PROGRAMMING FOR DATA MANAGEMENT & ANALYSIS II

Bachelor in Data and Business Analytics BDBA SEP-2023
PDMA-DBA.2.M.A
Area Data Science
Number of sessions: 30
Academic year: 23-24
Degree course: SECOND
Number of credits: 6.0
Semester: 1º
Category: COMPULSORY
Language: English

Professor: **ROBERT DAVID POLDING**

E-mail: rpolding@faculty.ie.edu

ROBERT DAVID POLDING

Dr Polding holds a PhD and MSc in Information Systems from The University of Sheffield and a BSc (Hons) in Media Science from Sheffield Hallam University. He has taken part in research projects involving augmented reality, e-commerce, web applications, RFID and database technologies. He has lectured on database design, information systems modeling, project management, programming, operating systems, international business administration and big data.

robert.polding@ie.edu

SUBJECT DESCRIPTION

??Data Science is driven by code; whether it's analyzing business data or big data, coding is the way to implement algorithms of the 21st century.

?Computers are one of the most configurable machines we humans have invented since the dawn of time, yet most of their users are constrained to uses designed by others (third-party software). In this course, you will learn how to code computer programs that will allow you to expand your computer's functionality up to its full potential.
?Knowing how to code is not something reserved for computer scientists; everyone can write programs and if you combine your acquired mathematical and data science knowledge with the ability to write great software, you could unleash a new set of opportunities for your career and boost your performance as an entrepreneur. If you want to know how to code software, this course is for you.

?This is a course for python beginners. The course is designed for students learning how to code the language for the first time.?

LEARNING OBJECTIVES

The main objectives for students are the following:

- Develop logical thinking by developing programs
- Be able to solve real problems through the use of programming languages
- Get strong capabilities in programming with Python

Learning Objectives

In this course you will learn programming terminology and will obtain a solid grasp of the basic mechanics of programming. This includes:

- Introduction to problem solving for programming (i.e., "how to think about solving the problem" including techniques such as pseudo-code or flowcharts)
- Understand object oriented programming and its importance in writing business software applications
- You will be able to write fully functional console or GUI python programs
- Basics of data science tasks in Python

- Develop logical thinking by developing programs.
- Be able to solve real problems through the use of programming languages.
- Get strong capabilities in programming with Python.
- Create an environmental science data processing application.

TEACHING METHODOLOGY

Learning Activity	Weighting	Estimated time a student should dedicate to prepare for and participate in
Lectures	30.0 %	45.0 hours
Discussions	10.0 %	15.0 hours
Exercises in class, Asynchronous sessions, Field Work	30.0 %	45.0 hours
Group work	20.0 %	30.0 hours
Individual studying	10.0 %	15.0 hours
TOTAL	100.0 %	150.0 hours

PROGRAM

IMPORTANT

The following program is tentative. Although we will attempt to cover all the listed topics, the pace of the class depends on group performance. All group projects/presentations must be submitted via Turnitin on Campus Online. No work will be accepted if submitted otherwise.

SESSION 1 (LIVE IN-PERSON)

Course introduction and overview of the computer programming environment

In this session we will review the course logistics and organization together with the computer programming environment (IDEs, compilers, etc…)

Format: Synchronous

Book Chapters: Python Programming: An introduction to Computer Science (See Bibliography)

SESSION 2 (LIVE IN-PERSON)

Data types & variables

In this session we will learn how to store information in our programs using variables and what are the different data-types for variables in Python. We will program our first "Hello World" program in the Python language.

"Hello World"

p. 9-20 - chaos.py

Format: Synchronous

Book Chapters: Python Programming: An introduction to Computer Science (See Bibliography)

SESSION 3 (LIVE IN-PERSON)

Writing Programs

In this session we will learn about types of programs, stages in programming and expressions. We will examine the differences between programming in R and Python

Format: Synchronous

Book Chapters: Python Programming: An introduction to Computer Science (See Bibliography)

SESSION 4 (LIVE IN-PERSON)

Definite loops and the Math library

In this session we will learn about using the for loop and the math library

p. 57-68 - numbers and math

p. 286-288 - random numbers

Format: Synchronous

Book Chapters: Python Programming: An introduction to Computer Science (See Bibliography)

SESSION 5 (LIVE IN-PERSON)

Decision Structures (Part 1)

In this session we will learn about simple and two-way decisions and error handling

p. 209-235 - decisions and error handling

Format: Synchronous

Book Chapters: Python Programming: An introduction to Computer Science (See Bibliography)

SESSION 6 (LIVE IN-PERSON)

Decision Structures (Part 2)

In this session we will look at different strategies with decision structures

p. 27-50 - writing programs

Book Chapters: Python Programming: An introduction to Computer Science (See Bibliography)

SESSION 7 (LIVE IN-PERSON)

Indefinite loops

In this session we will introduce indefinite loops

p. 243-256 - loop structures

Format: Synchronous

Book Chapters: Python Programming: An introduction to Computer Science (See Bibliography)

SESSION 8 (LIVE IN-PERSON)

Boolean Algebra

In this session we will introduce boolean algebra and the underlying concepts

p. 243-256 - Booleans

Book Chapters: Python Programming: An introduction to Computer Science (See Bibliography)

SESSION 9 (LIVE IN-PERSON)

Functions, arguments and return types

In this session we will start encapsulating our code in methods that enable us to reuse the code preventing duplication and errors in large projects

p. 175-203 - functions

Format: Synchronous

Book Chapters: Python Programming: An introduction to Computer Science (See Bibliography)

SESSION 10 (LIVE IN-PERSON)

Lists

In this session we will learn how to store multiple items in lists in Python

129-164 - strings and lists

Format: Synchronous

Book Chapters: Python Programming: An introduction to Computer Science (See Bibliography)

SESSION 11 (LIVE IN-PERSON)

Reading and Writing Files

In this session we will learn how to use files to store information on the long term memory

p. 158-168 - files

Book Chapters: Python Programming: An introduction to Computer Science (See Bibliography)

SESSION 12 (LIVE IN-PERSON)

GUI programming

In this session we will use the graphic library form Python to develop our first simple GUI based program.

p. 83-112 - GUIs

Format: Synchronous

Book Chapters: Python Programming: An introduction to Computer Science (See Bibliography)

SESSION 13 (LIVE IN-PERSON)

GUI programming

In this session we will use the graphic library form Python to develop our first simple GUI based program.

p. 83-112 – GUIs

Format: Synchronous

Book Chapters: Python Programming: An introduction to Computer Science (See Bibliography)

SESSION 14 (LIVE IN-PERSON)

GUI programming

In this session we will use the graphic library form Python to develop our first simple GUI based program.

p. 83-112 – GUIs

Book Chapters: Python Programming: An introduction to Computer Science (See Bibliography)

SESSION 15 (LIVE IN-PERSON)

Midterm exam

Book Chapters: Python Programming: An introduction to Computer Science (See Bibliography)

SESSION 16 (LIVE IN-PERSON)

Introduction to Pandas

In these sessions we will learn about connecting to data sources, multidimensional lists, dictionaries and sets.

Recommended Reading

Python for Data Science for Dummies Chapter 7 (Conditioning Your Data)

Format: Synchronous

Book Chapters: Python for Data Science For Dummies (See Bibliography)

SESSION 17 (LIVE IN-PERSON)

Pandas

In these sessions we will learn the basics of the Pandas library as this library is the core of the data science stack in Python

Recommended Reading

Python for Data Science for Dummies Chapter 7 (Conditioning Your Data)

Book Chapters: Python for Data Science For Dummies (See Bibliography)

SESSION 18 (LIVE IN-PERSON)

Matplotlib

In this session we will learn how to develop data visualizations

Recommended Reading

Python for Data Science for Dummies, Chapter 10

Format: Synchronous

Book Chapters: Python for Data Science For Dummies (See Bibliography)

SESSION 19 (LIVE IN-PERSON)

Matplotlib Part 2

In this session we will learn how to develop data visualizations

Recommended Reading

Python for Data Science for Dummies, Chapter 10

Book Chapters: Python for Data Science For Dummies (See Bibliography)

SESSION 20 (LIVE IN-PERSON)

SQL

An introduction to SQL

Format: Synchronous

Other / Complementary Documentation: SQL Tutorial

SESSION 21 (LIVE IN-PERSON)

SQL in Python

More advanced SQL, and how to connect and use a database in Python

Format: Synchronous

Other / Complementary Documentation: Python MySQL

SESSION 22 (LIVE IN-PERSON)

Advanced SQL

Advanced use of SQL, including joins and table management

Other / Complementary Documentation: Python MySQL

SESSION 23 (LIVE IN-PERSON)

PySpark

Introduction to PySpark, including setting up accounts/computers and how to use Python to interact with Spark

Recommended Reading

Frank Kane's Taming Big Data with Apache Spark and Python – Chapter 1

Format: Synchronous

Book Chapters: Frank Kane's Taming Big Data with Apache Spark and Python (See Bibliography)

SESSION 24 (LIVE IN-PERSON)

PySpark – The Resilient Distributed Dataset, mapping and lambda functions

Recommended Reading

The fundamentals of using RDDs and mapping

Frank Kane's Taming Big Data with Apache Spark and Python – Chapter 2

Format: Synchronous

Book Chapters: Frank Kane's Taming Big Data with Apache Spark and Python (See Bibliography)

SESSION 25 (LIVE IN-PERSON)

PySpark – RDD Actions and key/value datastores

In these sessions we will be setting programming challenges and practicing what we have learnt in the course.

Recommended Reading

Frank Kane's Taming Big Data with Apache Spark and Python – Chapter 2

Book Chapters: Frank Kane's Taming Big Data with Apache Spark and Python (See Bibliography)

SESSION 26 (LIVE IN-PERSON)

Practical – Filtering

An example that shows how to count words and filter

Recommended Reading

Frank Kane's Taming Big Data with Apache Spark and Python – Chapter 2

Format: Synchronous

Book Chapters: Frank Kane's Taming Big Data with Apache Spark and Python (See Bibliography)

SESSION 27 (LIVE IN-PERSON)

Practical – Sorting

We improve the previous session's example and sort it to give better results

Recommended Reading

Frank Kane's Taming Big Data with Apache Spark and Python – Chapter 2

Format: Synchronous

Book Chapters: Frank Kane's Taming Big Data with Apache Spark and Python (See Bibliography)

SESSION 28 (LIVE IN-PERSON)

Practical – Popular Superheroes

Develop an application that analyses a superhero social network

Recommended Reading

Frank Kane's Taming Big Data with Apache Spark and Python – Chapter 3

Format: Synchronous

Book Chapters: Frank Kane's Taming Big Data with Apache Spark and Python (See Bibliography)

SESSION 29 (LIVE IN-PERSON)

Demonstrations of software developed in class

Short practical demonstrations from each group of the software they created

SESSION 30 (LIVE IN-PERSON)

 Final exam

EVALUATION CRITERIA

Throughout this course, you will be asked to read material related to the sessions, participate in discussions, complete individual assignments, participate in in-class quizzes, implement a group project (creating a Python program which provides a solution to a business challenge) and present the outcome of this project in class. Specifically, grading will be based on the following criteria:

criteria	percentage	Learning Objectives	Comments
Final Exam	35 %		
Class Participation	15 %		
Midterm	20 %		
Group Assingment	30 %		

RE-SIT / RE-TAKE POLICY

1. Class participation – discussion

   You are expected to attend every class and participate in the discussions and class activities (games, etc.). The basic criteria in grading your participation are: a) your presence in each session, b) your (quality) contributions to the group discussion. Lively discussions in the classroom are always encouraged, however, make sure that you provide constructive comments which contribute to the learning experience of the whole class.

2. Midterm Exam

   There will be a midterm test to cover basic principles. This will be group based working to solve a programming problem, and individual students will have to submit their contribution.

3. Group project implementation

   A major part of this course's learning experience consists of a group project which will be focusing on designing and implementing an algorithmic solution to a business problem. The definition of the problem as well as the main requirements will be provided to you during the sessions, so that you have only to focus on designing and implementing the best solution. Remember, creativity is always rewarded!

4. Final exam

   At the end of the course you will have to pass an individual exam. This will be 30% theoretical and 70% practical. The minimum pass grade is 3.5.

Minimum passing grade

To ensure quality, we will set a minimum passing grade in all final exams (3.5). If your score is lower than this minimum you will have to go to June retake, irrespective of your overall course grade. Also keep in mind that the overall passing course grade is

5.0.

The dates of the final and retake exams are set at the beginning of the course and cannot be changed under any circumstance.

Late Assignments/Presentation:

Will be penalized 2% per 24-hour period, starting on the day they are due. Only in cases of emergency or illness can changes be made to due dates of assignments or projects. ALL such arrangements are the full responsibility of the student and must be made PRIOR to the due date. Failure to confirm any changes to the due date with the professor prior to the due date will result in a grade of zero.

As per University Policy:

Each student has 4 chances to pass any given course distributed in two consecutive academic years (regular period and July period).

It is mandatory to attend 100% of the classes. Students who do not comply with at least 70% attendance will lose their 1st and 2nd chance, and go directly to the 3rd one (they will need to enroll again in this course the next academic year).

Grading for retakes will be subject to the following rules:

- Those students who failed the subject in the first regular period will have to do a retake in July(except those not complying with attendance rules who are banned from this possibility).
- Dates and location of the July retakes will be posted in advance and will not be changed. Please take this into consideration when planning your summer.
- The maximum grade that a student may obtain in the 2nd exam session is 8 out of 10. Those students in the 3rd call will be required to attend 50% of the classes. If due to schedule overlap, a different option will be discussed with the professor in order to pass the subject.

BIBLIOGRAPHY

Compulsory

- John Paul Mueller. (2019). *Python for Data Science For Dummies.* 2nd edition. John Wiley & Sons, Inc.. ISBN 9781119547624 (Printed)
- Frank Kane. (2017). *Frank Kane's Taming Big Data with Apache Spark and Python.* Pakt. ISBN 1787287947 (Printed)
- John Zelle. (2016). *Python Programming: An introduction to Computer Science.* 3rd. Franklin, Beedle & Associates inc. ISBN 9781590282755 (Printed)

BEHAVIOR RULES

Please, check the University's Code of Conduct here. The Program Director may provide further indications.

ATTENDANCE POLICY

Please, check the University's Attendance Policy here. The Program Director may provide further indications.

ETHICAL POLICY

Please, check the University's Ethics Code [here](). The Program Director may provide further indications.