TU 257 – Fundamentals of Data Science

Data Analytics

L0 - Module Admin

Brendan Tierney

Agenda

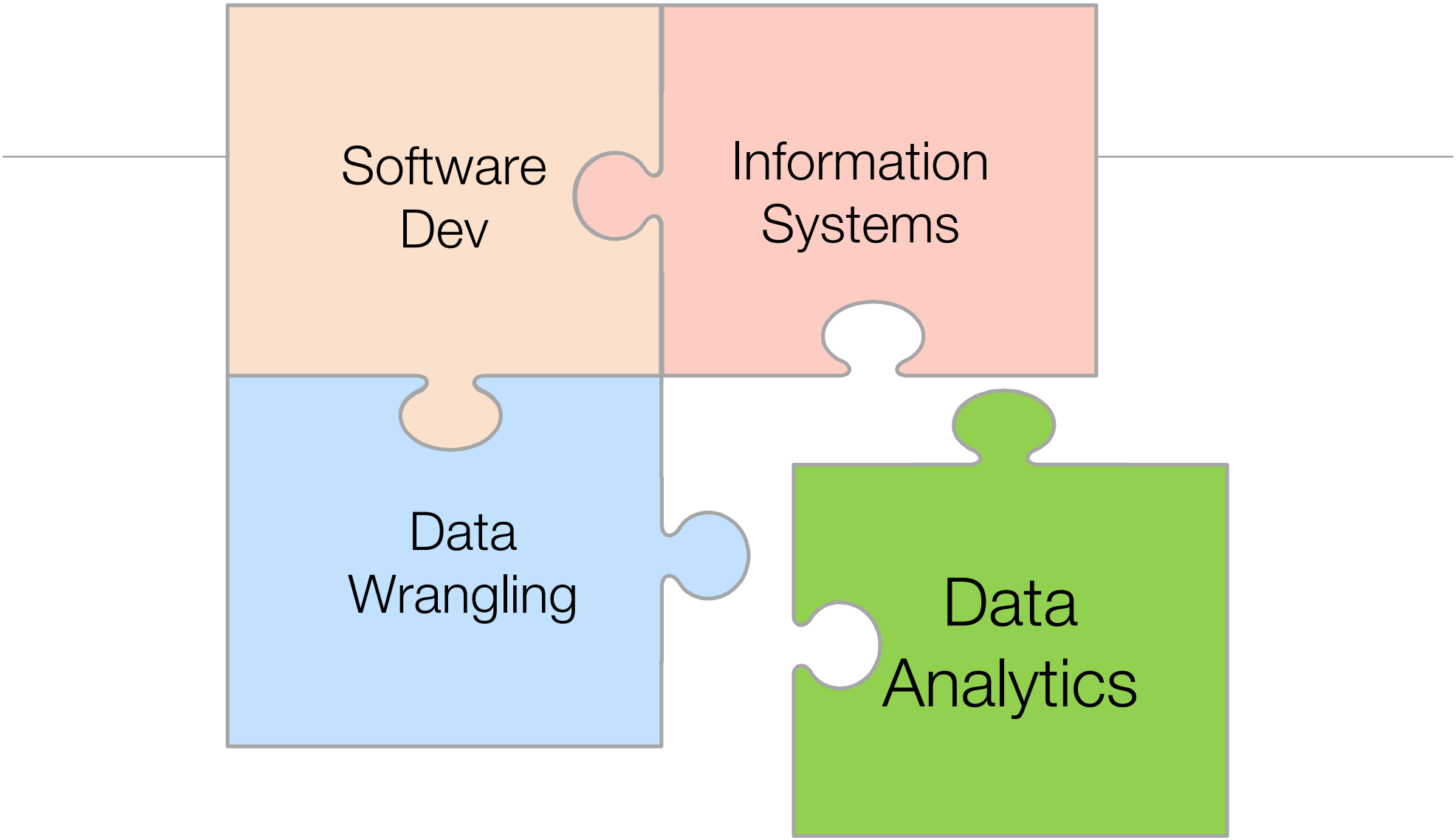
- What this module is about
- Learning Outcomes
- Weekly Schedule
- Lecture / Lab / Activities
- Assessments
- Useful Resources

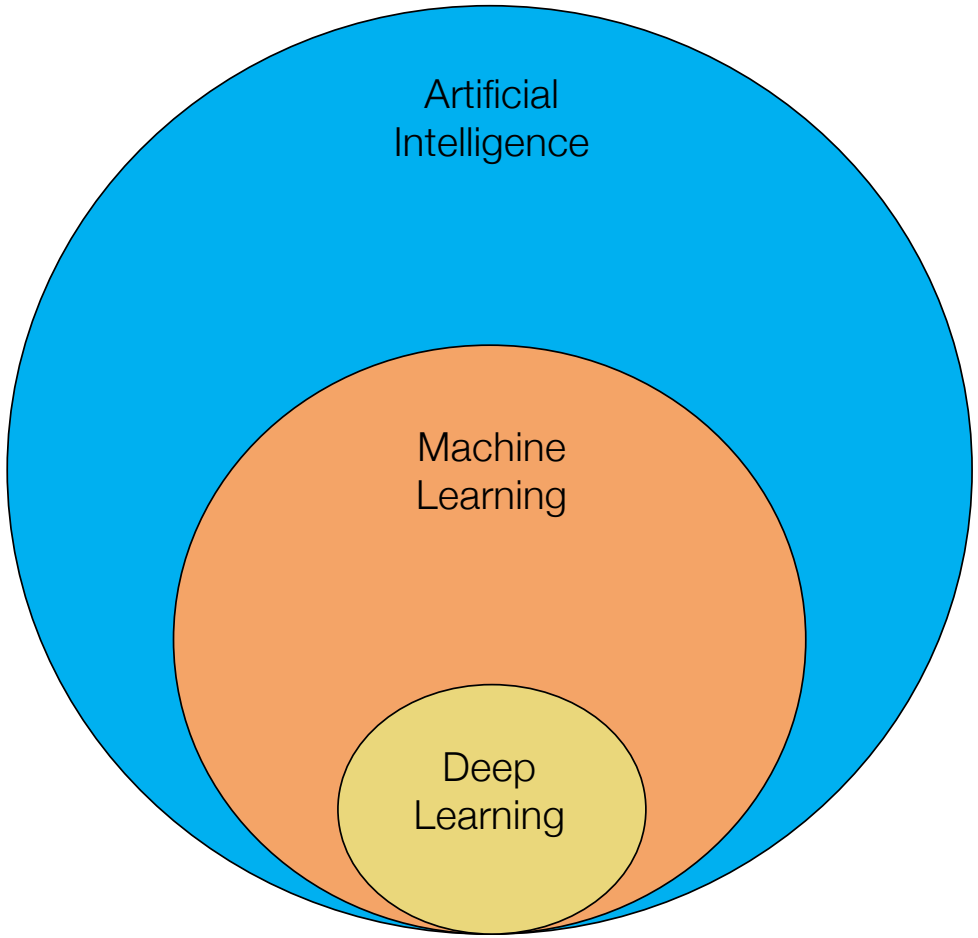
What is this module about

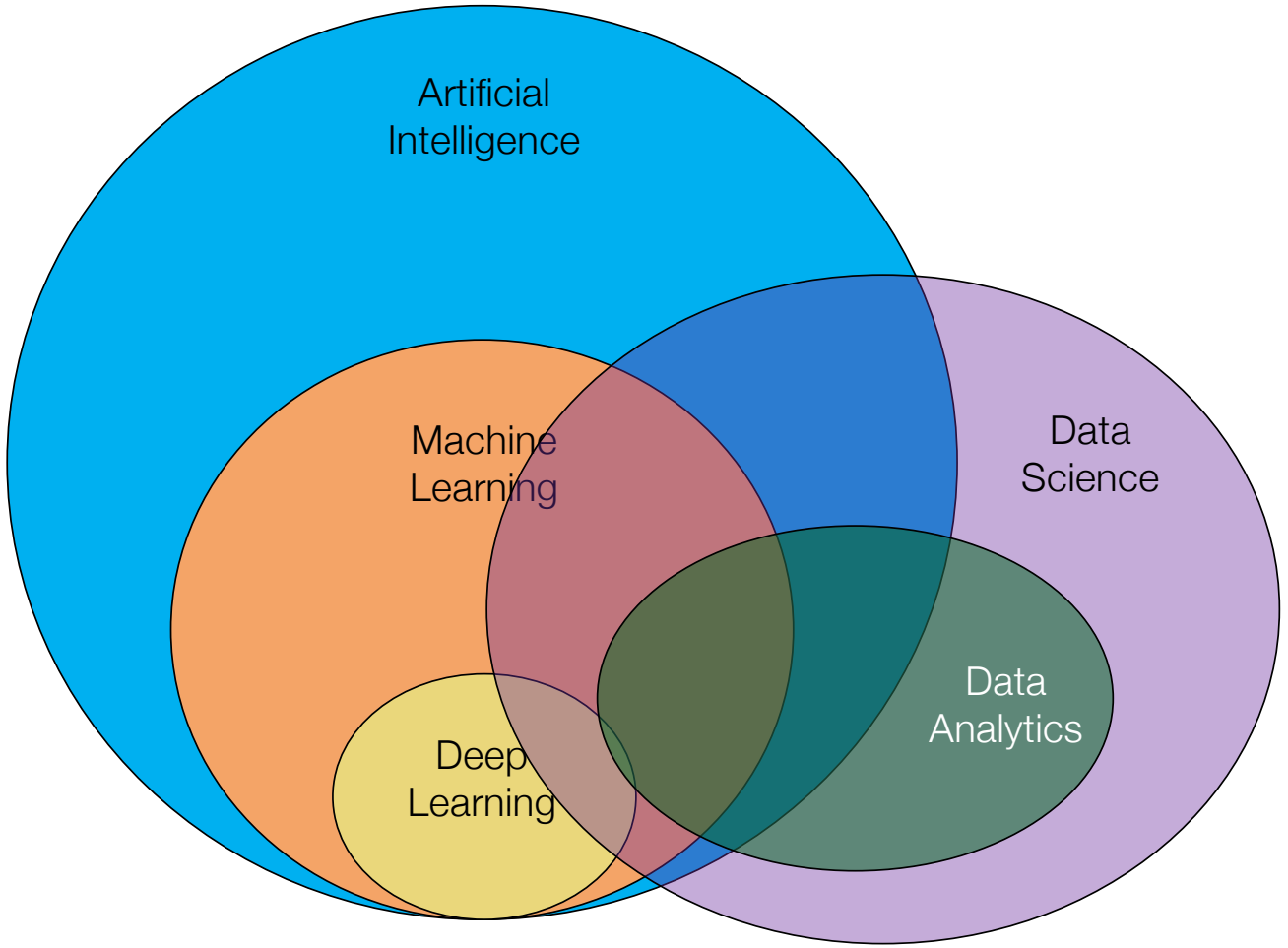
- Data analytics is an area of increasing importance and interest to organisations.

- Data analytics techniques offer huge potential in the creation of new knowledge products and services and the enhancement of existing products and services.

- Rather than focus on the details of specific data analytics techniques, this module addresses the application of data analytics techniques (from simple descriptive analytics techniques to more complex predictive analytics techniques) to real business problems.


Software
Dev
Information
Systems
Data
Wrangling
Data
Analytics

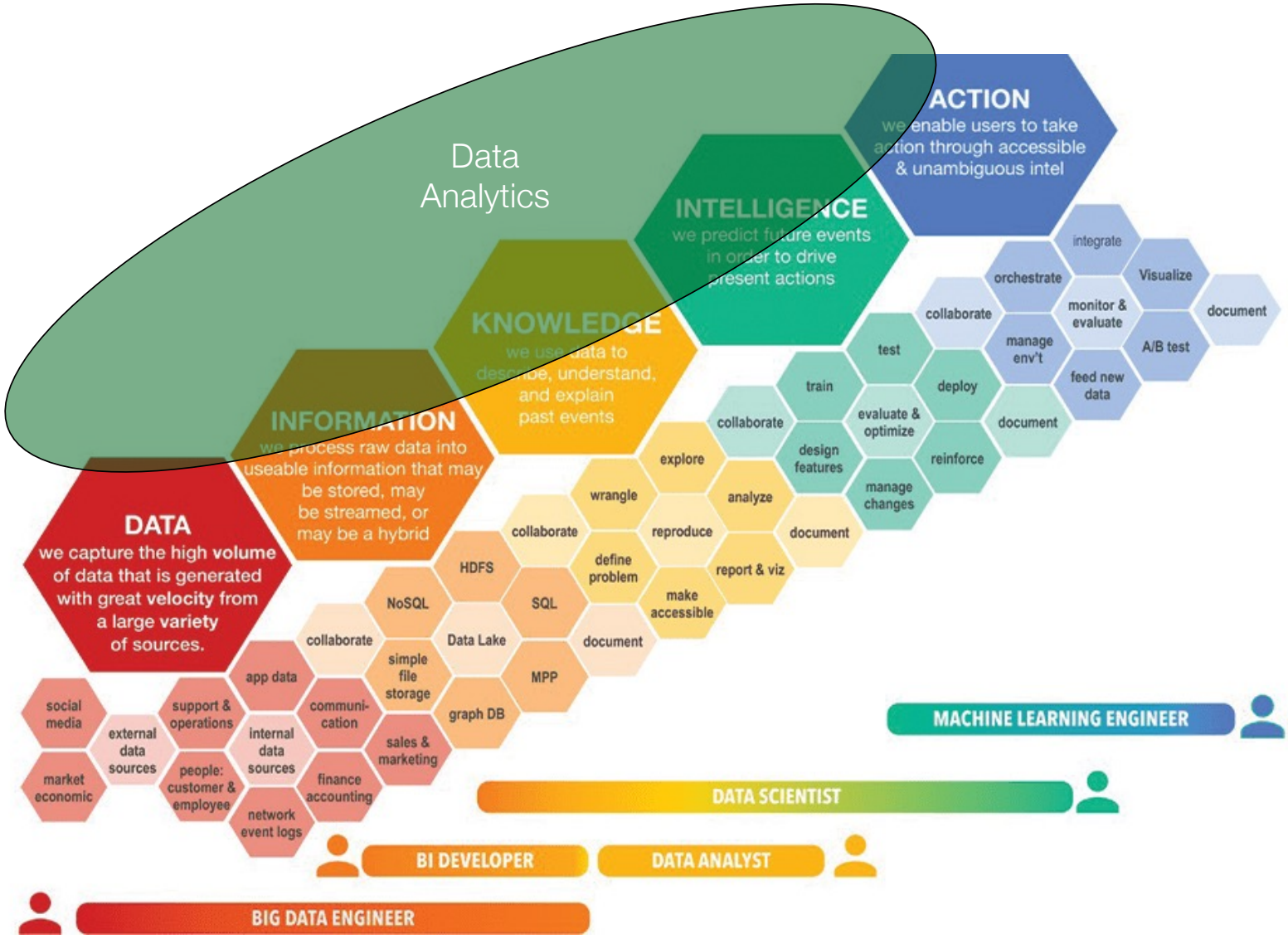

Artificial
Intelligence
Machine
Learning
Deep
Learning


Artificial
Intelligence
Machine
Learning
Deep
Learning
Data
Science
Data
Analytics

Important

- Is not Machine Learning
  - Machine Learning is focused on internals of algorithm
  - This will be covered in a Machine Learning module
- Will not look at the internal Mathematics of the algorithms
- Will not examine the theory behind individual algorithms
  - This will be covered in a Machine Learning module
- This module will take an Applied approach
  - You don't need to be a mechanic, or electrician, or engine designer to drive a car
- This module will focus on Process, Tasks, How to use, Applied, …


ACTION
we enable users to take action through accessible & unambiguous intel
INTELLIGENCE
we predict future events in order to drive present actions
KNOWLEDGE
we use data to describe, understand, and explain past events
INFORMATION
we process raw data into useable information that may be stored, may be streamed, or may be a hybrid
DATA
we capture the high volume of data that is generated with great velocity from a large variety of sources.
integrate
orchestrate
Visualize
collaborate
monitor & evaluate
document
manage env't
A/B test
test
feed new data
train
deploy
collaborate
evaluate & optimize
document
explore
design features
reinforce
wrangle
analyze
manage changes
collaborate
reproduce
document
HDFS
define problem
report & viz
NoSQL
SQL
make accessible
collaborate
Data Lake
document
app data
simple file storage
MPP
social media
support & operations
communi-cation
graph DB
external data sources
internal data sources
sales & marketing
market economic
people: customer & employee
finance accounting
network event logs
MACHINE LEARNING ENGINEER
DATA SCIENTIST
BI DEVELOPER
DATA ANALYST
BIG DATA ENGINEER


Data
Analytics
DATA
we capture the high volume of data that is generated with great velocity from a large variety of sources.
INFORMATION
we process raw data into useable information that may be stored, may be streamed, or may be a hybrid
KNOWLEDGE
we use data to describe, understand, and explain past events
INTELLIGENCE
we predict future events in order to drive present actions
ACTION
we enable users to take action through accessible & unambiguous intel
social media
market economic
external data sources
support & operations
people: customer & employee
app data
internal data sources
network event logs
communi-cation
finance accounting
sales & marketing
collaborate
simple file storage
NoSQL
HDFS
Data Lake
SQL
graph DB
MPP
collaborate
define problem
make accessible
document
collaborate
wrangle
reproduce
explore
analyze
report & viz
document
collaborate
design features
train
evaluate & optimize
test
deploy
reinforce
manage changes
document
collaborate
manage env't
orchestrate
monitor & evaluate
integrate
feed new data
A/B test
Visualize
document
MACHINE LEARNING ENGINEER
DATA SCIENTIST
BI DEVELOPER
DATA ANALYST
BIG DATA ENGINEER

Learning Outcomes

- Discuss the role of data analytics in an organisation
- Develop appropriate data analytics solutions to business problems
- Perform data analytics tasks
- Evaluate the range of different data analytics techniques and their data requirements
- Analyse and evaluate the suitability of different data analytics techniques
- Analyse case studies on how data analytics is used in different organisations
- Discuss the role of data management in data analytics and associated legal and ethical issues

Weekly Schedule

- Provisional Weekly Schedule
  - Subject to change - Check main Module webpage/VLE for any changes.

Week	Topic
Week 1	Introduction & Overview
Week 2	Bank Holiday
Week 3	Data Analytics & Data Science Lifecycle
Week 4	Data Exploration and Data Preparation
Week 5	Data Classification
Week 6	Data Classification
Week 7	Model Tuning & AutoML
Week 8	Bank Holiday
Week 9	Association Rules & Market Basket Analysis - Assessment-A Due
Week 10	Clustering
Week 11	Text Mining
	Spring Break
Week 12	Assessment-B Q&A Support session Self-Study topic on Legal & Ethical Issues for DS, ML & AI
Week 13	Bank Holiday - Assessment-B Due

4 hour Time Slot !

1 to 1.5 hours on Lecture + Demos

2 hours on Lab work
- Review & Rerun Demo code/notebooks
- Answer questions on Demo code/notebooks
- Expand (write code) on Demo code/notebooks
- Apply skills to a different dataset

Generally, most people will be finished by 21:00, others some-time after

Ok to do labs and work at a later time (in the week)

Lecture / Lab / Activities

- Mixture of Lecture + Lab
  - Slides
  - Activities/Exercises
  - Discussions, Share with class group
  - Take home exercises – Contribute/share at next class
  - Group Assessment activities

  - Breaks : 2-3 breaks through the class time

  - All materials are available on module Webpage/VLE

- Additional Tasks & Reading – for those who want to learn/do more
  - Links to additional readings are on module Webpage/VLE
  - Papers, articles, blog post, videos, etc
  - Not compulsory, but recommended to broaden your knowledge

Lecture / Lab / Activities

ANACONDA

python™

jupyter

jupyterLab

- Lectures
  - Slides
  - Activities/Exercises
  - Discussions, Share with class group
  - Demo notebook(s)
- Lab work
  - Rerun the Demo notebook(s)
  - Examine what was done, the outcomes and understand each step
  - Answer questions on certain parts of notebook – you might need to write some code
  - Expand notebooks to perform certain tasks – you will have to write code for this
  - Apply topics, examples, etc to a different dataset – apply your learning and skills
    - This might/will require some additional research/learning to complete
  - Take home exercises, complete lab exercises, etc
- All materials are available on module Webpage/VLE

Assessments

- Module = 100% Continuous Assessment
  - Assessment A - 40% -- Group (2) Assessment – Can use same-ish groups as Data Wrangling module
  - Assessment B - 60% -- Individual Assessment (multiple components)
  - Some lab time (most weeks) can be used to work on Assessments
  - Some independent learning is needed to complete Assessments.

- Plagiarism = 0%
  - See Student Handbook for more details and links to TU Dublin Plagiarism policy.
    - This also covered using tools like ChatGPT and similar LLMs
  - Fail, Repeat Module, May not be allowed to continue studies, etc. -> Lots of undesirable outcomes
  - It's Simple to avoid – Just don't do it!

- Assessments can be used to showcase your work to potential employers
  - eg. Post them on Github, Website, Blog, etc

10 Credit Module

- More topics covered, in more depth.
- More Complex topics.
- Some topics can be challenging to understand.

Contact Hours	4 hours per week over 11-12 weeks Will vary from week to week Mixture of Lectures, Lab Exercises, Support, Assessment Support, Q&As, MCQs, etc.
Self-Directed Learning Hours	148 hours - Approx. 11.4 hours per week Independent Learning, reviewing materials, redoing lab exercises, exploring materials & exercises in more detail, research, reading, working on assessments, etc.

from Module Descriptor

- You might / will need to do some additional reading.
  - Lots of links on the webpage for each week
  - Some weeks have links to additional Videos & other Resources
- Assessments test your understanding
  - Do you understand what you are doing and why
  - They are not about copying/pasting code
  - Can you explain what is happening, what the results mean, how by changing certain parameters you get a different outcome and why
  - Can you Apply to another problem set, Apply appropriate algorithms/code, Analyse, Understand, Improve and Explain

Website – Module Materials

- Everything is there.
- Allows you to work at your own pace or work ahead or to catchup
- Notes, Labs, Videos, Notebooks,
  - Additional readings
  - Videos and Resources
- Pre-Recorded Videos
  - Mainly of Lab Demonstrations and Notebooks
  - Lectures are recorded on BrightSpace/Bongo
  - There might be some differences between the content of the pre-recorded videos and what we cover in class - differences will generally be minor
- Assessments are at fixed dates - Cannot be changed i.e. no extensions
  - Coordinated with your other module(s)

FAQ Webpage

FAQ for TU257 Data Analytics

: My chart This page will contain Questions I've received about the module, the topics, labs, and assignments.

It can be challenging to ask questions, in class or after class etc. This FAQ attempts to address these many challenges and allows for the sharing of knowledge.

Students can contact me with their questions. I will attempt to respond promptly with answers. To assist with knowledge sharing with the whole class, I will post the questions I receive on this page, along with the answers. [The names of the student asking the question will not be listed]

IMPORTANT: When I say, I will attempt to respond promptly, it means I will endeavor to respond within a day or two of getting the question. If I don't respond as quickly as you would like, just remember I have other classes, task and roles to perform each day. If I haven't responded with three days, then Yes Please get onto me and Gently remind me :-)

IMPORTANT: There will be a delay to any questions asked during weekends, holidays and vacation periods. Questions will be answered after such periods.

If you see a week with No questions, that means no one asked me a question.

IMPORTANT: It is important that students check this page regularly for new Q&A. There will be no notices posted to the class group when new Q&A are added.

Week 0 - Course Admin

Q: Is there an exam?

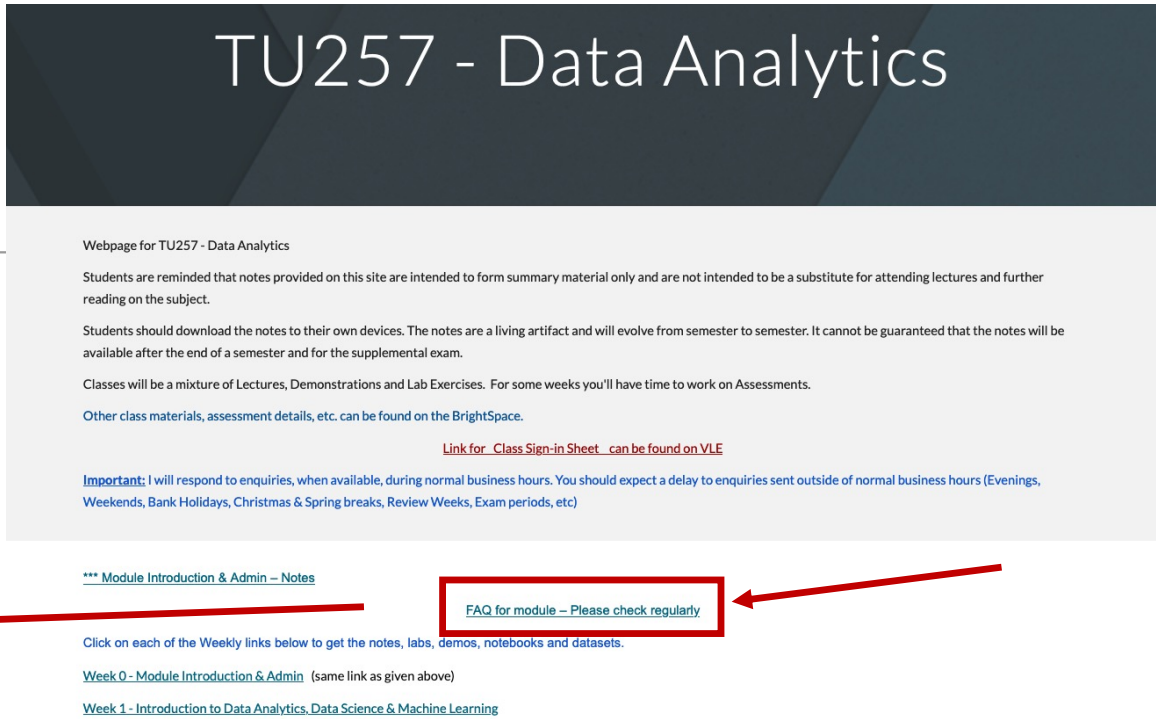
A: Nope, no exam. There are two assessments. The combination of marks from these makes up your final mark for the module. See the Module Introduction & Admin for the breakdown of these marks.

Q: How quickly will we get feedback on the assessment?

A: Typically within 2-3 working weeks. Depending on dates/timing, etc it might be a little longer. Feedback consists of a short paragraph on your assessment highlighting good things and things that needed some additional work

Week 1 - Introduction

Q: Will we be coding or doing lab work during the Week 1



- Check the FAQ first. If your question or answer is not there then email me
- FAQ will be updated with any New Questions and Answers.
- Covers weekly topics and Assessments.
- Lots of Q&As there already

Notices & Questions

- Anything Important, or Changes to something, etc
  - BrightSpace Email

  and/or

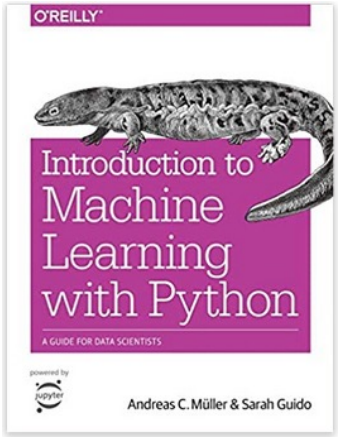
  - BrighSpace Module Notice/Announcements

- You have a Question
  - "It Depends!"
  - Ask during class - Use the Chatbox, or Turn on Microphone to ask
    - I'll ask "*Any Questions*" before/after breaks
    - Sometimes the answer to your question will be answered in a couple of slides!
  - Talk to me after class
  - Email me. Subject = TU257 <your subject title>
  - If needed, we can schedule a Teams call

Knowledge Sharing

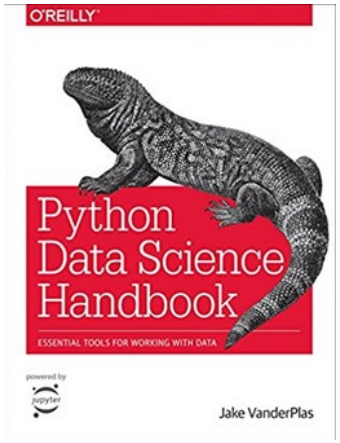
- Have you found something useful you could share with the Class?
  - Websites
  - Blogs
  - Tutorials
  - Articles
  - Books
  - Anything else
- Send me the details and I'll post it on Website
- Crowd sourcing materials/resources/etc

Books



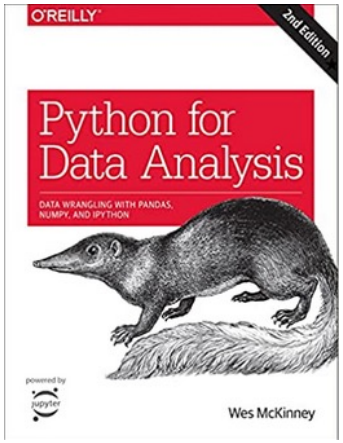
Introduction to Machine Learning with Python: A Guide for Data Scientists

https://amzn.to/3sxk3tT



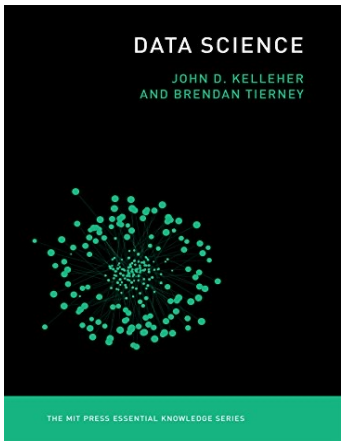
Python Data Science Handbook: Tools and Techniques for Developers: Essential Tools for Working with Data

https://amzn.to/3L7ilFY



Python for Data Analysis

https://amzn.to/3sxCoHg



Data Science : MIT Press essentials series
John Kelliher & Brendan Tierney

International Best Seller !

Available in 9 languages !

https://amzn.to/3FDFCy4

See links to these & Other Resources on module Webpage/VLE

Any Questions ?

What Now/Next ?