Week 1 Assignment

By: Valon Tika

Summary and Problem Definition

The MSPS software survey was conducted in December of 2016 to ask students information regarding data science curriculum planning for the program. The overall problem definition of this survey was to gauge the overall curriculum of the program from the surveyors. It also provided an opportunity to change the curriculum based on surveyor feedback of needs and interests expressed for the curriculum.

Research Design, Measurement and Statistical Method

Design of this analysis started with creating a survey for current students to participate in that were enrolled within the MSPA program. The survey provided 15 question for students to answer. The questions were geared for the results to provide answers on any software languages that students thought were important to the program and potential changes to the overall curriculum for better representation of the market need for data scientists. The survey asks students to first distribute 100 points across multiple software tools, such as Python, R and Java based on their desire to learn based on personal, professional and industrial needs. Then, the survey asked for their level of interest of potential future courses that the program could offer.

Overview of Programming Work

The analysis of the survey results were completed in Python, which is a programming language that provides a vast array of tools and packages for various programming needs, such as data analysis and engineering. The analysis used the Python packages pandas, numpy, matplotlib and

seaborn for data wrangling and visualization. In this analysis, descriptive analytics, exploratory modelling and multiple variables were chosen that provided insight to the recommendation to management for data-driven decision making for the future trajectory of the program.

Review of Results – Recommendations

The analysis that was completed from the surveys provided great insight as to some key recommendations for the program. Throughout all the surveys, there was a strong recommendation that the need to know both R and Python was important in personal, professional and industrial development. When analyzing the average response rate of the need to learn a programming language for data science, R was the most preferred language out of the 5 languages, with Python being close in second place. SAS still had some significance for a need to learn in both professional and industrial, but not as much for personal development.

It is also noted the interest of courses heavily favored Python for data analysis over analytics, foundations data engineering, and systems analytics. The consensus was that data analysis courses with Python should be added within the curriculum as core courses.