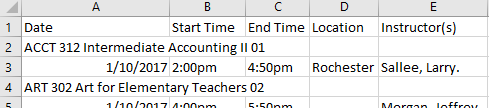
HW: Merging Data Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

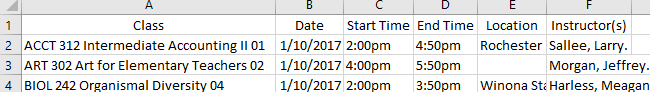
|  |  |
| --- | --- |
| For this homework assignment, we will consider data regarding the cancellation of WSU courses due to weather, i.e. snow days. There are three sources of data for this assignment – all of which can be found on our course website.   * List of Courses Cancelled on Jan 10: [Link](http://statsclass.org/dsci210/HW/WSU_Cancelations_Jan10.html) * List of Courses Cancelled on Jan 25: [Link](http://statsclass.org/dsci210/HW/WSU_Cancelations_Jan25.html) * Auxiliary Course Data: [Link](http://statsclass.org/dsci210/HW/HW2_AuxilaryData_SP17.html) | Image result for google images snowday friday |

Complete the following.

1. Download the Jan 10 data from the course website. Copy the contents of the Class Cancellations HTML table. In Excel, paste the contents of this table as text. Pasting as text will remove the HTML formatting that is needed for a browser, but not needed when working with the data.



Use the =OFFSET() function to rearrange the data into a rectangular format where each row contains a information for a class and the columns identify the Class, Date, Start Time, End Time, Location, and Instructors. (7 pts)



Note: You must accomplish this task using =OFFSET(). You will not receive credit with methods such as Cutting/Pasting, Filtering, etc.

Turn In: A screen shot of the OFFSET() formulas used for each column. You can place a ‘ mark in front of the =OFFSET() function to show the formula in the cell.

1. Download the Jan 25 data onto a second sheet in Excel. Repeat Problem #1 for the Jan 25 data – this too can be downloaded from course website. (4 pts)

Turn in: I don’t need to see another screen shot; however, you should provide a discussion of what you needed to do differently for the Jan 25 data compared to the Jan 10 data.

1. Use the =OFFSET() to append the data from Jan 25 to the bottom of the Jan 10 data. That is, create a single dataset that includes all data. (4 pts)

Turn in: A screen shot of the =OFFSET() formulas used for appending the Jan 25 to the bottom of Jan 10 data.

1. Use various string functions, i.e. =FIND(), =MID() to pull off the Department from each row. Next, use a PivotTable to obtain a table of counts for the number of classes cancelled for each Department. (3 pts)

Turn in: A screenshot of the PivotTable.

1. Use various string functions, i.e. =FIND(), =MID() to identify the Course Level, i.e. 100, 200, 300, or 400, etc. from each row. Next, use a PivotTable to obtain a table of counts for the number of classes cancelled by Course Level. (3 pts)

Turn in: A screenshot of the PivotTable.

1. One data issue in Problem #4 is that there are many more courses offered by some Departments compared to others. Thus, a percentage of courses cancelled (relative the number of courses offered) should be used when making comparison between Departments. Use the Auxiliary Data provided on our course website to compute the appropriate percentages for each Department. Sort the list of Departments from largest to smallest according to the percentage of classes cancelled. Which Department cancelled the largest percentage of their courses? (2 pts)
2. Akin to Problem #6, compute percentage for course level. Which course level had the most cancellations? (2 pts)