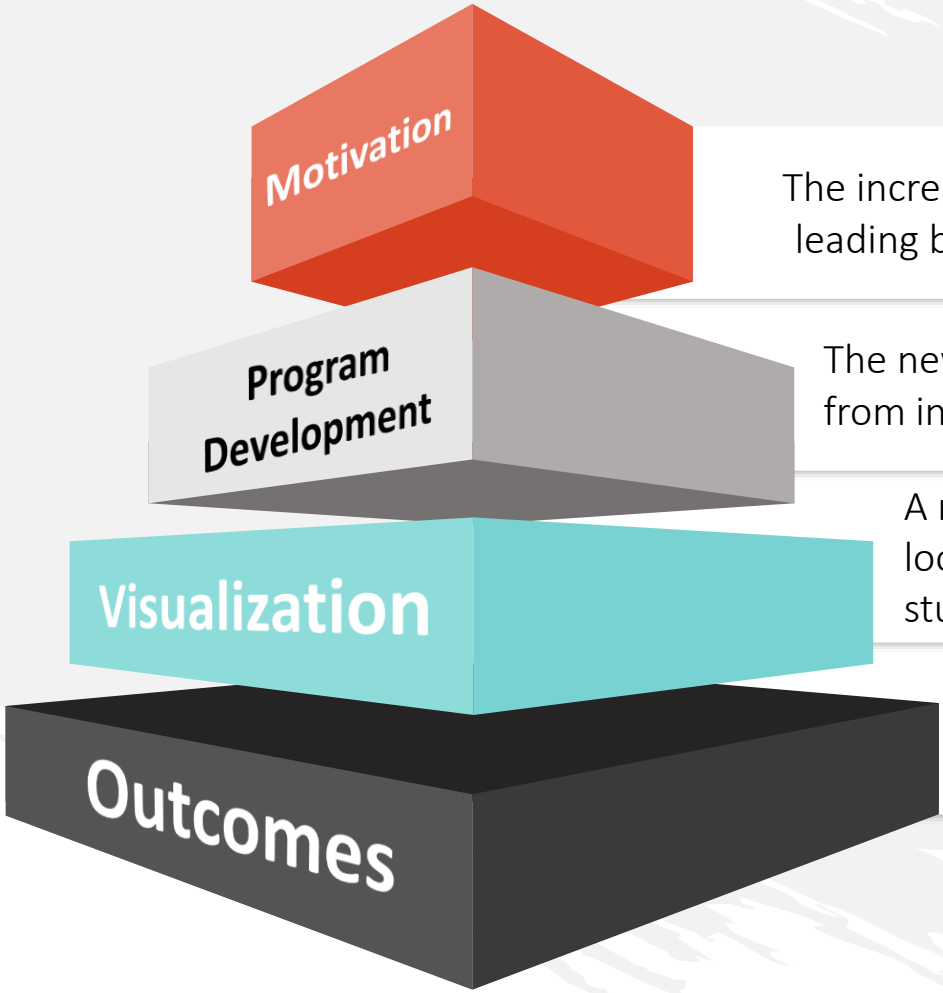
SMART-ED

- Education Made Smarter -

MIE 1624 Winter 2020 Group 11 – Startup Project

- Simon Faux | Akshat Mathur | Rami Al-Sahar | Gautam Dawar

Business Ideology



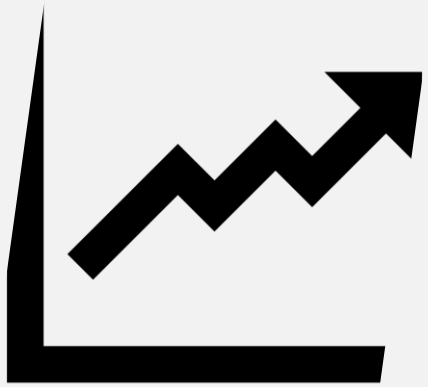
The increasing demand for a world leading business analytics degree

The new degree program was built on inputs from industry and the academic world

A representation of what the new program looks like and how it will meet the needs of students

The new program will be a world leader in preparing students for the field

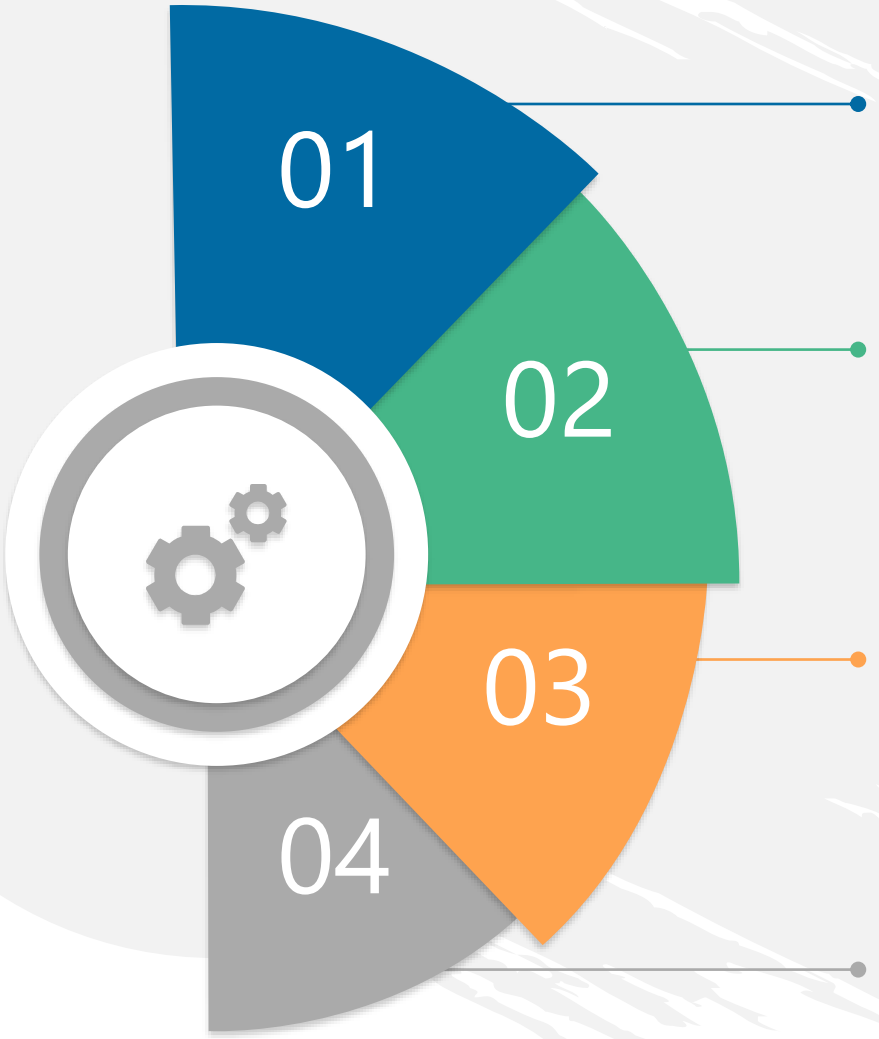
A Program For Students, Designed By Industry



In Demand Skills	Leading Courses	Data Analysis	Our Program
Data extracted from over 300 data scientist job postings	We picked top Universities which designed business focused courses	Used Novel models and scoring strategies to create best course for you!	A curriculum based to prepare students to become industry leader in Analytics

Part 1:
Job Skills Analysis and Redesign MIE 1624

Part 1: Model Functioning Overview



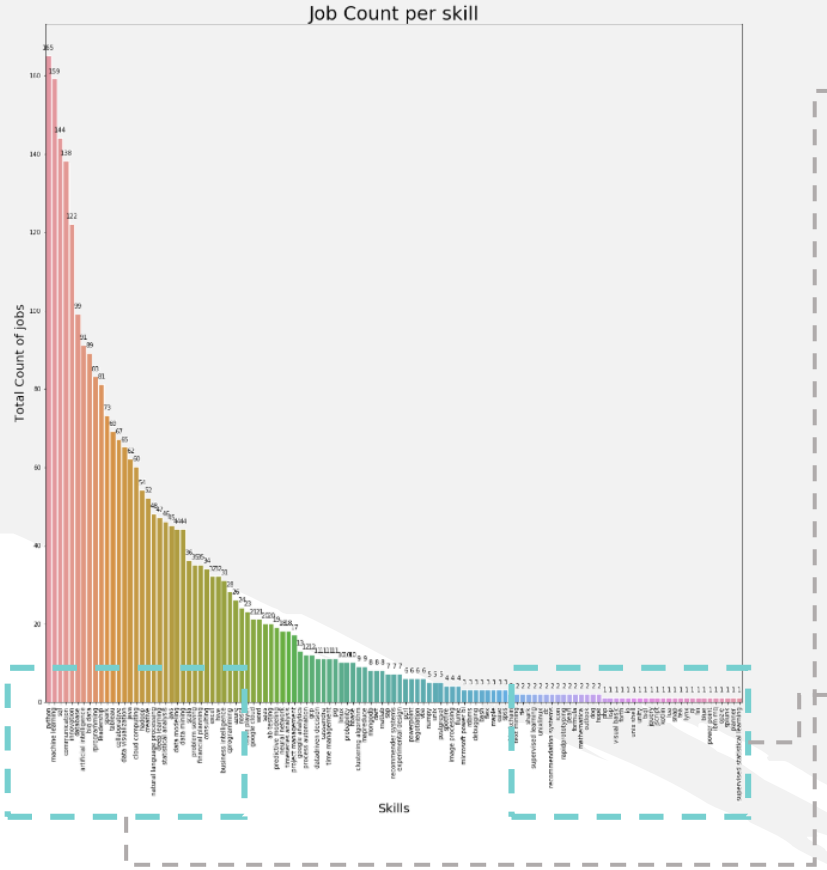
Data of **350 Unique Jobs** in Greater Toronto Area was scrapped from famous job posting websites; **LinkedIn and Indeed** and approx. **Unique 50 University and Online Courses**

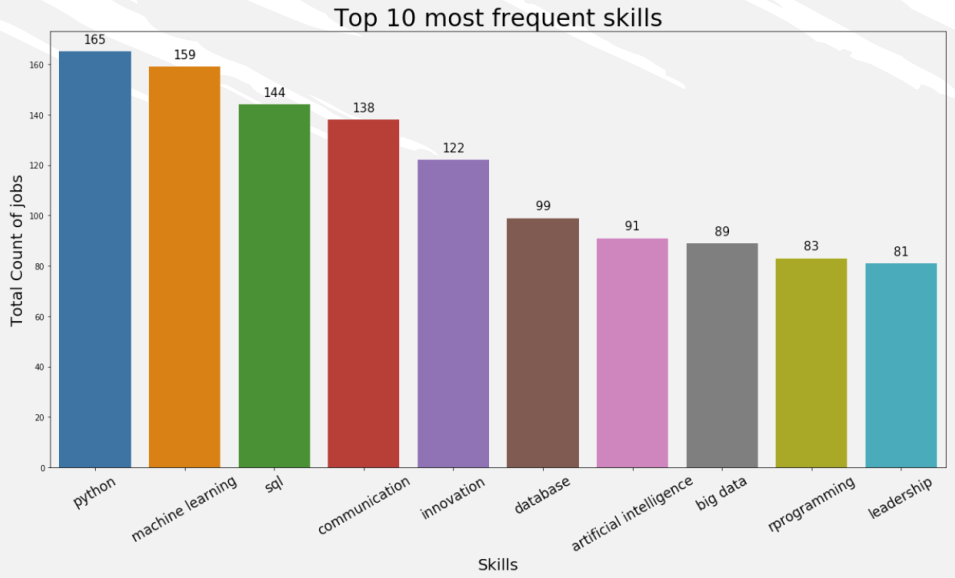
A list of approx. **800 skills** was built by scraping the internet and manually parsing jobs which are related to **data science** and **business intelligence**

The skills list was **mapped** with job description and courses using **bag of N-grams** to get an insight into the data scientist jobs and **clustering** was similar skills was performed

Learnings visualized and accordingly **MIE 1624** was course was **redesigned**

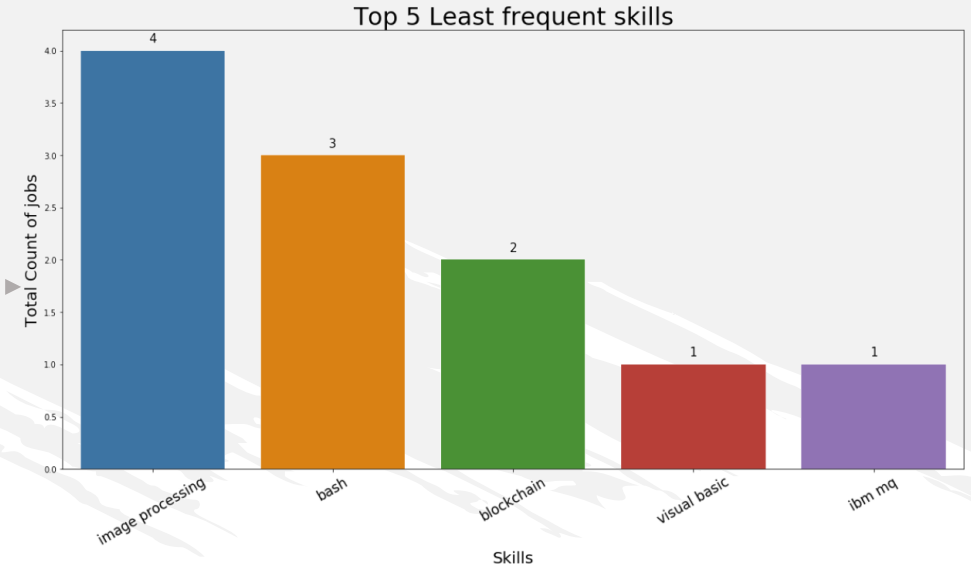
Part 1: Skill Insights





Top 10 Most Frequent Skills

Stats and coding skills like python, r, Artificial Intelligence and **DBMS skills** like sql, database, big data and **soft skills** like communication *and* leadership are **most popular** among employers for Data Scientists



Top 5 Least Frequent Skills

Image Processing, blockchain and Bash are among least poplar skills in demand by the employers for Data Scientist

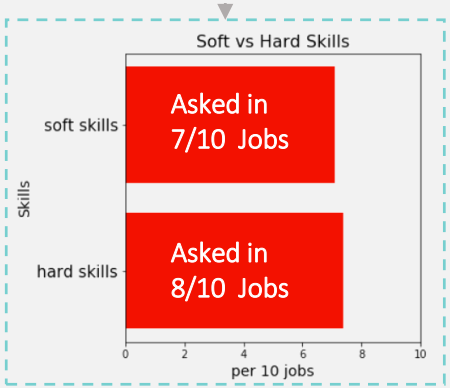
Part 1 Results:
Fully Redesigned MIE 1624

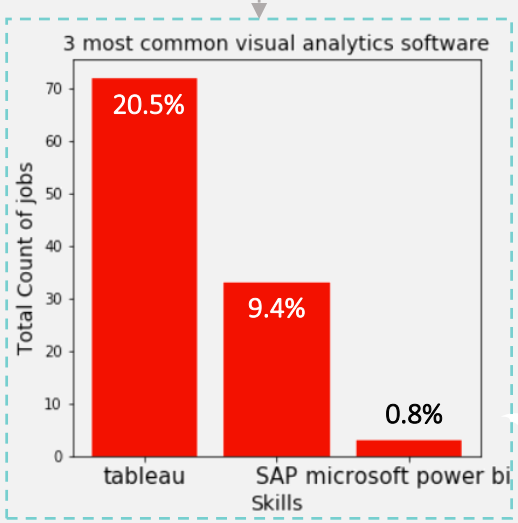
Redesigned MIE 1624 – Introduction to Data Science and Analytics

Week	Outline	Skills Gained
1	Introduction to data science and analytics	Quantitative modeling | basic statistics | linear algebra | Python
2	Modeling techniques, regression	Mathematical modeling | Linear regression | Logistic regression
3	Data visualization and visual analytics	IBM Watson Analytics | Storytelling | analytical decision | Tableau
4	Data mining & machine learning	Classification | Clustering | supervised machine learning | Data mining
5	Optimization	non-linear optimization | constrained optimization
6	Simulation modeling	Monte Carlo | Random number generation
7	Cognitive computing & artificial intelligence	neural networks | natural language processing | Reinforcement learning
8	Enterprise Hadoop & Hadoop basics	HDFS | YARN | data management | big data | business case
9	SQL: Hive and Impala	scripting | Spark | Hive | Impala | SQL: most important commands
10/11	Spark, introduction	PySpark | time-series - Spark | Model development

Homework & Skills

- **#1 Assignment** Python Basics | data exploration | Statistics | linear algebra
- **#2 Assignment** logistic regression | Tableau Dashboard
- **#3 Group Poject** NLP | collaboration | leadership | innovation | communication
- **#4 Assignment** data mining | Hive | SQL | Spark





43% approx.

Jobs require knowledge on SQL and Big - data

Existing Curriculum

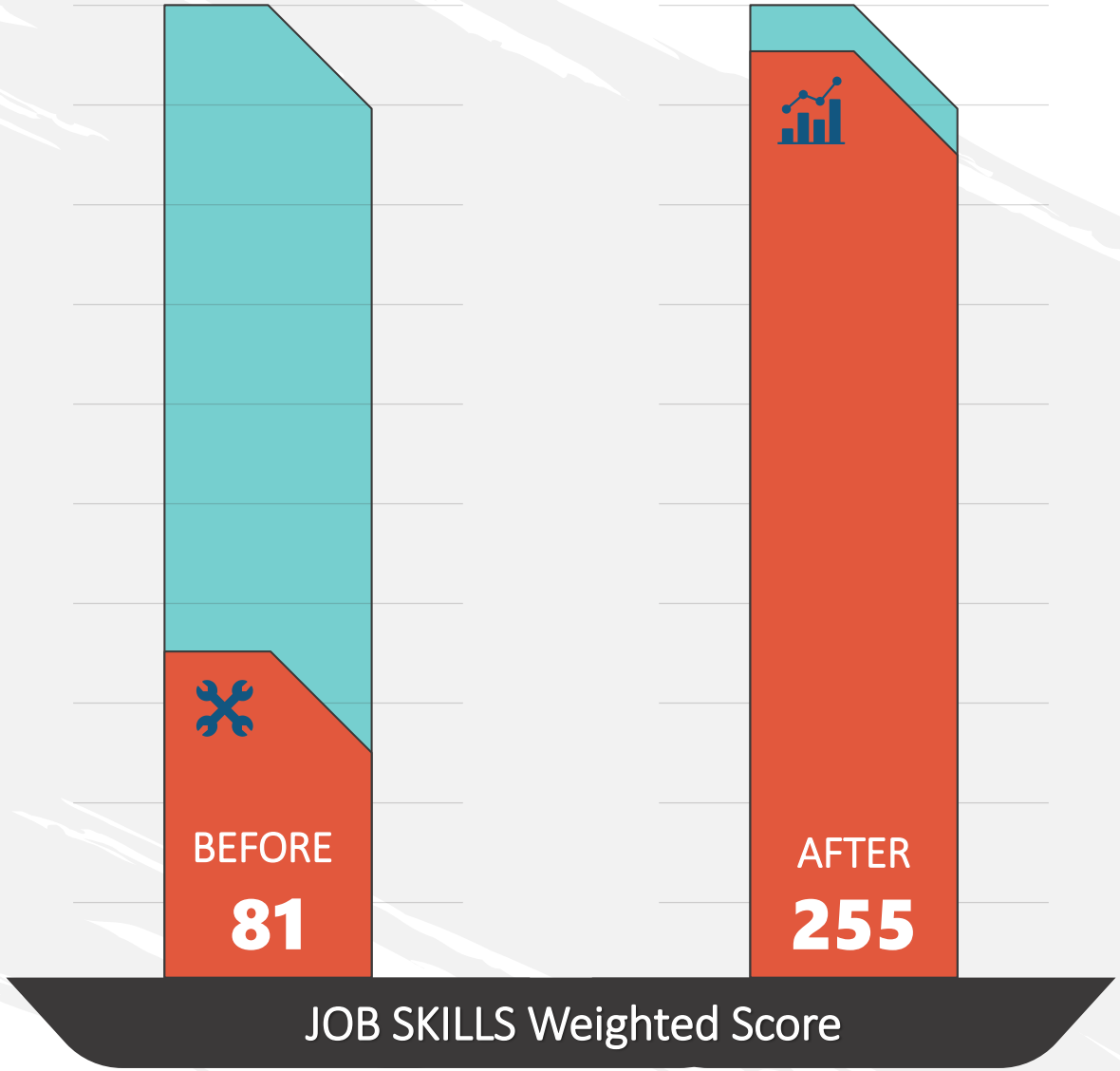
Modified Curriculum

Part 1: Course Comparison

The following is a comparison between the number of jobs qualified for by studying the original **MIE 1624 course** vs. **the updated MIE 1624 course.**

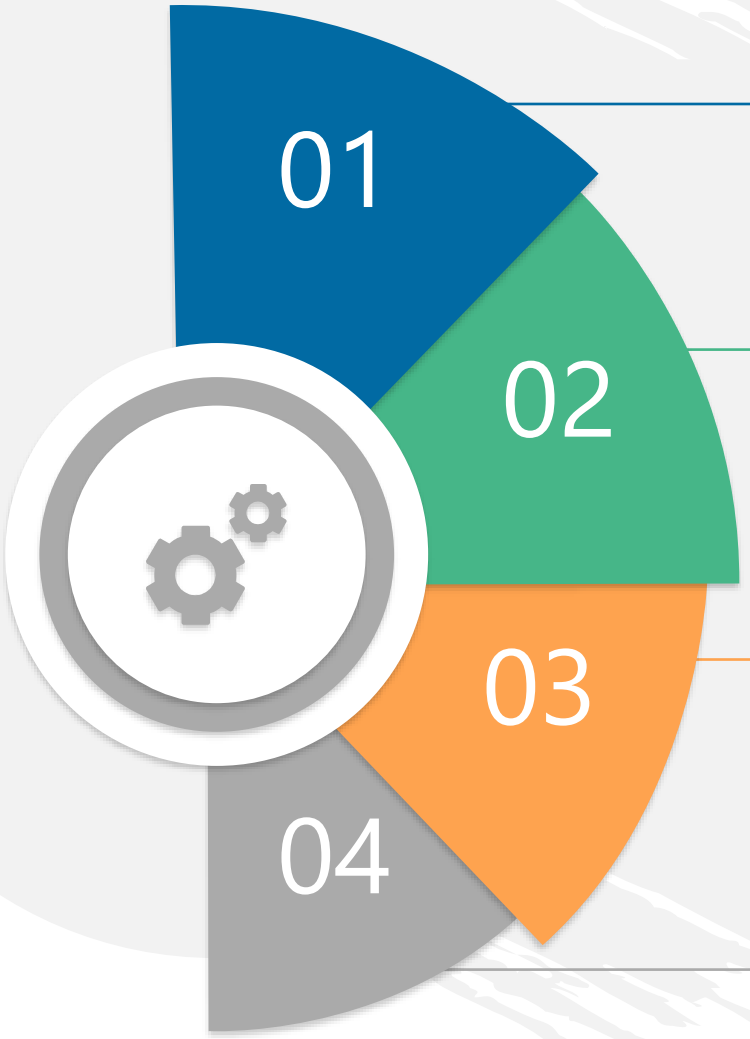
# of jobs qualified for by taking the original course.

# of jobs qualified for by taking the redesigned course



Part 2:
Masters Program Design –
Master of Data Science and Artificial Intelligence

Part 2: Model Functioning Overview



Performed **Clustering of the skills** to obtain **9 clusters** using **cosine distance similarity** of the skill vectors obtained by **tfidf Vectorization**

46 Unique course were identified from **Top Universities** around the world related to Data Science and Analytics

Courses extracted from the universities were **mapped with the skills** obtained from part 1 to match the 9 clusters with the 45 courses

Courses were selected based on their occurrence in the cluster and their individual presence in the job skill data frame. Also Iterative Learning was kept in mind

Part 2: Skill Cluster Identification

01

'python', 'sql', 'communication', 'database', 'rprogramming', 'tableau', 'data visualization'

02

'saas', 'ab testing', 'google analytics', 'datadriven decision'

03

'nosql', 'cassandra'

04

'big data', 'spark', 'java', 'hadoop', 'scala', 'hive', 'hbase'

05

'problem solving', 'consulting', 'business intelligence', 'word', 'timeseries analysis', 'project management', 'process automation', 'time management'

06

'innovation', 'leadership', 'collaborative', 'creative', 'data modeling', 'financial planning', 'excel', 'team player'

07

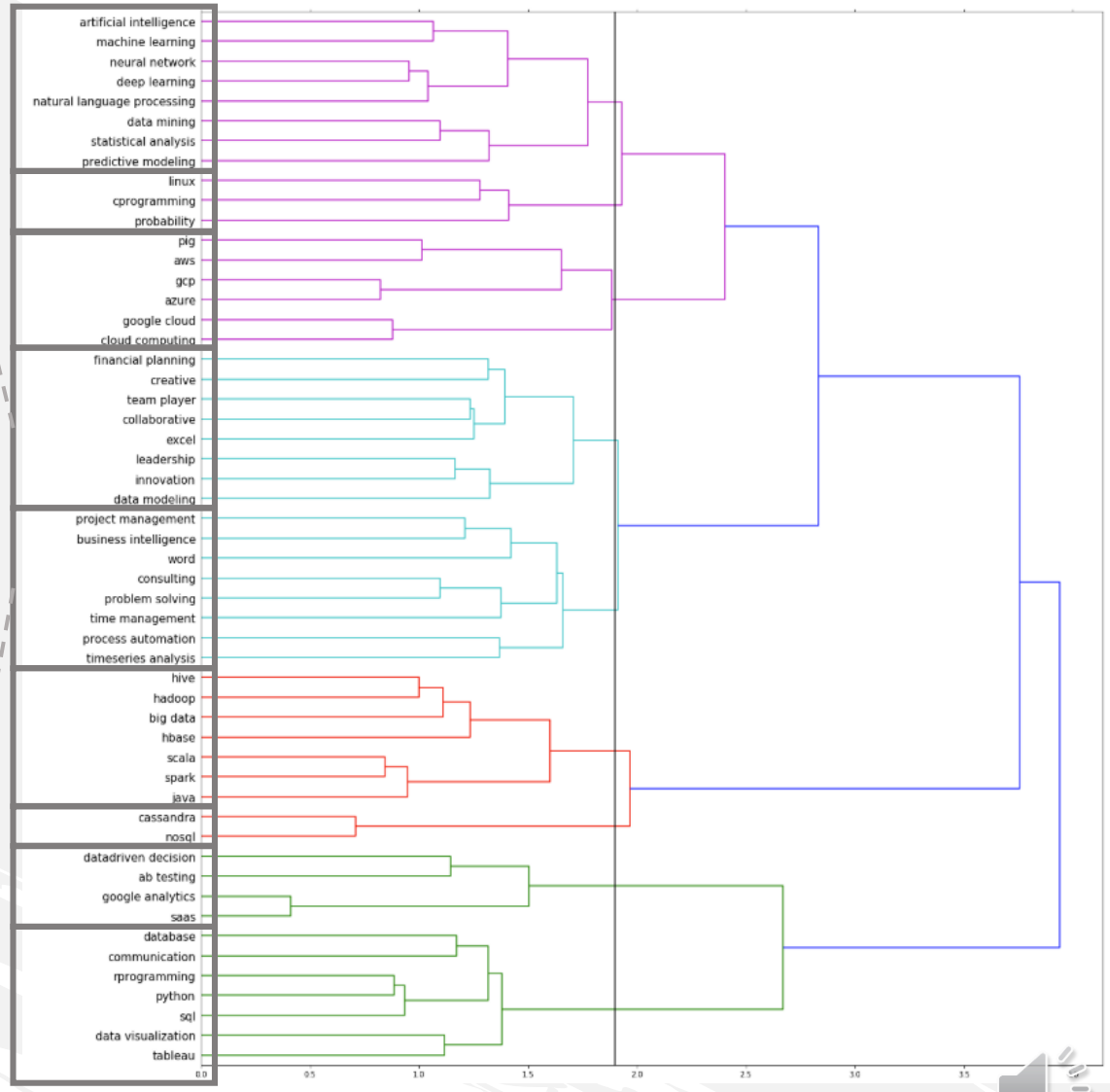
'cloud computing', 'aws', 'azure', 'google ', 'gcp', 'picloudg'

08

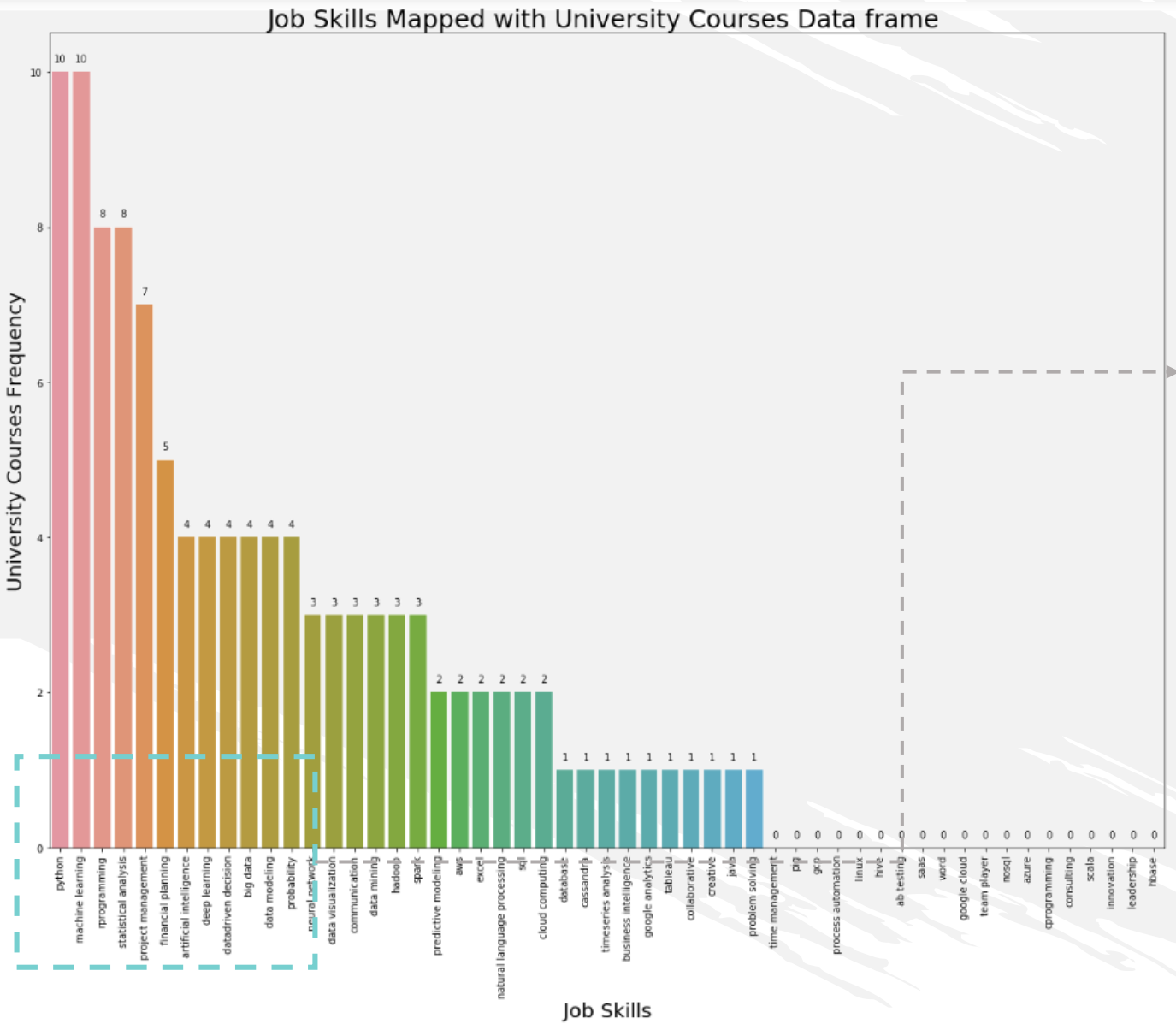
'cprogramming', 'probability', 'linux'

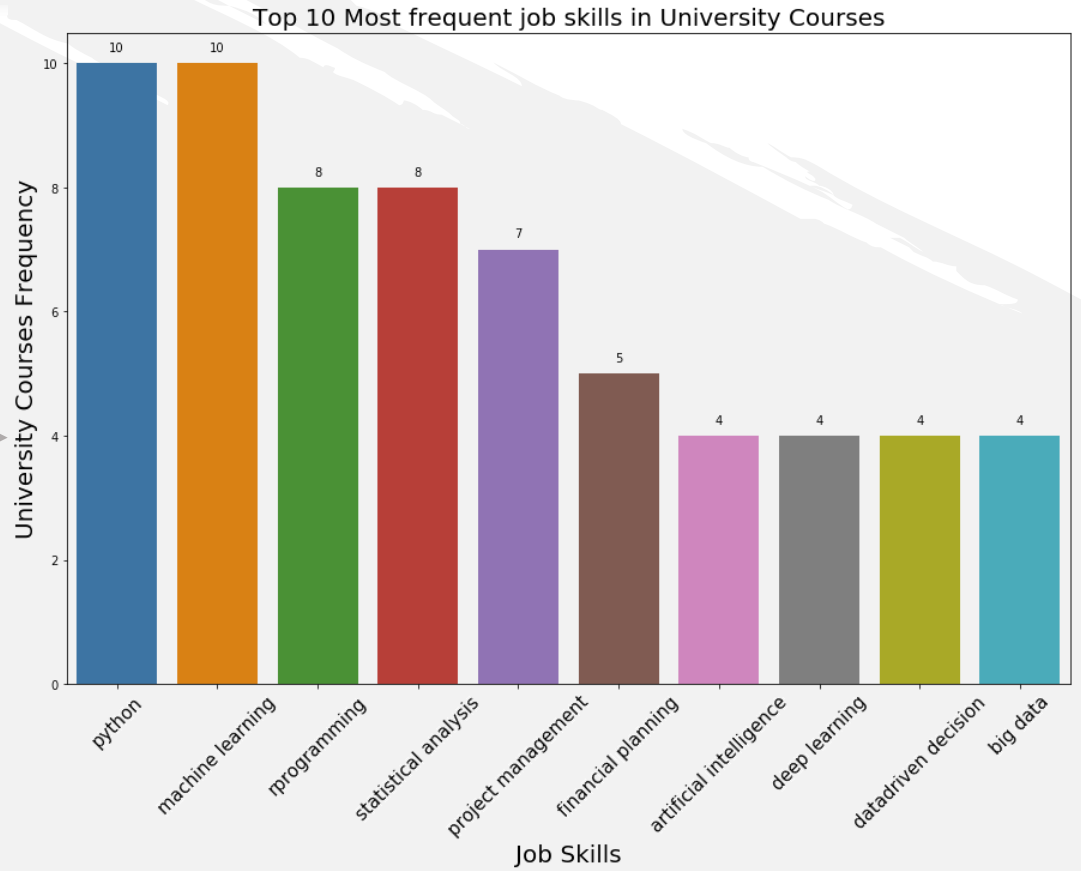
09

'machine learning', 'artificial intelligence', 'natural language processing', 'deep learning', 'statistical analysis', 'data mining', 'predictive modeling', 'neural network'



Part 2: Job Skill Mapping with University Courses

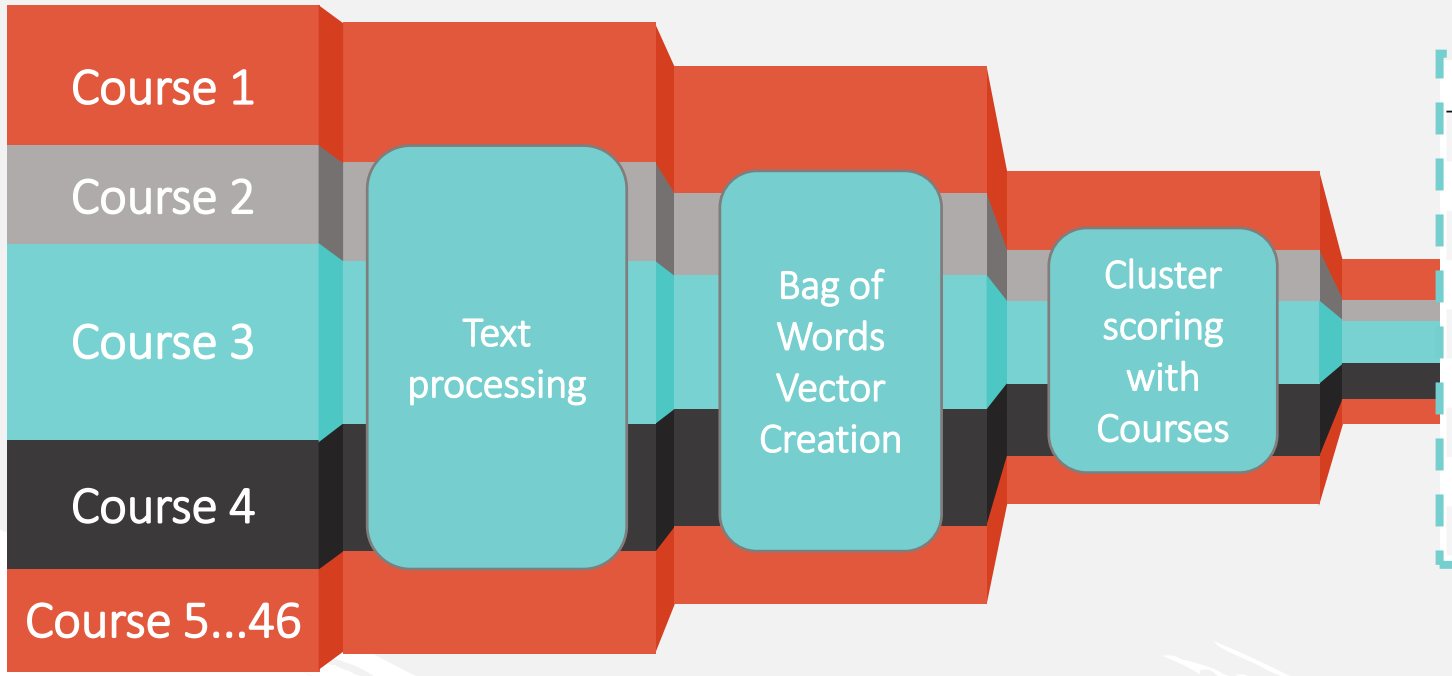




Top 10 Most Frequent Job Skills in University Courses

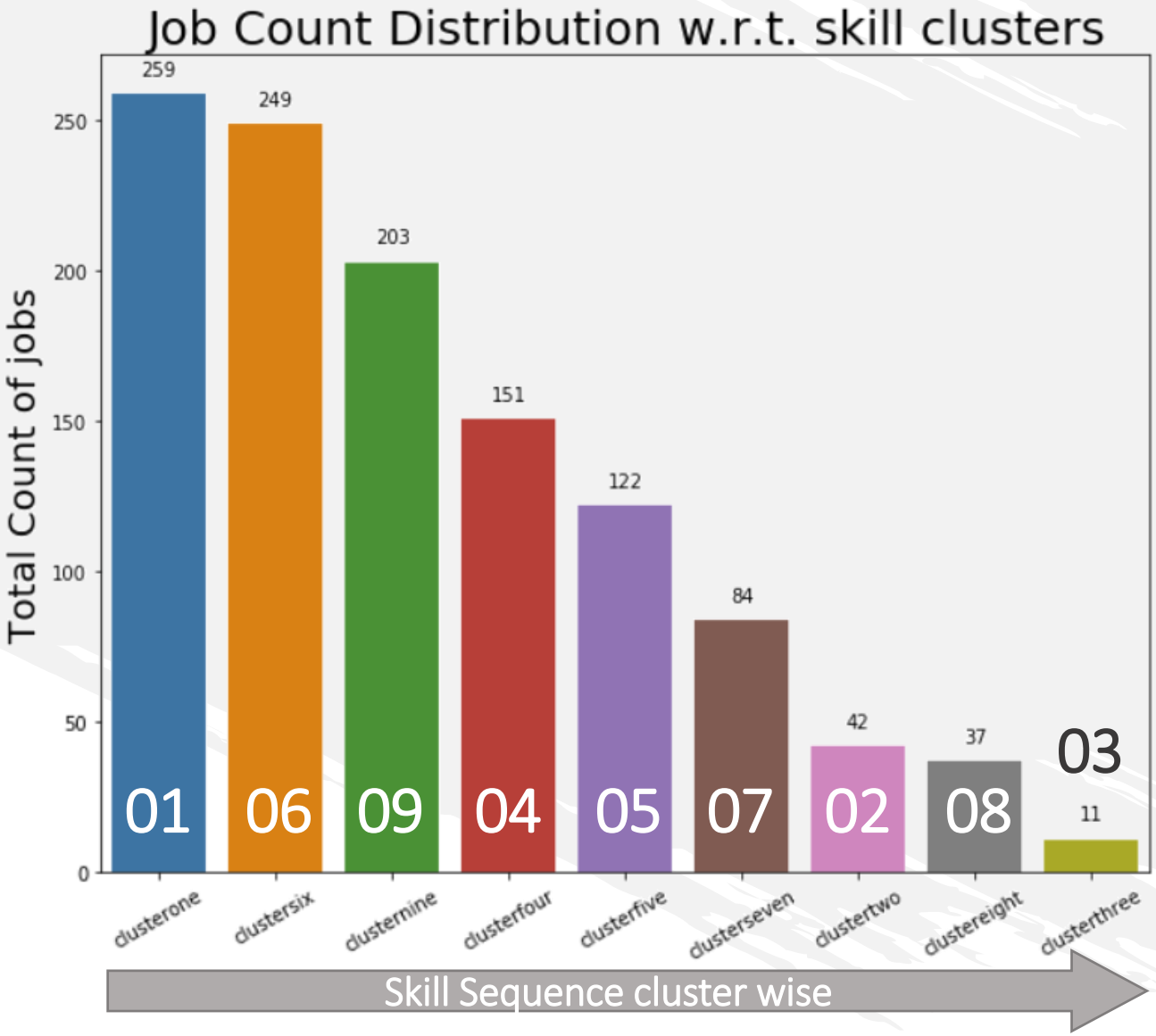
Skills like python | Machine Learning | R | Statistical Analysis | Project Management | Financial Planning | Artificial Intelligence | Deep Learning | Data Driven Decision | Big Data are quite common

Part 2: University Courses Mapping to Job Skill Clusters



	cluster_number	Title
11	1	Data Visualization
1	2	Research Design and Application for Data and Analysis
7	3	Deep Learning in the Cloud and at the Edge
9	4	Machine Learning at Scale
8	5	Statistical Methods
2	6	Statistics for Data Science
3	7	Fundamentals of Data Engineering
4	8	Applied Machine Learning
21	9	Leveraging AI and Deep Learning Tools in Marketing (RSM8521)

Part 2: Skill Cluster Sequencing based on Job Skills Distribution



CLUSTERS

01
74.0 %
Data visualization and programming

02
12.0 %
Business Intelligence Data Driven Decision

03
3.1 %
Cloud Computing with DBMS

04
43.1 %
Big Data Java Hadoop Hive

05
34.9 %
Statistical Methods Business Intelligence

06
71.1 %
Stats Data modelling and Financial Planing

07
24.0 %
Cloud computing and Data Engineering

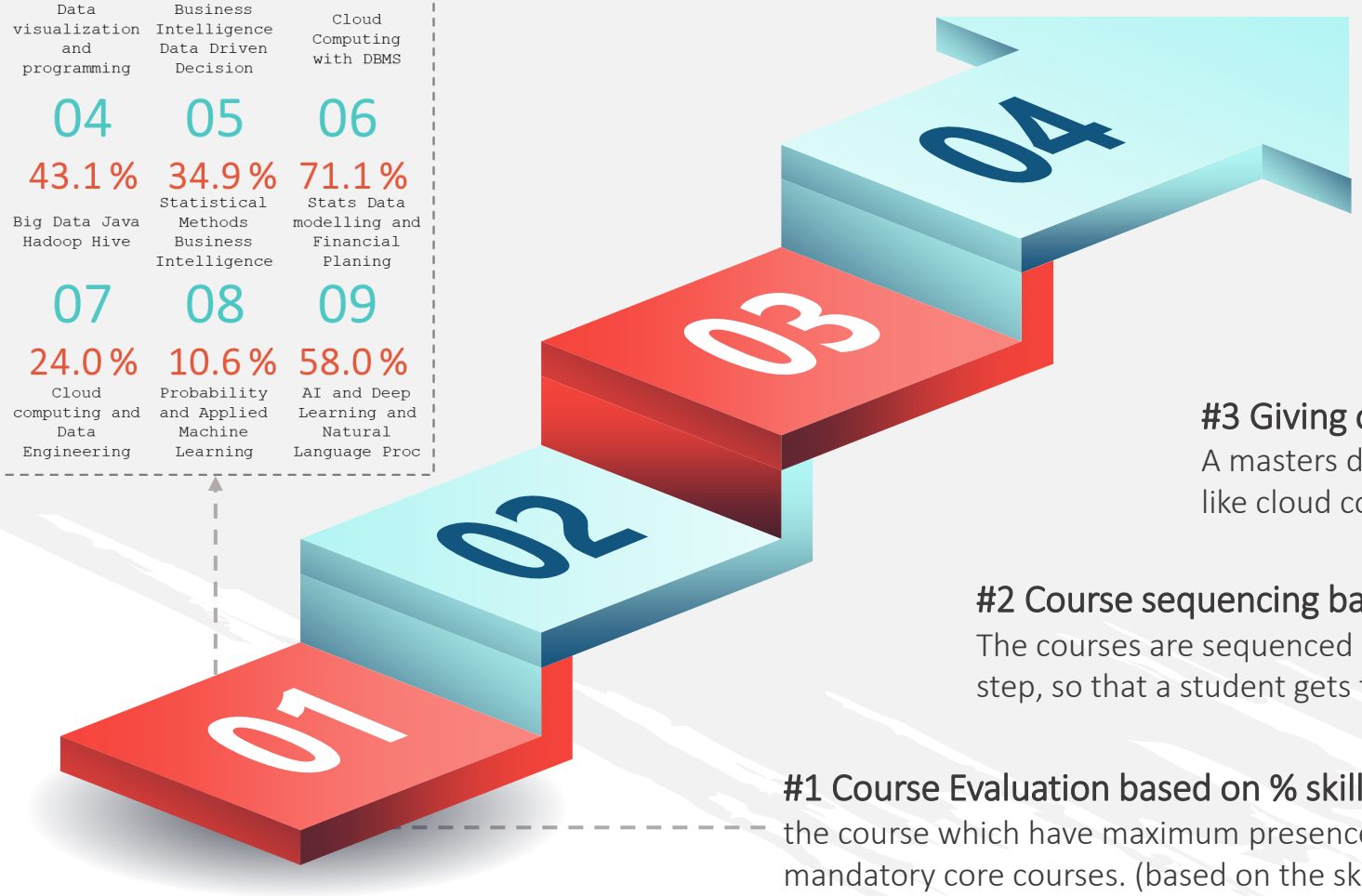
08
10.6 %
Probability and Applied Machine Learning

09
58.0 %
AI and Deep Learning and Natural Language Proc

Part 2: Masters Course Sequencing and Design Ideology

CLUSTERS

01	02	03
74.0 %	12.0 %	3.1 %
Data visualization and programming	Business Intelligence Data Driven Decision	Cloud Computing with DBMS
04	**05**	**06**
43.1 %	34.9 %	71.1 %
Big Data Java Hadoop Hive	Statistical Methods Business Intelligence	Stats Data modelling and Financial Planing
07	**08**	**09**
24.0 %	10.6 %	58.0 %
Cloud computing and Data Engineering	Probability and Applied Machine Learning	AI and Deep Learning and Natural Language Proc



#4 Adding Coop / Internship –
The internship would help to gain the soft skills and industrial experience so to be prepared for full time jobs

#3 Giving option of specialization -
A masters degree with a specialization certificate in skills like cloud computing can be of a great asset in job market

#2 Course sequencing based on agglomerative learning –
The courses are sequenced in a way to build up knowledge step by step, so that a student gets full understanding of the subject matter

#1 Course Evaluation based on % skills presence in jobs –
the course which have maximum presence in jobs are made mandatory core courses. (based on the skill clusters

Part 2 Result:
Masters Program Design –
Master of Data Science and Artificial Intelligence

Sequence	Course Title	Skills Gained	Industrial Job Scope Covered
Pre-Requisite 0	NEW MIE1624 – Intro. to Data Sc. & Analytics	Python | Regression | Classification | Neural Networks | Classification	--Pre-Requisite--
Core Courses 1	Data Visualization and Data Management	SQL | Database | Tableau | R Program | Data Visualization	74%
Core Courses 2	Statistics for data Science	Financial Planning | Data Modelling | Statistical Analysis	71%
Core Courses 3	Machine Learning at scale with Big Data	Big Data | Hadoop | Spark | Scala | Hive | HBase	43%
Core Courses 4	AI and deep learning tools in Marketing	NLP | Deep Learning | Predictive modeling | Neural Networks	58%
Core Courses 5	Applied machine learning	Probability | Artificial Intelligence | Linear Algebra	10%
Specialization: Could Computing 1	Deep Learning in the Cloud and at the Edge	NoSQL | Cassandra | Information Retrieval | AWS	3%
Specialization: Could Computing 2	Fundamentals of Data Engineering	Azure | Google Cloud Platform | PIG | Cloud Computing | Data Manage	24%
Specialization: Business Intelligence 1	Research Design and App. for Data Analysis	AB Testing | Google Analytics | Data Driven Decision making	12%
Specialization: Business Intelligence 2	Statistical Methods and Business Intelligence	Time Series Analysis | Process Automation | Statistical modeling	35%

Extra Curricular

#1. Capstone Project

One capstone project to be completed testing your skills gained from all the courses taken so far.

Skills Gained –

Innovation | Creativity | Time Management | Project Management

#2. Industrial Internship

A 5-month internship in the field of your specialization which guarantees you to gain valuable experience.

Skills Gained –

Leadership | Collaborative | Communication | Team Player

Part 2 Result:

Master of Data Science and Artificial Intelligence

Core Courses

Course 1:
Data Visualization and Data Management

Topics Covered

01 Fundamentals of Visualization Essential Design Principles for Tableau

02 Visual Analytics and Case Studies

03 Creating Dashboards and Storytelling

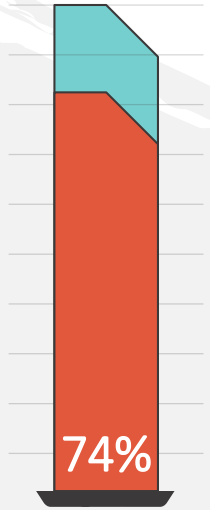
04 Data-driven Decision Making

Course Length : 10 – 11 Weeks

Skills Gained

Data Analysis | Data Visualization (DataViz) | Storyboarding | Computer Graphics | Tableau | R Programming | SQL

Job Scope Covered



Homework

#1 Assignment :
Data Visualization
#2 Assignment :
Story Telling – Data
#3 Group Project :
Data Dashboard

Course 2:
Statistics for Data Science

Topics Covered

01 Statistical Inference and Data Oriented Strategies

02 Linear models, ANOVA and ANCOVA and Analysis of residuals and variability

03 Building and applying prediction functions like regression, classification trees, Naive Bayes

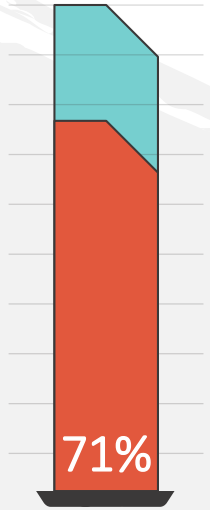
04 Production output from a statistical analysis and automate complex analysis

Course Length : 10 – 11 Weeks

Skills Gained

R Programming| Regression Analysis| Statistical Inference| Hypothesis Testing| Financial Planning| Model Selection

Job Scope Covered



Homework

#1 Assignment : Hypothesis Testing
#2 Assignment : Model Selection
#3 Group Project : Automate Analysis

Course 3:
Machine Learning at Scale with Big Data

Topics Covered

01 Data Manipulation at Scale: Systems and Algorithms – Big Data

02 Practical Predictive Analytics: Models and Methods

03 Graph Analysis in the Cloud, in which you will use Elastic MapReduce and the Pig language

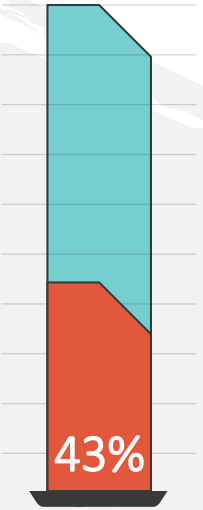
04 Use cloud computing and Big Data to analyze large datasets in a reproducible way.

Course Length : 10 – 11 Weeks

Skills Gained

Mathematical Optimization| SQL| Relational Algebra| Statistics| Hadoop| Spark | Scala | Hive | Hbase

Job Scope Covered



Homework

#1 Assignment :
Big Data Manipulate
#2 Assignment :
Cloud Graph Analysis
#3 Group Project :
Data Science at Scale

Course 4:
AI and Deep Learning tools in Marketing

Topics Covered

01 Customer Analytics - buying patterns and customer data collection

02 Artificial Intelligence and Deep Learning in marketing applications

03 People Analytics and Performance Evaluation

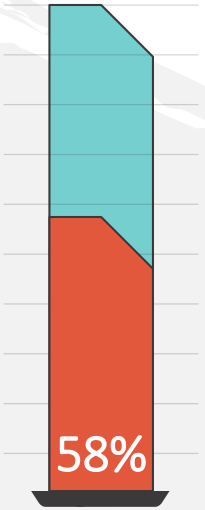
04 Accounting Analytics : financial statement data and non-financial metrics

Course Length : 10 – 11 Weeks

Skills Gained

Financial Analytics |Neural Networks| Deep Learning| Predictive Modeling| Natural Language Processing|

Job Scope Covered



Homework

#1 Assignment : Customer Analytics
#2 Assignment : Accounting Analytics
#3 Group Project : Business Analytics

Course 5:
Applied machine learning

Topics Covered

01 Fundamentals of Machine Learning

02 Mathematical approach to supervised learning methods for both classification and regression

03 Evaluation and model selection methods

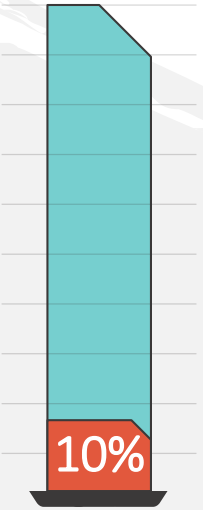
04 Mathematical approach to advanced supervised learning methods that include ensembles and Neural Nets

Course Length : 10 – 11 Weeks

Skills Gained

Probability | Artificial Intelligence | Linear Algebra | Mathematical simulation of Predictive Modeling

Job Scope Covered



Homework

#1 Assignment : Model Selection
#2 Assignment : Mathematical Proofs
#3 Group Project : Research Paper

Part 3:
Engineering Analytics Emphasis Course Selection Utility Recommender System

Course Selection Utility

With the tool, the best courses for are recommended to get the skills for a desired job

	Transportation Planner with Entuitive	Information Systems Analyst with Humber River Hospital	Data Scientist I with TD Bank
Required Course	APS1070: Foundations of Data Science and ML	APS1070: Foundations of Data Science and ML	APS1070: Foundations of Data Science and ML
Core Course	MIE1624: Introduction to Data Science and Analytics	MIE1624: Introduction to Data Science and Analytics	MIE1624: Introduction to Data Science and Analytics
Elective 1	CIV1507: Public Transport	MIE1623: Introduction to Healthcare Engineering	MIE1628: Big Data Science
Elective 2	APS1070: Supply Chain Mgmt and Logistics	APS502: Financial Engineering	APS1022: Financial Engineering II
Elective 3	CIV1532: Fundamentals of ITS and Traffic Mgmt	APS1005: Operations Research for Eng Mgmt	APS1040: Quality Control For Engineering Management

Thank You