|  |  |
| --- | --- |
| **Course Code and Title** | |
| **MLI26C744 Tools for Data Analysis** | **6 cr** |
| **Learning Outcomes and Content** | |
| Learning outcomes for this course, upon successful completion, include the ability to: 1) understand principles of programming using the Python programming language, 2) use Python to collect data from various sources for analysis, 3) employ Python for data cleaning, 4) implement statistical and predictive models in Python using business data, 5) understand how to choose the correct statistical or predictive model based on the available data and business context, and 6) understand how the information resulting from data analysis leads to improved business decision-making.  Content:  This course is intended to introduce the student to programming languages as tools for conducting data analysis, focusing on Python in particular. The course will cover basic principles of programming languages, as well as libraries useful in collecting, cleaning and analyzing data in order to answer research questions. Students will learn to use Python to apply forecasting tools and predictive models to business settings. The course will be divided between lecture and lab time, and labs will be focused on teaching students how to implement the programming techniques and statistical models discussed in lectures. | |

|  |
| --- |
| **InstructorName and Profile** |
| Dustin White  Associate Professor of Economics, University of Nebraska at Omaha  I study sports, labor, and health economics, and teach lots of data-driven coursework. While I grew up in the Seattle, WA area, I also lived for two years in southern Brazil, and speak Spanish and Portuguese (though I am a bit out of practice!). I love learning languages (I am currently learning Finnish), doing data science, and I love to play Pokémon! |

|  |
| --- |
| **Email Address** |
| drwhite@unomaha.edu |

|  |
| --- |
| **Office Hours** |
| Office hours/lab time will take place from 1300 to 1400 on Monday-Thursday. All portions of the course will occur in person. During lab, you will be able to break out into small groups to work together on the material for a given class period. I will spend lab time answering questions and checking in with you and other groups as you work through exercises related to the course material.  **Class will be held from 0900 to 1200 in Classroom 1** |

|  |
| --- |
| **Required Reading** |
| All required reading will be provided within the course and will be part of the course notes. If you would like more material to supplement what I provide, please let me know and I can give you book lists to suit your interests. |

|  |  |  |
| --- | --- | --- |
| **Course Schedule** | | |
| **Deduction due to an absence on the first day of the course:** 5 points (on a 100-point scale) will be deducted from the student’s final raw score before converting it to the final grade. If a student is absent on the first day due to illness, and provides the Manager of Academic Operations with a medical certificate, the 5-point deduction will be waived. The Manager of Academic Operations will then inform the instructor of the waived deduction. | | |
| Session # and Date | Topic/s | Assignment/s |
| Session 1: July 29th | Introduction to using Python. We will cover opening notebooks, and basic functions in Python. | No reading or assignments due |
| Session 2: July 30 | Loops and Conditions. We will focus on creating logical conditions for our programs to meet, as well as looping through code to streamline repeated processes. | Assignment 1 due at 0800 |
| Session 3: July 31 | Functions. Creating functions in a programming language allows us to reuse code in many contexts and to solve new problems. We will explore how to do this in Python so that we better understand the code we will be using moving forward. | Assignment 2 due at 0800 |
| Session 4: August 1 | Data Frames and Pandas. We will practice importing and utilizing data in Python. This is the basis for being able to conduct analysis in Python. | Assignment 3 due at 0800 |
| Session 5: August 2 | Regular Expression and text analysis. Sometimes it is advantageous to be able to process text into quantifiable information. Regex provides us the capability to transform text and quickly extract patterns from raw data. | Assignment 4 due at 0800 |
| Session 6: August 5 | Plotting in Python. We will create visuals using Python to be able to supplement the stories that we tell with data through visual media. | Assignment 5 due at 0800 |
| Session 7: August 6 | Introducing Linear Regression and its implementation in Python. Linear regression provides a jumping-off point for statistical analysis, and gives us a chance to prepare our data for analysis. | Assignment 6 due at 0800 |
| Session 8: August 7 | Classification and Regression Trees. Decision trees will give us a chance to discuss machine learning and why it differs from regression analysis. | Assignment 7 **and project proposal** due at 0800 |
| Session 9: August 8 | Random Forests and ensemble methods. Ensemble methods provide improved accuracy and robustness relative to single machine learning models. We will explore these properties through random forest models. | Assignment 8 due at 0800 |
| Session 10: August 9 | Natural Language Processing (NLP). When regular expression isn’t enough, we can do even more with our text data. Using NLP, we can assess things like the sentiment implied in the text, identify parts of speech to extract keywords, or map out topics and entities described in the text. | Assignment 9 due at 0800 |
| Session 11: August 12 | Intro to Neural Networks. While basic machine learning models work exceptionally well on tabular data, sometimes we need to analyze other types of information. Neural Networks allow us to build models capable of performing tasks like NLP, computer vision, and generative models. This is an advanced topic, so we will focus on the basics and some introductory models. | Assignment 10 due at 0800 |
| Session 12: August 13 | Web scraping allows an analyst to collect data from nearly any resource that can be accessed online. This powerful tool allows for the examination of complex problems and the creative collection of resources to address many different needs. | Assignment 11 due at 0800 |
| Session 13: August 14 | Where possible, the use of Web APIs to streamline data collection is a valuable tool. Data collected by API is typically clean and standardized, unlike the data that is collected through web scraping. | Assignment 12 due at 0800 |
| Session 14: August 15 | GUEST SPEAKER: Joona Astrom (On doing data projects in the real world!) & Project Workday. | Assignment 13 due at 0800 |
| Session 15: August 16 | Project presentations. Each student group will present a brief summary of a research question they have answered during the term, and policy implications from the results that they have uncovered. | **Final Project due at 0800** |

|  |  |  |
| --- | --- | --- |
| **Grading** | | |
| **Course Requirements** | | Weighting (%) or maximum points |
| Homework Assignments (aggregated) | | 66 |
| Group Project | | 20 |
| Group Project Presentation | | 10 |
| Discussion and Participation | | 4 |
| **Total** | | **100** |
| |  |  | | --- | --- | | **Conversion scale** | **Final grade**  **(official scale)** | | 90 - 100 | 5 | | 80 - 89 | 4 | | 70 - 79 | 3 | | 60 - 69 | 2 | | 50 - 59 | 1 | | 0 - 49 | 0 | |  | | | |  |
|  | |  |
|  | |  |
|  |  | |

|  |  |
| --- | --- |
| **ECTS STUDENT WORKLOAD** | |
| This course is a 6 ECTS unit course, following the ECTS (European Credit Transfer System) guidelines of Aalto University School of Business. The number of hours the average student is expected to work in the course is 160 (including in-class and out-of-class work). | |
| **Types of Hours** | **Number of Hours** |
| **Contact hours (on- and off-campus):** | **45** |
| **Out-of-class hours:** | **115** (Sum of fields below) |
| Work with course materials, eg required reading | **40** |
| Exam preparation | **0** |
| Individual research & writing | **50** |
| Team projects (meetings, research, preparation, etc.) | **25** |
| Other | **0** |
| **Total of all student workload (contact and out-of-class) hours:** | **160** |

**ACADEMIC POLICY STATEMENTS**

|  |
| --- |
| **CODES OF CONDUCT** |
| Academic excellence and high achievement levels are only possible in an environment where the highest standards of academic honesty and integrity are maintained. Students are expected to abide by the Aalto University Code of Academic Integrity, other relevant codes and regulations, as well as the canons of ethical conduct within the disciplines of business and management education.  In addition, the BScBA Program has strict exam regulations in force which must be followed in all test-taking situations. |

|  |
| --- |
| **TEXTBOOK POLICY** |
| All required textbooks and other course materials are the responsibility of the student. It is the expectation of faculty that all students will have access to the textbooks and other reading material. If a student is not able to purchase his/her own copy of the textbook or other required reading materials, it is nonetheless the student’s responsibility to find a way to complete the reading for the course. |

|  |
| --- |
| **CLASS ATTENDANCE AND PARTICIPATION** |
| Class attendance and participation are considered integral parts of teaching and learning at the BScBA program in Mikkeli. Therefore, regular class attendance is required of all students and attendance records are kept for each class. Students are also expected to be in class on time.  If the student participates in the final exam/assessment, it will be graded and counted towards the final grade.  The attendance policy of the BScBA program provides that:     1. **A maximum of three absences of any kind** is allowed for a 3-week, 6-credit course. Four or more absences will result in being dropped from the course. 2. Whenever taking an absence, **the student bears the risk of missing class**, and the consequences, which may include a lower participation grade, missing a graded activity, etc. It is up to the course instructor to decide whether or not a graded activity can be completed later. 3. **An absence on the first day of the course** will result in 5 points (on a 100-point scale) being deducted from the student’s final raw score before converting it to the final grade. If a student is absent on the first day due to illness, and provides the Manager of Academic Operations with a medical certificate, the 5-point deduction will be waived. The Manager of Academic Operations will then inform the instructor of the waived deduction. 4. **A student getting to class after the session has started** will not be able to enter the classroom until the first break and will get an absence for the day. 5. It is expected that **students marked present for the day are in class the entire time.** Students leaving class early may be marked absent. 6. **The instructor may include class participation as a component of the grade;** up to 15% of the total points that can be earned toward the final grade. 7. **The instructor may identify up to three days of the course (in addition to the first day) as mandatory,** ie taking an absence on those days would have a direct impact on the course grade.   The instructor for the course will take attendance in classes. The decision to drop a student from a course will be made by the instructor, who will inform Mari Syväoja, Manager of Academic Operations: [mari.syvaoja@aalto.fi](mailto:mari.syvaoja@aalto.fi).  **Addition to the attendance policy of the BScBA Program, Mikkeli Campus:**   * This addition concerns absences in addition to the normal maximum of three that would fall under a category called **Medical and Family Emergency cases**. * Students who want to use this option to complete a course must fulfil these criteria:   + The total absences of the student will exceed the normally allowed three absences due to a major medical problem or family emergency.   + The student will be absent no more than 5 days; exceeding that number of days will result in dropping the course.   + Documentation or a detailed explanation concerning the entire period of the emergency (such as a medical certificate) is provided to the Manager of Academic Operations. * The case-by-case solution will be coordinated by the Manager of Academic Operations, who will deal with the documentation and discuss with the instructor to find a pedagogical solution enabling the student to continue in the course. In case the MAO is on leave, the student should contact the other study office staff. * The solution must not cause a significant increase in the instructor’s workload. The grading elements for the course may be reviewed, and additional assignments may be arranged if feasible. However, a shifting of grading proportions may occur. The course grade might be affected due to the student missing some in-class activities. |
|  |

|  |
| --- |
| **Additional Information** |
| I will provide all content in a format that can be used on Google’s Colaboratory platform, where you can run code at no cost. This will require you to have and use a Google account, but has been found to be useful by previous cohorts. This tool is available at the following link: <https://colab.research.google.com/>  Google Colab can be used with Aalto user account. **Please see the instructions on how a Google account can be created and used with an Aalto University user ID (**[**firstname.lastname@aalto.fi**](mailto:firstname.lastname@aalto.fi)**).**  [Google Drive: registration and closing of an account | Aalto University](https://www.aalto.fi/en/services/google-drive-registration-and-closing-of-an-account) |