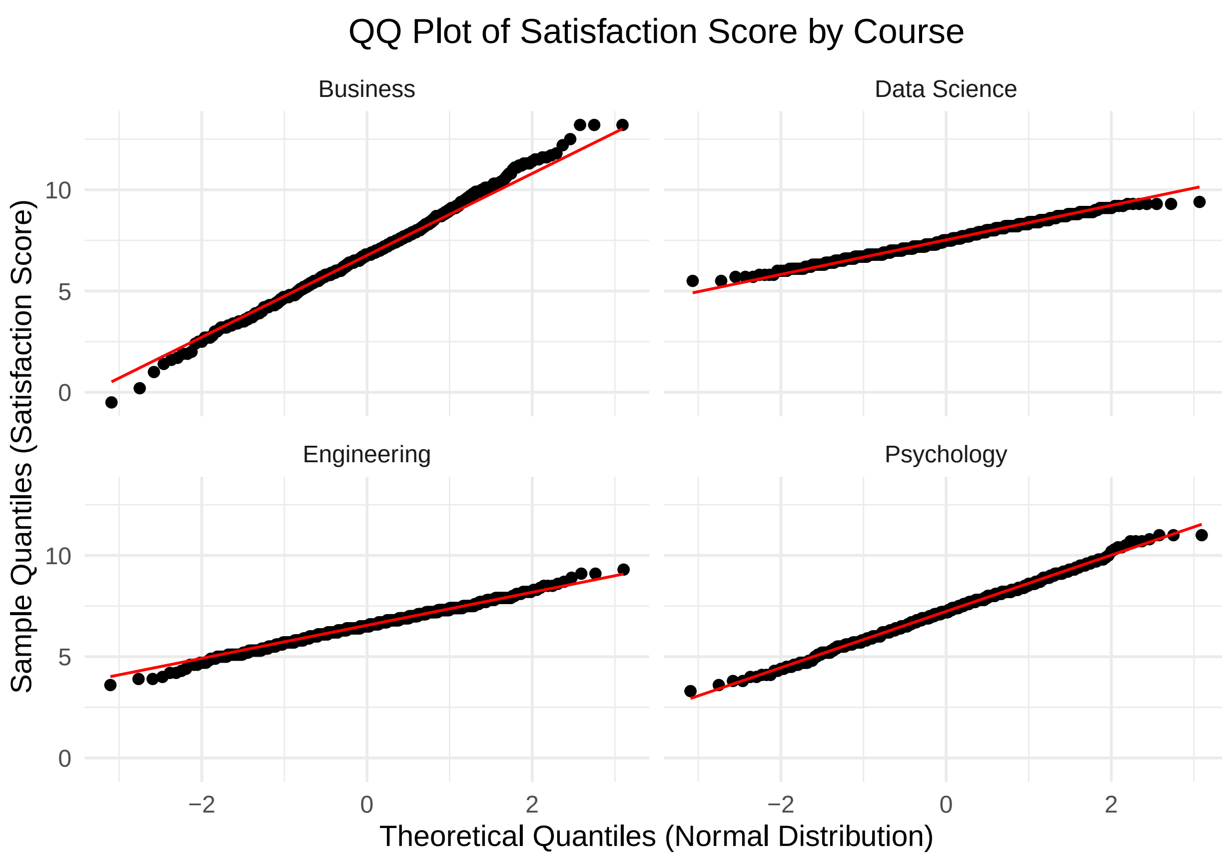
$$\begin{aligned} H_0 :& \quad \text{median}_{\text{DS}} = \text{median}_{\text{BS}} = \text{median}_{\text{PS}} = \text{median}_{\text{EG}} \quad \text{vs.} \\ H_1 :& \quad \text{at least one median}_i \neq \text{median}_j \end{aligned}$$

```r
feedback_data %>%
  group_by(course) %>%
  summarise(
    mean_satisfaction = mean(satisfaction_score, na.rm = TRUE),
    median_satisfaction = median(satisfaction_score, na.rm = TRUE),
    count = n()
  )
```

```
## # A tibble: 4 x 4
##   course       mean_satisfaction median_satisfaction count
##   <chr>                    <dbl>               <dbl> <int>
## 1 Business                  6.80                 6.8   504
## 2 Data Science              7.52                 7.5   463
## 3 Engineering               6.51                 6.5   527
## 4 Psychology                7.22                 7.2   506
```

Since students' satisfaction scores do not follow a normal distribution based on previous analysis, median values were chosen to test the hypothesis. This statistics table shows mean satisfaction scores, median satisfaction scores, and total number of samples across courses. To assess whether the median satisfaction scores across courses are statistically significant, a Kruskal-Wallis rank sum test test should be conducted.

```r
ggplot(feedback_data, aes(sample = satisfaction_score)) +
  stat_qq() +
  stat_qq_line(color = "red") +
  facet_wrap(~ course) +
  labs(
    title = "QQ Plot of Satisfaction Score by Course",
    x = "Theoretical Quantiles (Normal Distribution)",
    y = "Sample Quantiles (Satisfaction Score)"
  ) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```




```r
feedback_data %>%
  group_by(course) %>%
  summarise(
    shapiro_p_value = shapiro.test(satisfaction_score)$p.value
  )
```

```
## # A tibble: 4 x 2
##   course       shapiro_p_value
##   <chr>                  <dbl>
## 1 Business              0.599
## 2 Data Science          0.0142
## 3 Engineering           0.113
## 4 Psychology            0.493
```

Normality tests were conducted to confirm the absence of normality. According to the Q-Q plots and Shapiro-Wilk normality test results, satisfaction scores by course follow a normal distribution ($p > 0.05$), except for satisfaction scores of data science ($p < 0.05$), justifying the use of a Kruskal-Wallis rank sum test.

```
kruskal.test(satisfaction_score ~ course, data = feedback_data)
```

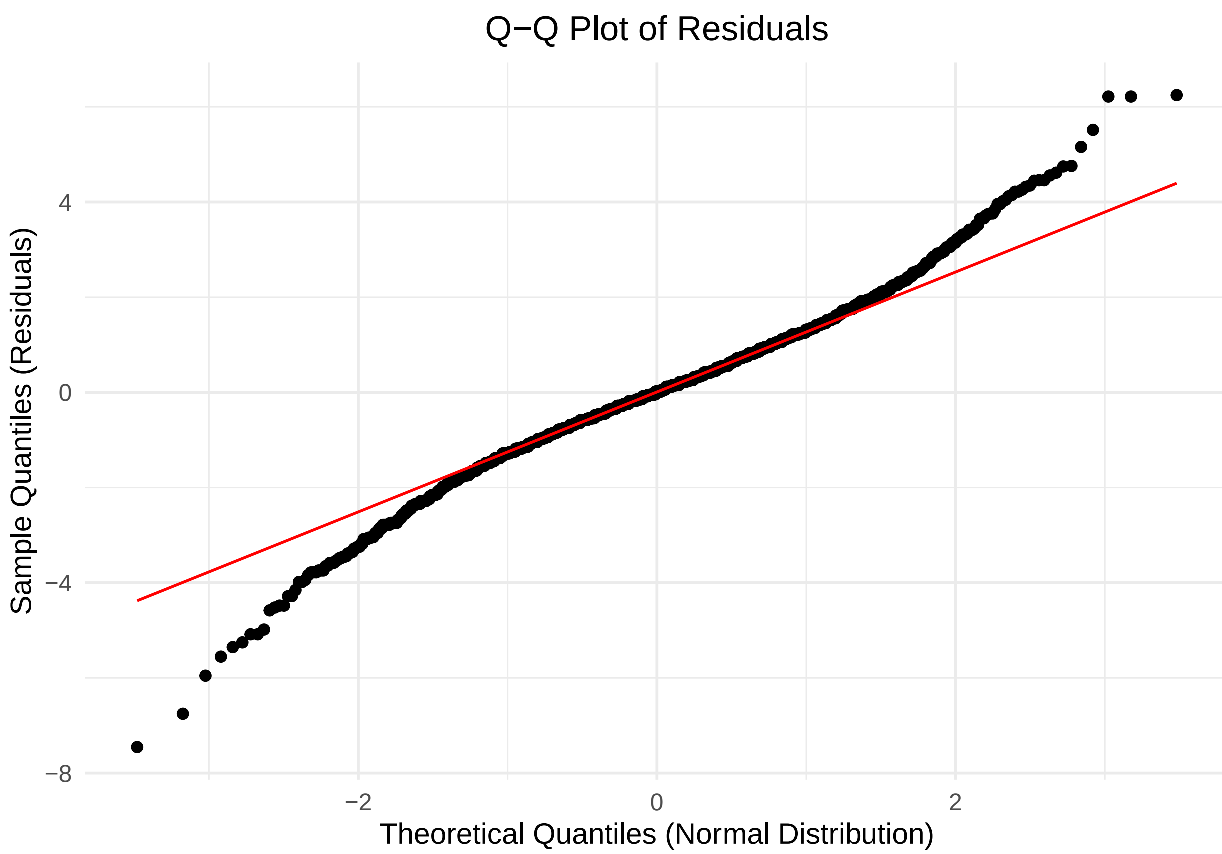
```
##
##  Kruskal-Wallis rank sum test
##
## data:  satisfaction_score by course
## Kruskal-Wallis chi-squared = 209.9, df = 3, p-value < 2.2e-16
```

According to the Kruskal-Wallis rank sum test result, the null hypothesis is rejected. The median satisfaction scores are statistically significantly different across courses ($p < 0.05$).

Appendix 6. Is there an interaction between gender and level on satisfaction?

$$H_0 : \beta_{\text{gender : level}} = 0 \quad \text{vs.} \quad H_1 : \beta_{\text{gender : level}} \neq 0$$

```
anova_model <- aov(satisfaction_score ~ gender * level, data = feedback_data)

anova_residuals <- residuals(anova_model)

residuals <- data.frame(residuals = anova_residuals)

ggplot(residuals, aes(sample = residuals)) +
  stat_qq() +
  stat_qq_line(color = "red") +
  labs(
    title = "Q-Q Plot of Residuals",
    x = "Theoretical Quantiles (Normal Distribution)",
    y = "Sample Quantiles (Residuals)"
  ) +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```




```
shapiro.test(residuals$"residuals")
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals$residuals
## W = 0.9827, p-value = 7.619e-15
```

To assess normality, ANOVA residuals were used. According to the Q-Q plot and Shapiro-Wilk normality test results, the residuals do not follow a normal distribution ($p < 0.05$), suggesting the use of a GLM model.

```
glm_model <- glm(satisfaction_score ~ gender * level, data = feedback_data,
                 family = gaussian())

summary(glm_model)
```

```
##
## Call:
## glm(formula = satisfaction_score ~ gender * level, family = gaussian(),
##     data = feedback_data)
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       6.98051    0.06888 101.343   <2e-16 ***
```

```
## genderMale                         0.06133    0.09685   0.633    0.527
## genderOther                        0.23949    0.22149   1.081    0.280
## levelUndergraduate                -0.02740    0.09516  -0.288    0.773
## genderMale:levelUndergraduate     -0.03085    0.13639  -0.226    0.821
## genderOther:levelUndergraduate -0.15260    0.32979  -0.463    0.644
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 2.215682)
##
##     Null deviance: 4422.8  on 1999  degrees of freedom
## Residual deviance: 4418.1  on 1994  degrees of freedom
## AIC: 7274.9
##
## Number of Fisher Scoring iterations: 2
```

According to the GLM result, the null hypothesis is not rejected. Gender has no statistically significant effect on satisfaction (the p-values of genderMale and genderOther > 0.05). Likewise, level has no statistically significant effect on satisfaction (the p-value of levelUndergraduate > 0.05). Then, the interaction of gender and level has no statistically significant effect on satisfaction (the p-values of genderMale:levelUndergraduate and genderOther:levelUndergraduate > 0.05).

```r
anova_model <- aov(satisfaction_score ~ gender * level, data = feedback_data)
summary(anova_model)
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## gender          2      3  1.5132   0.683  0.505
## level           1      1  1.1758   0.531  0.466
## gender:level    2      1  0.2576   0.116  0.890
## Residuals    1994   4418  2.2157
```

To support the outcome, a two-way ANOVA test was conducted. Since the residuals do not follow a normal distribution, the accuracy of this test may not be reliable. However, the result also shows that gender, level, and their interaction have no statistically significant effect on satisfaction (all p-values > 0.05), indicating the same findings from the GLM result.

Appendix 7. Is the proportion of males equal to that of females?

$$H_0: \; p_{\text{male}} = p_{\text{female}} \quad \text{vs.} \quad H_1: p_{\text{male}} \neq p_{\text{female}}$$

```r
male_female_data <- feedback_data$gender[feedback_data$gender %in% c("Male", "Female")]
male_female_data <- droplevels(as.factor(male_female_data))

gender_counts <- table(male_female_data)

print(gender_counts)
```

```
## male_female_data
## Female   Male
##    981    929
```

The total number of females is 981, while the total number of males is 929.

```r
chisq_result <- chisq.test(gender_counts, p = c(0.5, 0.5))

chisq_result$expected
```

```
## Female   Male
##    955    955
```

Based on the expected frequency results, each cell is more than 5. Then, a Chi-squared test was conducted.

```r
chisq_result <- chisq.test(gender_counts, p = c(0.5, 0.5))

print(chisq_result)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  gender_counts
## X-squared = 1.4157, df = 1, p-value = 0.2341
```

According to the Chi-squared test result, the null hypothesis is not rejected (P > 0.05). Therefore, the difference in gender proportions is not statistically significant.

Appendix 8. Are course and recommendation in the male population associated?

$$H_0: \ P(\text{Recommendation} \mid \text{Course}) = P(\text{Recommendation}) \quad \text{vs.}$$
$$H_1: \ P(\text{Recommendation} \mid \text{Course}) \neq P(\text{Recommendation})$$

```r
male_data <- subset(feedback_data, gender == "Male")

male_table <- table(male_data$course, male_data$recommended)

expected_frequency <- chisq.test(male_table, simulate.p.value = FALSE)$expected

print(expected_frequency)
```

```
##
##                      No      Yes
##   Business     64.58342 164.4166
##   Data Science 61.19914 155.8009
##   Engineering  67.68568 172.3143
##   Psychology   68.53175 174.4682
```

Based on the expected frequency results, each cell is more than 5. Then, a Chi-squared test was conducted.

```r
chi_test <- chisq.test(male_table)

print(chi_test)
```

```
##
##  Pearson's Chi-squared test
##
## data:  male_table
## X-squared = 2.7424, df = 3, p-value = 0.4331
```

According to the Chi-squared test result, the null hypothesis is not rejected ($p > 0.05$). Therefore, there is no statistically significant association between course and recommendation in the male population.

Appendix 9. Are gender and perceived improvement associated?

$$H_0: \; P(\text{Improvement} \mid \text{Gender}) = P(\text{Improvement}) \quad \text{vs.}$$
$$H_1: \; P(\text{Improvement} \mid \text{Gender}) \neq P(\text{Improvement})$$

```r
gender_improvement_table <- table(feedback_data$gender, feedback_data$improved)

expected_frequency <- chisq.test(gender_improvement_table,
                                 simulate.p.value = FALSE)$expected

print(expected_frequency)
```

```
##
##                No      Yes
##   Female 493.9335 487.0665
##   Male   467.7515 461.2485
##   Other   45.3150  44.6850
```

Based on the expected frequency results, each cell is more than 5. Then, a Chi-squared test was conducted.

```r
chi_test <- chisq.test(gender_improvement_table)

print(chi_test)
```

```
##
##  Pearson's Chi-squared test
##
## data:  gender_improvement_table
## X-squared = 0.53499, df = 2, p-value = 0.7653
```

According to the Chi-squared test result, the null hypothesis is not rejected ($p > 0.05$). Therefore, there is no statistically significant association between gender and improvement.

Appendix 10. What are the key predictors of students' satisfaction score?

a. Find the best linear model to explain students' satisfaction score.

```
full_feature_model <- lm(satisfaction_score ~ gender + level + course +
                         study_hours + improved + recommended,
                         data = feedback_data)

summary(full_feature_model)
```

```
##
## Call:
## lm(formula = satisfaction_score ~ gender + level + course + study_hours +
##     improved + recommended, data = feedback_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.2677 -0.7817 -0.0094  0.7801  6.4412
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           6.796724   0.111830  60.777  < 2e-16 ***
## genderMale            0.041126   0.065988   0.623  0.53319
## genderOther           0.257547   0.158817   1.622  0.10504
## levelUndergraduate -0.019380   0.064572  -0.300  0.76411
## courseData Science  0.722163   0.092863   7.777 1.19e-14 ***
## courseEngineering  -0.286454   0.089778  -3.191  0.00144 **
## coursePsychology    0.424402   0.090708   4.679 3.08e-06 ***
## study_hours        -0.001482   0.003650  -0.406  0.68480
## improvedYes         0.041769   0.064439   0.648  0.51694
## recommendedYes     -0.029196   0.070693  -0.413  0.67965
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.439 on 1990 degrees of freedom
## Multiple R-squared:  0.06792,    Adjusted R-squared:  0.0637
## F-statistic: 16.11 on 9 and 1990 DF,  p-value: < 2.2e-16
```

According to the result, the best linear model for explaining satisfaction score may include only course ($p < 0.05$). The other features do not have statistically significant associations with satisfaction.

```
best_model <- lm(satisfaction_score ~ course * gender, data = feedback_data)

summary(best_model)
```

```
##
## Call:
## lm(formula = satisfaction_score ~ course * gender, data = feedback_data)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1749 -0.7749 -0.0114  0.7708  6.5251
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     6.67490    0.09078  73.529  < 2e-16 ***
## courseData Science              0.88187    0.13143   6.710 2.53e-11 ***
## courseEngineering              -0.18902    0.12788  -1.478   0.1395
## coursePsychology                0.53648    0.12903   4.158 3.35e-05 ***
## genderMale                      0.24038    0.13143   1.829   0.0675 .
## genderOther                     0.26260    0.30729   0.855   0.3929
## courseData Science:genderMale  -0.32711    0.18931  -1.728   0.0842 .
## courseEngineering:genderMale   -0.19710    0.18440  -1.069   0.2853
## coursePsychology:genderMale    -0.27481    0.18492  -1.486   0.1374
## courseData Science:genderOther -0.23113    0.47448  -0.487   0.6262
## courseEngineering:genderOther  -0.10786    0.40887  -0.264   0.7920
## coursePsychology:genderOther    0.44367    0.47382   0.936   0.3492
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.438 on 1988 degrees of freedom
## Multiple R-squared:  0.07025,    Adjusted R-squared:  0.06511
## F-statistic: 13.66 on 11 and 1988 DF,  p-value: < 2.2e-16
```

The model including course and its interaction with gender performs slightly better than the model with only course. However, this model only explains approximately 6.5% (adjusted R-squared) of the satisfaction variance.

b. Comment on the effect size of the significant predictor

```r
summary(lm(formula = satisfaction_score ~ course * gender, data = feedback_data))
```

```
##
## Call:
## lm(formula = satisfaction_score ~ course * gender, data = feedback_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1749 -0.7749 -0.0114  0.7708  6.5251
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     6.67490    0.09078  73.529  < 2e-16 ***
## courseData Science              0.88187    0.13143   6.710 2.53e-11 ***
## courseEngineering              -0.18902    0.12788  -1.478   0.1395
## coursePsychology                0.53648    0.12903   4.158 3.35e-05 ***
## genderMale                      0.24038    0.13143   1.829   0.0675 .
## genderOther                     0.26260    0.30729   0.855   0.3929
## courseData Science:genderMale  -0.32711    0.18931  -1.728   0.0842 .
## courseEngineering:genderMale   -0.19710    0.18440  -1.069   0.2853
```

```
## coursePsychology:genderMale    -0.27481    0.18492  -1.486   0.1374
## courseData Science:genderOther -0.23113    0.47448  -0.487   0.6262
## courseEngineering:genderOther  -0.10786    0.40887  -0.264   0.7920
## coursePsychology:genderOther    0.44367    0.47382   0.936   0.3492
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.438 on 1988 degrees of freedom
## Multiple R-squared:  0.07025,    Adjusted R-squared:  0.06511
## F-statistic: 13.66 on 11 and 1988 DF,  p-value: < 2.2e-16
```

According to the result, female students in business showed an average 6.7 satisfaction score ($p < 0.05$). Female students in data science showed a 0.88 points higher average satisfaction score than female students in business ($p < 0.05$). Female students in psychology showed a 0.54 points higher average satisfaction score than female students in business ($p < 0.05$). In addition, male and the interaction of male and data science have marginally significant associations with satisfaction. Males in business showed 2.4 points higher average satisfaction score than females in business. Then, the difference in average satisfaction scores between males and females in data science is approximately 0.33 points lower than the difference between males and females in business.

c. Comment on how much of the variation in satisfaction can be explained by the significant predictors.

```r
model_exp <- lm(satisfaction_score ~ course * gender, data = feedback_data)

summary(model_exp)$adj.r.squared
```

```
## [1] 0.06510587
```

The result indicates that approximately 6.5% of the variation in satisfaction can be explained by course and its interaction with gender. The figure is statistically significant ($p < 0.05$), as shown in the appendix 10 (b). For future research, additional predictors that can explain the remaining 93.5% should be identified and incorporated.