INDIVIDUAL ASSIGMENT

TECHNOLOGY PARK MALAYSIA

CT127-3-2-PFDA

PROGRAMMING FOR DATA ANALYSIS

APD2F2202CS(IS)

HAND OUT DATE : 11 April 2022

HAND IN DATE : 23 May 2022

WEIGHTAGE : 50%

NAME : Mr. Amardeep Singh AL Uttam Singh

NAME : Faithlin Hoe Wei Xin

TP NUMBER : TP055347

Table of Contents

1.0 Introduction and Assumptions

The assignment demands us to investigate a dataset including data from three-year final scores of degree students from multiple schools in order to understand what impacted the outcome of their performance. This dataset contains 33 columns with 16 integer datatype values,9 string values and 8 boolean values that consists the personal details, academic performance, family backgrounds and daily routines. The goal of this research is using RStudio to find out the factors that influenced their academic performance and provide useful information for decision making.

1.1 Assumptions

The dataset is quite unclear due to insufficient explaination. There were multiple factors that might've affected the data either. Thus, multiple assumptions were made during the research.

The outcome of the research is to be observed and used with care.

2.0 Data Import / Cleaning / Pre-Processing / Transformation

2.1 Data Import

```
#Importing Data
dataOfStudent = read.csv("D:\\APU Stuff\\Bachelor of Computer Science (Intelligent Systems) [APD2F2202CS(IS)]\\
                          Semester 1\\Programming for Data Analysis [CT127-3-2-PFDA]\\Assignment\\student.csv" ,header=TRUE)
View(dataOfStudent)
```

Diagram 1: The Dataset is Imported

The dataset is imported with the format of csv into RStudio. The “read.csv” function was used to read the csv file. The “header=TRUE” is to accept the headers in the csv file as parameters.

2.2 Data Cleaning

```
data_cleaning = function() {
  ducplicate_data = function(){
    tidied_data %>% group_by(Index) %>%
      filter(n() > 1) %>% summarise(count = n()) %>% print()
  }
  ducplicate_data() # no duplicate data
}
data_cleaning()
```

Diagram 2: The data set has been cleaned and checked for any duplicated data

The data cleaning process is placed in a function block. It checks for any duplicated data.

2.3 Data Pre-Processing

```
#Data Pre-Processing
names(dataOfStudent)=c("Index","School_Name","Sex","Age","Home_Area","Family_Size","Parents_Maritial_Status",
                       "Mothers_Education_Level","Fathers_Education_Level","Mothers_Job","Fathers_Job","Reason_Of_School_Chosen",
                       "Guardian","Travelling_Time_to_School","Weekly_Study_Time","Amount_of_Fails_in_Class",
                       "Extra_Educational_Support","Family_Educational_Support","Extra_Paid_Classes",
                       "Participation_in_Extra_Activities","Attendance_in_Nursery_School","Intention_to_Further_Higher_Education",
                       "Internet_Access_at_Home","Romantic_Relationship","Quality_of_Family_Relationship","Free_Time_After_School",
                       "Goes_Out_with_Friends","Workday_Alcohol_Consumption","Weekend_Alcohol_Consumption","Current_Health_Status",
                       "Number_of_Absences","Math_First_Period_Grade","Math_Second_Period_Grade","Math_Final_Grade")
tidied_data <- dataOfStudent
```

Diagram 3: Appropriate Headers Have been Assigned

Appropriate headers have been assigned to the dataset with the "name" function. Then, it is placed into a variable named "tidied_data".

```
#The combined grades is to use as a measurement for the students' performance
combined_grades <- tidied_data$Math_First_Period_Grade + tidied_data$Math_Second_Period_Grade + tidied_data$Math_Final_Grade
```

Diagram 4: A variable named "combined_grades" has been prepared for later analysis

As the research is to understand how certain factors affect a student's performance. A variable named "combined_grades" is prepared to act as a measure to check if the student is a performer.

```
tidied_data <- cbind(tidied_data, combined_grades)

write.csv(tidied_data, file = "D:\\APU Stuff\\Bachelor of Computer Science (Intelligent Systems) [APD2F2202CS(IS)]\\
          Semester 1\\Programming for Data Analysis [CT127-3-2-PFDA]\\Assignment\\Tidied_Data.csv", row.names=FALSE)
```

Diagram 5: Data Being Saved into the CSV File

After the dataset has been preprocessed and the new variable "combined_grades" is prepared, they are added into a new data frame named "tidied_data" so that it made be used in further analysis.

The function "write.csv" is used to assign and save the data into a new csv file.

2.4 Data Exploration

```
#Exploring the Data
class(dataOfStudent) #data frame type
length(dataOfStudent) #34 Attributes
summary(dataOfStudent) #Summarize data of all attributes, no missing values are found
```

Diagram 6: Dataset being Explored

Some simple functions like “class”, “length” and “summary” were used to further understand the dataset.

Then, it was to decide how it should be tidied.

3.0 How Does the Family Background Affects Students' Performance?

3.1 Family Size vs Marks

```
familyAnalysis = function(){
  summary(tidied_data$combined_grades)

  data4Family <- select(tidied_data, "Family_Size","combined_grades")
  data4Family

  good_result_GT3 <- nrow(data4Family[(data4Family$Family_Size == "GT3") & (data4Family$combined_grades > 40.00),])
  good_result_LE3 <- nrow(data4Family[(data4Family$Family_Size == "LE3") & (data4Family$combined_grades > 40.00),])

  GT3 <- nrow(data4Family[data4Family$Family_Size == "GT3",])
  LE3 <- nrow(data4Family[data4Family$Family_Size == "LE3",])

  percentageOfPerformers_GT3 <- (good_result_GT3/GT3)*100
  percentageOfPerformers_LE3 <- (good_result_LE3/LE3)*100

  good_result <- c(percentageOfPerformers_GT3,percentageOfPerformers_LE3)

  barplot(good_result,
          main = "Family Size vs Percentage of Performers",
          xlab = "Family Size",
          ylab = "Percentage of Performers(%)",
          names.arg = c("Greater Than 3", "Less Than or Equal To 3"),
          col = "darkred",
          horiz = FALSE)

}

familyAnalysis()
```

Diagram 7: The Analysis to Find the Relationship Between Family Size and Students' Performance

First, we have found out the 1st quartile and 3rd quartile of the "combined_grades" of the students with the "summary" function. This was done to find the performers' grades.

For ease of analyzing, a new data frame "data4Family" was made, it consists of data from "Family_Size" and "combined_grades".

We then found the number of students that come from a family with members greater than 3, and also performed well in their studies. The number of students that come from a family with members less than or even to 3, and also performed well in their studies is also found.

Then, the percentage of performers that come from the different family sizes were calculated by dividing by the total numbers of students from the respective family sizes.

It was placed in a bar graph to see the relationship between family sizes and performers.

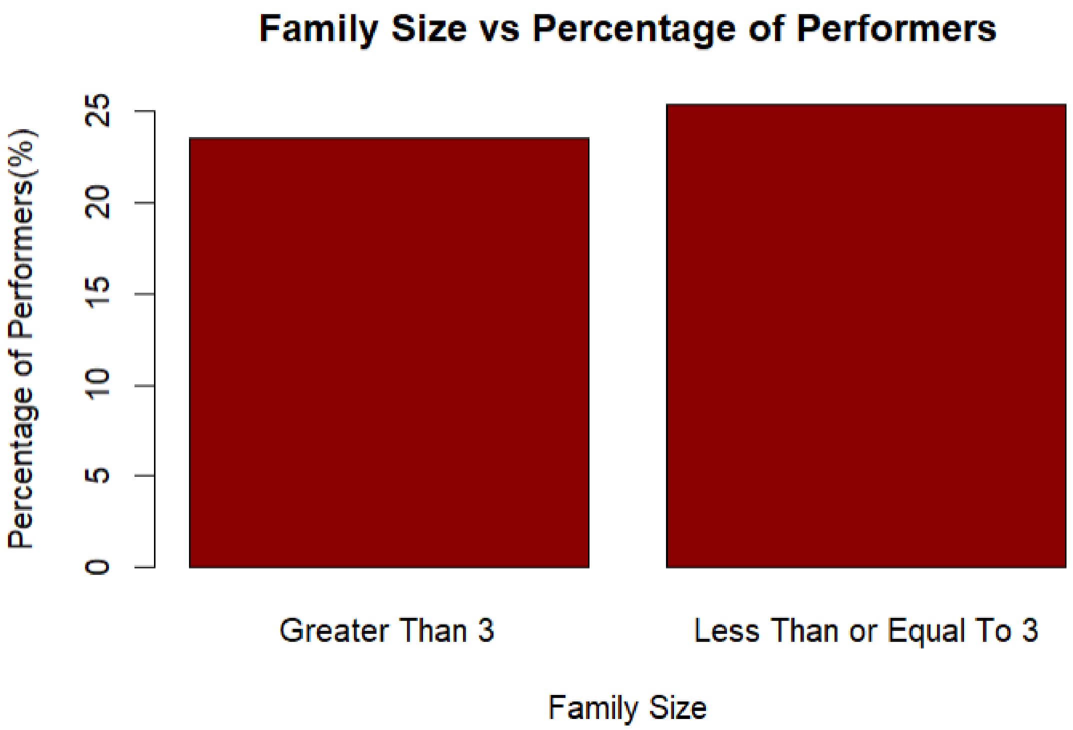

Diagram 8: Bar Graph from the Analysis "Family Size vs Performers"

From this graph, we can see that students with a smaller family size has a higher percentage in performing well in their studies.

As said by Onzaberigu(2017), as the size of the family grows, parents tend to fail to pay attention to every individual child, fail to provide sufficient financial support and many other reasons.

3.2 Parent's Marriage Status vs Absences vs Marks

```
marriageStatusAnalysis = function(){
  data4MarriageStatus <- select(tidied_data, "Parents_Maritial_Status","Number_of_Absences","combined_grades")
  data4MarriageStatus

  summary(tidied_data$Number_of_Absences)

  T <- nrow(data4MarriageStatus[data4MarriageStatus$Parents_Maritial_Status == "T",])
  A <- nrow(data4MarriageStatus[data4MarriageStatus$Parents_Maritial_Status == "A",])

  T_HAR <- nrow(data4MarriageStatus[(data4MarriageStatus$Parents_Maritial_Status == "T") & (data4MarriageStatus$Number_of_Absences > 8.0),])
  A_HAR <- nrow(data4MarriageStatus[(data4MarriageStatus$Parents_Maritial_Status == "A") & (data4MarriageStatus$Number_of_Absences > 8.0),])

  AbsenceRate_T <- (T_HAR/T)*100
  AbsenceRate_A <- (A_HAR/A)*100

  Parents_Maritial_Status_vs_Absence_Rate <- c(AbsenceRate_T,AbsenceRate_A)

  good_result_T <- nrow(data4MarriageStatus[(data4MarriageStatus$Parents_Maritial_Status == "T") & (data4MarriageStatus$combined_grades > 40.00),])
  good_result_A <- nrow(data4MarriageStatus[(data4MarriageStatus$Parents_Maritial_Status == "A") & (data4MarriageStatus$combined_grades > 40.00),])

  percentageOfPerformers_T <- (good_result_T/T)*100
  percentageOfPerformers_A <- (good_result_A/A)*100

  Parents_Maritial_Status_vs_Performance <- c(percentageOfPerformers_T,percentageOfPerformers_A)

  totalAbsence <- sum(data4MarriageStatus$Number_of_Absences)
   performer_HAR <- nrow(data4MarriageStatus[(data4MarriageStatus$Number_of_Absences > 8.0) & (data4MarriageStatus$combined_grades > 40.00),])
   performer_LAR <- nrow(data4MarriageStatus[(data4MarriageStatus$Number_of_Absences < 4.0) & (data4MarriageStatus$combined_grades > 40.00),])

   percentageOfPerformers_HAR <- (performer_HAR/totalAbsence)*100
   percentageOfPerformers_LAR <- (performer_LAR/totalAbsence)*100

   Absence_Rate_vs_Performance <- c(percentageOfPerformers_HAR, percentageOfPerformers_LAR)

   marriageStatusAnalysis <- cbind(Parents_Maritial_Status_vs_Absence_Rate, Absence_Rate_vs_Performance, Parents_Maritial_Status_vs_Performance)

   barplot(marriageStatusAnalysis,beside=T)

}

marriageStatusAnalysis()
```

Diagram 9: The Analysis of Marriage Status and Absence and Marks

This analysis was done to find the relationship of how the marriage status of the parents of the student can affect the students' absence and marks. It also analyzed how absence can affect the performance of the student.

First, a data frame "data4MarriageStatus" was made for the ease of the analysis. It has the data from "Parents_Maritial_Status", "Number_of_Absences" and "combined_grades".

High absence is determined by the 3rd quartile of the "Number_of_Absences".

Then, the number of students that have parents that are together and also highly absent in their classes are calculated. The number of students that have parents that are apart and also high absent in their classes are also calculated.

Its percentage is then calculated by dividing with its respective total number of parents that are either together or apart. The percentages are then calculated and placed into variables for further analysis.

Next, the number of students that have parents that are together and also a performer are calculated. The number of students that have parents that are apart and also a performer are also calculated.

Its percentage is then calculated by dividing with its respective total number of parents that are either together or apart. The percentages are then calculated and placed into variables for further analysis.

Lastly, the number of students that are highly absent in their class and also a performer is calculated. Then, the number of students that are highly present in their class and also a performer is calculated.

Its percentage is then calculated by dividing with the total number of absences. The percentages are then calculated and placed into variables for further analysis.

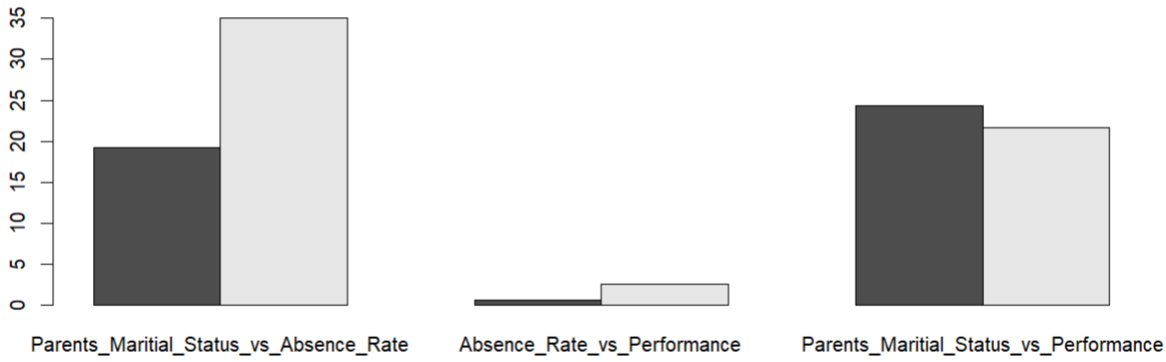

Diagram 10: The Graphs for Parents’ Maritial Status and Absence Rate and Performance

From this analysis, we can see that students with parents that are not together have a higher absence rate. We can also see that students with higher absence rate have a higher chances in performing poorly in their studies. As conclusion, we can say that students with parents that are not together have higher chances in performing poorly in their studies. This hypothesis is proven by the bar graph.

It is said that a child's mental state can be greatly disrupted by their parent's split. Student with divorced parents have higher percentage of skipping class and dropping out of classes, which is also a reason why they perform poorer in their studies. As the parents have split, the students usually lose financial stability either, which causes their performance to drop. (Marripedia, n.d.)

3.3 Parent's Education Level vs Marks

```
parentELAnalysis = function(){

  combined_level <- tidied_data$Fathers_Education_Level + tidied_data$Mothers_Education_Level

  tidied_data <- cbind(tidied_data, combined_level)

  write.csv(tidied_data, file = "D:\\APU Stuff\\Bachelor of Computer Science (Intelligent Systems) [APD2F2202CS(IS)]\\Semester 1
            \\Programming for Data Analysis [CT127-3-2-PFDA]\\Assignment\\Tidied_Data.csv", row.names=FALSE)

  summary(tidied_data$combined_level)

  data4EduLevel <- select(tidied_data, "combined_level","combined_grades")
  data4EduLevel

  good_result_HEL <- nrow(data4EduLevel[(data4EduLevel$combined_level > 7.0) & (data4EduLevel$combined_grades > 40.00),])
  good_result_LEL <- nrow(data4EduLevel[(data4EduLevel$combined_level < 4.0) & (data4EduLevel$combined_grades > 40.00),])

  HEL <- nrow(data4EduLevel[data4EduLevel$combined_level > 7.0,])
  LEL <- nrow(data4EduLevel[data4EduLevel$combined_level < 4.0,])

  percentageOfPerformers_HEL <- (good_result_HEL/HEL)*100
  percentageOfPerformers_LEL <- (good_result_LEL/LEL)*100

  good_result <- c(percentageOfPerformers_HEL,percentageOfPerformers_LEL)

  barplot(good_result,
          main = "Parent's Education Level vs Percentage of Performers",
          xlab = "Parent's Education Level",
          ylab = "Percentage of Performers(%)",
          names.arg = c("High Education Level", "Low Education Level"),
          col = "blue",
          horiz = FALSE)
}

parentELAnalysis()
```

Diagram 11: The Analysis of Parents’ Education Level and Students’ Performance

First, a variable “combined_level” is made to combine the education level of the students’ father and mother. It is then added into the “tidied_data” csv for further analysis.

The “combined_level”’s 1st quartile is used as a measurement of low education level and 3rd quartile is used as a measurement of high education level.

A data frame “data4EduLevel” is made and has data from “combined_level” and “combined_grades”.

The students with parents that have high level of education and are also a performer are calculated. The students with parents that have a lower level of education and are also a performer are calculated. It is then divided with total numbers of its respective education level to find the percentage.

It is then analyzed and placed into a graph.

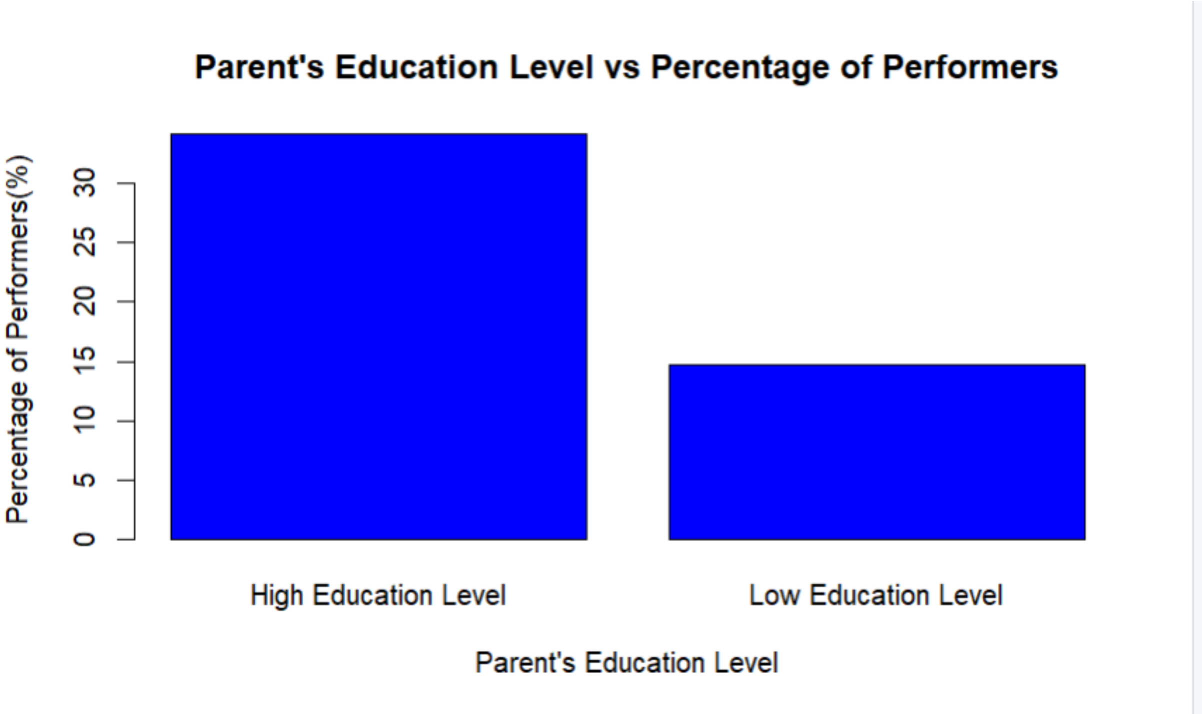

Diagram 12: The Relationship between Parents’ Education Level and Performers

As the graph shown, students that have parents with higher education level perform better in their studies.

As said, parents with high education level tend to model achievement-oriented behavior and provide opportunities to their child to model it. The child may develop the belief that achievement-oriented behavior is valued and expected. The parents also tend to spend more time with their children and developing their talents and skills. (Clearinghouse, 2020)

3.4 Father's Job vs Marks

```
fatherJobAnalysis = function(){

  data4FatherJob <- select(tidied_data, "Fathers_Job","combined_grades")
  data4FatherJob

  checkCategories <- factor(data4FatherJob$Fathers_Job)
  levels(checkCategories)

  good_result_home <- nrow(data4FatherJob[(data4FatherJob$Fathers_Job == "at_home") & (data4FatherJob$combined_grades > 40.00),])
  good_result_health <- nrow(data4FatherJob[(data4FatherJob$Fathers_Job == "health") & (data4FatherJob$combined_grades > 40.00),])
  good_result_services <- nrow(data4FatherJob[(data4FatherJob$Fathers_Job == "services") & (data4FatherJob$combined_grades > 40.00),])
  good_result_teacher <- nrow(data4FatherJob[(data4FatherJob$Fathers_Job == "teacher") & (data4FatherJob$combined_grades > 40.00),])
  good_result_other <- nrow(data4FatherJob[(data4FatherJob$Fathers_Job == "other") & (data4FatherJob$combined_grades > 40.00),])


  nHome <- nrow(data4FatherJob[data4FatherJob$Fathers_Job == "at_home",])
  nHealth <- nrow(data4FatherJob[data4FatherJob$Fathers_Job == "health",])
  nServices <- nrow(data4FatherJob[data4FatherJob$Fathers_Job == "services",])
  nTeacher <- nrow(data4FatherJob[data4FatherJob$Fathers_Job == "teacher",])
  nOther <- nrow(data4FatherJob[data4FatherJob$Fathers_Job == "other",])

  percentageOfPerformers_Home <- (good_result_home/nHome)*100
  percentageOfPerformers_Health <- (good_result_health/nHealth)*100
  percentageOfPerformers_Services <- (good_result_services/nServices)*100
  percentageOfPerformers_Teacher <- (good_result_teacher/nTeacher)*100
  percentageOfPerformers_Other <- (good_result_other/nOther)*100

  good_result <- c(percentageOfPerformers_Home,percentageOfPerformers_Health,percentageOfPerformers_Services,percentageOfPerformers_Teacher,
                   percentageOfPerformers_Other)

  barplot(good_result,
          main = "Father's Job vs Percentage of Performers",
          xlab = "Father's Job",
          ylab = "Percentage of Performers(%)",
          names.arg = c("Stay at Home","Health","Services","Teacher","Others"),
          col = "green",
          horiz = FALSE)
}

fatherJobAnalysis()
```

Diagram 13: Analysis for Father's Occupation and Performance of Students

First, a data frame "data4FatherJob" is made and has "Fathers_Job" and "combined_grades". The categories in "Fathers_Job" is checked using the function "levels" and used for further analysis.

Then, the numbers of performers with respective father jobs are calculated. They are then divided by the total number of students with respective father jobs. The rates are used to see the relationship of father's jobs and performers.

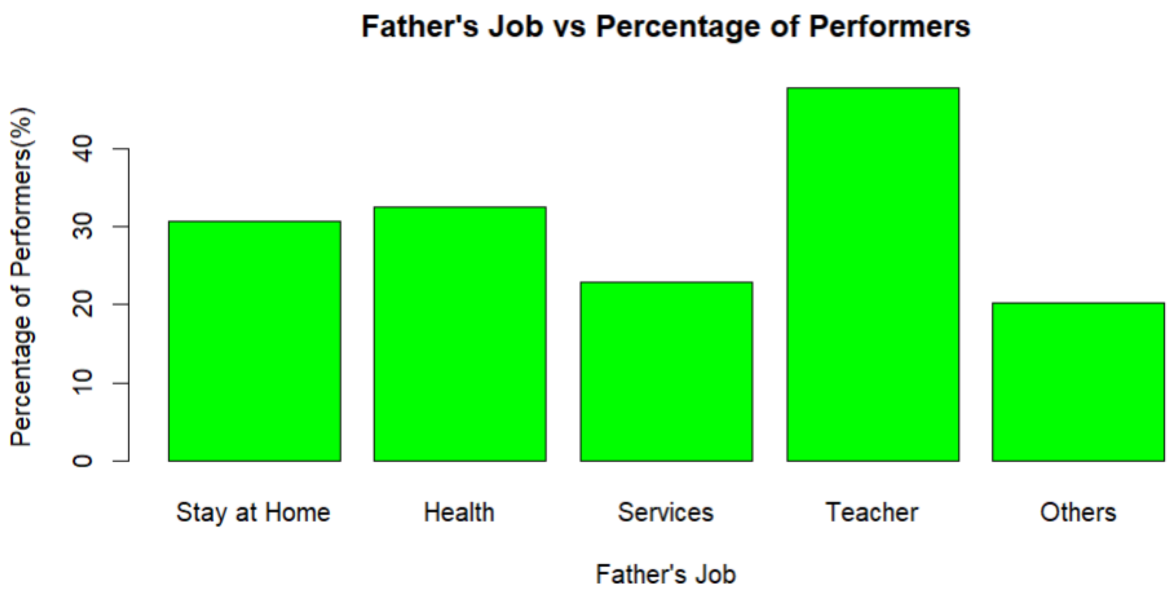

Diagram 14: The Graph that Shows the Relationship of Father’s Job and Percentage of Performers

As seen in the graph, it is shown that students with fathers that are teachers have the highest probability of performing well in the studies. Followed by health, stay at home, services then others.

It is said that children with parents as educators have a higher probability of performing well in their studies. It is found that their parents, as educators, provide a pro-education environment, which can help with their children's performance in their studies. It is also found the parents that are educators help out their children in their schoolwork, which helps with their performance in school. (Denny K., 2011)

3.5 Mother's Job vs Marks

```
motherJobAnalysis = function(){

  data4MotherJob <- select(tidied_data, "Mothers_Job","combined_grades")
  data4MotherJob

  checkCategories <- factor(data4MotherJob$Mothers_Job)
  levels(checkCategories)

  good_result_home <- nrow(data4MotherJob[(data4MotherJob$Mothers_Job == "at_home") & (data4MotherJob$combined_grades > 40.00),])
  good_result_health <- nrow(data4MotherJob[(data4MotherJob$Mothers_Job == "health") & (data4MotherJob$combined_grades > 40.00),])
  good_result_services <- nrow(data4MotherJob[(data4MotherJob$Mothers_Job == "services") & (data4MotherJob$combined_grades > 40.00),])
  good_result_teacher <- nrow(data4MotherJob[(data4MotherJob$Mothers_Job == "teacher") & (data4MotherJob$combined_grades > 40.00),])
  good_result_other <- nrow(data4MotherJob[(data4MotherJob$Mothers_Job == "other") & (data4MotherJob$combined_grades > 40.00),])


  nHome <- nrow(data4MotherJob[data4MotherJob$Mothers_Job == "at_home",])
  nHealth <- nrow(data4MotherJob[data4MotherJob$Mothers_Job == "health",])
  nServices <- nrow(data4MotherJob[data4MotherJob$Mothers_Job == "services",])
  nTeacher <- nrow(data4MotherJob[data4MotherJob$Mothers_Job == "teacher",])
  nOther <- nrow(data4MotherJob[data4MotherJob$Mothers_Job == "other",])

  percentageOfPerformers_Home <- (good_result_home/nHome)*100
  percentageOfPerformers_Health <- (good_result_health/nHealth)*100
  percentageOfPerformers_Services <- (good_result_services/nServices)*100
  percentageOfPerformers_Teacher <- (good_result_teacher/nTeacher)*100
  percentageOfPerformers_Other <- (good_result_other/nOther)*100

  good_result <- c(percentageOfPerformers_Home,percentageOfPerformers_Health,percentageOfPerformers_Services,percentageOfPerformers_Teacher,
                   percentageOfPerformers_Other)

  barplot(good_result,
          main = "Mother's Job vs Percentage of Performers",
          xlab = "Mother's Job",
          ylab = "Percentage of Performers(%)",
          names.arg = c("Stay at Home","Health","Services","Teacher","Others"),
          col = "darkgreen",
          horiz = FALSE)
}

motherJobAnalysis()
```

Diagram 15: Analysis for Mother’s Occupation and Performance of Students

First, a data frame “data4MotherJob” is made and has “Mothers_Job” and “combined_grades”. The categories in “Mothers_Job” is checked using the function “levels” and used for further analysis.

Then, the numbers of performers with respective mother jobs are calculated. They are then divided by the total number of students with respective mother jobs. The rates are used to see the relationship of mother’s jobs and performers.

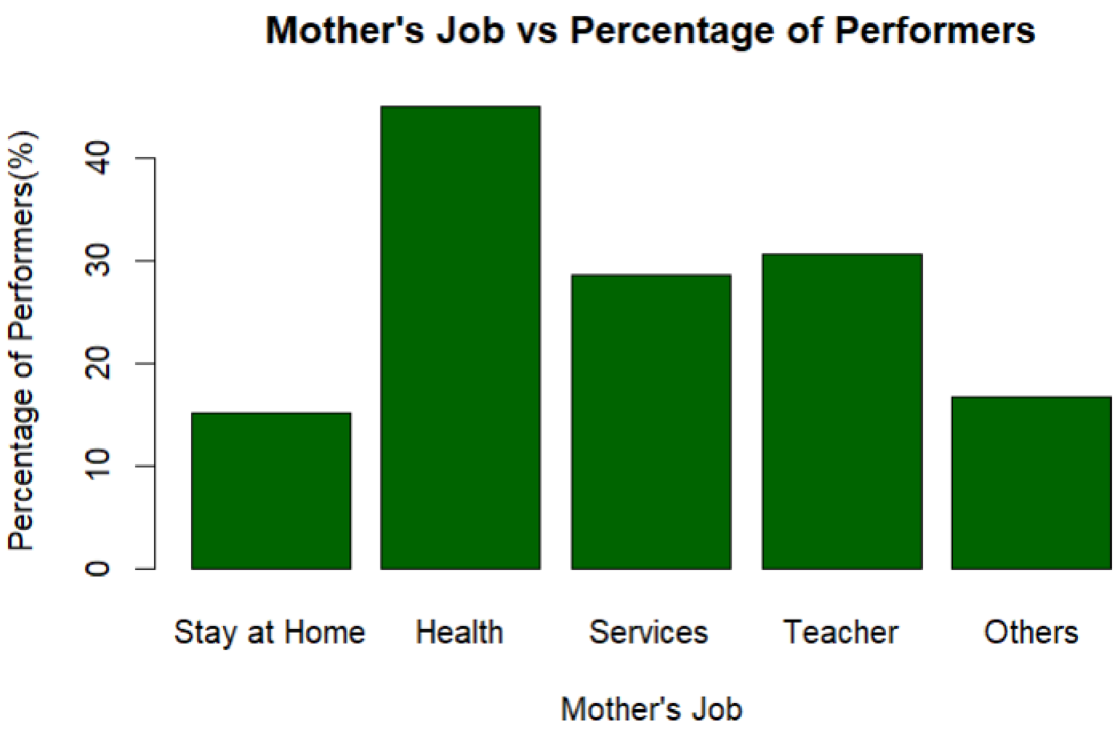

Diagram 16: The Graph that Shows the Relationship of Mother's Job and Percentage of Performers

It is shown that most performers have mothers that have work related to health. It is then followed by teacher, services, others then stay at home.

It is shown that parents that have health related work earn more. Thus, providing financial stability to the family. Children that study under a stable financial condition can perform better. As a conclusion, children with parents from health related work earn good money, provides a good and stable environment, and students performs better in such condition. (CJ Heinrich, 2014)

4.0 How Does Educational Support Affects Students' Performance?

4.1 School Supplement vs Marks

```
supplementAnalysis = function(){

  data4Supplement <- select(tidied_data, "Extra_Educational_Support","combined_grades")
  data4Supplement

  good_result_YS <- nrow(data4Supplement[(data4Supplement$Extra_Educational_Support == "yes") & (data4Supplement$combined_grades > 40.00),])
  good_result_NS <- nrow(data4Supplement[(data4Supplement$Extra_Educational_Support == "no") & (data4Supplement$combined_grades > 40.00),])

  YS <- nrow(data4Supplement[data4Supplement$Extra_Educational_Support == "yes",])
  NS <- nrow(data4Supplement[data4Supplement$Extra_Educational_Support == "no",])

  percentageOfPerformers_YS <- (good_result_YS/YS)*100
  percentageOfPerformers_NS <- (good_result_NS/NS)*100

  good_result <- c(percentageOfPerformers_YS,percentageOfPerformers_NS)

  barplot(good_result,
          main = "Extra Educational Support (School Supplements) vs Percentage of Performers",
          xlab = "Extra Educational Support (School Supplements)",
          ylab = "Percentage of Performers(%)",
          names.arg = c("Yes", "No"),
          col = "purple",
          horiz = FALSE)
}

supplementAnalysis()
```

Diagram 17: Analysis for School Supplement and Marks

First, a data frame “data4Supplement” that has “Extra_Educational_Support” and “combined_grade” is made.

The number of performers that has supplement and not have are calculated. It is then divided by the total number of students with its respective supplement status to find the percentage.

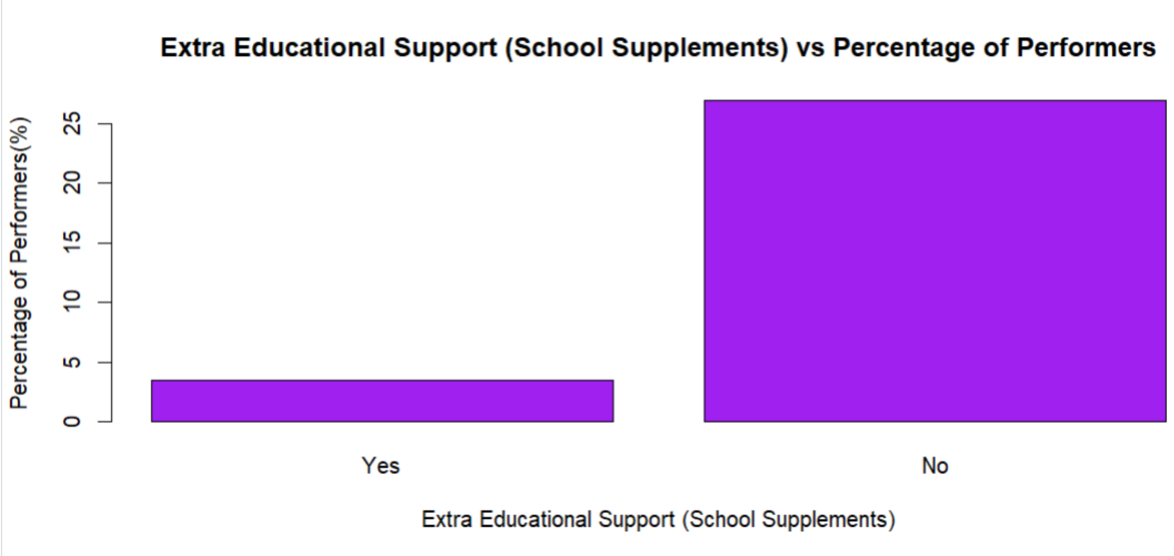

Diagram 18: The Graph with School Supplements and Percentage of Performers

Based on the graph, we can see that students that took extra educational support has a lower chance of performing in their studies.

4.2 Family Support vs Marks

```
fSupportAnalysis = function(){

  data4fSupport <- select(tidied_data, "Family_Educational_Support","combined_grades")
  data4fSupport

  good_result_YFS <- nrow(data4fSupport[(data4fSupport$Family_Educational_Support == "yes") & (data4fSupport$combined_grades > 40.00),])
  good_result_NFS <- nrow(data4fSupport[(data4fSupport$Family_Educational_Support == "no") & (data4fSupport$combined_grades > 40.00),])

  YFS <- nrow(data4fSupport[data4fSupport$Family_Educational_Support == "yes",])
  NFS <- nrow(data4fSupport[data4fSupport$Family_Educational_Support == "no",])

  percentageOfPerformers_YFS <- (good_result_YFS/YFS)*100
  percentageOfPerformers_NFS <- (good_result_NFS/NFS)*100

  good_result <- c(percentageOfPerformers_YFS,percentageOfPerformers_NFS)

  barplot(good_result,
          main = "Family Educational Support vs Percentage of Performers",
          xlab = "Family Educational Support",
          ylab = "Percentage of Performers(%)",
          names.arg = c("Yes", "No"),
          col = "Yellow",
          horiz = FALSE)
}

fSupportAnalysis()
```

Diagram 19: Analysis for Family Support and Marks

First, a data frame “data4fSupport” that has “Family_Educational Support” and “combined_grade” is made.

The number of performers that has family educational support and not have are calculated. It is then divided by the total number of students with its respective family educational support status to find the percentage.

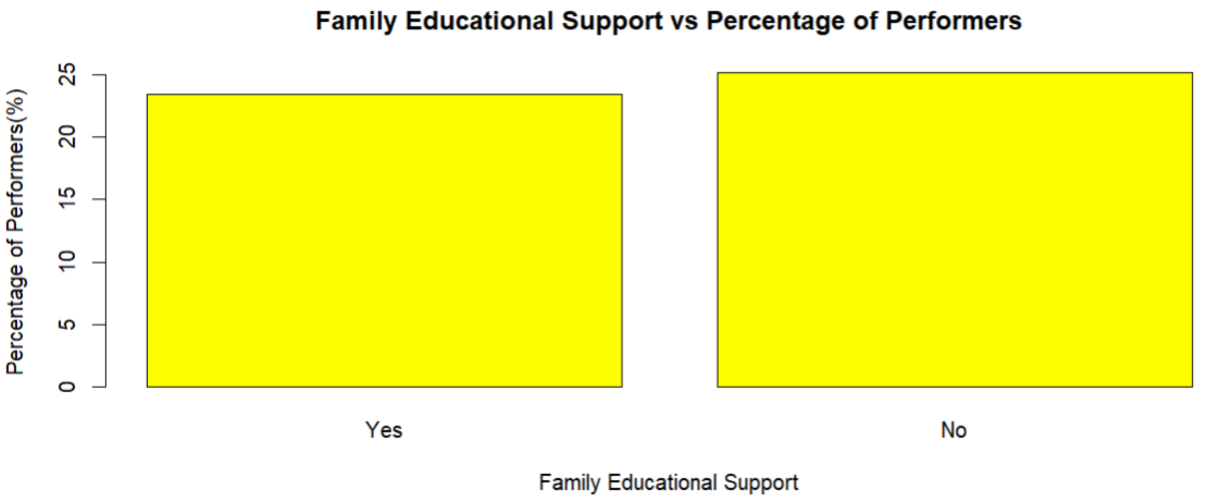

Diagram 20: The Graph with Family Educational Support and Percentage of Performers

Based on the graph, we can see that students that have family educational support has a lower chance of performing in their studies.

However, there is not a distinct difference in the rates. Thus, we can say that family members support does not really affect students’ performance.

4.3 Extra Paid Classes vs Marks

```
ePaidAnalysis = function(){

  data4EPClasses <- select(tidied_data, "Extra_Paid_Classes","combined_grades")
  data4EPClasses

  good_result_Y <- nrow(data4EPClasses[(data4EPClasses$Extra_Paid_Classes == "yes") & (data4EPClasses$combined_grades > 40.00),])
  good_result_N <- nrow(data4EPClasses[(data4EPClasses$Extra_Paid_Classest == "no") & (data4EPClasses$combined_grades > 40.00),])

  Y <- nrow(data4EPClasses[data4EPClasses$Extra_Paid_Classes == "yes",])
  N <- nrow(data4EPClasses[data4EPClasses$Extra_Paid_Classes == "no",])

  percentageOfPerformers_Y <- (good_result_Y/Y)*100
  percentageOfPerformers_N <- (good_result_N/N)*100

  good_result <- c(percentageOfPerformers_Y,percentageOfPerformers_N)

  barplot(good_result,
          main = "Extra Paid Classes vs Percentage of Performers",
          xlab = "Extra Paid Classes",
          ylab = "Percentage of Performers(%)",
          names.arg = c("Yes", "No"),
          col = "White",
          horiz = FALSE)
}

ePaidAnalysis()
```

Diagram 21: Analysis for Extra Paid Classes within the Course and Marks

First, a data frame “data4EPClasses” that has “Extra_Paid_Classes” and “combined_grade” is made.

The number of performers that has extra paid classes within the course and not have are calculated. It is then divided by the total number of students with its respective extra paid classes status to find the percentage.

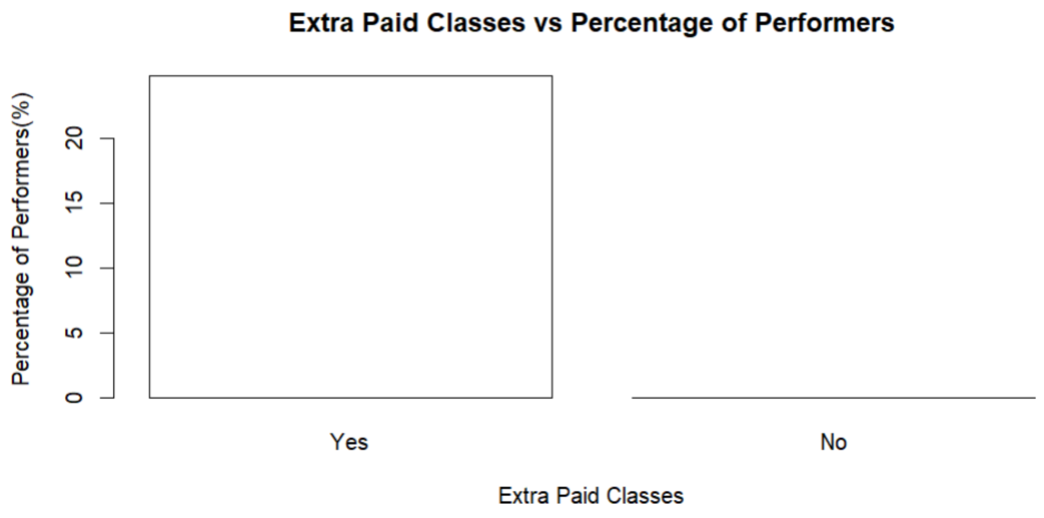

Diagram 22: The Graph with Family Educational Support and Percentage of Performers

As we can see in the graph, the percentage of performers are a lot higher for students that took extra paid classes within the course.

It is said that students get focused attention from teachers within the extra classes. As students learn, they might get confused. Having a teacher focused on them gives more time, and teachers can go into detail and explain better to the student. Not only that, the environment in the extra classes is usually pressure-free. In extra classes, students are given more time to completely immerse into the new materials until they fully understand them. Thus, they might feel more relaxed and do not need to worry about not being to catch up to the others. Lastly, it frees the student from worrying. Extra classes help students feel more secure in their grasp of the materials. All of these benefits have proven to help students perform better. (abaKus, n.d.)

5.0 How Does Relationships Affects Students' Performance?

5.1 Romantic Relationship vs Marks

```
romanticAnalysis = function(){

  data4RR <- select(tidied_data, "Romantic_Relationship","combined_grades")
  data4RR

  good_result_Y <- nrow(data4RR[(data4RR$Romantic_Relationship == "yes") & (data4RR$combined_grades > 40.00),])
  good_result_N <- nrow(data4RR[(data4RR$Romantic_Relationship == "no") & (data4RR$combined_grades > 40.00),])

  Y <- nrow(data4RR[data4RR$Romantic_Relationship == "yes",])
  N <- nrow(data4RR[data4RR$Romantic_Relationship == "no",])

  percentageOfPerformers_Y <- (good_result_Y/Y)*100
  percentageOfPerformers_N <- (good_result_N/N)*100

  good_result <- c(percentageOfPerformers_Y,percentageOfPerformers_N)

  barplot(good_result,
          main = "Romantic Relationships vs Percentage of Performers",
          xlab = "Romantic Relationships",
          ylab = "Percentage of Performers(%)",
          names.arg = c("Yes", "No"),
          col = "Black",
          horiz = FALSE)

}

romanticAnalysis()
```

Diagram 23: Analysis for Romantic Relationship and Marks

First, a data frame "data4RR" that has "Romantic_Relationship" and "combined_grade" is made.

The number of performers that has romantic relationship and not have are calculated. It is then divided by the total number of students with its respective romantic relationship status to find the percentage.

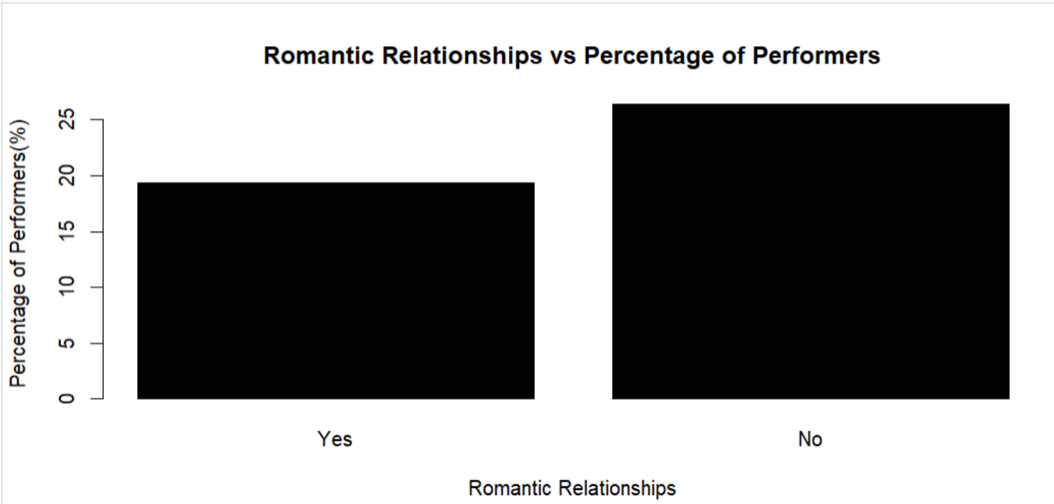

Diagram 24: The Graph with Family Educational Support and Percentage of Performers

As shown in the graph, we can see that students that are not in a romantic relationship perform better than students that are in one.

It is said that as romantic relationships cause students to lose concentration. Students would spend time with their romantic partners instead of studying. Not only that, the couple might get into arguments. This causes stress which would further distract the students from learning. These are the reasons why romantic relationship are not advisable for students that want to perform well in their studies. (Muychrea L., Vatanak S., Sopheavatey T., n.d.)

5.2 Family Relationship vs Marks

```
familyRelationshipAnalysis = function(){

  data4FR <- select(tidied_data, "Quality_of_Family_Relationship","combined_grades")
  data4FR

  good_result_1FR <- nrow(data4FR[(data4FR$Quality_of_Family_Relationship == 1) & (data4FR$combined_grades > 40.00),])
  good_result_2FR <- nrow(data4FR[(data4FR$Quality_of_Family_Relationship == 2) & (data4FR$combined_grades > 40.00),])
  good_result_3FR <- nrow(data4FR[(data4FR$Quality_of_Family_Relationship == 3) & (data4FR$combined_grades > 40.00),])
  good_result_4FR <- nrow(data4FR[(data4FR$Quality_of_Family_Relationship == 4) & (data4FR$combined_grades > 40.00),])
  good_result_5FR <- nrow(data4FR[(data4FR$Quality_of_Family_Relationship == 5) & (data4FR$combined_grades > 40.00),])

  FR1 <- nrow(data4FR[data4FR$Quality_of_Family_Relationship == 1,])
  FR2 <- nrow(data4FR[data4FR$Quality_of_Family_Relationship == 2,])
  FR3 <- nrow(data4FR[data4FR$Quality_of_Family_Relationship == 3,])
  FR4 <- nrow(data4FR[data4FR$Quality_of_Family_Relationship == 4,])
  FR5 <- nrow(data4FR[data4FR$Quality_of_Family_Relationship == 5,])

  percentageOfPerformers_FR1 <- (good_result_1FR/FR1)*100
  percentageOfPerformers_FR2 <- (good_result_2FR/FR2)*100
  percentageOfPerformers_FR3 <- (good_result_3FR/FR3)*100
  percentageOfPerformers_FR4 <- (good_result_4FR/FR4)*100
  percentageOfPerformers_FR5 <- (good_result_5FR/FR5)*100

  good_result <- c(percentageOfPerformers_FR1,percentageOfPerformers_FR2,percentageOfPerformers_FR3,percentageOfPerformers_FR4,
                   percentageOfPerformers_FR5)

  barplot(good_result,
          main = "Quality of Family Relationship vs Percentage of Performers",
          xlab = "Quality of Family Relationship",
          ylab = "Percentage of Performers(%)",
          names.arg = c("1", "2","3","4","5"),
          col = "pink",
          horiz = FALSE)
}

familyRelationshipAnalysis()
```

Diagram 25: Analysis for Quality of Family Relationship and Marks

First, a data frame "data4FR" is made and has "Quality_of_Family_Relationship" and "combined_grades".

Then, the numbers of performers with respective quality of family relationship are calculated. They are then divided by the total number of students with respective quality of family relationship. The rates are used to see the relationship of family relationship and performers.

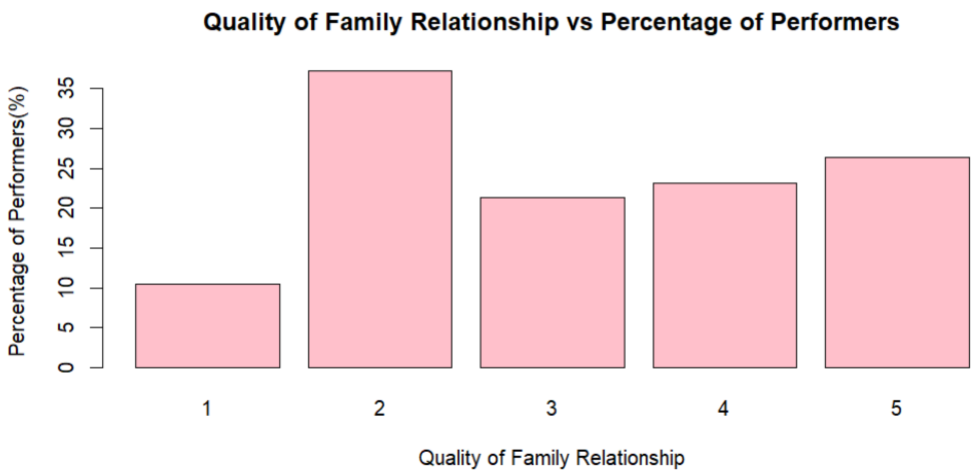

Diagram 26: The Graph with Quality of Family Relationship and Percentage of Performers

As shown in the graph, students with family quality - 2, performs the best in their studies. As followed, family quality - 5, 4, 3, then 1.

6.0 How Do the Students Spend Their Time Affects Their Performance?

6.1 Free time vs Marks

```
freeTimeAnalysis = function(){

  data4FT <- select(tidied_data, "Free_Time_After_School","combined_grades")
  data4FT

  good_result_1FT <- nrow(data4FT[(data4FT$Free_Time_After_School == 1) & (data4FT$combined_grades > 40.00),])
  good_result_2FT <- nrow(data4FT[(data4FT$Free_Time_After_School == 2) & (data4FT$combined_grades > 40.00),])
  good_result_3FT <- nrow(data4FT[(data4FT$Free_Time_After_School == 3) & (data4FT$combined_grades > 40.00),])
  good_result_4FT <- nrow(data4FT[(data4FT$Free_Time_After_School == 4) & (data4FT$combined_grades > 40.00),])
  good_result_5FT <- nrow(data4FT[(data4FT$Free_Time_After_School == 5) & (data4FT$combined_grades > 40.00),])

  FT1 <- nrow(data4FT[data4FT$Free_Time_After_School == 1,])
  FT2 <- nrow(data4FT[data4FT$Free_Time_After_School == 2,])
  FT3 <- nrow(data4FT[data4FT$Free_Time_After_School == 3,])
  FT4 <- nrow(data4FT[data4FT$Free_Time_After_School == 4,])
  FT5 <- nrow(data4FT[data4FT$Free_Time_After_School == 5,])

  percentageOfPerformers_FT1 <- (good_result_1FT/FT1)*100
  percentageOfPerformers_FT2 <- (good_result_2FT/FT2)*100
  percentageOfPerformers_FT3 <- (good_result_3FT/FT3)*100
  percentageOfPerformers_FT4 <- (good_result_4FT/FT4)*100
  percentageOfPerformers_FT5 <- (good_result_5FT/FT5)*100

  good_result <- c(percentageOfPerformers_FT1,percentageOfPerformers_FT2,percentageOfPerformers_FT3,
                   percentageOfPerformers_FT4,percentageOfPerformers_FT5)

  barplot(good_result,
          main = "Free Time vs Percentage of Performers",
          xlab = "Free Time",
          ylab = "Percentage of Performers(%)",
          names.arg = c("1", "2","3","4","5"),
          col = "red",
          horiz = FALSE)
}

freeTimeAnalysis()
```

Diagram 27: Analysis for Free Time and Marks

First, a data frame “data4FT” is made and has “Free_Time_After_School” and “combined_grades”.

Then, the numbers of performers with respective free time after school are calculated. They are then divided by the total number of students with respective free time after school. The rates are used to see the relationship of free time after school and performers.

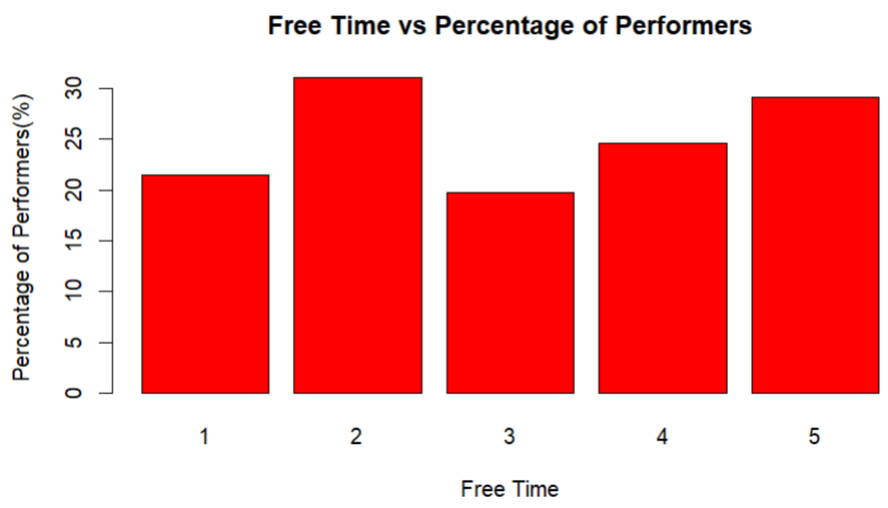

Diagram 28: The Graph with Free Time and Percentage of Performers

As shown in the graph, we can see that students with not a lot of free time after school perform better. As followed is students with a lot of free time, students with quite much of free time, student with none to very little free time, and lastly students with moderate free time.

The reason might be that student with not a lot of free time after school spend more time in school studying or revising, thus resulting in them getting a better result. However, people with a lot of free time comes next. It might be because they spend a lot of time studying in their free time.

Further Analysis is done to prove this hypothesis.

6.2 Weekly Study Time vs Marks

```
studyAnalysis = function(){

  data4FTS <- select(tidied_data, "Free_Time_After_School","Weekly_Study_Time","combined_grades")
  data4FTS

  checkCategories <- factor(data4FTS$Weekly_Study_Time)
  levels(checkCategories)

  good_result_2n1ST <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 2) & (data4FTS$Weekly_Study_Time == 1) & (data4FTS$combined_grades > 40.00),])
  good_result_2n2ST <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 2) & (data4FTS$Weekly_Study_Time == 2) & (data4FTS$combined_grades > 40.00),])
  good_result_2n3ST <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 2) & (data4FTS$Weekly_Study_Time == 3) & (data4FTS$combined_grades > 40.00),])
  good_result_2n4ST <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 2) & (data4FTS$Weekly_Study_Time == 4) & (data4FTS$combined_grades > 40.00),])

  ST2n1 <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 2) & (data4FTS$Weekly_Study_Time == 1),])
  ST2n2 <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 2) & (data4FTS$Weekly_Study_Time == 2),])
  ST2n3 <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 2) & (data4FTS$Weekly_Study_Time == 3),])
  ST2n4 <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 2) & (data4FTS$Weekly_Study_Time == 4),])

  percentageOfPerformers_ST2n1 <- (good_result_2n1ST/ST2n1)*100
  percentageOfPerformers_ST2n2 <- (good_result_2n2ST/ST2n2)*100
  percentageOfPerformers_ST2n3 <- (good_result_2n3ST/ST2n3)*100
  percentageOfPerformers_ST2n4 <- (good_result_2n4ST/ST2n4)*100

  good_result_2FST <- c(percentageOfPerformers_ST2n1,percentageOfPerformers_ST2n2,percentageOfPerformers_ST2n3,percentageOfPerformers_ST2n4)

  good_result_5n1ST <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 5) & (data4FTS$Weekly_Study_Time == 1) & (data4FTS$combined_grades > 40.00),])
  good_result_5n2ST <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 5) & (data4FTS$Weekly_Study_Time == 2) & (data4FTS$combined_grades > 40.00),])
  good_result_5n3ST <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 5) & (data4FTS$Weekly_Study_Time == 3) & (data4FTS$combined_grades > 40.00),])
  good_result_5n4ST <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 5) & (data4FTS$Weekly_Study_Time == 4) & (data4FTS$combined_grades > 40.00),])

  ST5n1 <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 5) & (data4FTS$Weekly_Study_Time == 1),])
  ST5n2 <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 5) & (data4FTS$Weekly_Study_Time == 2),])
  ST5n3 <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 5) & (data4FTS$Weekly_Study_Time == 3),])
  ST5n4 <- nrow(data4FTS[(data4FTS$Free_Time_After_School == 5) & (data4FTS$Weekly_Study_Time == 4),])

  percentageOfPerformers_ST5n1 <- (good_result_5n1ST/ST5n1)*100
  percentageOfPerformers_ST5n2 <- (good_result_5n2ST/ST5n2)*100
  percentageOfPerformers_ST5n3 <- (good_result_5n3ST/ST5n3)*100
  percentageOfPerformers_ST5n4 <- (good_result_5n4ST/ST5n4)*100

  good_result_5FST <- c(percentageOfPerformers_ST5n1,percentageOfPerformers_ST5n2,percentageOfPerformers_ST5n3,percentageOfPerformers_ST5n4)

  FreeTimeStudyAnalysis <- cbind(good_result_2FST,good_result_5FST)

  barplot(FreeTimeStudyAnalysis,beside=T)
}

studyAnalysis()
```

Diagram 29: The Analysis of Studying in Free Time (2 and 5) and Marks

This analysis was done to find the relationship of how the students that does not have a lot of time and have a lot of time, how much do they study during their free time and how it affects their performance in school.

First, a data frame “data4FTS” was made for the ease of the analysis. It has the data from “Free_Time_After_School”, “Weekly_Study_Time” and “combined_grades”.

The numbers of categories in the “Weekly_Study_Time” are confirmed by using the “levels” function.

Then, the performers that does not have a lot of free time, with different study time are each calculated and stored in different variables. It is then divided by total numbers of students in its respective study time to find the rate.

Next, the performers that have a lot of free time, with different study time are each calculated and stored in different variables. It is then divided by total numbers of students in its respective study time to find the rate.

Lastly, they are placed in a graph to see the relationship of the performers that have different amount of free time and different amount of study hours.

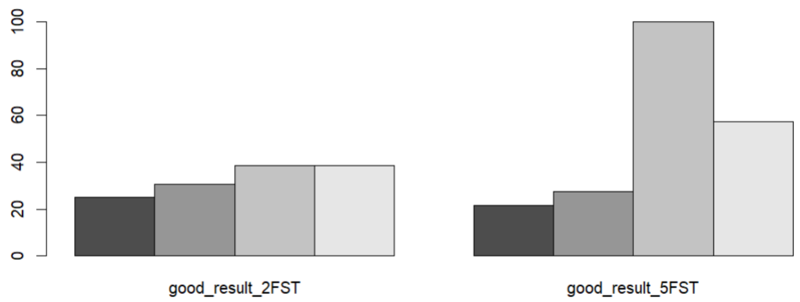

Diagram 30: The Graphs of Studying in Free Time (2 and 5) and Marks

As shown in the graphs, students that have not a lot of free time and have a lot of free time after school share a common aspect. Which is that, students that perform well in their studies both spend a lot of time studying during their free time. As we can see in the second graph, students with a lot of free time, which also spend a lot of said free time studying performs a lot better in their studies.

As a conclusion, we can say that students who perform well in their studies spend a lot of their free time studying.

6.3 Frequency of Hanging Out with Friends vs Marks

```r
friendsAnalysis = function(){

  data4FTF <- select(tidied_data, "Free_Time_After_School","Goes_Out_with_Friends","combined_grades")
  data4FTF

  checkCategories <- factor(data4FTF$Goes_Out_with_Friends)
  levels(checkCategories)

  good_result_2n1F <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 2) & (data4FTF$Goes_Out_with_Friends == 1) & (data4FTF$combined_grades > 40.00),])
  good_result_2n2F <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 2) & (data4FTF$Goes_Out_with_Friends == 2) & (data4FTF$combined_grades > 40.00),])
  good_result_2n3F <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 2) & (data4FTF$Goes_Out_with_Friends == 3) & (data4FTF$combined_grades > 40.00),])
  good_result_2n4F <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 2) & (data4FTF$Goes_Out_with_Friends == 4) & (data4FTF$combined_grades > 40.00),])
  good_result_2n5F <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 2) & (data4FTF$Goes_Out_with_Friends == 5) & (data4FTF$combined_grades > 40.00),])

  F2n1 <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 2) & (data4FTF$Goes_Out_with_Friends == 1),])
  F2n2 <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 2) & (data4FTF$Goes_Out_with_Friends == 2),])
  F2n3 <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 2) & (data4FTF$Goes_Out_with_Friends == 3),])
  F2n4 <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 2) & (data4FTF$Goes_Out_with_Friends == 4),])
  F2n5 <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 2) & (data4FTF$Goes_Out_with_Friends == 5),])


  percentageOfPerformers_F2n1 <- (good_result_2n1F/F2n1)*100
  percentageOfPerformers_F2n2 <- (good_result_2n2F/F2n2)*100
  percentageOfPerformers_F2n3 <- (good_result_2n3F/F2n3)*100
  percentageOfPerformers_F2n4 <- (good_result_2n4F/F2n4)*100
  percentageOfPerformers_F2n5 <- (good_result_2n5F/F2n5)*100

  good_result_2FFT <- c(percentageOfPerformers_F2n1,percentageOfPerformers_F2n2,percentageOfPerformers_F2n3,percentageOfPerformers_F2n4,percentageOfPerformers_F2n5)
  good_result_5n1F <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 5) & (data4FTF$Goes_Out_with_Friends == 1) & (data4FTF$combined_grades > 40.00),])
  good_result_5n2F <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 5) & (data4FTF$Goes_Out_with_Friends == 2) & (data4FTF$combined_grades > 40.00),])
  good_result_5n3F <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 5) & (data4FTF$Goes_Out_with_Friends == 3) & (data4FTF$combined_grades > 40.00),])
  good_result_5n4F <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 5) & (data4FTF$Goes_Out_with_Friends == 4) & (data4FTF$combined_grades > 40.00),])
  good_result_5n5F <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 5) & (data4FTF$Goes_Out_with_Friends == 5) & (data4FTF$combined_grades > 40.00),])

  F5n1 <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 5) & (data4FTF$Goes_Out_with_Friends == 1),])
  F5n2 <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 5) & (data4FTF$Goes_Out_with_Friends == 2),])
  F5n3 <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 5) & (data4FTF$Goes_Out_with_Friends == 3),])
  F5n4 <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 5) & (data4FTF$Goes_Out_with_Friends == 4),])
  F5n5 <- nrow(data4FTF[(data4FTF$Free_Time_After_School == 5) & (data4FTF$Goes_Out_with_Friends == 5),])


  percentageOfPerformers_F5n1 <- (good_result_5n1F/F5n1)*100
  percentageOfPerformers_F5n2 <- (good_result_5n2F/F5n2)*100
  percentageOfPerformers_F5n3 <- (good_result_5n3F/F5n3)*100
  percentageOfPerformers_F5n4 <- (good_result_5n4F/F5n4)*100
  percentageOfPerformers_F5n5 <- (good_result_5n5F/F5n5)*100

  good_result_5FFT <- c(percentageOfPerformers_F5n1,percentageOfPerformers_F5n2,percentageOfPerformers_F5n3,percentageOfPerformers_F5n4,percentageOfPerformers_F5n5)

  FreeTimeFriendAnalysis <- cbind(good_result_2FFT,good_result_5FFT)

  barplot(FreeTimeFriendAnalysis,beside=T)
}

friendsAnalysis()
```

Diagram 31: The Analysis of Students that Hangs Out with Friends in Free Time (2 and 5) and Marks

This analysis was done to find the relationship of how the students that does not have a lot of time and have a lot of time, how much do they hang out with their friend during their free time and how it affects their performance in school.

First, a data frame “data4FTF” was made for the ease of the analysis. It has the data from “Free_Time_After_School”, “Goes_Out_with_Friends” and “combined_grades”.

The numbers of categories in the “Goes_Out_with_Friends” are confirmed by using the “levels” function.

Then, the performers that does not have a lot of free time, with different going out with friend time are each calculated and stored in different variables. It is then divided by total numbers of students in its respective going out with friend time to find the rate.

Next, the performers that have a lot of free time, with different hanging out with friend time are each calculated and stored in different variables. It is then divided by total numbers of students in its respective hanging out with friend time to find the rate.

Lastly, they are placed in a graph to see the relationship of the performers that have different amount of free time and different amount of hanging out with friend hours.

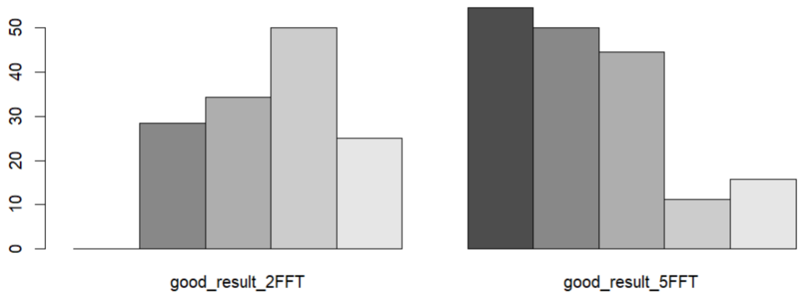

Diagram 32: The Graphs of Going Out with Friends in Free Time (2 and 5) and Marks

As shown in the graphs, students that have not a lot of free time and have a lot of free time both do not spend a lot of their free time hanging out with their friends.

We can most likely assume that it is because that the students spend a lot of their free time studying rather than hanging out with friends.

6.4 Alcohol Consumption vs Marks

```r
alcoholAnalysis = function(){

  combined_alcohol_consumption <- tidied_data$Workday_Alcohol_Consumption + tidied_data$Weekend_Alcohol_Consumption

  tidied_data <- cbind(tidied_data, combined_alcohol_consumption)

  write.csv(tidied_data, file = "D:\\APU Stuff\\Bachelor of Computer Science (Intelligent Systems) [APD2F2202CS(IS)]\\Semester 1\\
            Programming for Data Analysis [CT127-3-2-PFDA]\\Assignment\\Tidied_Data.csv", row.names=FALSE)

  data4FTA <- select(tidied_data, "Free_Time_After_School","combined_alcohol_consumption","combined_grades")
  data4FTA

  checkCategories <- factor(data4FTA$combined_alcohol_consumption)
  levels(checkCategories)

  good_result_2n2A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 2) & (data4FTA$combined_grades > 40.00),])
  good_result_2n3A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 3) & (data4FTA$combined_grades > 40.00),])
  good_result_2n4A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 4) & (data4FTA$combined_grades > 40.00),])
  good_result_2n5A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 5) & (data4FTA$combined_grades > 40.00),])
  good_result_2n6A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 6) & (data4FTA$combined_grades > 40.00),])
  good_result_2n7A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 7) & (data4FTA$combined_grades > 40.00),])
  good_result_2n8A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 8) & (data4FTA$combined_grades > 40.00),])
  good_result_2n9A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 9) & (data4FTA$combined_grades > 40.00),])
  good_result_2n10A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 10) & (data4FTA$combined_grades > 40.00),])

  A2n2 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 2),])
  A2n3 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 3),])
  A2n4 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 4),])
  A2n5 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 5),])
  A2n6 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 6),])
  A2n7 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 7),])
  A2n8 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 8),])
  A2n9 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 9),])
  A2n10 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 2) & (data4FTA$combined_alcohol_consumption == 10),])
 percentageOfPerformers_A2n2 <- (good_result_2n2A/A2n2)*100
 percentageOfPerformers_A2n3 <- (good_result_2n3A/A2n3)*100
 percentageOfPerformers_A2n4 <- (good_result_2n4A/A2n4)*100
 percentageOfPerformers_A2n5 <- (good_result_2n5A/A2n5)*100
 percentageOfPerformers_A2n6 <- (good_result_2n6A/A2n6)*100
 percentageOfPerformers_A2n7 <- (good_result_2n7A/A2n7)*100
 percentageOfPerformers_A2n8 <- (good_result_2n8A/A2n8)*100
 percentageOfPerformers_A2n9 <- (good_result_2n9A/A2n9)*100
 percentageOfPerformers_A2n10 <- (good_result_2n10A/A2n10)*100

 good_result_2FAT <- c(percentageOfPerformers_A2n2,percentageOfPerformers_A2n3,percentageOfPerformers_A2n4,percentageOfPerformers_A2n5,percentageOfPerformers_A2n6,
                       percentageOfPerformers_A2n7,percentageOfPerformers_A2n8,percentageOfPerformers_A2n9,percentageOfPerformers_A2n10)

 good_result_5n2A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 2) & (data4FTA$combined_grades > 40.00),])
 good_result_5n3A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 3) & (data4FTA$combined_grades > 40.00),])
 good_result_5n4A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 4) & (data4FTA$combined_grades > 40.00),])
 good_result_5n5A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 5) & (data4FTA$combined_grades > 40.00),])
 good_result_5n6A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 6) & (data4FTA$combined_grades > 40.00),])
 good_result_5n7A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 7) & (data4FTA$combined_grades > 40.00),])
 good_result_5n8A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 8) & (data4FTA$combined_grades > 40.00),])
 good_result_5n9A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 9) & (data4FTA$combined_grades > 40.00),])
 good_result_5n10A <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 10) & (data4FTA$combined_grades > 40.00),])

 A5n2 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 2),])
 A5n3 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 3),])
 A5n4 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 4),])
 A5n5 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 5),])
 A5n6 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 6),])
 A5n7 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 7),])
 A5n8 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 8),])
 A5n9 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 9),])
 A5n10 <- nrow(data4FTA[(data4FTA$Free_Time_After_School == 5) & (data4FTA$combined_alcohol_consumption == 10),])
  percentageOfPerformers_A5n2 <- (good_result_5n2A/A5n2)*100
  percentageOfPerformers_A5n3 <- (good_result_5n3A/A5n3)*100
  percentageOfPerformers_A5n4 <- (good_result_5n4A/A5n4)*100
  percentageOfPerformers_A5n5 <- (good_result_5n5A/A5n5)*100
  percentageOfPerformers_A5n6 <- (good_result_5n6A/A5n6)*100
  percentageOfPerformers_A5n7 <- (good_result_5n7A/A5n7)*100
  percentageOfPerformers_A5n8 <- (good_result_5n8A/A5n8)*100
  percentageOfPerformers_A5n9 <- (good_result_5n9A/A5n9)*100
  percentageOfPerformers_A5n10 <- (good_result_5n10A/A5n10)*100


  good_result_5FAT <- c(percentageOfPerformers_A5n2,percentageOfPerformers_A5n3,percentageOfPerformers_A5n4,percentageOfPerformers_A5n5,percentageOfPerformers_A5n6,
                        percentageOfPerformers_A5n7,percentageOfPerformers_A5n8,percentageOfPerformers_A5n9,percentageOfPerformers_A5n10)

  FreeAlcoholTimeAnalysis <- cbind(good_result_2FAT,good_result_5FAT)

  barplot(FreeAlcoholTimeAnalysis,beside=T)
}

alcoholAnalysis()
```

Diagram 33: The Analysis of Consuming Alcohol in Free Time (2 and 5) and Marks

This analysis was done to find the relationship of how the students that does not have a lot of time and have a lot of time, how much do they consume alcohol during their free time and how it affects their performance in school.

First, a variable “combined_alcohol_consumption” is made to combine the weekend and workday alcohol consumption of the students. It is then added into the “tidied_data” csv for further analysis.

A data frame “data4FTA” was made for the ease of the analysis. It has the data from “Free_Time_After_School”, “combined_alcohol_consumption” and “combined_grades”.

The numbers of categories in the “combined_alcohol_consumption” are confirmed by using the “levels” function.

Then, the performers that does not have a lot of free time, with different alcohol consumption time are each calculated and stored in different variables. It is then divided by total numbers of students in its respective alcohol consumption time to find the rate.

Next, the performers that have a lot of free time, with different alcohol consumption time are each calculated and stored in different variables. It is then divided by total numbers of students in its respective alcohol consumption time to find the rate.

Lastly, they are placed in a graph to see the relationship of the performers that have different amount of free time and different amount of alcohol consumption hours.

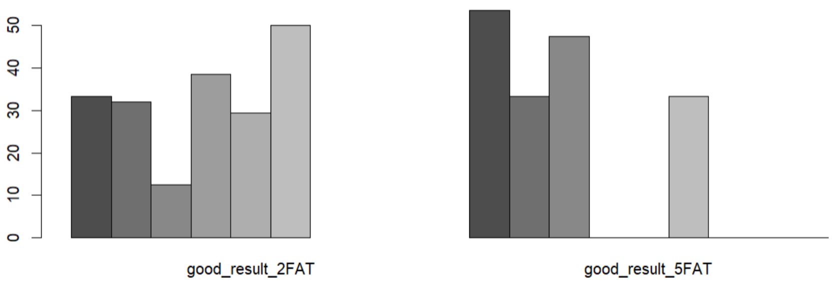

Diagram 34: The Graphs of Consuming Alcohol in Free Time (2 and 5) and Marks

As shown in the 1st graph, we can see that performers with not a lot of free time after school mostly consume a lot of alcohol. It might be because that they needed alcohol to relief their stress after spending a lot of time at school. People that are less stressed are usually more healthy mentally, and being healthy mentally means that they can study and focus better. As shown in the 2nd graph, we can see that performers with a lot of free time after school mostly only consume a little of alcohol on their free time. It might be because they spend a lot of their time studying in their free time.

6.5 The Factors Mentioned in the Previous 3 Analysis vs Free Time

```
fFactorsAnalysis = function(){

  combined_alcohol_consumption <- tidied_data$Workday_Alcohol_Consumption + tidied_data$Weekend_Alcohol_Consumption
  tidied_data <- cbind(tidied_data, combined_alcohol_consumption)
  write.csv(tidied_data, file = "D:\\APU Stuff\\Bachelor of Computer Science (Intelligent Systems) [APD2F2202CS(IS)]\\
          Semester 1\\Programming for Data Analysis [CT127-3-2-PFDA]\\Assignment\\Tidied_Data.csv", row.names=FALSE)


  studyPercentage <- tidied_data$Weekly_Study_Time/tidied_data$Free_Time_After_School
  friendPercentage <- tidied_data$Goes_Out_with_Friends/tidied_data$Free_Time_After_School
  alcoholPercentage <- tidied_data$combined_alcohol_consumption/tidied_data$Free_Time_After_School

  tidied_data <- cbind(tidied_data, studyPercentage, friendPercentage, alcoholPercentage)
  write.csv(tidied_data, file = "D:\\APU Stuff\\Bachelor of Computer Science (Intelligent Systems) [APD2F2202CS(IS)]\\
          Semester 1\\Programming for Data Analysis [CT127-3-2-PFDA]\\Assignment\\Tidied_Data.csv", row.names=FALSE)

  data43A <- select(tidied_data, "Free_Time_After_School","studyPercentage","friendPercentage","alcoholPercentage","combined_grades")
  data43A

  performers_2F <- subset(data43A, Free_Time_After_School == 2 & combined_grades > 40.0)
  performers_2F

  totalStudyTime2F <- sum(performers_2F$studyPercentage)
  totalFriendTime2F <- sum(performers_2F$friendPercentage)
  totalAlcoholTime2F <- sum(performers_2F$alcoholPercentage)

  sum <- nrow(tidied_data)

  percentageStudy2F <- totalStudyTime2F/sum
  percentageFriend2F <- totalFriendTime2F/sum
  percentageAlcohol2F <- totalAlcoholTime2F/sum

  FT2 <- c(percentageStudy2F, percentageFriend2F, percentageAlcohol2F)

  performers_5F <- subset(data43A, Free_Time_After_School == 5 & combined_grades > 40.0)
  performers_5F
  totalStudyTime5F <- sum(performers_5F$studyPercentage)
  totalFriendTime5F <- sum(performers_5F$friendPercentage)
  totalAlcoholTime5F <- sum(performers_5F$alcoholPercentage)

  sum <- nrow(tidied_data)

  percentageStudy5F <- totalStudyTime5F/sum
  percentageFriend5F <- totalFriendTime5F/sum
  percentageAlcohol5F <- totalAlcoholTime5F/sum

  FT5 <- c(percentageStudy5F, percentageFriend5F, percentageAlcohol5F)

  FreeTimeAnalysis <- cbind(FT2,FT5)

  barplot(FreeTimeAnalysis,beside=T)
}

fFactorsAnalysis()
```

Diagram 35: The Analysis of How Different Actions Done During Free Time (2 and 5) Can Affect the Students' Performance

This analysis was done to find the relationship of the actions of the performers and their free time.

First, variables "combined_alcohol_consumption", "studyPercentage", "friendPercentage" and "alcoholPercentage" are made to combine the weekend and workday alcohol consumption of the students, calculate the percentage of time spent on studying on their free time, the percentage of time spent on hanging out with friends on their free time and the percentage of time spent on consuming alcohol on their free time. It is then added into the “tidied_data” csv for further analysis.

A data frame “data43A” was made for the ease of the analysis. It has the data from “Free_Time_After_School”, “studyPercentage”,”friendPercentage”,”alcoholPercentage”, and “combined_grades”.

Then, the performers that does not have a lot of free time are filtered out and placed in a variable “performers_2F” for the ease of analysis.

Next, the total percentage of performers that spent their free time studying, going out with friends and consuming alcohol are summed up. It is used to find the percentage of performers that does not have a lot of free time and study during their free time and number of performers.

After that, the performers that does have a lot of free time are filtered out and placed in a variable “performers_5F” for the ease of analysis.

The total percentage of performers that spent their free time studying, going out with friends and consuming alcohol are then summed up. It is used to find the percentage of performers that have a lot of free time and study during their free time and number of performers.

Lastly, they are placed in a graph to see the relationship of the performers that have different amount of free time and how did they spent it.

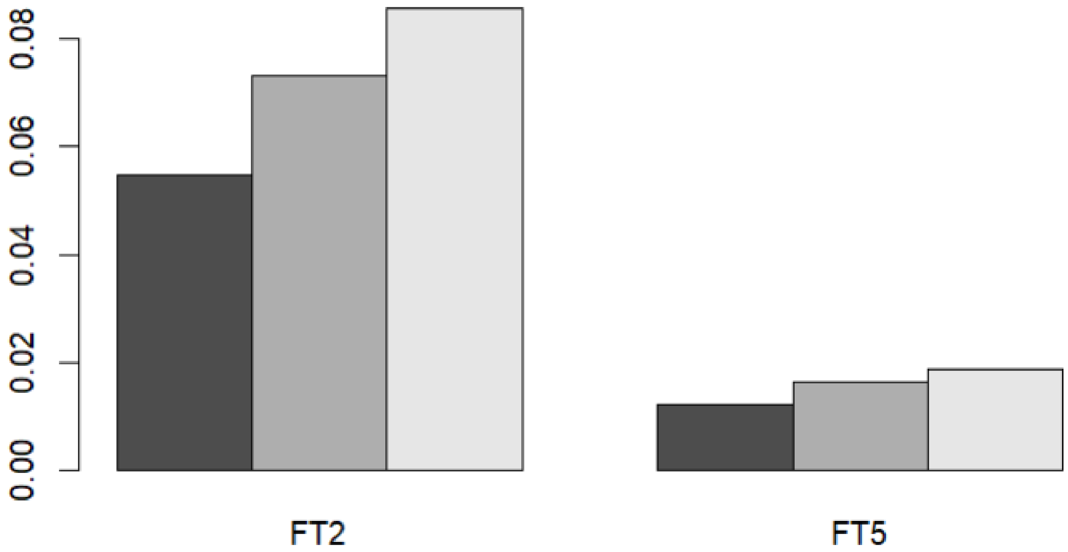

Diagram 36: The Graph on How Different Actions Done During Free Time Can Affect the Students' Performance

As shown in the graphs, the performers spend a lot of time consuming alcohol and hanging out with friends. Then only, studying.

It is said that consuming alcohol and hanging out with friends can reduce stress and freshen up our mind. Thus, performers usually freshen up their mind and also study to keep them focused and may perform better.

7.0 How Do the Students' Education Background Affect Their Performance?

7.1 School vs Marks

```
schoolAnalysis = function(){

  data4School <- select(tidied_data, "School_Name","combined_grades")
  data4School

  checkCategories <- factor(data4School$School_Name)
  levels(checkCategories)

  good_result_GP <- nrow(data4School[(data4School$School_Name == "GP") & (data4School$combined_grades > 40.00),])
  good_result_MS <- nrow(data4School[(data4School$School_Name == "MS") & (data4School$combined_grades > 40.00),])

  GP <- nrow(data4School[data4School$School_Name == "GP",])
  MS <- nrow(data4School[data4School$School_Name == "MS",])

  percentageOfPerformers_GP <- (good_result_GP/GP)*100
  percentageOfPerformers_MS <- (good_result_MS/MS)*100

  good_result <- c(percentageOfPerformers_GP,percentageOfPerformers_MS)

  barplot(good_result,
          main = "School vs Percentage of Performers",
          xlab = "School",
          ylab = "Percentage of Performers(%)",
          names.arg = c("GP", "MS"),
          col = "brown",
          horiz = FALSE)
}

schoolAnalysis()
```

Diagram 37: The Analysis to Find the Relationship Between School and Students' Performance

For ease of analyzing, a new data frame "data4School" was made, it consists of data from "School_Name" and "combined_grades".

Then, the categories of schools are found.

We then found the number of students that come from the school "GP", and also performed well in their studies. The number of students that come from the school "MS", and also performed well in their studies is also found.

Then, the percentage of performers that come from the different schools were calculated by dividing by the total numbers of students from the respective schools.

It was placed in a bar graph to see the relationship between schools and performers.

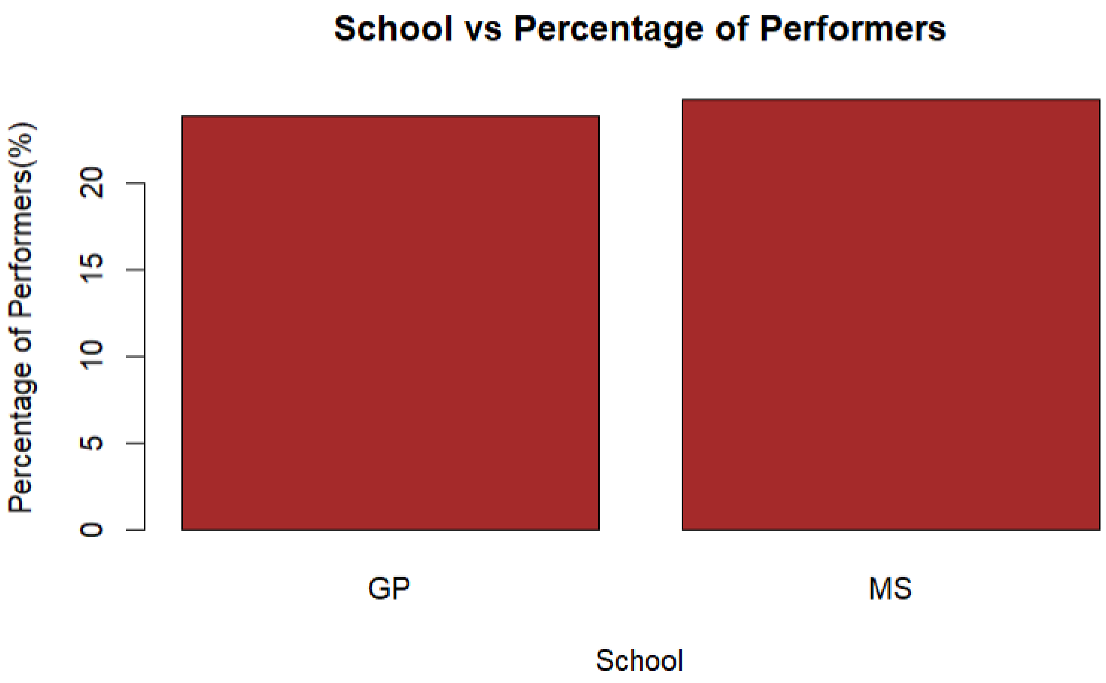

Diagram 38: Bar Graph from the Analysis "School vs Percentage of Performers"

It is shown that students from MS school have a higher chance in performing in their studies.

It is said that there are many factors that could affect a student's performance in their studies. While an in-depth analysis on the schools is much needed, we can assume that MS is better than GP regarding the factors on how it can affect a student's performance. (Shetty N., 2020)

7.2 Failures in Classes vs Marks

```
failuresAnalysis = function(){

  data4Fails <- select(tidied_data, "Amount_of_Fails_in_Class","combined_grades")
  data4Fails

  checkCategories <- factor(data4Fails$Amount_of_Fails_in_Class)
  levels(checkCategories)

  good_result_0 <- nrow(data4Fails[(data4Fails$Amount_of_Fails_in_Class == 0) & (data4Fails$combined_grades > 40.00),])
  good_result_1 <- nrow(data4Fails[(data4Fails$Amount_of_Fails_in_Class == 1) & (data4Fails$combined_grades > 40.00),])
  good_result_2 <- nrow(data4Fails[(data4Fails$Amount_of_Fails_in_Class == 2) & (data4Fails$combined_grades > 40.00),])
  good_result_3 <- nrow(data4Fails[(data4Fails$Amount_of_Fails_in_Class == 3) & (data4Fails$combined_grades > 40.00),])

  F0 <- nrow(data4Fails[data4Fails$Amount_of_Fails_in_Class == 0,])
  F1 <- nrow(data4Fails[data4Fails$Amount_of_Fails_in_Class == 1,])
  F2 <- nrow(data4Fails[data4Fails$Amount_of_Fails_in_Class == 2,])
  F3 <- nrow(data4Fails[data4Fails$Amount_of_Fails_in_Class == 3,])

  percentageOfPerformers_0 <- (good_result_0/F0)*100
  percentageOfPerformers_1 <- (good_result_1/F1)*100
  percentageOfPerformers_2 <- (good_result_2/F2)*100
  percentageOfPerformers_3 <- (good_result_3/F3)*100

  good_result <- c(percentageOfPerformers_0,percentageOfPerformers_1,percentageOfPerformers_2,percentageOfPerformers_3)

  barplot(good_result,
          main = "Failures in Class vs Percentage of Performers",
          xlab = "Failures in Class",
          ylab = "Percentage of Performers(%)",
          names.arg = c("0","1","2","3"),
          col = "grey",
          horiz = FALSE)
}

failuresAnalysis()
```

Diagram 39: Analysis for Failures in Class and Marks

First, a data frame “data4Fails” is made and has “Amount_of_Fails_in_Class” and “combined_grades”.

Then, the numbers of performers with respective amount of fails in class are calculated. They are then divided by the total number of students with respective amount of fails in class. The rates are used to see the relationship of amount of fails in class and performers.

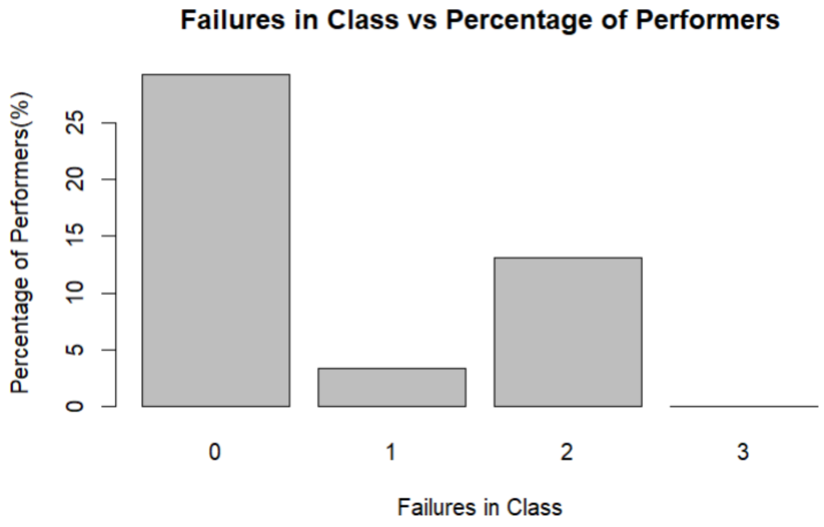

Diagram 40: The Graph with Failures in Class and Percentage of Performers

As shown in the graph, most of the performers have never once failed in their class.

Class tests are used to determine whether students have learned what they are expected to learn or degree to which students have learned the material. It is used to measure learning progress and achievement. As explained, students that have not failed in classes means that they have learned and understood the materials which were provided. Since they have understood clearly, it is a given that they would perform well in their total grades. (Kelly M., 2019)

7.3 Nursery School vs Marks

```
nurseryAnalysis = function(){

  data4Nursery <- select(tidied_data, "Attendance_in_Nursery_School","combined_grades")
  data4Nursery

  good_result_Y <- nrow(data4Nursery[(data4Nursery$Attendance_in_Nursery_School == "yes") & (data4Nursery$combined_grades > 40.00),])
  good_result_N <- nrow(data4Nursery[(data4Nursery$Attendance_in_Nursery_School == "no") & (data4Nursery$combined_grades > 40.00),])

  Y <- nrow(data4Nursery[data4Nursery$Attendance_in_Nursery_School == "yes",])
  N <- nrow(data4Nursery[data4Nursery$Attendance_in_Nursery_School == "no",])

  percentageOfPerformers_Y <- (good_result_Y/Y)*100
  percentageOfPerformers_N <- (good_result_N/N)*100

  good_result <- c(percentageOfPerformers_Y,percentageOfPerformers_N)

  barplot(good_result,
          main = "Nursery School vs Percentage of Performers",
          xlab = "Attended Nursery School?",
          ylab = "Percentage of Performers(%)",
          names.arg = c("Yes","No"),
          col = "black",
          horiz = FALSE)
}

nurseryAnalysis()
```

Diagram 41: Analysis for Attendance in Nursery School and Marks

First, a data frame “data4Nursery” that has “Attendance_in_Nursery_School” and “combined_grade” is made.

The number of performers that has been to a nursery school and not have are calculated. It is then divided by the total number of students with its respective nursery school status to find the percentage.

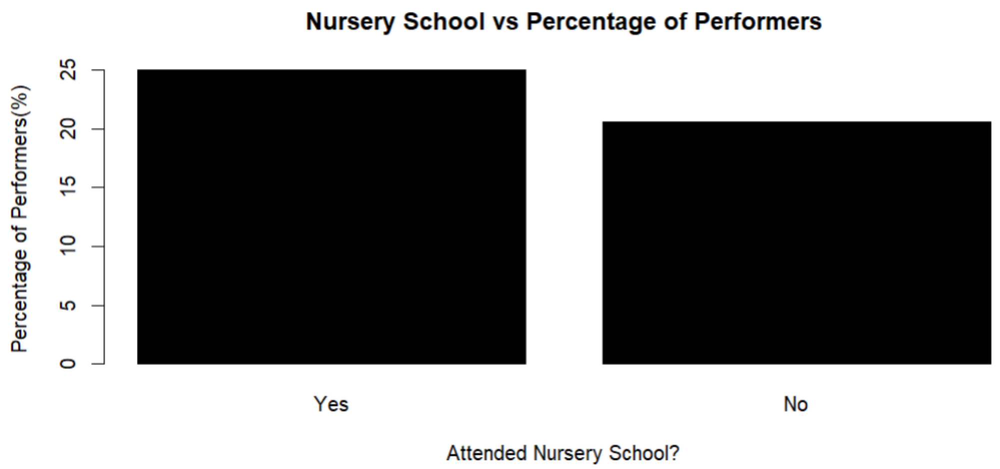

Diagram 42: The Graph with Nursery School and Percentage of Performers

As shown in the graph, most of the performers have previously attended nursery school.

Nursery school is the first point of experience for a child in a learning environment. It helps students prepare for the curriculum at structured schools, helping them to not be overwhelmed. Not only that, it helps with students' cognitive development. It teaches the students to practise and interact in big groups. It also teaches students how to manage time, which we now know that how students manage their time will impact their performance in studies. (Peter S. P. S., 2019)

7.4 Desire for Higher Education vs Marks

```
higherEduAnalysis = function(){

  data4Higher <- select(tidied_data, "Intention_to_Further_Higher_Education","combined_grades")
  data4Higher

  good_result_Y <- nrow(data4Higher[(data4Higher$Intention_to_Further_Higher_Education == "yes") & (data4Higher$combined_grades > 40.00),])
  good_result_N <- nrow(data4Higher[(data4Higher$Intention_to_Further_Higher_Education == "no") & (data4Higher$combined_grades > 40.00),])

  Y <- nrow(data4Higher[data4Higher$Intention_to_Further_Higher_Education == "yes",])
  N <- nrow(data4Higher[data4Higher$Intention_to_Further_Higher_Education == "no",])

  percentageOfPerformers_Y <- (good_result_Y/Y)*100
  percentageOfPerformers_N <- (good_result_N/N)*100

  good_result <- c(percentageOfPerformers_Y,percentageOfPerformers_N)

  barplot(good_result,
          main = "Intention to Further Higher Education vs Percentage of Performers",
          xlab = "Intention to Further Higher Education",
          ylab = "Percentage of Performers(%)",
          names.arg = c("Yes","No"),
          col = "lightyellow",
          horiz = FALSE)
}

higherEduAnalysis()
```

Diagram 43: Analysis Intention to Further Higher Education and Marks

First, a data frame “data4Higher” that has “Intention_to_Further_Higher_Education” and “combined_grade” is made.

The number of performers that have an intention to further their studies and not have are calculated. It is then divided by the total number of students with its respective intention status to find the percentage.

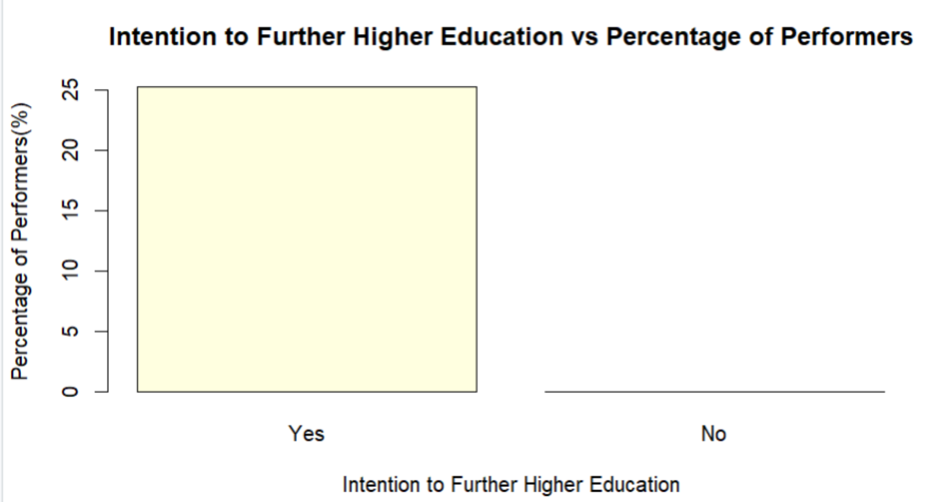

Diagram 44: The Graph with Intention to Further Higher Education and Percentage of Performers

As shown in the graph, all of the performers have an intention to further into higher education.

It is because for every higher education institution, there are entry requirements. Students with intention to further their studies will study harder and perform better to ensure a slot in those institutions.

8.0 Extra Features

8.1 Dplyr

Diagram 45: The logo of the package "dplyr"

Dplyr is a package that provides functions with a consistent set of verbs that help us solve the most common data manipulation challenges.

The functions used in this research is select() and summarise(). Select() allows us to pick variables based on their names. Meanwhile, summarise() allows us to reduce multiple values down to a single summary.

```
#Installing Packages
install.packages("dplyr")

#Loading Packages
library(dplyr)
```

Diagram 46: The Installation and Loading of the package "dplyr"

9.0 Conclusion

As a conclusion, we found that there are many factors that can affect the performance of a student in their studies, which are their family size, parent's marriage status, parent's education level, father's and mother's job, extra educational support (school supplement), extra paid classes, romantic relationship, family relationship, weekly study time, time spent going out with friends, time spent consuming alcohol, school, failures in class, nursery school and desire for higher education.

It is safe to say that a student that has a small family size, parents that are still together, parents that have higher education level, father as a teacher, mother working in health, does not take extra educational support (school supplement), has extra paid classes within the course, has no romantic relationship, has poorer quality of family relationship, spends their free time studying, does not spend a lot of time studying, consumes alcohol frequently, studies in MS, has no failures in class, has been to nursery school and has intention to further their studies will perform well in their studies.

Thus, to perform better in studies, students need to pay attention to the factors mentioned and carefully adjust them.

10.0 References

1. abaKus {n.d.) *What are the Benefits of Extra Classes?* [Online] Available at: https://abakusenrichment.com/what-are-the-benefits-of-extra-classes/#:~:text=Extra%20classes%20in%20math%20and,the%20benefits%20of%20extra%20classes. [Accessed on 17 May 2022]

2. Clearinghouse (2020) *Parents' Educational Levels Influence on Child Educational Outcomes: Rapid Literature Review* [Online] Available at: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://militaryfamilies.psu.edu/wp-content/uploads/2020/01/Parents-Educational-Levels-Influence-on-Child-Educational-Outcomes.20Jan06.final_.pdf [Accessed on 17 May 2022]

3. Denny K. (2011) *Do Teachers' Children Do Better at School?* [Online] Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.822.8784&rep=rep1&type=pdf [Accessed on 17 May 2022]

4. Heinrich C.J. (2014) *Parents' Employment and Children's Wellbeing* [Online] Available at: https://files.eric.ed.gov/fulltext/EJ1029033.pdf [Accessed on 17 May 2022]

5. Kelly M. (2019) *School Testing Assesses Knowledge Gains and Gaps* [Online] Available at: https://www.thoughtco.com/the-purpose-of-tests-7688#:~:text=They%20are%20used%20to%20determine,the%20effectiveness%20of%20educational%20programs. [Accessed on 17 May 2022]

6. Marripedia (n.d.) *Effects of Divorce on Children's Education* [Online] Available at: https://marripedia.org/effects_of_divorce_on_children_s_education [Accessed on 17 May 2022]

7. Muychrea L., Vatanak S., Sopheavatey T. (n.d.) *Cambodian University Students' Perspectives on the Effects of Having a Romantic Relationship on Academic Performance* [Online] Available at: https://uc.edu.kh/userfiles/image/2018/Cambodian_University_Students%E2%80%99_Perspectives.pdf [Accessed on 17 May 2022]

8. Onzaberigu J. N. (2017) *The Effects of Family Size On the Investment of Child Education.* [Online] Available at: https://www.researchgate.net/publication/325972109_The_Effects_Of_Family_Size_On_The_Investment_Of_Child_Education_Case_Study_At_Atonsu-Buokro_Kumasi [Accessed on 17 May 2022]

9. Peter S. P. S. (2019) *Nursery Introduces Time Management* [Online] Available at: https://stpetersprep.co.uk/why-nursery-important/#:~:text=Nursery%20helps%20children%20learn%20time,will%20need%20it%20much%20more! [Accessed on 17 May 2022]

10. Shetty N. (2020) *Importance of Choosing the Right School for your Children* [Online] Available at: https://www.zedua.com/blog/importance-of-choosing-the-right-school-for-your-children/ [Accessed on 17 May 2022]