****

Faculty of Computing

MCSD 1123 - Big Data Management

SECTION 01

Assignment 1: Report 1a

**Lecturer:**

PM DR. Mohd Shahizan Othman

Group name: TheBoys

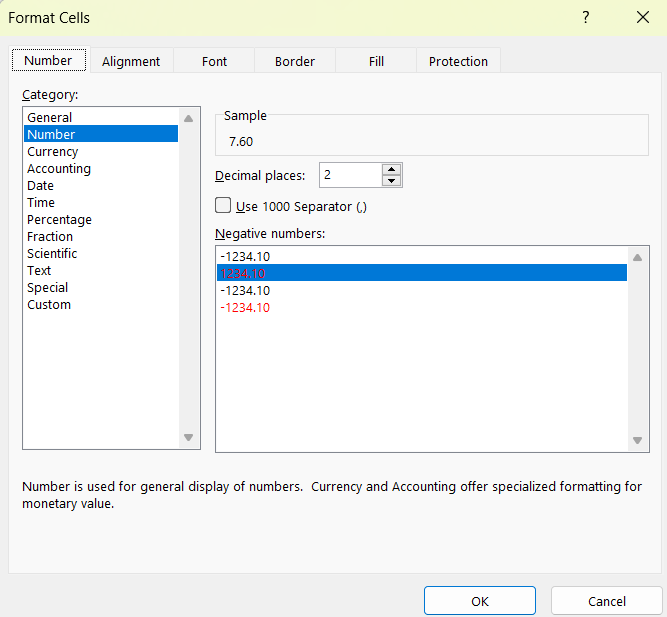
|  |  |
| --- | --- |
| Name | Matric No. |
| Pang Chern Hong | MCS231006 |
| Nian Cong | MCS231007 |
| Wu JiaMing | MCS221033 |
| Liu KaiYuan | MCS231020 |

INTRODUCTION

This dataset is about the assessment of the students' performance based on their scores academic, sports, co-curriculum and the tests. We need to process the data and analyze the results using the visualization through the dashboard to give a clear picture on the student’s performance on this assessment.

Data Processing

First, we convert the value on the data by using the format for number to make the value possess 2 decimal places.



Then we use a formula for columns of P1 to P5 to standardize the score.

For G2(the second cell of P1 column): =(B2/max(B:B))\*3.33

Then, we drag the formula from P1 to P5 to copy the formula and transform parameter to the following cells, then double click the plus sign located at the lower right of the P5 second cell to cover 5 columns with the formula.

Disussion

As We can see in this dataset, there are 5 types of scores which are Academic, Sports, Co-curriculum, Test 1 and Test 2. To give the marks to the students according to their performance, we only choose the 3 best scores they obtained and standardize these scores by dividing the score by the highest score obtained in each type, then make these scores have the equal proportion to the total marks, which is 1 over 3 each.

Only the total marks which are higher or equal to 65% counted as Pass. The grading scheme is as the figure below:

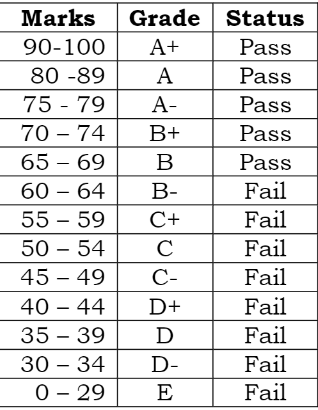


Figure 1: Grading scheme

In this case, the number of students who passed is 72611 while the number of students who failed is 38908. Table 2 will show the specific number of students for each grade.

|  |  |
| --- | --- |
| **Row Labels** | **Count of Id\_No** |
| Fail | 38908 |
| Pass | 72611 |

Table 1: Number of students passed and failed.

|  |  |
| --- | --- |
| **Grade** | **Data** |
| A+ | 8439 |
| A | 24229 |
| A- | 12503 |
| B+ | 13938 |
| B | 13502 |
| B- | 10714 |
| C+ | 7762 |
| C | 7542 |
| C- | 6006 |
| D+ | 3470 |
| D | 1677 |
| D- | 1028 |
| E | 709 |
| **Total** | **112 k** |

Table 2: Number of Students for each grade.

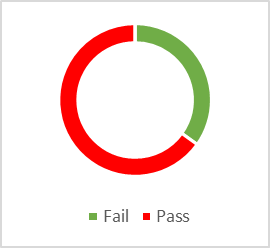


Figure 2: Pie chart of the ratio of students that failed and Passed.

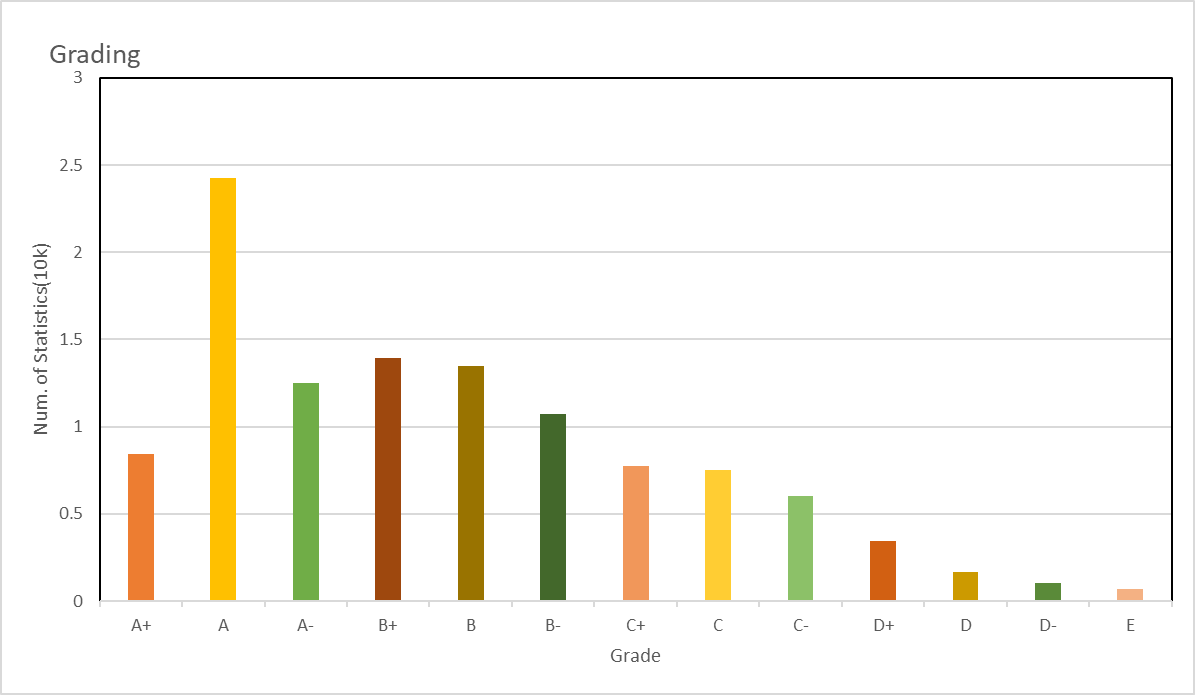


Figure 3: Statistic of the overall result

From this chart we can see that the number of students who obtained grade A is the highest, B+ is the second highest, followed by B, A-, B- and then A+, while the number of students who obtained grade E is the lowest. Nearly 2/3 of students passed while the rest failed. We can say that this assessment is reasonable for students to pass and there is a high chance to perform better than B- and among the students that passed that is more likely they get a result of at least A- and most probably they got A.

The Maximum score obtained is 98.57%, the minimum score is 14.82% and the average score is 69.87%. The average score is actually above the passing score 65%, which can be another evidence that most of the students are able to pass, although this can be due to the standardization we did to the data.

We also used conditional formatting to color the rows of the students passed pink and also give their total marks cells a green color by applying the rules of greater than or equal to 65% by using the format only the cells that contain, and by using a formula to determine which cells to format: =$R2=”Pass” to achieve the effect we want.

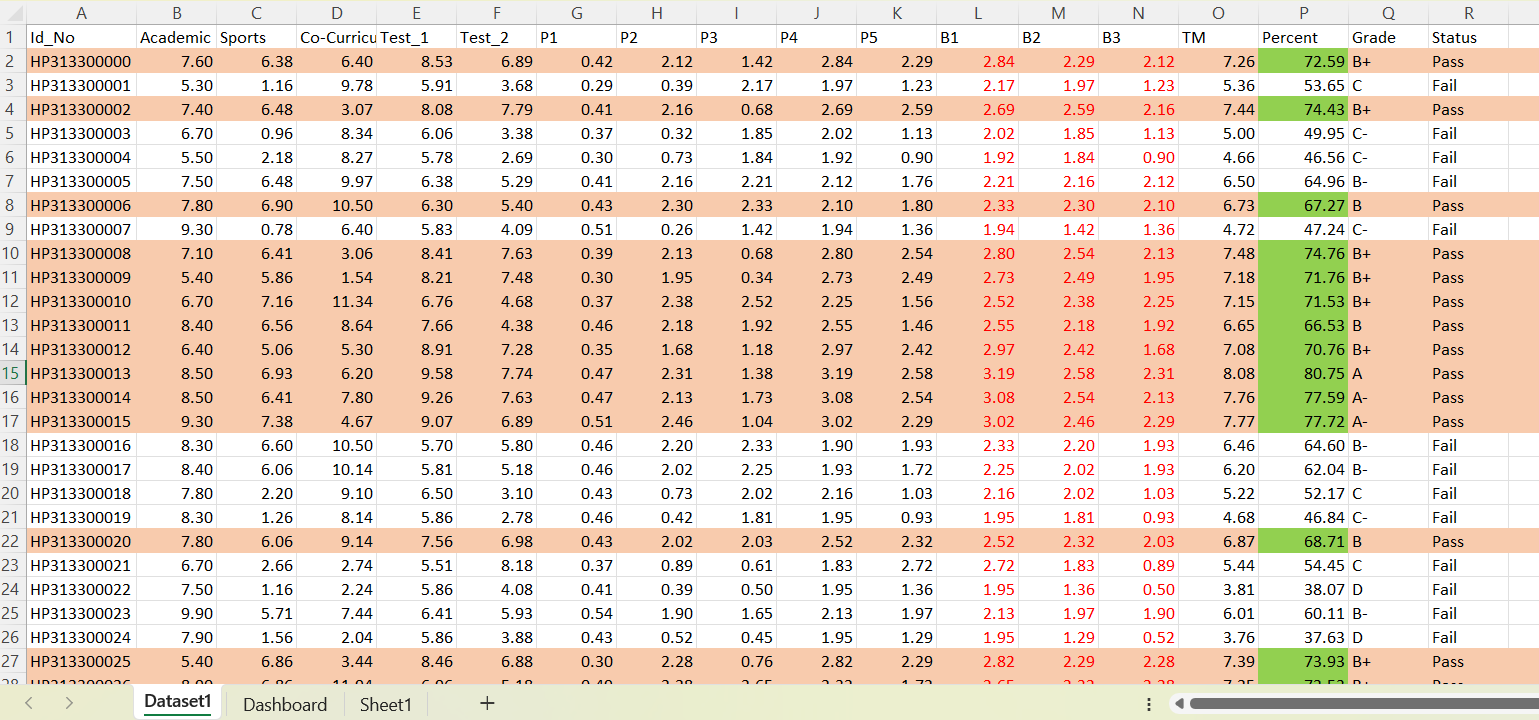


Figure 4: Processed dataset

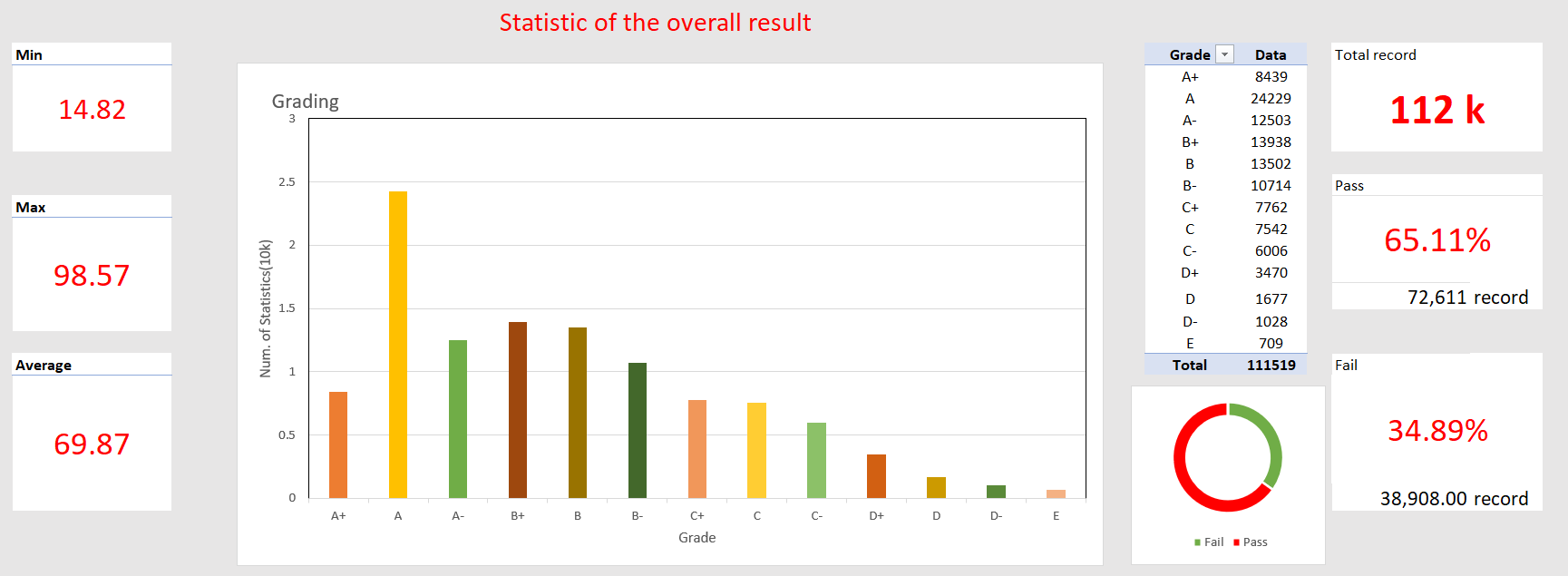


Figure 5: Dashboard

Conclusion

To summarize the whole report, this assessment is legitimate for students to pass as nearly 2/3 of the students passed and the average score is 69.87% which is higher than the passing score of 65%. From another aspect of evaluation, we standardize the score so that even if the maximum value they obtained in one item is far lower than the real maximum score of that item, they still can get a comparably good result based on their peer’s performance. The probability of getting a grade A is also not to be considered low because most of the students actually get these grades with a number of 24229 which is nearly equal to the 20 percent more of the total students.