# Intermediate Data Analytics – Final Project

You have been charged with deciding if any changes need to be made to the cafeteria at your college. You conducted a survey and got results back from 126 students who attend your college. The questions are in the “codebook\_food.docx” file. The responses are in the “food\_coded.csv” file.

Use the results of the survey to answer the following questions. Add any additional features to the dataset that you think would be useful in your analysis. At least one of your questions should be answered using text analytics (NLP) techniques.

You can use Excel, Python, Tableau, Power BI, or a tool (or tools) of your choice.

**Submit**

* A short (2-3 page) summary of the techniques you used to analyze the dataset. Review the following rubric and ensure that you have answered the questions in the rubric.
* Any scripts that you wrote to conduct your analysis. If you did not write scripts, include screenshots of any analysis that you did.
* Ann 8-10 slide presentation to the management of the college with your recommendations for changes to the cafeteria. Answer the following questions as part of those recommendations:

1. How health conscious are your students?
2. Have students’ eating habits changed since they came to college? For those who said habits got worse, what were the top two most reported changes? For those who said habits got better, what were the top two most reported changes?
3. If you offered ethnic food in your cafeteria, would students be likely to eat it? If so, what is the most popular cuisine?
4. You’d also like to offer some comfort foods for your students. What are the 5 most popular comfort foods?
5. You also want to bring some nostalgia to your cafeteria. What were your students top 5 childhood favorites?
6. Should you offer fruit in the cafeteria?
7. Should you offer vegetables in the cafeteria?
8. Based on the student’s ideal diet, what other changes would you recommend?
9. Do enough students live on campus who would come to your cafeteria?
10. Based on what your students are willing to pay for a meal out, how should your price your meals?

Rubric:

1. Answering the question (5 pts)

* Did you specify the type of data analytic questions (e.g. exploration, association causality) before touching the data?
* Did you consider whether the question could be answered with the available data?

1. Checking the data (10 pts)

* Did you plot summaries of the data?
* Did you check for outliers?
* Did you identify missing data?

1. Tidying the data (10 pts)

* Is each variable one column?
* Is each observation one row?
* If the data is not tidy, did you record the recipe for moving from raw to tidy data?
* Did you create a code book?
* Did you record all parameters, units, and functions applied to the data?

1. Exploratory analysis (25 pts)

* Did you identify missing values?
* Did you make univariate plots (histograms, boxplots)?
* Did you consider correlations between variables (scatterplots)?
* Did you check the units of all data points to make sure they are in the right range?
* Did you try to identify any errors or miscoding of variables?
* Would a scatterplot be more informative?
* Did you include at least one example that uses NLP/text analytics?

1. Written analyses (10 pts)

* Did you describe the question of interest?
* Did you describe the data set and question you are answering?
* Did you specify the type of data analytic question you are answering?

1. Visualizations (15 pts)

* Does each visualization communicate an important piece of information or address a question of interest?
* Do all your visualizations include plain language axis labels?
* Is the font size large enough to read?
* Does every visualization have a detailed caption that explains all axes, legends, and trends in the figure?

1. Presentation (20 pts)

* Did you lead with a brief, understandable to everyone, statement of your problem?
* Did you explain the data?

1. Reproducibility (5 pts)

* Did you avoid doing calculations manually?
* Did you create a script that reproduces all your analyses?
* Did you save the raw and processed versions of your data?