CIDM/ECON 6308 HW1: Data Analytics in Microsoft Excel

Learning Objectives:

- Refresh your understanding of basic statistic terms;
- Get familiar with common data analysis functions/formulas in Excel;
- Perform descriptive analytics by making charts, tables, or basic PivotTable in Excel.

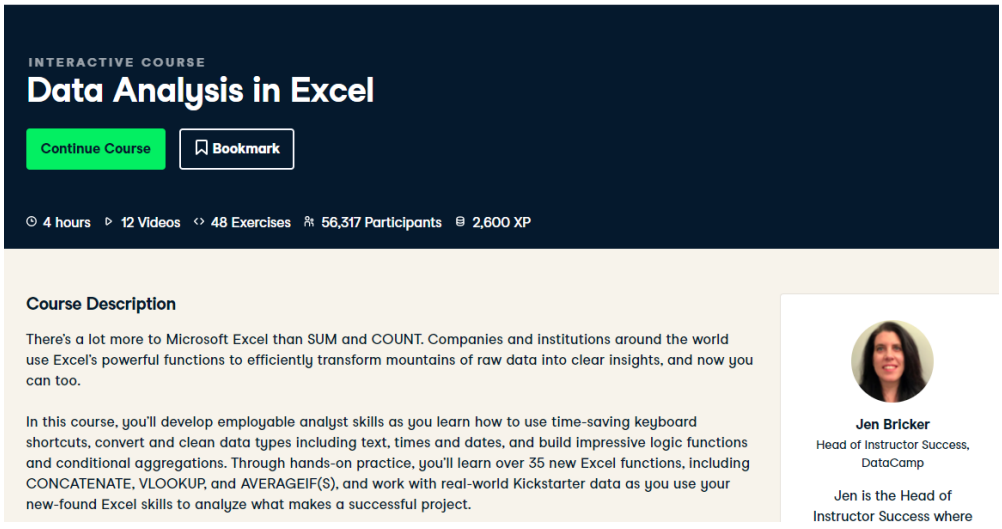
Submission

Please submit all your responses via HW1 Submission on WTClass so that you can view your grade immediately to decide if you want to resubmit your answers. You have up to two attempts and the higher one will be counted into your final grade.

Part 1: DataCamp Course (20 points)

Please complete a DataCamp course, Data Analysis in Excel. This course includes three chapters, Exploring Data, Preparing Data, and Analyzing Data.

Attention: Before starting this course, please make sure that you have joined our class group on DataCamp.



You don't need to provide any proof for this section as I can verify your completion status on DataCamp. If you have previously completed this course, please review it to refresh your memory.

Part 2: Data-driven Decision Making in Excel (40 points)

Business Understanding

The instructor of CIDM/ECON 6308 wants to understand his students' background on data analytics (including their current capability, what analytics skills they are eager to learn, what analytics software they have used) and time commitment (the amount of time that students can dedicate to this course) and allocation (how students allocate their committed time to this course each day). Therefore, in the beginning of each semester, all the CIDM/ECON students are invited to complete a short survey.

Data Understanding

The Excel file, HW1.xlsx, includes three worksheets:

- The first worksheet, Student_Response_2024, consists of 44 valid responses obtained in the spring semester of 2024: The Question ID (or variable name) in the first row and all the responses from 44 students to the questions in the survey in rows 2-45.
- The second worksheet, Student_Response_2023, consists of 41 valid responses obtained in the spring semester of 2023: The Question ID (or variable name) in the first row and all the responses from 41 students to the questions in the survey in rows 2-42.
- The third worksheet, Variable_Definition, describes all the variables in the previous two worksheets and measurement scales of some variables.

Please follow each step to complete all the required questions. Attention: the graphic illustrations in this instruction are only for your reference about how because they are not based on this year's data.

Please use the 2024 spring data to answer the following questions.

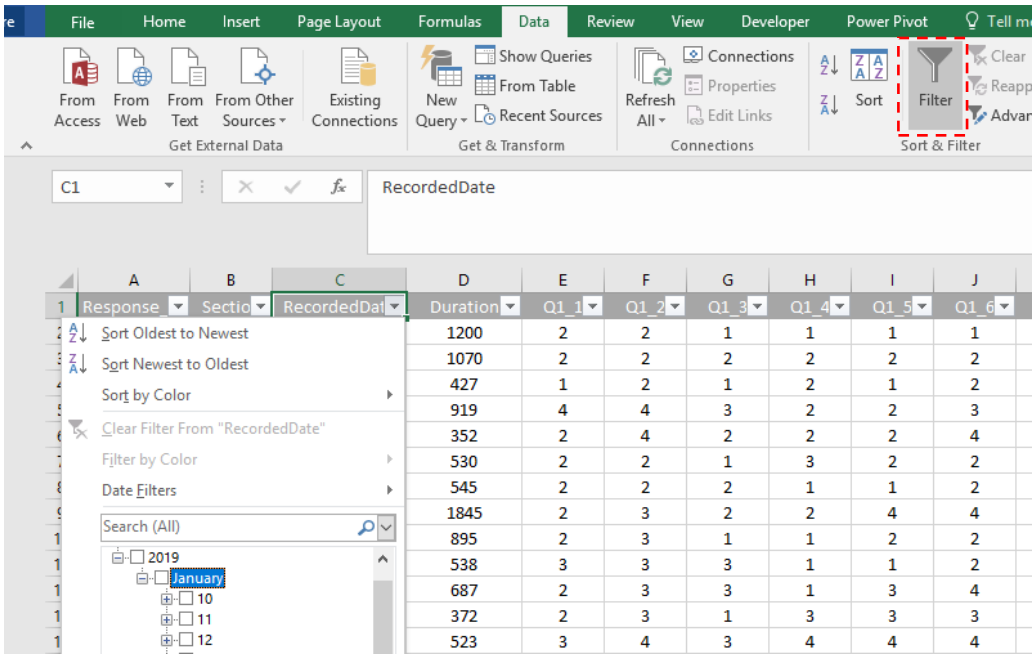
1. Analyzing the basic information of the 2024 spring survey (10 points)
   1.1. Take a look at the second column (Column B), Section. This semester, this course is cross-listed as CIDM 6308 and ECON 6308.

   1.2. The third column (Column C) shows the program in which each student is enrolled. Among the nine programs listed in the survey, which one contributes the highest responses? Type MBA CIS, MBA General, MBA Healthcare, MBA Management, MBA Marketing, MS CISBA, MSFE, MBA Procurement, or MPA. How many responses were obtained from this program? (2 points).

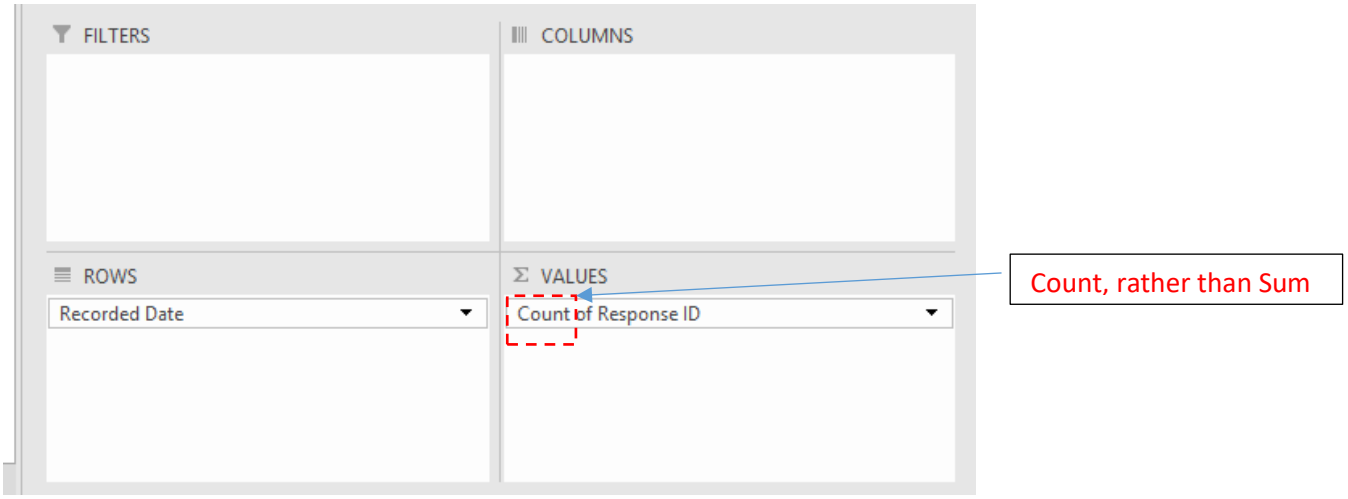
   1.3. Take a look at the fourth column (Column D), RecordedDate. Please indicate that on which date we obtained the largest number of responses? How many responses were received on that day? (2 points)

   You can use whatever method for the two questions above in Excel, even though two methods are provided below (Use Q1.3 as an example).

   1) The first way is using the Filter function (you can click the hyperlink for more instruction): Then click RecordedDate and select each date to find the number of records received in each date.



2) The second or the easiest way is to insert a Pivottable (you can click the hyperlink for more instruction if you do not know how to insert a PivotTable): and drag two variables into it as below.



1.4. The fifth column (Column E) shows how much time (in seconds) each student spent on the survey. Please compute:

1.4.1. The minimum value and the maximum value of the variable duration.

1.4.2. The mean, median, and sample standard deviation of the duration. Round your answers to a whole number (i.e., integer).

Note: You can use the formulas MIN, MAX, AVERAGE, MEDIAN, and STDEV.S in Excel to compute the above-mentioned measures, respectively.

Pause and Think: You will notice that mean and mean in this case are very different and that the sample standard deviation is extremely large. This is usually caused by extreme values or outliers.

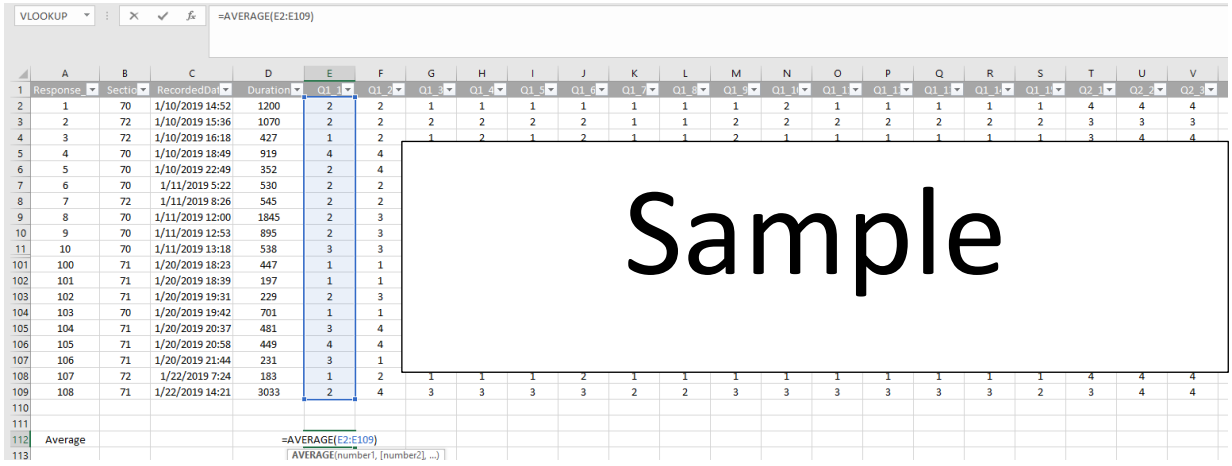
1.4.3. Successful completion of a survey requires the respondents to invest sufficient time. Typically, if a participant spent too little or too much time on the survey, their response may be invalid. Based on the following two criteria, how many responses in total will be considered as invalid?

- On one hand, any responses with less than 4 minutes (240 seconds) will be counted as invalid because they might not take the survey seriously.
- On the other hand, completing this short survey needs no more than 1 hour. If someone spent more than one hour (i.e., 3,600 seconds), s/he probably was distracted when taking this survey, so we count such responses invalid as well.

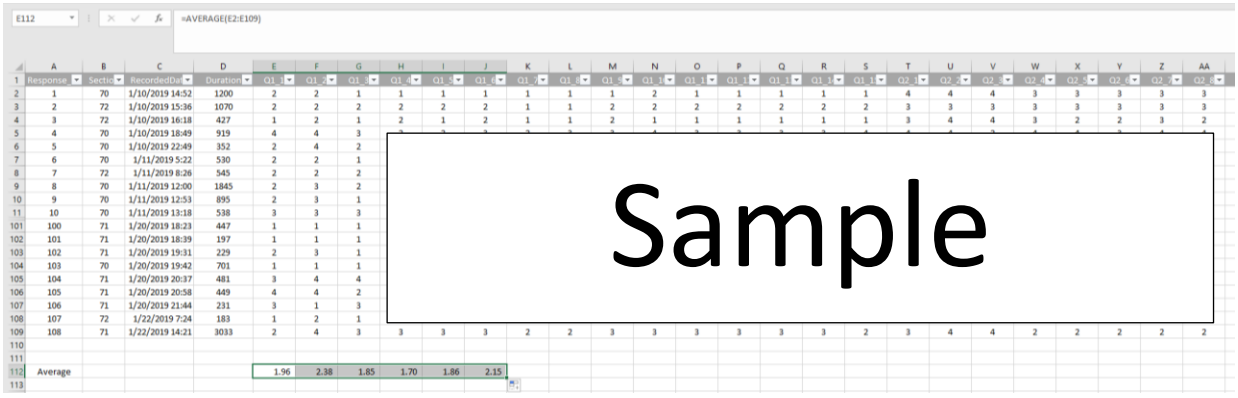
1.4.4. Invalid response(s) must be further examined (or even deleted) in practice. However, in this homework assignment, **we would rather keep them in our dataset** (just in case that a student did not get the correct answers to Q1.4.3, s/he would miss all the following questions).

2. Please use the first worksheet, Student_Response_2024, to compute students' average current capability in each skill (this involves columns Q1_1 to Q1_16) and then generate a column or bar chart to show the average capability in each skill. Based on this, please answer a few questions. **The following steps provide a detailed instruction which you are going to follow; however, please notice that the data is different from what you are using now.**

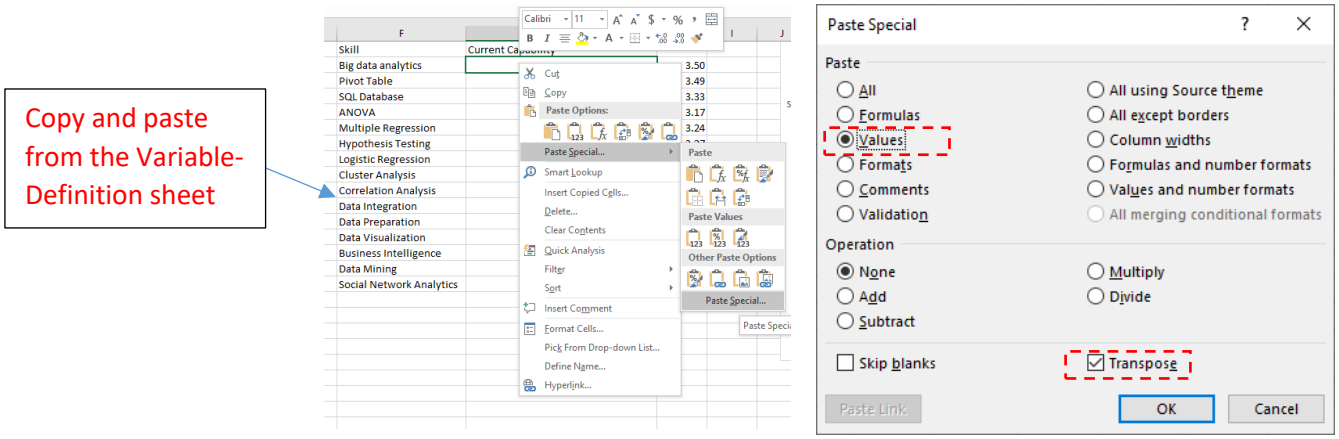
2.1. Step 1: **compute the average and round the average to the second decimal place using the Average formula in Excel.**



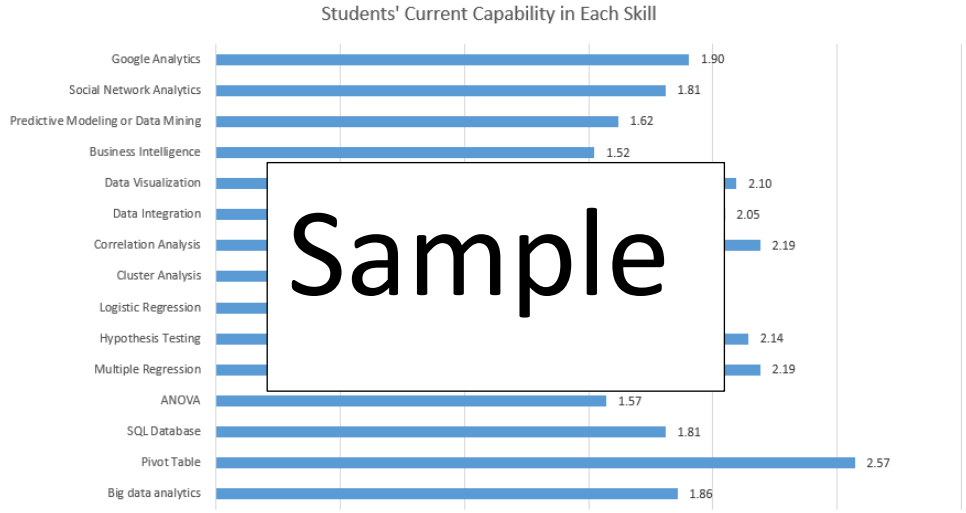
2.2. Step 2: copy the formula to the other skills (drag the cell all the way to the right)



2.3. Step 3: copy all the averages to another worksheet and match the column ID with each skill (hint: when copying and pasting those cells, please use paste special (check both Values and Transpose).



2.4. Step 4: generate a column chart or bar chart (please show the data label in the chart; if you do not know how to create a chart, please click the hyperlink for more instruction). You do not need to submit this chart, but it is always a good manner to visualize the data before actual analysis.



Hint: you may consider sorting the value in a descending order to help you quickly identify the answers.

2.5. Step 5: observe the chart (or the sorted result) and then identify the skill in which our students currently exhibit the highest capability and the one in which they are least capable. Also, provide the respective averages for each. (4 points).

3. Following the same procedure above (Steps 1-5), please compute students' average eagerness to learn each skill (Columns Q2_1 to Q2_16 in the first worksheet, Student_Response_2024) and then generate a column or bar chart to show the average eagerness in each skill. Round all the numbers to the second decimal place such as 0.12. Please identify the top three skills that our students are most eager to learn from this course and provide their averages. (6 points)

4. The column Q3 in the dataset describe what statistical or analytic software packages have been used by each student (7 points).

4.1. We would like to determine the frequency of each software used by our students. Please complete the following frequency table. If you are unsure of the process, you can refer to the instructional video available on WTClass.

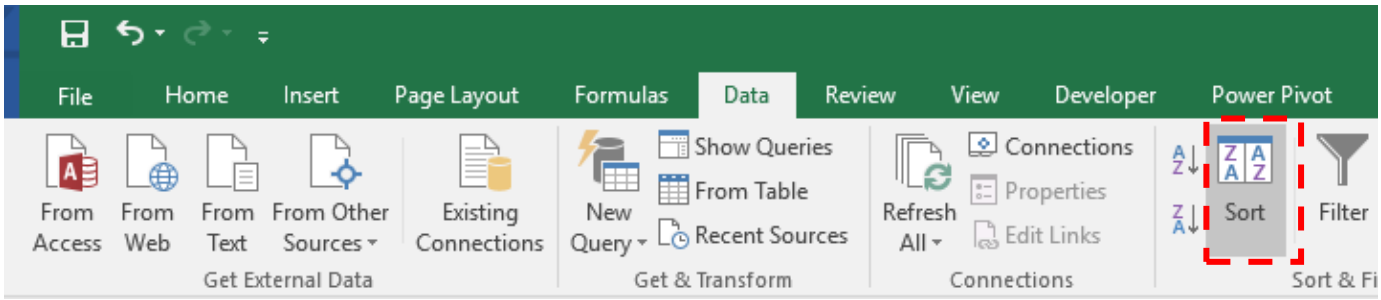
Scale	Software Name	Frequency
1	MicroStrategy	0
2	JMP	3
…	…	…
99	I have never used any software listed above.	

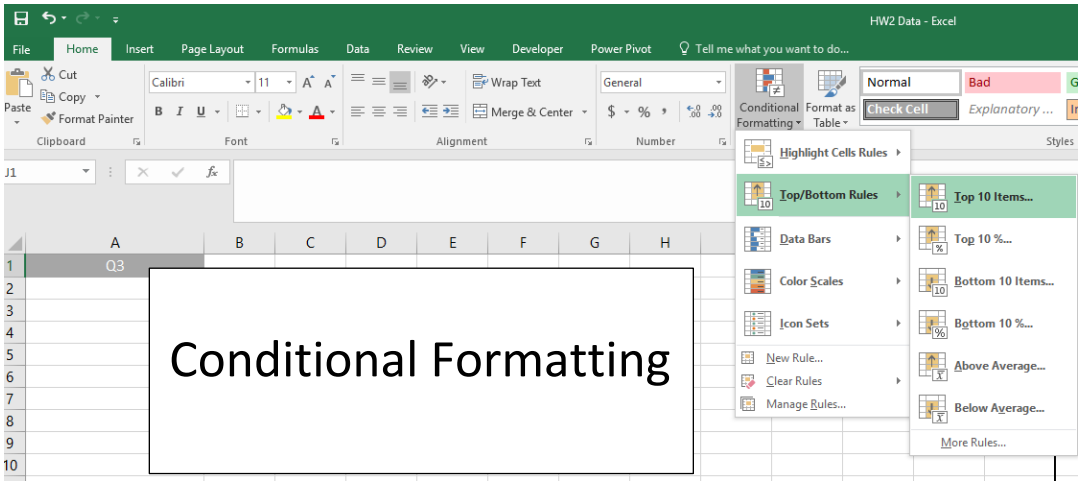
4.2. Based on the frequency table, please answer the following questions:

4.2.1. Please find the frequency of the four required software packages listed in our syllabus.

Software Name	Frequency
Microsoft Excel	
SQL	
Tableau	
RapidMiner	

4.2.2. Identify the two analytics software packages that our students most frequently used. You can utilize either the sort function or the conditional formatting function in Excel for this task.





4.2.3. Among all the software packages listed in the frequency table, how many have never been used by students (i.e., frequency =0)? Type a whole number here.

4.3. **Pause and Think**: Pick one software package that you have never used before but are interested in learning, then do some research about it. You do not need to submit your answer.
- The basic information about the software, including its name, brief description, provider, and main purposes and functions.
- Your interest in this software package, including why you are interested in it and how you may plan to use it in the future.
- Share your response in our community discussion.

5. Our course is quite practice-intensive (esp., starting from Week 5), so please reserve sufficient amount of time (e.g., 8 to 15 hours, depending on your analytical and statistical skills) for each class, including discussion, assigned readings, quiz, lab section, and homework. Typically, the amount of time spent on a course largely determines a student's grade and other learning performance (e.g., analytical capability which cannot be measured by the grade only).
   5.1. Based on Q4 (Column AM), please compute the average time (in hours) our students can spend on this course each week. (1 point). Round to the second decimal place such as 8.22.
   5.2. Then, compute the average time that CIDM 6308 and ECON 6308 students can spend on this course each week, respectively below (2 points). Round to the second decimal place such as 8.22.

Section	Average time students can spend on this course each week
CIDM 6308	Mean for students in CIDM 6308
ECON 6308	Mean for students in ECON 6308

   5.3. **Pause and Think**: can you spend the average amount of time in this course? Discuss why you think the amount of time you are willing to commit to this course is sufficient for you to achieve your goal in this course. The goal of this question is to encourage you to think about your study plan for this course, and you do not need to submit your answer.

6. The last seven columns (Q4_1 to Q4_7) in the survey show the percentage of time each student can spend on this course each day). This is relevant to the scheduling of course. Please compute the average percentage of time that our students can spend on this course each day, respectively (4 points).
   6.1. Complete the following table to display the percentage of time that our students can spend on this course each day (round all the numbers to integers). Some results have already been provided below.

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Percentages	_____	_____	9	13	_____	_____	23

   6.2. **Pause and Think**: think about how you manage your time to make great use of each small or big piece of your time. Do not wait until the weekend or due day to start an important homework. Read through an assignment during the week to clear any technical obstacles, problems, confusion, and/or misunderstanding. Take a look at the success strategies, shared in Class 00 and you will find those top performers in previous semesters are quite good at time management and scheduling.

7. Comparison between 2023 and 2024 Results (6 points).
    - Identify the skill in which the 2023 Spring students, found in the second worksheet 'Student_Response_2023,' were most capable. Determine if this skill aligns with the one identified in 2024 in Step 2.5. Type Yes or No here.
    - Identify the top three skills that the 2023 Spring students are most eager to learn. Indicate how many of these skills are also in the top three list for 2024 Spring, as identified in Step 3. Note: Consideration does not need to be given to the rank or order in the top three list. Type a whole number here.
    - Identify the top two software packages that the 2023 Spring students had ever used. Determine whether these packages are the same as those identified in 2024 Spring in Step 4.2.2. Type Yes or No here.
    - Calculate the average amount of time that the 2023 Spring students indicated spending on this course. Compare it with the value identified in 2024 Spring in Step 5.1. If the difference is within 0.5 hours, consider them as the same. Type Yes or No here. Of course, you can conduct a two-sample t-test to determine whether there is a significant difference between them, but such a test is not required here.