**Introduction to Python for Data Analysis**

**CourseID:** 120973B **Spring**-2023 **Total Sessions:** 48 **Lecture Sessions**: 32 **Lab Sessions**: 16

**Credits：**3 **Type of Exam**: 考查

**Prerequisites:** Computer Fundamentals.

**Teaching Objectives:**

It’s a beginner’s level Python programming course for data analysis. The teaching content for this course will be, focusing on learning Python programming fundamentals and basic packages used for data analysis. The major topics include, Fundamentals and data structures of Python, Data pre-processing, Python data visualization, Python Exploratory Data Analysis (EDA), Inference statistics, Regression, Multi-variable analysis, Time series and so on, so that students can master the fundamentals of python programming and some data analysis common algorithms and packages, and have the ability to apply them. This course will strengthen the ability of students to design programs which will lays a solid foundation for programming required in subsequent courses such as Bigdata and Machine learning. The specific objectives of the course are as follows:

**Goal 1**: To become familiar with python programming.

**Goal 2**: Master the packages of python for data analysis.

**Goal 3**: Have a certain data analysis ability using python.

**Major Topics Covered in the Course:** This course is a basis for data science and big data technology. The course focuses on the skill training for analyzing data by using Python. Detailed contents include:

* Overview of Python Language
* Basic Elements of Python
* Control Structure
* Python Data Structure
* File and Exception Handling
* Basic Packages
* Data Pre-processing
* Python Exploratory Data Analysis (EDA)
* Python Data Visualization
* Inference Statistics
* Regression
* Multi-variable Analysis
* Time Series

**Tentative Class Schedule:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Week** | **Lecture Topic** | **Credit Hours** | | |
| **Lectures** | **Lab work** | **Total** |
| 1 | **Chapter 1 Overview of the Python Language**  1.1 Course Introduction  1.2 Overview of Python Language | 2 | 1 | 3 |
| 1 | **Chapter 2 Python Language Basics**  2.1 Basic Syntax  2.2 Variables and Assignments  2.3 Data Types  2.4 Type Casting  2.5 Basic Arithmetic Operators  2.6 Expressions | 2 | 1 | 3 |
| 2 | **Chapter 3 Control Structure**  3.6 Control Statements  3.7 Loops |  |  |  |
| 3 | **Chapter 4 Python Data Structures**  4.1 Arrays  4.2 List  4.3 Dictionary  4.4 Tuple  4.5 Set | 2 | 1 | 3 |
| 4 | **Chapter 5 File and Exception Handling**  5.1 Creating a File  5.2 Opening, Closing and Deleting a File  5.3 Modes of Reading and Writing to Files | 2 | 1 | 3 |
| 5 | **Chapter 6 Packages for Data Analysis**  6.1 Numpy  6.2 Pandas  6.3 Matplotlib  6.4 Scikit-learn  6.5 Jupyter-Lab and Jupyter-Notebook  **6.6** [SciPy](https://www.scipy.org/getting-started.html) | 2 | 1 | 3 |
| 6-7 | **Chapter 7 Data Pre-processing**  7.1 Import Libraries and Dataset  7.2 Take Care of Useless and Missing Data  7.3 Encoding/Transformation of Data  7.4 Rescale the Data  7.5 Standardize the Data  7.6 Normalize the Data | 2 | 1 | 3 |
| 8 | **Chapter 8 Exploratory Data Analysis (EDA)**  8.1 Graphical Approach  8.2 Non-graphical Approach | 2 | 1 | 3 |
| 9-10 | **Chapter 9 Data Visualization**  9.1 Matplotlib , Seaborn, Plotly  9.2 Line Charts  9.3 Bar Graphs  9.4 Histograms  9.5 Scatter Plots  9.6 Heat Maps | 2 | 1 | 3 |
| 11-12 | **Chapter 10 Inference Statistics**  10.2 Inferential versus Descriptive Statistics  10.2 Sampling with Python  10.3 Measures of Central Tendency   * Mean, Weighted mean, Geometric mean * Harmonic Mean, Median, Mode   10.4 Measures of Variability   * Variance, Standard deviation * Skewness, Percentiles, Ranges   10.5 Outliers | 2 | 1 | 3 |
| 13 | **Chapter 11 Regression with Scikit-learn**  11.1 Simple Linear Regression  11.2 Multiple Linear Regression  11.3 Polynomial Regression | 2 | 1 | 3 |
| 14 | **Chapter 12 Multi-variable Analysis**  12.1 Overview of Multivariate Analyses  12.2 Reading and Plotting Multivariate Data  12.3 Summary Statistics for Multivariate Data  12.4 Principal Components Analysis (PCA)  12.5 Linear Discriminant Analysis (LDA) | 2 | 1 | 3 |
| 15 | **Chapter 13 Introduction to Time-Series Analysis**  13.1 Introduction to Time-Series Analysis  13.2 Components of a Time-Series  13.3 Stationary and Non-Stationary Time Series |  |  |  |
| 16 | **Final review** | 2 | 1 | 3 |

**Teaching Methods and Means:** This course is characterized by a wide range of knowledge, both theory and practice, and requires a lot of programming practice to master the language. Therefore, the teaching of this course adopts a combination of classroom teaching and a hand on computer programming. The classroom teaching contain easy-to-understand examples to explain the theory to students clearly, and at the same time multimedia facilities such as computers and large-screen projections will be used to demonstrate on the spot, so that students can strengthen their conceptual and logical understanding; According to the teaching content, practical programming sessions and homework assignments will be arranged to further strengthen students' ability to design computer programs.

**Course Instructor:**

**Instructor Name**: Dr. Zakir Ullah

**School Name:** School of Data Science, Capital University of Economics and Business.

**Office Hours:** Tuesday, Wednesday and Thursday 2:00-5:00pm

**Email:** zakirullah98@gmail.com

**Class Time/Place:**

Lectures Time: **To be decided.** Building/Class Room: **To be decided.**

**Textbooks:**

[1] Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, 2nd edition. by Wes McKinney. O’Reilly Media (2017). ISBN: 978-1491957660.

**Grading Schema:** Grades will be based on Quizzes, homework and Attendance according to the following formula:

* Homework assignments = 30%
* Quizzes = 10%
* Semester Project=30%
* Lab=30%

**Grade Cutoffs** : A [90-100], B+ [85-90), B [82-85), C+ [78-82), C [71-78), D [61-71), F [0-60)

**Programming Homework Assignments and Quizzes:**

* There will be regular programming assignments which must be submitted electronically by the announced due date and time. All code must compile, code that does not compile will not be graded. Assignments will be graded based on program performance and documentation. You may not submit any assignment late, late programming work will not be graded.
* All assignments must have the following information listed clearly in documentation, at the beginning of each file: your name, the course ID, your class/section, the programming assignment number, and comments in your program code.
* Along with Homework assignments there will be surprised quizzes during the lectures or lab hours.

**Make-up Quiz Policy:** Do not miss a Quiz. Make-up Quiz will be given only in extenuating circumstances (e.g., doctor's note stating that you were ill and unfit to take the Quiz). Students who miss Quiz for a valid reason may need to take a make-up Quiz; specific arrangements will be made on a case-by-case basis.

**Re-grading Policy:** For re-grading of an assignment or Quiz, please meet with the instructor responsible for the grading. Please arrange a re-evaluation within one week of receiving the graded work. All such requests that are later than one week from the date the graded work is returned to the class will not be entertained.

To promote consistency of grading, questions and concerns about grading should be addressed the course instructor, via email or meet in person during the office hours

**Special Rules about Grading**:

* If all your grades, including homework assignments and quizzes, grades are above the respective class averages, you're guaranteed to receive a grade of C or higher for this class.
* The final grade you receive in this class will reflect, as far as possible, the extent to which you have mastered the concepts and their applications. How much someone needs a grade, or how close they are to the next higher grade, will have no effect on grade.
* As the instructor, I want everyone to do well in this course, and will make every reasonable effort to help you understand the material taught. However, the grades provided at the end of the semester are final, except for rare situations involving grading errors. They will not be altered for any reason, so please do not ask me to do so.

**Communications:** For any private matter or concern about the course students can directly email the instructor via email.

**Email Etiquette:**

* If you have a private matter to discuss, when emailing your instructor about the course, use the following guidelines to ensure a timely response:
* Use a descriptive subject line that includes "Course ID" and a brief note on the topic
* Begin email body with a proper greeting, such as "Hi Sir/Dear Sir" briefly explain your question or concern or request, end with a proper closing that includes your full name, Student ID number.

**Academic Integrity Statement:**

Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Faculty is required to report any suspected instances of academic dishonesty to the school/university Academic Judiciary.