

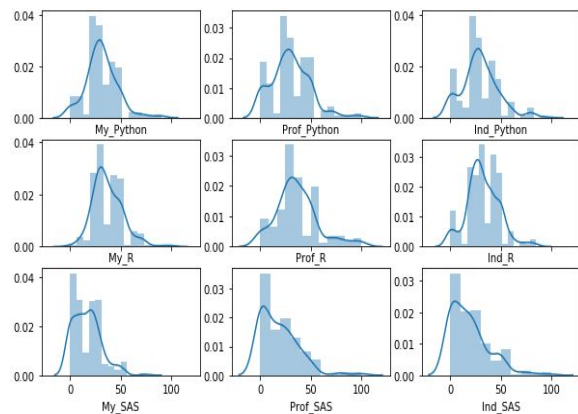
## Introduction

For the Data Science graduate program to remain competitive in the future while meeting the needs of our students, we must assess student skill development relative to industry needs. Within this analysis, we seek to have a deeper look at feedback from students regarding their individual preferences as they relate to perceived industry needs.

This analysis is based on data from the “MSPA Software Survey”, which was a survey given to students in the MSPA program during December of 2016. The survey focused on student preferences and perception of value of five tools: Python, R, SAS, Java/Scala/Spake, and Javascript/HTML/CSS.

## Discussion: Research Design and Statistical Methods

This survey was given to 207 students currently enrolled in the ‘MSPS’ program. The survey consisted of 14 separate questions, each querying the student about a specific part of the program.

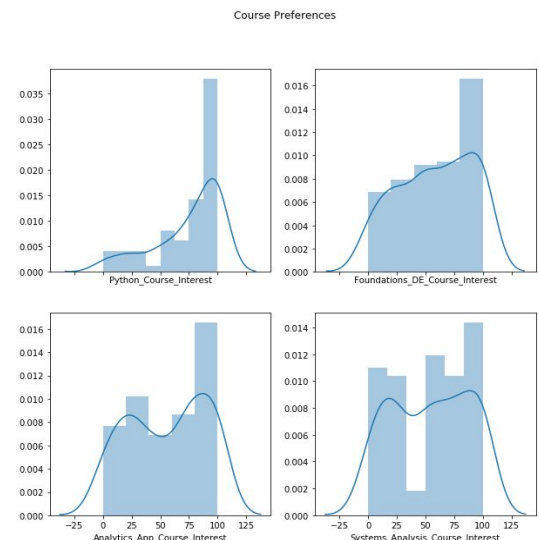


First, students were asked to distribute 100 points between 5 tool options: Python, R, SAS, Java/Scala/Spake, and Javascript/HTML/CSS with respect to three different categories: personal desire learn the tool, professional need to learn the tool, and their perception of the tool’s value within the industry. Next, students were asked about their interest in a variety of potential elective courses relating

to data engineering: Python, Foundations in DE, Analytics Apps, and Systems Analysis.

Students’ preferences and perceived value of languages heavily favored Python and R. If we look at distributions of each language’s score, we can see that Python and R both have much more normally distributed curves, as opposed to the more right-skewed SAS.

This preference is also reflected in the students’ strong interest in a future data engineering elective course focusing on Python skills. Note

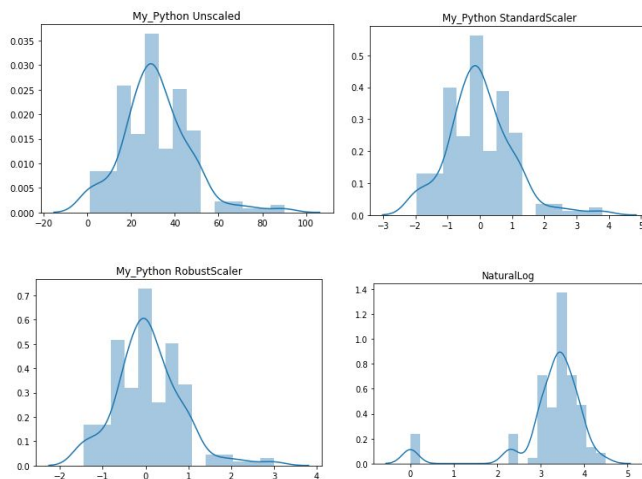


the scale of the histograms: students scored their interest in a Python elective at almost twice the interest in the other three courses.

When we look at an intercorrelation heat map between the different student language preferences, we see that the highest correlations between students' personal and professional preference for Python skills, and their perceived value of Python skills within the data science industry. Furthermore, if we examine the intercorrelation between preferred language and course interest, we unsurprisingly see a correlation between interest in Python and interest in a Python course.

## Programming Work

The analysis of this data was completed in Python, a high-level language used for general purpose programming. This analysis leans on some of the most popular data analysis packages in the world, such as Pandas, Matplotlib, and Numpy.



Within the notebook is a review of the effect of transformation against a variable that had significant magnitude: 'My\_Python'. Using the Seaborn.distplot function, the notebook displays the effect of the distribution of the My\_Python data using different scaling techniques. The distribution seems relatively unaffected in shape when using the 'Standard' or 'MinMax' scaler (we do see some change in the amplitude of the distribution), but then the

distribution becomes much more skewed to the right when using a natural log transformation.

## Results and Recommendation

The data suggest that there is a strong preference for using Python as the primary tool -- not only in the MSDS program, but also in students' professional lives. Students also rank Python as the most valuable technology within their industry.

Within the MSPS curriculum, there's an equal offering of SAS, Python, and R. It would be wise of the MSPS program to consider ways to incorporate more Python-intensive courses and tools into the curriculum in order to stay abreast of industry demand, and also to continue to offer a valuable knowledge set to attract students.