DSIC 210: HW #1 Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
Spring 2016
Points: 25

Consider the following variable description from the RateMyProfessor dataset. This data is provided on our course website.

|  |  |
| --- | --- |
| Variable / Field | Description |
| Table\_ID | Used to uniquely identify each row of the dataset, 7448 unique values: 1,2, …, 7448 |
| College | College for which the faculty member teaches. 11 unique values: Chicago, CMU, Tufts, Duke, UTexas, Harvard, Yale, MIT, Stanford, Princeton, Rice |
| Department | Department in which the faculty member teaches. 140 unique values: Accounting, Advertising, …, Writing  |
| Name | Name of Faculty Member, 7413 unique values: Abbott, Andrew, Achar, Pramod, …, Zodrow, George |
| Total Ratings | Number of Rate My Professor reviews. Min=0, Max = 116 |
| Overall Quality | Average overall quality score. Min=1, Max=5 |
| Easiness | Average easiness score. Min=1, Max=5 |
| Hot | An indicator of whether or not the faculty member is considered hot by at least one reviewer |

 Table name: Professors



1. Consider the following applications of the =COUNT() function in Excel. (4 pts)

Version A: =COUNT(Professors[College])

Version B: =COUNT(Professors[Total Ratings])

Version C: =COUNT(Professors[Overall Quality])

1. Which version produces the correct number of rows in this dataset? Discuss.
2. Figure out what Version A is doing. Why does this not produce the correct value?
3. Find an alternative to the COUNT() function to obtain the count using the Professors[College] column. Verify that this function produces the desired value.
4. Why does Version C produce a value that is different than Version B?
5. What is the average overall quality rating for all professors? (1 pt)
6. Using the =AVERAGEIF() function, obtain the average overall quality rating by college. Rank/sort the colleges from highest average overall score to lowest. (3 pts)
7. Using the =AVERAGEIF() function, obtain the average overall quality rating by Hot. Can we say that a faculty member that is considered hot by at least one reviewer tends to have a higher average overall quality score? How much does Hotness appear to affect the average overall quality score? Briefly discuss. (3 pts)
8. Using the =COUNTIF() function in Excel to obtain the total number of reviews by Department. (2 pts)



1. Use the =COUNTIFS() function in Excel, obtain the total number of Hot faculty members by Department. Repeat to obtain the total number of Not Hot faculty members by Department. Use these value to obtain the % of faculty members that are consider Hot by Department. Finally, use the % Hot column to sort the Departments from highest % Hot to lowest. Identify the top 5 and bottom 5 departments from this sorted list. (6 pts)

|  |  |  |
| --- | --- | --- |
| Top 5 Departments for Hotness |  | Bottom 5 Departments for Hotness |
| 1 |  |  | 1 |  |
| 2 |  |  | 2 |  |
| 3 |  |  | 3 |  |
| 4 |  |  | 4 |  |
| 5 |  |  | 5 |  |

1. Obtain a table akin to the one provided above, but here only include Departments that have a minimum of 10 reviews. This will prevent low counts from adversely influencing the rankings. (3 pts)

|  |  |  |
| --- | --- | --- |
| Top 5 Departments for Hotness\* |  | Bottom 5 Departments for Hotness\* |
| 1 |  |  | 1 |  |
| 2 |  |  | 2 |  |
| 3 |  |  | 3 |  |
| 4 |  |  | 4 |  |
| 5 |  |  | 5 |  |
| \*Departments with 10 or more reviews |

1. Use a Stacked Google Bar Chart (code provided below) to create a visualization of the table provided in the problem above. In order to successfully make this chart, you will need to calculate the % Hot and % Not Hot for each Department that is included on this graph.



Note: The code requires specification of a label for the Department, the % Hot, and % Not Hot. An example is provided here.

 ['Accounting', 45, 55],

 ['Chemistry', 40, 60],

 ['Biology', 30, 70],

 ['Mathematics', 20, 80]

If you’d like, you can use the =CONCATENATE() function as shown below to create the vector elements required for the Javascript code. The comma on the **last vector element must be removed**, e.g. Mathematics does not have a comma after the right square bracket.



|  |
| --- |
| HTML / Javascript Code for creating Google Chart |
| <html> <head> <script type="text/javascript" src="https://www.gstatic.com/charts/loader.js"></script> <script type="text/javascript"> google.charts.load('current', { packages: ['corechart', 'bar'] });  google.charts.setOnLoadCallback(drawBarColors); function drawBarColors() { var data = google.visualization.arrayToDataTable([  ['Department', '% Hot', '% Not Hot'], ['Accounting', 45, 55], ['Chemistry', 40, 60], ['Biology', 30, 70], ['Mathematics', 20, 80]  ]); var options = { title: 'Rate My Professor Hotness Percentages by Department', chartArea: { width: '50%' }, colors: ['#b0120a', '#ffab91'], hAxis: { title: 'Percent', minValue: 0 }, vAxis: { title: 'Department' }, isStacked: true, height: 600 };  var chart = new google.visualization.BarChart(document.getElementById('chart\_div')); chart.draw(data, options); }  </script>  </head> <body>  <script type="text/javascript" src="https://www.gstatic.com/charts/loader.js"></script> <div id="chart\_div"></div>  </body></html> |