# DATA SCIENCE PROGRAM OBJECTIVES

## EXECUTIVE SUMMARY

To ensure that our Data Science program graduate program meets the needs of our students from both technological and course content perspectives we must assess what our program currently offers and analyze our students feedback on their individual preferences from both technological and offered course content.

We need to understand where there are potential areas of improvement in course material and software preferences that students have. We must also be mindful of what the job market is demanding, as it is important that our graduates are adequately prepared for a job in the highly completive data science market.

## RESEARCH DESIGN

For this analysis we designed a survey for students currently enrolled in the program to solicit feedback on various parts of the program. The survey consists of fourteen separate items, each constructed with a specific part of the program in mind. The overall goal of the survey is to get detailed feedback on the technologies used in the program, as well as where students would like to see the curriculum expanded in the form of new courses and specializations.

Data Science is heavily dependent on the use of various software systems to facilitate almost every aspect of the learning objectives, from general purpose programming, software-based modeling, and of course the latest database system technology from both the programming / integration side as well as higher level concepts such as what types of database technology are best suited for specific business use cases. In this survey will we assess students programming language exposure through their course work, as well as their personal preference, perceived industry need and professional need.

The survey also has sections about what courses students would have interest in taking to enhance our curriculum. The courses offered in the program must align with need in the data science industry, and the survey will have four new proposed classes, each focusing on a specific sub-domain of the data science discipline.

## TECHNICAL OVERVIEW

The analysis of the survey data was completed in Python, a high-level language used for general purpose programming. The solution leverages some of the most popular data analysis packages in the world, such as Pandas, Matplotlib and Numpy to quickly and efficiently take the idea from prototype to actionable results as quickly as possible.

In this analysis, we made use of traditional statistical methods and general-purpose data transformations to visualize the survey results so that we can make informed suggestions on how to improve current and future students experience in the Data Science program.

## CONCLUSION

After performing the aforementioned data analysis on the student survey data, there are a few key take-aways that I believe are strongly supported by the data.

One hypothesis I had was that the future course interest would be highly variate when categorized by students who just started the program, and students who were nearing completion. There was a clear outlier in the four proposed courses, and that's the course dedicated to Python programming. The students interest in this class was universally strong across all program tenures, indicating that even students who were about to complete the program would still have interest in this class, and it should be one of the first classes students take given the foundational aspects if offers in the program. Adding this course to the curriculum should be an immediate action item.

Secondly, the analysis of student software preferences was a bit surprising in that overall R was the dominate language, although it was very close overall with Python. The two languages dominated the survey, and for good reason as they're universally considered the core toolkits of any data scientist. I would recommend that SAS be worked out of the program, and the time/courses spent working with SAS would be better allocated by increasing the proficiency of our students with both R and Python as there is clear interest by the students to have a dedicated Python course which tells me that students feel they could potentially be leaving the program without their desired proficiency in arguably the most fundamental tool in the field.

These recommendations are clearly supported by the student survey, and the impacts to the program could be critical.