

# 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.

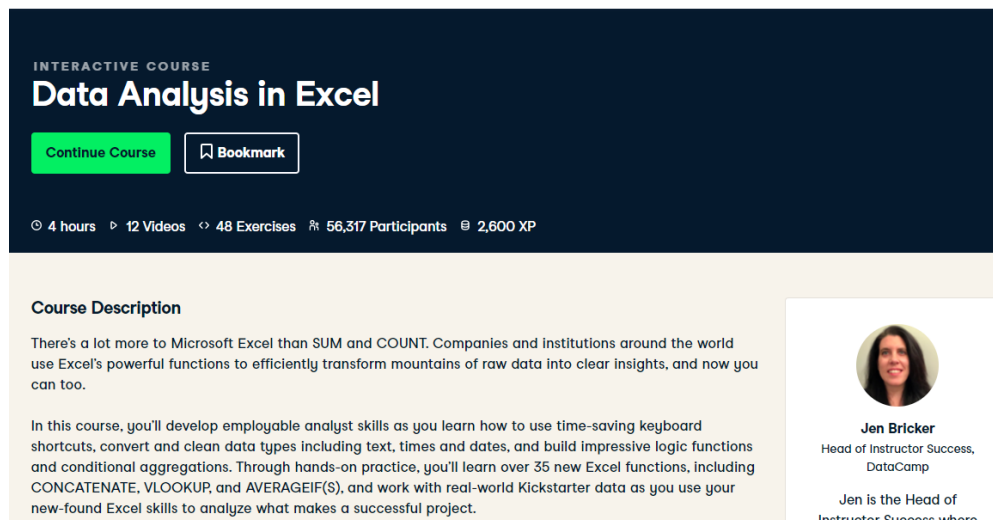
## Submission

Please submit all your responses via HW1 Submission on WTCClass 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 group via the link provided under Class 02 folder.**



The screenshot shows the DataCamp course page for "Data Analysis in Excel". At the top, it says "INTERACTIVE COURSE" and "Data Analysis in Excel". Below this are two buttons: "Continue Course" (green) and "Bookmark" (white with a bookmark icon). Underneath the buttons, it displays course statistics: "4 hours", "12 Videos", "48 Exercises", "56,317 Participants", and "2,600 XP". The "Course Description" section follows, stating: "There's a lot more to Microsoft Excel than SUM and COUNT. Companies and institutions around the world use Excel's powerful functions to efficiently transform mountains of raw data into clear insights, and now you can too." It then describes the course content: "In this course, you'll develop employable analyst skills as you learn how to use time-saving keyboard shortcuts, convert and clean data types including text, times and dates, and build impressive logic functions and conditional aggregations. Through hands-on practice, you'll learn over 35 new Excel functions, including CONCATENATE, VLOOKUP, and AVERAGEIF(S), and work with real-world Kickstarter data as you use your new-found Excel skills to analyze what makes a successful project." On the right side, there is a profile picture of Jen Bricker, followed by her name "Jen Bricker" and her title "Head of Instructor Success, DataCamp". Below that, it says "Jen is the Head of Instructor Success where".

You do not need to submit any proof for this part because I can check your completion status on DataCamp. If you have already completed this course before, please notify the instructor at least two days before the deadline and then a substitute DataCamp course will be assigned to you.

## 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\_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 second worksheet, Student\_Response\_2022, consists of 68 valid responses obtained in the spring semester of 2022: the Question ID (or variable name) in the first row and all the responses from 68 students to the questions in the survey in rows 2-69.
- 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 2023 spring data to answer the following questions.

1. Analyzing the basic information of the 2023 spring survey (10 points)
  - 1.1. Take a look at the second column (Column B), Mode. This semester, this course is offered both online and on campus.
  - 1.2. The third column (Column C) shows the category of participants. **Among the eight categories listed in the survey, which one contributes the highest responses?** Type MBA CIS, MBA General, MBA Healthcare, MBA Management, MBA Marketing, MS CISBA, MSFE, or MPA. **How many responses were obtained from this category?** (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.

	A	B	C	D	E	F	G	H	I	J
	Response	Section	RecordedDate	Duration	Q1_1	Q1_2	Q1_3	Q1_4	Q1_5	Q1_6
1				1200	2	2	1	1	1	1
2				1070	2	2	2	2	2	2
3				427	1	2	1	2	1	2
4				919	4	4	3	2	2	3
5				352	2	4	2	2	2	4
6				530	2	2	1	3	2	2
7				545	2	2	2	1	1	2
8				1845	2	3	2	2	4	4
9				895	2	3	1	1	2	2
10				538	3	3	3	1	1	2
11				687	2	3	3	1	3	4
12				372	2	3	1	3	3	3
13				523	3	4	3	4	4	4

- 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.

Filters	Columns	Rows	Values
		Recorded Date	Count of Response ID

- 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 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\_2023, 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.

VLOOKUP

X

✓

fx

=AVERAGE(E2:E109)

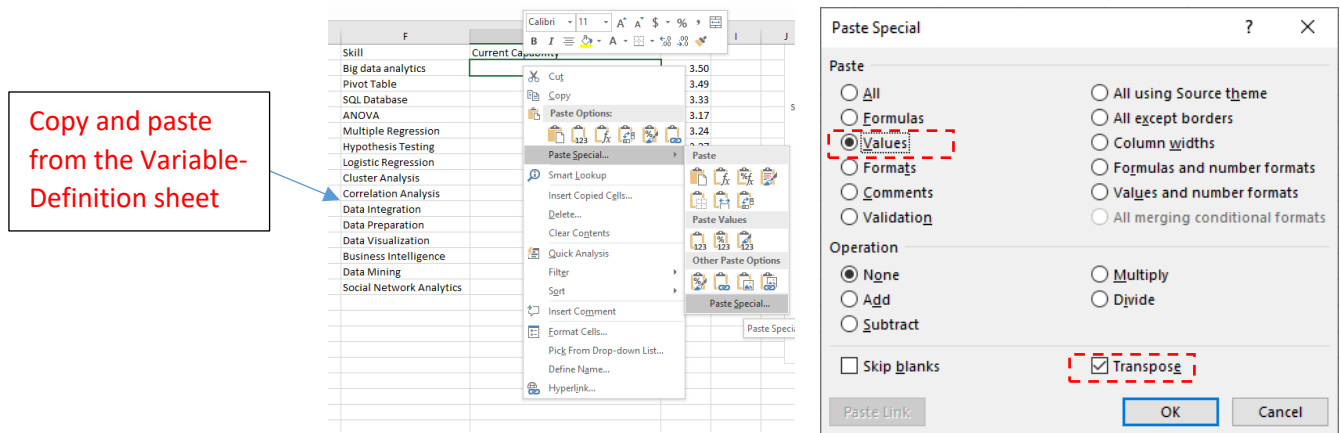
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Response	Section	RecordedDate	Duration	Q1_1	Q1_2	Q1_3	Q1_4	Q1_5	Q1_6	Q1_7	Q1_8	Q1_9	Q1_10	Q1_11	Q1_12	Q1_13	Q1_14	Q1_15	Q2_1	Q2_2	Q2_3
2	1	70	1/10/2019 14:52	1200	2	2	1	1	1	1	1	1	1	2	1	1	1	1	1	4	4	4
3	2	72	1/10/2019 15:36	1070	2	2	2	2	2	1	1	2	2	2	2	2	2	2	2	3	3	3
4	3	72	1/10/2019 16:18	427	1	2	1	2	1	2	1	1	2	1	1	1	1	1	1	3	4	4
5	4	70	1/10/2019 18:49	919	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
6	5	70	1/10/2019 22:49	352	2	4																
7	6	70	1/11/2019 5:22	530	2	2																
8	7	72	1/11/2019 8:26	545	2	2																
9	8	70	1/11/2019 12:00	1845	2	3																
10	9	70	1/11/2019 12:53	895	2	3																
11	10	70	1/11/2019 13:18	538	3	3																
101	100	71	1/20/2019 18:23	447	1	1																
102	101	71	1/20/2019 18:39	197	1	1																
103	102	71	1/20/2019 19:31	229	2	3																
104	103	70	1/20/2019 19:42	701	1	1																
105	104	71	1/20/2019 20:37	481	3	4																
106	105	71	1/20/2019 20:58	449	4	4																
107	106	71	1/20/2019 21:44	231	3	1																
108	107	72	1/22/2019 7:24	183	1	2																
109	108	71	1/22/2019 14:21	3033	2	4																
110																						
111																						
112	Average				=AVERAGE(E2:E109)																	
113					[AVERAGE(number1,[number2],...)]																	

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

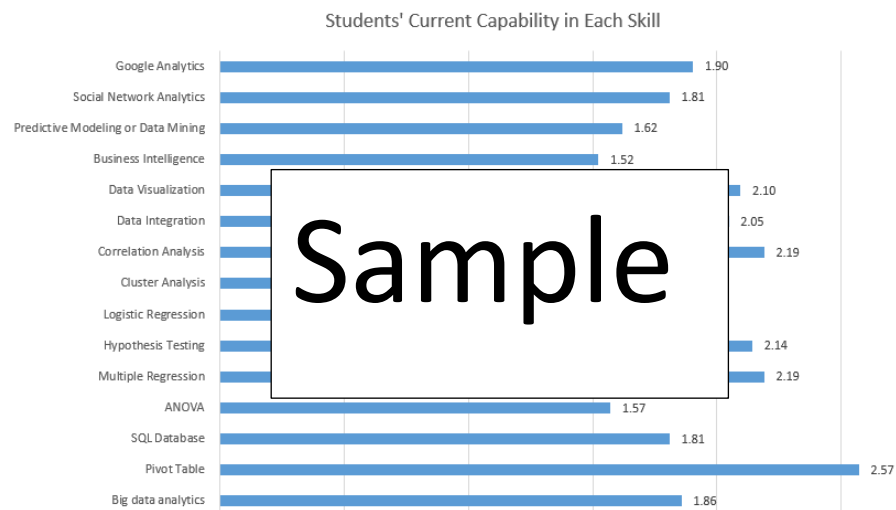
E112

</

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 answer the questions: Which skill our students are the most capable in currently and which one our students are the least capable in currently? What are their averages? (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\_2023) 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 find the top three skills our students are the most eager to learn from this course. What are their averages? (6 points)

4. The column Q3 in the dataset describe what statistical or analytic software packages have been used by each student (9 points).

- 4.1. We want to know the frequency of each software used by our students. Please complete the following frequency table. If you do not know how to do this, you may refer to the instruction video provided on WTClass.

Scale	Software Name	Frequency
1	MicroStrategy	1
2	JMP	0
...	...	...
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. Please find the top three analytics software packages used by our students (you can use either the sort function or the conditional formatting function in Excel).

The top screenshot shows the Microsoft Excel ribbon with the 'Data' tab selected. The 'Sort' button, located in the 'Sort & Filter' group, is highlighted with a red dashed box. The bottom screenshot shows the 'Conditional Formatting' menu open, with the 'Top 10 Items...' option highlighted. A large white box with the text 'Conditional Formatting' is overlaid on the bottom screenshot.

- 4.2.3. Among all the software packages (1-31) listed in the frequency table, how many have never been used by students (i.e., frequency = 0)? Type a whole number here.

4.2.4. Among all the software packages, which one is mentioned in Class 03 lecture? Type it 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.

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) students from each mode can spend on this course each week (2 points).** Round your answer to the second decimal place such as 8.22. Ignore any missing values in this step.

Mode	Average time students can spend on this course each week
On Campus	Mean for students on campus
Online	Mean for students online

5.2. **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 on-campus and online students can spend on this course each day, respectively **(6 points)**.

6.1. Complete the following table and then **type the percentage of time that on-campus and online students can spend on Sunday and Wednesday**, respectively (round all the numbers to integers). Some results have already been provided below.

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
On Campus students	_____		17	_____	10		19
Online students	_____	12		_____		11	

6.2. Based on the table above, you can compare the time distribution of on-campus and online students.

- \_\_\_\_\_ students spend a higher percentage of their weekend time (Friday, Saturday, and Sunday) on this course. **Type Online or On campus.**
- Both on-campus students and online students spend almost the same percentage of their time on \_\_\_\_\_. **Type Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, or Saturday.**

6.3. **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 2022 and 2023 Results (3 points).**

- Please find the skill that the 2022 Spring students (which you can find from the second worksheet, Student\_Response\_2022) were the most capable in. Is this skill the same as the one in 2023 that you identified in Step 2.5? **Type Yes or No here.**
- Please find the top three skills that the 2022 Spring students are most eager to learn and then indicate how many of them are also in the top three list in 2023 Spring that you identified in Step 3. **Type a whole number here.** Note: You do not need to consider the rank or order in the top three list.
- Please find the top three software packages that the 2022 Spring students had ever used and then indicate how many of them are also in the top three list in 2023 Spring that you identified in Step 4.2.2. **Type a whole number here.** Note: You do not need to consider the rank or order in the top three list.