

## Assignment 3

Your task is based on scraping the following searches:

The first screenshot shows a Yelp search for "mexican food" in "Washington, DC". The results list "Ad Agua 301" with a 4.5-star rating, 296 reviews, and a price range of \$\$ (Mexican). The address is 301 Water St SE, Washington, DC 20003, and the phone number is (202) 484-0301. The second screenshot shows a Yelp search for "chinese food" in "Washington, DC". The results list "Ad TNR Cafe" with a 4.5-star rating, 365 reviews, and a price range of \$\$ (Chinese, Asian Fusion, Soup). The address is 2049 Wilson Blvd, Arlington, VA 22201, and the phone number is (703) 875-0428. Both screenshots include a map on the right side of the results.

### Step 1

Scrape the following data from ALL restaurants in both searches

Restaurant Name	Street Address	City	State	Zip	Phone	Number of reviews	Rating	Price Range
-----------------	----------------	------	-------	-----	-------	-------------------	--------	-------------

Store the data from Mexican restaurants as **mexican.csv**

Store the data from Chinese restaurants as **chinese.csv**

### Step 2

Draw the histogram of **Rating** for both Mexican and Chinese restaurants and save them in separate pdf files names **histRatingChinese.pdf** and **histRatingMexican.pdf**. Make sure that the plots are labelled - axis labels for both axes and plot title.

### Step 3

For Chinese restaurants:

Plot the relationship between Rating (Y) and Price Range (X). Label it and save it as **CXY1.pdf**.

Plot the relationship between Rating (Y) and Number of reviews (X). Label it and save it as **CXY2.pdf**.

For Mexican restaurants:

Plot the relationship between Rating (Y) and Price Range (X). Label it and save it as **MX1.pdf**.

Plot the relationship between Rating (Y) and Number of reviews (X). Label it and save it as **MX2.pdf**.

### Step 4

Combine **mexican.csv** and **chinese.csv** into one one. Regress Rating (Y) on Price Range (X) and Number of reviews (X). Print out the coefficients and the  $R^2$ . Make a 3D plot (with both axes labelled and a title) and save it as **regress.pdf**,

Submit two files named: **A03\_gwid.py** and **A03Mod\_gwid.py**. The **A03Mod\_gwid.py** file should contain *at least* four functions (associated with the four steps). Name those four functions **step1**, **step2**, **step3**, and **step4**.

### How to submit your assignment

1. You need to submit two Python files to Blackboard to the Assignment 03 link.
2. The programs should be commented well enough so that the TA or I should not have to struggle with understanding variable names and codes and what statements or code blocks do.
3. The grading rubric is shown on the last page.
4. Your program headers for each program should look something like

```
# -*- coding: utf-8 -*-  
"""
```

Created on Fri Sep 04 09:09:54 2016

@author: kanungo  
GWID: G19860011

A brief description of the program not exceeding two lines  
"""

### Rubric for Grading the Programming Assignment

	Unsatisfactory	Satisfactory	Good	Excellent
<b>Delivery</b>	<ul style="list-style-type: none"> <li>Completed less than 70% of the requirements.</li> <li>Not delivered on time or not in correct format (Blackboard or git)</li> </ul>	<ul style="list-style-type: none"> <li>Completed between 70-80% of the requirements.</li> <li>Delivered on time, and in correct format (Blackboard or git)</li> </ul>	<ul style="list-style-type: none"> <li>Completed between 80-90% of the requirements.</li> <li>Delivered on time, and in correct format (Blackboard or git)</li> </ul>	<ul style="list-style-type: none"> <li>Completed between 90-100% of the requirements.</li> <li>Delivered on time, and in correct format (Blackboard or git)</li> </ul>
<b>Coding Standards</b>	<ul style="list-style-type: none"> <li>No name, date, or assignment title included</li> <li>Poor use of white space (indentation, blank lines).</li> <li>Disorganized and messy</li> <li>Poor use of variables (many global variables, ambiguous naming).</li> </ul>	<ul style="list-style-type: none"> <li>Includes name, date, and assignment title.</li> <li>White space makes program fairly easy to read.</li> <li>Organized work.</li> <li>Good use of variables (few global variables, unambiguous naming).</li> </ul>	<ul style="list-style-type: none"> <li>Includes name, date, and assignment title.</li> <li>Good use of white space.</li> <li>Organized work.</li> <li>Good use of variables (no global variables, unambiguous naming)</li> </ul>	<ul style="list-style-type: none"> <li>Includes name, date, and assignment title.</li> <li>Excellent use of white space.</li> <li>Creatively organized work.</li> <li>Excellent use of variables (no global variables, unambiguous naming).</li> </ul>
<b>Documentation</b>	<ul style="list-style-type: none"> <li>No documentation included.</li> </ul>	<ul style="list-style-type: none"> <li>Basic documentation has been completed including descriptions of all variables.</li> <li>Purpose is noted for each function.</li> </ul>	<ul style="list-style-type: none"> <li>Clearly documented including descriptions of all variables.</li> <li>Specific purpose is noted for each function and control structure.</li> </ul>	<ul style="list-style-type: none"> <li>Clearly and effectively documented including descriptions of all variables.</li> <li>Specific purpose is noted for each function, control structure, input requirements, and output results.</li> </ul>
<b>Runtime</b>	<ul style="list-style-type: none"> <li>Does not execute due to errors.</li> <li>User prompts are misleading or non-existent.</li> <li>No testing has been completed.</li> </ul>	<ul style="list-style-type: none"> <li>Executes without errors.</li> <li>User prompts contain little information, poor design.</li> <li>Some testing has been completed.</li> </ul>	<ul style="list-style-type: none"> <li>Executes without errors.</li> <li>User prompts are understandable, minimum use of symbols or spacing in output.</li> <li>Thorough testing has been completed</li> </ul>	<ul style="list-style-type: none"> <li>Executes without errors excellent user prompts, good use of symbols, spacing in output.</li> <li>Thorough and organized testing has been completed and output from test cases is included.</li> </ul>
<b>Efficiency</b>	<ul style="list-style-type: none"> <li>A difficult and inefficient solution.</li> </ul>	<ul style="list-style-type: none"> <li>A logical solution that is easy to follow but it is not the most efficient.</li> </ul>	<ul style="list-style-type: none"> <li>Solution is efficient and easy to follow (i.e. no confusing tricks).</li> </ul>	<ul style="list-style-type: none"> <li>Solution is efficient, easy to understand, and maintain.</li> </ul>