

A survey was conducted in December 2016 by Northwestern University Master's in Predictive Analytics (MSPA) program to determine programming language and software interest. The five computer languages or software systems that were evaluated were: (a) Java/Scala/Spark; (b) JavaScript/HTML/CSS; (c) Python; (d) R; and (e) SAS. In addition, four questions evaluated the interest in new courses for data engineering. Respondents were also asked to indicate which courses in the program had been completed and their expected graduation date. This information was used to assist in the curriculum planning for the program and allow the program to remain a desirable and relevant degree for future enrollments.

The survey responses were collected into a comma separated value file (.csv). The .csv file was opened in Python and saved into a data frame called `valid_survey_input`. The data frame contains responses from 207 individuals on 40 variables specified in the survey. The `RespondentID` served as an index, or row identifier, and was not evaluated. An exploratory data analysis (EDA) was conducted on the data set utilizing various data manipulation and visualization packages. A profile analysis of the data set was conducted with the `pandas_profiling` module. This includes information on the basic data type for each variable, a descriptive statistical analysis with quantiles, histograms and correlations. A correlation matrix of survey variables was constructed into a graphical display called a heat map when applicable. The correlation coefficient value is displayed, and color coordinated. Positive correlation is red and negative correlation is blue.

Respondents have completed an average of six courses. While there is no specific course that all students have completed, most have completed the core courses of 400, 401, and 420. Two other courses that had a high number of completions were the data modeling components

of 410 and 411. Participation is intermittent in the remaining courses. Most of the respondents had an expected graduation date of Winter 2017 or later.

A majority of respondents had a higher personal desire to learn the programming language R followed by Python and then SAS. Interest in both Java/Scala/Spark and JavaScript/HTML/CSS were the lowest of the five programming languages. Results for professional need and importance and prevalence in respondent industry were similar to the personal desire outcome. An indication of a personal desire to learn a specific programming language had a high correlation to professional need and industry importance. There was a negative correlation between a personal desire to learn the R programming language and the other programming languages. Another notable negative correlation was observed between a personal desire to learn SAS and Python.

Respondents provided favorable rating for all four proposed courses. There was high interest in a new course on Python for Data Analysis, followed by Foundations of Data Engineering. There was a high correlation between interest in Analytics Application Development and Data Science Systems Analysis courses.

Based on the survey results, it is recommended that all four new courses be developed and incorporated into the curriculum. The program could implement more Python into the courses, maintain options available in R, and focus less on SAS. If the changes were implemented as soon as possible, then students enrolled in the program would benefit professionally and cultivate programming skills relevant to their respective industries. The MSPA program should implement these changes in order to remain relevant in the fast-changing and constantly evolving field of data science.