

**INDIVIDUAL ASSIGNMENT**

**TECHNOLOGY PARK MALAYSIA**

**CT127-3-2-PFDA**

**PROGRAMMING FOR DATA ANALYSIS**

**APD2F2106CS(IS)**

**HAND OUT DATE: 5 JULY 2021**

**HAND IN DATE: 23 AUGUST 2021**

**WEIGHTAGE: 50%**

**STUDENT NAME: SIM YOKE SHIN**

**STUDENT ID: TP059851**

# **Introduction**

In this assignment, researcher has been required to explore and analysis a set of data that store in student.csv. This csv file has stored 33 columns and 922 rows of information about the family background, academic, and daily routine of the students which studying at Gabriel Pareira and Mourinho da Silveira. Besides that, the csv file also stores the student’s final grades of math for 3year. It can enable researcher to analyse the other attributes by comparing with the marks of final grades. The researcher has been required to use several techniques like data exploration, manipulation, transformation, and visualization to analyse the data into a meaningful representation.

# **Assumption**

To start with cleaning the data, researcher found that the naming of the attribute is unmeaningful. So, as the table 1 that show below, researcher has renamed the whole attribute name to become more readable. Researcher has assumed the failure is the times of grade repetition

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Original Name | New Attribute Name | No | Original Name | New Attribute Name |
|  | school | stu\_school |  | paid | ext\_paid |
|  | sex | stu\_sex |  | activities | ext\_cur\_activity |
|  | age | stu\_age |  | nursery | attend\_nursery\_sch |
|  | address | stu\_address |  | higher | higher\_edu |
|  | famsize | fam\_size |  | internet | home\_internet |
|  | Pstatus | parent\_living\_status |  | romantic | romantic |
|  | Medu | Mo\_edu |  | famrel | fam\_rel\_quality |
|  | Fedu | Fa\_edu |  | freetime | free\_time |
|  | Mjob | Mo\_job |  | goout | out\_wth\_friend |
|  | Fjob | Fa\_job |  | Dalc | workday\_alc |
|  | reason | reason\_chose |  | Walc | weekend\_alc |
|  | guardian | guardian |  | health | stu\_health |
|  | traveltime | time\_HometoSch |  | absences | num\_absences |
|  | studytime | study\_time |  | G1 | Grade\_1st |
|  | failures | num\_failure |  | G2 | Grade\_2nd |
|  | schoolsup | ext\_edu\_sup |  | G3 | Final\_Grade |
|  | famsup | fam\_edu\_sup |  |  |  |

# **Data import/Cleaning/pre-processing/ transformation**

Data import

Graphical user interface, text, application

Description automatically generated



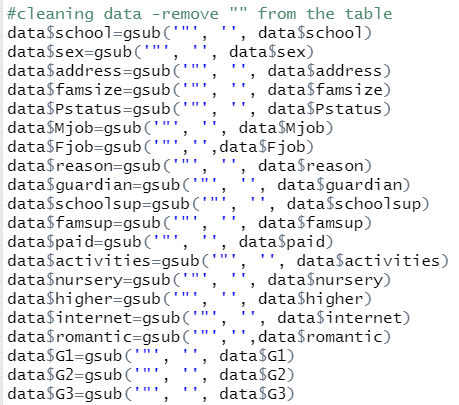
Data exploration(找summary)

Text

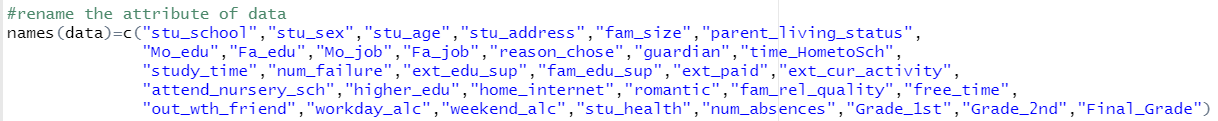
Description automatically generated

Data cleaning

After view the report, there is 杂乱, so researcher remove the “”.



Data pre\_processing



Data transformation

Text

Description automatically generated

A picture containing text, newspaper

Description automatically generated

PFDA

i think i can say some general things la (from my tutor at least)

question (example, how to improve student grade)

introduce the question : why you wanna study this, what attrs you wan use?

analysis 1 : relationship between A and B

why A ? why B ?

show code, results graph etc, then explain on it

discuss the result, if can include a research to support your point

repeat for analysis 23456 etc

then provide a conclusion on how these 123456 analyses can help in answering your main question

Data Analysis

Question 1 Does drinking habit will have bad affect to student?

In general know, drinking habit is a bad habit that will bring a negative effect to people. Most people believe people who drink more will hurt your brain. So, researcher is interest in does the student which has a drinking habit will bring a bad effect to student. For making the data become easier to use, researcher has added a now column which call Total\_alc that has combined the workday\_alc and weekend\_alc. Also, researcher has added a new column named AVG\_grade which is the mean of Grade\_1st, Grade\_2nd, and Grade\_3rd.

Analysis 1-1: Comparison of workday\_alc and weekend\_alc

The table has shown the summary of the data, it states that the mean of weeked alc is greater than workday\_alc.

Text

Description automatically generatedChart, pie chart

Description automatically generatedChart, bar chart, histogram

Description automatically generated

In the graph, it has shown that there have 1/3 of student among the whole 922 student which is drink frequently(total drink frequency more than 5).

Analysis1.1 find the relationship between drink frequency and health

While talking about the drink, most people think that it will make our health become worse. So, researcher has made 2 graphs about the relationship of drink frequency and health in this analysis. The attribute that has used in this analysis is stu\_health,workday\_alc,weekend\_alc.

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

As this graph state, it has use geom\_jitter to show the relaithonship of student’s health and drink frequency. Refer to the graph, researcher has found that the student who drink alchoho in weekend frequenly will healthy then those student who drink frequenly in workday. There has lot of student which drink in workday is having a bad health than the student who drink in weekend. Beside that, the graph shown that students who drink on workday are drink lesser and more self-disciplined.

Chart, bar chart

Description automatically generated Chart, bar chart

Description automatically generated

After compare with geom\_jitter, researcher has displayed the result again with using geom\_col to get the result which is more clearly. As this graph state, the is more student which like to drink in weekend. Besides that, it has shown among the student who drink at weekend and workday, there is most student which drink in weekend can get the highest healthy rank.

\*\*So, researcher can make a small conclusion that if student want to drink, it is better to drink in weekend.

Analysis1.2 find the relationship between drink frequency and score

In this analysis, researcher want to discuss about the relationship of drink’s frequency and score. First The attribute that researcher has use in this analysis is AVG\_Score and Total\_alc.

Chart, scatter chart

Description automatically generated

As the graph has shown in upside, it has use the geom\_count to show the data. Researcher has found that there has a pattern that student score decreases with increasing frequency of drinking.

Analysis1.3 Find the relationship between high frequency of drink alcohol with absence and failure

Drink more become absence frequenly, done

Drink more ppl failure times

For doing this analyse, researcher has use summary() to check the range of the num\_absences. As this figure shown, the median of num\_absences is 4 but the maximum value of the dataset is 75 and the 1st quarter value is same with the minimum. For drawing a clearly graph, researcher has reset the range of selection dataset to 1 until 30.



Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

As this graph show, we can know that most student who drink in weekend absence more than student who drink in workday. And compare between both graph, researcher has found that the higher amount of them is gather in 2 and 4 days. So, researcher can say most student will not absence more than 5 days.

\*\* student who drink in weekend will absence more time.

After that, researcher is interested in the failure of them because absence frequently will make the student hard to keep up with the teaching’s process.

Chart

Description automatically generated

Analysis1.4 Is student which go out frequently will drink more? Chart, bar chart, histogram

Description automatically generated

Question 2

How the study environment affects student marks?

Analysis 2.1 Compare the two schools, which school teaches students better?

Analysis 2.2 Compare with 2 student address, which place’s student smatter?

Analysis 2.2 Find the relationship between having home internet with study time

Is doesn’t have home internet will make student study more?

Analysis 2.3Find the relationship between Father or mother job is teacher and the student marks

Question 3

How to score higher mark?

Analysis 3.1What is the common feature of the higher mark’s student?

(eg. Read more, go out less, have tuition,)

Analysis 3.2 Find the relationship between want to go for higher education and marks.

As people say, having a goal can make people motivate to fight. Therefore, is wanting to go to a higher education will affect students' scores higher than those without goals?

Analysis 3.3 Find the relationship between romantic and student marks?

Analysis 3.4 Find the relationship between frequent social activities and marks?

Does frequent social activities for example go out with friend and having extra curriculum activity make students ignore learning?

Analysis 3.5 Find the relationship between absences and marks

Is student absences frequency will affect their marks?

Question 4

How does parent become a factor of affect student mark?

Analysis 4.1 Is a student who born in a high edu family will make he/she become smarter?

Analysis 4.2 Find the relationship between family relationship quality and parent living status with marks

Is family relationship quality will affect student mark?

Analysis 4.3 Find the relationship of family education support or extra education support with family relationship quality

Is parent which give their children education support or extra education support mean that the relation of family is good?