Design of a Data Science Education System

Group 8

March 2020

Data Preparation

indeed
LinkedIn

- 1469 job-descriptions for Data Scientist
- 1714 job-descriptions for Data Manager

kaggle

- 2018 Kaggle Survey

Beautiful Soup
Web Scraping

coursera

- 672 online course-descriptions and 3325 chapter-descriptions in Data Science area
- 220 online course descriptions and 1071 chapter-descriptions in AI area

Mechanical & Industrial Engineering
UNIVERSITY OF TORONTO

- All 43 courses in Analytics emphasis from MIE department

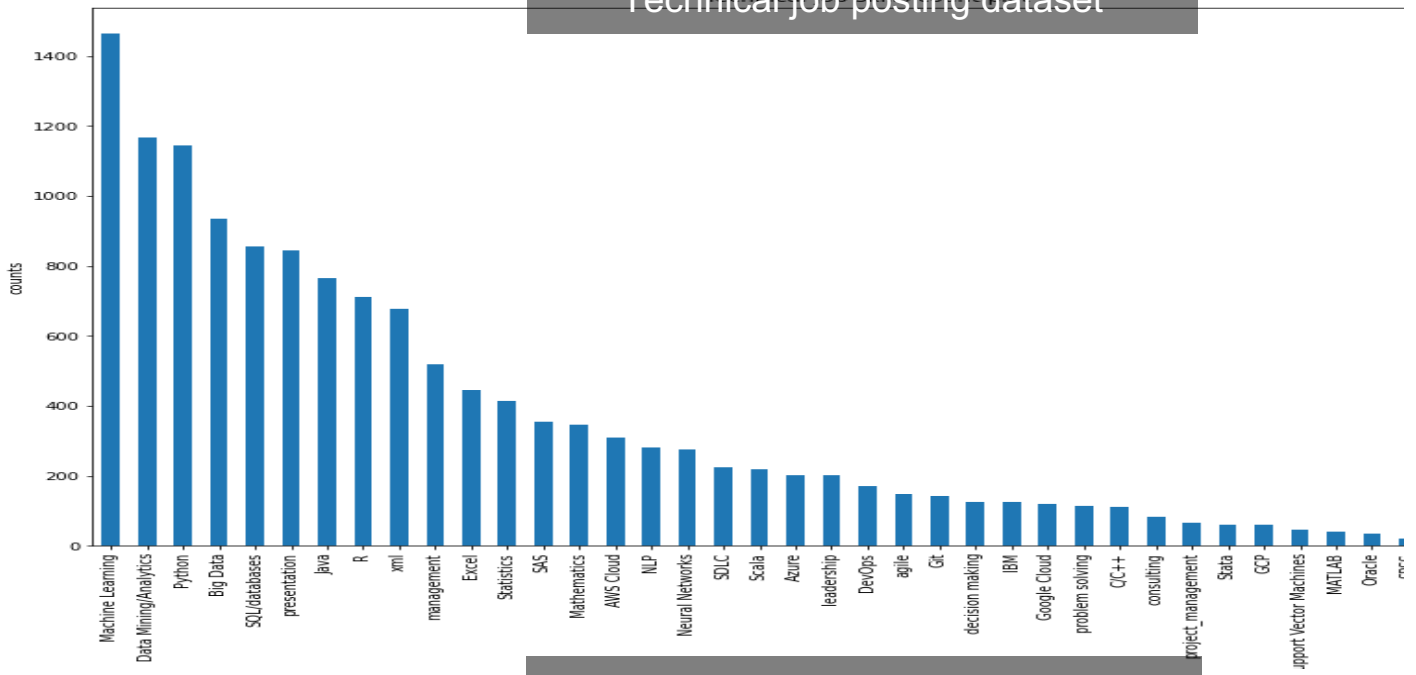
- 70 courses in Master of Business Analytics or Management Analytics from famous business schools across Canada and USA

HARVARD BUSINESS SCHOOL
IVEY Business School
WESTERN UNIVERSITY · CANADA
Queen's Smith SCHOOL OF BUSINESS
Rotman
Schulich School of Business
York University

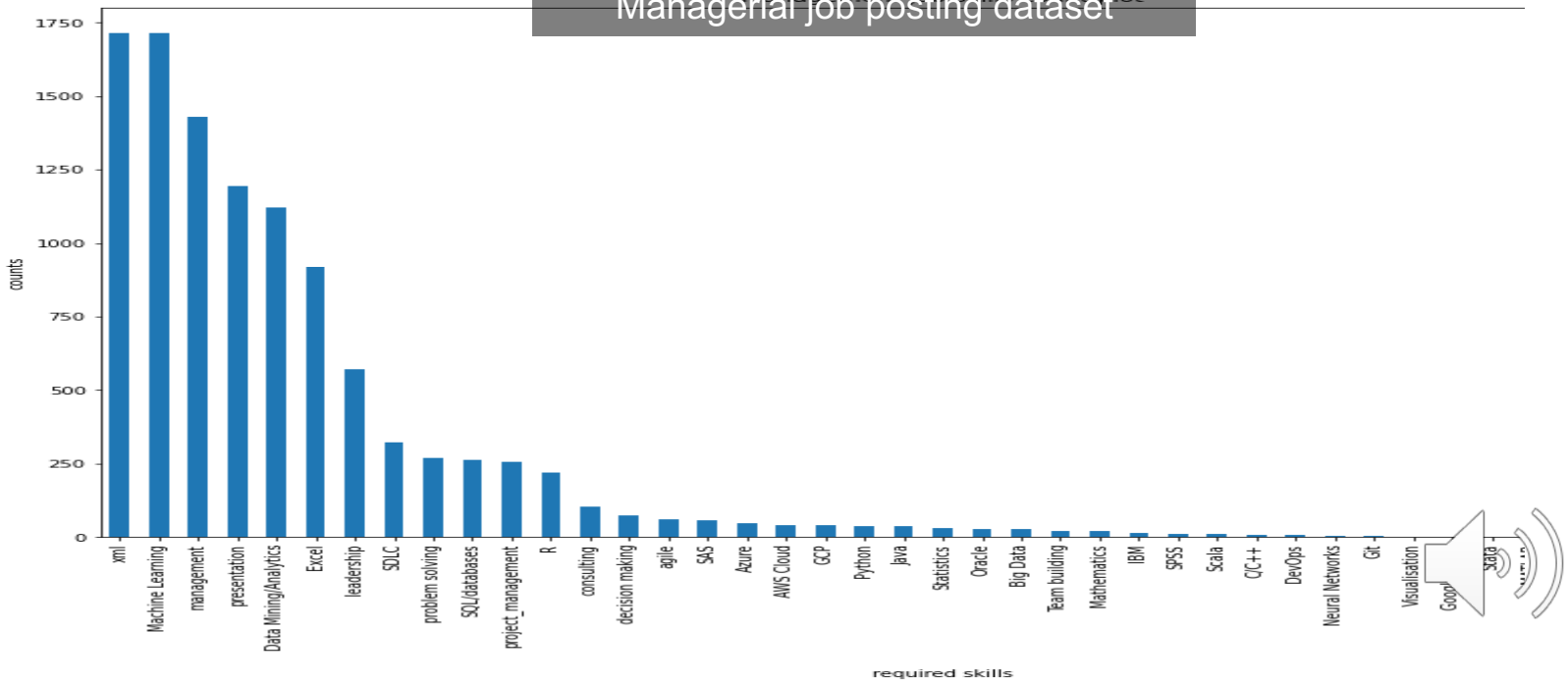
MIE 1624 Curriculum Design

Required Skills Count Plot of Job Posting Web-Scrapping Datasets

Technical job posting dataset



Managerial job posting dataset



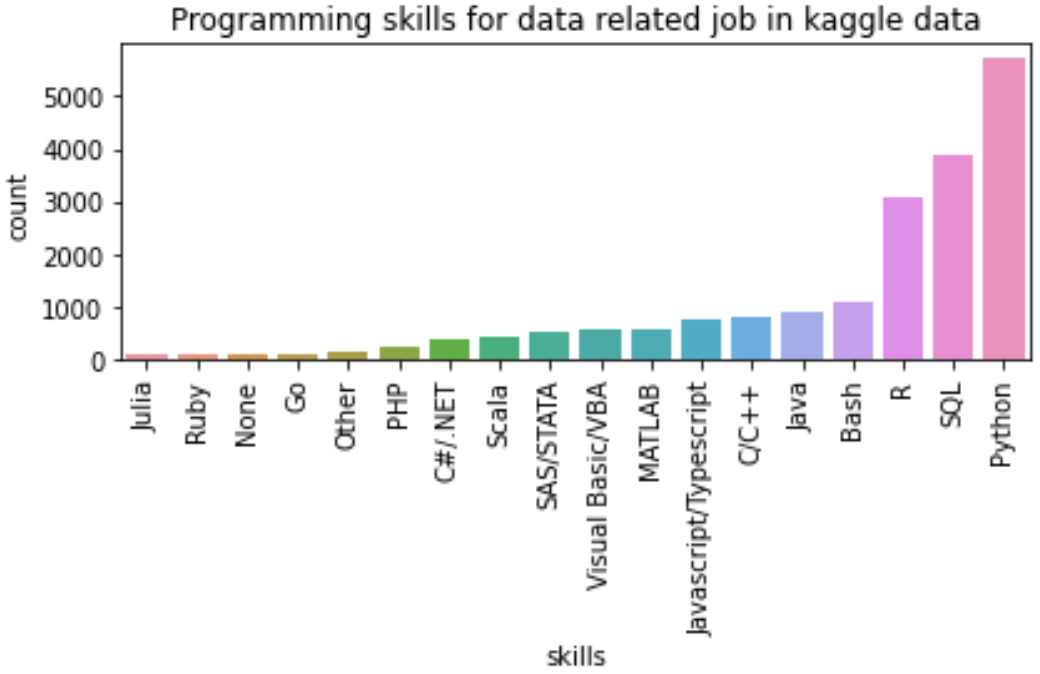
The above two bar charts contain top 43 required skills from technical job posting website and managerial job posting website web scraping datasets.

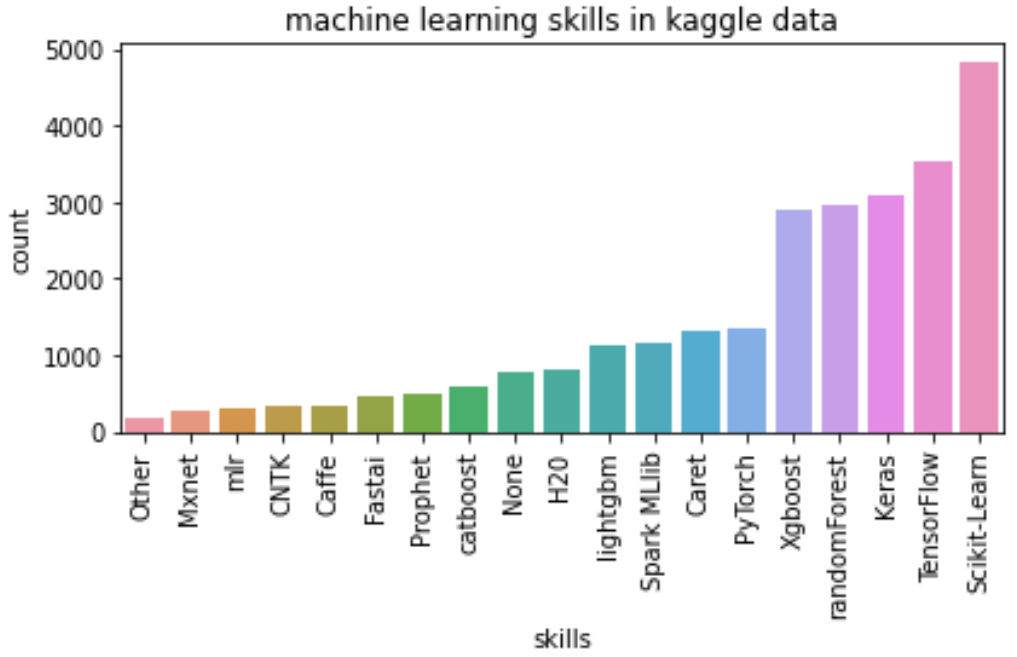
Those technical skills from two dataset can be divided into seven topics intuitively:

Programming languages	Mathematics and Statistics	Machine learning	Data Mining/Analytics	Natural Language Processing	Visualisation	Big data

For soft skills from managerial dataset :

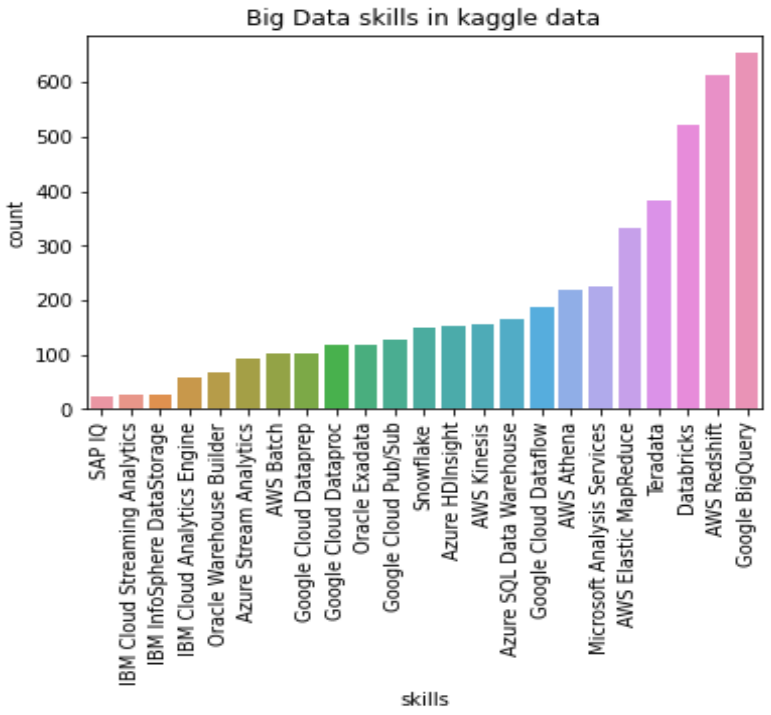
management, presentation, team building, decision making, consulting, project management, problem solving, leadership

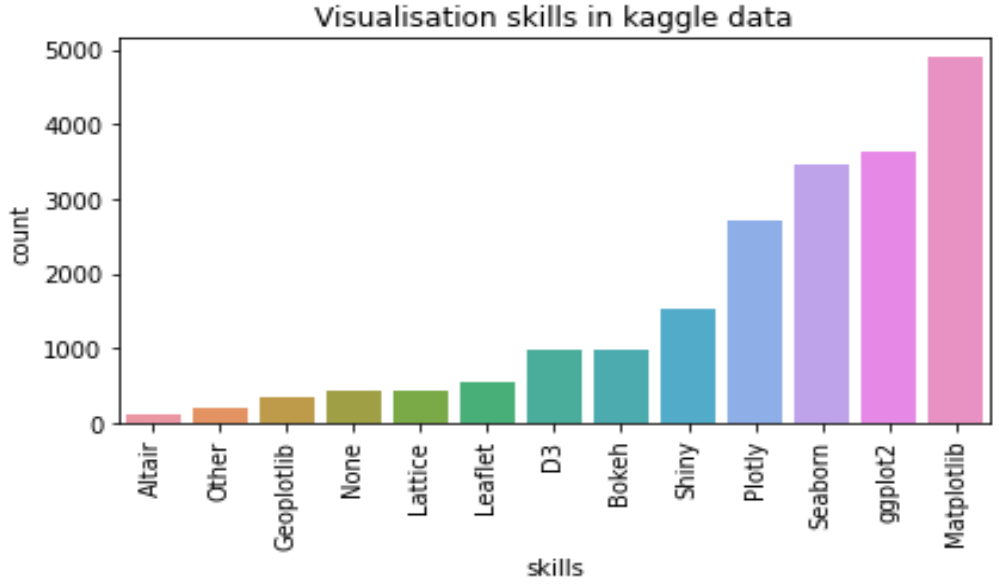




Kaggle ML and Data Science Survey Analysis

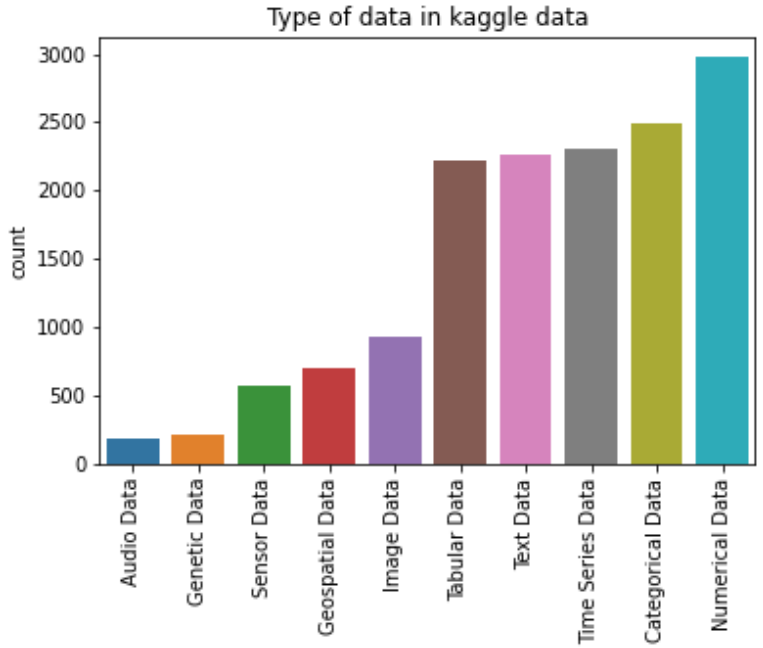
- Kaggle survey dataset will help us to find out what content should be designed under each topic.
- Machine learning skills:
  - scikit-learn, Tensorflow, Keras, RandomForest, XGBoost.
- Programming languages:
  - Python, SQL, R are the most popular programming skills among those data related workers.

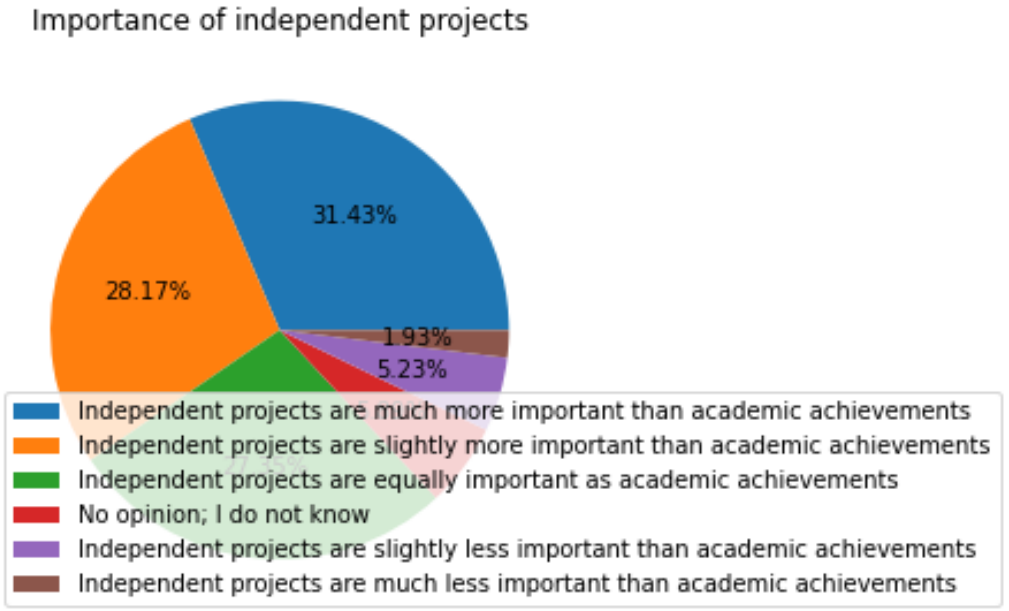




Kaggle ML and Data Science Survey Analysis

- Visualization: Matplotlib, seaborn should be added for this topic
- Big data: Google BigQuery, AWS Redshift, Databricks





Kaggle ML and Data Science Survey Analysis

- Pie chart shows that independent projects are much more important than academic achievements, therefore we will add two individual assignments for MIE1624

- **Course assignment/project design:**
- Numerical data, categorical data, time series data, text data.

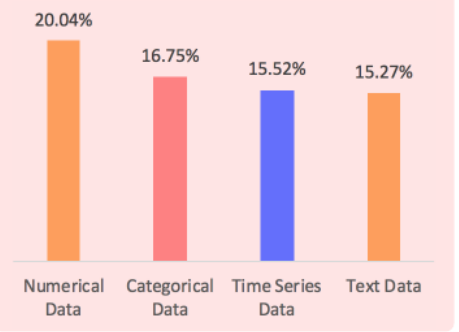
MIE 1624: Introduction to Data Science & Analytics

Topics	Subtopics
Introduction to Data Science	Intro to AI, Intro to Data Analytics, Course Description, Data Science Career Insight
Fundamental Mathematics	Probability, Statistics, Linear Algebra
Programming Basics	Python: list, dictionary, Pandas, Numpy, SciPy
Data Mining/Analytics	Data Type, Data Cleaning, Feature Selection
Introduction to Machine Learning	Basic: Linear Regression, Logistic Regression, SVM
	Advanced: Naive Bayes, Neural Network, NLP
	Tools: scikit-learn Implementation, Hyperparameter Tuning
Assignment 1	Numerical Data, Categorical Data, Classification, Regression Analysis, Visualization, Story Telling
Optimization	Convex Function, Gradient Descent, Local Minima, Slow Progress, Oscillations
Data Visualization	Matplotlib, Seaborn, ggplot2
Overview of Big Data	Google Bigquery, Hadoop, Mapreduce
Assignment 2	Time Series Data, Machine Learning Model Implementation, Visualization, Story Telling
Course Project	Text Data, Natural Language Processing, Machine Learning Models, Team Work, Presentation, Report

Support Analytics

About 78% of data science related jobs require working knowledge of Python.

TOP 3 visualization skills used by data scientists.



Both **technical and managerial** jobs mention leadership, management and presentation skills very frequently.

MIE 1624 Course Curriculum

Master of Data Science & Artificial Intelligence

Master of Data Science & Artificial Intelligence (MDSAI) Program

- **Program Admission requirements:**
- 1. Have a Science or Engineering bachelor's degree
- 2. Have completed and received at least B in courses:

  - ❑ Linear algebra
  - ❑ Multivariable calculus
  - ❑ Statistics
  - ❑ Programming

Master of Data Science & Artificial Intelligence (MDSAI) Program

- **Program Requirements:**

- 1. Complete 4 core and 6 elective courses in the first 3 semesters

- 2. Complete a 4-month internship in the last semester

  - Can be either industrial-based or research-based
  - Solve problems using professional tools (SAS, Google Cloud etc.)
  - Submit a report summarizing problem solving procedures

Master of Data Science & Artificial Intelligence (MDSAI)

Course Chapters:

1. Data Science

2. Artificial Intelligence

3. Business Analytics

Master of Data Science & Artificial Intelligence (MDSAI)

Data science:

1. (Core) Understand the fundamental mathematical knowledge and python programming skills

Course content: linear algebra, statistics, Python

2. (Core) Develop working knowledge of database & SQL

Course content: Select/join/manage table, SQL function, SQL programming interface

2. (Elective) Study other popular data analysis tools and techniques

Course content: R, Excel, NLP

3. (Elective) Develops basics for web programming

Course content: HTML, CSS, JavaScript

Master of Data Science & Artificial Intelligence (MDSAI)

Artificial Intelligence:

1. **(Core) Understand common machine learning & data mining algorithms**

   Course content: supervised learning, unsupervised learning, reinforcement learning

2. **(Elective) Develop basics of computer vision and cloud computing**

   Course content: image recognition & segmentation, IBM & Google & Amazon clouds

3. **(Elective) Investigate deeply in decision theory and game theory**

   Course content: game trees, Bayesian models, Nash Equilibrium

4. **(Elective) Understand reinforcement learning techniques**

   Course content: Markov decision process, value function, Q-learning

Master of Data Science & Artificial Intelligence (MDSAI)

Business Analytics:

1. (Core) Understand how to apply data science skills to make marketing decisions

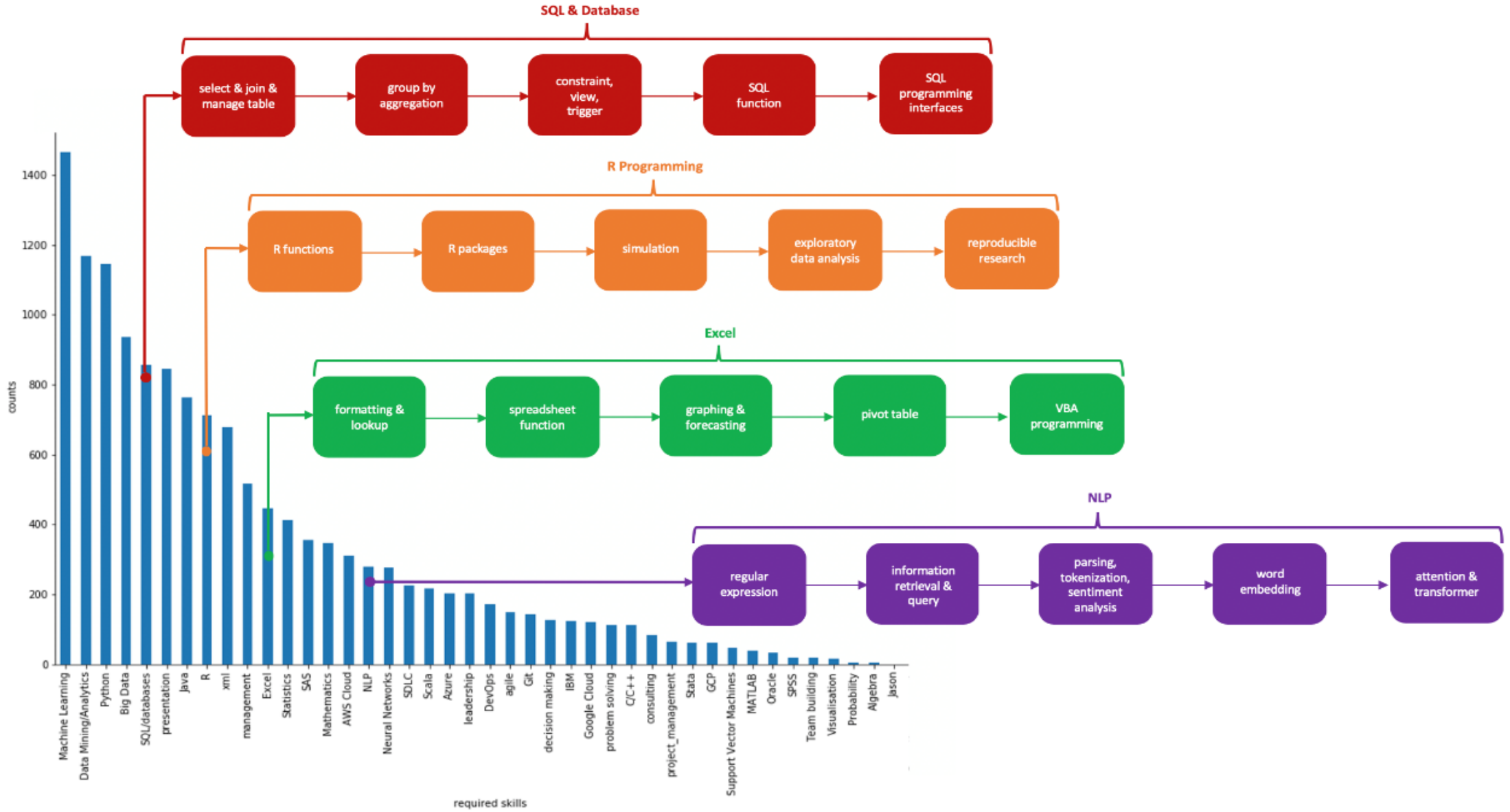
Course content: demand & supply simulation, finance

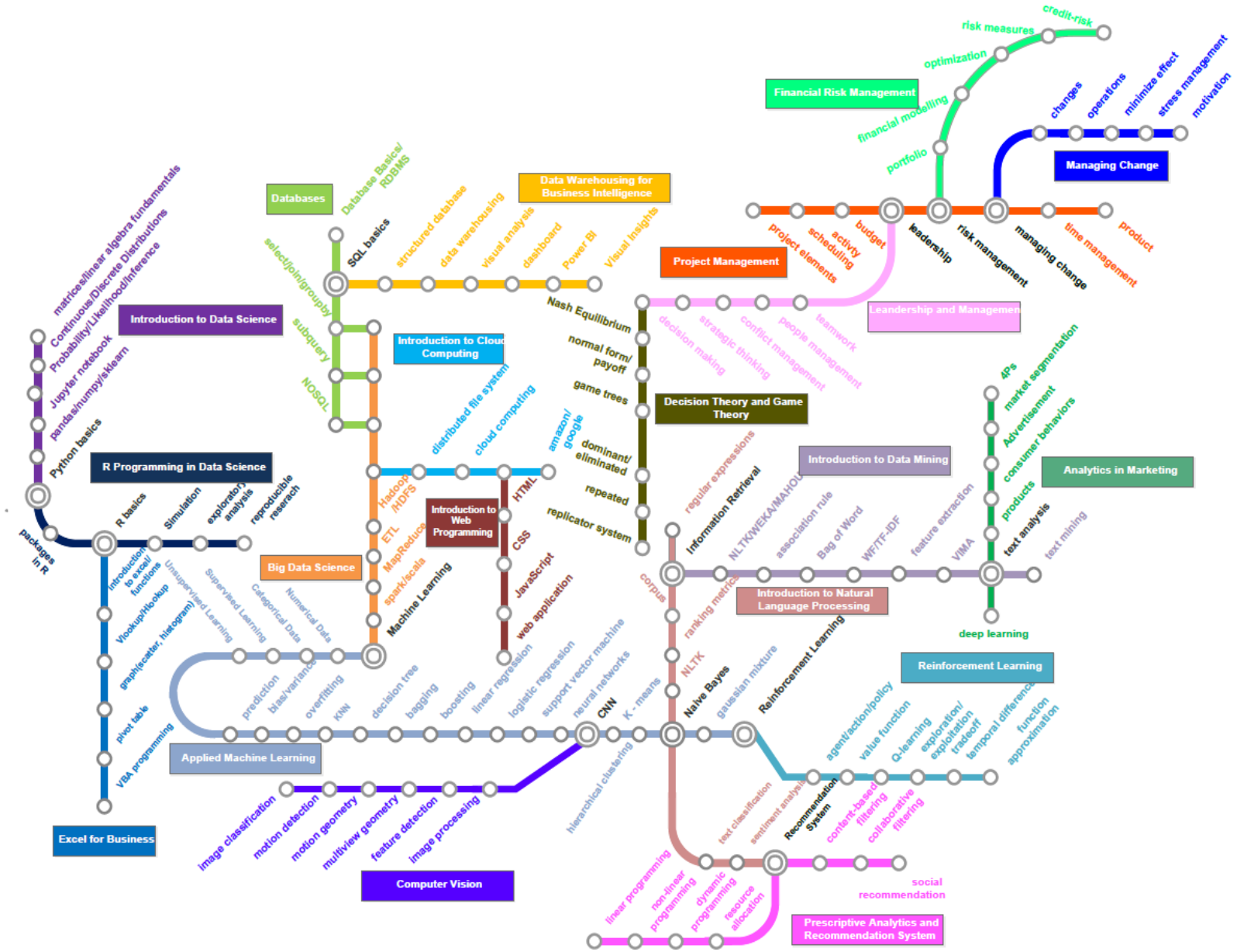
2. (Elective) Improve management and leadership skills

Course content: project & risk & team management

3. (Elective) Study advanced business intelligence tools

Course content: data warehouse, recommendation system, prescriptive analytics

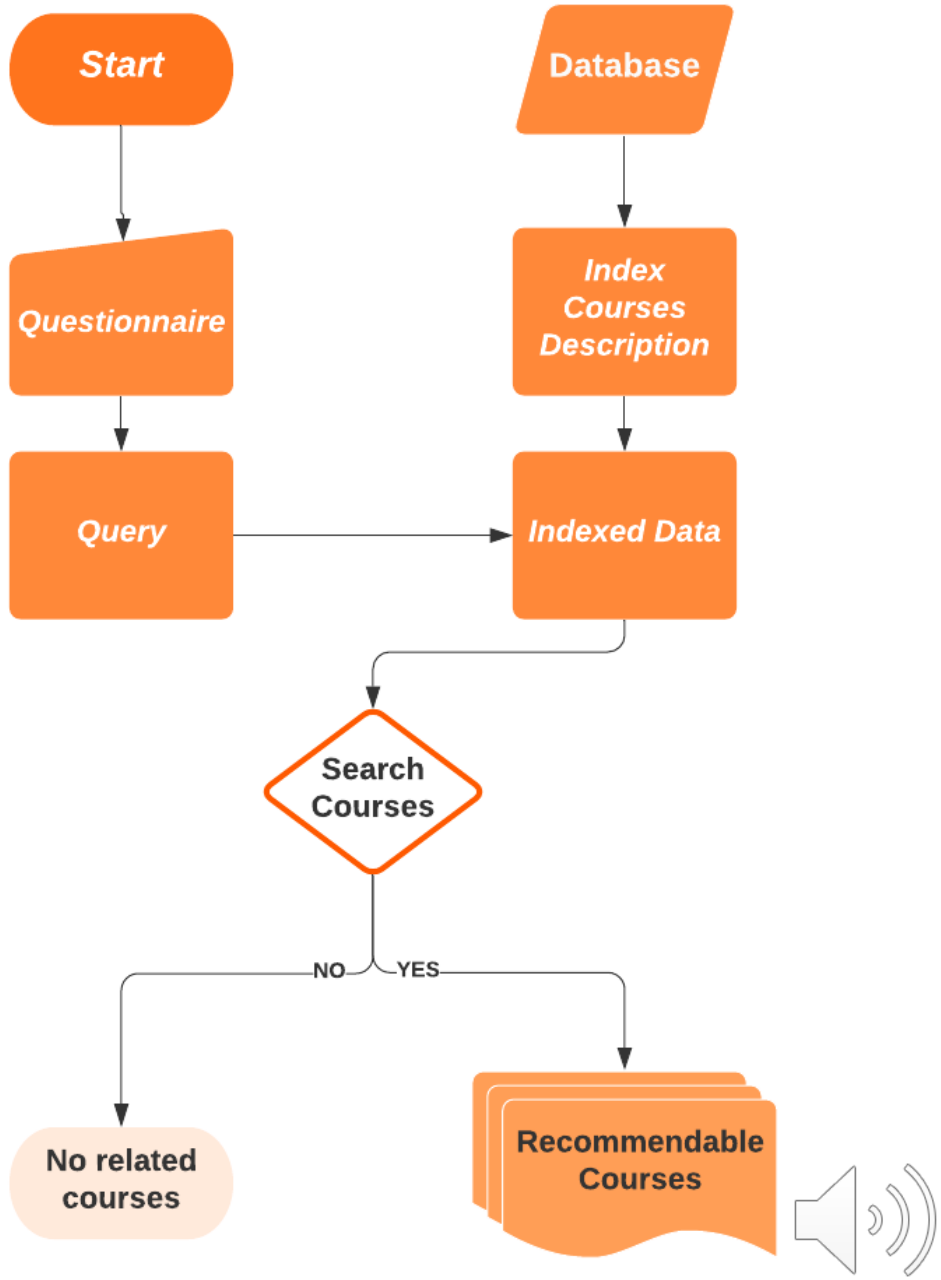

SQL & Database
select & join & manage table
group by aggregation
constraint, view, trigger
SQL function
SQL programming interfaces
R Programming
R functions
R packages
simulation
exploratory data analysis
reproducible research
Excel
formatting & lookup
spreadsheet function
graphing & forecasting
pivot table
VBA programming
NLP
regular expression
information retrieval & query
parsing, tokenization, sentiment analysis
word embedding
attention & transformer
counts
0
200
400
600
800
1000
1200
1400
Machine Learning
Data Mining/Analytics
Python
Big Data
SQL/databases
presentation
Java
R
xml
management
Excel
Statistics
SAS
Mathematics
AWS Cloud
NLP
Neural Networks
SDLC
Scala
Azure
leadership
DevOps
agile
Git
decision making
IBM
Google Cloud
problem solving
C/C++
consulting
project_management
Stata
GCP
Support Vector Machines
MATLAB
Oracle
SPSS
Team building
Visualisation
Probability
Algebra
Jason
required skills


Introduction to Data Science
matrices/linear algebra fundamentals
Continuous/Discrete Distributions
Probability/Likelihood/Inference
Jupyter notebook
pandas/numpy/sklearn
Python basics
R Programming in Data Science
packages in R
R basics
Simulation
exploratory analysis
reproducible research
Excel for Business
Introduction to excel/ functions
Unsupervised Learning
Supervised Learning
Categorical Data
Numerical Data
VlookupHlookup
graph(scatter, histogram)
pivot table
VBA programming
Databases
Database Basics/ RDBMS
SQL basics
structured database
data warehousing
visual analysis
dashboard
Power BI
Visual Insights
select/join/groupby
subquery
NOSQL
Data Warehousing for Business Intelligence
Introduction to Cloud Computing
distributed file system
cloud computing
amazon/ google
Big Data Science
Hadoop /HDFS
ETL
MapReduce
spark/scala
Introduction to Web Programming
HTML
CSS
JavaScript
web application
Machine Learning
prediction
bias/variance
overfitting
KNN
decision tree
bagging
boosting
linear regression
logistic regression
support vector machine
neural networks
CNN
K - means
hierarchical clustering
Applied Machine Learning
image classification
motion detection
motion geometry
multiview geometry
feature detection
image processing
Computer Vision
Financial Risk Management
credit-risk
risk measures
optimization
financial modelling
portfolio
Project Management
project elements
activity scheduling
budget
leadership
risk management
Managing Change
changes
operations
minimize effect
stress management
motivation
managing change
time management
product
Leandership and Management
decision making
strategic thinking
conflict management
people management
teamwork
Nash Equilibrium
normal form/ payoff
game trees
dominant/ eliminated
repeated
replicator system
Decision Theory and Game Theory
Introduction to Data Mining
regular expressions
Information Retrieval
NLTK/WEKA/MAHOUT
association rule
Bag of Word
WF/TF-IDF
feature extraction
VIMA
Introduction to Natural Language Processing
corpus
ranking metrics
NLTK
Naive Bayes
gaussian mixture
Analytics in Marketing
4Ps
market segmentation
Advertisement
consumer behaviors
products
text analysis
text mining
deep learning
Reinforcement Learning
Reinforcement Learning
agent/action/policy
value function
Q-learning
exploration/ exploitation tradeoff
temporal difference
function approximation
Recommendation System
text classification
sentiment analysis
content-based filtering
collaborative filtering
social recommendation
Prescriptive Analytics and Recommendation System
linear programming
non-linear programming
dynamic programming
resource allocation

Recommender System

Goal of our recommender system:

A recommender system is built to help potential users select courses to satisfy their demands. Users may not have sufficient technical skill, and management skill to fit the requirement of data scientist jobs. The demand could come from the job requirements, the interested skills and ability of users.



Question 1: What kind of job do you want to find? (eg. big data, machine learning, natural language processing, ex..)

machine learning

Question 2: what kind of programming language do you want to learn? (eg. python, java, R, SQL, ex..)

python

Question 3: What kind of skills do you need to learn for your future career? (eg. visualization, mathematics, SAS, ...ex)

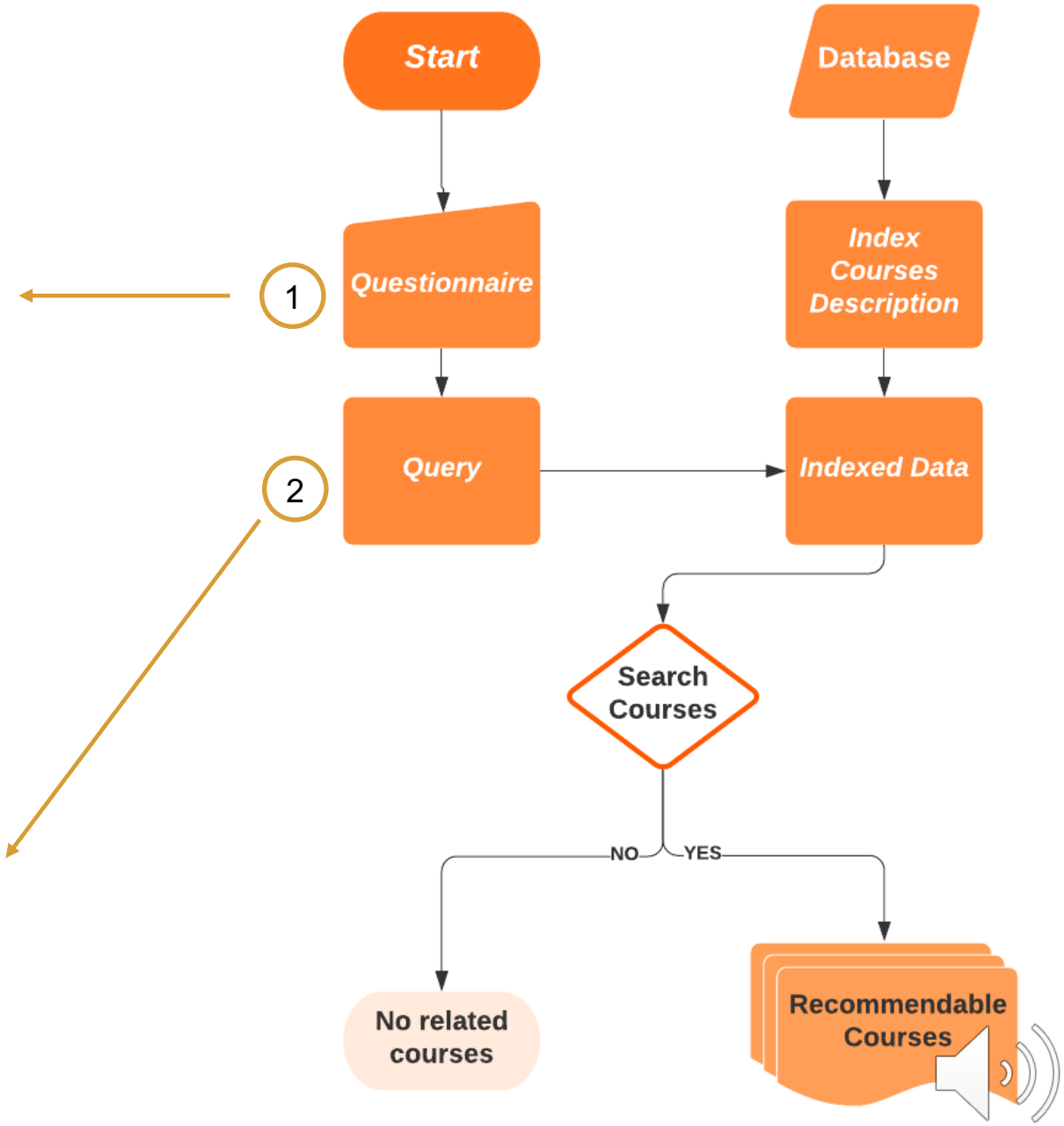
visualization and linear optimization

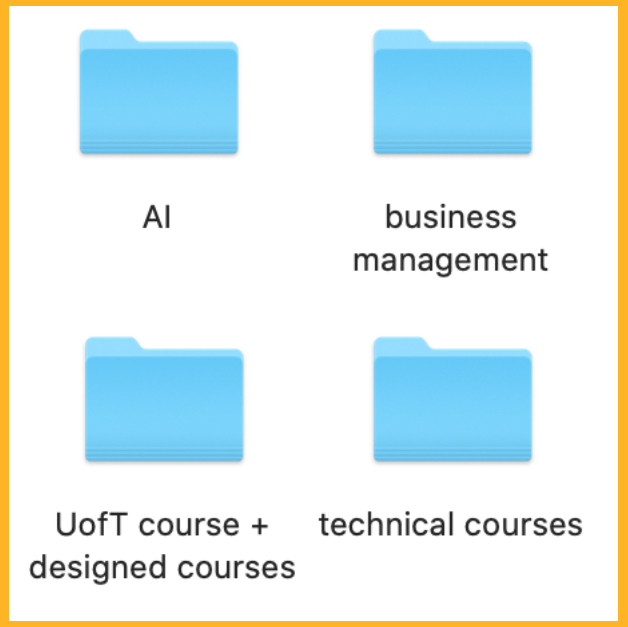
Question 4: What kind of management ability do you need to learn for your future career? (eg. decision making, teamwork, presentation,...ex)

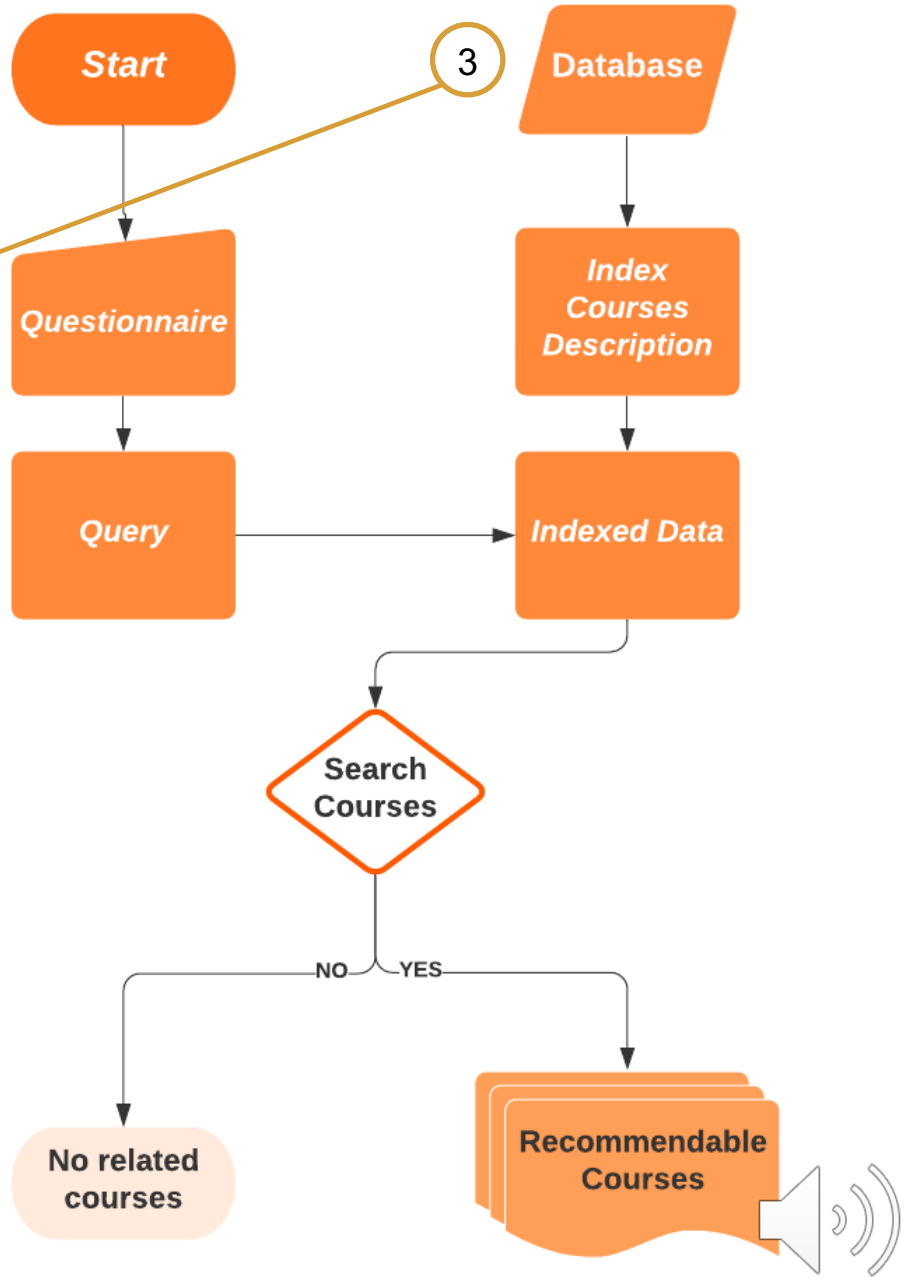
presentation

Question 5: What else knowledge do you want to learn to improve yourself? (eg. you can type nothing or linear optimization ex..)

Machine learning, python, visualization and linear optimization, presentation




AI
business management
UofT course + designed courses
technical courses


Start
3
Database
Questionnaire
Index Courses Description
Query
Indexed Data
Search Courses
NO
YES
No related courses
Recommendable Courses

Online courses:

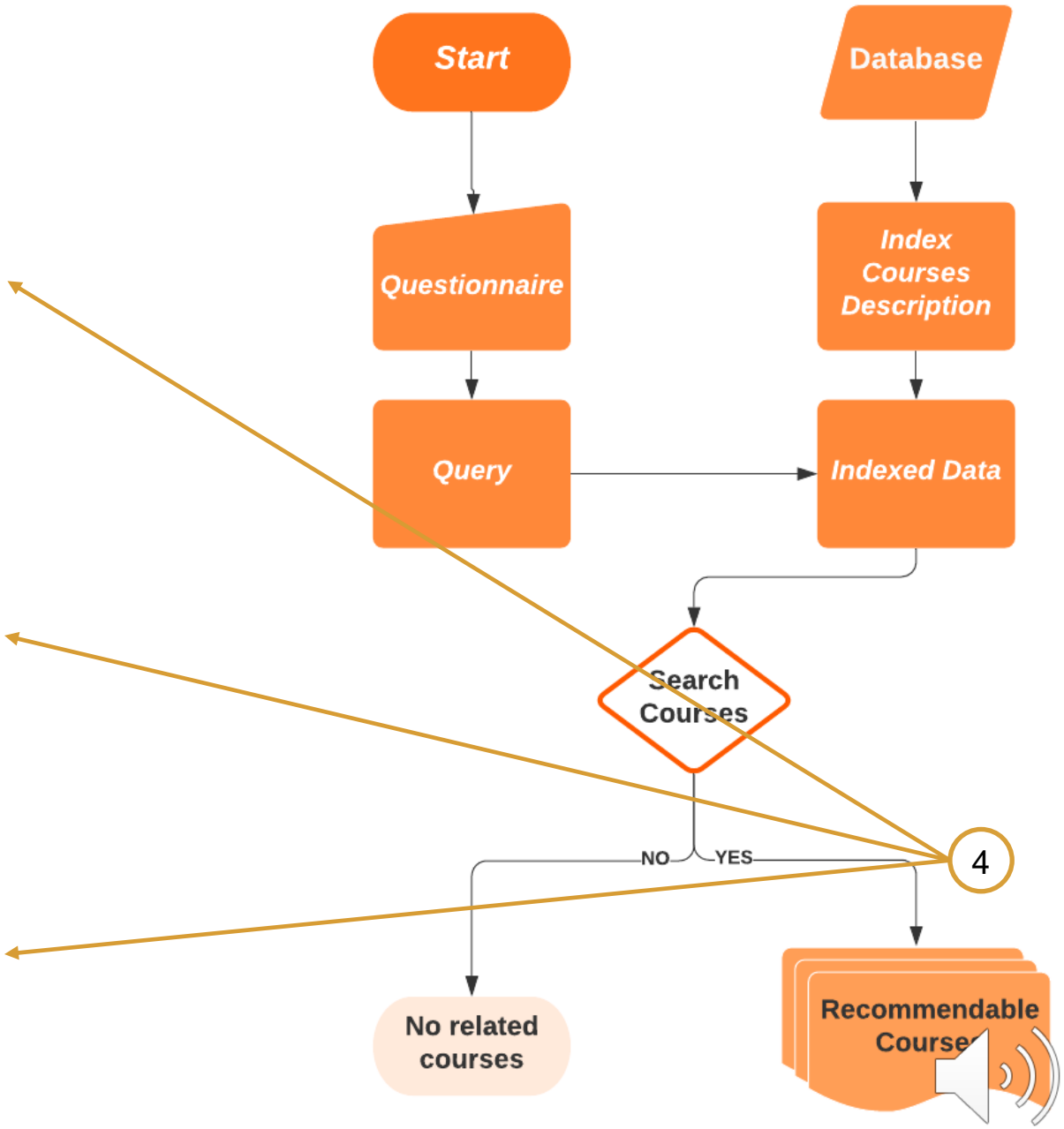
	Course Names	Course Descriptions	Chapter Links
470	Mathematics for Machine Learning Specialization	For a lot of higher level courses in Machine L...	https://www.coursera.org/learn/linear-algebra-...
17	Prescriptive Analytics and Optimization	This course covers formulations, applications ...	NaN
1150	Interactive Data Visualization with Bokeh	Bokeh is an interactive data visualization lib...	NaN
582	Machine Learning for Trading Specialization	This Specialization is for finance professiona...	https://www.coursera.org/learn/introduction-tr...

UofT Elective courses at Emphasis in Analytics

	Course Names	Course Descriptions
1	MIE 1624H Introduction to Data Science and Ana...	The objective of the course is to learn analyt...
10	APS 1052H: A.I. in Finance	In this course we'll give an overview of sever...
6	APS 1022H: Financial Engineering II	The course presents two important topics in fi...
2	ECE 1513H Introduction to Machine Learning (ex...	An Introduction to the basic theory, the funda...
19	ECE1504H: Statistical Learning (new course off...	This course will provide a mathematical introd...

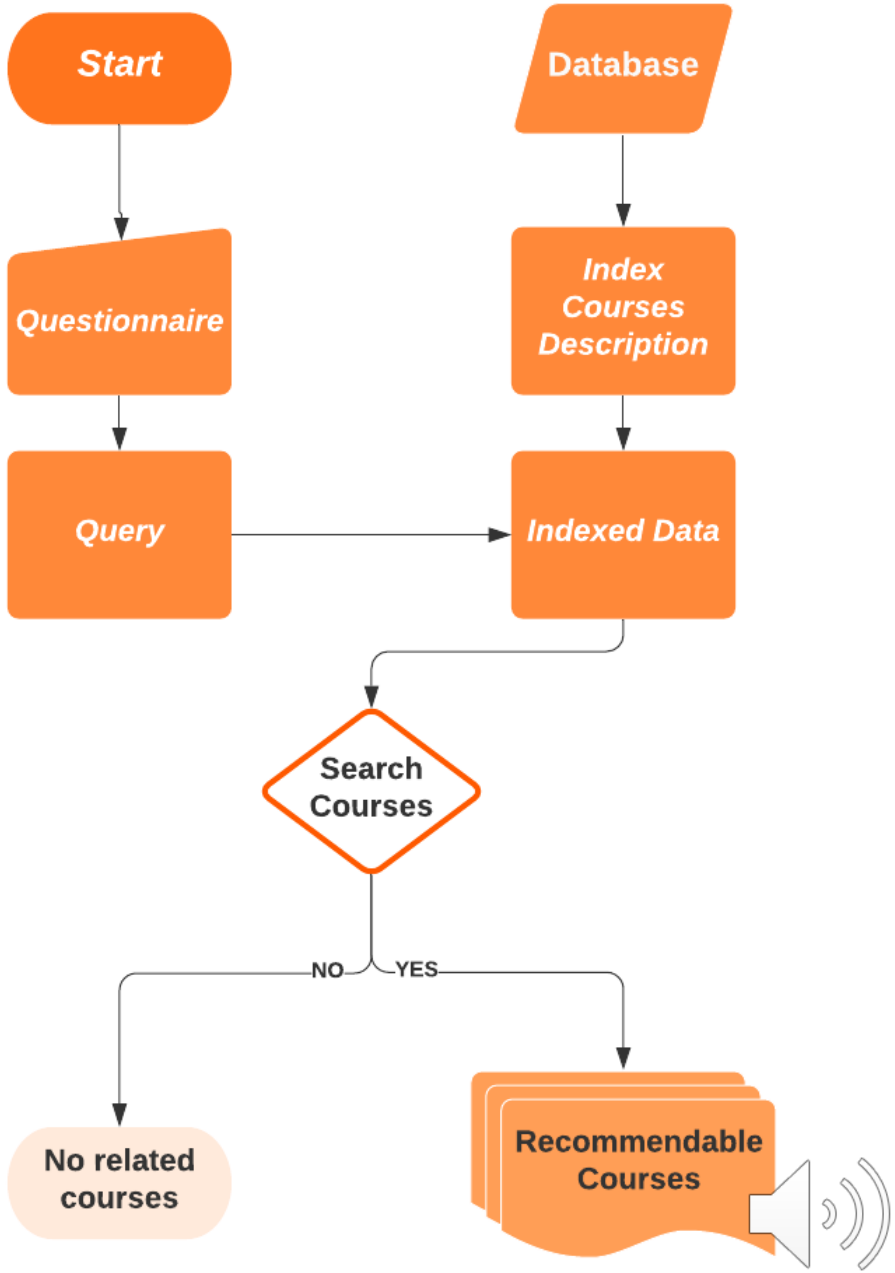
Redesigned courses

	Course Names	Course Descriptions
7	Applied Machine Learning	Machine learning brings together computer scie...
3	Big Data Science	You will gain an understanding of what insight...
0	Introduction to Data Science	This course will help students not only review...
2	Introduction to Natrual Language Processing	This course will cover most of the important t...
14	Financial Risk Management	This course combines the knowledge of finance ...



Conclusion:

- more advanced and more professional than other tools
- give a combo of selected courses, which includes most demand of a user
- provide the more suitable courses



Thanks for Listening !