

The International College of Economics and Finance

Syllabus for “Programming and Data Processing”

Lecturer: Sergey G. Efremov

Class teacher: Sergey G. Efremov

Course description:

In the modern, highly technological world computer skills have become essential for specialists in almost all possible areas. Programming in particular has gone beyond its traditional borders of being just a prerogative of IT specialists, turning into an element of computer literacy. In the last 10 years programming languages and tools have evolved significantly, which now enables people even without a solid technical background to successfully master related skills.

The present course is offered to 1st year ICEF students as an alternative to the entry level “Information Computer Systems” course. Students wishing to enter the course can take a test to verify the required knowledge and skills. In particular, candidates are assumed to have a solid user-level understanding of the Windows (or other desktop) operating system (GUI, file system, running and installing applications, using standard applications: text editor, browser, mail client, etc.) and programming basics (any high-level language). Knowledge of Excel fundamentals is required for the second part of the course.

The course is split into two distinct parts. The first part focuses on programming and data processing techniques using the Python language. The second part covers advanced Excel features that can be useful in later ICEF courses and economics-related applications.

The course is not part of the University of London international programme.

Course Objectives

The aim of the course is twofold: on the one hand, it provides students with knowledge of fundamental programming principles and the corresponding practical skills. Although based on a particular toolset (Python), the course aims to give a general view of what can be done using a modern general-purpose programming language. On completing the first part students are expected to know several techniques of automated data acquisition, including web queries, and basic data processing.

On the other hand, students will learn several advanced Excel skills that are useful in many practical applications in economics.

Methods

The course is practice-oriented and requires active student involvement in its activities. The following methods and forms of study are used in the course:

- Practice sessions (4 hours a week, conducted in a computer class)
- Regular homework assignments. Each assignment takes from 2 to 4 hours to complete including the required readings.

The final assignment will be in the form of a small project (2 weeks)

- Online consultations from course instructor(s)
- Self-study activities: completing homework assignments, studying recommended resources, experimenting with the toolset, solving advanced tasks

In total the course comprises 72 hours of practice sessions and 80 hours of self-study activities.

Main Reading:

The main textbook “Think Python” [1] is a well-structured introductory resource for Python programming. Students can study the book chapter by chapter, as the course closely follows its structure. Special attention should be paid to the glossary at the end of each chapter. “Fundamentals

of Python” my K. Lambert [2] is a more advanced textbook and only selected reading (see course outline) is recommended in the beginning. For Excel-related topics student guides and textbooks [4, 5] written by ICEF staff are the primary recommended resource.

1. Allen B. Downey. Think Python. How to think like a computer scientist. Green Tea Press, 2008. Electronic version available for free download at: <http://greenteapress.com/wp/think-python-2e/>
Interactive version with program execution in the browser available at: <http://interactivepython.org/runestone/static/thinkcspy/index.html>
2. Kenneth A. Lambert. Fundamentals of Python: From First Programs Through Data Structures. – Course Technology, 2010.
3. Walkenbach J. Excel 2010 Bible. - Wiley, 2010.
4. Акиншин А.А., Белоусова С.Н., Бессонова И.А. Руководство для студентов по курсу «Информационные компьютерные системы». Москва: МИЭФ, 2014. 68 с.
5. Акиншин А.А., Белоусова С.Н., Бессонова И.А. Специальные возможности MS Excel для работы с большими массивами информации (2-е издание исправленное и дополненное). Москва: МИЭФ, 2010. 162 с.

Supplementary reading

6. Лутц М. Изучаем Python, 4-е издание. – Пер. с англ. – СПб.: Символ-Плюс, 2011. – 1280 с., ил.
7. W. McKinney. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. – O’Reilly Media, 2012.

Internet resources

8. Python practice book. <http://anandology.com/python-practice-book/index.html>
9. Coursera online course: Interactive Python. <https://www.coursera.org/course/interactivepython1>
10. CodeAcademy. Python track. <https://www.codecademy.com/tracks/python>
11. Python 3 installation guides for different OS: <https://pragprog.com/book/gwpy2/practical-programming>
12. Some of the basic differences between Python 2 and Python 3:
<http://www.cs.carleton.edu/faculty/jgoldfea/cs201/spring11/Python2vs3.pdf>
13. Python regular expressions tutorial.
http://www.tutorialspoint.com/python/python_reg_expressions.htm
14. Python requests library. <http://docs.python-requests.org/en/latest/>
15. API integration in Python – Part 1. <https://realpython.com/blog/python/api-integration-in-python/>
16. NumPy tutorial. http://wiki.scipy.org/Tentative_NumPy_Tutorial
17. OpenPyXL library. <https://openpyxl.readthedocs.org/en/latest/>
18. RESTful web services: <http://www.drdoobbs.com/web-development/restful-web-services-a-tutorial/240169069>
19. Python database access: http://www.tutorialspoint.com/python/python_database_access.htm

Required software

The first part of the course is based on Python 3. The default toolset, which can be downloaded from <https://www.python.org/downloads/>, is open-source and cross-platform.

The second part of the course requires installation of Excel 2010 or higher.

Grade determination

The final grade for the course is calculated as a weighted average of homework assignments, mid-term and final exams. The percentage breakdown is the following:

<i>Component</i>	<i>Weight (%)</i>
Homework assignments (HW)	20
Mid-term exam (MT)	25
Final exam (EX)	55

$$FG = 0,2(HW) + 0,35(MT) + 0,45(EX)$$

Course Outline

Part 1. Programming in Python

1. Introduction to programming and the Python language.

Purpose of programming. Source code and executable files. Programming languages. Python application areas. Python versions 2 and 3. Overview of resources and development tools. Software installation. Interactive shell

[1: P. 1-7], [2: P. 2-29], [12]

2. Python basics

Types and variables. Integer, float and string types. Conversion between types. “type” operator. Arithmetic operators. Console input-output. Formatted output.

[1: P. 9-15], [2: P. 342-375]

3. Program flow

Boolean expressions. Conditional execution. Code formatting. “while” loop. Code editor. Debugging programs.

[1: P. 39-47, 63-69], [2: P. 75-120]

4. Functions and modules

Importing modules. Calling standard functions. Functions and methods. Math and random modules. Defining custom functions. Installing new packages.

[1: 17-26], [2: 63-69, 201-210]

5. Data structures

Mutable and immutable types. Lists: creating a list, adding and removing elements, retrieving elements, sorting lists, slices. Dictionaries: creating a dictionary, adding and removing items, querying items by key. Tuples. Conversions between data structures. “for” loop and “in” operator. The datetime type. Comprehensions.

[1: P. 89-134], [2: P. 159-200]

6. Processing text. Regular expressions

Specialized methods for string manipulation and processing. Regular expression language: main capabilities and samples. Extracting data to groups.

[1:71-88], [2: P. 121-140], [13]

7. File input-output

File system. Absolute and relative paths. File access modes. Formats. Standard file operations.
[1: P. 137-140], [2: P. 141-147]

8. Data acquisition

Human-readable and machine-readable formats. Integration with the Web: basics of HTTP, extracting data from HTML, downloading files, querying RESTful services. Integration with SQL DBMS.
[13, 14, 17, 18]

9. Python for data analysis and visualization

Overview of available packages and their features. Integration with Excel.
[15, 16]

Part 2. Advanced Excel

Topic 1. Functions

Functions in Excel vs Functions in programming. Excel Functions Syntax. Cell references. Names in formulas. Computational and financial Excel functions. Conditional formatting.
[3: P. 213-243, 467-486], [4: P. 30-34]

Topic 2. Graphical Data Visualization and Analysis in MS Excel

Charts, graphs, and their properties. Customizing charts. Smoothing. Graphical data analysis. Sparklines for visual representation of data.
[3: P. 389-422, 487-498]

Topic 3. Processing large series of data.

Excel database. Sorting, searching and editing. Filtering, AutoFilter. Creating custom filters using Excel Advanced Filter. Database functions.
Vertical and horizontal lookup functions. Subtotalling the data. Data Consolidation.
Pivot Tables and Charts. Sorting and filtering subtotals. Calculations in pivot tables: additional calculations, calculated fields and objects. Pivot charts.
[3: P. 311-328, 665-712], [5: P. 45-91, 95-112, 113-154]

Topic 4. MS Excel Add-ins

Microsoft Excel add-ins for statistical tasks (Analysis ToolPak) and optimization (Solver).
Analysis ToolPak for Microsoft Excel: finance, statistics and engineering functions. Solver Add-In.
What-If analysis. Using Solver for solving systems of linear and non-linear equations.
Goal Seek. Solving system of equations.
[3: P. 727-744], [5: P. 5-30]

Topic-wise course plan

No	Topics	In class	Self-study
Part 1. Programming in Python			
1	Introduction to programming, the Python language and IDE	2	4
2	Python basics	2	6
3	Program flow	4	6
4	Functions and modules	6	6
5	Data structures	10	8
6	Processing text. Regular expressions	6	8
7	File input-output	8	4
8	Data acquisition	12	12
9	Python for data analysis and visualization	4	4
Part 2. Advanced Excel			
1	Built-in functions	6	6
2	Graphical analysis	4	4
3	Working with large series of data	4	6
4	Add-ins for economic tasks	4	6
	Total:	72	80

Preliminary course schedule (2018-2019)

Week	Dates	Topic / activity	HW
Module 1			
1	1/09	Course presentation to students	
2	3/09 – 9/09	Introduction to programming. Python basics	HW0 (not graded): install IDE, experiment with Python
3	10/09 – 16/09	Program flow	HW1: conditional execution and loops
4	18/09 – 23/09	Functions and modules	HW2: functions and modules
5	24/09 – 30/09	Data structures: lists, tuples	
6	1/10 – 7/10	Data structures: dictionaries, representing dates, comprehensions	HW3: data structures
7	8/10 – 14/10	Processing text, regular expressions	
8	16/10 – 21/10	Regular expressions, file input-output	HW4: processing text and regular expressions
9	22/10 – 28/10	Mid-term test	
Module 2			
1	29/10 – 4/11	Downloading files, web queries	Project (HW5)
2	6/11 – 11/11	Python and SQL	
3	12/11 – 18/11	Python for data analysis and visualization. <i>Project presentation</i>	
4	19/11 – 25/11	Excel: Functions.	HW6: Excel functions
5	26/11 – 2/12	Excel: Data visualization	HW7: Excel data visualization
6	3/12 – 9/12	Excel: processing large series of data	HW8: Excel: large series of data
7	10/12 – 19/12	Excel: Add-ins	

8	20/12 – 31/12	<i>Final exam</i>
---	---------------	-------------------