

## CUSP-GX-5008.001: Big Data Management & Analysis SPRING 2016

# Homework 1

Please read Chapter 2 of the "Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data" textbook. (Note: an electronic version of this book is also available online within NYU network. Please check the Resources page on NYU Classes to access the book).

Your task is to come up with a data science project and an analytical plan based on the Data Analytics Lifecycle process that we discussed in the Week 2 lecture. You are expected to provide, at least, the followings.

1. **Project description:** please provide a general context and motivation of the project; clearly stating its main goals and the broader impact from a stakeholder or a project sponsor's point of view.
2. **Phase 1** must include potential team members + their roles, how you're framing the business problems to analytic problems as a data scientist. You must include at least **two** initial hypotheses. You are encouraged to use the CUSP Data Facility to find potential data sets to be used in the project.
3. **The rest of the other phases** must clearly document the roles and personnel involved, along with their contributions, in addition to the phase objectives.

Please use the Case study in Section 2.8 and the Yoyodyne Bank example in class as a guideline. The homework will be graded mainly based on the correctness in assigning tasks, roles, and personnel in each phase, though the data sets and model will be also be taken into account.

Please submit your assignment either as a PDF or a Word document on NYU Classes. (PDFs are preferable!)