**Week 2 Study Plan**

**Titanic: Machine Learning from Disasters I**

Objective:

1. Understand the problem description.
2. Given the data from Kaggle:
   1. Determine the most important factors which should be included during modeling
   2. Demonstrate the relationship between potential factors using plot or numeric prove with Pandas and Numpy as tools
   3. Preprocess data for future modeling (fill missing data, transform string to numeric value etc.)
3. Find potential models for this project.

Study Materials:

1. Getting Started with Python (https://www.kaggle.com/c/titanic/details/getting-started-with-python) :  
   This tutorial shows how to process data using Numpy.
2. Getting Started with Python II (https://www.kaggle.com/c/titanic/details/getting-started-with-python-ii):  
   This tutorial shows how to process date using Pandas package and more advanced data wrangling tools.
3. Python for data analysis – By McKinney  
   Refer this book for more techniques about data loading, wrangling, aggregation and visualization.

Requirement:

Please upload a report demonstrates the process for data wrangling and conclusion for next step modeling.

Note:

IPython Notebook can be used to write report which is an integrate tool and can include text, Python code and figures. This is suggested, but not required. Please use whatever editor you want.