DSIC 210: HW #3 Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
Spring 2016  
Points: 35

For this assignment, we will consider the OpenBeer database. (Original Source: OpenBeerDB.com). The files in this database have been cleaned up and have been provided on our course website. There are five tables/datasets I

[beers.csv](http://course1.winona.edu/cmalone/dsci210/datasets/OpenBeerDB/beers.csv) | [breweries.csv](http://course1.winona.edu/cmalone/dsci210/datasets/OpenBeerDB/breweries.csv) | [categories.csv](http://course1.winona.edu/cmalone/dsci210/datasets/OpenBeerDB/categories.csv) | [geocodes.csv](http://course1.winona.edu/cmalone/dsci210/datasets/OpenBeerDB/geocodes.csv) | [styles.csv](http://course1.winona.edu/cmalone/dsci210/datasets/OpenBeerDB/styles.csv)

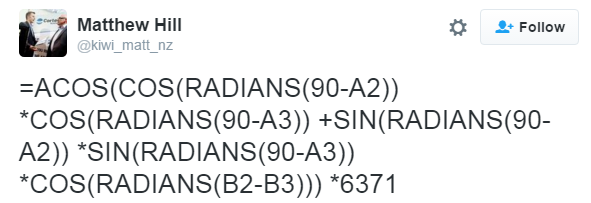
1. The beers table is a listing of all beers in the database. Starting with this table, use the INDEX() and MATCH() functions to merge the following into the beer table. (10 pts)

* Brewery Name
* Brewery City
* Brewery State
* Brewery Latitude
* Brewery Longitude
* Beer Style

Note: If information about a beer or brewery is missing, leave the cell contents blank. That is, you should fix any/all #N/A issues. The ISERROR() will be useful here.

*Turn in*: Save a copy of your complete/cleaned dataset in the Class Storage Directory (T Drive).

1. I found the following formula for computing the distance between two GPS locations in Excel from a tweet by Matthew Hill.

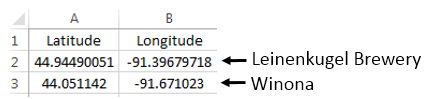


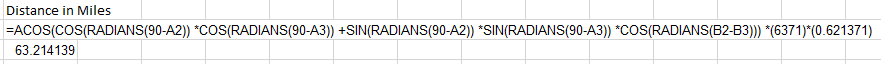
1. Find two additional sources from the internet that will collaborate that this is the correct formula. Provide links to these sources here. (2 pts)

Additional Source #1:

Additional Source #2:

Consider the following example that calculates the distance between Leinenkugel Brewery and Winona.





1. I had to slightly modify the formula provided by Matthew Hill’s tweet. I needed to multiple an additional constant of 0.621371. What is the purpose of this constant? Discuss. (2 pts)

An alternative to the formula provided above is to create your own Excel function. The following Visual Basic code will create a new function called getDistance() which can be used to directly compute distance between two GPS locations.

Step 1: Open the Visual Basic Editor.

PC: <http://www.techonthenet.com/excel/macros/visual_basic_editor2013.php>

MAC: <http://www.techonthenet.com/excel/macros/visual_basic_editor2011.php>

Step 2: Select Insert > Module. Paste the function code into this module.

Step 3: Select File > Save. You must **save** your file as an **Excel Macro-Enabled Workbook**.

getDistance() function code was corrected on Wednesday, 2/24, please use this updated code

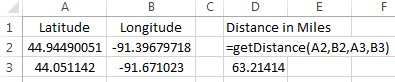
|  |
| --- |
| getDistance() function to compute distance between two GPS locations |
| Option Explicit  Dim earth\_radius, Pi, deg2rad, dLat, dLon, a, c, d  Public Function getDistance(latitude1, longitude1, latitude2, longitude2)    ' Use this for Km  'earth\_radius = 6371  'Use this for Miles  earth\_radius = 6371 \* 0.621371  Pi = 3.14159265  deg2rad = Pi / 180    dLat = deg2rad \* (latitude2 - latitude1)  dLon = deg2rad \* (longitude2 - longitude1)    a = Sin(dLat / 2) \* Sin(dLat / 2) + Cos(deg2rad \* latitude1) \* Cos(deg2rad \* latitude2) \* Sin(dLon / 2) \* Sin(dLon / 2)  c = 2 \* WorksheetFunction.Asin(Sqr(a))    d = earth\_radius \* c    getDistance = d    End Function |

Use the following to return to your workbook.

PC: Select File > Close and Return to Microsoft Excel

MAC: Select Excel > Close and Return to Excel.

Once you return to your workbook, you should now be able to use your custom function. The following is using the =getDistance() function to obtain the distance between two GPS points.



1. Use the process outlined above to create your own getDistance() function in Excel. You must save your working file as a Macro-Enabled File so that your function gets properly saved. -- the standard excel file format does not save visual basic code. Put this macro-enabled excel file on the class storage directory. (5 pts)

Now, use your getDistance() to compute distances between all breweries and Winona. If you cannot get the getDistance() function to work properly, simply use the code provided in Problem #2.

1. A couple of my favorite beers are Leinenkugel's Sunset Wheat and Leinenkugel's Honey Weiss. Find the 10 closest breweries to Winona that have a beer whose name contains the word “Wheat.” Provide the list of breweries in the table below. (3 pts)

|  |  |
| --- | --- |
| 10 Closest Breweries to Winona who have a Wheat Beer | |
| 1 (Closest) |  |
| 2 |  |
| 3 |  |
| 4 |  |
| 5 |  |
| 6 |  |
| 7 |  |
| 8 |  |
| 9 |  |
| 10 (Furthest of top 10) |  |

1. Next, suppose I want to drive to all these breweries using the shortest route possible. The SpeedyRoute website (Link: <https://www.speedyroute.com/> ) will map out the shortest driving path through multiple destinations. Provide a map for the shortest route to all 10 of these breweries. Include a screen shot of this map. (4 pts)
2. Leinenkugel's Sunset Wheat is considered a Belgian-Style White beer (Style\_ID = 67). The following is a list of the 5 closest beers (from Winona) in this category. I have excluded the beers whose Brewery ID = 100. Figure out why these beers should indeed be excluded from this list? Discuss. (3 pts)



1. Leinenkugel's Honey Weiss is considered an American-Style Lager (Style = 95). Provide a list of all beers that are of this style that are within 100 miles of Winona. You will have to add more rows to this table for this problem. (3 pts)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Beer ID | Beer Name | Brewery Name | City / State | Distance |
|  |  |  |  |  |
|  |  |  |  |  |

1. The internal structure of the OpenBeer database can be reviewed from some of their online resources. When this database was setup, the default value for abv (alcohol by volume) is 0. Thus, when a beer is entered into the database and the value of abv is not known, the abv will be set to 0. This is bad database/table design as one of the possible styles is non-alcoholic which would also have an abv value of 0.

Consider the following summaries of the average abv by Style -- sorted from highest to lowest. From this table, a Belgian-Style Quadrupel style beer has the highest alcoholic content by volume in this database. However, these averages are incorrect because 0 is the default value for missing values and thus zeros are being used to compute these averages.

|  |  |
| --- | --- |
| PivotTable Setup for initial abv analysis | PivotTable sorted by average abv |
|  |  |

Place abv field in the FILTERS box on the PivotTable setup. Now, use this Filter to figure out how to remove the zeros from the calculations for average abv. Screen shot the updated table of the first few styles of beer that have the highest average abv. (3 pts)