

Dalya Adams  
MSDS 422 section 59  
Survey Exploration and Visualization

In creating a successful data science graduate program, delivered to students across the globe, in a multitude of industries, ensuring that the coursework keeps up with the shifting technology is extremely important. To address the ever-changing industry software, as well as ensure that coursework is targeted to the wants and needs of students, a MSPA Software survey was collected in December 2016. The results of this survey were used to inform data science curriculum planning for the MSDS program.

As academic administrators define the future of the MSPA/MSDS graduate program, they must keep many things in mind, including software preferences and interest in potential new courses. They must guide software and systems planning for current and future courses and curriculum. They must also weigh the wants of the current and future students against the shifting software landscape in the data science world. This problem is what this survey was intended to help address. By identifying the current preferences of students, as well as identifying the changes in the industry and professional landscapes, the survey was intended to inform decisions for the MSPA/MSDS program.

The survey consisted of 14 items, taking approximately 15 minutes of the respondent's time. It covered topics from personal, professional and industry language preferences, to the respondent's level of interest in four different courses. The survey was delivered through Survey Monkey and was taken by 207 students, with graduation dates ranging from Fall 2016 until 2020 and later. The software choices presented to respondents were: Java, JavaScript, Python, R and SAS and interest was gauged on 4 new courses: Python for Data Analysis, Foundations of Data Engineering, Analytics Application Development, and Data Science Systems Analysis.

In order to assist academic administrators in steering the MSPA/MSDS program in the best direction, we looked at the interest in proposed courses as well as the changes in preferences

Dalya Adams  
MSDS 422 section 59  
Survey Exploration and Visualization

of students, personally, professionally and industry wise, across graduation dates. The assumption being that students with a later graduation date are more recent entrants into the program and the changing preferences of these new entrants may indicate changes in the industry. This indicator, if accurate, allows the program to adjust its software and courses, allowing it to remain relevant and provide the best education to the students in the MSPA/MSDS program.

Across all 207 respondents, R was, on average, the most preferred language in all three categories, personally, professionally and industry wise, followed by Python, SAS, Java and JavaScript. Separating the respondents into three groups, those graduating in 2017 or before, 2018 graduation dates, and 2019 or later graduation dates, we see the same ranking of languages across all 3 subsets of graduation date. In looking by graduation date, those graduating in 2019 or later show a higher than average interest in Python and a lower than average interest in SAS. Those graduating in 2018 showed the highest average interest in R among the three groups. Next, in looking at the descriptive statistics of student interest in four different courses, the highest average interest is in Python for Data Analysis, followed by Foundations of Data Engineering, Analytics Application Development and then Data Science Systems Analysis.

In conclusion, focusing resources on classes utilizing R, Python, and SAS would ensure that the languages with the highest personal, professional and industry demand are accounted for. Also, increasing the prevalence of Python courses, from the 3 provided in the 2016 MSPA curriculum, would provide students with the opportunity to become more familiar with Python. Possibly allowing students to choose between taking a course in R, Python or SAS would solve both the limited number of Python courses, as well as the high interest in Python specific coursework.

Dalya Adams  
MSDS 422 section 59  
Survey Exploration and Visualization

Appendix

Descriptive statistics for Personal Preferences of all respondents

	My_Java	My_JS	My_Python	My_R	My_SAS
count	207.000000	207.000000	207.000000	207.000000	207.000000
mean	10.135266	4.797101	31.304348	37.125604	16.637681
std	11.383477	6.757764	15.570982	14.576003	13.626400
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	20.000000	30.000000	5.000000
50%	9.000000	0.000000	30.000000	35.000000	15.000000
75%	20.000000	10.000000	40.000000	50.000000	25.000000
max	70.000000	30.000000	90.000000	100.000000	75.000000

Descriptive statistics for Professional Preferences of all respondents

	Prof_Java	Prof_JS	Prof_Python	Prof_R	Prof_SAS
count	207.000000	207.000000	207.000000	207.000000	207.000000
mean	9.251208	5.840580	30.028986	36.415459	18.463768
std	13.167505	10.812555	19.144802	20.847606	18.831841
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	20.000000	25.000000	0.000000
50%	5.000000	0.000000	30.000000	33.000000	15.000000
75%	15.000000	10.000000	40.000000	50.000000	30.000000
max	80.000000	100.000000	100.000000	100.000000	100.000000

Descriptive statistics for Industry Preferences of all respondents

	Ind_Java	Ind_JS	Ind_Python	Ind_R	Ind_SAS
count	207.000000	207.000000	207.000000	207.000000	207.000000
mean	11.942029	6.966184	29.772947	32.434783	18.884058
std	14.706399	10.030721	17.959816	15.912209	19.137623
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	20.000000	22.500000	0.000000
50%	5.000000	0.000000	30.000000	30.000000	15.000000
75%	20.000000	10.000000	40.000000	40.000000	30.000000
max	70.000000	50.000000	95.000000	85.000000	100.000000

Dalya Adams  
MSDS 422 section 59  
Survey Exploration and Visualization

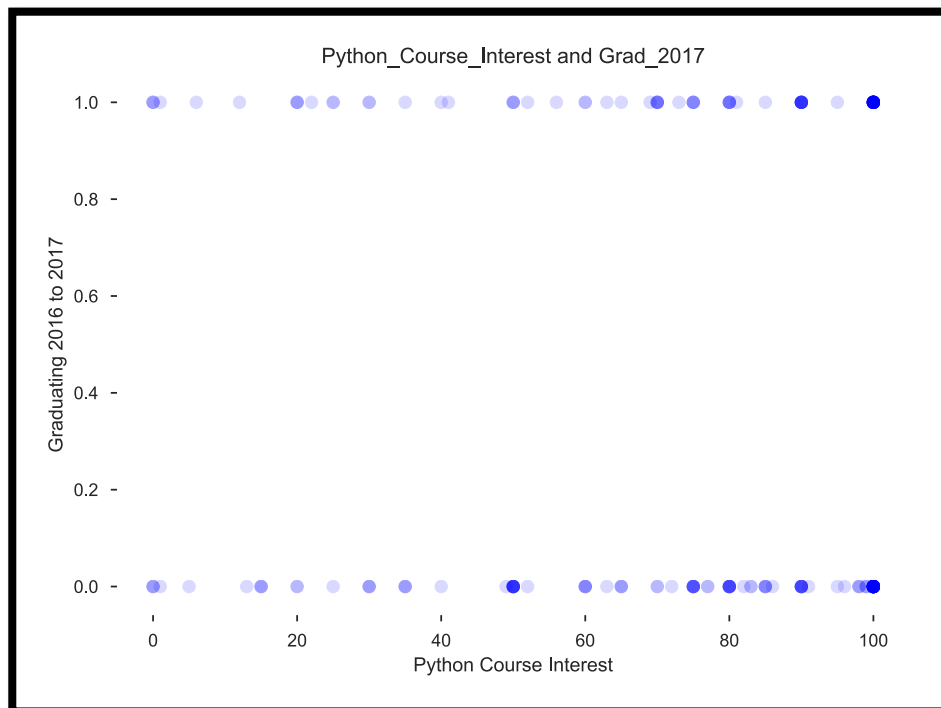
Descriptive Statistics for Course Interest

	Python_Course_Interest	Foundations_DE_Course_Interest	Analytics_App_Course_Interest
count	206.000000	200.000000	203.000000
mean	73.529126	58.045000	55.201970
std	29.835429	32.588079	34.147954
min	0.000000	0.000000	0.000000
25%	53.000000	29.500000	25.000000
50%	82.500000	60.000000	60.000000
75%	100.000000	89.250000	85.000000
max	100.000000	100.000000	100.000000

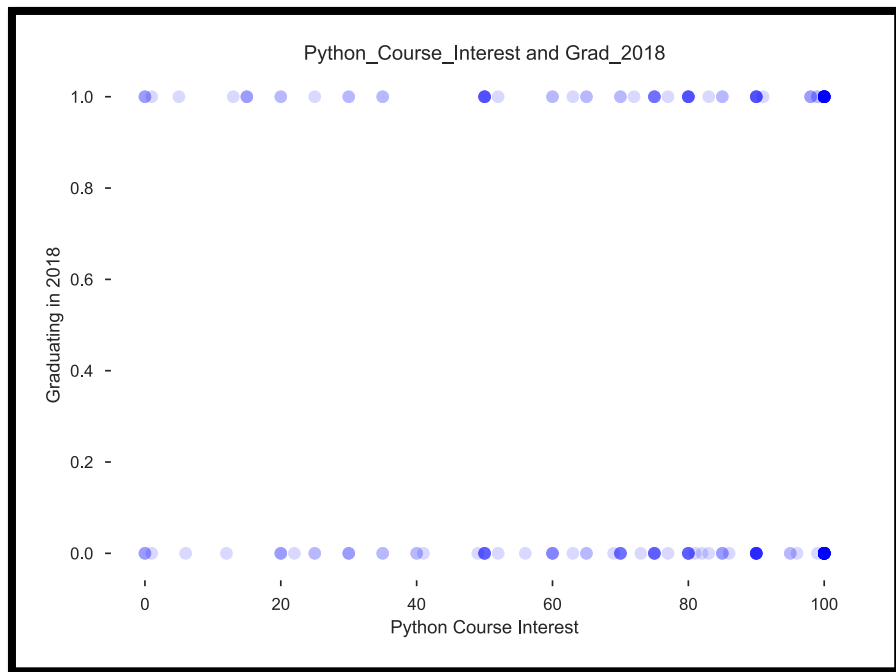
	Systems_Analysis_Course_Interest
count	200.000000
mean	53.630000
std	33.539493
min	0.000000
25%	21.500000
50%	51.500000
75%	80.250000
max	100.000000

Scatter Plot of Interest in Python Course by students graduating in 2017 or sooner



Dalya Adams  
MSDS 422 section 59  
Survey Exploration and Visualization

Scatter Plot of Interest in Python Course by students graduating in 2018



Scatter Plot of Interest in Python Course by students graduating in 2019 or later

