Class Project:

This class project worth 40% of your course grade. Select one of the designated Kaggle competition projects, or a non-Kaggle project with instructor's approval. The emphasis of this course is practicing data science, not theoretical understanding of methodologies. The project should have enough depth and breadth to illustrate your analytic skills and your understanding of statistical/machine learning methodologies. You could form a team or work alone on this project. I highly recommend you consider the Kaggle projects because they have been vetted and have some reasonable structure to problem definition.

The project will be submitted in 3 parts.

Due on Nov 16

Part 1: Define the objective and scope of the project. Gather and organize the data for the project.

   a) Conduct exploratory data analysis such as visualizing the data through graphs, tables, summary statistics, and other means to understand the data.
   b) Identify any issues associated with data gap, data size, data type, data manipulation, data storage and data retrieval for analysis. Structured or unstructured data?
   c) Describe the high level analytic problem needs to be resolved: supervised learning, unsupervised learning.

Due on Nov 30

Part 2: Model construction and evaluation

   a) Construct analytic model(s) to address the project objective
   b) Evaluate the model outcomes
   c) Iterate and improve the model when necessary
   d) Justify the final model and its output

Due on Dec 7

Part 3: Document the findings

   a) Write a report summarizing the previous two parts. Clearly summarize your steps and your analytic process. Don't just provide screen shots of algorithm outputs.
   b) Compare your results with the Kaggle score board if you select a Kaggle project

You will be graded on your contribution to this project. Please indicate your work clearly in your team report so that credit will be given to the individual. At times, it is hard to separate a tight collaboration. Hence, team members would share the credit. Please indicate share credit in your submission.