MIE1624 Course Project

Master of Data Science and Artificial Intelligence
Program Design

Group 19

Hao Liu 1006864163
Junyu Jiang 1002916561
Keyang Zhao 1003866649
Tongyuan Zhu 1003944388
Xinyi Yao 1005217766
Yijia Gao 1002717241
Zhiting Zhang 1002955064

Contents

Introduction

Background

Data science is one of the most popular areas of study nowadays, and demand for data analyst, scientist, engineer, and data manager on major job-posting websites has been rising for years. Therefore, many educational institutions construct and update their data analysis program frequently to cope with the change of the times.

Project Overview

University of Toronto is now re-design the course curriculum for “MIE1624: Introduction to Data Science and Analytics”, in which students can use various types of machine learning algorithms to solve business problems. We will illustrate the redesign of “MIE1624: Introduction to Data Science and Analytics" in part 1. A curriculum for a new program named “Master of Data Science and Artificial Intelligence” will be introduced. In part 2, the details of program structure and course descriptions are specified. The visualization is created in order to show the details of major course design in part 3. Finally, a recommender system is designed for students to personalize their study. To further improve the learning experience of users, DataPro APP is designed to assist student’s course selection and track their assignments.

Part 1 - MIE 1624 Redesign

Problem Definition

MIE1624 is an introductory course of data science and analytics. Data science is an interdisciplinary field focused on extracting knowledge from data sets, which are typically large and applying the knowledge and actionable insights from data to solve problems in a wide range of application domains. The objective of this course is to prepare the students with the fundamental skills of data science and analytics for future study and career. The goal of re-design this course is to equip the students with the most relevant technical and non-technical skills in data science.

Data Source

To redesign the course MIE1624, the Kaggle Machine Learning & Data Science Survey 2021 was used as our resource since this survey presents a truly comprehensive view of the state of data science and machine learning which is related to our course. There are 42 questions in total. The team has analyzed 12 questions which seem to be most relevant to the topic of MIE1624 and the skills that a data scientist needs. The team only focused on the answers from the participants who have professional experiences on data analysis or machine learning. The count of each response in question is calculated to determine the most important skills and techniques.

In addition, similar courses to MIE1624 were found on platforms such as Coursera, WeCouldData and some other universities' websites (e.g., McGill University, UBC, Stockholm University, etc.). Skills that are listed on their course outline and description were used as reference for our course redesign.

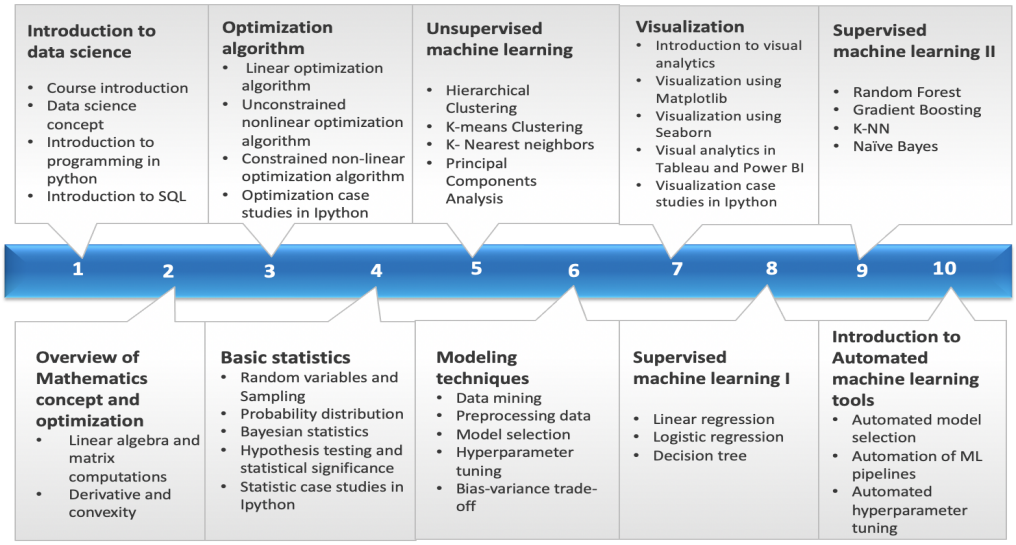
Analysis

Figure 1 shows the distributions of participants' responses to the 3 survey questions. According to the pie chart one the bottom left corner, the most popular computer programming languages are Python, SQL and R. Thus, MIE 1624 will use Python as its core language and all the course projects are required to write in python. Furthermore, the course has one module that includes an introduction on SQL. The top pie chart suggests that the most frequently used machine learning algorithms are linear or logistic regression, decision trees or random forest and Gradient Boosting Machines. The final pie chart shows that Matplotlib and Seaborn are the two most frequently used visualization libraries. It is not surprising since they are Python libraries and Python is the most popular programming language in Data Science.

The team has also created heatmaps describing the relation between job titles and skills and tools. Figure 2 represents 4 survey question responses and their significant level with each job title. The upper left heatmap illustrates that the Jupyter notebook is used by most of the data relative jobs except Statistician and Developer relations/Advocacy. According to this founding, MIE 1624 instructor should suggest students to use Jupyter notebook as their integrated development environments. The heatmap on the upper right corner suggests that the machine learning framework that people in the data science industry use most frequently is Scikit-learn. To make data analysis simpler, the team would like to add some automated machine learning tools to the course module. From the fourth heatmap, the team constructed automated model selection automation of ML pipeline and automated hyperparameter tuning are three most frequently used automated machine learning tools.

Result

In the MIE 1624 course, many machine learning algorithms will be covered, such as linear, logistic regression, decision tree, random forest, gradient boosting etc. Students will also learn other skills like data mining, visualization, and hyperparameter tuning. This course will also include 3 individual assignments and 2 group projects since most of the data relevant jobs will face the situation of working in a group. All the course projects are required to write in python and the suggested integrated development environment is Jupyter Notebook. All assignments should be submitted through GitHub. The course outline is shown below:



Part 2 - Data Science program curriculum design

Methodology:

In the process of designing the new program: Master of Data Science and Artificial Intelligence, we use below main methodologies.

Program curriculum and structure design: The program curriculum and structure design is based on the significant skills identified from the skill frequency using indeed.com, the insights gained from the kaggle dataset used in part1 and the reference to the other professional programs worldwide.

In order to design this technically and business-oriented program, we need to identify the significant relevant technical oriented skills and business-oriented skills. For identifying significant skills, we firstly create relevant three skill lists based on our recognition toward data professionals, which are coding & software skill list, the professional skill list and the soft skill list. The technical - oriented skills are included in the software & coding and professional skill lists, and the business - oriented skills are included in the soft skill list. Then we apply NLP techniques to calculate the frequency of skills in each skill list based on the job descriptions of data analyst, data scientist, data engineer and data manager from indeed.com. The relevant data results are shown in the appendix. Please refer to the **appendix B - part 2**, **Section1: The relevant results of Skill frequency for four data-related positions respectively** and **Section2: The relevant results of skill frequency for four data-related positions aggregately.**

Based on the skill frequency in three lists, we identify the significant technical oriented skills and business-oriented skills shown below:

Technical – oriented skills:

Coding & Software Skills: Python, SQL, R, Tableau, Cloud.

Professional skills: data analysis, machine learning, database management, data modeling, data visualization, statistics, mathematics, data mining.

Business – oriented skills:

Business decision making skills, Business intelligence, Business analytics, Business management, leadership skills, communication skills, teamwork, writing skills, research skills, presentation skills

Besides the above methodology, to pursue the unbiased and detailed course design, we also conduct the deeper research into data-related courses, certificates, and master's programs from other universities around the world, including McMaster University, Northeastern University, University of Chicago, etc.

Program Overview - Master of Data Science and Artificial Intelligence

In the era of big data, most enterprises are using big data. Every link of the data industry needs to be completed by professionals. However, the ability to uncover business insights based on data is highly specialized skills processed by too few people. The supply of data professionals who can derive business insights and make informed decisions from data is far from meeting the market demand.

Master of Data Science and Artificial Intelligence is a 16 - months full-time program designed to address this workforce gap by providing theoretical as well as practical learning, committed to help them become

excellent data professionals. Through learning in the program, the students can equip themselves with the important technical skills, business skills, practical experience in the data field and most importantly, the confidence to seize opportunities in an ever - expanding field.

This program is a career - oriented program. In order to enable students to find the job they particularly want, according to the different requirements of positions in the data field, the program also sets up four specializations, which are data analysis specialization, data science specialization, data engineering specialization and data management specialization. Each specialization will set different core courses and elective course recommendations, which makes students learn relevant skills more targeted according to their career goals and enhance their future workplace competitiveness pertinently. The recommended core courses and elective courses for each specialization are shown in the appendix. Please refer to the **appendix A**, **Section1: Recommended core and elective courses for each specialization.**

The courses for the program are listed below. The course descriptions of core courses are shown in the appendix. Please refer to the **appendix A**, **Section2: The descriptions of courses offered by the program.**

Introductory courses:	
Programming for Data Science	Statistics for Data Science

Core Courses:	
Technical – Oriented Courses:	**Business – oriented courses:**
• Machine Learning with Applications In Python	• Business Leadership and Communication
• Data Modeling and Database Management	• Data Mining for Business Applications
• Data Analytics: Methods and Practical Approaches	• Business Problem Analysis and Management
• Introduction to Artificial Intelligence	
• Applications with Deep Learning	

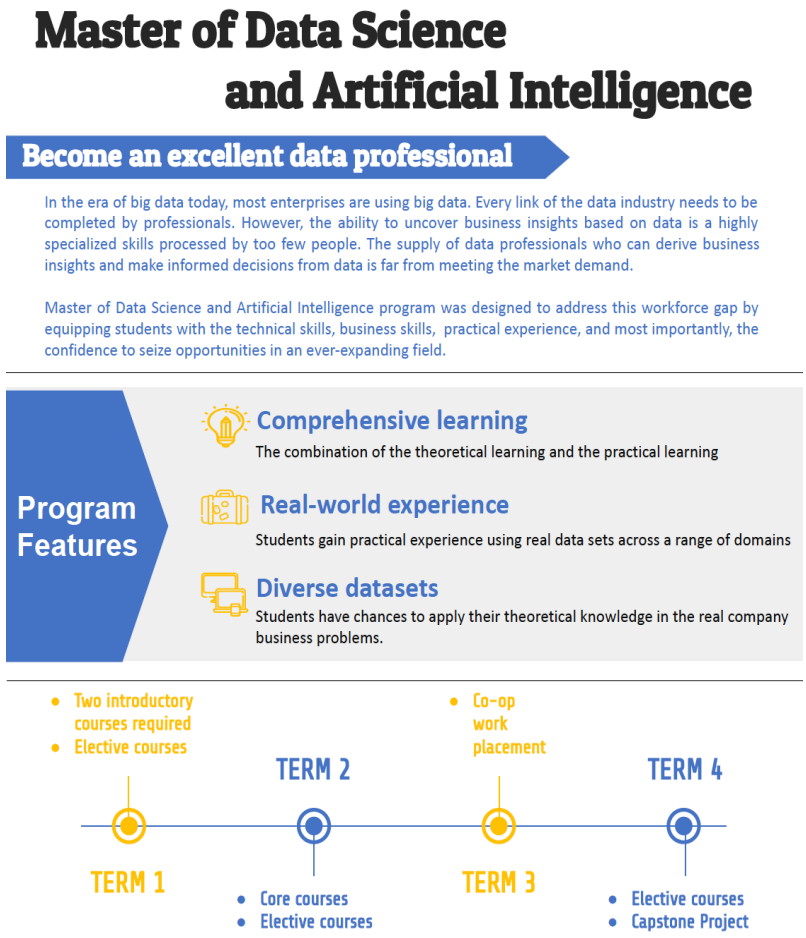
Elective Courses:		
Technical – oriented Courses:		**Business – oriented Courses:**
• Modeling Tools for Predictive Analytics	• R for Data Science	• Business Immersion
• Experimental Design for Data Science	• Optimization Techniques	• Analytics in Management
• Storytelling with Data using Tableau	• Stochastic Modeling	• Supply Chain Management
• Research Data and Research Operations	• Reinforcement Learning	• Data Driven Investments
• Forecasting and Time Series Analytics		• Simulation and Risk Analytics
• Cloud Technology in Data Science		• Analytics for Marketing Strategy

To complete the program studies and gain the degree, the students must complete up to 7 credits for 16 months, which include 2 introductory courses, 4 core courses for their respective specialization, one Co-op work term (the co-op work term accounts for 1 credit) and 6 elective courses (or 4 elective courses + capstone project).

The program holds the activities including Co-op program, Workshops, Capstone Project and provides the services including the writing center. The descriptions of each activity are shown in the latter visualization in part 3 and the appendix. Please refer to the **appendix A**, **Section3: The descriptions of the activities & services offered by the program.**

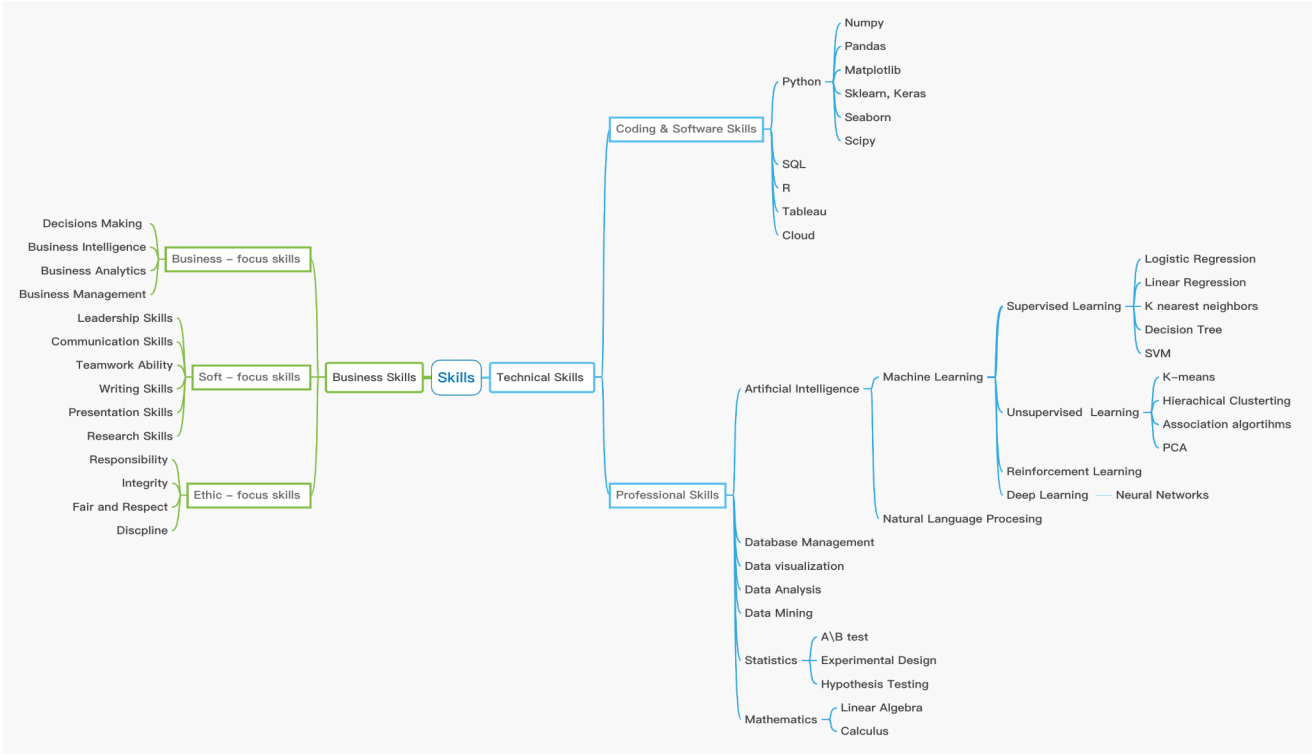
Part 3 - Visualization of Course Curriculum

Below is the visualization of the whole program - Master of Data Science and Artificial Intelligence. This clearly displays the essential components of our program, the characteristics of our program and the purpose plus vision of our program.

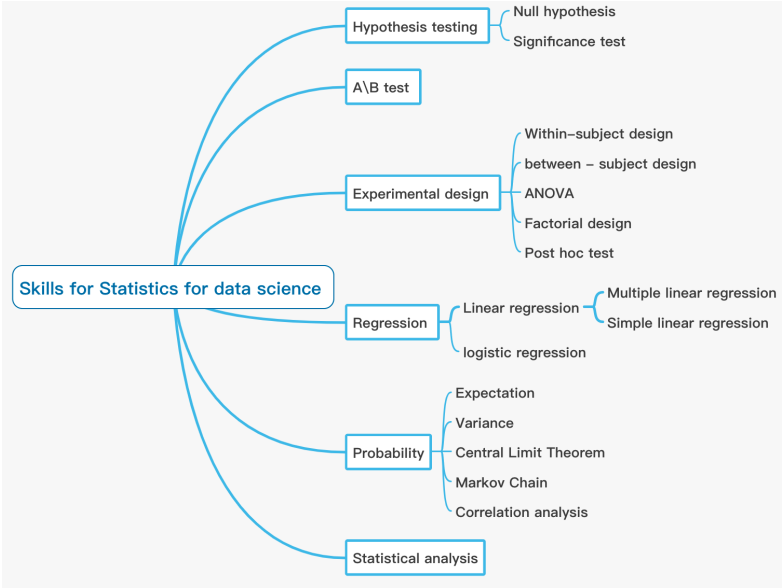


Below are shown the Visualization of skills learned over the whole program and visualization of skills for the selected 6 courses provided by the program. The skills to be learned identified for each course provided by the program (e.g., All core courses and elective courses in the program) are shown in the appendix. Please refer to the **appendix A**, **Section4: List of skills for <u>core courses</u> and <u>elective courses</u> in Master of Data Science and Artificial Intelligence**.

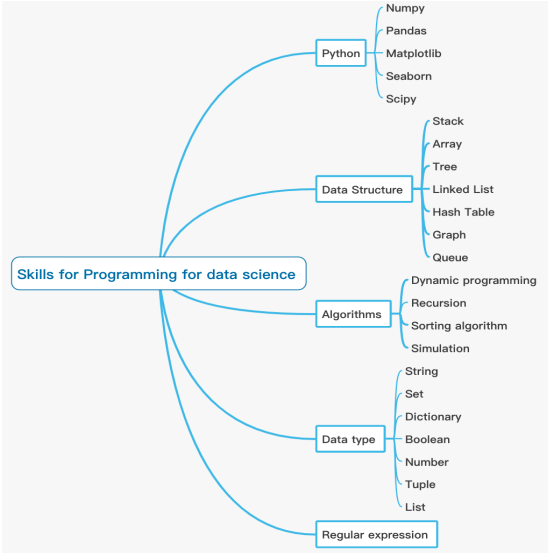
Tree Diagram1 : Visualization of skills learned over the program



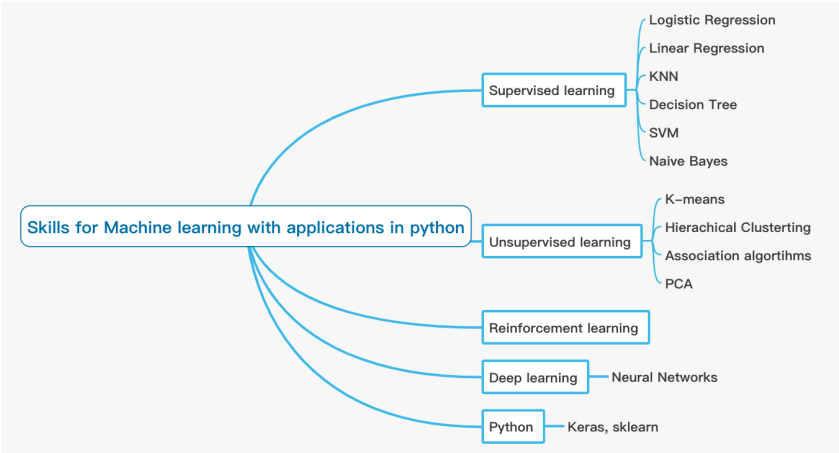
Tree Diagram 2: Visualization of Skills for Course 1 - Statistics for Data Science



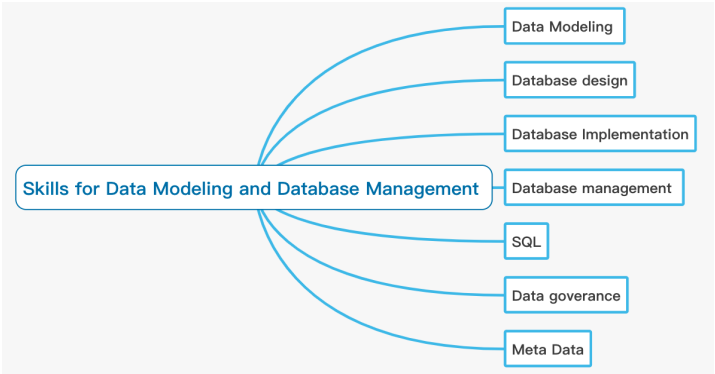
Tree Diagram 3: Visualization of Skills for Course 2 - Programming for Data Science



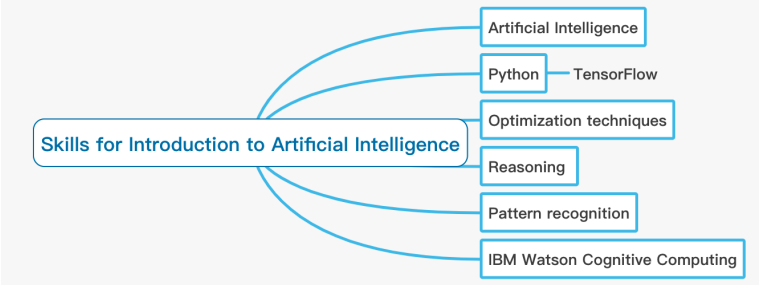
Tree Diagram 4: Visualization of Skills for Course 3 - ML with Application in Python



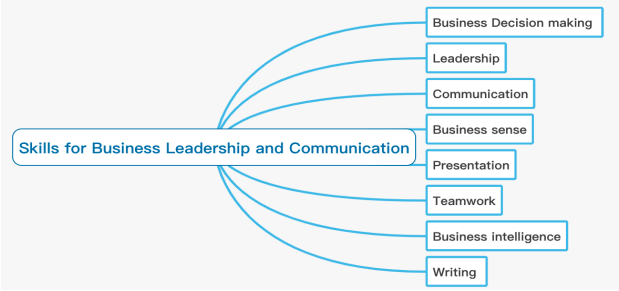
Tree Diagram 5: Visualization of Skills for Course 4 - Data Modeling and Database Management(SQL)



Tree Diagram 6: Visualization of Skills for Course 5 - Introduction to Artificial Intelligence



Tree Diagram 7: Visualization of Skills for Course 6 - Business Leadership and Communication



Part 4 – Recommender system for personalized study

When we enroll in the program to study data science, we should always ask ourselves some questions: What kind of data careers should I choose? What courses should I take? In order to solve these problems to have a better learning experience, a recommender system is designed for students to personalize their study.

First, students need to select a career path or field that they want to get into. Then, the system will recommend 4 core courses for each career path. These core courses are fixed for each path and are designed according to the work done in part 2 and part 3. For example, if a student chooses data analyst as his/her future career, the core courses he/she needs to take are Course 4,5,8 and 9.

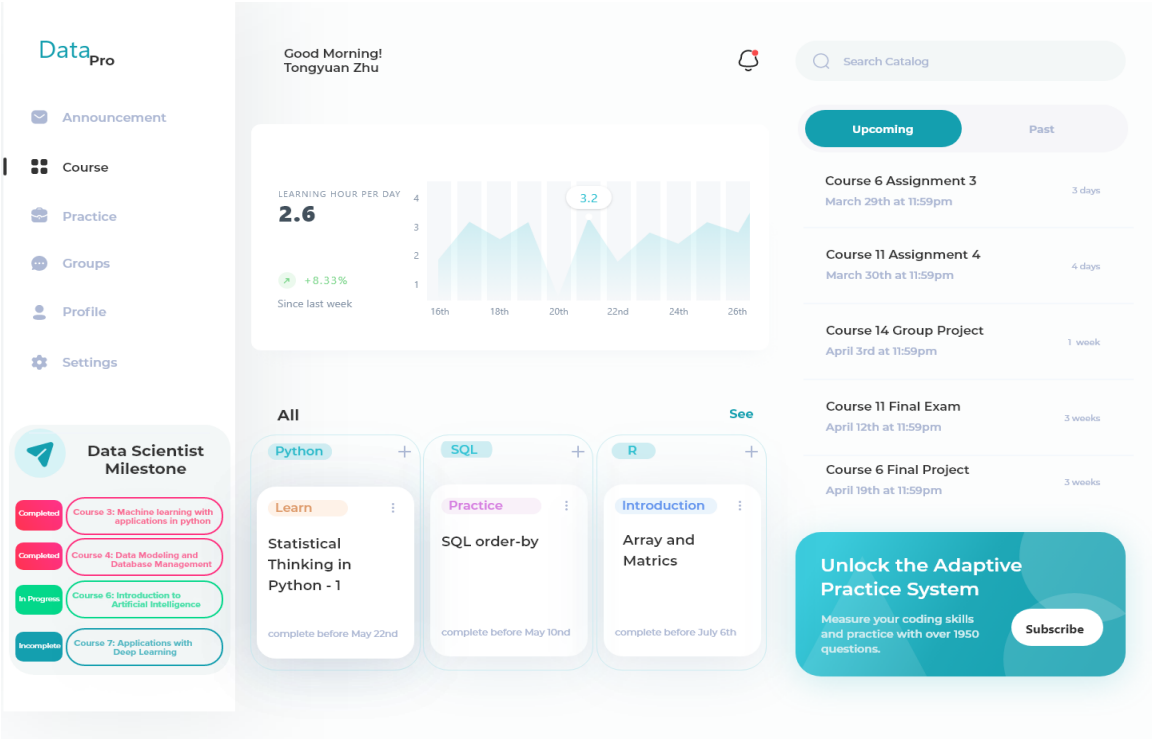
After that, an elective course recommending system is designed based on the core courses the students choose. The logic of designing the system is to use content based method and cosine similarity as the parameter to find the most related elective course based on the core courses. The dataset we choose is the course names of the elective courses and the skills they will teach. During data preprocessing, we try to eliminate all of the punctuations and capital letters. Also PorterStemmer is used for the stemming process and the words in course name and skills are vectorized for applying cosine similarity test. The input of this system is the course name and description of one of the core courses, and the output is 3 recommended elective courses that are related to the core course. Moreover, students can fill in their preferred keywords, such as the skills they are willing to learn, into the system to get the recommended elective courses.

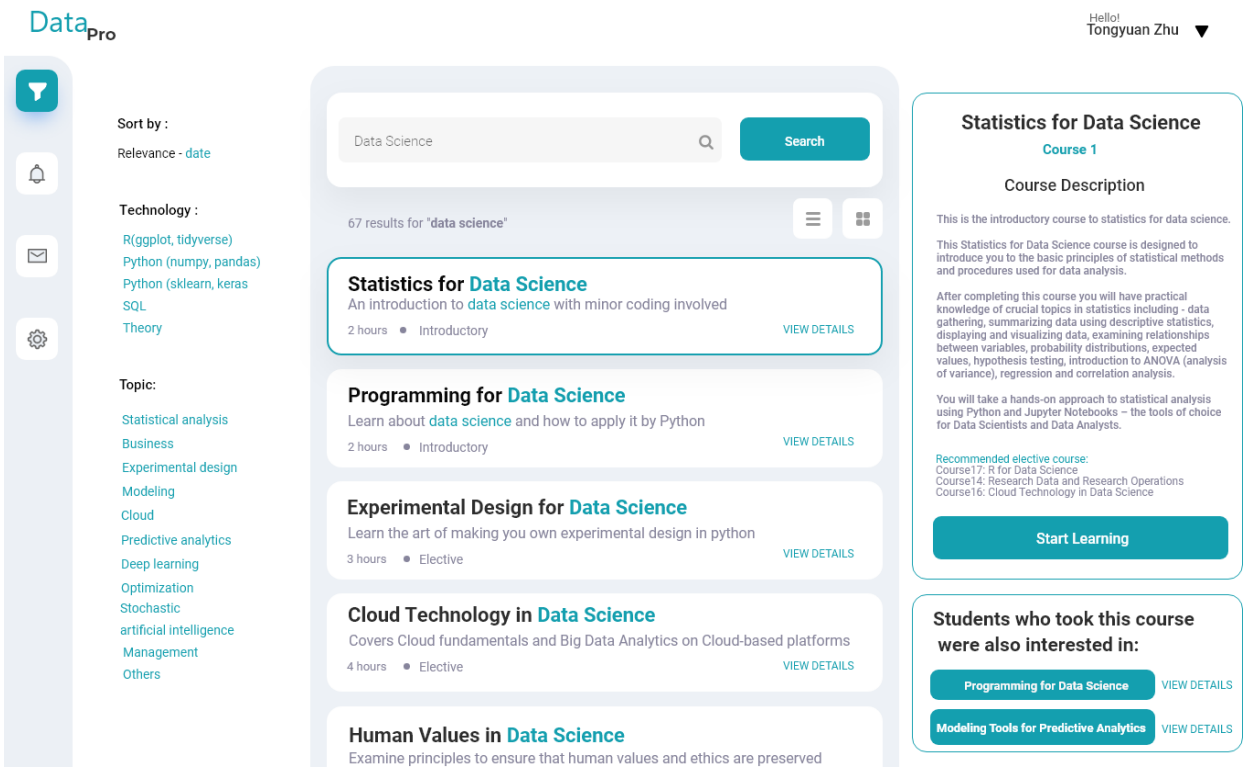
The code is shown in Part 4 code.ipynb.

In order to create a good user experience for students who enroll in this program, DataPro App is designed to assist students to choose courses through the recommender system mentioned above. The below two figures show the interfaces of the main page and search page of the app.In order to create a good user experience for students who enroll in this program, DataPro App is designed to assist students to choose courses through the recommender system mentioned above. The below two figures show the interfaces of the main page and search page of the app.

The first figure is the main page of the app, which shows some general information about the program such as studying hours and studying process.

The second figure shows the search page of the app which integrates the course recommending system. Recommended elective courses for each core or introductory course are shown under course description. Also, students can type keywords in the search bar if they have certain preferences when selecting elective core courses.




Data Pro
Hello!
Tongyuan Zhu
Sort by :
Relevance - date
Technology :
R(ggplot, tidyverse)
Python (numpy, pandas)
Python (sklearn, keras
SQL
Theory
Topic:
Statistical analysis
Business
Experimental design
Modeling
Cloud
Predictive analytics
Deep learning
Optimization
Stochastic
artificial intelligence
Management
Others
Data Science
Search
67 results for "data science"
Statistics for Data Science
An introduction to data science with minor coding involved
2 hours • Introductory
VIEW DETAILS
Programming for Data Science
Learn about data science and how to apply it by Python
2 hours • Introductory
VIEW DETAILS
Experimental Design for Data Science
Learn the art of making you own experimental design in python
3 hours • Elective
VIEW DETAILS
Cloud Technology in Data Science
Covers Cloud fundamentals and Big Data Analytics on Cloud-based platforms
4 hours • Elective
VIEW DETAILS
Human Values in Data Science
Examine principles to ensure that human values and ethics are preserved
Statistics for Data Science
Course 1
Course Description
This is the introductory course to statistics for data science.
This Statistics for Data Science course is designed to introduce you to the basic principles of statistical methods and procedures used for data analysis.
After completing this course you will have practical knowledge of crucial topics in statistics including - data gathering, summarizing data using descriptive statistics, displaying and visualizing data, examining relationships between variables, probability distributions, expected values, hypothesis testing, introduction to ANOVA (analysis of variance), regression and correlation analysis.
You will take a hands-on approach to statistical analysis using Python and Jupyter Notebooks – the tools of choice for Data Scientists and Data Analysts.
Recommended elective course:
Course17: R for Data Science
Course14: Research Data and Research Operations
Course16: Cloud Technology in Data Science
Start Learning
Students who took this course were also interested in:
Programming for Data Science
VIEW DETAILS
Modeling Tools for Predictive Analytics
VIEW DETAILS

Reference:

https://masterdatascience.ubc.ca/
https://www.sauder.ubc.ca/programs/masters-degrees/mban/curriculum
https://ischool.utoronto.ca/
https://uwaterloo.ca/data-analytics/deep-learning
https://www.coursera.org/learn/predictive-modeling-analytics
https://ischool.utoronto.ca/course/experimental-design-for-data-science/
https://continuingstudies.mcgill.ca/search/publicCourseSearchDetails.do?method=load&courseId=6666007&mc_cid=f9d9f6e8f4&mc_eid=[UNIQID]
https://www.coursera.org/learn/data-management
https://www.rotman.utoronto.ca/Degrees/PhD/Academics/MajorAreasofStudy/OperationsManagementandStatistics
https://www.coursera.org/lecture/demand-prediction-using-time-series/arima-models-7UQJG
https://learn.utoronto.ca/programs-courses/certificates/cloud-computing?gclid=CjwKCAjwiuuRBhBvEiwAFXKaNH98EKPXzCB9buvUv85y7J5MAGlz3QJIAHxExFTmDfgkApWIBBUL7hoCSIEQAvD_BwE
https://www.datacamp.com/courses/free-introduction-to-r
https://www.datacamp.com/tracks/r-programming?irclickid=wD4U0tyCNxyITb2w14VOxwykUkGQ3SR3UU6gT80&irgwc=1&utm_medium=affiliate&utm_source=impact&utm_campaign=357605&
https://online.stanford.edu/courses/mse221-stochastic-modeling
https://cse.buffalo.edu/~avereshc/rl_spring20/Reinforcement_Learning_Course_Syllabus_2020_Updated.pdf
https://www.ischool.berkeley.edu/courses/datasci/231
https://www.ischool.berkeley.edu/courses/datasci/231
http://blogs.ubc.ca/comm120businessimmersion/files/2015/08/COMM-120-Course-Outline-Jan-2016-FINAL-updated-30Mar2016.pdf

https://cisl-onlinecourses.getsmarter.com/presentations/lp/university-of-cambridge-sustainable-supply-chain-management-online-short-course/?ef_id=c:433334258859_d:c_n:g_ti:kwd-391954643477_p:_k:%2Bsupply%20%2Bchain%20%2Bmanager%20%2Bcourse_m:b_a:82241924091&gclid=CjwKCAjwiuuRBhBvEiwAFXKaNOJtUlLW2rw64ur-cSCna9Sz2E251jQZJEb-qb9L9JlAK1AQ1VGaMRoCGkoQAvD_BwE&gclsrc=aw.ds
https://gradstudies.engineering.utoronto.ca/files/2020/05/APS1017-Course-Outline.pdf
https://learn.utoronto.ca/programs-courses/courses/2121-principles-operations-supply-chain-management?gclid=CjwKCAjwiuuRBhBvEiwAFXKaNOzSsKRb8jXeRtmLhCddIdCQ1bj722hlcaXIGbDCWophyLUGEFb-EhoCi70QAvD_BwE
https://www.ivey.uwo.ca/mma/courses/simulation-and-risk-analysis/
https://online.stanford.edu/courses/cee242r-project-risk-analysis
https://ug.degroote.mcmaster.ca/files/2020/07/4MI3-Marketing-Analytics-Course-Outline-Fall-2020.pdf
https://mbastudent.degroote.mcmaster.ca/wp-content/uploads/sites/69/2021/10/M733-Marketing-Analytics-Winter-2020-BLENDED-syllabus-18dec19.pdf

Appendix A – Part 2&3 Program overview and visualization

Section 1: Recommended core and elective courses for each specialization.

Data Analyst	
Core Course Recommendations	**Elective Course Recommendations**
Course 4: Data Modeling and Database Management (SQL) Course 5: Data Analytics: Introduction, Methods and Practical Approaches Course 8: Business Leadership and Communication Course 9: Data Mining for Business Applications	Course 11: Modeling tools for predictive analytics Course 13: Storytelling with data (Tableau) Course 15: Forecasting and Time series application Course 17: R for data science Course 21: Human values in data science Course 22: Information values and ethics Course 23: Business immersion Course 24: Analytics in management Course 25: Supply chain management Course 28: Analytics for marketing strategy

Data Scientist	
Core Course Recommendations	**Elective Course Recommendations**
Course 3: Machine learning with applications in python Course 4: Data Modeling and Database Management (SQL) Course 6: Introduction to Artificial Intelligence Course 7: Applications with Deep Learning	Course 11: Modeling tools for predictive analytics Course 12: Experimental Design for Data Science Course 16: Cloud technology in data science Course 17: R for data science Course 18: Optimization Techniques Course 19: Stochastic Modeling Course 20: Reinforcement Learning Course 21: Human values in data science Course 24: Analytics in management Course 27: Simulation and risk analytics

Data Engineer	
Core Course Recommendations	**Elective Course Recommendations**
Course 3: Machine learning with applications in python Course 4: Data Modeling and Database Management (SQL) Course 9: Data Mining for Business Applications Course 10: Business problem analysis and management	Course 13: Storytelling with data (Tableau) Course 14: Research data and research operations Course 15: Forecasting and Time series application Course 16: Cloud technology in data science Course 18: Optimization Techniques Course 19: Stochastic Modeling Course 20: Reinforcement Learning Course 21: Human values in data science Course 24: Analytics in management Course 27: Simulation and risk analytics

Data Manager	
Core Course Recommendations	**Elective Course Recommendations**
Course 5: Data Analytics: Introduction, Methods and Practical Approaches Course 8: Business Leadership and Communication Course 9: Data Mining for Business Applications Course 10: Business problem analysis and management	Course 13: Storytelling with data (Tableau) Course 16: Cloud technology in data science Course 18: Optimization Techniques Course 22: Information values and ethics Course 23: Business immersion Course 24: Analytics in management Course 25: Supply chain management Course 26: Data Driven investments Course 27: Simulation and risk analytics Course 28: Analytics for marketing strategy

Section2: The descriptions of courses offered by the program.

Course 1 - Statistics for data science: This is the introductory course to statistics for data science. This Statistics for Data Science course is designed to introduce you to the basic principles of statistical methods and procedures used for data analysis and data mining. After completing this course you will have practical knowledge of crucial topics in statistics including - data wrangling, data summary statistics, displaying and visualizing data, performing regression and correlation analysis, probabilities and density distributions, expected values, hypothesis testing, introduction to ANOVA (analysis of variance). You will gain a hands-on approach to statistical analysis using Python and Jupyter Notebooks – the tools of choice for Data professionals.

Course 2 - Programming for data science : This is the introductory course to programming for data science. In this course, I will use the most popular coding language Python to look at the design and implementation of algorithms that are relevant to Data Science tasks. The student will look into the

specifics of Python and teach them with libraries such as Numpy, Pandas, Matplotlib, data types, algorithms that will add to their background in Data Science.

Course 3 - Machine learning with applications in python : Machine learning has recently become the dominant field in AI research and constitutes the main part of the tools applied in industry-based AI positions. Data Analysts, data scientists, data managers and data engineers are required to know machine learning at different levels.
This course will dive into the essential topics of machine learning using the most popular coding language, Python. In this course, we will get the general overview of machine learning topics such as supervised learning vs unsupervised learning. For supervised learning, this course focuses on teaching logistic regression, k-nearest neighbor, decision trees, support-vector machines. For unsupervised learning, this course focuses on teaching K-means clustering, Hierarchical clustering, principal component analysis. Furthermore, this course will cover the basics of deep learning and natural language processing. More importantly, in this course, you can transform your theoretical machine learning knowledge into practical skills using hands – on labs.

Course 4 - Data Modeling and Database Management (SQL) : In fact, during the era of big data, database systems are central to most organizations' information systems strategies. At any organizational level, users are expected to face frequency contact with the use of database systems. Therefore, skills in interacting with these database systems, which include knowing how to retrieve data and utilize information effectively become essential in any industry vertical. This course focuses on learning different database systems, database design, database implementation and database management. Furthermore, this course will also focus on teaching the advanced knowledge and skills in SQL used for data science, which help you communicate with database systems better.

Course 5 - Data Analytics: Introduction, Methods and Practical Approaches: This course provides a comprehensive explanation to the field of analytics, including but not limited to knowledge extraction, data mining, data cleaning techniques, statistical and quantitative analysis. In this course, we will discuss various data analytical methods that can be applied on both structured and unstructured data and the relevant decisions which can be made in specific situations. The students will actively participate in the delivery of this course through case and project presentations.

Course 6 - Introduction to Artificial Intelligence : Since Artificial intelligence techniques are adopted widely by businesses nowadays, this course provides the students with an introduction to basic knowledge and skills they will need to work in the fields related to AI. It takes a variety of useful topics into account, such as optimization, reasoning and pattern recognition, and shows the students how these concepts are implemented under academic and commercial settings. They are going to be familiar with some of the popular AI tools, such as TensorFlow, and the services like IBM Watson cognitive computing. A prerequisite of programming experience using Python needs to be fulfilled before enrolling into this course.

Course 7 - Applications with Deep Learning: Building upon the fundamentals of machine learning, this course advances to deep learning and neural networks. It explores deep learning applications in depth and introduces how deep neural networks (DNNs) work on classification, recognition and

localization. There will be interesting and popular deep learning topics including perceptron, back propagation, regularization, sum-product networks, recurrent neural networks, and convolutional networks. The students are expected to spend a considerable amount of time on reading publications, understanding deep learning methods and getting their hands on solving real-world problems with their groups. They will have opportunities to gain experiences on implementing and testing the methods they learnt, which will enhance their career.

Course 8 - Business Leadership and Communication: To be a qualified data professional, you are not only required to be a strong problem solver, but also required to be an excellent communicator. Accompanied with the utilization of data analytic skills in business, data professionals are also required to understand the business comprehensively, communicate the analysis results very clearly in both writing and presentation. In this course, students will learn how to communicate effectively with different audiences and how to use logical persuasive techniques in writing and presenting. Students have an opportunity of developing the written, oral, interpersonal and team skills while developing an understanding of leadership communication in the different business contexts.

Course 9 - Data Mining for Business Applications : This course introduces data mining problems and tools to enhance managerial decision making. Students will learn how to ask the right questions and how to draw inferences from data by using the appropriate data mining tools. This course will enable students to approach business problems data-analytically, envision data mining opportunities in organizations, and follow up on ideas or opportunities that present themselves. The emphasis of this course is on understanding the application of a wide range of modern data techniques to specific business decision-making situations, rather than mastering the theoretical underpinnings of the data techniques. Upon successful completion of the course, you should possess valuable practical analytical skills that will equip you with a competitive edge in almost any contemporary workplace.

Course 10: Business problem analysis and management: As more and more business issues can be addressed by using techniques of Data Science and Artificial Intelligence, this course is structured to provide the students with insight into the key challenges within business problem analysis and the practical process used in approaching these challenges, including problem framing, project plan developing, resource scheduling, risk management, decision making, visualization and delivery to the clients. The examined problems are selected from a wide range of business issues. The students can choose the topics they are interested in. Pre-requisites of basic statistics and programming experience need to be fulfilled before enrolling into this course.

Besides courses provided by the program, the program also offers other activities and services to achieve the students' comprehensive learning and increase their Master learning experience. Among them, the main activities and services are Co-op program, workshops, writing center, and capstone project.

Section3: The descriptions of the activities & services offered by the program.

Co-op program:
A co-operative education (Co-op) is an experiential learning program obtained in partnership between students, employers and the program 'master of Data Science and Artificial Intelligence'. Through Co-op, students complete up to paid, full- time co-op work terms and gain professional experience in their fields of study. It is an opportunity to apply knowledge in a professional situation and gain academic credits. To complete studies in Master of Data Science and Artificial Intelligence and gain the degree, students are required to complete 1 term of Co-op (i.e. 4 months).

Events & Workshops:
The faculty provides various workshops to the students to improve their comprehensive quality and help them better become a data professional after graduation. These workshops include but are not limited to resume customization workshop, interview mock workshop, cover letter customize workshop, data visualization workshop, data presentation workshop. The events include but are not limited to the career information session, Diversity's talk, aluminis' talk, career fair.

Writing Center:
The writing center is aimed to help students improve their writing skills, including the grammar correction, language expression polish, etc.

Capstone Project:
The optional capstone project is equivalent to two electives. It needs to be completed over two terms and involves research, report writing and presentation. It will provide the students with real-world problems and allow them to apply the tools and methodologies they learnt creatively.

Section4: List of skills for core and elective courses in Master of Data Science and Artificial Intelligence.

Course 1 - Statistics for data science:
statistics, statistical analysis, data gathering, descriptive statistics, visualization, probability distribution, hypothesis testing, ANOVA, regression, correlation analysis

Course 2 - Programming for data science:
Python, programming principles, object-oriented programming, iterable, regular expression, algorithm

Course 3 - Machine learning with applications in python:
python, machine learning, logistic regression, KNN, decision tree, SVM, K-means clustering, hierarchical clustering, PCA, deep learning, Naive Bayes

Course 4 - Data Modeling and Database Management (SQL):
data modeling, database design, database implementation, database management, SQL, MetaData, data governance

Course 5 - Data Analytics: Introduction, Methods, and Practical Approaches:
statistical analysis, quantitative analysis, exploratory data analysis, predictive model, data mining, decision making, presentation

Course 6 - Introduction to Artificial Intelligence:
artificial intelligence, python, machine learning, optimization, reasoning, pattern recognition, TensorFlow, IBM Watson cognitive computing

Course 7 - Applications with Deep Learning:
machine learning, deep learning, neural networks, DNN, perceptron, back propagation, regularization, sum-product networks, recurrent neural networks, and convolutional networks, teamwork
Course 8 - Business Leadership and Communication:
business, problem solving, communication, writing, presentation, teamwork, leadership
Course 9 - Data Mining for Business Applications:
business, data mining, decision making
Course 10 - Business problem analysis and management:
business, scheduling, risk management, visualization, communicatio
Course 11 - Modeling tools for predictive analytics:
modeling, predictive analytics, statistical analysis, machine learning, exploratory data analysis, Excel, business
Course 12 - Experimental Design for Data Science:
experimental design, planning, research, qualitative analysis, quantitative analysis
Course 13 - Storytelling with data (Tableau):
Tableau, storytelling, delivery, visualization, robust visualization
Course 14 - Research data and research operations:
research, data management, data storage, security of data, decision making
Course 15 - Forecasting and Time series application:
predictive analytics, time series analytics, machine learning, ARIMA model, stationarity, cyclicality, seasonality, autocorrelation
Course 16 - Cloud technology in data science:
cloud technology, cloud computing, cloud architecture, big data, security of data, IaaS, PaaS, SaaS, risk management
Course 17 - R for data science:
R, business, big data, statistical analysis, benchmarking, profiling, ggplot, tidyverse
Course 18 - Optimization Techniques:
optimization, linear optimization, non-linear optimization, theory-based mathematics, convexity, duality, sensitivity analysis, simplex method, Newton's method, gradient-based method, stochastic gradient descent
Course 19 - Stochastic Modeling:
stochastic modeling, optimization, two-stage stochastic programming, Bender's method, SAA, Monte Carlo, static robust optimization, two-stage robust optimization, Markov chain, queuing theory, renewal process
Course 20 - Reinforcement Learning:
reinforcement learning, dynamic programming, Monte Carlo, temporal-difference, behavioral cloning, inverse reinforcement learning, meta-learning, multi-agent learning
Course 21 - Human values in data science:
ethics, reading, writing, policy, privacy, discrimination
Course 22 - Information values and ethics:
ethics, reading, writing, policy, privacy, discrimination
Course 23 - Business immersion:
business, organizational management, planning, personal branding, writing, presentation
Course 24 - Analytics in management:

management, management analytics, business management, business intelligence

Course 25 - Supply chain management:

business, supply chain management, supply chain analytics, capacity management, six sigma, VUCA, deterministic model, stochastic model, transportation, contract design

Course 26 - Data Driven investments:

business, investment, predictive analytics, bonds, theory of interest, IRR, portfolio management, portfolio modeling, linear programming, dynamic programming, integer programming, risk management

Course 27 - Simulation and risk analysis:

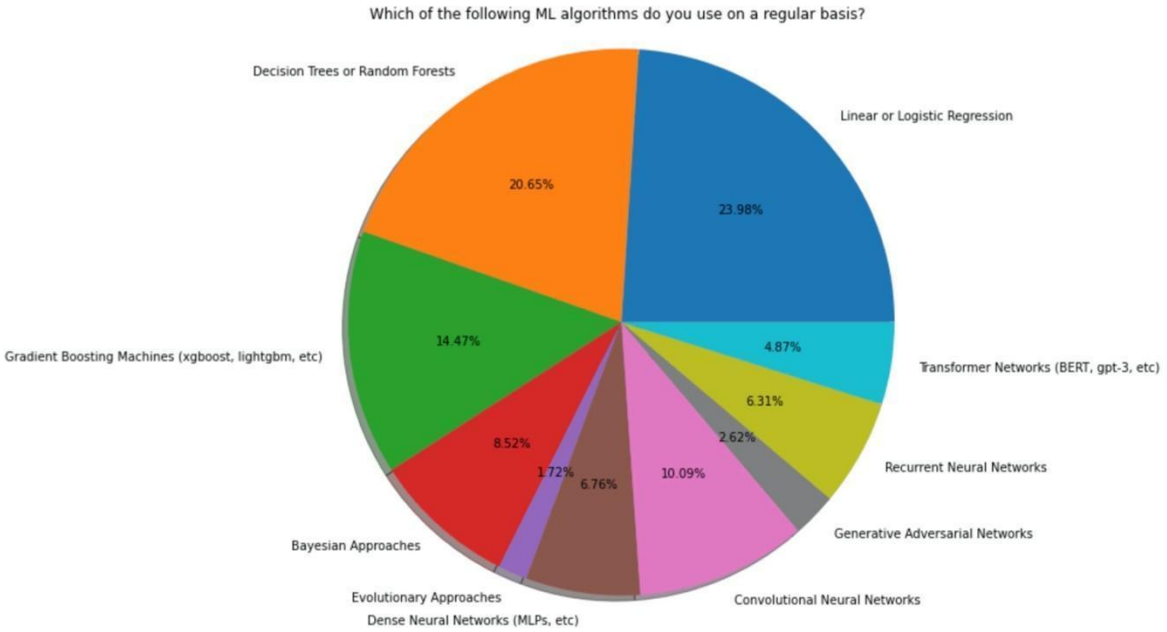
statistical analysis, risk management, time series analytics, cyber security, discrete event simulation, optimization, Monte Carlo, Markov chain, Excel, SIMUL8

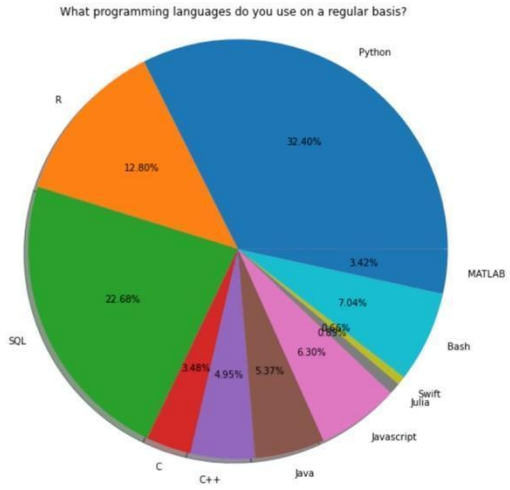
Course 28 - Analytics for marketing strategy:

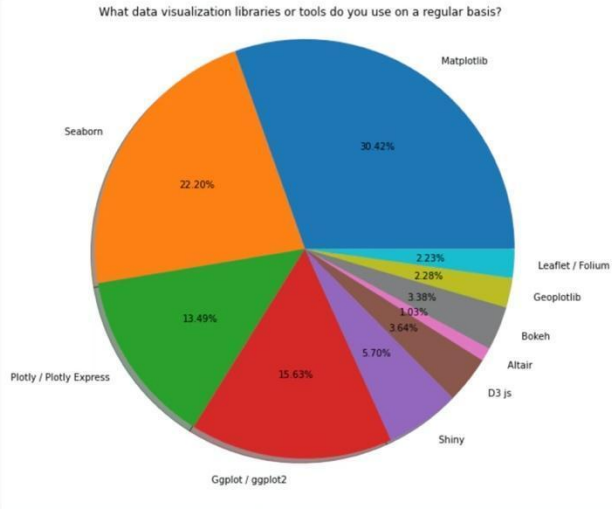
business, marketing, factor analysis, cluster analysis, classification analysis, multivariate regression analysis, deep learning, web scraping, R, ME>X

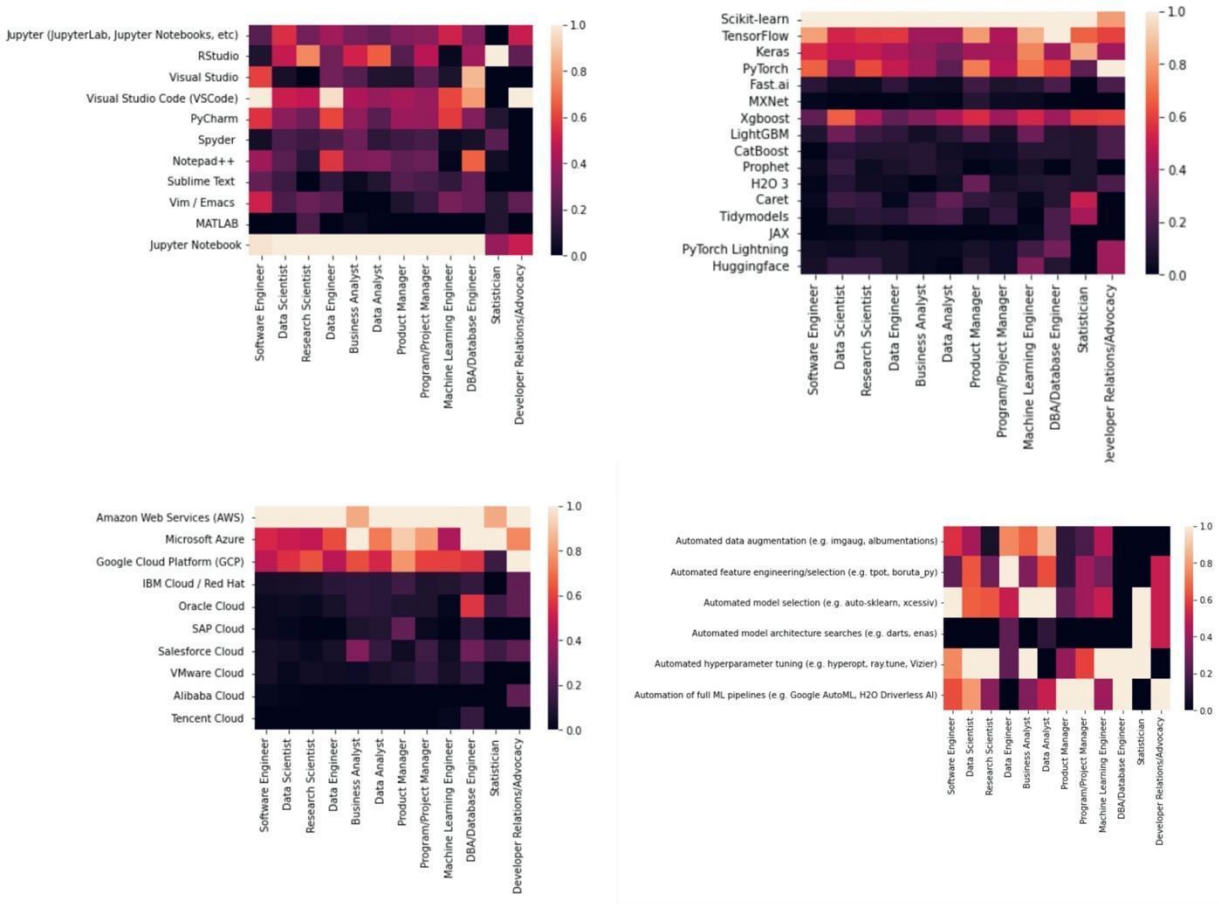
Appendix B – Table & Figure

Part 1:



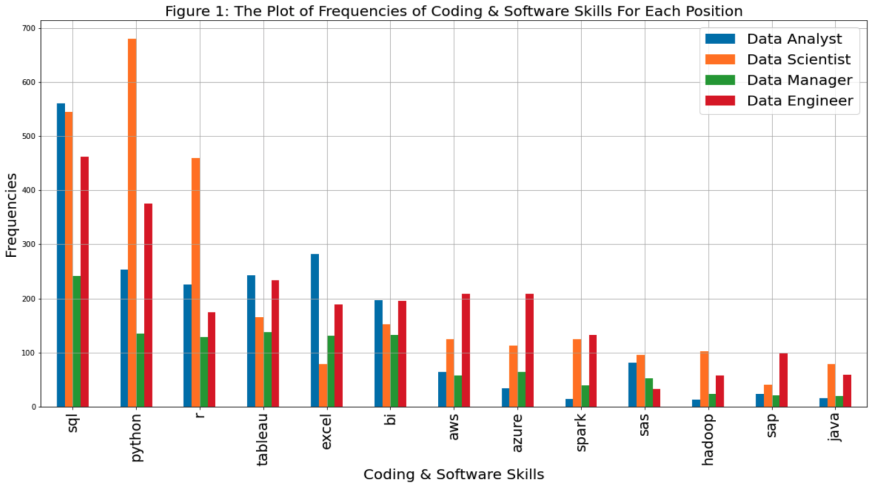






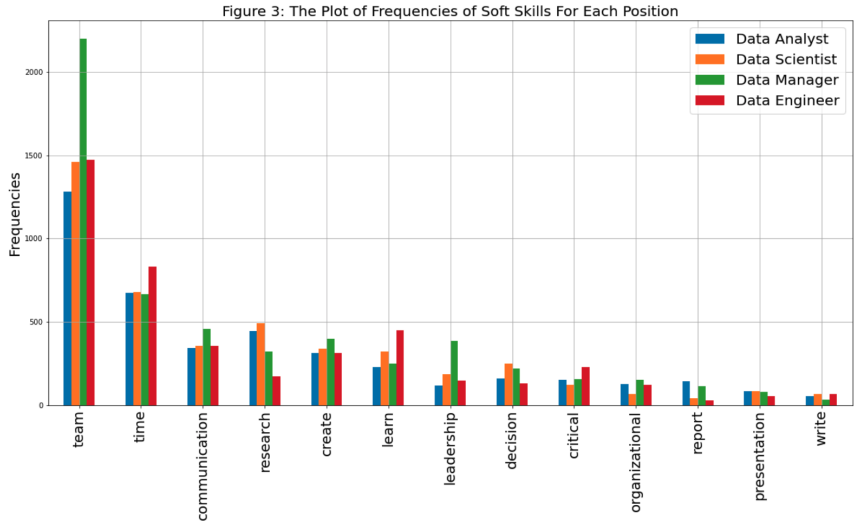
Part 2:

Section 1: The relevant results of Skill frequency for four data-related positions respectively.



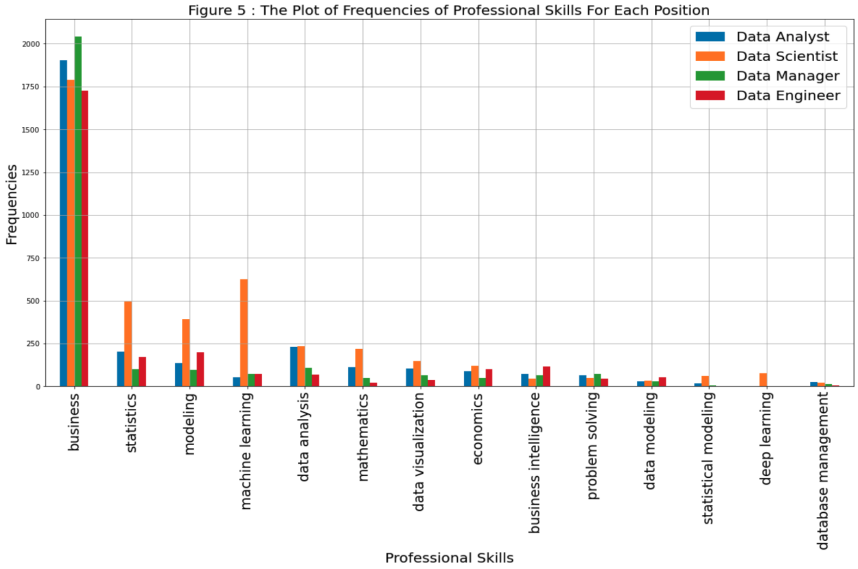
Bar Chart1: The frequencies of coding & software skills for each position.

Bar Chart1 shows for data analyst, sql and excel are the most frequent coding & software skills to be used. For data scientists and data engineers, sql and python are the most frequent skills to be used. For data managers, sql is the most frequent skill to be used in the work placement .



Bar Chart2: The frequencies of soft skills for each position.

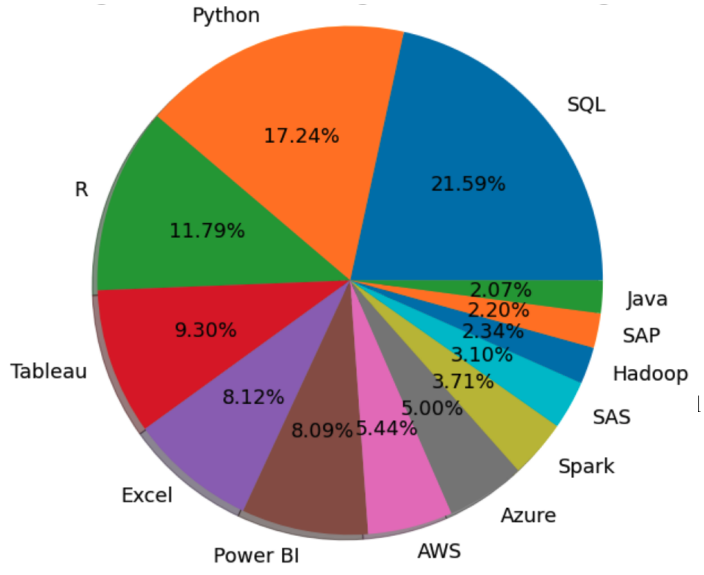
Bar Chart2 shows for all four positions, teamwork ability and time management skills are the most frequent skills to be used in the work placement. Further, leadership skills are also important to data managers.



Bar Chart3: The frequencies of professional skills for each position.

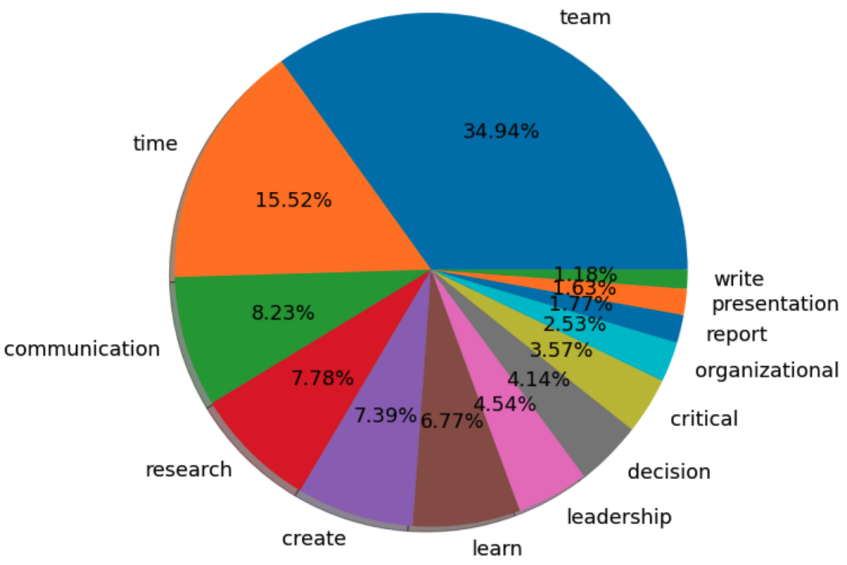
Bar Chart3 shows that the business skills and the ability to gain the business insights are important to all four data-related positions. Further, the statistics knowledge, data modeling skills and machine learning skills are very important to the data scientists in the work placement.

Section 2:The relevant results of skill frequency for four data-related positions aggregately.



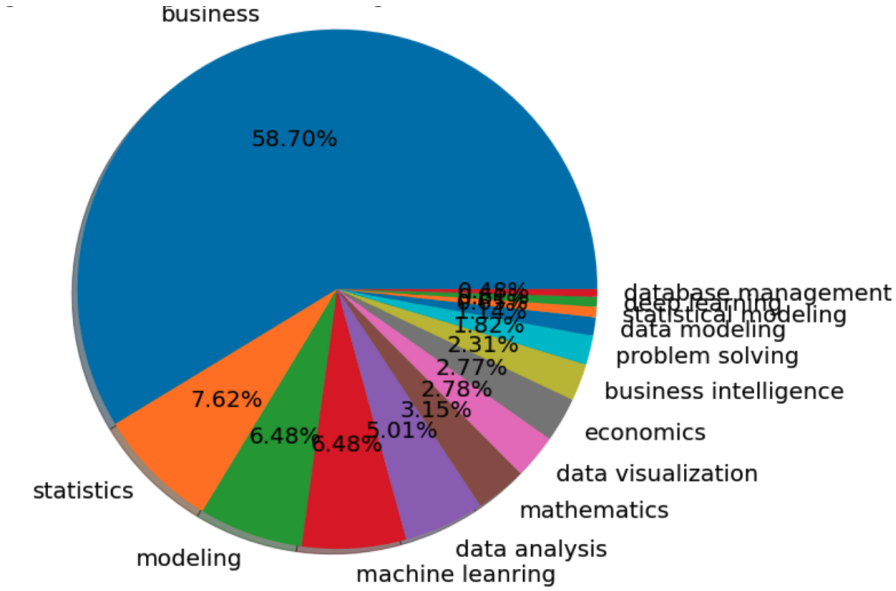
Pie Chart 1: Percentage of each skill among all selected coding & software skills

Pie Chart1 shows Python, SQL and R are the most important coding & software skills for all four data-related positions.



Pie Chart 2: Percentage of each skill among all selected soft skills

Pie Chart2 shows that Teamwork ability, time management skills and communication skills are the most important soft skills for all four data-related positions.



Pie Chart 3: Percentage of each skill among all selected professional skills

Pie Chart3 shows that the business skills, statistics knowledge, modeling skills and machine learning skills are the most important professional skills for all four data-related positions.