

Data Science with Python

COM6018

Varvara Papazoglou

About me

- Varvara Papazoglou (email address is v.papazoglou@sheffield.ac.uk)
- University Teaching Associate
- Research interests
 - Natural Language Processing and Music Information Retrieval
 - Song lyrics, music playlists, metadata, user studies
 - Music recommendation, topic classification/clustering, human evaluation
 - Artificial Intelligence in Higher Education
 - AI literacy, educational software
- You can find out more [on my webpage](#)

Data Science with Python

- Why is this module called Data Science with Python
- Data Science is a broad subject encompassing areas including statistics, machine learning, data visualisation, data mining, and data engineering (i.e., your entire degree!)
- The module focuses on the with Python part of the title. We will be showing you how to use Python to do data science.

COM6018 Course aims

The module has the following aims:

- to provide an introduction to the Python programming language
- to introduce a range of data analysis techniques using the latest Python tools and an understanding of how to apply the appropriate techniques for data analysis
- to understand how to present data science clearly, rigorously and reproducibly

Note, the module focuses on practical data analysis and will be less theoretical than the `Machine Learning and Adaptive Intelligence` module.

Schedule

- Lectures on Monday 13:00–15:00
 - Diamond Lecture Theatre 7
- Labs on Tuesday 11:00–13:00
 - Diamond Computer Room 3
 - Bring Your Own Device

Teaching Weeks 1–5 and 7–11; Weeks 6 and 12 are Readings Weeks (for *this* module!)

Assessment

Two pieces of coursework:

- Assessment 1 (40%)
 - Similar in style to the labs
 - Released start of Week 5
 - Due end of Week 6
- Assessment 2 (60%)
 - A larger piece of work with a written report
 - Released start of Week 9
 - Due Wednesday of Week 12

Feedback

- Feedback during lab classes
- Solutions will be provided for all lab classes
- Feedback will be provided for both assignments

Feedback from *you* to us

- In person
- Online (e.g., email, Blackboard Discussions)
- Through the TellUs survey

Resources

Course content will include

- slides (like these) in PDF format for each lecture,
- tutorials,
- lab classes,
- lab class solutions in jupyter notebook format.

Recommended reading

- [Python Data Science Handbook](#), Jake VanderPlas, O'Reilly, 2016 (available in the library, electronic holdings)
- [Practical data science for information professionals](#), David Stuart, Facet Publishing, 2020 (available in the library, electronic holdings)
- [Python for Data Analysis](#), Wes McKinney, O'Reilly, 2017
- [Practical data science with Python](#), Nathan George, Packt Publishing, 2021

Content Delivery

- Blackboard site
 - Discussion boards, links to course content
 - Links to the reading list
- GitHub repository
 - <https://github.com/UOS-COM6018-2627/COM6018>
 - Allowing you to download the course content
 - Updated throughout the course
- COM6018 Jupyter book
 - <https://uos-com6018-2627.github.io/COM6018/>
 - Allows you to view the Jupyter notebook content online

Course structure

Part I

- Week 1: Getting Started with Python
- Week 2: Reading and Writing Data
- Week 3: Numerical Computing with NumPy
- Week 4: Processing Structured Data with Pandas
- Week 5: Exploratory Data Analysis and Visualization

READING WEEK – Assignment 1

Course structure (continued)

Part II

- Week 7: Introducing Scikit-Learn
- Week 8: Classification with Scikit-Learn
- Week 9: Classifier evaluation and statistical testing
- Week 10: Regression with Scikit-Learn
- Week 11: Unsupervised learning with Scikit-Learn

READING WEEK - Assignment 2