**Exam 1 – Part 2 Submission**

Megan Moore

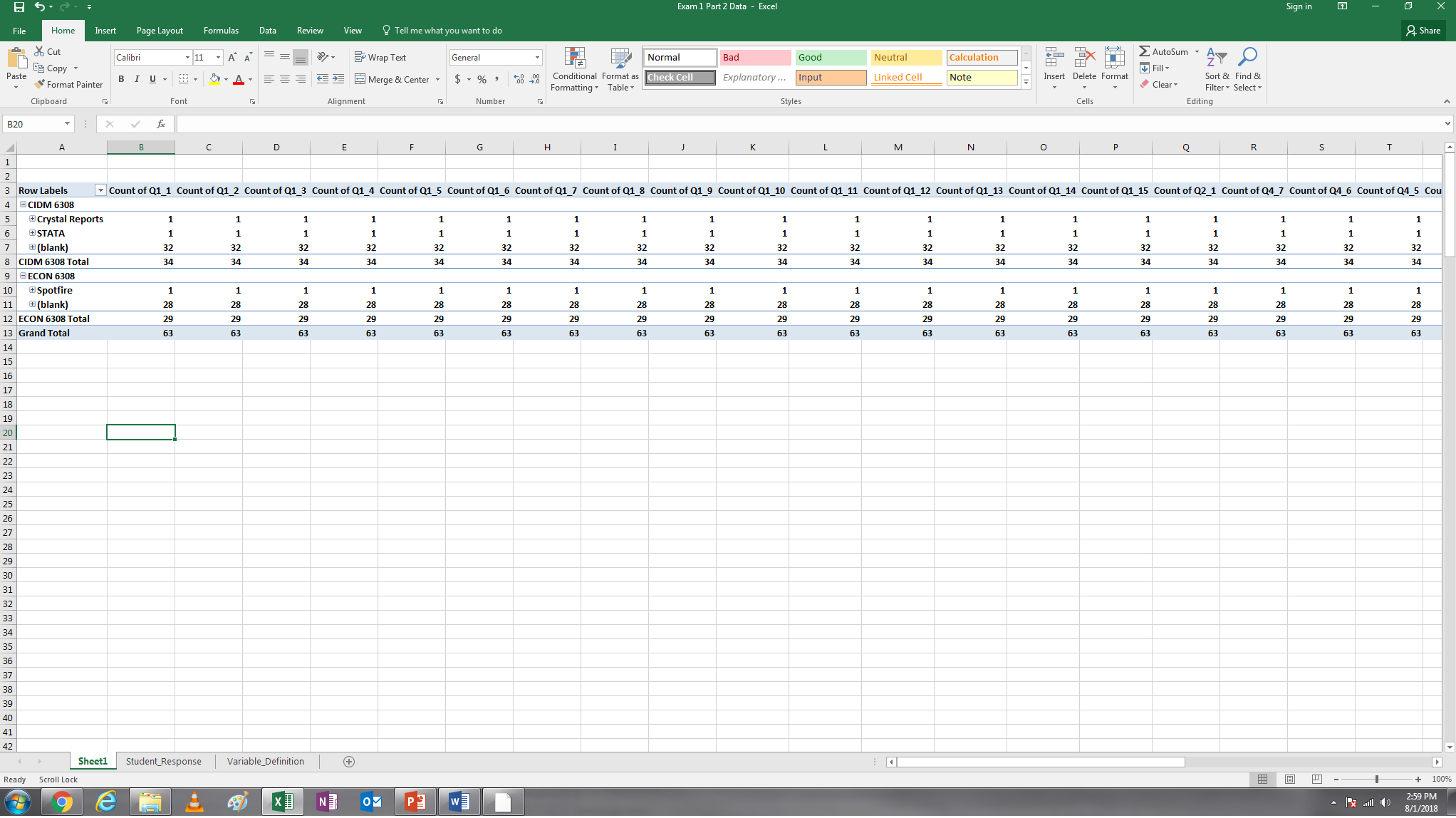
[Mmmoore2@buffs.wtamu.edu](mailto:Mmmoore2@buffs.wtamu.edu)

1. Analyzing the basic information of the survey itself.

1.1. Take a look at the first column, Class. Please indicate the number of responses from CIDM 6308 and ECON 6308, respectively (4 points).

CIDM 6308 : 34

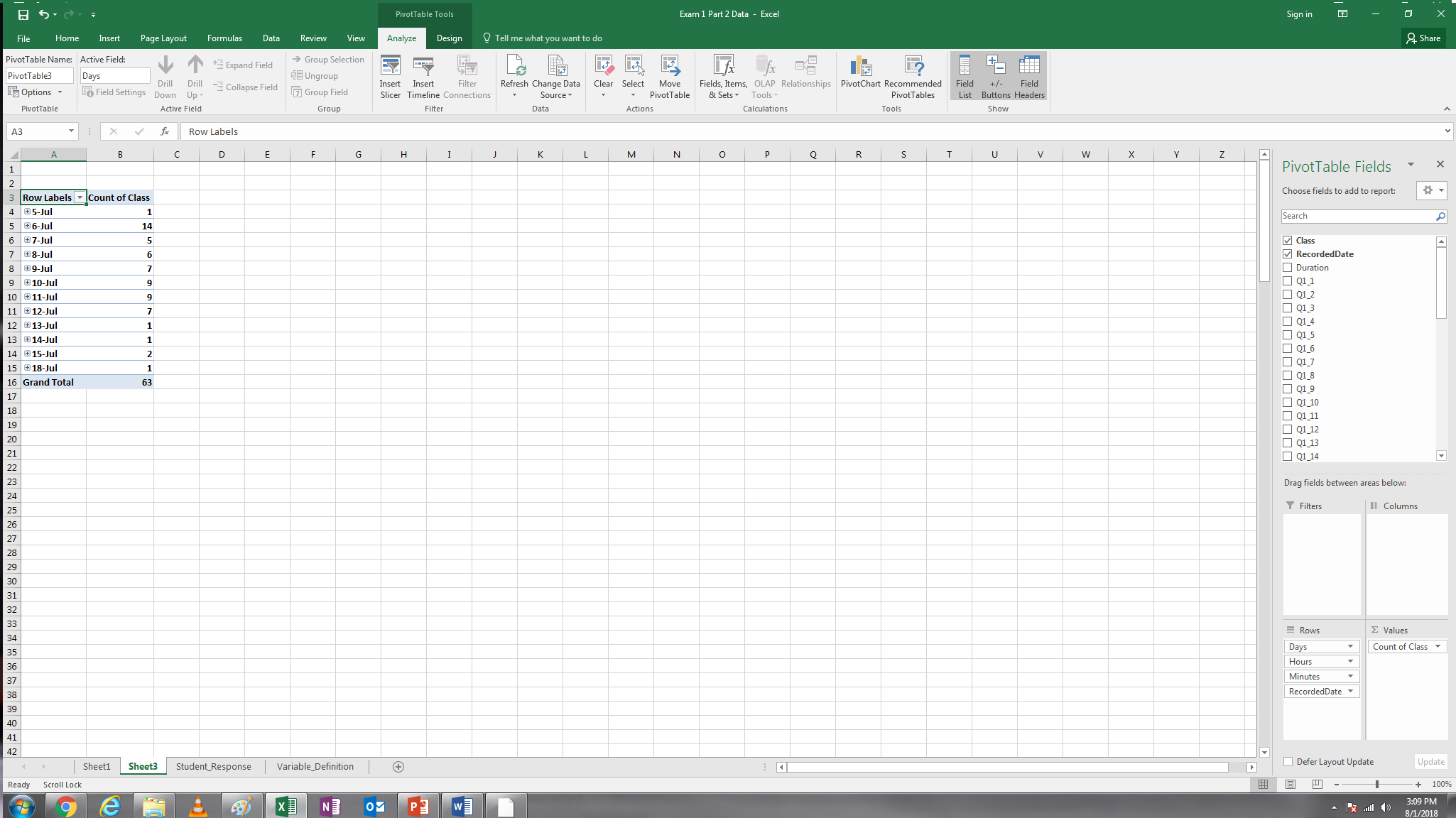
ECON 6308 : 29



1.2. Take a look at the second column, RecordedDate. Please indicate that which date obtained the largest number of responses? July 6, 2018

How many responses were received on that day? 14

(4 points) Take a screenshot of your Pivottable or any supporting evidence generated by other tools (Screenshot 1).

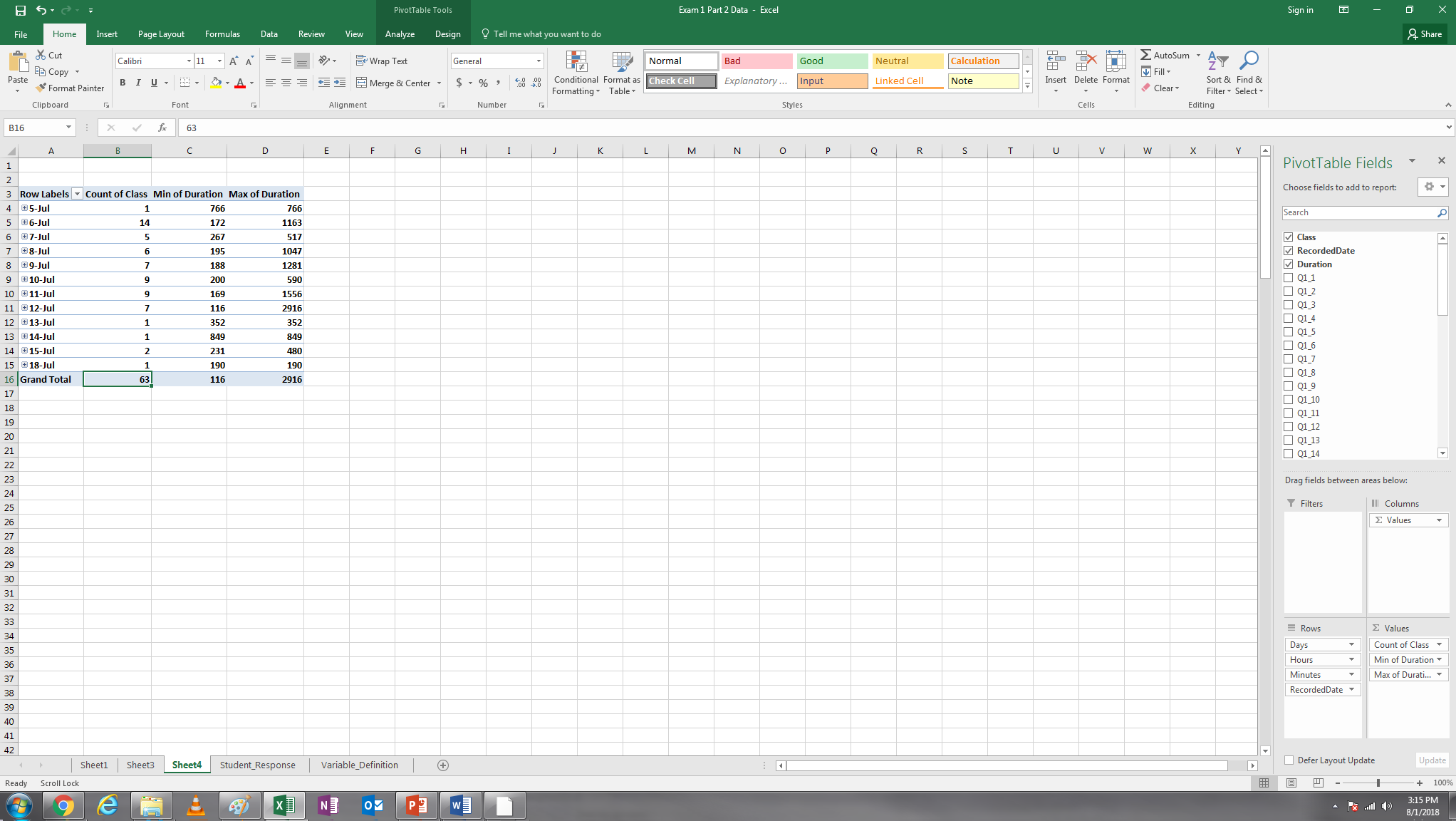


1.3. The third column shows how much time (in seconds) each student spent on the survey. Please compute:

1.3.1. The minimum and maximum values of the variable duration (4 points)

Min: 116

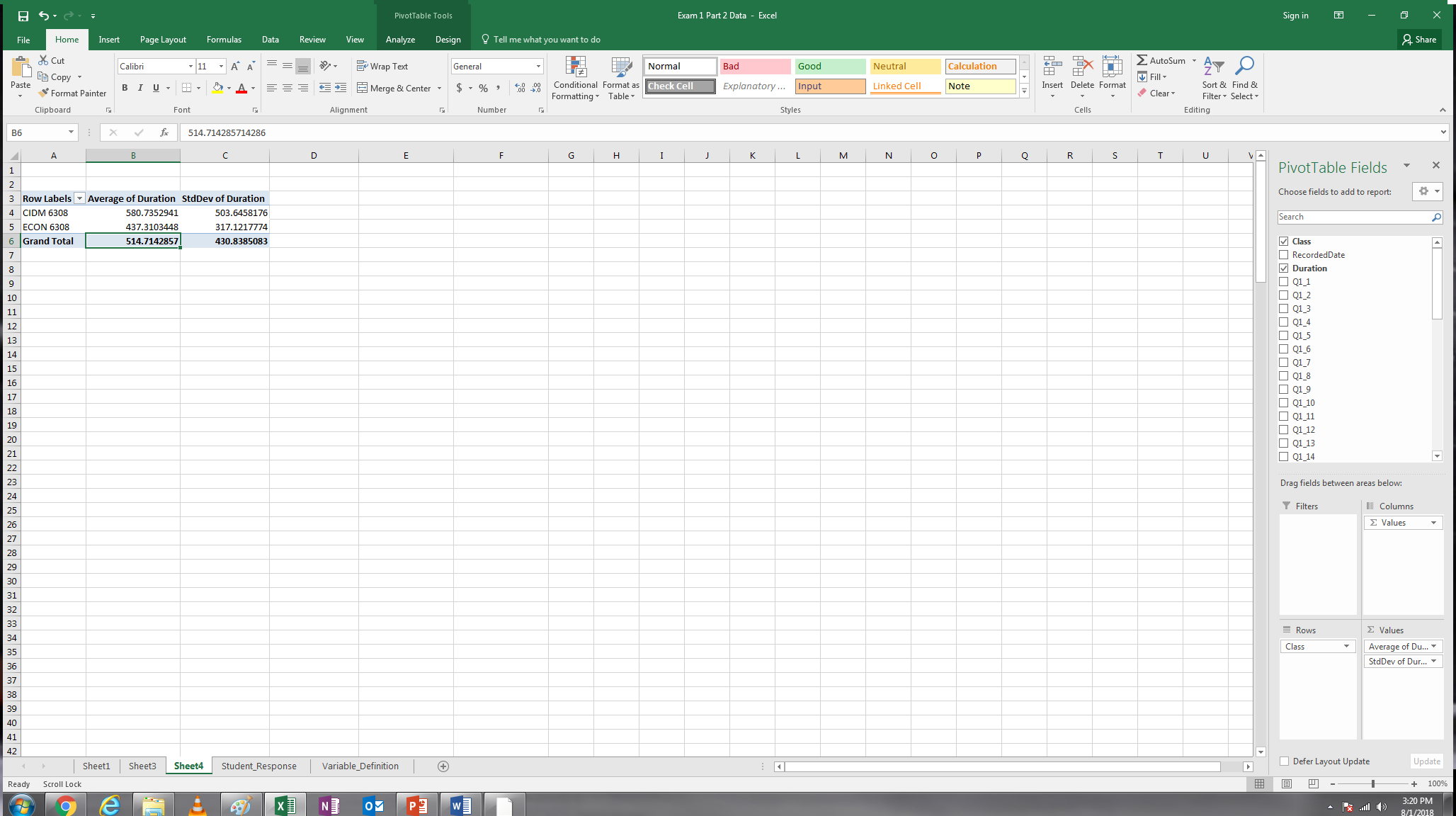
Max: 2916



1.3.2. The mean and sample standard deviation of the duration. Round your answer to a whole number (i.e., integer). (4 points)

Mean: 515

Standard Deviation: 431



1.3.3. Typically, if a participant spent too little or too much time on the survey, their response may be invalid (6 points). Based on the length of the survey, any responses with less than 2.5 minutes (150 seconds) will be counted as invalid. With this criterion, how many responses will be counted as invalid?

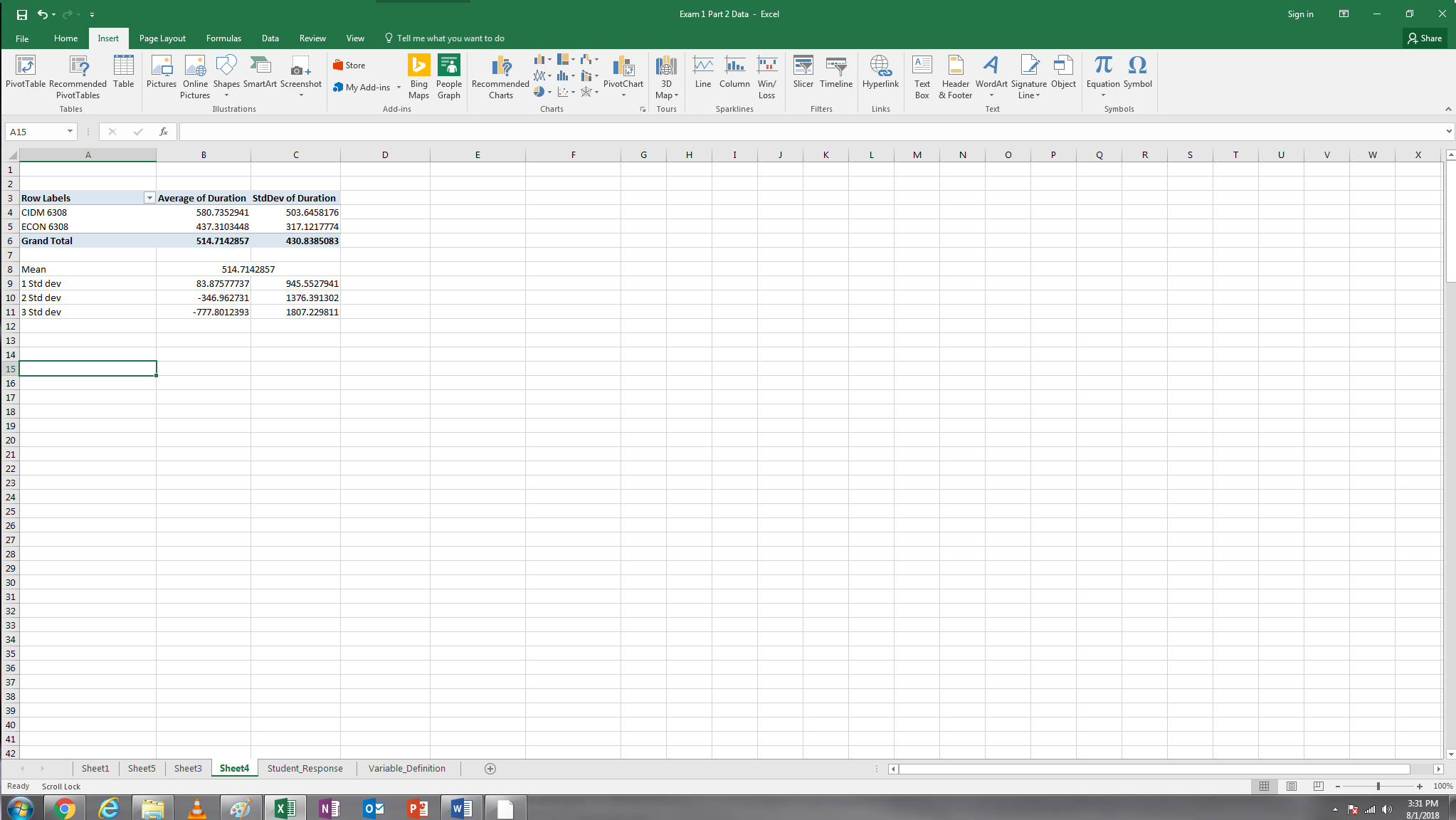
Two

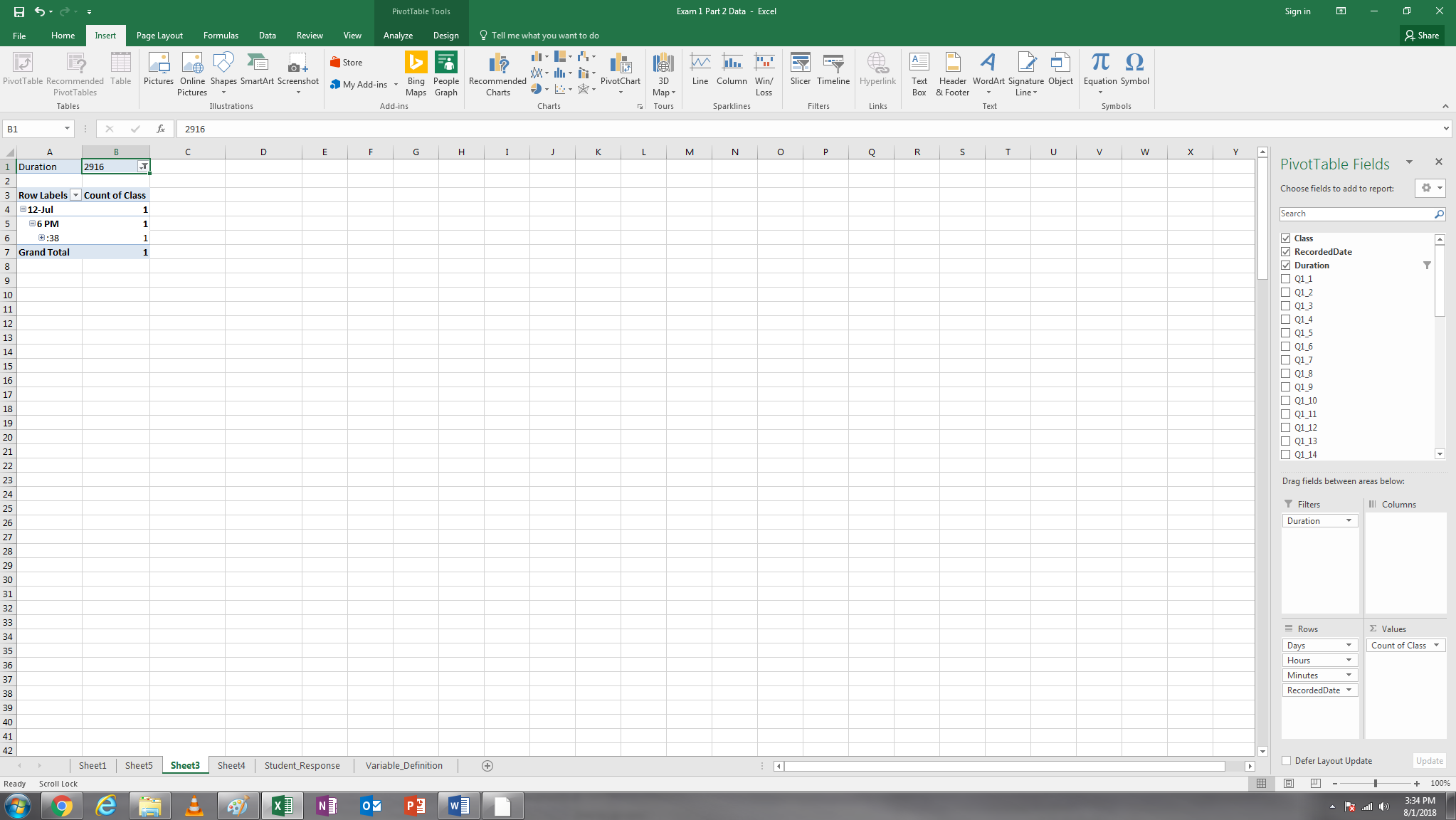
Additionally, Class 05 suggests that data values beyond three standard deviations can be identified as outliers (i.e., invalid responses). Therefore, the responses with duration beyond three standard deviations are also counted as invalid. Using your computations in Q1.3.2, please indicate the number of outliers (beyond three standard deviation) in the dataset in terms of duration. (Hint: you may use the formula mean ± 3 × sample standard deviation to determine the thresholds of outliers and then check how many values in the third column beyond the thresholds).

In total, how many responses will be counted invalid?

Please take a screenshot or picture to show any evidence to support your answers to this question (Screenshot 2).

Threshold calculation shows anything above 1807 to be an outlier. This removes one response bringing the total removed responses counted as invalid to 3.





1.3.4. The rows with invalid responses must be deleted in practice. However, in this exam, **we would rather keep them in our dataset** (just in case that a student did not get the correct answers to Q1.3.3, s/he would miss all the following questions).

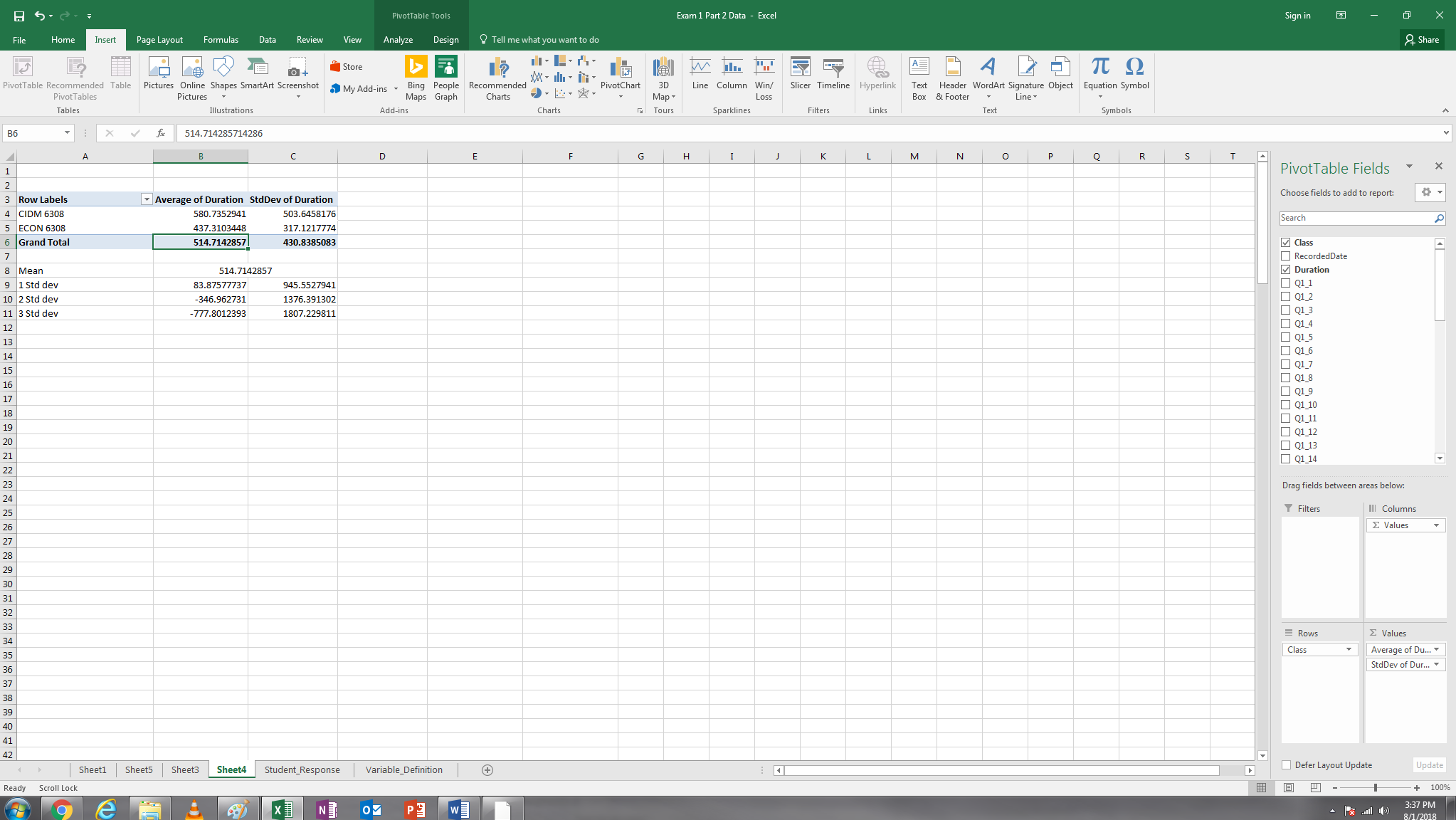
1.3.5. Please compute the average time spent on the survey by students in CIDM 6308 and ECON 6308, respectively (round your answers to integers). (4 points)

Hint: The easiest way is to insert a Pivottable and drag two variables into it as below.

Take a screenshot of your Pivottable or any supporting evidence generated by other tools (Screenshot 3). Your screenshot must directly support your answer.

Average in CIDM 6308: 581

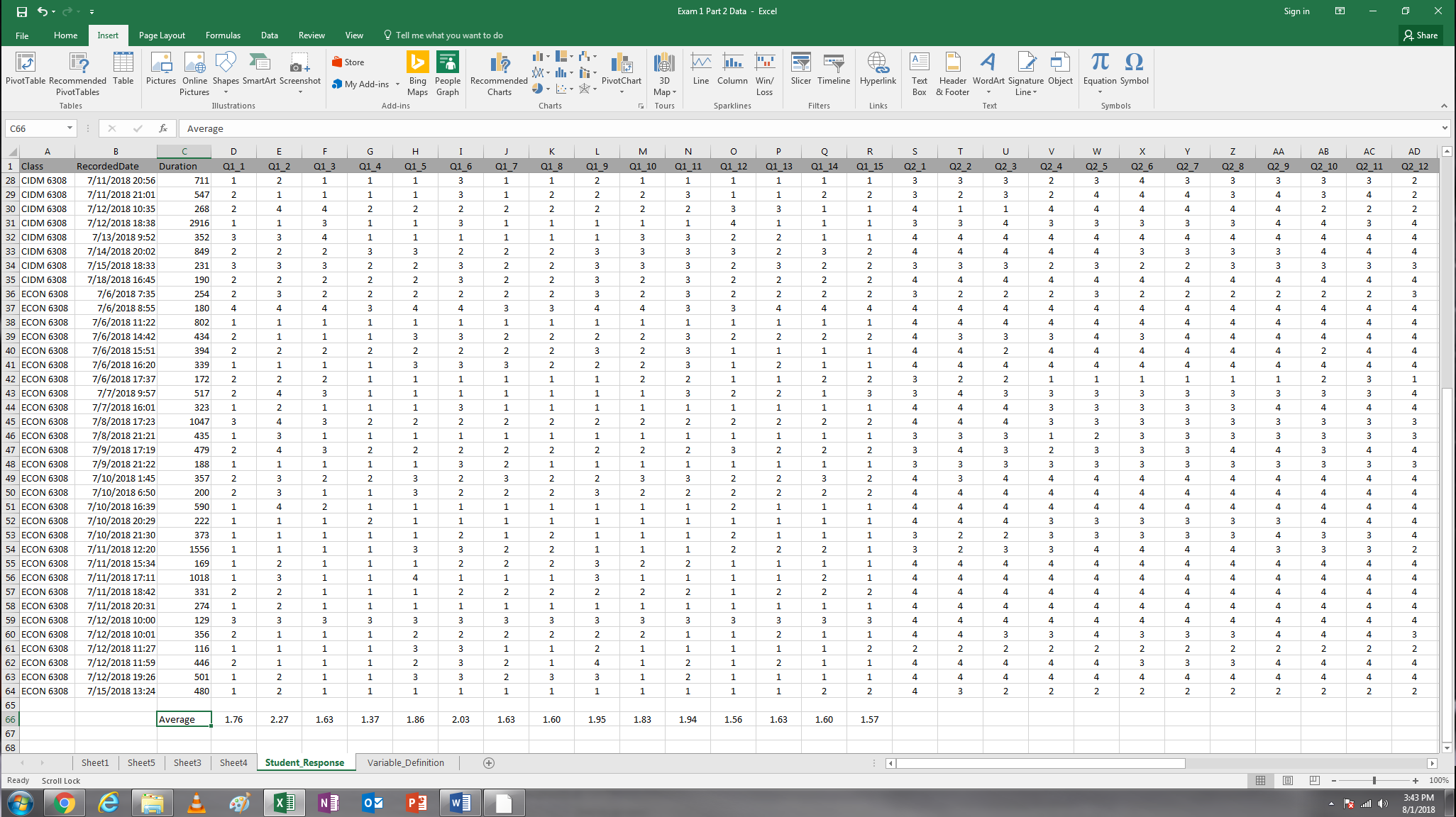
Average in ECON 6308: 437



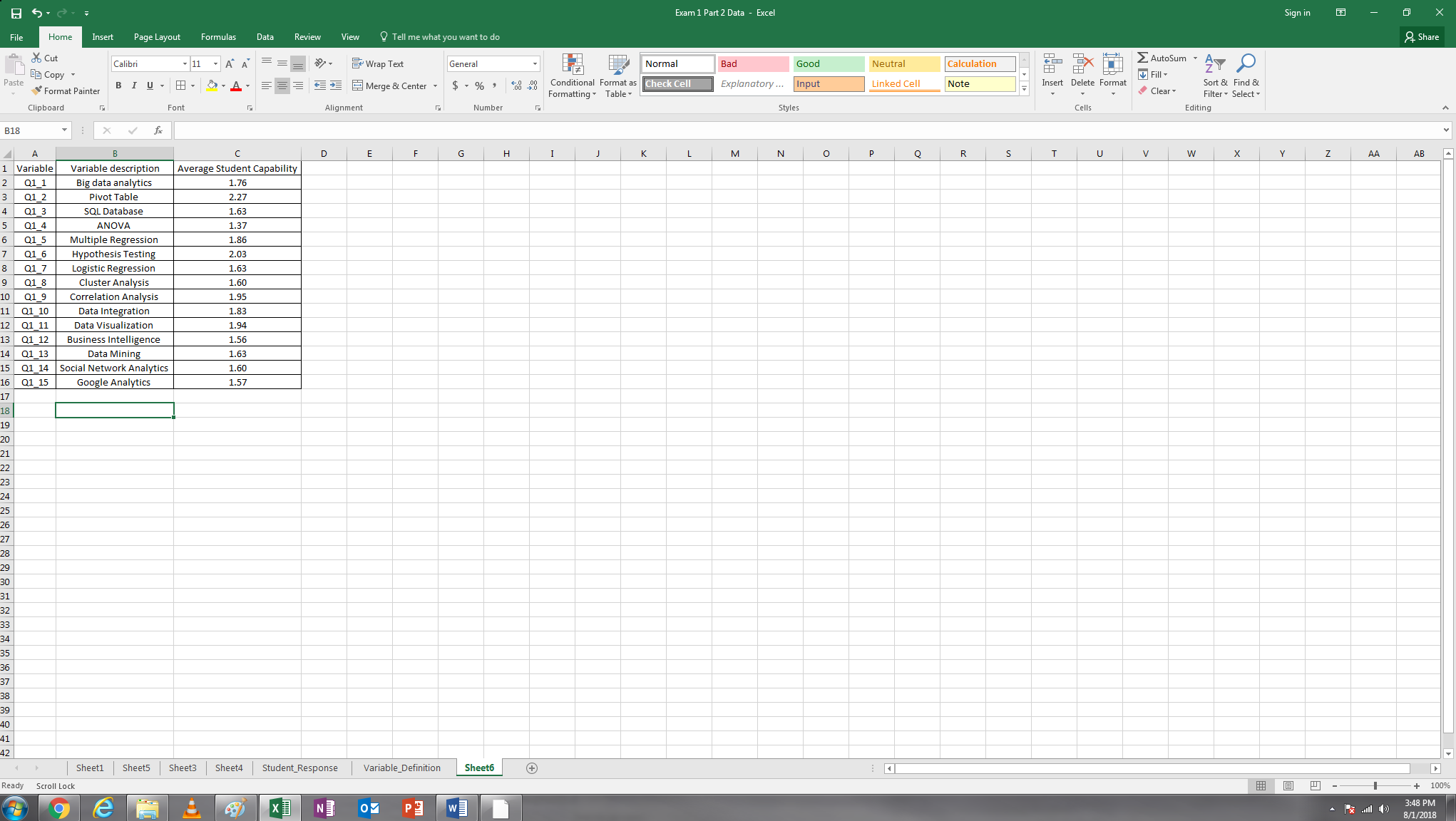
2. Please compute students’ average current capability in each skill (this involves columns Q1\_1 to Q1\_15) and then generate a column or bar chart to show the average capability in each skill. Based on this, please answer a few questions. The following steps provide a detailed instruction which you 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.

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



2.3. Step 3: copy all the averages to another worksheet and match the column ID with each skill



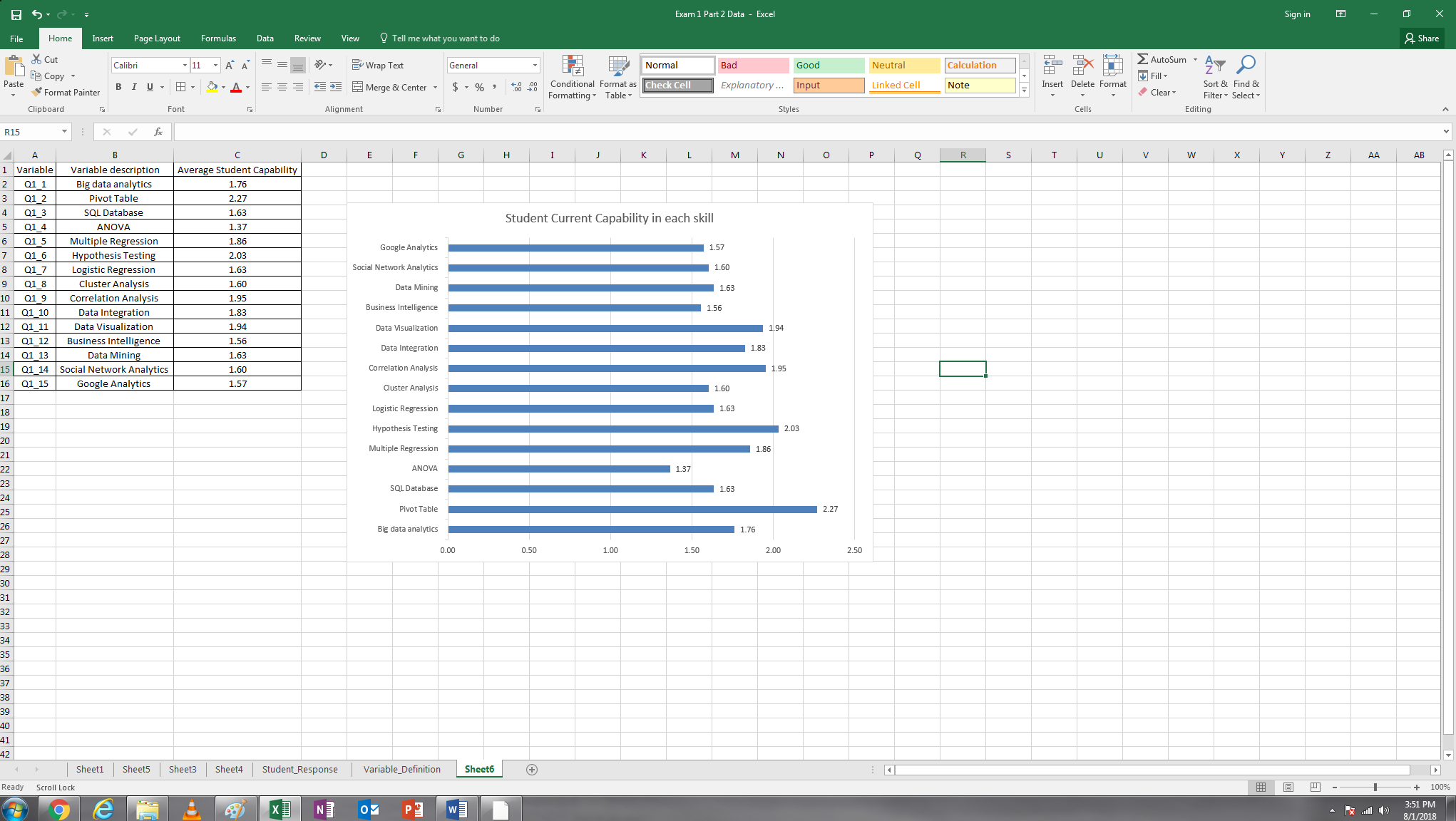
2.4. Step 4: generate a column chart or bar chart (please show the data label in the chart).

2.5. Step 5: observe the chart and then answer the questions. Among the 15 skills, which one our students are most capable in currently and which one our students are least capable in currently? (5 points)

Most capable skill: Pivot Table

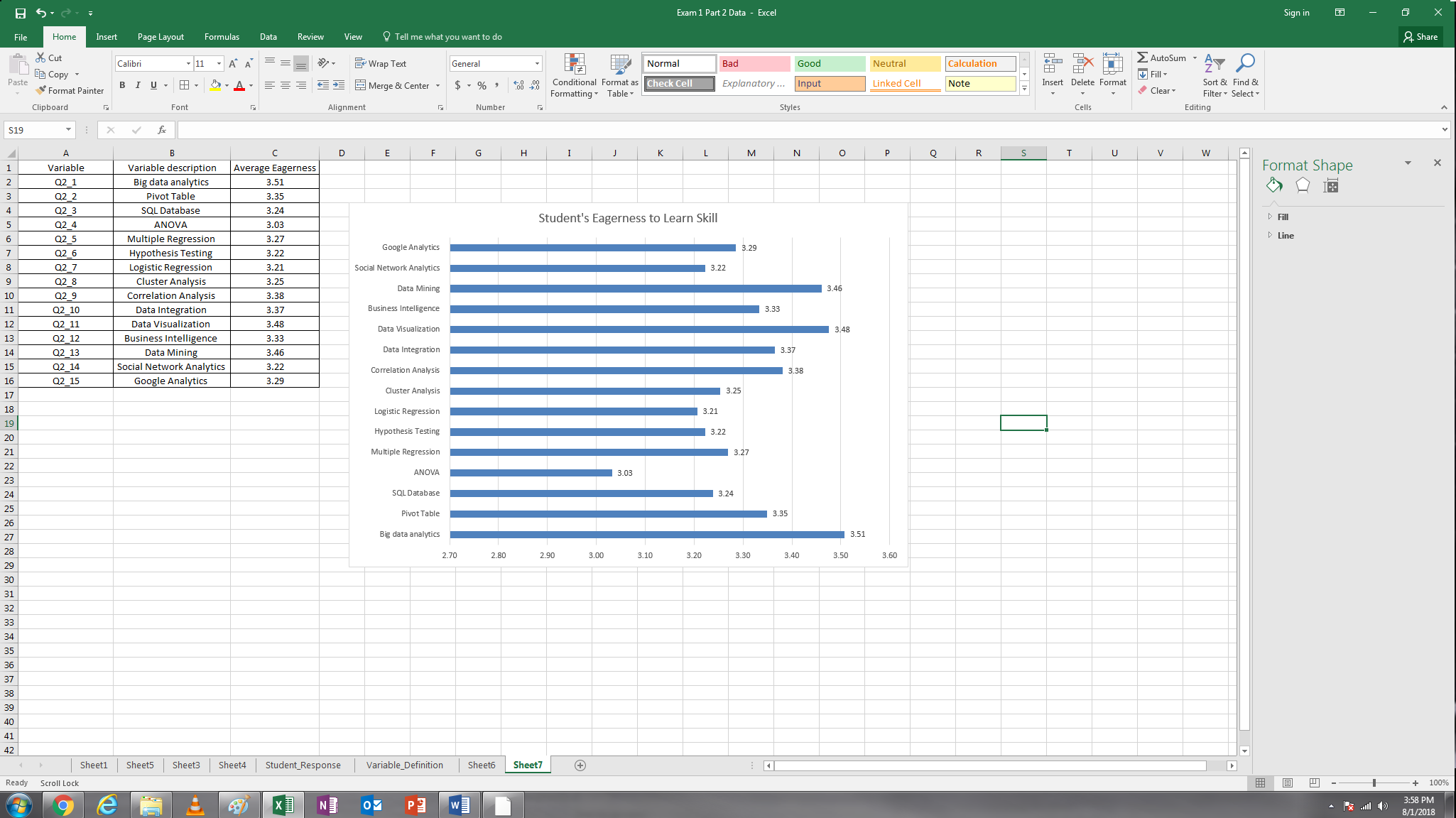
Least capable skill: ANOVA

Take a screenshot of your bar or column chart (Screenshot 4).



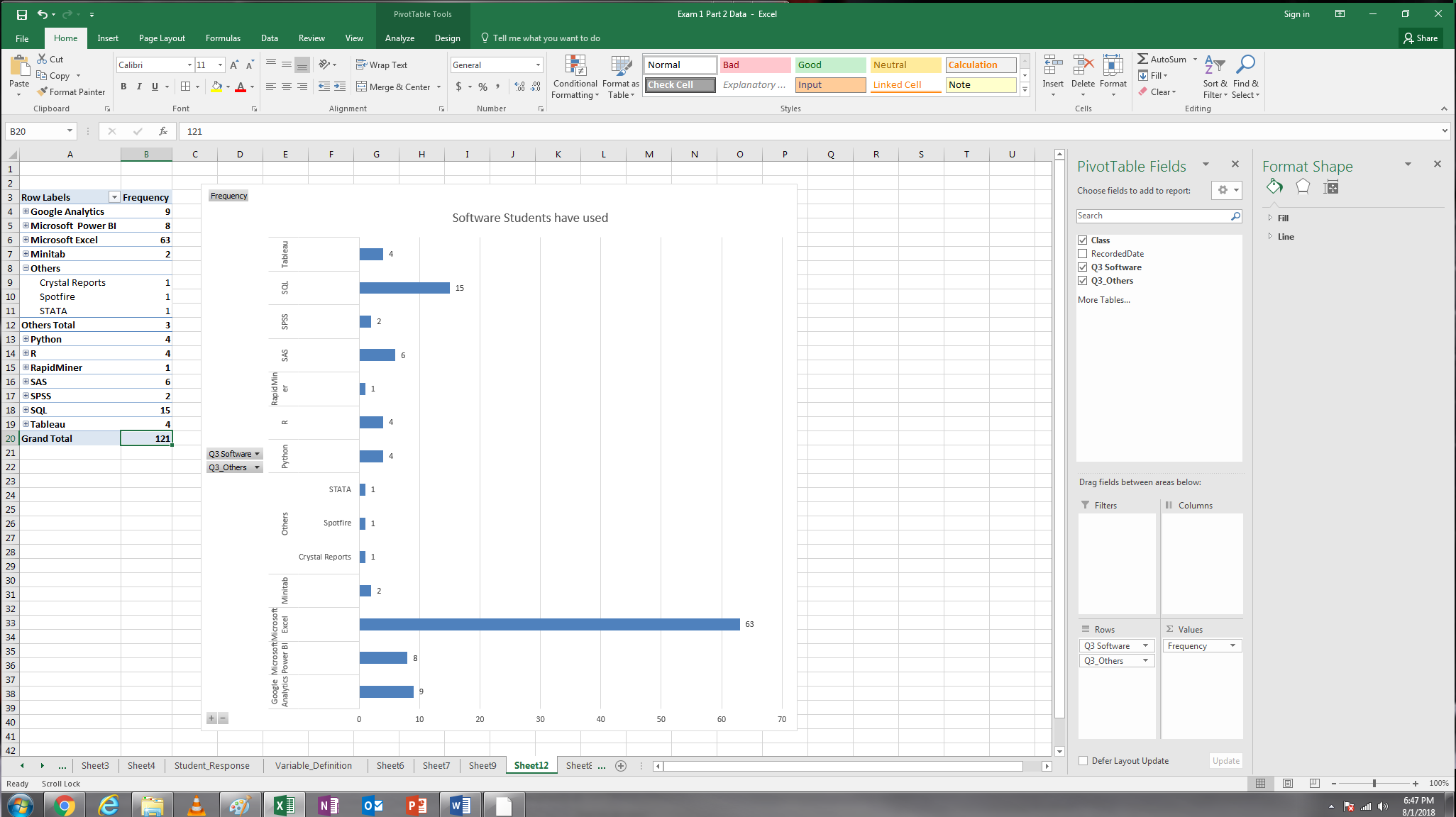
3. Follow the same procedure, please compute students’ average eagerness to learn each skill (this involves columns Q2\_1 to Q2\_15) and then generate a column or bar chart to show the average eagerness in each skill. Based on this, please find the top three skills our students are eager to learn from this course. (5 points) Take a screenshot of your bar or column chart (Screenshot 5).

Top Three skills Students are eager to learn: Big data Analytics (3.51), Data Visualization (3.48), Data Mining (3.46)



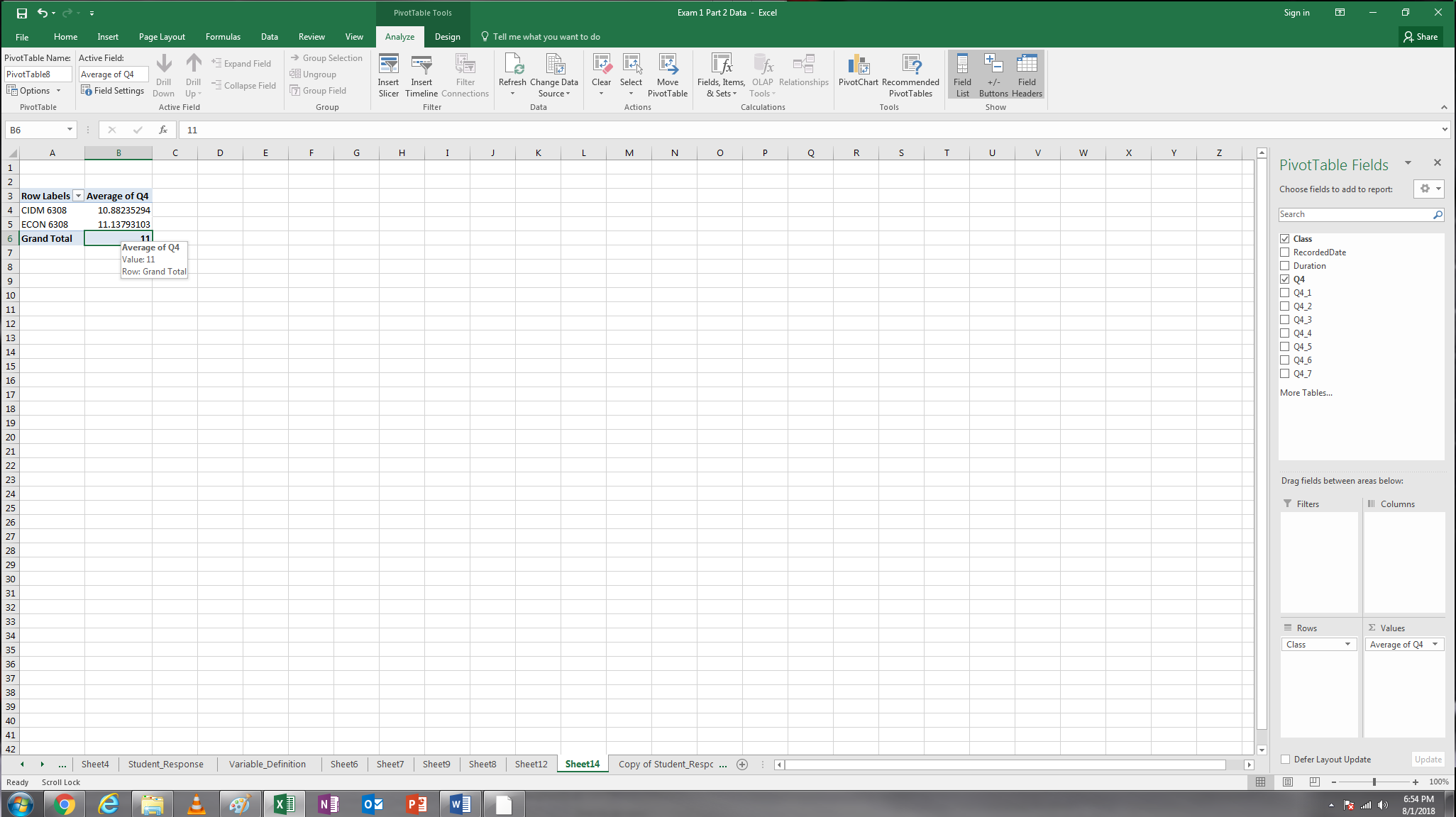
4. Based on Q3 and Q3\_Others in the dataset (i.e., which of the following statistical or analytic software have you used before?), please compute the frequency of each software in our dataset and then organize it using a table as below **or** visualize it using a bar or column chart. Based on this, please find the top three software which most students have used before. Take a screenshot of your table or chart (Screenshot 6). Note: Q3 is a multiple-answer question, so you must count the frequency of each software separately (5 points).

Top three software: Microsoft Excel (63), SQL (15) , Google Analytics (9)



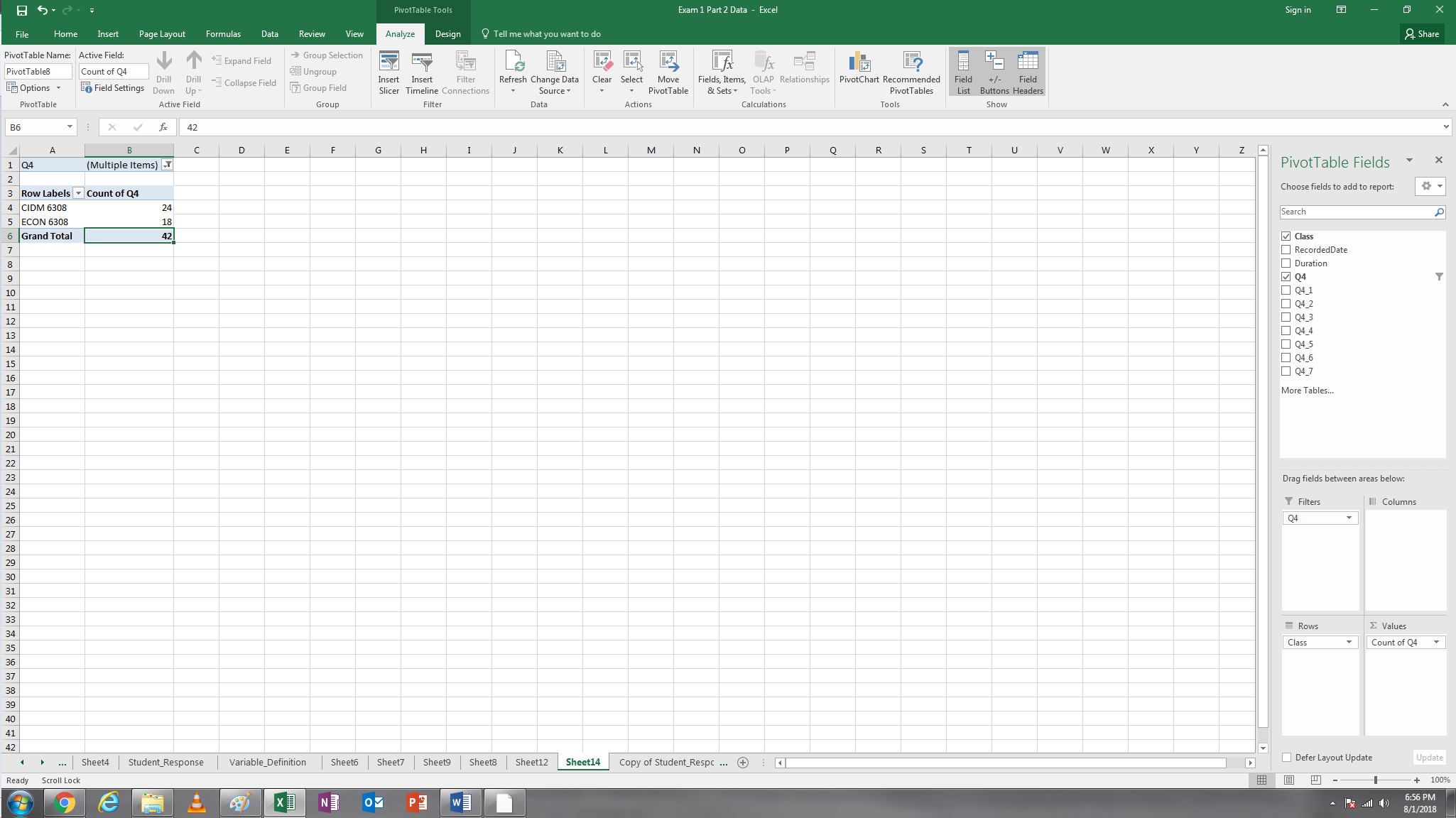
5. Based on Q4 (Column AJ), please compute the average time (in hours) students can spend on this course each week. Round your answers to an integer.

11



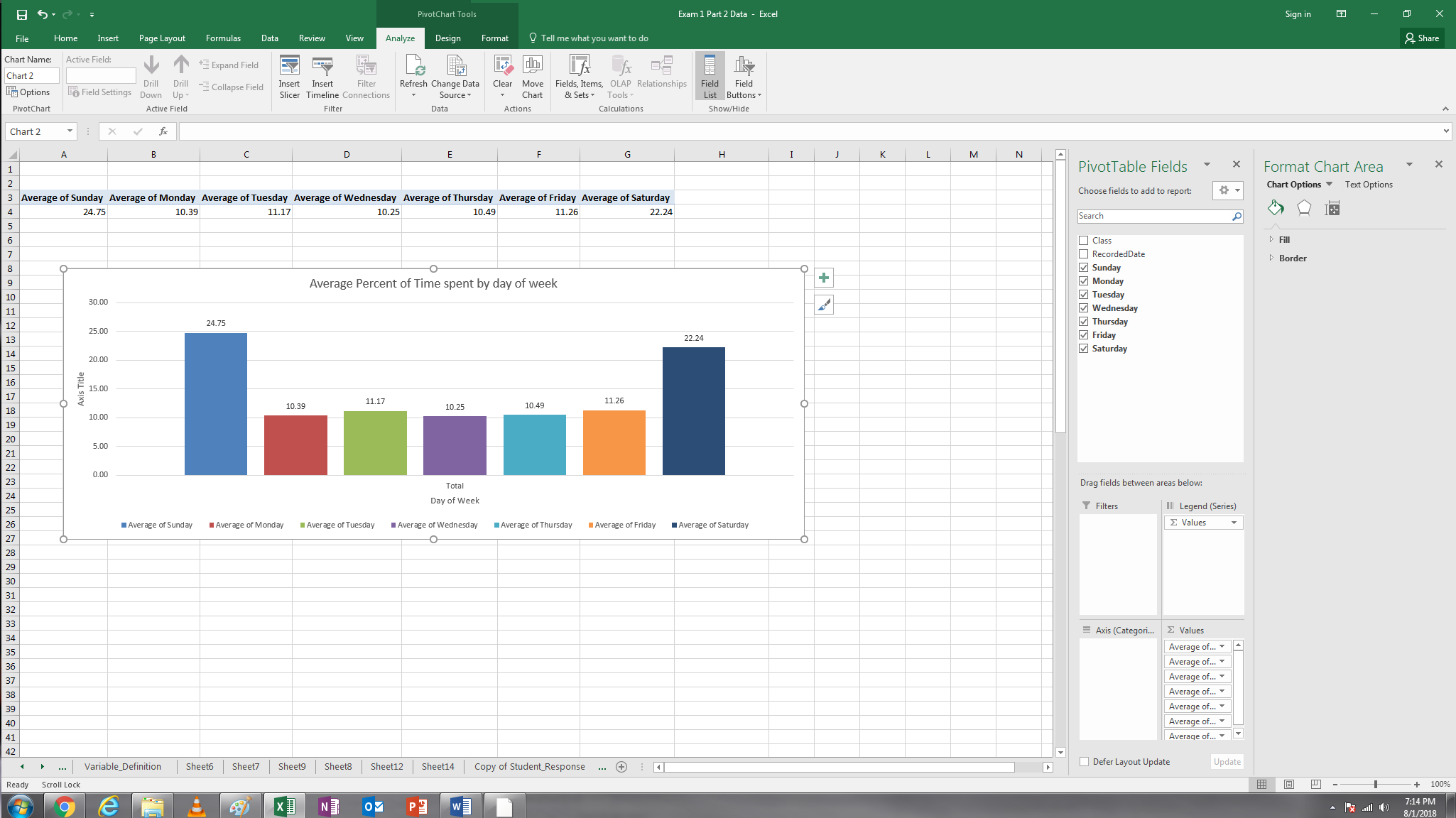
As indicated in our syllabus (page 3), “this is a practice intensive course, so please reserve at least five hours for each class (i.e., 10 hours per week), including online participation, assigned readings, quiz, lab section, and homework”. How many students indicate that they can spend 10 or more hours on this course weekly? (4 points).

42



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 (5 points).

6.1. Please compute the average percentage of time our students spend on this course each day and visualize it using a bar or column chart. Take a screenshot of your chart (Screenshot 7). A sample is provided as below. Note: the sum of the seven average percentages is equal to 100.



6.2. As shown in page 5 of our syllabus, this summer course divide a week into two portions: Sunday to Wednesday, and Wednesday to Saturday. Each portion covers one class. Among the seven days, Wednesday is overlapped in two portions and the instructor’s office hours are also set on that day. Students may adjust their schedule on Wednesdays to make each portion equal. Based on the survey data, please compute the total percentage of time our students can spend on each portion (round your answer to the second decimal place such as 1.11)?

Using the average percentages.

Sunday to Wednesday: 56.56%

Wednesday to Saturday: 54.24%

It is OK the sum of the two portions is over 100 because Wednesday is double-counted. Are they very close? Does the instructor’s scheduling method make sense based on the distribution of students’ time on each day?

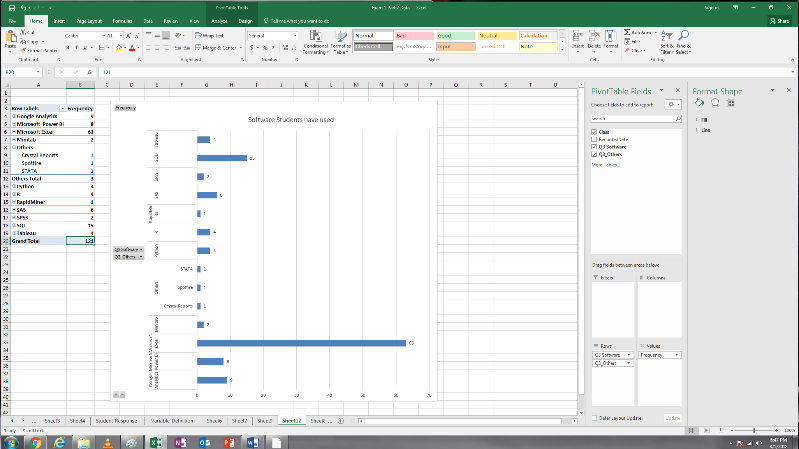
Yes, very comparable. I would say yes the scheduling method can make sense based on this observation however some of the records in the chart showed zero time available on Wednesdays for some students. Also, something to note is the lowest average of time percentage in the chart was Wednesday. This could be reviewed more to find a possible better alternative day that a larger percentage of students are available, or consider breaking the office hours to two sessions.

7. This is a bonus question. If you were Leon, based on the student’s background information, what conclusions or decision would you make? Please provide two conclusions or decisions within 50 – 100 words (5 bonus points). Remember your conclusions or decisions must be based on evidence. Submit your answers with your screenshots.

|  |
| --- |
| My first conclusion is …….. (See screenshot 6).  My third conclusion is ……. (See screenshots 7 and 5) |

Referencing screenshot 6, my first conclusion is every student in each class has experience using Microsoft Excel and therefore has performed some level of data analysis. This program is used across many fields and industries so this is not too surprising and focusing on ways to teach and challenge students on excel skills related to enhancing their ability in data analytics would be key. All of the other programs had far less students with background using them, possibly due to access to such software through their employers, etc. While having knowledge of them all is useful, that is very time intensive and possible not relevant for all parties.

Screenshot 6



Referencing screenshot 5 and 7, my second conclusion is eagerness in the class to learn and understand big data analytics is high, along with many of the other skills. However, students have a limited percentage of time to dedicate for this course due to other time obligations. The weekends compose the largest percentages of time available for students to study or do their homework, etc. I would use these two charts combined to find the most desired skills compared to available time and plan a course strategy to accomplish those objectives.

Screenshot 5 and 7

