DSCI 210: Data Science – Excel Take-home Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
Spring 2016
Points:

Consider the Baby Names dataset on the course webpage. This file contains the unique names of over 600 babies born in Rochester MN over the course of a year. There were substantially more babies born during this time period – this is just a list of the unique names. For example, there was likely several baby girls named Abigail born over this time period.



1. Use the LEFT() and RIGHT() function in Excel to pull off the leftmost and rightmost character for each name.
	1. Write an IF() statement to identify whether or not the leftmost and rightmost character match. (2 pts)



* 1. Using the PivotTable feature, identify the percentage of names in this list that start and end with the same character. (2 pts)
1. Next, use the MATCH() function in Excel to determine whether or not the first letter in their name is a vowel. Do the same for the last letter in their name. Use an IF() statement to identify names who first and last letter are vowels. (3 pts)



1. Consider my output for attributes Leftmost Vowel?, Rightmost Vowel?, and Both Vowels?



* 1. What does the value in cell F2 mean? How about cell G5? Discuss. (2 pts)
	2. I used the following function in cell H5 to identify a match, =IF(F5 = G5, “Yes”, “No” ). My function returned a “No” in cell H5, but clearly the first letter (A) and the last letter (e), are both vowels. Why did my function not work? Discuss (2 pts)
	3. Fix the formula in column H to correctly identify matches. A Google search suggests that you could use the =ISNUMBER() function somehow. I’m sure other approaches will work as well. I don’t care which approach you use, just fix the formula. (3 pts)
	4. Use a PivotTable again to identify the percentage of names in this list that start and end with a vowel. (2 pts)
1. In the English language, a palindrome is a word that is spelled exactly the same forward and backward. For example, the name ANNA is a palindrome.



Your task for this problem is to create two different procedures by which each name is to be verified as a palindrome or not. The first procedures (part a) will be via brut force, and the second procedure (part b) will be using the new function called Reverse.

1. Ad
2. For this part, we will create a custom formula in Excel. This custom formula will be called Reverse() and uses the StrReverse function provided in visual basic. Use the following process for creating this custom function in Excel.

|  |  |  |
| --- | --- | --- |
| The Excel Visual Basic Editor can be obtained using Alt + F11. This editor can also be found on the Developer ribbon – which may have to be added to the list of visible ribbons.

|  |  |
| --- | --- |
| Insert > Module will provide anew module window | Creating a custom formula named =Reverse()Function Reverse(str As String) As String Reverse = StrReverse(Trim(str))End Function |

Save your Excel file as a macro-enabled file. This is required for your new function to work. You should be able to use your new function. A simple application is shown here. |

1. Use the MID() function in Excel to pull apart each letter of Name into its own column. Now, as mentioned in class repeatedly each time a formula is entered into Excel, we create an opportunity to make a mistake in our work. To alleviate this issue, you can only enter the formula once. I entered my formula into Cell D2. You can copy across and down to obtain all necessary output; however, your formula can only be entered once. You will have some cells that are blank as not all names are the same length. (5 pts)



For this problem, we will consider the ability of a voicemail system to successfully translate a message into text. The data for this investigation are provided in the VoiceMail file on our course website.



Descriptions of attributed in this data file

* Message ID: Represents an ID for the 6 messages under investigation.
* School Reach Message: An indicator of whether or not the message is from School Reach. School Reach is a system that lets parents know of school cancellations.
* Human translation of voicemail to text: I listened to each message and wrote out the contents of each message
* Computer translation of voicemail to text: This is the text provided by the voicemail system regarding the contents of each message. For convenience, I separated each word into its own column.

The goal here is to develop a measurement or metric that can be used to measure the effectiveness of my voicemail system to correctly translate a voicemail message to text. You should develop your own metric. I provide some guidance here as to the makeup of a potential metric.

* The FIND() or SEARCH() functions in Excel can be used to identify how many words match between the two messages. FIND() is case-sensitive, so I’d encourage SEARCH() instead.
* Your metric should take into consideration the number of words in the message as 20 matches out of 25 total words is very different than 20 matches in a 150 word message.
* For simplicity, you need not worry about the order in which the words appear.
1. Discuss in detail the metric you created to measure the effectiveness of my voicemail system to translate a voicemail message to text. (5 pts)
2. Use the Voice Mail data provided to calculate your metric for each message. (3 pts)
3. Use your metric to provide (3 pts)
	1. An overall quality value for all 6 messages
	2. A quality value for the School Reach messages and a quality value for the non School Reach messages.
4. Consider the following question.

Question: Is there evidence to suggest that my voicemail system does a better job at the beginning of a message compared against the end of a message?

Use your metric to provide an answer to this question. You should carefully explain what you did to obtain your values to answer this question. (3 pts)