

## **SUMMARY**

A MSPA Software Survey was collected in December 2016 and it was given to graduate students within the MSPA program. Data from the survey was used to help with the data science program curriculum planning. The objectives from the survey were to learn about current student software preferences, student interest in potential new courses, guide software and systems planning for current and future courses, and to guide data science curriculum planning. These objectives were created in mind to ensure that the program is capturing both the preferences and meeting the needs of the students in a technological and course material aspect. In addition, we will also be segmenting our data to see what software and languages are in high demand (high usage) in the industry so we will be able to prepare graduates of this program to have a competitive advantage in their job search, leading to a successful career.

## **RESEARCH DESIGN**

In order to collect enough data to meet our curriculum planning needs, we created a fourteen-item survey that initially consisted of the current state of the program, indicating the amount of courses with a particular program language/software a typical student should expect to take. The first section we will establish our segmentation of personal preference, professional need, and industry prevalence, where students will distribute 100 points throughout the languages and software of their preferences/need for each segment. Next the survey asks for the students to rank their level of interest 0-100 for a few classes that the program is thinking of offering, with 100 meaning that the student will certainly include that course in their curriculum. A few other personal questions about the student's graduating class and how many and which courses the

student has taken to gauge the current state of the most popular courses and to assess for survey bias (more newer students are taking the survey, etc).

## **TECHNICAL OVERVIEW**

Analysis on the survey data was performed using traditional statistical methods and Python, a high-level general program language. Several Python packages were used, including NumPy, pandas, matplotlib, and seaborn. The latter two packages were used to help create visualizations to guide analysis and curriculum planning.

## **RESULTS & CONCLUSION**

Among survey takers, it seems like there was a good mix of students, almost equally for those halfway through the program (>6 courses) and those not yet completed more than half of the program (<6 courses). Most of the students that took the survey were from the more immediate graduating classes, with most of them graduating in Spring 2018. Across the three segments of personal preference, professional need, and industry prevalence, R was the predominant preferred language, with Python closely following second, and JS, HTML, and CSS used the least. In terms of Course Interest, Python for Data Analysis had the highest percentage of course interest with 73.5%, followed by Foundations of Data Engineering (58.0%).

My recommendation would be to implement more coursework into the program that uses Python and implement “Python for Data Analysis”. Not only is Python a course and personal interest, but data shows that usage of Python is high for the industry and professional segments. With approximately half of the survey takers not having completed half of the program, interest in Python is not fleeting and it may be possible

to implement more Python in the program, and possibly parallel courses in R if the budget and enough interest allows. If debating between Python or R for program usage, an additional survey can be sent out for survey takers to indicate a narrowed preference.