

University of Toronto
Department of Mechanical and Industrial Engineering

Course Project

MIE 1624: Introduction to Data Science and Analytics

Group 8

Bo Peng (998834711)
Haotian Zhao (1002601669)
Harshil Patel (100757672)
Shuyue Zhang (1007538185)
Tuhin Ranjan (1006633555)
Tushar Aggarwal (999356913)

Due Date: November 30, 2020

Executive Summary

This is the report for the group project of MIE 1624: Introduction to Data Science and Analytics. This project aims to first redesign the course curriculum for "MIE1624: Introduction to Data Science and Analytics"; second, to design a curriculum for a new technically- and business-oriented program: Master of Data Science and Artificial Intelligence; and third, develop analytical-based solutions to some topics related to data science education.

The main tool used for this project is Python 3 and its libraries. Web scraping is performed to obtain job skills requirements from Indeed and LinkedIn, and skills taught by existing courses on Coursera. More data on skills used in data science-related jobs are obtained from the Kaggle Survey dataset. Exploratory data analysis is performed to find the frequency of occurrence and relation between those skills.

The new courses are designed based on the correlation of skills in job requirements and existing courses. Detailed description, visualizations, and reasonings are provided for the Mandatory courses: Introduction to Data Science and Analytics, and core courses: Introduction to Machine Learning, Introduction to Deep Learning, Big Data, Statistical Analysis, and Professional Communication and Leadership. Simple descriptions are provided for the elective courses: Basic Coding, Advanced Data Analysis, Introduction to Project Management in Data Science, Professional Communication and Leadership, and Operations Research.

As a full master program, it also contains a capstone project, where two sample capstone project topics are proposed in the report. Further efforts are put into researching for companies that offer internship opportunities, and visualizations are done for the top 10 companies that most frequently offer internships. Finally, an EdTech Startup ideation is proposed to help students pick courses and acquire skills that are required according to the current market.

Table of Contents

Executive Summary.....	ii
1.0 Introduction	1
2.0 Data Retrieval & Analysis	1
3.0 Curriculum Design.....	2
3.1 Courses.....	2
3.1.1 Introduction to Data Science & Analytics (MIE 1624).....	2
3.1.2 Introduction to Machine Learning	3
3.1.3 Big Data	4
3.1.4 Professional Communication & Leadership.....	4
3.1.5 Business Intelligence.....	5
3.1.6 Introduction to Deep Learning.....	5
3.1.7 Statistical Analysis	5
3.1.8 Advanced Data Analysis	6
3.1.9 Full Stack Development	6
3.1.10 Introduction to Project Management in Data Science	6
3.1.11 Data Science Programming Basics	7
3.1.12 Introduction to Computational Finance	7
3.1.13 Operations Research.....	7
3.2 Capstone Project Option	7
3.3 Internship Option.....	8
4.0 EdTech Startup Ideation.....	9
5.0 Conclusion.....	9
6.0 References	11
Appendix A: Occurrence of the Skills in Job Postings	12
Appendix B: Hierarchical Clustering for Skills Required for Technical Roles	13
Appendix C: Hierarchical Clustering for Skills Required for Management	14
Appendix D: Industry View on Academic Achievements vs. Independent Projects.....	15
Appendix E: Most Popular Platforms used for Data Science Related Courses	16
Appendix F: Most Common Skills Taught with Each Topic in Coursera.....	17
Appendix G: Number of Survey Participants in Each Industry.....	18
Appendix H: EdTech Startup Ideation Flow Chart	19

List of Figures and Tables

Figure 1. Curriculum Timeline Structure.....	2
Figure 2. Most Popular Languages used in Industry, Kaggle 2019	3
Figure 3. Most Interacted Types of Data, Kaggle	6
Figure 4. Top Technical Skills in DS Industry	6
Table 1. Recommended Pathway for Technical & Business Emphasis.....	2
Table 2. Sample Capstone Project Recommendations	8
Table 3. Companies & Industries for Internship Opportunities.....	8

1.0 Introduction

Data science is an interdisciplinary field that uses mathematics, statistics, computer science, and many other areas of knowledge to extract information and insights from data. As data science, analytics, AI, and big data have developed rapidly in recent years, demands in data science education also increases rapidly.

Currently, at the University of Toronto, the Data Science and Artificial Intelligence topics are included in the Master of Mechanical and Industrial Engineering degree as an Emphasis in Analytics. The objective of the project is to apply the skills learned in the course to design a curriculum for a new technical- and business-oriented “Master of Data Science and Artificial Intelligence” program. First, the course curriculum for “MIE1624: Introduction to Data Science and Analytics” is re-designed. Second, other courses and capstone projects are designed for the program. And lastly, analytical-based solutions are developed for some topics related to data science education.

2.0 Data Retrieval & Analysis

To design a curriculum that meets the needs of the current job market, web scraping tools were used to collect and parse out job postings from web job portals like Indeed and LinkedIn with job titles such as Data Analyst, Data Scientist, and Data Engineer. Analysis of the retrieved job postings was performed for both technical and business skills required for the jobs. The most in-demand skills in the current job market were found via this process. The data was further analyzed to find the most common skills in the current job market, shown in **Appendix A**, and hierarchical clusters were performed on the skills that were required for technical roles, as shown in **Appendix B**. Similarly, **Appendix C** contains the analysis of hierarchical clustering for management roles.

Furthermore, the team used the 2018 and 2019 Kaggle ML and Data Science Survey (Kaggle Surveys) [1] [2] to study the tools and techniques used by data scientists in practice. It was also found that the industry would value more practical experiences than just good marks in the courses, as shown in **Appendix D**. Thus, in the program, the students would be asked to do either a 4-month internship or a capstone project to gain hands-on experiences and to demonstrate their abilities. The instructors would also be encouraged to include project assignments in the course if possible.

Besides understanding the skills required in the job market, the programs and courses currently offered in the market and universities were also studied. The team researched Data Science Programs offered by other faculties or universities, such as Rotman UofT, UBC, and Waterloo to learn of how they structured their programs. Also, in analyzing the Kaggle Surveys, it was found that Coursera [3] is the most commonly used platform that the participants took data science courses from, which is included in **Appendix E**. As a result, the team scrapped the data related courses from Coursera and analyzed how the Coursera courses were designed, especially, what are the skills that are covered in each topic. The figures and graphs from the analysis are included in **Appendix F**.

Consequently, the Master of Data Science and Artificial Intelligence curriculum is designed based on the understanding of the current job market demand and the current programs and courses offered by other platforms and universities.

3.0 Curriculum Design

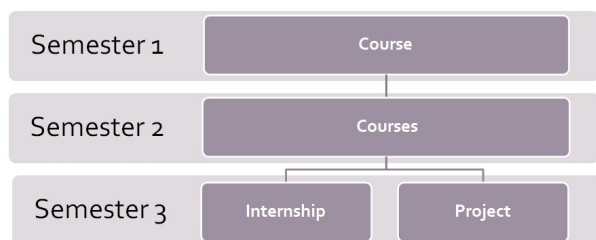


Figure 1. Curriculum Timeline Structure

The program is a full-time professional Master's Program with 3 4-month semesters as shown in **Figure 1**. To receive the degree, students must have completed a total of 7 courses and completed either a 4-month internship or an individual project supervised by a faculty member in the last semester.

For the first two semesters, students must take at least 3 courses each semester. By the end of the second semester, students must notify the faculty which option they have chosen and submit either a copy of their internship employment contract or a project advisor agreement form to the faculty.

Students enrolled in the program have 2 options to focus their degree: the technically focused pathway or the business-focused pathway. The recommendation of course selection for each pathway is shown in **Table 1**.

Table 1. Recommended Pathway for Technical & Business Emphasis

Technical	Business
Introduction to Data Science & Analytics*	Introduction to Data Science & Analytics*
Introduction to Machine Learning*	Professional Communication & Leadership*
Big Data*	Business Intelligence*
Introduction to Deep Learning	Introduction to Project Management in Data Science
Statistical Analysis	Data Science Programming Basics
Advanced Data Analysis	Introduction to Computational Finance
Full Stack Development	Operations Research
<i>Courses with * are courses that are mandatory for the pathway as the skills taught in these courses are essential in the job market. Other courses are recommendations. Students can pick and design their degree as they like, however, all course requirements should be met.</i>	

3.1 Courses

The following section provides the course description, learning outcomes, and justification of each course offered in the designed Master of Data Science and AI program.

3.1.1 Introduction to Data Science & Analytics (MIE 1624)

The main update in terms of content for this course is the elimination of deep learning in-depth theory, instead, it will just briefly be covered. As well as simulation modeling will be briefly introduced but the focus will be on Python programming, statistical analysis, linear algebra, regression modeling, data visualization, and some of the top advanced machine learning algorithms. The grading scheme will also be changed to allocate more marks for technical communication via presentations and report writing. Each assignment will also require students to submit a 2-3-minute video presentation summarizing their objectives along with the notebooks, and PowerPoint presentation slides. The course project will be a group project with a formal presentation as well.

Course Description: The course is divided into two major sections [4]:

1. Introduction to computational, mathematical, and statistical tools to be able to build data science models
2. Introduction and application of machine learning and data science algorithms

Updated Course Goals & Learning Outcomes:

- **Python Programming:** Introduction to python programming, and comparison of python to R, and SQL for data science applications.
- **Advanced Machine Learning:** Introduction to classification techniques (decision trees), advanced supervised machine learning algorithms, ensemble learning algorithms, text analytics and natural language processing, clustering, and a very brief overview of neural networks and deep learning algorithms, dimensional reducibility, and reinforcement learning.
- Other learning outcomes for Introduction to data science and analytics, Basic Statistics, Linear Algebra, Optimization, Modelling techniques & Regression, and Data Visualization will remain the same as the 2020 curriculum.

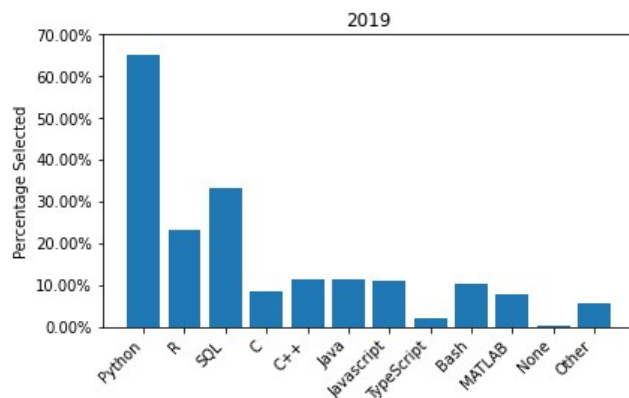


Figure 2. Most Popular Languages used in Industry, Kaggle 2019

R and SQL will be introduced in the course and compared to Python for data science applications since “Python”, “SQL”, and “R” are the top 3 most popular languages in the Kaggle survey as shown in **Figure 2**. More marks are allocated for presentations and report writing as communication and project management is a very valuable skill for employers according to the top skills from LinkedIn, Indeed, etc.

3.1.2 Introduction to Machine Learning

Course Description: Machine Learning is an important component of data science. This course will provide students with basic skills and broad knowledge of several machine learning algorithms. The course will start with very popular machine learning algorithms such as Linear Regression, Logistic Regression, and as the course progresses students will be introduced to advanced machine learning algorithms such as Deep Learning (Neural Network), Reinforcement Learning. The course will be taught using Python. Students will be given a review tutorial on basic mathematical skills such as statistics, linear algebra, probability theory. Moreover, students will be provided with Tensorflow and Pytorch framework as they were some of the popular frameworks currently used in the industry [5].

Pre-requisites: Students are expected to have taken some basic statistics, probability theory, and linear algebra in undergraduate courses as this course heavily relies on math. Exception: Students taking any Statistics/Linear Algebra Course.

Learning: Linear Regression, Logistic Regression, Ridge Regression, Linear classification, Gradient Descent, Decision Tree, Random Forest, Support Vector Machine, Optimization, Hyperparameter optimization, Principal Component Analysis, K-Means, Monte Carlo simulations, Introduction to Neural Network, Introduction to Convolution Network, Introduction to Reinforcement Learning.

3.1.3 Big Data

Course Description: This course provides a thorough, introductory understanding of Big Data, which would cover many technologies that are used in practice. You will gain an understanding of what insights big data can provide through hands-on experience with the tools and systems used by big data scientists and engineers [6] [7].

Course Goals & Learning Outcomes:

- Big Data Infrastructure: Introduction to Hadoop Cluster, HDFS and MapReduce, Resource Management with Yarn
- Big Data Storage and Querying: Introduction to SQL Recap and Hive
- Machine Learning with Big Data: Implementing Spark: PySpark to ML
- Streaming Data in Real Time: Introduction to Kafka and Spark Streaming

Additional Topics:

- NoSQL Database
- Programming in Scala

In the Kaggle survey for 2017 and 2018, over 20% of surveyors have never used any Big Data programs, however, in the job posting, big data is one of the top skills required. Thus, Big Data is included as a mandatory course in the technical pathway.

3.1.4 Professional Communication & Leadership

Course Description: Critical thinking, communication, and leadership all become essential skills of data science jobs. Critical thinking enables us to avoid common obstacles, test our beliefs and assumptions, and correct distortions in our thought processes. Communication allows us to deliver our ideas powerfully in the workplace. Leadership is essential for managing, inspiring, and leading your team to achieve greater success. This course is designed to teach you those communications, leadership, and other soft skills, which are beneficial for your professional career [8].

Course Goals & Learning Outcomes:

- Leadership: You will learn essential leadership skills, including how to inspire and motivate individuals, manage talent, influence without authority, and lead teams.
- Writing & Presentation: You will learn how to communicate effectively in English in professional contexts, by formulating and communicating your thoughts and ideas more clearly, in both written and speaking.
- Decision Making, Critical Thinking, Problem Solving: You will learn how to use analysis, synthesis, and positive inquiry to address individual and organizational problems and develop the critical thinking skills needed in today's turbulent times.

From the dendrogram, in **Appendix C**, of job skills requirement (obtained from indeed), we can see that soft skills are also required by the data science jobs, as well as business and marketing skills, and they are related quite closely.

3.1.5 Business Intelligence

Course Description: This course would provide students the skills they need to become a BI Analyst. The topics would include how to use data for improved business decision-making, present information in the form of metrics, KPIs, reports, and dashboards, and analyze current and historical data to find trends, market conditions, and research competitor positioning. This course would also introduce the students to use Tableau for visualization and dashboard [9].

As Tableau is in high demand for management jobs, we would like to design a course to allow the student to use Tableau and learn the essential skills to become a BI analyst.

3.1.6 Introduction to Deep Learning

Students taking this course are required to have taken Introduction to Machine Learning in a previous term as a prerequisite and the course will be taught using Tensorflow as the core programming framework.

Course Description: This course provides in-depth knowledge of many advanced machine learning algorithms. This course will prepare students to work in the data science and artificial intelligence industry as was shown during our research. Students at the end of the course will be able to work on projects from the healthcare industry, natural language processing, computer vision, object detection. Apart from theoretical background students will also be introduced to many current popular algorithms such as YoloV4, SIFT keypoint detectors, etc. [10].

Learning: Dense Neural Network, Recurrent Neural Network, Convolution Neural Network, Generative Adversarial networks, Autoregressive and Reversible Models, Introduction to Generative Adversarial Network, Auto-Encoder and Decoders, Attention Models

3.1.7 Statistical Analysis

Course Description: Statistical methods are key parts of data science. In this course, students will learn how to utilize data for estimation and assessing theories, construct confidence intervals, interpret inferential results, and apply more advanced statistical modeling procedures. They will then relate the statistics with data science, learn about where data comes from, what types of data can be collected, study data design, data management, and how to effectively carry out data exploration and visualization.

Course Goals & Learning Outcomes:

- Statistics: estimation and assessing theories, construct confidence intervals, interpret inferential results, and apply more advanced statistical modeling procedures.
- Data Analytics: where data comes from, what types of data can be collected, study data design, data management, and how to effectively carry out data exploration and visualization.

Statistical Analysis is one of the top 25 skills taught on Coursera, and it is a common skill used in Data Science and Machine learning. On the dendrogram of job skills requirement (obtained from indeed), **Appendix B**, both statistics and probability are all common skills required by data science jobs. Statistics (and probability) is a key part of data science, where it can be used in exploratory data analysis, bias reduction, estimation of outcomes, etc.

3.1.8 Advanced Data Analysis

Course Description: This course introduces methodologies to represent data in forms suitable for computer processing, particularly in audio and video.

Statistical techniques addressing classification, clustering, and sequence analysis will be addressed. Application of theories will be discussed and study the capabilities to perform complex tasks.

From the Kaggle Survey shown in **Figure 3**, it was found that people would often handle Numerical, Tabular, Time-Series, and Text Data in their daily practice. These data types would be covered in our core course. Besides, about 25% of people would often interact with Image Data, and nearly 10% of people would interact with Video Data. Thus, this course is designed as an elective.

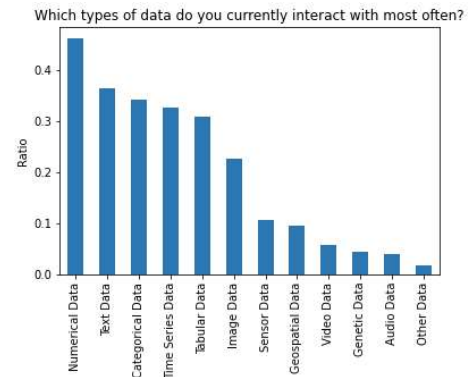


Figure 3. Most Interacted Types of Data, Kaggle

3.1.9 Full Stack Development

Course Description: This course should introduce full-stack app development. This course combines lectures and hands-on open-ended programming assignments to expose the students to the most popular Android app development tools. Previous programming experience is a prerequisite for this course. This course will develop an Android app instead of a website because mobile and wearable devices, combined with web-based software, have given rise to an explosion in creativity and possibility.

Having a full-stack development experience would be precious no matter what kind of technical positions are inquired. An important usage of data science and data analysis is to analyze user data and improve the application's performance, which is a part of full-stack development. Most of the full-stack development (especially mobile app development) related skills are shown in the top 20 of the required skills listed in job postings, **Appendix A**.

3.1.10 Introduction to Project Management in Data Science

Course Description: This course will introduce the students to key functional areas of management and the typical decisions they face. It will illustrate how each functional area approaches some common managerial problems, and where data and analytics may be usefully employed. It's designed to ensure all students from all backgrounds or experience are equipped to place the management of AI, data science, and data analytics in the context of general business management [11].

Based on the web-scraped data for technical skills, in particular management skills, communication and basic management skills are among the topmost valued skills in the industry as shown in **Figure 4**. Critical thinking, writing, and leadership skills are also among the top 5 skills. Therefore, this course is an elective course to expose interested students to management skills that would equip them with valuable skills to be successful in the industry.

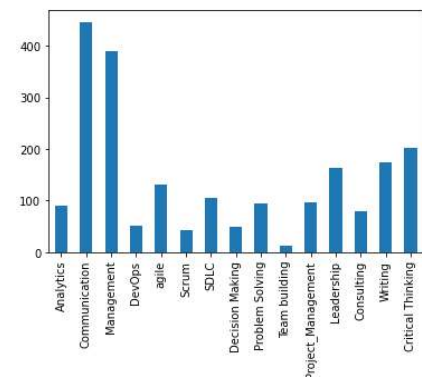


Figure 4. Top Technical Skills in DS Industry

3.1.11 Data Science Programming Basics

Course Description: This course is designed for students who entered the program with none or very minimal background in data analysis. The course is a practical course that introduces basic data analysis skills and tools including Python (Numpy, Pandas), Excel (VBA), and SQL. The first topic will be Python as it is widely used in other courses.

From the skills gained from the job posting as shown in **Figure 2**, it is seen that Python, Excel, and SQL are the top ones. We understand that most students entering this master's program have learned these skills before. However, we believe that students with a bachelor's in management or art faculties would be also interested in coming to our master's program, therefore this course is offered as an elective.

3.1.12 Introduction to Computational Finance

Course Description: The objective of the course is to introduce students to the use of mathematical models and numerical techniques in finance. Thus, this course will cover both financial economics topics and relevant mathematical/statistical topics [12].

Based on the skills that occurred in the job postings, Finance is one of the top skills needed in the market. Even though not everyone would work in the Financial Sector, we believe having a Finance course as an elective is necessary.

3.1.13 Operations Research

Course Description: Operations Research is a well-defined field of study involving mathematics, computer modeling, and optimization. It is one of the various disciplines used in the emerging interdisciplinary field of study of Data Science. This course aims to introduce optimization techniques applicable to solving various engineering problems. These techniques are widely used in engineering design, optimal control, production planning, reliability engineering, and operations management [13].

Operations (Research) is one of the common skills in data science jobs as shown on the dendrogram in **Appendix B**. Python or other computer languages can be useful tools for creating mathematical models and solving optimization problems. However, it was not a popular skill for all data science jobs, hence made as an elective option.

3.2 Capstone Project Option

Students who choose the project option need to consult with their advisors on the project topics and scopes. The project should allow students to apply a large range of skills they have learned in the courses. **Table 2** includes two types of sample capstone projects that students could potentially pursue.

Table 2. Sample Capstone Project Recommendations

Capstone Project 1	Capstone Project 2
Project Description Design a system to analyze customer’s reviews on products/services.	Project Description Design a system to track, analyze, and predict the stock market price.
Project Goals and Outcomes <ul style="list-style-type: none"> • Able to identify the sentiment score (positive, negative) based on words used. • Able to extract keywords (reasons for complaints/suggestion for growth) from customer’s reviews. • Able to detect fraudulent reviews. 	Project Goals and Outcomes <ul style="list-style-type: none"> • Able to track the stock market price. • Analyze the trend of some selected stocks. • Able to make predictions on future prices.
(Possible) Involved Techniques <ul style="list-style-type: none"> • Web scraping • Natural language processing • Sentiment analysis • Model creating and tuning 	(Possible) Involved Techniques <ul style="list-style-type: none"> • Web scraping • Model creating and tuning

3.3 Internship Option

With the 2018 Kaggle Survey, we found that the industries that offer the most data-related jobs are Computer/Technologies, Academic/Education, and Accounting/Finance, which is shown in **Appendix G**. Students who want to work in the Academic/Education Industry would be likely to find research opportunities in their school. Thus, we would focus on Computer/Technologies and Accounting/Finance Industries.

During our research, we found 187 internship options in various positions that were related to the Data Science program. Internships will help students to put their skills into practice. **Table 3** provides a list of companies that we can approach to help students to find their internship jobs. The data in **Table 3** is based on the current job postings. To make our program grow, we will continue monitoring the job market and expanding the lists of companies we can approach for internships.

Table 3. Companies & Industries for Internship Opportunities

Industry	Companies
Computer/Technologies	Facebook, IBM, South Hound, Celestica
Accounting/Finance	Clearbanc, RBC, Scotiabank
Other	Halifax Regional Municipality, Sunrise International Inc., QuadReal

4.0 EdTech Startup Ideation

From our research, we came to know that different course designs like Introduction to Machine Learning, Big Data, Statistical Analysis, Advance Data Analysis, etc. are currently in high demand which the employers are seeking for. We already designed some course structures based on our research, but the problem is how to provide easy access to the upcoming students to understand the need of the market and the courses they should opt for. Therefore, we are suggesting a recommendation system on a website platform that will provide an easy understanding of the fast-changing data science jobs in the market, which can be referred to Appendix H.

Key Points:

- Initially, it will check the student's current skill set; background in data science, analytics, finances, management, etc.
- It will then ask for the future pathways the student wishes to pursue; either technical or business pathways.
- Based on the prior selections (current skill sets and future pathways), the system will recommend the courses which are being offered by University or any third-party organizations such as Data camp and Coursera.
- The list of the courses offered by the university which are predefined based on our research for the current market can be updated every year based on the requirements.
- After the student selects any “X” number of courses, then our recommendation system will define a “Y” number of skillsets for the selected courses.
- Based on the “Y” skillsets, the recommendation system will search the web from the different job-seeking websites like LinkedIn, indeed, glassdoor, etc., and give the student a probable number of jobs based on the prior selections.

Advantages:

- Courses offered will be based on current and required skill sets.
- Easy connection between the job opportunities and the offered course option without referring to multiple resources.
- Convenient, time-saving, and easy to use.
- Data set for the courses can be updated by the university or any third-party organization which will make sure that the recommendation system is always up to date.

5.0 Conclusion

The main objective was to design a new Master of Data Science and Artificial Intelligence program that is business-oriented and update the current MIE1624 curriculum. Data retrieval and analysis were performed by web scrapping job descriptions on Indeed and LinkedIn with job titles such as Data Analyst, Data Scientist, and Data Engineer. The skills were analyzed according to technical skills and business skills and research the top skill of each type. Analysis of the skills, along with hierarchical clusters for technical and management roles was created which are located in **Appendices**. Additionally, the 2018 and 2019 Kaggle surveys were also used to study the tools and techniques used by data scientists in the industry. Upon further research, practical experiences were important to employers, thus, the inclusion of a capstone project or internship option in the last semester of the program. Moreover, programs and courses offered by other platforms such as Coursera were analyzed as well as similar programs at other universities.

The main changes in the MIE1624 curriculum will be a shift in focus from deep learning algorithms to the development of Python basics and linear algebra. There will also be more marks allocated to technical presentations and reports for assignments and projects as communication and management skills are important to employers in the data science industry.

The program will require students to take 7 courses plus a capstone or internship option in their last semester. Students can choose between two streams, technical or business. In the first two semesters, students will take 3 courses each, 3 core courses specific to the stream, and 3 electives of their choice.

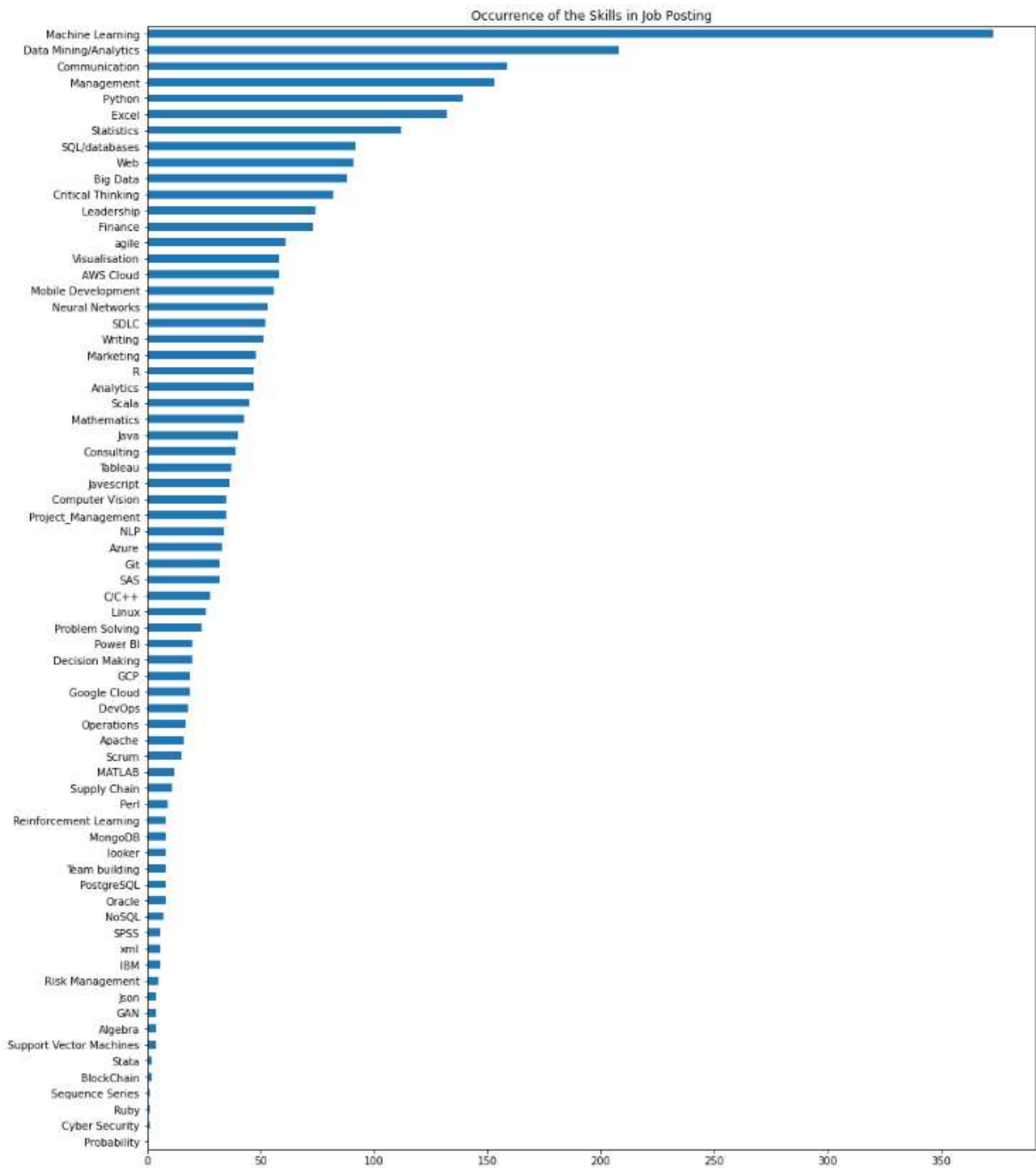
Companies such as Facebook, Clearbanc, QuadReal, and numerous others were found to offer the most opportunities for internship positions, thus will be approached by students. The industries with the most data science-related jobs are computer/technologies, academic/education, and accounting/finance. Such that, courses are meant to be structured to give students an optimal advantage over other students in other programs.

An EdTech Startup ideation is proposed to provide future students of the program the ability to understand the need of the market and courses they should take accordingly. The key ideas of the proposal include checking the current skillset and prior knowledge of the student and then learning about the pathway they want to pursue. Upon that information, the system will recommend courses offered by universities or third part teaching platforms to match the current industry needs.

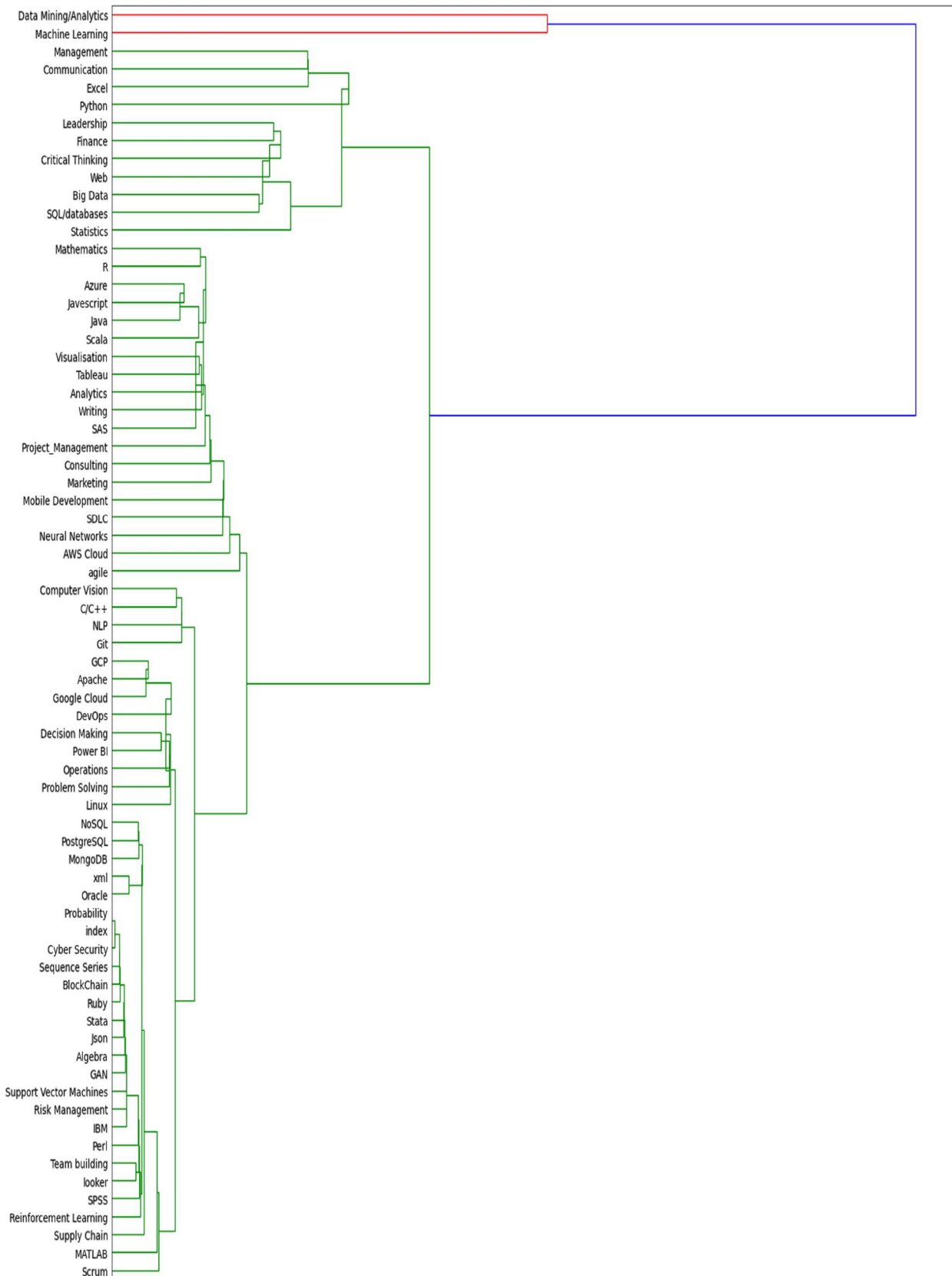
6.0 References

- [1] Kaggle, "2018 Kaggle ML & DS Survey," [Online]. Available: <https://www.kaggle.com/kaggle/kaggle-survey-2018>.
- [2] Kaggle, "2019 Kaggle ML & DS Survey," [Online]. Available: <https://www.kaggle.com/c/kaggle-survey-2019>.
- [3] Coursera. [Online]. Available: <https://www.coursera.org/>.
- [4] A. Romanko, "MIE-1624HF: Introduction to Data Science and Analytics," [Online]. Available: <https://www.mie.utoronto.ca/wp-content/uploads/2019/08/MIE1624H-Course-Outline-Fall-2019.pdf>.
- [5] Huang, "CSC2515 Fall 2020: Machine Learning," [Online]. Available: https://www.cs.toronto.edu/~huang/courses/csc2515_2020f/Lectures.html.
- [6] Udemy, "The Ultimate Hands-ON Hadoop: Tame your Bid Data," [Online]. Available: <https://www.udemy.com/course/the-ultimate-hands-on-hadoop-tame-your-big-data/>.
- [7] Coursera, "Big Data Specialization," [Online]. Available: <https://www.coursera.org/specializations/big-data#courses>.
- [8] Coursera, "Leading People and Team Specialization," [Online]. Available: <https://www.coursera.org/specializations/leading-teams>.
- [9] Udemy, "The Business Intelligence Analyst Course 2020," [Online]. Available: <https://www.udemy.com/course/the-business-intelligence-analyst-course-2018/>.
- [10] Coursera, "Deep Learning Specialization," [Online]. Available: <https://www.coursera.org/specializations/deep-learning>.
- [11] Udemy, "The Complete Data Science Project Management Course," [Online]. Available: <https://www.udemy.com/course/data-science-project-management/>.
- [12] A. Romanko, "MIE 1622H: Computational Finance and Risk Management," [Online]. Available: <https://www.mie.utoronto.ca/wp-content/uploads/2019/08/Romanko-MIE1622H-Outline-2020.pdf>.
- [13] U. o. Toronto, "APS1005: Operations Research for Engineering Management," [Online]. Available: https://gradstudies.engineering.utoronto.ca/files/2012/09/APS1005-Opertations-Reseach-for-Engineering-Management_updated.pdf.

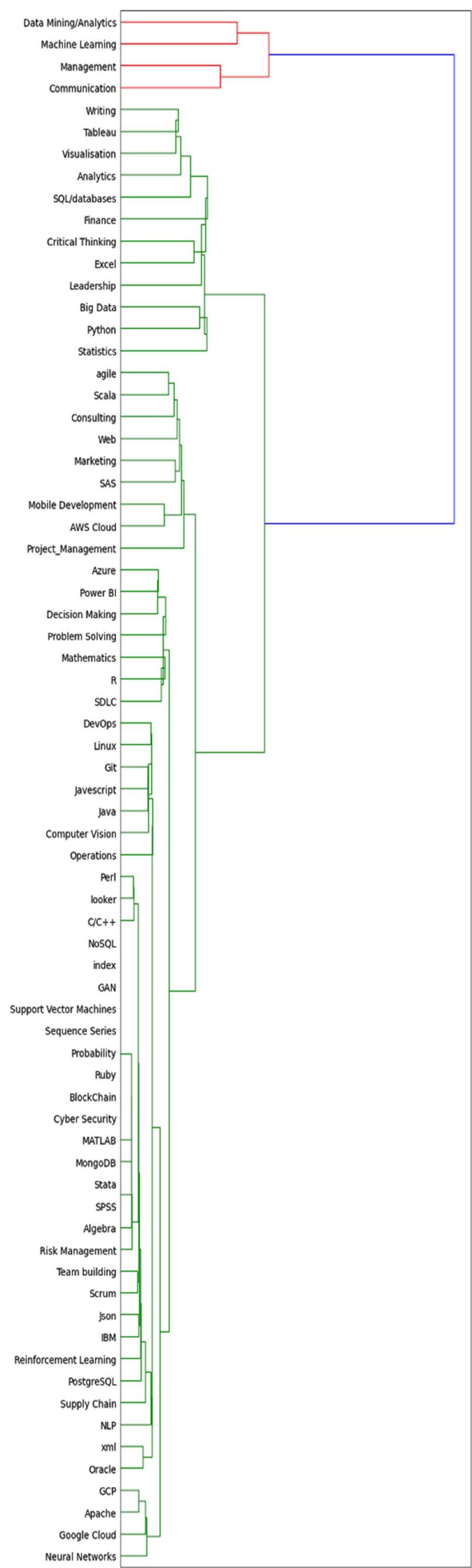
Appendix A: Occurrence of the Skills in Job Postings



Appendix B: Hierarchical Clustering for Skills Required for Technical Roles

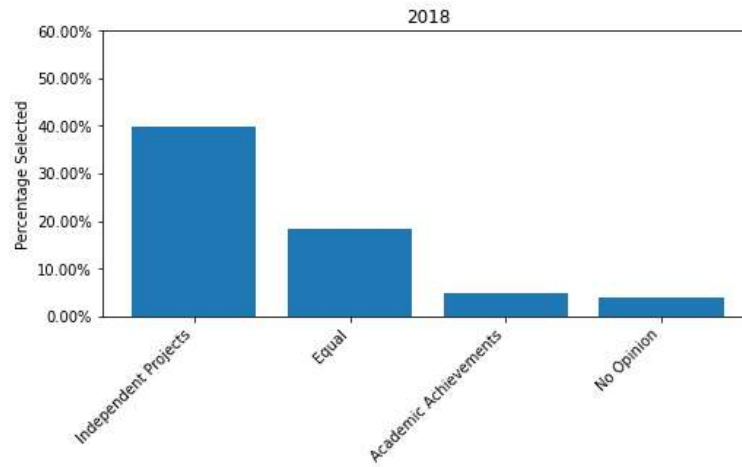


Appendix C: Hierarchical Clustering for Skills Required for Management



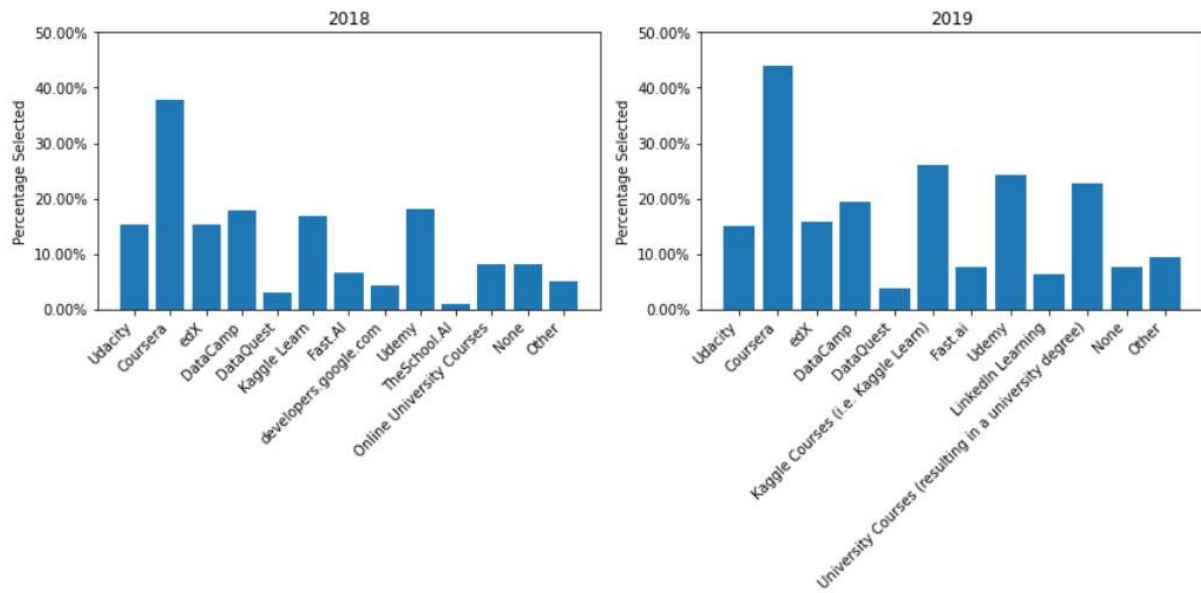
Appendix D: Industry View on Academic Achievements vs. Independent Projects

Which better demonstrates expertise in data science: academic achievements or independent projects?

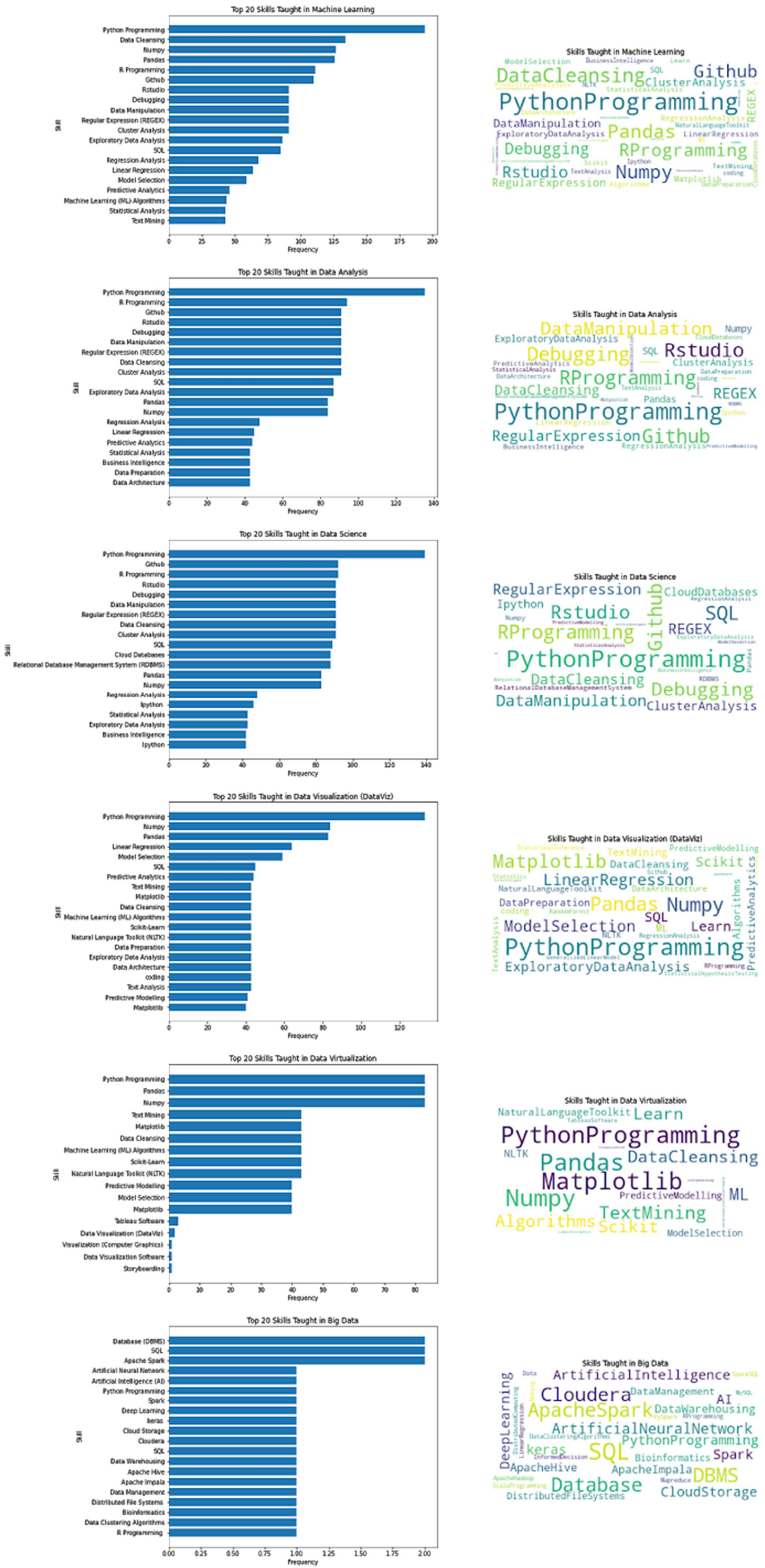


Appendix E: Most Popular Platforms used for Data Science Related Courses

On which platforms have you begun or completed data science courses?

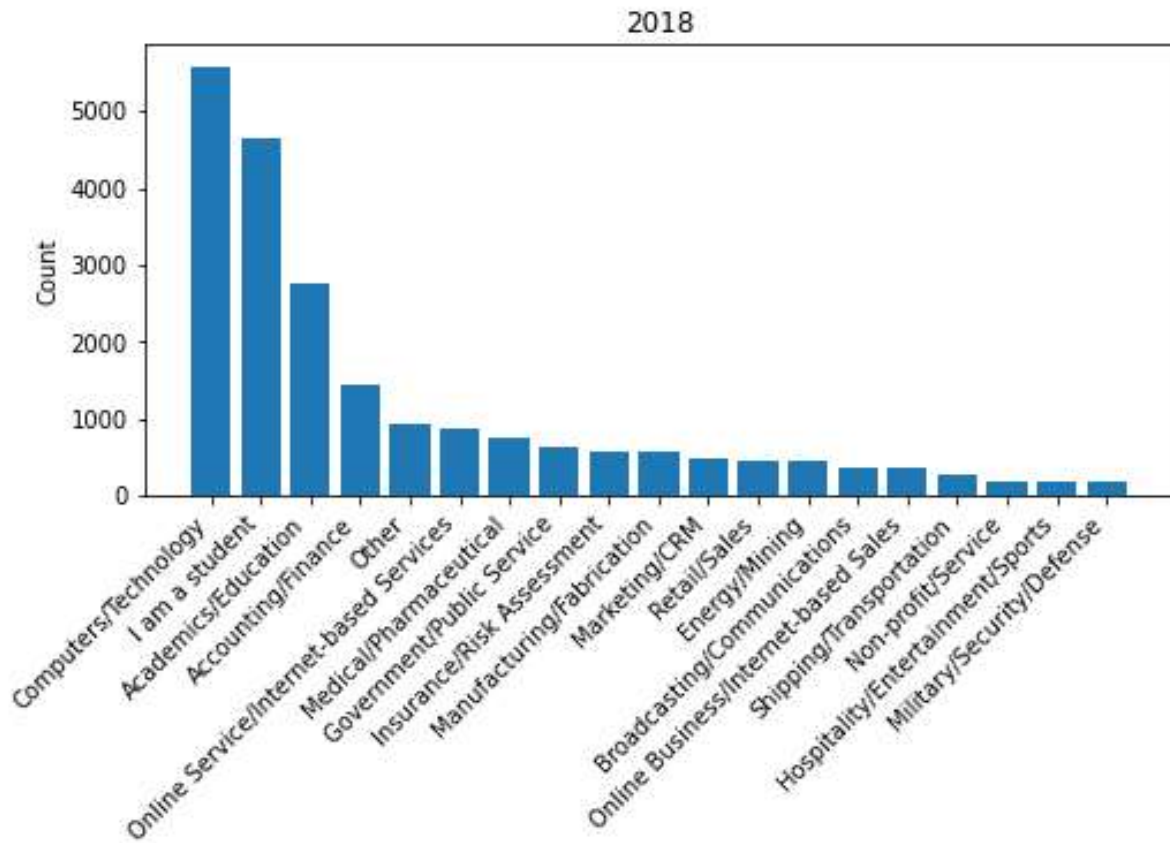


Appendix F: Most Common Skills Taught with Each Topic in Coursera



Appendix G: Number of Survey Participants in Each Industry

In what industry is your current employer/contract?



Appendix H: EdTech Startup Ideation Flow Chart

