Congratulations! You passed!

TO PASS 80% or higher

Keep Learning

GRADE
100%

Week 3 Final Assignment

LATEST SUBMISSION GRADE

100%

1. You have received some sales data from your company's database. The data is in a dirty format and you are required to clean it up in order to perform some analysis on it. Since your company will need to perform this analysis each month with new data, you are to use formulas to clean the data so that the workbook can be easily updated each month with new records from the database. Follow the instructions carefully and answer the questions as you go. **1 / 1 point**

   Download and open the assignment workbook. There are 2 versions of the workbook, depending on which version of Excel you are using. If you use Excel for Windows Office 365, 2016 or 2013, or Excel for Mac Office 365, 2016 or 2011, use the Unicode version:

   C4 W3 Final Assessment Unicode.xlsx

   If you use Excel for Windows 2010 or 2007, use this version:

   C4 W3 Final Assessment.xlsx

   Look at the **Data** worksheet and **Clean** worksheet. Make sure the **Check Sum** value in **Data** cell **B2** is 95858 (90898 on Excel for Mac 2011). Do not insert any rows or columns into the Data sheet during your work or the Check Sums may give different results than programmed.

   Our goal is to populate the columns of the **Clean** worksheet with the data in a form that we can easily analyse. This may look daunting, but we are told by the Data Manager that the raw data always follows a certain pattern. That pattern is:

   1. Some **/** characters
   2. The product type sold, with the prefix **PR:**
   3. Some "_" characters
   4. The Sales Person's name, with the prefix **SP:**
   5. The date of the sales, in the format **yyyymmd**. All sales are from October 2017
   6. Some non-breaking space characters
   7. The amount of sales, with the prefix **$**. All sales amounts are below $10000
   8. Some space characters
   9. A final random character that has an ASCII value between 28 and 31.

   We will clean up the data one item at a time, by using formulas to make copies of the data to the right of column **B**, with each copy being slightly cleaner than the previous. When we are done we will separate the data into 4 components on the **Clean** worksheet tab, and perform some basic analysis.

   Go to the **Data** worksheet. To begin, we want to verify what special characters might be at the end of each data entry. In column **C** (cells **C5:C500**) write a formula using the **RIGHT** and **CODE** functions to return the ASCII code of the final character in each record.

   How many records have the character 30 as the final character? (Hint: After writing your formula in column **C**, you can write a formula in cell **C3** that uses the **COUNTIFS** function to count how many times the value in column **C** is 30.)

   126

   Correct

2. Next, we want to remove some of these final characters with the **CLEAN** function. In column **D**, write a formula that applies the **CLEAN** function to the column **B** data. To check if this has been done correctly, enter the value of the Question 2 checksum from cell **D2**. **1 / 1 point**

   Enter just the 5 digits (6 on a Mac), without any commas or other thousands separator.

   99221

   Correct
   Yes, that the correct checksum if you are using Windows.

3. Now we want to verify that the **CLEAN** function did, in fact, remove all of the unwanted final characters. In column **E**, write a formula just like column **C** that returns the ASCII code of the final character from each record in column **D**. How many records have the character 32 as the final character? **1 / 1 point**

   496

Correct

4. Our next plan is to get rid of the non-breaking spaces between the dates and the sales amounts by replacing those characters with regular spaces. Which function can we use to change all of the instances of one character from a text string into a different character? (Type only the function name) **1 / 1 point**

SUBSTITUTE

Correct

5. To use the **SUBSTITUTE** function to replace non-breaking spaces with regular spaces, we need a way to refer to both of these character types in the arguments of the **SUBSTITUTE** function. For the space character, we can easily do this by typing a space in between quotation marks like this: **" "**. For the non-breaking space character, we can use the **CHAR** function to return a non-breaking space. If you are using the Unicode workbook, then you can use the **UNICHAR** function instead. **1 / 1 point**

What number is used as the argument of the **CHAR** function to return a non-breaking space?

160

Correct
Yes, that is the non-breaking space character on Windows.

6. In column **F**, use the **SUBSTITUTE** function (with a **CHAR** function in the second argument) to change all of the non-breaking spaces in column **D** into regular spaces. What is the value of the Check Sum in cell **F2** after you do this? **1 / 1 point**

92309

Correct

7. We could use the **TRIM** function now to get rid of the trailing spaces and the duplicate spaces between the date and price, but think ahead. We still have some "**/**" characters at the start of each record that we don't want, and we have some "_" characters between the product and sales person that we don't want. It makes sense to wait until after we remove these characters before we use the **TRIM** function, otherwise we may end up just needing to use it twice. Let's focus on the "**/**" characters at the start of each record next. **1 / 1 point**

In column **G**, use the **SUBSTITUTE** function to remove all of the "**/**" characters from the records in column **F**. However this time, instead of replacing the "**/**" characters with a space, replace them with nothing at all, so that each record will begin with **PR:** once you are done. (HINT: remember we can refer to blank text by typing two sets of quotation marks with nothing in between.)

What is the value of the Check Sum in cell **G2** after you do this?

82774

Correct

8. Now in column **H**, replace all of the "_" characters in our column **G** data with spaces, and then we will be ready to use the **TRIM** function. The reason we want to replace "_" with spaces instead of blank text is because having a space character between the Product data and the Sales Person data will make it much easier for us to extract those individual bits of data than if there was no space between them, as we will see later. **1 / 1 point**

Perform this replacement and provide the value of the Check Sum from cell **H2**.

66457

Correct

9. In column **I**, apply the **TRIM** function to our data. **1 / 1 point**

When you are done, instead of using the Check Sum to see if this has been done correctly, we're going to use the **LEN** function. In cells **J5:J500**, use the **LEN** function to find the length of each now-trimmed record from column **I**. Cell **J2** contains a formula that is summing all of these lengths. What is the value in cell **J2**?

15186

Correct

Correct

10. In cell **K5**, write a formula to find the ASCII code of the final character of each record and apply it to cells **K5:K500**. If we look at the results we will see that all values range between 48 and 57, which correspond to the digits 0 through 9. That's good news, since we expect every record to be ending with the sales amount at this point. The digits 0 through 9 for the final character is exactly what we would expect. Performing checks like this as we go can be helpful to make sure we are on the right track. **1 / 1 point**

    Now it is time to get rid of the of the prefixes in front of the product names and the Sales Person names. In column **L**, write a formula that refers to our records in column **I** but substitutes the text **PR:** for blank text (remember the double quotation marks). When you are done, enter the value of the Check Sum from cell **L2**.

    81796

    Correct

11. In column **M**, further clean the data by substituting **SP:** with blank text. After you do, what is the value of the check sum in cell **M2**? **1 / 1 point**

    70357

    Correct

12. Now we need to insert a space between the end of the Sales Person name and the start of the Date. Notice how every date begins with the text **201710**. So we can substitute "**201710**" with " **201710**" which will place a space in front of the start of the date. Now this particular formula would need to be updated each month, since if we were to use our workbook for data from November 2017, the dates would begin with **201711**. But that is a problem for another day. For now, go ahead and perform this substitution in column **N**. Finally, note that we don't just want to substitute **2** or **201** because that might alter some of the sales amounts which could contain those numbers. But since we know that all sales amounts are less than $10000, substituting the string **201710** will leave the sales amounts untouched. **1 / 1 point**

    If you have done this correctly, you should have the Check Sum value 65831 in cell **N2**.

    We now have clean data in column **N**. The only problem is, each record is still a continuous text string. Our next challenge is to separate the 4 pieces of information from each individual record (**Product**, **Sales person**, **Date** and **Sales Amount**). To do this, we are going to use the **LEFT** and **MID** functions to extract the desired pieces from each record, and the **FIND** function to help us enter the arguments for the **LEFT** and **MID** functions. We didn't include the **FIND** function in the videos. We will use it to locate where a specified text string (in our case, a single space character) occurs within another text string (the complete clean record).

    In cell **O5**, write the following formula and then apply it to cells **O5:O500**.

    =FIND(" ",N5,1). The 1 at the end of this formula is an optional argument, and it tells the **FIND** function to begin searching for the text string " " starting at the 1st character within **N5**. When you are done, write down the value in cell **O2**.

    3348

    Correct

13. In cell **P5**, write a formula to find the location of the second space character within column **N**, and apply it to cells **P5:P500**. Hint: you can use the location of the first space that we found earlier as the place to start our search. You will need to add 1 to the previous location so that it doesn't find the same space again. **1 / 1 point**

    What is the value in cell **P2**?

    6448

    Correct

14. There is one more space character we need the location of. The third space character in the string is between the date and the Sales Amount. In cell **Q5** we can write **=FIND(" ",N5,P5+1)**, or we can write **=FIND("$",N5,1)-1**. They should give equivalent results. Do you see why? The second of these formulas is looking for the location of the first dollar sign, and then subtracting one from that value. Since our data is structured in a way that the third space character is immediately before the **$** character, either formula should work. Apply one of these formulas in cells **Q5:Q500**, and then enter the value from cell **Q2**. **1 / 1 point**

    10768

    Correct

15. In column **R** we are going to extract the product names. Since they are at the beginning of the record, we can use the **LEFT** **1 / 1 point**

    function. The number of characters we want returned by the **LEFT** function will be one less than the location of the first

function. The number of characters we want returned by the **LEFT** function will be one less than the location of the first space. Column **O** tells us the location of the first space. Write a formula in cell **R5** to extract the product names and apply it to cells **R5:R500**.

What is the value in cell **R2**?

37696

Correct

16. In column **S** we will extract the Sales Person. For this, we need to use the **MID** function, and the location of our space characters to determine: 1 / 1 point

    1. the start point of extraction, and
    2. how many characters to extract.

    Write a formula in **S5** to extract the Sales Person and apply this down column **S**. What is the value in cell **S2**?

    35588

    Correct

17. In column **T** we will extract the date, as text. We can convert it into a value later. Using the same type of formula construction as we did in column **S**, but with references to the second and third space locations instead of the first and second space locations, write a formula in cells **T5:T500** that will return the date as a text string. There is no need to use the **TEXT** function at this point. When you are done, enter the value from cell **T2**. 1 / 1 point

    28624

    Correct

18. We're almost done! In column **U**, we want the Sales Amount, without the dollar sign. This formula is actually more simple since we do not need to worry about getting precisely the correct value for the length of the string to extract. If we enter a large number in the third argument of the **MID** function, say 99, and there are fewer than 99 characters remaining in the string, then the function will just return what is remaining in the string. This is useful when we want to extract text from the end of a string. Also, since we do not want the $ character returned, we will start our extraction 2 characters after the location of the final space. We would also like the Sales Amount to be returned as a numeric value, rather than a text string. Therefore we will also wrap our **MID** function inside a **VALUE** function. 1 / 1 point

    In cell **U5**, write a formula to extract the Sales Amount as a numeric value and apply this down column U. What is the value of the sum in cell **U2**?

    441522

    Correct

19. The final thing we will do on this sheet is to convert the dates in column **T** from text strings to date values recognised by Excel in column **V**. We can do this with the formula **=DATE(2017,10,??)** where the **??** represent the day of the month, which will be either the 7th, or the 7th and 8th characters from column **T**. Use a **MID** function to extract these day characters from column T, and embed that **MID** function within the **DATE** function in place of the **??**. So, the formula at column **V5** will look like **=DATE(2017,10,MID(...))** 1 / 1 point

    You will need to write appropriate arguments inside the **MID** function in place of "**...**". When you are done, enter the value from the Check Sum at cell **V2**.

    We have now finished cleaning our data! If you navigate over to the **Clean** worksheet tab, you should see the columns **B** to **E** filled in, ready for analysis to be performed. Save your work. Well done for completing the assignment for this week.

    28272

    Correct