# INFO5502 Assignment 8: Develop a Data Analysis Project

November 27, 2019

In this assignment, you pick your **own questions** and **datasets** to build a data analysis project following data science workflow. Specifically, you need complete the following tasks:

1. Develop a question of your choice that can be addressed by identifying, collecting, and analyzing relevant data. You need find relevant data by yourself, and describe the data such as the source, attributes, size, how the data were collected, is the dataset sample data or population data?, etc. The dataset should have at least six distinct variables (i.e. columns) and a sample size (i.e. rows) of 500 or more. (3 points)
2. Perform exploratory data analysis (EDA). Describe the EDA process and result with at least four data visualizations. Explain whether the data is sufficient to answer the question you developed based on EDA result. If it is not sufficient, how did you address the issue? (3 points)
3. Describe any data cleaning or transformations that you perform and why they are motivated by your EDA? (2 point)
4. Apply relevant inference or predication methods such as **linear regression** or **K-nearest neighborhood (KNN)** to analyze your processed data, and validate the analysis results using cross-validation. Explain the training process, and the loss functions used in the analysis. Using examples (i.e. the values of the loss functions) to explain how the minimal value(s) of the loss function is/are found. (7 points)
5. Summarize and interpret your results including at least four data visualizations. Provide an evaluation of your approach and discuss any limitations of the methods you used. (2 points)
6. Write a project report to describe all tasks. (1 point)
7. Submit the original datasets (or public links to the datasets, make sure they are accessible), the processed datasets (or public links to the datasets, make sure they are accessible), Python code, and the project report. (2 points)

This assignment is developed based on content from www.ds100.org/fa19/gradproject/