SCHOOL OF COMPUTING

IT8701 INTRODUCTION TO PROGRAMMING FOR DATA SCIENCE

**MODULE OVERVIEW**

# Introduction

Introduction to Programming for Data Science is a first year module for the Specialist Diploma in Data Science course.

# Module Aims

This module provides students with the fundamental skills to code applications to retrieve, manipulate, process and visualize data using the Python programming language. Students learn key concepts such as what structured and unstructured data are, and how they can create and manipulate relational and NoSQL databases to explore data and to create visualizations that can help them gain useful insights from it.

# Module Contents

The topics within the module and the project studies hours are listed as below:

|  |  |  |
| --- | --- | --- |
| Topic | **Title** | Hours |
|  | Python Basics | 12 |
|  | Data Manipulation using the *Numpy* package | 12 |
|  | Data Visualization using the *Matplotlib* package | 12 |
|  | Data Manipulation and Analysis using the *Pandas* package | 12 |
|  | Techniques to process structured and unstructured data | 12 |
|  | **Total** | 60 |

# Module Map & Teaching Plan

|  |  |  |  |
| --- | --- | --- | --- |
| **Date Start**  **(of the week)** | **Week** | **Activities for the Week** | **Things to submit this week** |
| 17-Oct-22 (Mon) | 1 | Module Overview LECTURE - Topic 1 LAB 1 - Python Basics |  |
| 24-Oct-22 (Mon) | 2 | LECTURE - Topic 1 LAB 1 - Python Basics |  |
| 31-Oct-22 (Mon) | 3 | LECTURE - Topic 1 LAB 1 - Python Basics | **Lab 1 (13 Nov Sun)** |
| 07-Nov-22 (Mon) | 4 | LECTURE - Topic 2 Lab 2 - Numpy |  |
| 14-Nov-22 (Mon) | 5 | LECTURE - Topic 2 Lab 2 - Numpy | **Lab 2 (27 Nov Sun)** |
| 21-Nov-22 (Mon) | 6 | LECTURE - Topic 3 Lab 3 - Matplotlib |  |
| 28-Nov-22 (Mon) | 7 | CA1 Brief LECTURE - Topic 3 Lab 3 - Matplotlib |  |
| 05-Dec-22 (Mon) | 8 | CA1 Consultation +  **Quiz 1** |  |
| 12-Dec-22 (Mon) | 9 | **Vacation** |  |
| 19-Dec-22 (Mon) | 10 | **Vacation** |  |
| 26-Dec-22 (Mon) | 11 | **Vacation** |  |
| 02-Jan-23 (Mon) | 12 | **CA1 Interviews** | **CA1 (02 Jan Mon)** |
| 09-Jan-23 (Mon) | 13 | CA2 Brief LECTURE - Topic 4 Lab 4 - Pandas |  |
| 16-Jan-23 (Mon) | 14 | LECTURE - Topic 4 Lab 4 - Pandas | **Lab 4 (29 Jan Sun)** |
| 23-Jan-23 (Mon) | 15 | LECTURE - Topic 5 Lab 5 SQL |  |
| 30-Jan-23 (Mon) | 16 | LECTURE - Topic 5 Lab 6 NoSQL |  |
| 06-Feb-23 (Mon) | 17 | CA2 Consultation +  **Quiz 2** |  |
| 13-Feb-23 (Mon) | 18 | **CA2 Interviews** | **CA2 (13 Feb Mon)** |

# Assessment

The assessment consists of three individual assignments. The weightage and format are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| 1. | CA1 | Assignment 1 | 40% |
| 2. | CA2 | Assignment 2 | 40% |
| 3. | CA3 | Assignment 3 | 20% |
| TOTAL | | | **100%** |

Notes:

|  |  |
| --- | --- |
| **CA1** | This is an individual assignment which requires the student to code a Python application that retrieves and combines data from multiple text files and perform basic data manipulation operations such as cleansing, transformation and visualization on the data. |
| **CA2** | This is an individual assignment which requires the student to code a Python application that retrieves and combines data from multiple data sources, including relational databases and NoSQL datastores and perform data cleansing, transformation, visualization and analysis on it. |
| **CA3** | This is an individual assessment component which consists of, but not limited to; quizzes, lab exercises, participation in class activities, learning attitude and attendance. |

# Resource Materials

|  |  |
| --- | --- |
| 1. | William Wesley McKinney (2017), *Python for Data Analysis*, 2nd Edition, O’Reilly |
| 2. | Martin Czygan, Phuong Vo.T.H (2015), *Getting Started with Python Data Analysis*, Packt |
| 3. | Luca Massaron, John Mueller (2015), *Python for Data Science For Dummies*, Wiley |
| 4. | Ivan Idris (2015), *NumPy: Beginner’s Guide – Third Edition*, Packt Publishing |