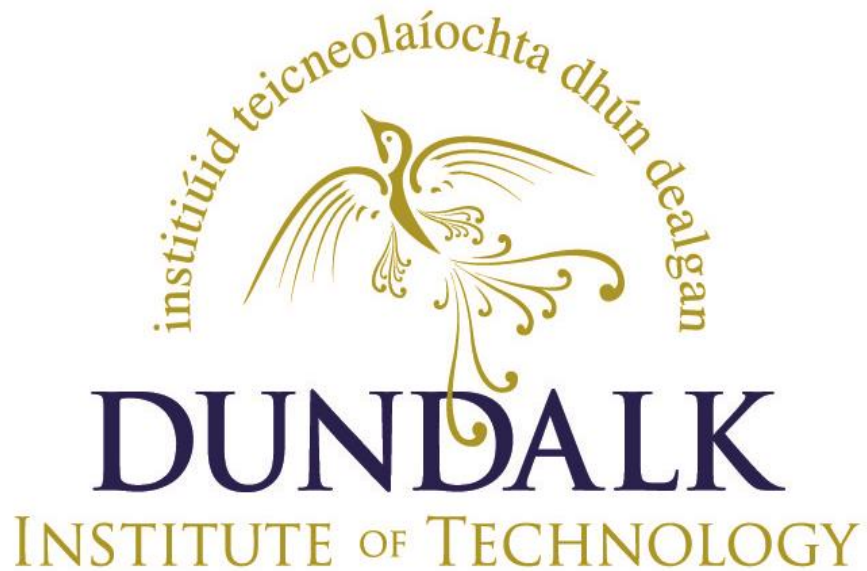


Study Design Continuous Assessment

Analysis of DkIT Student's Wellbeing Survey.



Students Names:

Israa Atike (D00262160)

Kaylee Jacqueline Wijaya (D00262128)

Table of Contents

Table of Contents	1
Introduction.....	2
Research Area	3
Research Question	3
Preliminary Research and Hypothesis	4
Survey	6
Data Cleaning.....	7
Data Analysis	10
Hypothesis Testing.....	15
Conclusion	18
References.....	19
Appendix.....	20

Introduction

Wellbeing is defined as a person experiencing health and happiness. Many factors can affect a person's well-being, such as physical and mental health (Dolan, P., Peasgood, T. & White, M., 2008). Students' wellbeing is important as it affects the ability of students to develop their potential (education.ec.europa.eu, n.d.).

In this report, we are going to analyze the well-being of students at Dundalk Institute of Technology. We will create an online survey to collect data from students and analyze the data using R programming language to gain statistical insights and create a conclusion.

Research Area

As university students we chose to study the area of student wellbeing since it is a highly relevant and impactful phenomenon and since we are also affected by that as much as other students are, we were curious to discover more about this phenomenon that has been impacting us, our classmates and friends who are pursuing their bachelor's degree inside of DkIT. We agreed that physical and mental health should be included in our research since both of them matter equally, embracing both of them will give us the ability to have more insights about this study as both of them impact academic performance which is considered one of the most crucial parts of a student's degree.

Research Question

Our research question is **“Do students in Dundalk Institute of Technology prioritize their wellbeing as in physical and mental health?”**

We settled with this question since the well-being of students is not only about physical health as some people think but also about mental health, also understanding if students prioritize their mental and physical health or not will shed light on the difficulties students find in prioritizing their wellbeing not only but it will also shed light on how students that prioritize their well-being manage to do that. We decided to include DkIT's students in this research question only since they are the most accessible people for us and also because we are part of them, thus it will make us comprehend them better.

Preliminary Research and Hypothesis

We did some research that is related to our research questions to further understand the research topic.

1. Prioritization of needs among students of the University of Medical Sciences: A needs assessment (Fattahi et al., 2020)

This journal was created to assess the needs of students from the Iran University of Medical Sciences. The research was done by using qualitative methods and they did interviews for the survey type. Then they proceeded to use a simple random sampling to select people to complete the questionnaires.

The result and conclusion of this journal:

- a. The students prioritize their well-being first and ranked their educational needs as secondary.
- b. Gender is not a factor in the result of the survey.
- c. Marital status might be affecting the students' economic needs.
- d. The education system should take into consideration the physical needs of its students.

2. Young people's perceptions of mental and physical health in the context of general wellbeing (Singletary et al., 2014)

This report conducts a study to what is the perception of health from the perspective of young people. The research was done by using qualitative methods and they did anonymous questionnaires for the survey type. They also considered gender as a factor for this survey.

The result and conclusion:

- a. Most young people correlate wellbeing with only physical health and only a few think that mental health can affect their wellbeing.
- b. The gender does not impact the response as both genders answered similarly, but more girls included the social aspect of mental health.

From both of the research that we did, we decided that we hypothesize that DkIT students will prioritize their well-being, especially their physical health but also their mental health.

Survey

For our survey, we decided to conduct an online questionnaire survey which is Qualitative research, and we will use Google Forms since it is easily accessible and a user-friendly website for both survey creators and respondents. The responses are collected in real-time, and they are immediately available for analysis, it also provides a variety of question formats. Thus, we decided to use it for our survey.

The sampling method that we used is a mixture between Convenience and Snowball Sampling. We sent the survey link to our closest friends who are easily accessible to us and we asked them to send it to their friends to get more responses.

We divided our survey into three parts and each part has several questions: (See Appendix A)

1. We focused on obtaining the general information of the respondent such as Age, Course name, Type of the course (full-time/part-time), and their GPA for the last academic year.
2. For the second part of our survey, we focused on making physical health questions and it included questions such as Sleep hours per night, Smoking or vaping habits, drinking habits, physical activity, diet type, and overall physical health rating.
3. The third part of our survey focused on the respondent's mental health. It had questions such as how much students prioritize their mental health and manage stress in their daily lives, how often they feel depressed or sad, whether have they ever had thoughts of self-harm or suicide, and the rating of their mental health.

We did not use any sampling method since we only had 35 responses, but our response rate was 0.875 (87.5%) since we sent the survey to 40 students and 35 of them responded to it.

For our data collection, we used the option of the CSV file that Google Forms provides to analyze our data since a CSV file makes it easier to access and use.

The time frame of the conduction of our data was a month since at first we did not have a lot of students to whom we could send the survey, but later we decided to use snowball sampling to get more responses since the convenience sampling was not enough.

Data Cleaning

The cleaning that we have done to our data is the following:

Column name	Cleaning made
Age Column:	We replaced the question 'How old are you?' with 'Age' We considered '<18' as 17 We considered '18-25' as 22 We considered '25-35' as 30 We considered '>35' as 36
Gender column:	We replaced the question 'What is your gender?' with 'Gender' We considered 'Females' as F We considered 'Males' as M We considered 'Prefer not to say' as P
Course column	We replaced the question 'In what field are you doing your degree?' with 'Course' We considered the Department of Business Studies a BS We considered the Department of Computing Science and Mathematics as CSM We considered the Department of Creative Arts, Media, and Music as CAMM We considered the Department of Electronic and Mechanical Engineering as EME We considered the Department of Engineering Trades and Civil Engineering as ETCE We considered the Department of Hospitality Studies as an HS We considered the Department of Humanities as H We considered Department of Life & Health Sciences as LHS We considered the Department of Management and Financial Studies as MFS We considered the Department of Nursing, Midwifery & Early Years as NMEY

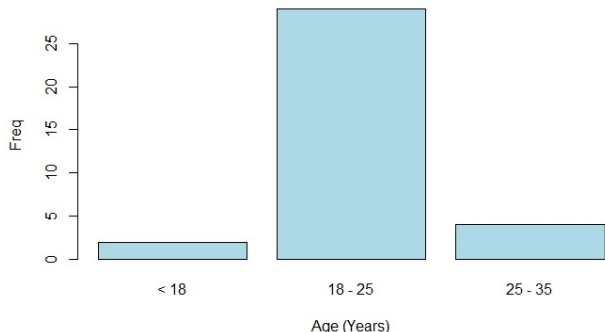
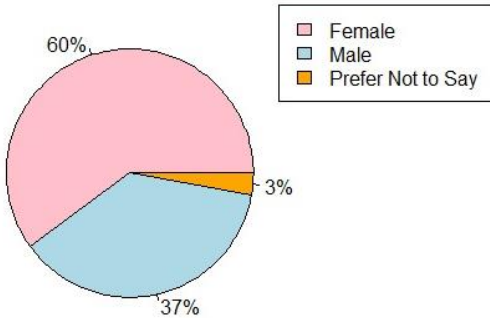
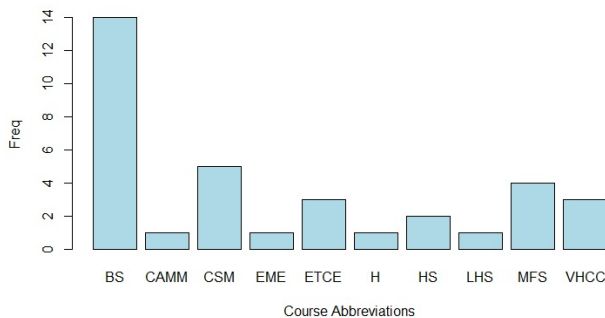
	<p>We considered the Department of the Built Environment as BE</p> <p>We considered the Department of Visual and Human-Centred Computing as VHCC</p> <p>We considered the Department of Agriculture, Food & Animal Health as AFAH</p>
FT/PT column	<p>We replaced the question ‘Are you a full-time or a part-time student’ with ‘FT/PT’</p> <p>We considered ‘full time’ as ‘FT’</p> <p>We considered ‘part-time’ as ‘PT’</p>
GPA column	<p>We replaced the question ‘What is your last year's GPA?’ with ‘GPA’</p> <p>We considered ‘40-60’ as ‘50’</p> <p>We considered ‘60-80’ as ‘70’</p> <p>We considered ‘More than 80’ as ‘90’</p>
Sleep hours column:	<p>We replaced the question ‘How many hours of sleep do you get usually per night?’ with ‘Sleep hours’.</p> <p>We considered ‘,<4 hours’ as ‘2 hours’</p> <p>We considered ‘4-6hours’ as ‘5 hours’</p> <p>We considered ‘6-8hours’ as ‘7 hours’</p> <p>We considered ‘8-10hours’ as ‘9 hours’</p>
Smoke/ vape column:	<p>We replaced the question ‘Do you smoke or vape?’ with ‘Smoke/vape’.</p>
Drink per week column:	<p>We replaced the question ‘How often do you drink weekly?’ with ‘Drink per week’.</p> <p>We considered ‘0 days’ as ‘0 days’.</p> <p>We considered ‘1-3 days’ as ‘2 days’.</p>

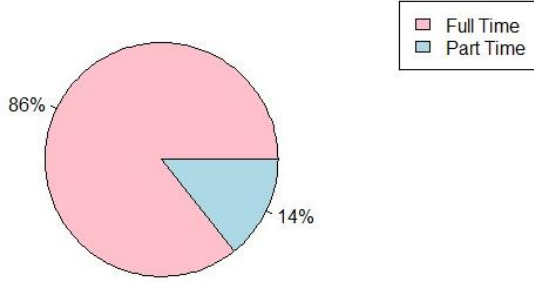
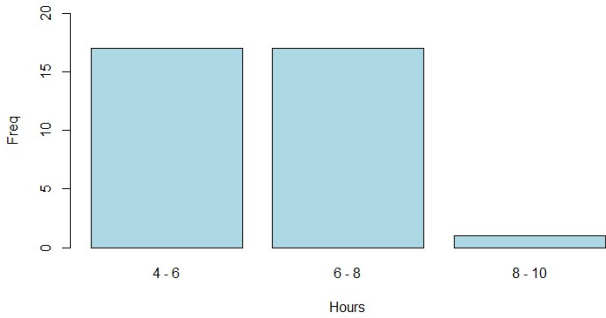
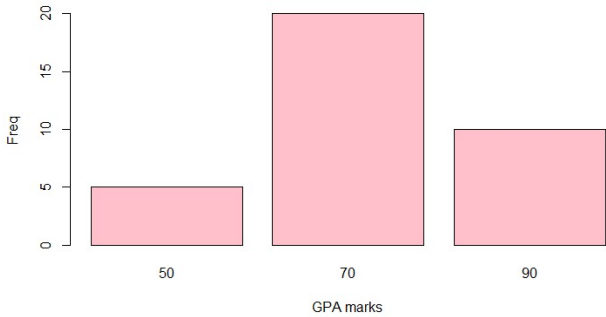
	We considered 'More than 3 days' as '4 days'.
Exercise per week column:	<p>We replaced the question 'How often do you exercise or engage in physical activity each week?' with 'Exercise per week'.</p> <p>We considered '0' as '0'</p> <p>We considered '1 to 3 times' as '2 times'</p> <p>We considered '3 to 5 times' as '4 times'</p> <p>We considered 'More than 5 times' as '6 times'</p>
Diet satisfaction column	We replaced the question 'Are you satisfied with your current diet and nutrition habits' with 'Diet satisfaction'
Physical health rating column	We replaced the question 'How would you rate your overall physical health in a scale of 1 to 5?' with 'Physical health rating'.
Mental health prioritization column	We replaced the question 'Do you prioritize mental health and stress management in your daily life?' with 'Mental health prioritization'.
Depression scale column	We replaced the question 'How often do you feel depressed in your daily life?' with 'Depression scale.
Suicidal thoughts column	We replaced the question 'Have you ever had thoughts of self-harm or suicide?' with 'Suicidal thoughts'.
Mental health rating column	We replaced the question 'How would you rate your current level of mental health on a scale of 1 to 5?' with 'Mental health rating.

Data Analysis

We used R programming language to create plots for the variables in the survey. (See Appendix B)

I. Plots for General Information

Variable	Plot	Analysis																						
Age	<p style="text-align: center;">Bar Plot of Age</p>  <table><caption>Data for Bar Plot of Age</caption><thead><tr><th>Age (Years)</th><th>Freq</th></tr></thead><tbody><tr><td>< 18</td><td>2</td></tr><tr><td>18 - 25</td><td>25</td></tr><tr><td>25 - 35</td><td>4</td></tr></tbody></table>	Age (Years)	Freq	< 18	2	18 - 25	25	25 - 35	4	From the bar plot, we notice that most of the respondents are in the age group of 18 to 25 years old.														
Age (Years)	Freq																							
< 18	2																							
18 - 25	25																							
25 - 35	4																							
Sex	<p style="text-align: center;">Pie Chart of Sex</p>  <table><caption>Data for Pie Chart of Sex</caption><thead><tr><th>Sex</th><th>Percentage</th></tr></thead><tbody><tr><td>Female</td><td>60%</td></tr><tr><td>Male</td><td>37%</td></tr><tr><td>Prefer Not to Say</td><td>3%</td></tr></tbody></table>	Sex	Percentage	Female	60%	Male	37%	Prefer Not to Say	3%	From the pie chart, we notice that most of the respondents are females with a percentage of 60%, after that males come with a percentage of 37%.														
Sex	Percentage																							
Female	60%																							
Male	37%																							
Prefer Not to Say	3%																							
Course name	<p style="text-align: center;">Bar Plot of Course</p>  <table><caption>Data for Bar Plot of Course</caption><thead><tr><th>Course Abbreviations</th><th>Freq</th></tr></thead><tbody><tr><td>BS</td><td>14</td></tr><tr><td>CAMM</td><td>1</td></tr><tr><td>CSM</td><td>5</td></tr><tr><td>EME</td><td>1</td></tr><tr><td>ETCE</td><td>3</td></tr><tr><td>H</td><td>1</td></tr><tr><td>HS</td><td>2</td></tr><tr><td>LHS</td><td>1</td></tr><tr><td>MFS</td><td>4</td></tr><tr><td>VHCC</td><td>3</td></tr></tbody></table>	Course Abbreviations	Freq	BS	14	CAMM	1	CSM	5	EME	1	ETCE	3	H	1	HS	2	LHS	1	MFS	4	VHCC	3	From the bar plot, we notice that most of the respondents study business with a frequency of 14.
Course Abbreviations	Freq																							
BS	14																							
CAMM	1																							
CSM	5																							
EME	1																							
ETCE	3																							
H	1																							
HS	2																							
LHS	1																							
MFS	4																							
VHCC	3																							

Course Type	<p>Pie Chart of Full time/ Part time</p>  <p>A pie chart titled 'Pie Chart of Full time/ Part time'. The chart is divided into two segments: a large pink segment representing 'Full Time' at 86%, and a smaller light blue segment representing 'Part Time' at 14%. A legend to the right of the chart identifies the colors: pink for 'Full Time' and light blue for 'Part Time'.</p>	From the pie chart, we notice that most of the respondents study full time with a percentage of 88%.
Hours of Sleep	<p>Bar Plot For Hours of Sleep</p>  <p>A bar plot titled 'Bar Plot For Hours of Sleep'. The x-axis is labeled 'Hours' and has three categories: '4 - 6', '6 - 8', and '8 - 10'. The y-axis is labeled 'Freq' and ranges from 0 to 20 in increments of 5. The bars are light blue. The frequencies are: 17 for '4 - 6', 17 for '6 - 8', and 1 for '8 - 10'.</p>	From the bar plot, we notice that most respondents sleep between 4 and 8 hours with a frequency of 32.
GPA	<p>Bar Plot of GPA</p>  <p>A bar plot titled 'Bar Plot of GPA'. The x-axis is labeled 'GPA marks' and has three categories: '50', '70', and '90'. The y-axis is labeled 'Freq' and ranges from 0 to 20 in increments of 5. The bars are pink. The frequencies are: 5 for '50', 20 for '70', and 10 for '90'.</p>	From our bar plot, we notice that most of the respondents, got a GPA of 70 (between 60 and 80) with a frequency of 20.

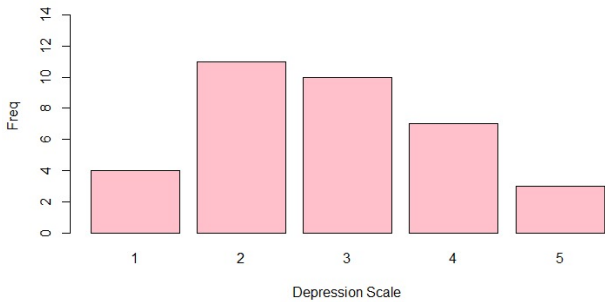
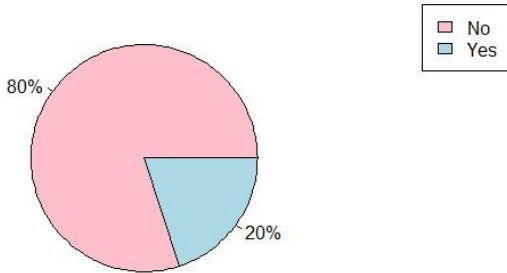
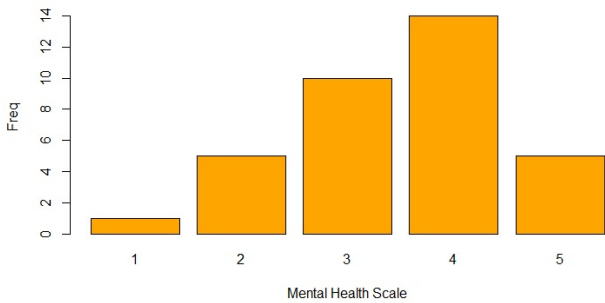
II. Plot for Physical Health Sector

Variable	Plot	Analysis
Smoking or vaping habit	<p>Pie Chart For Smoking/Vaping</p> <p>A pie chart titled 'Pie Chart For Smoking/Vaping' showing the distribution of smoking or vaping habits. The chart is divided into three segments: a large pink segment for 'No' at 66%, an orange segment for 'Yes' at 31%, and a small light blue segment for 'Sometimes' at 3%. A legend to the right of the chart identifies the colors: pink for 'No', light blue for 'Sometimes', and orange for 'Yes'.</p>	From the pie chart, we can notice that most of the respondents do not smoke or v with a percentage of 66%.
Drinking habit	<p>Bar Plot of Drinking Frequency per Week</p> <p>A bar plot titled 'Bar Plot of Drinking Frequency per Week'. The x-axis is labeled 'Days' and has three categories: '0', '1 - 3', and '> 3'. The y-axis is labeled 'Freq' and ranges from 0 to 15. There are three orange bars representing the frequency for each category: 9 for '0', 14 for '1 - 3', and 8 for '> 3'.</p>	From the bar plot, we can notice that most of the respondents drink between 1 to 3 days a week.
Diet Satisfaction	<p>Pie Chart For Diet Satisfaction</p> <p>A pie chart titled 'Pie Chart For Diet Satisfaction' showing the distribution of diet satisfaction. The chart is divided into three segments: a light blue segment for 'No' at 40%, a pink segment for 'Maybe' at 34%, and an orange segment for 'Yes' at 26%. A legend to the right of the chart identifies the colors: pink for 'Maybe', light blue for 'No', and orange for 'Yes'.</p>	From the pie chart, we can notice that most of the respondents are not satisfied by their diet with a percentage of 40%.

Exercise/ Physical	<p>Bar Plot of Exercising per Week</p> <p>A bar plot showing the frequency of exercise per week. The x-axis is labeled 'Days' and has categories: 0, 1 - 3, 3 - 5, and > 5. The y-axis is labeled 'Freq' and ranges from 0 to 20. The bars are light blue with black outlines. The frequencies are approximately 11 for 0 days, 16 for 1-3 days, 7 for 3-5 days, and 1 for more than 5 days.</p> <table><thead><tr><th>Days</th><th>Freq</th></tr></thead><tbody><tr><td>0</td><td>11</td></tr><tr><td>1 - 3</td><td>16</td></tr><tr><td>3 - 5</td><td>7</td></tr><tr><td>> 5</td><td>1</td></tr></tbody></table>	Days	Freq	0	11	1 - 3	16	3 - 5	7	> 5	1	From the bar plot, we can notice that most of the respondents exercise between 1 to 3 times a week with a frequency of 15.		
Days	Freq													
0	11													
1 - 3	16													
3 - 5	7													
> 5	1													
Physical Health Rating	<p>Bar Plot of Physical Health Rating</p> <p>A bar plot showing the frequency of physical health ratings. The x-axis is labeled 'Rating' and has categories: 1, 2, 3, 4, and 5. The y-axis is labeled 'Freq' and ranges from 0 to 14. The bars are light blue with black outlines. The frequencies are 2 for rating 1, 3 for rating 2, 14 for rating 3, 11 for rating 4, and 3 for rating 5.</p> <table><thead><tr><th>Rating</th><th>Freq</th></tr></thead><tbody><tr><td>1</td><td>2</td></tr><tr><td>2</td><td>3</td></tr><tr><td>3</td><td>14</td></tr><tr><td>4</td><td>11</td></tr><tr><td>5</td><td>3</td></tr></tbody></table>	Rating	Freq	1	2	2	3	3	14	4	11	5	3	From the bar plot, we can notice that most of the respondents have a rate of 3 regarding their physical health.
Rating	Freq													
1	2													
2	3													
3	14													
4	11													
5	3													

III. Plots for the Mental Health Sector

Variable	Plot	Analysis								
Mental health prioritization	<p>Pie Chart For Mental Health Prioritization</p> <table><thead><tr><th>Response</th><th>Percentage</th></tr></thead><tbody><tr><td>Yes</td><td>49%</td></tr><tr><td>Maybe</td><td>29%</td></tr><tr><td>No</td><td>23%</td></tr></tbody></table>	Response	Percentage	Yes	49%	Maybe	29%	No	23%	From the pie chart, we can notice that most of the respondents prioritize their mental health with a percentage of 49%
Response	Percentage									
Yes	49%									
Maybe	29%									
No	23%									

Depression scale	<p>Bar Plot of Depression Scale</p>  <p>A bar plot showing the frequency of depression scale ratings from 1 to 5. The y-axis is labeled 'Freq' and ranges from 0 to 14. The x-axis is labeled 'Depression Scale'. The bars show frequencies of 4 for scale 1, 11 for scale 2, 10 for scale 3, 7 for scale 4, and 3 for scale 5.</p> <table><thead><tr><th>Depression Scale</th><th>Freq</th></tr></thead><tbody><tr><td>1</td><td>4</td></tr><tr><td>2</td><td>11</td></tr><tr><td>3</td><td>10</td></tr><tr><td>4</td><td>7</td></tr><tr><td>5</td><td>3</td></tr></tbody></table>	Depression Scale	Freq	1	4	2	11	3	10	4	7	5	3	From the bar plot, we can notice that most of the respondents have a depression scale of 2 with a frequency of 11.
Depression Scale	Freq													
1	4													
2	11													
3	10													
4	7													
5	3													
Suicidal thoughts	<p>Pie Chart For Suicidal Thoughts</p>  <p>A pie chart showing the percentage of respondents with and without suicidal thoughts. The chart is divided into two segments: a large pink segment labeled '80%' for 'No' and a smaller blue segment labeled '20%' for 'Yes'. A legend on the right shows a pink square for 'No' and a blue square for 'Yes'.</p> <table><thead><tr><th>Response</th><th>Percentage</th></tr></thead><tbody><tr><td>No</td><td>80%</td></tr><tr><td>Yes</td><td>20%</td></tr></tbody></table>	Response	Percentage	No	80%	Yes	20%	From the pie chart, we can notice that most of the respondents do not have suicidal thoughts with a percentage of 80%.						
Response	Percentage													
No	80%													
Yes	20%													
Mental health rating	<p>Bar Plot of Mental Health Rating</p>  <p>A bar plot showing the frequency of mental health ratings from 1 to 5. The y-axis is labeled 'Freq' and ranges from 0 to 14. The x-axis is labeled 'Mental Health Scale'. The bars show frequencies of 1 for rating 1, 5 for rating 2, 10 for rating 3, 14 for rating 4, and 5 for rating 5.</p> <table><thead><tr><th>Mental Health Scale</th><th>Freq</th></tr></thead><tbody><tr><td>1</td><td>1</td></tr><tr><td>2</td><td>5</td></tr><tr><td>3</td><td>10</td></tr><tr><td>4</td><td>14</td></tr><tr><td>5</td><td>5</td></tr></tbody></table>	Mental Health Scale	Freq	1	1	2	5	3	10	4	14	5	5	From the bar plot, we can notice that most of our respondents have a mental health rating of 4 with a frequency of 14.
Mental Health Scale	Freq													
1	1													
2	5													
3	10													
4	14													
5	5													

Hypothesis Testing

We did a single and multiple regression test for the Physical and Mental health sectors to see what are the factors that affect each of them.

The multiple regression test is a part of the t-test as we will check the t-value of each test. (See Appendix C)

A. Physical Health

From the regression test, we first use the full model by adding all variables that we think might affect the physical health of students. The summary of the full model states that the only significant variable is the Exercise (with $p\text{-value} = 8.94\text{e-}07 < 0.05$). We did the added variable test to check if there are variables that do not affect Physical Health by looking at the plot and if there is no slope present by the blue line, it means that the variable is not significant and should be removed.

Considering the p-value from the multiple linear regression, we decided to use the Exercise variable only making it a single linear regression model for Physical Health. We did some tests using the AIC and BIC to check which of the models is preferred, the full model or the single model (reduced model). From AIC and BIC testing, we choose the smaller value as the preferred model and both AIC and BIC testing show that the single linear regression (reduced model) is preferred as it has a smaller value.

So we conclude that the variable that is affecting Physical Health the most is the amount of days of exercising.

B. Mental Health

From the regression test, we first use the full model by adding all variables that we think might affect the mental health of students. The summary of the full model states that the only significant variable is the Depression Scale (the p-value is the smallest among all other variables). We did the added variable test to check if there are variables that do not affect Mental Health by looking at the plot and if there is no slope present by the blue line, it means that the variable is not significant and should be removed.

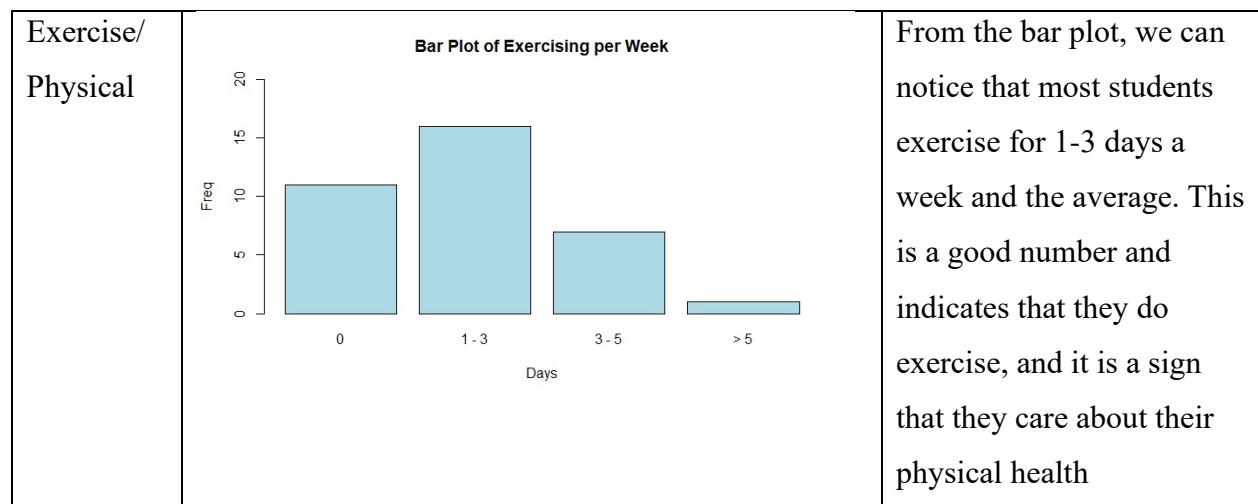
Considering the p-value from the multiple linear regression, we decided to use the Depression scale variable only making it a single linear regression model for Mental Health.

We did two tests using the AIC and BIC to check which of the models is preferred, the full model or the single model (reduced model). From AIC and BIC testing, we choose the smaller value as the preferred model and both AIC and BIC testing show that the single linear regression (reduced model) is preferred as it has a smaller value.

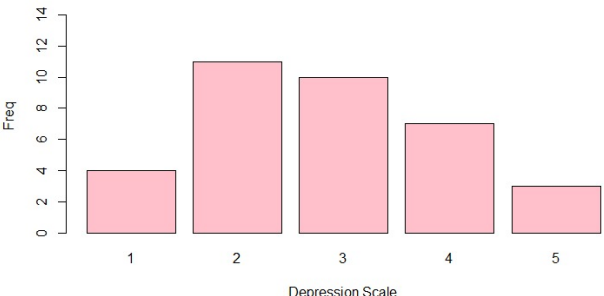
So we conclude that the variable that is affecting Mental Health the most is the scale of depression of the respondents.

To answer our hypothesis of “Do students in DkIT prioritize their well-being (physical and mental health)?”, we are going to look at the plot of the Exercise and Depression scale in addition to our multiple linear regression.

From the plot of Exercise per week:



From the plot of Depression Scale:

Depression scale	<p style="text-align: center;">Bar Plot of Depression Scale</p>  <table data-bbox="380 260 980 554"><tr><th>Depression Scale</th><th>Freq</th></tr><tr><td>1</td><td>4</td></tr><tr><td>2</td><td>11</td></tr><tr><td>3</td><td>10</td></tr><tr><td>4</td><td>7</td></tr><tr><td>5</td><td>3</td></tr></table>	Depression Scale	Freq	1	4	2	11	3	10	4	7	5	3	<p>From the bar plot, we can notice that most of the respondents have a depression scale of 2 and the average is 2.8.</p> <p>From this number, we conclude that the students do not have a high depression scale (the average is less than 3).</p> <p>This indicates that they care about their mental health.</p>
Depression Scale	Freq													
1	4													
2	11													
3	10													
4	7													
5	3													

Conclusion

From the hypothesis testing section, we conclude that the students from Dundalk Institute of Technology do prioritize their physical and mental health which also indicates that they prioritize their wellbeing.

References

Dolan, P., Peasgood, T. & White, M., 2008, Do we know what makes us happy A review of the economic literature on the factors associated with subjective well-being, *Journal of Economic Psychology*, 29(1), pp. 94-122

education.ec.europa.eu. (n.d.). Well-being at school | European Education Area. [online] Available at: <https://education.ec.europa.eu/education-levels/school-education/well-being-at-school>.

Fattahi, E., Solhi, M., Abbas, J., Kasmaei, P., Rastaghi, S., Pouresmaeil, M., Ziapour, A. and Gilan, H.D., 2020. Prioritization of needs among students of University of Medical Sciences: A needs assessment. *Journal of Education and Health Promotion*, 9.

Singletary, J.H., Bartle, C.L., Svirydzenka, N., Suter-Giorgini, N.M., Cashmore, A.M. and Dogra, N. (2014). Young people's perceptions of mental and physical health in the context of general wellbeing. *Health Education Journal*, 74(3), pp.257–269.
doi:<https://doi.org/10.1177/0017896914533219>.

Appendix

Appendix A

<https://forms.gle/K2L24kub9ZQfU9Y98>

Appendix B

The code for the plots (R programming):

```
data = read.csv("D:\\Kaylee\\Kuliah\\Study Design\\data.csv")
data
###
age.freq = table(data$Age)
barplot(age.freq, main = "Bar Plot of Age", names.arg = c('< 18','18 - 25','25 - 35'), ylab = 'Freq',
xlab = 'Age (Years)', col = "lightblue")

sex.freq = table(data$Sex)
percent.labels <- round(100*sex.freq/sum(sex.freq))
pie.labels <- paste(percent.labels, "%", sep = "")
pie(table(data$Sex), main = "Pie Chart of Sex", labels = pie.labels, col = c("pink",
"lightblue", 'orange'))
legend("topright", legend = c("Female", "Male", 'Prefer Not to Say'), fill = c("pink",
"lightblue", 'orange'))

course.freq = table(data$Course)
barplot(course.freq, main = "Bar Plot of Course", ylab = 'Freq', xlab = 'Course Abbreviations',
col = "lightblue")

ftpt.freq = table(data$FT...PT)
percent.labels <- round(100*ftpt.freq/sum(ftpt.freq))
pie.labels <- paste(percent.labels, "%", sep = "")
pie(ftpt.freq, main = "Pie Chart of Full time/ Part time", labels = pie.labels, col = c("pink",
```

```

"lightblue"))
legend("topright", legend = c("Full Time", "Part Time"), fill = c("pink", "lightblue"))

GPA.freq = table(data$GPA)
barplot(GPA.freq, main = "Bar Plot of GPA", ylab = 'Freq', xlab = 'GPA marks', col = "pink")

sleep.freq = table(data$Sleep..hrs.)
barplot(sleep.freq, main = "Bar Plot For Hours of Sleep", names.arg = c('4 - 6','6 - 8','8 - 10'),
        ylim = c(0,20), ylab = 'Freq', xlab = 'Hours', col = "lightblue")

smoke.freq = table(data$Smoke...Vape)
percent.labels <- round(100*smoke.freq/sum(smoke.freq))
pie.labels <- paste(percent.labels, "%", sep = "")
pie(smoke.freq, main = "Pie Chart For Smoking/Vaping", labels = pie.labels, col = c("pink",
"lightblue",'orange'))
legend("topright", legend = c("No", "Sometimes", "Yes"), fill = c("pink", "lightblue", 'orange'))

drink.freq = table(data$Drink.per.week..days.)
barplot(drink.freq, main = "Bar Plot of Drinking Frequency per Week", names.arg = c('0','1 -
3','> 3'), ylab = 'Freq', xlab = 'Days', col = "orange")

exercise.freq = table(data$Exercise.per.week..days.)
barplot(exercise.freq, main = "Bar Plot of Exercising per Week", names.arg = c('0','1 - 3','3 -
5','> 5'), ylim = c(0,20), ylab = 'Freq', xlab = 'Days', col = "lightblue")

diet.freq = table(data$Diet.Satisfaction)
percent.labels <- round(100*diet.freq/sum(diet.freq))
pie.labels <- paste(percent.labels, "%", sep = "")
pie(diet.freq, main = "Pie Chart For Diet Satisfaction", labels = pie.labels, col = c("pink",
"lightblue",'orange'))
legend("topright", legend = c("Maybe", "No", "Yes"), fill = c("pink", "lightblue", 'orange'))

```

```

physical.freq = table(data$Physical.Health.Rating)
barplot(physical.freq, main = "Bar Plot of Physical Health Rating", ylab = 'Freq', xlab = 'Rating',
col = "lightblue")

mental.freq = table(data$Mental.Health.Prioritization)
percent.labels <- round(100*mental.freq/sum(mental.freq))
pie.labels <- paste(percent.labels, "%", sep = "")
pie(mental.freq, main = "Pie Chart For Mental Health Prioritization", labels = pie.labels, col
=c("pink", "lightblue", 'orange'))
legend("topright", legend = c("Maybe", 'No', "Yes"), fill = c("pink", "lightblue", 'orange'))

depression.freq = table(data$Depression.Scale)
barplot(depression.freq, main = "Bar Plot of Depression Scale", ylim = c(0,15), ylab = 'Freq', xlab
= 'Depression Scale', col = "pink")

suicide.freq = table(data$Suicide.Thoughts)
percent.labels <- round(100*suicide.freq/sum(suicide.freq))
pie.labels <- paste(percent.labels, "%", sep = "")
pie(suicide.freq, main = "Pie Chart For Suicidal Thoughts", labels = pie.labels, col = c("pink",
"lightblue"))
legend("topright", legend = c('No', "Yes"), fill = c("pink", "lightblue"))

mental.freq = table(data$Mental.Health.Rating)
barplot(mental.freq, main = "Bar Plot of Mental Health Rating", ylim = c(0,15), ylab = 'Freq',
xlab = 'Mental Health Scale', col = "orange")

```

Appendix C

Loading the data

```

data = read.csv("D:\\Kaylee\\Kuliah\\Study Design\\data.csv")
#data

```

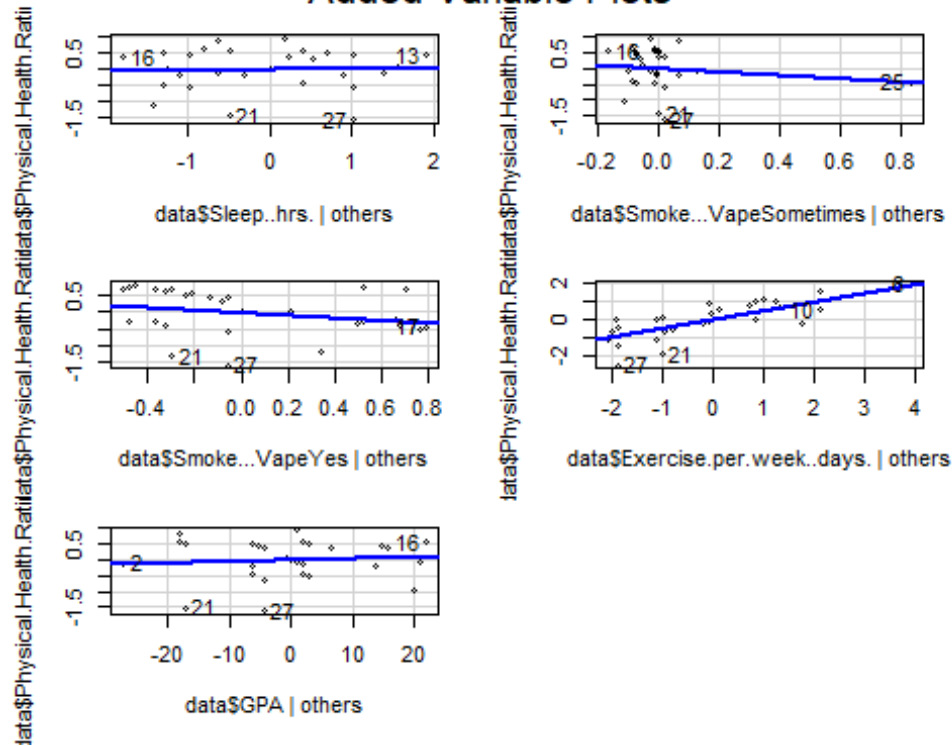
Testing the model for Physical Health

```
library(car)
#creating the full model
physical_model=lm(data$Physical.Health.Rating~data$Sleep..hrs.+data$Smoke...Vape+
  data$Exercise.per.week..days.+data$GPA)
summary(physical_model)

##
## Call:
## lm(formula = data$Physical.Health.Rating ~ data$Sleep..hrs. +
##     data$Smoke...Vape + data$Exercise.per.week..days. + data$GPA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.60118 -0.44470  0.02124  0.47005  0.94620
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.004459   0.819041    2.447   0.0207 *
## data$Sleep..hrs.    0.039654   0.113734    0.349   0.7299
## data$Smoke...VapeSometimes -0.535886   0.714461   -0.750   0.4593
## data$Smoke...VapeYes    -0.376890   0.256799   -1.468   0.1530
## data$Exercise.per.week..days.  0.461414   0.074287    6.211 8.94e-07 ***
## data$GPA            0.004559   0.009394    0.485   0.6311
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6533 on 29 degrees of freedom
## Multiple R-squared:  0.6076, Adjusted R-squared:  0.54
## F-statistic: 8.981 on 5 and 29 DF, p-value: 3.042e-05

#checking the added variable plot
avPlots(physical_model)
```


Added-Variable Plots



#Not satisfied with the added variable plot, so proceed with reduced model (single linear model)

```
slm=lm(data$Physical.Health.Rating~
        data$Exercise.per.week..days.)
summary(slm)
```

```
##
## Call:
## lm(formula = data$Physical.Health.Rating ~ data$Exercise.per.week..days.)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4687 -0.3655 -0.2624  0.6345  0.7376
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.46867    0.17095   14.441 8.05e-16 ***
## data$Exercise.per.week..days.  0.44843    0.06946    6.456 2.54e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6499 on 33 degrees of freedom
## Multiple R-squared:  0.5581, Adjusted R-squared:  0.5447
## F-statistic: 41.68 on 1 and 33 DF, p-value: 2.538e-07
```

#Checking the AIC and BIC

```
AIC(physical_model,slm)
```

```
##           df      AIC
## physical_model  7 76.94270
## slm             3 73.10179
```

```
BIC(physical_model,slm)
```

```
##           df      BIC
## physical_model  7 87.83014
## slm             3 77.76783
```

#both AIC and BIC prefer the reduced model

Testing the model for Mental Health

#creating the full model

```
mental_model=lm(data$Mental.Health.Rating~data$Mental.Health.Prioritization+d
ata$Depression.Scale+data$Suicide.Thoughts)
```

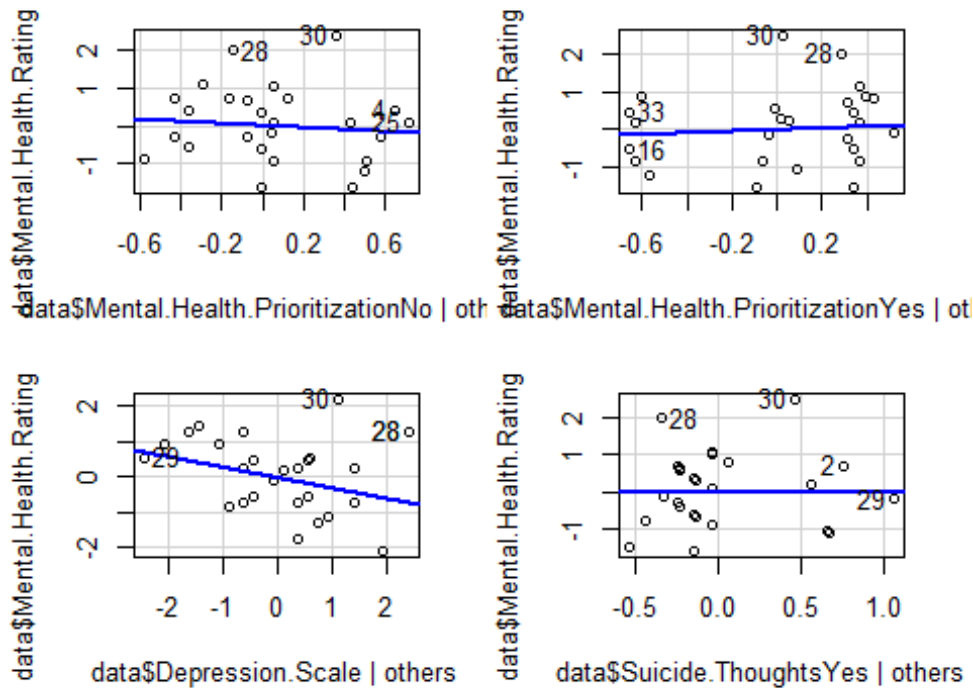
```
summary(mental_model)
```

```
##
## Call:
## lm(formula = data$Mental.Health.Rating ~ data$Mental.Health.Prioritization
+
##   data$Depression.Scale + data$Suicide.Thoughts)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.62245 -0.65274  0.08222  0.59765  2.49813
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   4.27370     0.49775   8.586 1.4e-09
## data$Mental.Health.PrioritizationNo -0.26602     0.47713  -0.558  0.5813
## data$Mental.Health.PrioritizationYes  0.23475     0.38963   0.603  0.5514
## data$Depression.Scale          -0.29534     0.15688  -1.883  0.0695.
## data$Suicide.ThoughtsYes       -0.02914     0.44313  -0.066  0.9480
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9713 on 30 degrees of freedom
## Multiple R-squared:  0.1853, Adjusted R-squared:  0.07667
## F-statistic: 1.706 on 4 and 30 DF,  p-value: 0.1748
```

#checking the added variable plot

```
avPlots(mental_model)
```

Added-Variable Plots



#Not satisfied with the added variable, proceed with the reduced model without Suicide

```
reduced_model = lm(data$Mental.Health.Rating~data$Depression.Scale)
summary(reduced_model)
```

```
##
## Call:
## lm(formula = data$Mental.Health.Rating ~ data$Depression.Scale)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.75731 -0.59593 -0.09276  0.57179  2.24269
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.4346     0.4310  10.290 7.87e-12 ***
## data$Depression.Scale -0.3355     0.1414  -2.372  0.0237 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9484 on 33 degrees of freedom
## Multiple R-squared:  0.1457, Adjusted R-squared:  0.1198
## F-statistic: 5.626 on 1 and 33 DF, p-value: 0.02368
```

#Checking the AIC and BIC

```
AIC(mental_model,reduced_model)
```

```
##           df      AIC
## mental_model  6 103.89518
## reduced_model  3  99.55784

BIC(mental_model,reduced_model)

##           df      BIC
## mental_model  6 113.2273
## reduced_model  3 104.2239

#Both AIC and BIC prefer a reduced model
```

Appendix D:

Research Area: Israa and Kaylee

Research Question: Israa and Kaylee

Preliminary Research and Hypothesis: Kaylee

Survey: Israa

Data Cleaning: Israa and Kaylee

Data Analysis: Israa and Kaylee

Hypothesis Testing: Israa and Kaylee