

**MCSD1123- 01 BIG DATA MANAGEMENT**

**Assignment 1A Report**

**Lecturer:**

PM DR. Mohd Shahizan Othman

| **Members** | **Matric NO.** |
| --- | --- |
| Hazem Taher Omran Fenneer | MCS231019 |
| Shivanesh | MCS231014 |
| Mustafa | MCS212012 |
| Nur Shahirah | MEC233005 |

INTRODUCTION

The case study focuses on the data processing and visualization of a dataset named dataset1.txt. This dataset contains information on academic performance, sports participation, co-curricular activities, and test scores. The primary objectives of this case study are to transform and analyze the data in Google Sheets and create a visualization dashboard to gain insights into the dataset.

Data Entry

Importing the dataset into Google Sheets for further processing

Data Processing

To convert the Academic, Sports, Co-Curriculum, Test\_1, and Test\_2 data values to two decimal places in Google Sheets, we used the "Format" option. Select the cells containing the data values. In our case, select all the cells in column B to F. Right-click on the selected cells, or go to the "Format" menu at the top.

1. Choose "Number" from the dropdown menu.
2. In the sub-menu that appears, select "Number."
3. A dialog box may open, allowing you to set the number of decimal places. Enter "2" in the "Decimal places" field.
4. Click "Apply" or "Done."

To provide a new maximum value of 3.33 for columns B (Academic) through F (Test 2) and update the values in columns G (P1) to K (P5) while displaying values to two decimal places in Google Sheets, you can use the following steps:

1. Since we want to scale the values to a maximum of 3.33, we divided 3.33 by the maximum value in each column.
2. We used the formula ‘=MAX(B:B)’ for column B to ‘MAX(F:F)’ for column F.
3. Once we get the maximum value for each column, we can give new values in columns G to K. We used the formula,

* =(B2/60.30)\*3.33
* =(C2/10)\*3.33
* =(D2/15)\*3.33
* =(E2/10)\*3.33
* =(F2/10)\*3.33

A table with numbers and letters

Description automatically generated

After applying the formula to cells in columns G to K, this is the result

To find the top three values, you can use the LARGE function. In cell L2, enter:

1. =LARGE(G2:G,1) to find the highest value.
2. In cell M2: =LARGE(G2:G,2) to find the second highest value.
3. In cell N2: =LARGE(G2:G,3) to find the third highest value.

A screenshot of a spreadsheet

Description automatically generated

After using the formula in columns L to N, here are the results.

To compute total points by combining the data from columns L to N. The total mark value is entered in column O(TM ) . Use the formula SUM for this task.

1. Enter ‘=SUM(L2:N2)’ and copy down the formula.

To calculate the percentage of the total mark, we will use the formula below:

 =(O2/10)\*100  (since the maximum possible score is 10)

Now, the task is to assign the grade in column Q and status( Pass or Fail) in column R.

For the grade, ‘IF’ function is used in column Q.

=IF(AND(P2 >= 90, P2 <= 100), "A+", IF(AND(P2 >= 80, P2 < 90), "A", IF(AND(P2 >= 75, P2 < 80), "A-", IF(AND(P2 >= 70, P2 < 75), "B+", IF(AND(P2 >= 65, P2 < 70), "B", IF(AND(P2 >= 60, P2 < 65), "B-", IF(AND(P2 >= 55, P2 < 60), "C+", IF(AND(P2 >= 50, P2 < 55), "C", IF(AND(P2 >= 45, P2 < 50), "C-", IF(AND(P2 >= 40, P2 < 45), "D+", IF(AND(P2 >= 35, P2 < 40), "D", IF(AND(P2 >= 30, P2 < 35), "D-", IF(AND(P2 >= 0, P2 < 30), "E", "")))))))))))))

For the status, ‘IF’ and ‘OR’ function is used in column R.

=IF(OR(Q17="A+", Q17="A", Q17="A-", Q17="B+", Q17="B"), "Pass", "Fail")

A screenshot of a test

Description automatically generated

After copying down the formula in column Q and R. This is the outcome.

To apply green color to column P when the status is "Pass”, we used conditional formatting. We followed these steps:

1. Select the range of cells in column P.
2. Go to "Format" in the top menu.
3. Select "Conditional formatting."
4. In the Conditional Format Rules panel on the right, choose the following settings:

* Format cells if
* "Text is exactly"
* Enter "Pass" without quotes in the text box
* Choose the green color in the formatting style.

1. Click “Done”

Now, column P will be colored green whenever the status is "Pass."

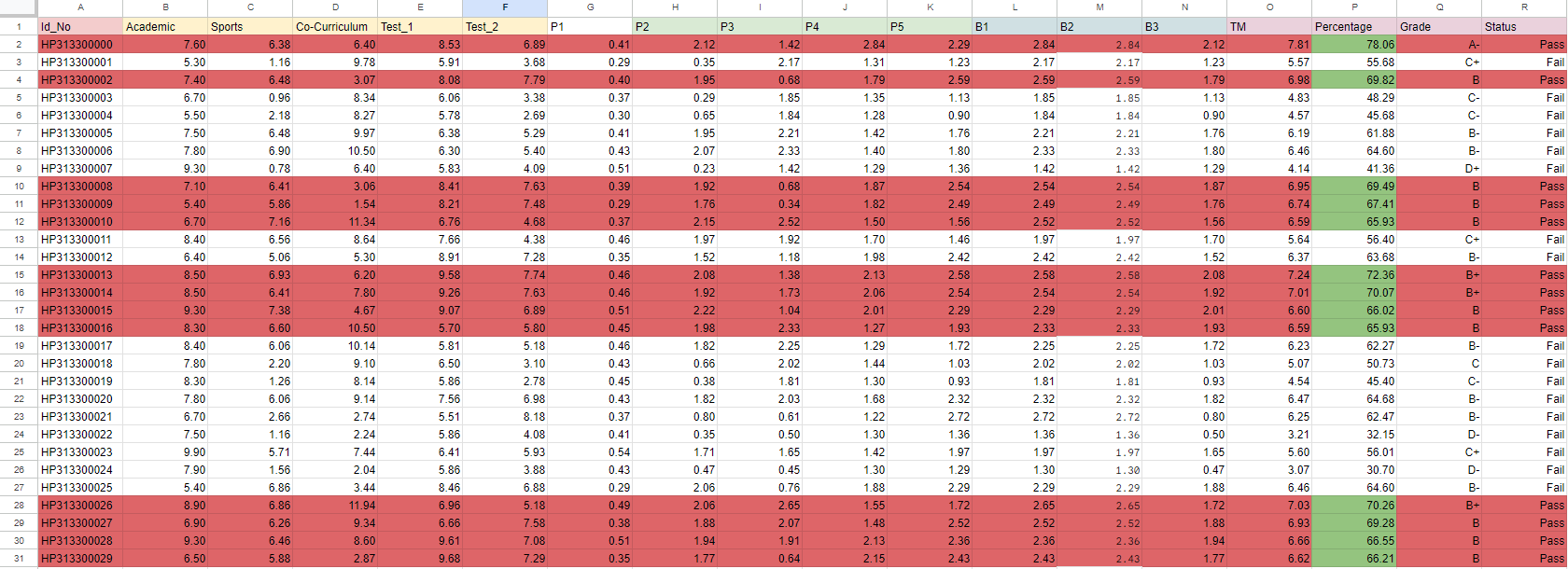
To color the entire row with light red when the status is "Pass," we used the following steps:

1. Select the entire row that corresponds to the "Pass" status. To do this, you can click on the row number (e.g., row 2) on the left-hand side of the spreadsheet.
2. Go to "Format" in the top menu.
3. Select "Conditional formatting."
4. In the Conditional Format Rules panel on the right, choose the following settings:

* Format cells if
* “Custom formula is”
* =$R2=”Pass”
* Choose light red fill color in the formatting style
* This formula checks if the status in column R is "Pass" for the current

1. Click "Done."

Now, all the rows with status as “Pass” will be filled with light red.



The final output of the data processing.

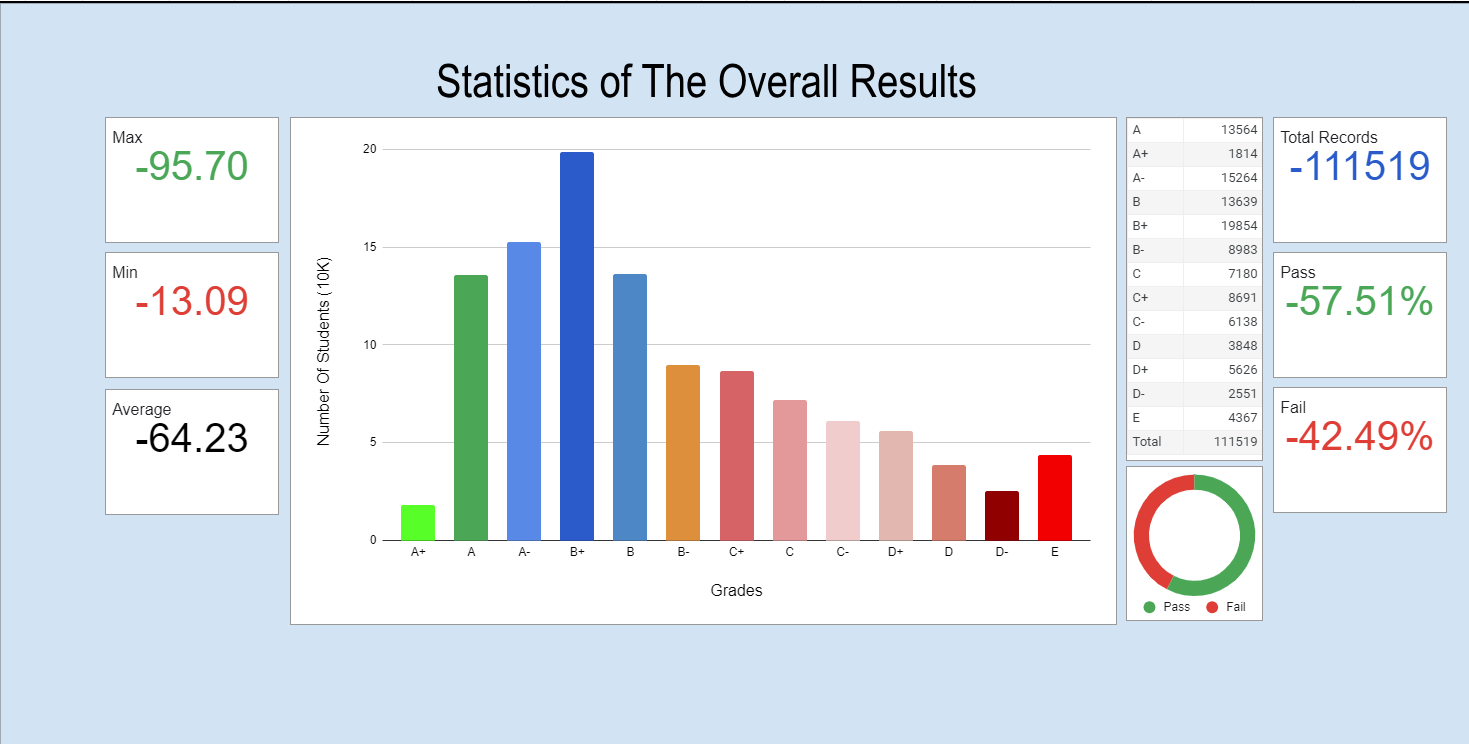
Data Visualization

Data visualization is the main objective of this task, to do so we have implemented the dashboard using Google Sheets to visualize the data based on the most important aspect of it.

First we determine what is needed to visualize, which was:

* The maximum, minimum and average grades
* The grades distribution for all students
* A table containing the number of students and their grades
* The total records of students with the pass and fail records number
* The contrast of pass and fail grades

For that we have implemented columned graphs, pie chart, table, and scorecard charts, which can be displayed in the figure below:



Conclusion:

In conclusion, the assignment goal was to articulate the diagram and create a dashboard that correctly visualizes the data, which was achieved. The final dashboard correctly visualizes the data contained in the dataset.