**Introduction:**

A survey was created to understand and analyze the curriculum and tool sets used in the MSPA program and to closely relate to the industry requirements and standards. One of the important attributes is the programming language required/recommended/preferred for a course. The survey is created to keep the program up to date and keep improving on courses and tool sets used and to also improve the curriculum, based on the courses preferred by students. The data collected is based on the responses from current students and faculty. The goal of the project is to analyze the survey data and provide feedback on the value of current courses and tool sets.

**Discussion:**

The survey included several questions relating to the courses completed, the programming languages that were used for the course and their personal, professional preference on the tool sets along with their industry experience in the industry.

A total of 207 responses were collected in December 2016. From the data, we see that many students were expected to graduate in 2017 and 2018. Each student that responded has completed/registered at least one course that involved Python, R or SAS. The table on the left gives a breakdown of the courses offered and the total number of respondents that completed the course. On an average, 6 courses were completed by the students. Figure 3 implies that these tools showed a close to normal distribution.

| | Course | Title | Software | Count |
|---|---|---|---|---|
| 0 | PREDICT400 | PREDICT 400 Math for Modelers (Python) | Python | 163 |
| 1 | PREDICT401 | PREDICT 401 Introduction to Statistical Analys... | R | 171 |
| 2 | PREDICT410 | PREDICT 410 Regression and Multivariate Analys... | SAS | 145 |
| 3 | PREDICT411 | PREDICT 411 Generalized Linear Models (SAS) | SAS | 113 |
| 4 | PREDICT413 | PREDICT 413 Time Series Analysis and Forecasti... | R | 59 |
| 5 | PREDICT420 | PREDICT 420 Database Systems and Data Preparat... | Python | 127 |
| 6 | PREDICT422 | PREDICT 422 Practical Machine Learning (R) | R | 48 |
| 7 | PREDICT450 | PREDICT 450 Marketing Analytics (R) | R | 17 |
| 8 | PREDICT451 | PREDICT 451 Risk Analytics (R) | R | 7 |
| 9 | PREDICT452 | PREDICT 452 Web Analytics and Network Data Sci... | Python | 13 |
| 10 | PREDICT453 | PREDICT 453 Text Analytics (Python) | Python | 11 |
| 11 | PREDICT454 | PREDICT 454 Advanced Modeling Techniques (R) | R | 5 |
| 12 | PREDICT455 | PREDICT 455 Data Visualization (R) | R | 30 |
| 13 | PREDICT456 | PREDICT 456 Sports Performance Analytics (R) | R | 6 |
| 14 | PREDICT457 | PREDICT 457 Sports Management Analytics (R) | R | 4 |
| 15 | OtherPython | Other Course with Python as the Primary Language | Python | 5 |
| 16 | OtherR | Other Course with R as the Primary Language | R | 14 |
| 17 | OtherSAS | Other Course with SAS as the Primary Language | SAS | 2 |
| 18 | Other | Other | Other | 26 |

**Figure 1: Courses and Tools**

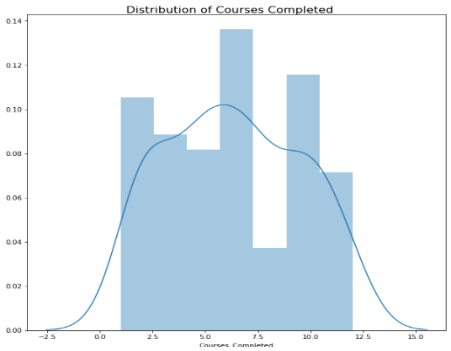


**Figure 2: Distribution of Courses Completed**

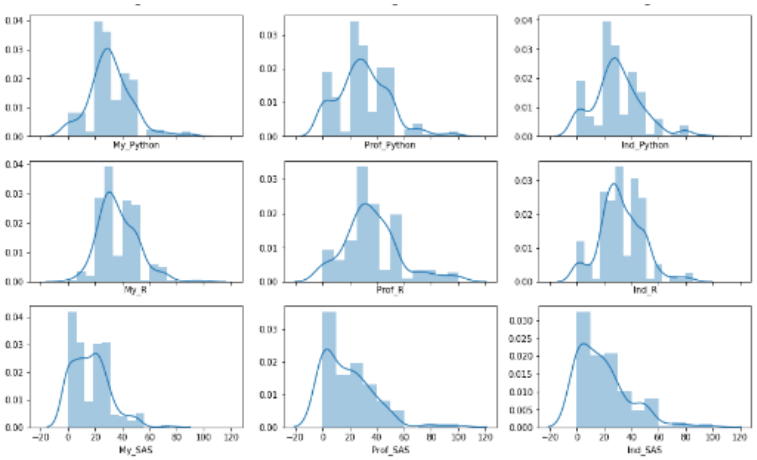


**Figure 3: Distribution plots for Python, R and SAS**

On a 100-point scale, when the tool sets were rated, Python, R and SAS were the most preferred compared to Java and JS. The distribution plots in

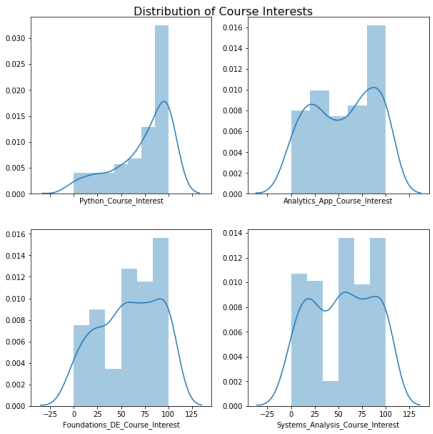


**Figure 5: Distribution of Course Interests**

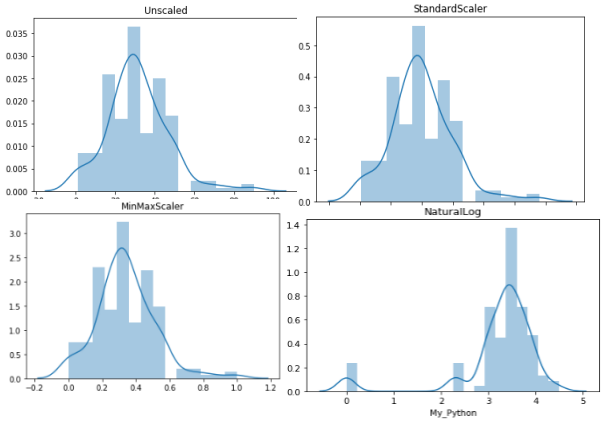


**Figure 4: Distplot of My_Python - Unscaled, StandardScaler, MinMaxScaler and Natural Log**

Figure 4 represents the effect of transformation on My_Python variable which has the strongest correlation value. The distribution is similar when Standard and MinMax scalers are used, but the distribution is skewed to the right when the NatualLog transformation is used. The distribution plots in Figure 5, demonstrate that Python is the most preferred tool with 30% of interest, while other courses range at 15%.

**Insights and recommendations:**

The review of data collected and analyzed so far, suggests that there is a preference for Python as per personal, professional and industry standards.

**Appendix:**

The ipynb notebook and an html version of the notebook along with the output and graphs are included in the submission.