# Modified SQL Data

**Original Dataset:** https://www.kaggle.com/kunal28chaturvedi/covid19-and-its-impact-on-students

Link to my modified dataset for all the graphs:
https://docs.google.com/spreadsheets/d/126nid5bJwU_uKkEWwHIkAJt1o5yZCnvk1z3Z3uNinR0/edit?usp=sharing
I modified the original dataset to make it easier to use for creating graphs.

- For the modified dataset of Graph 1, I only copy the data that will be used in Graph 1, including “timeSpentOnFitness”, “numberOfMealsPerDay”, and “changeInWeight”. In Excel, I use the IF function to predict the change in weight. After that, I use the IF function again to see if the predictions are the same as the actual weight change for each student. If the values are equal, the cell prints "Yes". Otherwise, the cell prints "No". Last, I find out the pattern of the data, and then I use the COUNTIF function to get the results under certain conditions (the chart on the right).
- For the modified dataset of Graph 2, I only copy the data that will be used in Graph 2, including “id”, “timeSpentOnOnlineClass”, “timeSpentOnSelfStudy”, “timeSpentOnSocialMedia”, and “timeSpentOnTv”. First, in Excel, I use the SUM function to add student’s time spent on online class and self-study together to a new column. Then, I use the SUM function again to add student’s time spent on social media and TV together to a new column. In the end, I also create a new column to show the time difference between the time spent on study (online class + self-study) and leisure activities (social media + TV) for each student.
- For the modified dataset of Graph 3, I only copy the data that will be used in Graph 3, including “id”, “moreConnectionWithFamilyCloseFriendsOrRelatives”, and “healthIssueDuringLockdown”. In Excel, I use the IF function to see if the answers of “moreConnectionWithFamilyCloseFriendsOrRelatives” are the same as the answers of “healthIssueDuringLockdown”. If the values are equal, the cell prints "Yes". Otherwise, the cell prints "No". Last, I use the COUNTIF function to get the results I need under certain conditions (the chart on the right).

# Data Visualization

Graph 1 (Depicts my Q1): https://www.datawrapper.de/_/9pArb/

a. The graph shows the relationship between students' fitness time & the number of meals and students' weight.
b. By looking at the graph, the reader can know that the fitness time and the number of meals **WILL NOT** affect the students' weight.
c. I use the "Grouped Column" chart.
    - I chose this sort of graph because it allows the reader to quickly see the differences between the count of "Increased Weight" and the count of "Decreased Weight".
    - X-axis: 2 different sets of criteria ("Eat More Meals & Do Less Exercise" and "Eat Fewer Meals & Do More Exercise")
        - In each group, there are 2 values ("Increased Weight" and "Increased Weight").
    - Y-axis: Number of people

Graph 2 (Depicts my Q2): https://www.datawrapper.de/_/pLvzv/

a. The graph shows how long each student spends on study (online class and self-study) and leisure activities (social media and TV).
b. By looking at the graph, the reader can see the difference between the time each student spends on study and leisure activities.
c. I use the "Bullet Bars" chart.
    - I chose this sort of graph because it allows the reader to easily and clearly see the difference between the time each student spends on study and leisure activities. I also think that using this chart to read the differences is more interesting than looking at a bunch of numbers in tables like SQL and Excel.
    - X-axis: Time spent per day (in hours)
        - The reader can observe the difference in time use of each student by looking at the gap between the color bars (purple and yellow).
    - Y-axis: Each student's ID

Graph 3 (Depicts my Q3): https://www.datawrapper.de/_/1m4OE/

a. The graph shows the relationship between students' health and students' connection with those who are close to them.
b. By looking at the graph, the reader can know that the students' connection with their family, close friends, and relatives **WILL NOT** cause their health issues during the lockdown.
c. I use the "Pie" chart.
    - I chose this sort of graph because it allows the reader to immediately know the result I want to present.
    - Light Purple Area: The count of "Yes" (if both answers in the survey are the same)
    - Dark Purple Area: The count of "No" (if both answers in the survey are different)