FACULTY OF ENGINEERING

SCHOOL OF COMPUTING

SEMESTER 1/20212022

SECP1513 – TECHNOLOGY INFORMATION SYSTEM

SECTION 03

ASSIGNMENT 4

DATA ANALYTICS ON STUDENT'S PERFORMANCE IN MATHEMATIC BY USING MICROSOFT POWER BI

Lecturer: PM Dr Mohd Shahizan Othman

Name	Matric ID
LOO ZHI YUAN	A21EC0197
NG ZI XING	A21EC0213

TABLE OF CONTENT

GROUP DETAILS

Title of the report: DATA ANALYTICS ON STUDENT'S PERFORMANCE IN MATHEMATIC BY USING MICROSOFT POWER BI

Group leader: NG ZI XING

Group member: LOO ZHI YUAN

Task assigned: NG ZI XING - Data analytics interpretation and Data Board for the chosen data

LOO ZHI YUAN – Introduction part(Trends of Data Analytics in Different Sectors and Industrial Talk7)

Trends of Data Analytics in Different Sectors

Introduction

Emerging Trends: Data Analytics

In this modern-day age, data play a significant role in helping people make decisions in daily life. Based on Investopedia, data analytics is the science of analysing raw data to make conclusions about that information. Many of the techniques and processes of data analytics have been automated into mechanical methods and algorithms that work over raw data for human consumption. Like data, data analytics also act as a crucial role in human life by helping to optimise and improve the processes in different sectors such as education, healthcare, manufacturing, communication and media, and banking and securities.

In the education sector, data analytics can assist the university and schools analyse and tracking the students' data in the university or school, such as students' performance and interest in certain subjects or courses based on the data collected. Furthermore, data analytics can also assist the school's management in making decisions. Data analytics can make decisions that consider students, staff, and the school itself based on the resources provided. Hence, data analytics can allow the education sector to become complete and more systematic with a better understanding of the student, staff, and teachers.

Data analytics provides a way to collect health data from the public and use it with further analysis in the healthcare sector. In this COVID-19 pandemic that gets into more severe conditions day by day, collecting public health data also contributes to the quick responses to the new or mutated virus such as Omicron. Therefore, the trend of data analytics is significantly impacting the healthcare sector by improving efficiency and effectiveness.

In the manufacturing sector, data analytics helps the manufacturer discover something new and find better patterns that improve the manufacturing processes. The data analytics analyse the manufacturing processes with the number of processed and made products. Hence, data analytics enhance the manufacturing section's productivity in the best way they read.

Data analytics helps some companies know the customers and users better in the communication and media sector. The companies can plan better marketing ideas and activities based on the data's needs. Data analytics can also apply to real-time reports and push notifications of the latest trends, events, and more. Therefore, the experience of customers on the product will be improving.

In the banking and securities sector, data analytics can act as a "CCTV" on the stock market to detect and monitor illegal and abnormal trading activities. Data analytics can also make predictions on the market trends and growth to decrease the risk taken. Therefore, data analytics can provide safety and convenience in banking and securities.

Reflection

From our point of view, we truly believe that data analytics is becoming the hottest emerging trend in this 21st century. The values of data analytics are reflected in its contributions in different sectors such as education, healthcare, manufacturing, communication and media, and banking and securities. Data analytics enhances productivity, efficiency, and effectiveness in most mentioned sectors. This shows that data analytics can make human life easier with its advantages. Data analytics can also provide safety and convenience to the public and market, improving the securities in the banking sector. In short, data analytics trends are closely related to the development of technology and Industrial Revolution 4.0 (IR4.0). We can firmly believe that the emerging trends of data analytics will lead humans to higher technological development and enter the completed IR4.0 era.

Reference

1. https://www.investopedia.com/terms/d/data-analytics.asp

Industrial Talk 7: Introduction to Data Visualization (iCEP)

Introduction

Industrial Talk 7: Introduction to Data Visualization by Mr Isma Redha from iCEP

On 23 December 2021, we attended the informative industrial talk by Mr Isma Redha from iCEP about data visualisation and Microsoft Power BI. iCEP is formerly known as WGC 2021 Organising Committee, which aspires to be the professional conference, exhibition and event organiser. Before starting using Microsoft Power Pi, we should master the basics of data visualisation that taught us to understand types of data that can be visualised, such as quantitative, discrete, continuous, and categorical. The speaker also provides us with a better understanding of the different data relationships and chart types that we should know before using Microsoft Power PI.

What, When and Where to use Microsoft Power BI

Microsoft Power BI is a platform that shares, analyses, visualises, and connects data that provides Microsoft's cloud-based Business Intelligence service. Microsoft Power BI is a data visualisation tool that visualises the data collected in various ways and offers a unique insight. Microsoft Power BI enables us to present and deliver our data and information simpler for better understanding.

People often use Microsoft Power BI to connect all the data, especially when the data is in huge numbers and complicated to understand. Microsoft Power BI can provide convenience to the users by using the Microsoft Powe BI platform that provides better accessibility.

As a Business Intelligence platform provided by Microsoft, Microsoft Power BI is accessible on several platforms such as desktop, IOS, Windows, and Android devices available for free. Microsoft Power BI on the desktop provides a better experience than others while collecting,

visualising, and analysing the data. Other than that, the paid versions such as Power BI Pro and Power BI Premium are also available for users or companies with further needs.

Reflection

Based on the exciting talk by Mr Isma Redha from iCEP, we had gained the information, deeper understanding and further knowledge on the data visualisation and Microsoft Power BI. As a foundation knowledge before using Microsoft Power BI, the speaker introduces data visualisation, such as the several types of data, data relationships, and chart types that we should know. Other than that, we learned the basic skills on using Microsoft Power BI that we can use to connect our data in the reporting or presenting. Furthermore, the talk also makes us realise the data visualisation trends with all the benefits it provided. It makes the complex understanding and detailed data more familiar and convenient to most people with greater accessibility, organisation, and visibility. In conclusion, we sincerely hope that the data visualisation and Microsoft Power BI can be popularised to most people nowadays to notice the charm of technology development impact our lives.

DATA VISUALIZATION AND ANALYTIC

Introduction

Data is a worth explored resource that drives the development of various aspects as the information inside the data could let us realize the fact accurately and concretely. The topic discussed below is the factors that affect students' performance in math by data visualization technique to give us about the intuition. Various arguments are filled on the internet to claim what factors will affect students' performance, and our goal is to verify them by descriptive analysis. To be concrete, the three main questions that will be tried to be answered are:

1. Do extra paid classes can enhance student's performance in math?
2. How does student's time allocate will affect student's performance?
3. Whether student's family status will affect student's performance?
4. Most people consider male has greater performance than females in math, is this true?

Let us guess first and use data visualization to explore the fact for validating our assumptions.

About Data

This data is collected from [1], and the object of investigation is secondary education of Portuguese schools. The data is collected by using school reports and questionnaires. They are 395 samples available in this data. The attribute contained is over 33 features, which is too much to integrate into one page only dashboard. Thus, we use the data that will be helpful to answer our questions. Briefly, the italics font in the article below would be the feature's name in original datasets. Besides, the data collected three grade periods in the range of 0 to 20. To be convenient in visualise it, the grade will be converted into an average of them or an excellent grade when the grade is above 15. They are called *gred* and *gredA* in the following discussion.

For the question1, we use the attribute *paid* to study it. The data classified into “yes” and “no” instead of a specific amount of paid. To visualise it, clustered horizontal bar chart is used to show the relationship between mean, standard deviation and median against to paid.

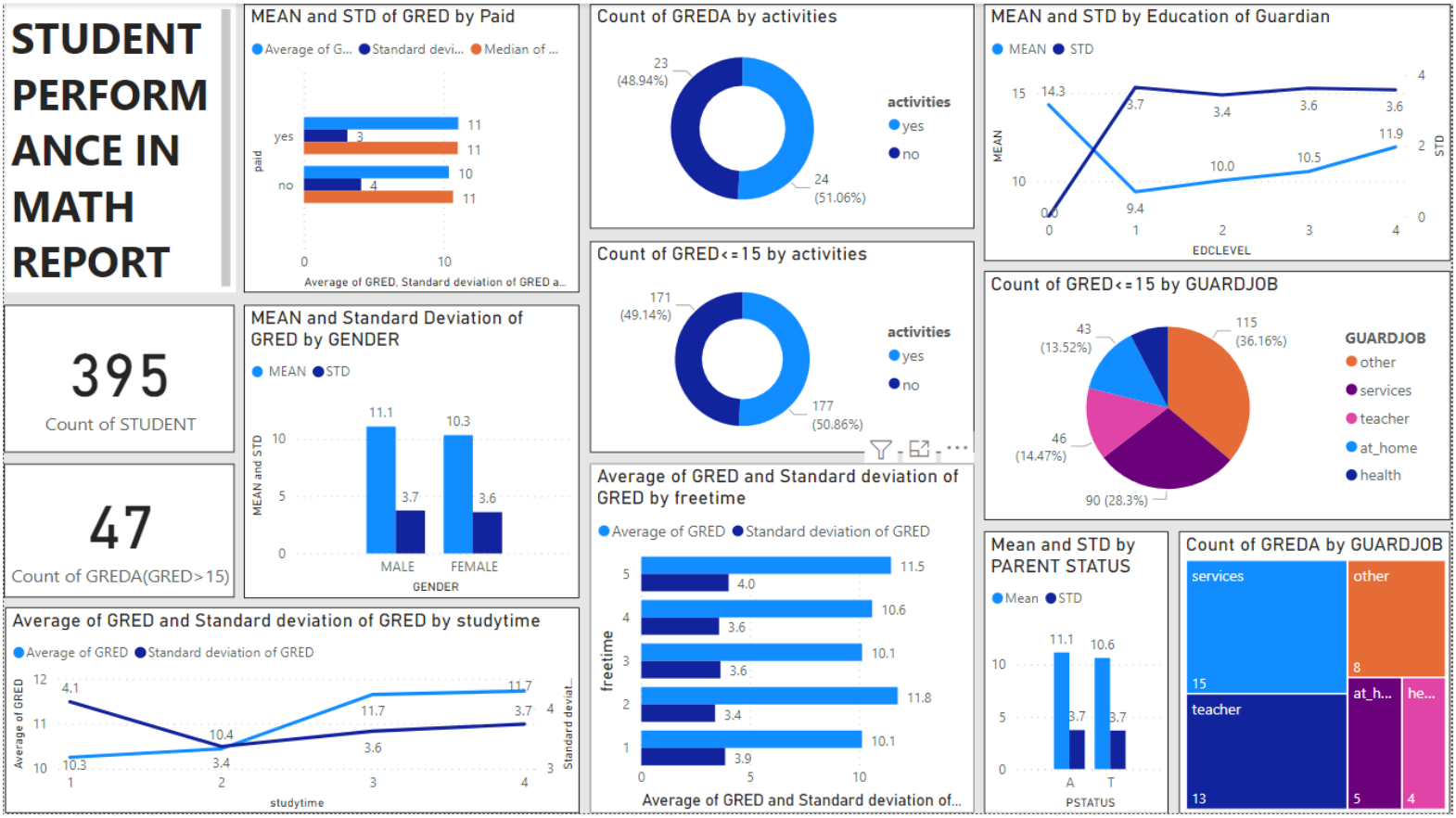
For question2, three features which are *activities*, *freetime* and *studytime*. The *activities* are categorised into “yes” and “no”. We will use a donut chart to investigate the percentage of students who are qualified *gredA* or below against activities. Next, the *freetime* is divided into 5 levels. The highest one implicates the students who have the freest time after school. The horizontal clustered bar chart will be graphed to show the relationship between the standard deviation(std) and the means of *gred* of each level with the level of free time. After that, the *studytime* is also be divided into 4 levels according to the standard of weekly study time: <2 hours -1, 2 to 5 hours -2, 5-10 hours -3, >10 hours -4. We graph the mean and standard deviation of *gred* of each level against *studytime* into a line chart. The three features in this paragraph can well represent how student’s after school time is allocated. The *freetime* show student has how much free time and the activities and *studytime* show how student use that time.

For question3, six features are used to answer this question, including *Pstatus*(Parent Status), guardian, *Medu*(Mother’s education), *Fedu*(Father’s education), *Mjob*(Mother’s job) and *Fjob*(Father’s job). The *Pstatus* shows the parent’s cohabitation status, “A” stands for the parent is apart and “T” implies that the parent is living together. The relationship of mean and standard deviation of each category with *Pstatus* are represented clustered column chart. After that, the guardian has consisted of “father”, “mother” and “other” but this attribute will not be shown directly in data dashboard. This feature is used to simplify *Medu*, *Fedu*, *Mjob* and *Fjob* as they are too many features and lead to the difficulty of visualising it. To be *more straightforward*, we convert them into the *Guardedu*(Guardian’s education level) and *GuardJob*(Guardian’s job), which stand for if the guardian is mother, we just directly replace the feature of education level and job with mother’s education level and job vice versa. The most obvious disadvantage to doing this is we have to discard the information about “other“guardians because there is no guardian’s education and job available in the data, but we still have 363 samples that are still in the acceptable range. To continue, the *guardedu* is categorised into five levels by the norm of 0- none, 1- primary education (4th grade), 2- 5 to 9th grade, 3- secondary education, 4- higher education. The *guardedu* attribute with the mean and standard deviation of *gred* will also be plotted in a line

chart. Then, for the *guardjob*, which is a nominal data including "teacher", "health" care related, civil "services" (e.g. administrative or police), "at home" and "other". Due to the numerous categories, we draw it in *gredA* against *guardjob* in Treemap by the number of students who got *gredA* in each category.

For question4, the sex (gender of the student) will be utilised, which "F" stand for female and "M" stand for male. A clustered bar chart plotted the mean and standard deviation of *gred* against gender.

ONLINE DASHBOARD REPORT



Interpretation of the output

1. **Will health affect the performance to get excellent results?**

Based on the graph "MEAN and STD of GRED by Paid", we can observe that the mean is similar with the median, which implies the distribution of paid and not paid to follow the normal distribution. Thus, we can use some hypothesis testing techniques to verify our result. For the standard deviation, the student with extra pay to study (3.119) is lower than the student with not paid (4.105), which implies student with paid has more stable performance in the math exam. Besides, the student's mean with paid (11.039) is also higher than the student with not paid (10.375). But due to the relative value and different amounts of samples, we can use a two-sample t-test to verify this result because they follow a normal distribution.

t-Test: Two-Sample Assuming Unequal Variances

	yes	*no*
Mean	11.03867	10.37539
Variance	9.732447	16.85216
Observations	181	214
Hypothesized Mean Difference	0	
df	389	
t Stat	1.822053	
P(T<=t) one-tail	0.034607	
t Critical one-tail	1.64878	
P(T<=t) two-tail	0.069215	
t Critical two-tail	1.966081	

From the table, we can see the t Stat lower than t Critical one-tail. Thus, we could claim that student with paid has higher mean performance than those not paid indeed. In conclusion, *paid* will increase the student's performance in math with higher marks and more stability.

2. **How does student's time allocate will affect student's performance?**

The bar chart "average of *gred* and standard deviation of *Gred* by freetime" shows that the student's free time after school with level 2 has the highest mean of *gred* (11.78) and the lowest standard deviation (3.40). The level 3 and 1 have the lowest mean of *gred* (10.1), but level 1 with higher standard deviation (3.86) than level 3 (3.65). Then, the highest standard deviation (4.02) is associated with the student having the free time of level5, but the mean is second highest. From the information above, we can see that the data trend is not clear. But what we can conclude is that student has more free time does not mean their performance in math will grow accordingly. The students who have the freest time maybe could have more time to do revision and use that time to play with a friend also is possible, that is the reason why the free time has the largest standard deviation because some students

can manage themselves well instead of the mindset of “I have so much time so I could waste time” but some cannot. So, the middle level of free time could lead to students having higher performance, and the data also verify that level 2 has the highest mean and lowest standard deviation. Besides, the student in level1 has the lowest mean and second high variance, which implies the too low free time is not helpful in student’s performance. In short, a suitable amount of student’s free time is the most beneficial to raise student performance.

After that, from the graph “Average *gred* and standard deviation of *Gred* by *studytime*”, we can see clearly the trend of means is growing according to the increasing of *studytime* then go stable. For the standard deviation, the overall trend is decreasing, but a little bit bounce up after level2 but not much. From the graph, we can conclude that the greater the time student spend in study, his/ her performance in math also will growing and their performance also is more stable due to the standard deviation lower implies that the greater the study time, the higher the probability the student can get a high mark. Conversely, students who have *studytime* of level one (weekly study time lower than 2 hours) have the lowest mean (10.25) and highest standard deviation (4.12), implicating their performance are both low and unstable.

Lastly, there are two pictures which are “Count of *GREDA* by activities” and “Count of *GRED*<=15 by activities” is related to activities. We discovered from these two graphs that the percentage of students who have or have no extracurricular activities is almost the same (51.06 % of student with activities in *GREDA*, 50.86% in *GRED*<=15). Hence, we could conclude that the extracurricular has a lesser relation with students' performance in math overall but not particularly as we don’t have a clue to prove that.

In short, the answer to this question is that students with a moderate or suitable amount of free time and more study time have a positive effect on students' performance in math. But the activities have not much influenced in it.

3. Whether student's family status will affect student's performance?

t-Test: Two-Sample Assuming Equal Variances

	"A"	*"T"*
Mean	11.1463415	10.6252354
Variance	14.2447154	13.61112089
Observations	41	354
Pooled Variance	13.6756089	
Hypothesized Mean	0	
df	393	
t Stat	0.85417685	
P(T<=t) one-tail	0.19676368	
t Critical one-tail	1.64874011	
P(T<=t) two-tail	0.39352737	
t Critical two-tail	1.96601861	

Firstly, from the graph "Mean and STD by Parent Status", we can see that their standard deviation is similar and the mean of categories "A" (apart) (11.146) is higher than "T" (together) (10.625). So, is that mean a student with a parent who apart has higher performance? We can do a two-sample hypothesis t-test to verify this claim by excel.

From the result, we can see t Stat < t Critical two-tail. Thus, their performance has not much different the reason is the sample amount of "T" is significantly larger than "A" which is 354>41. Thus, the result tells us the parent status does not influence students' math performance from the dataset we collected.

Secondly, by the graph" MEAN and STD by Education of Guardian", we can observe that the mean trend decreases sharply at first and grows stably after level 1. The trend is sharply increasing for the variance and becoming stable after level 1. We observe the movement after level 1 first, as common sense expected. The higher the education level of student's guardian, the higher the student's performance because the guardian has related knowledge and experience that can teach their children and give a better environment of studying to study their children. Besides, that is one outlier inside our data. For level 0, that is only one sample with *GUARDEDU*=1, it has zero standard deviation, and this sample has a high mark. Due to the lack of data about *GUARDEDU*=1, we ignore it. In conclusion, the greater the parent's education level, the greater the student's performance in math.

Lastly, there is two pictures regarding *GuardJob*, which is "Count of GRED<=15 by *GUARDJOB*" and "Count of *GREDA* by *GUARDJOB*". We can observe that the area covered in both graphs is slightly different, implying that in either GREDA or GRED<=15, the percentage of each job is influenced by the GRED. To verify that, we could do a chi-square test:

	GREDA		GRED<=15	
Category	real ratio	Frequency	Real ratio	Expected Frequency
other	0.1778	8	0.3616	16.272
service	0.3333	15	0.283	12.735
teacher	0.2889	13	0.1447	6.5115
at_home	0.1111	5	0.1352	6.084
health	0.0889	4	0.1447	6.5115
Chi-square:	0.015684			
Significance:	0.05			

We can observe that chi-square is lower than significance (0.0157<0.05), meaning the percentage of jobs distributed in each picture is different. Thus, we can conclude that the student's performance is related to *GUARDJOB*.

In brief, family's status does affect students' math scores, especially if the higher guardian's education level, the greater student's performance. Besides, guardian's job also does relate to student's performance, but it is nominal data that lead to the difficulty of analysis. Lastly, parent's status has lesser or have no influence compared to the other two factors.

4. Most people consider male has greater performance than females in math, is this true?

Based on the graph "MEAN and Standard Deviation of *GRED* by GENDER", the fact we observed is that the average performance of males (11.07) in math is higher than the of females (10.33), which implies that most of the time the male student has a higher likelihood to get the high mark to compare to female. But the other fact is female students' performance is more stable than males because the standard deviation of females (3.61) is lower than males (3.75). This difference may be caused by the difference of mindset between gender make them have different advantages in math performance. The conclusion is gender does affect the student's performance in math.

Conclusion

In short, many factors will affect students' performance in math. For question 1, we find out that students with extra paid classes have a higher-grade mean. Besides, a suitable amount of free time to give to students and students should be assigned to study because this could provide a more stable and higher performance. Other than that, in question 3, the fact discovered is family's status also contributes to student's math grade, because *GUARDEDU* which the higher the guardian's education level, the higher the performance of the student and the *GUARDJOB* is a related factor. Lastly, the male gender truly has a higher performance mean than females, but females' version is more stable than males. The last two questions are about the fact that students cannot change it, but this does not mean students

cannot have a good result in math. The most important is, like we mention in questions 1 and 2, spending more effort and time studying is the best approach to getting high performance in math.

SOURCE OF DATA

1. https://archive.ics.uci.edu/ml/datasets/student+performance