

Faculty of Computer Technology and Cybersecurity
Department of Mathematical and Computer Modeling



Approved
Dean of faculty FCTC
ITU JSC
Seilova N.A.
2023

**SYLLABUS
(ACADEMIC PROGRAM)**

Course (code, title): SFT6503 Python for data analytics
(code, title):

Major (code, title): B057 Information Technologies
(code, title):

Educational program 6B06111 Financial Mathematics, 6B06112 – Data science
(code, title)

Year: 3 **Semester:** 5 **Number of credits:** 5 ECTS

Lectures: 15 hours

Practical classes: 30 hours

T/SIS: 105 hours

Total: 150 hours

Cycle: BS

Final assessment form: Examination

Almaty 2023

«ITU» JSC

Academic program of the course (code, title) SFT6503 Python for data analytics has been developed based on Standard Academic Program.

Academic program has been reviewed at the meeting of Mathematical and Computer Modeling department.

Minutes № 12 dated «12» June 2023.

Head of the Department



signature

Ydyrys A. Zh., PhD, Assistant-Professor

full name, title, degree

Author



signature

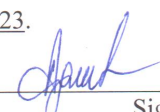
Tokmukhamedova F. K., Msc, senior - lecturer

full name, title, degree

The working academic program was approved at the meeting of the Council for Academic Quality of the Faculty of Computer Technology and Cybersecurity

Minutes № 6 dated "23" June 2023.

Head of the Department



Signature

A. Ajibayeva

for Educational and Methodological Affairs

1. GENERAL INFORMATION	
Faculty	Information Technology
Major code and title	B057 Information technology
Educational program code and title	6B06111 Financial Mathematics, 6B06112 Data Science
Year, semester	4th year, 7th semester
Subject category	Basic
Number of credits (ECTS)	4
Prerequisites	Algorithms and Programming Languages, Linear Algebra, Discrete Mathematics, Introduction to Probability and Statistics
Postrequisites	Data science, Machine learning, Deep Learning and Neural Networks
Lecturer	Tokmukhamedova Fatima Kadyrovna, Msc, senior-lecturer, f.tokmukhamedova@iitu.edu.kz , 807 room
2. GOALS, OBJECTIVES AND LEARNING OUTCOMES OF THE COURSE	
<p>The course goal is</p> <p>Students with a minor in Introduction to Programming in Python.</p> <p>Students will develop relevant programming abilities.</p> <p>Students will demonstrate proficiency with visualization of data.</p> <p>Students will develop the ability to build and assess data-based models. Students will execute statistical analyses with professional statistical software. Students will demonstrate skill in data management.</p> <p>At the end of the course, students will have a solid grasp of Python programming basics and have been exposed to the entire data science workflow, starting from interacting with SQL databases to query and retrieve data, through data wrangling, reshaping, summarizing, analyzing and ultimately reporting their results. The course will introduce and use popular Python libraries such as pandas and NumPy, and all analyses will be performed using Jupyter notebooks.</p>	
<p>The objectives of the course are</p> <ul style="list-style-type: none"> – Basic process of data science – Python and Jupyter notebooks – An applied understanding of how to manipulate and analyze uncured datasets – Basic statistical analysis and machine learning methods – How to effectively visualize results <p>By the end of the course, you should be able to find a dataset, formulate a research question, use the tools and techniques of this course to explore the answer to that question, and share your findings.</p>	
<p>Learning outcomes of the course</p> <ul style="list-style-type: none"> – Query DataFrame structures for cleaning and processing – Explain distributions, sampling, and t-tests – Describe common Python functionality and features used for data science – Understand techniques such as lambdas and manipulating csv files 	

- Create a visualization using matplotlib
- Identify the functions that are best for particular problems
- Understand best practices for creating basic charts
- Describe what makes a good or bad visualization

Students successfully completing the course will be able to:

- How to import data sets, clean and prepare data for analysis, summarize data, and build data pipelines
- Use Pandas DataFrames, Numpy multidimensional arrays, and SciPy libraries to work with various datasets
- Load, manipulate, analyze, and visualize datasets with pandas, an open-source library

Build machine-learning models and make predictions with scikit-learn, another open-source library

3. COURSE DESCRIPTION

This course will introduce you to the field of data science. First, and foremost, you'll learn how to conduct data science by learning how to analyze data. That includes knowing how to import data, explore it, analyze it, learn from it, visualize it, and ultimately generate easily shareable reports. We'll also introduce you to two powerful areas of data analysis: machine learning and natural language processing. To conduct data analysis, you'll learn a collection of powerful, open-source, tools including: ● python ● jupyter notebooks ● pandas ● numpy ● matplotlib ● scikit learn ● nltk ● And many other tools And you won't be learning these tools in isolation. You will learn these tools all within the context of solving compelling data science problems

4. COURSE POLICY

You can expect me:

- To start and end class on time.
- To reply to e-mails within 24 hours on weekdays and 48 hours on weekends.
- To assign homework that adequately covers the material and meets the learning objectives of the course.
- To give exams that accurately reflect the material covered in class and assigned in homework.

I can expect you:

- To come to class on time.
- To be attentive and engaged in class.
- To refrain from using laptops, cell phones and other electronic devices during class.
- To spend an adequate amount of time on the homework each week, making an effort to solve and understand each problem.
- To submit any tasks (quizzes, tests, homework) on time.
- To not cheat. Plagiarized papers shall be graded by "0".
- To engage with both the abstract and computational sides of the material.
- To seek help when appropriate.

5. LITERATURE

Basic literature:

1. Python for Data Analysis: DATA WRANGLING WITH PANDAS, NUMPY, AND IPYTHON / McKinney Wes.- Tokyo, 2018.

Supplementary literature:

2. Nelli, F. (2015). Python Data Analytics : Data Analysis and Science Using Pandas, Matplotlib and the Python Programming Language. [Berkeley, CA]: Apress.

Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&site=eds-live&db=edsebk&AN=1056488>

3. Nelli, F. (2018). Python Data Analytics : With Pandas, NumPy, and Matplotlib (Vol. Second edition). New York, NY: Apress. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&site=eds-live&db=edsebk&AN=1905344>

6. Course schedule

Week/ date	Course topics	References	Lectures (1 h/w)	Practical sessions (2 h/w)	TSIS (1 h/w)	SIS (6 h/w)
1	Introduction to Data Science. Python Language Basics, IPython, and Jupyter Notebooks	Basic [1]-[2]	1	2	1	6
2	Built-in Data Structures, Functions and Files	Basic [1]-[2]	1	2	1	6
3	NumPy Basics: Arrays and Vectorized Computation	Basic [1]-[2]	1	2	1	6
4	Getting Started with pandas	Basic [1]-[2]	1	2	1	6
5	Data Loading, Storage and File Formats	Basic [1]-[2]	1	2	1	6
6	Data Cleaning and Preparation	Basic [1]-[2]	1	2	1	6
7	Data Wrangling: Join, Combine and Reshape	Basic [1]-[2]	1	2	1	6
8	Plotting and Visualization Midterm	Basic [1]-[2]	1	2	1	6
9	Data Aggregation and Group Operations	Basic [1]-[2]	1	2	1	6
10	Time Series	Basic [1]-[2]	1	2	1	6
11	Advanced pandas	Basic [1]-[2]	1	2	1	6
12	Introduction to modeling Libraries in Python	Basic [1]-[2]	1	2	1	6
13	Data Analysis Examples	Basic [1]-[2]	1	2	1	6
14	Final Project Endterm	Basic [1]-[2]	1	2	1	6
15	Overview of the course	Basic [1]-[2]	1	2	1	6
Total hours:		120	15	30	15	90

7. List of topics/ assignments for practical classes

№	Topic Title	Number of hours	References	Form of reporting	Deadline
1	2	3	4	5	6
1	Built-in Data Structures, Functions and Files	3	Basic [1]	Submission to DL	Next lesson
2	NumPy Basics: Arrays and Vectorized Computation	3	Basic [1]	Submission to DL	Next lesson
3	Getting Started with pandas	3	Basic [1]	Submission to DL	Next lesson
4	Data Loading, Storage and File Formats	3	Basic [1]	Submission to DL	Next lesson
5	Data Cleaning and Preparation	3	Basic [1]	Submission to DL	Next lesson
6	Data Wrangling: Join, Combine and Reshape	3	Basic [1]	Submission to DL	Next lesson
7	Plotting and Visualization	3	Basic [1]	Submission to DL	Next lesson
8	Data Aggregation and Group Operations	3	Basic [1]	Submission to DL	Next lesson
9	Time Series	3	Basic [1]	Submission to DL	Next lesson
10	Advanced pandas	3	Basic [1]	Submission to DL	Next lesson
	Total hours:	30			

8. List of topics/assignments for Student Independent Study

Proper organization of students' independent study is the key to the formation of skills in mastering, learning, assimilation and systematization of acquired knowledge, ensuring a high level of academic performance in the learning process

№	Topic/Assignment title	Number of hours	References	Form of reporting	Deadline
1	2	3	4	5	6
1	Built-in Data Structures, Functions and Files	9	Basic [1]	Submission to DL	Next lesson
2	NumPy Basics: Arrays and Vectorized Computation	9	Basic [1]	Submission to DL	Next lesson
3	Getting Started with pandas	9	Basic [1]	Submission to DL	Next lesson
4	Data Loading, Storage and File Formats	9	Basic [1]	Submission to DL	Next lesson
5	Data Cleaning and Preparation	9	Basic [1]	Submission to DL	Next lesson
6	Data Wrangling: Join, Combine and Reshape	9	Basic [1]	Submission to DL	Next lesson
7	Plotting and Visualization	9	Basic [1]	Submission to DL	Next lesson
8	Data Aggregation and Group Operations	9	Basic [1]	Submission to DL	Next lesson

9	Time Series	9	Basic [1]	Submission to DL	Next lesson
10	Advanced pandas	9	Basic [1]	Submission to DL	Next lesson
	Total hours:	90			

9. List of topics/assignments for TSIS

№	Topic/Assignment title	Number of hours	References	Form of reporting	Deadline
1	2	3	4	5	6
1	Introduction to Data Science. Python Language Basics, IPython, and Jupyter Notebooks	1	Basic [2] p.15	Report	Next lesson
2	Built-in Data Structures, Functions and Files	1	Basic [2] p.51	Report	Next lesson
3	NumPy Basics: Arrays and Vectorized Computation	1	Basic [2] p.85	Report	Next lesson
4	Getting Started with pandas	1	Basic [2] p.123	Report	Next lesson
5	Data Loading, Storage and File Formats	1	Basic [2] p.167	Report	Next lesson
6	Data Cleaning and Preparation	1	Basic [2] p.191	Report	Next lesson
7	Data Wrangling: Join, Combine and Reshape	1	Basic [2] p.221	Report	Next lesson
8	Plotting and Visualization	1	Basic [2] p.253	Report	Next lesson
9	Data Aggregation and Group Operations	1	Basic [2] p.287	Report	Next lesson
10	Time Series	1	Basic [2] p.317	Report	Next lesson
11	Advanced pandas	1	Basic [2] p.363	Report	Next lesson
12	Introduction to modeling Libraries in Python	1	Basic [2] p.383	Report	Next lesson
13	Data Analysis Examples	1	Basic [2] p.403	Report	Next lesson
14	Final Project	2	Basic [2]	Report	Next lesson
	Total hours:	15			

10. System for evaluating student performance in a discipline:

Each type of educational work is evaluated on a 100-point scale and is included in the average assessment of the current control, taking into account the weighting coefficient in accordance with the table

Period	Assignments	Maximum score	Weighting coefficient	Total
1 st attestation	Laboratory practice	100	0,5	100
	Midterm	100	0,5	
2 nd attestation	Laboratory practice	100	0,5	100
	Midterm	100	0,5	
Exam				100
Total	0,3*1stAtt+0,3*2ndAtt+0,4*Ex			100

*If the number of absences exceeds 20%, student will be automatically scheduled for a Retake (summer semester)

11. Assessment criteria:

The point-rating letter system for assessing the educational achievements of students with their interpretation in the traditional grading scale:

Letter Grade	Numerical equivalent	Points (%)	Traditional system assessment	General description of grading criteria
A	4,0	95-100	Excellent	The student has knowledge of the subject in the full scope of the curriculum, understands the discipline deeply enough; shows a high level of knowledge that exceeds the volume provided by the syllabus, gives an exhaustive answer
A-	3,67	90-94		The student has knowledge of the subject in the full scope of the curriculum, understands the discipline deeply enough; gives an exhaustive answer
B+	3,33	85-89	Good	The student shows a complete, well-founded knowledge of the subject, but the answers did not always highlight the main idea, rational methods of calculation were not always used; the answers were mostly brief and sometimes unclear.
B	3,0	80-84		
B-	2,67	75-79		
C+	2,33	70-74	Satisfactory	The student demonstrates sufficient knowledge of the subject, but without proper depth and justification, the answers are unclear and without proper logical sequence.
C	2,0	65-69		
C-	1,67	60-64		
D+	1,33	55-59		
D	1,0	50-54	Unsatisfactory	The student demonstrates insufficient knowledge of the subject, positive answers were not given to individual questions.
FX	0,5	25-49		

F	0	0-24		The student demonstrates a very low level of knowledge of the subject.
---	---	------	--	--

12. Assessment and evaluation materials (exam questions)

List of exam lecture topics:

- Python Language Basics
- Built-in Data Structures, Functions and Files
- NumPy Basics: Arrays and Vectorized Computation
- Pandas
- Data Loading, Storage and File Formats
- Data Cleaning and Preparation
- Data Wrangling: Join, Combine and Reshape
- Plotting and Visualization
- Data Aggregation and Group Operations
- Time Series

Exam duration is 90 minutes.

Task 1. [10p] Write a NumPy program to convert a list of numeric value into a one-dimensional NumPy array.

Task 2. [10p] Write a NumPy program to test whether none of the elements of a given array is zero.

Task 3. [20p] Write a NumPy program to compute the multiplication of two given matrixes.

Task 4. [20p] Write a Pandas program to convert a dictionary to a Pandas series.

Sample Series:

Original dictionary: {'a': 100, 'b': 200, 'c': 300, 'd': 400, 'e': 800}

Converted series:

a 100

b 200

c 300

d 400

e 800

dtype: int64

Task 5. [20p] Write a Pandas program to split the following dataframe by school code and get mean, min, and max value of age for each school.

Test Data:

school	class	name	date_Of_Birth	age	height	weight	address
S1	s001	V	Alberto Franco	15/05/2002	12	173	35 street1
S2	s002	V	Gino Mcneill	17/05/2002	12	192	32 street2
S3	s003	VI	Ryan Parkes	16/02/1999	13	186	33 street3
S4	s001	VI	Eesha Hinton	25/09/1998	13	167	30 street1
S5	s002	V	Gino Mcneill	11/05/2002	14	151	31 street2
S6	s004	VI	David Parkes	15/09/1997	12	159	32 street4

Task 6. [20p] Write a Python program to draw a scatter plot comparing two subject marks of Mathematics and Science. Use marks of 10 students.

Sample data:

Test Data:

math_marks = [88, 92, 80, 89, 100, 80, 60, 100, 80, 34]

science_marks = [35, 79, 79, 48, 100, 88, 32, 45, 20, 30]

marks_range = [10, 20, 30, 40, 50, 60, 70, 80, 90, 100]