**Syllabus of Python for Data Analysis**

Course ID: 120973B

**Types of courses:**

Compulsory courses in general education

Elective courses in general education

Disciplinary basic courses

Professional core courses

Professional Enhancement Course

Professional Development Course

**Total hours:** 48 (including 16 teaching hours and 32 experiment hours)

**Credits:** 3

**Exam Type:** Exam

**Applicable objects:** Statistics, Data Science and Big Data Technology, Mathematics and Applied Mathematics (Finance).Suitable as a personalized elective for students of other majors

**Prerequisites:** Computer Fundamentals

**Teacher**: Dr.zakir, [zakirullah98@gmail.com](mailto:zakirullah98@gmail.com), wechat Id:13051358386, Office #313

**I. Objectives**

This course is aimed at undergraduate students majoring in statistics, data science and big data, and its main purpose is to teach students to be able to use the python language to process and analyze real-world data. This course uses data in actual work for practical training. The teaching objectives are: to enhance students' practical ability to solve problems, master the basics and advanced programming techniques of python programming, and learn to use python to process and analyze actual data.

**II. Content and its correspondence with graduation requirements**

It focuses on the basic ideas and software implementation of python programming, advanced object-oriented programming, data preprocessing, and common statistical analysis methods. Python object-oriented programming is the difficult content of this course. The course intends to combine the case data in practical work to infer this part of the content to strengthen students' learning and train students' data analysis thinking. The content of the course is mainly based on multimedia courseware teaching. At the same time, python is used on the computer to implement various common methods of statistical analysis, and the exercises for each method are completed on the computer. Classroom exercises and homework are used to assess students' knowledge. It is recommended to use open-book or thesis writing for course assessment. The distribution of grade is calculated as usual attendance \* 10% + classroom Quiz and homework \* 20% + final exam \* 70%.

**III. Schedule (in credits)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Content | in Class | Experiments | Others | Total |
| 1 | Python Environment and Programming Theory | 2 | 4 |  | 6 |
| 2 | Data Structures and Sequences | 2 | 4 |  | 6 |
| 3 | Statements and Control Flow | 2 | 4 |  | 6 |
| 4 | Function | 4 | 8 |  | 12 |
| 5 | Iterators, Generators, Decorators | 2 | 4 |  | 6 |
| 6 | O-O Programming | 2 | 4 |  | 6 |
| 7 | Data Organizing and Analysis | 2 | 4 |  | 6 |
| **Total** |  | **16** | **32** |  | **48** |

**IV. Course Outline**

**Chapter 1 Python Environment and Programming Theory**

1.1 Python system configuration

1.2 Basic knowledge of python

1.2.1 Python help

1.2.2 Python identifiers

1.2.3 Lines and indentation

1.2.4 Python objects

1.2.5 Numbers and expressions

1.2.6 Operators

1.2.7 Strings

1.2.8 Date and time

1.3 Programming Theory

Teaching focus and difficulty: Python environment configuration and object-oriented programming foundation

The assessment requirements of the course: Understanding the python environment, interface, and operation mode; understanding that everything in python is an object, master the basics of python programming, and flexibly apply the basic components of programming.

Review questions:

1. What is a program? What are the characteristics of object-oriented programming?

2. Collect industry demand data and understand the status of python in the field of big data analysis.

3. How to understand "everything is an object" in python?

**Chapter 2 Data Structures and Sequences**

2.1 List

2.2 Tuple

2.3 Dict

2.4 Set

2.5 Comprehension

Teaching focus and difficulty: Python core data structure: sequence

Assessment requirements of the course: understanding the basic characteristics of sequences; understanding the basic characteristics of list, tuple, dictionary and set in python and their generation methods, and mastering the basic use of derivation.

Review questions:

1. Understand how comprehension works.

2. Why does the data structure of sequence play an extremely important role in big data analysis?

**Chapter 3 Statements and Control Flow**

3.1 Conditional Statements

3.2 Loop Statement

Teaching focus and difficulties: Python's conditional statements and loop statements and their control.

Assessment requirements of the course: understanding the working mechanism of loop statements in python, and mastering the basic usage of conditional statements and loops.

Review questions:

1. What are the similarities and differences between a loop statement and a comprehension?

**Chapter 4 Functions**

4.1 Parameters of functions

4.2 Global and local variables

4.3 Anonymous functions

4.4 Recursion and closures

4.5 Currying

4.6 Commonly used built-in functions

Teaching focus and difficulties: the basic characteristics of python functions, the generation, operation and calling methods of commonly used functions

Assessment requirements of the course: understanding the working mechanism of python functions, mastering the basic methods of generating functions, be familiar with various forms of function parameters, and understanding function operations such as recursion, closure, and currying.

Review questions:

1. What is currying and un-currying, and what are the similarities and differences between un-currying and closures?

2. Give an example to illustrate how functions can improve the efficiency of data analysis?

**Chapter 5 Iterators, Generators, Decorators**

5.1 Iterables and Iterators

5.2 Generators

5.3 Decorators

Teaching focus and difficulty: python iterators, generators, decorators

Assessment requirements of the course: understanding the working mechanism of iterators and generators, and mastering the usage of iterators and generators. Mastering the use of decorators.

Review questions:

1. How do iterators and generators improve the efficiency of data analysis?

2. What is the mechanism of action of the decorator?

**Chapter 6 Object-Oriented Programming Techniques**

6.1 I/O

6.2 Modules

6.3 Class

6.3.1 Declaration

6.3.2 Features

6.3.3 Methods

6.3.4 Properties

6.4 Packages

Teaching focus and difficulty: basic techniques of python object-oriented programming

Assessment requirements of the course: understanding python file I/O operations, mastering the basic usage of classes and modules, and understanding the generation and invocation of packages.

Review questions:

1. What is the difference and connection between functions and classes?

2. How do the characteristics of the class use methods to achieve control properties?

**Chapter 7 Data Organizing and Analysis**

7.1 Numpy

7.1.1 Vectors

7.1.2 Arrays

7.1.3 Matrices

7.1.4 File reading and writing

7.2 Pandas

7.2.1 Data structure of pandas (Series and DataFrame)

7.2.2 Data manipulation of pandas

7.3 Matplotlib and Visualization Basics

Teaching focus and difficulties: data structure and data analysis functions of numpy and pandas, basic drawing syntax of matplotlib

Course assessment requirements: mastering numpy ndarray objects and matrix objects, mastering the basic syntax of using numpy for data analysis; mastering pandas Series and DataFrame objects, mastering the basic techniques of data analysis and visualization using pandas; be familiar with the basic drawing skills of matplotlib .

Review questions:

1. Why is it more efficient to use numpy'sndarray to analyze the same data object?

2. In what ways can a DataFrame object be constructed?

**V. Examination methods and performance evaluation**

This course is generally assessed by exam or dissertation, with 70% of the paper and 30% of attendance, classroom presentations or homework.

**VI. Main reference books and other contents**

[1]Ruan Jing, LiuShuai, Python数据分析基础（第3版）The Elements of Data Analysis Using Python (3rd Edition) (14th Five-Year National Statistical Planning Textbook), China Statistics Press, 2022

[2]Ruan Jing, The Elements of Data Analysis Using Python (2nd Edition) (Thirteenth Five-Year National Statistical Planning Textbook), China Statistics Press, 2018

[3]Ruan Jing, Programming Foundation for Data Science, Peking University Press, Capital University of Economics and Business Press, 2019

**Professor:** Ruan Jing

**Department Teaching Director Verification and Signature**: