

**МИНОБРНАУКИ РОССИИ**  
**Федеральное государственное автономное образовательное**  
**учреждение высшего образования "Пермский**  
**государственный национальный исследовательский**  
**университет"**

Авторы-составители: **Арбузов Вячеслав Олегович**  
**Радионова Марина Владимировна**

Рабочая программа дисциплины  
**ANALYZING DATA IN PYTHON**  
Код УМК 97822

Утверждено  
Протокол №10  
от «09» июня 2021 г.

Пермь, 2021

## **1. Наименование дисциплины**

Analyzing Data in Python

## **2. Место дисциплины в структуре образовательной программы**

Дисциплина входит в обязательную часть Блока « Б.1 » образовательной программы по направлениям подготовки (специальностям):

Направление: **38.03.05** Бизнес-информатика

направленность Информационные системы и большие данные

### **3. Планируемые результаты обучения по дисциплине**

В результате освоения дисциплины **Analyzing Data in Python** у обучающегося должны быть сформированы следующие компетенции:

**38.03.05** Бизнес-информатика (направленность : Информационные системы и большие данные)

**ПК.3** Способен к планированию и организации аналитических работ, в том числе, с использованием технологий больших данных

#### **Индикаторы**

**ПК.3.3** Определяет необходимые ресурсы для проведения аналитических работ

#### 4. Объем и содержание дисциплины

<b>Направления подготовки</b>	38.03.05 Бизнес-информатика (направленность: Информационные системы и большие данные)
<b>форма обучения</b>	очная
<b>№№ триместров, выделенных для изучения дисциплины</b>	6
<b>Объем дисциплины (з.е.)</b>	3
<b>Объем дисциплины (ак.час.)</b>	108
<b>Контактная работа с преподавателем (ак.час.), в том числе:</b>	42
<b>Проведение лекционных занятий</b>	14
<b>Проведение лабораторных работ, занятий по иностранному языку</b>	28
<b>Самостоятельная работа (ак.час.)</b>	66
<b>Формы текущего контроля</b>	Входное тестирование (1) Итоговое контрольное мероприятие (1) Письменное контрольное мероприятие (2)
<b>Формы промежуточной аттестации</b>	Зачет (6 триместр)

## 5. Аннотированное описание содержания разделов и тем дисциплины

### **Input control**

The input control is designed to test students' knowledge on general topics related to algorithmization and programming:

variables, parameters, functions; cycles; classes of objects and their properties; code compilation; algorithmization of calculations; vectors, data arrays;

### **Introduction in Python**

Python is a high-level general-purpose programming language with dynamic typing and automatic memory management, focused on improving developer productivity, code readability and quality, as well as ensuring the portability of programs written in it. The language is completely object-oriented — everything is objects. An unusual feature of the language is the allocation of code blocks with whitespace indents.

### **Libraries in Python. Pip, Conda.**

The syntax of the language core is minimalistic, due to which in practice there is rarely a need to refer to documentation, the language itself is known as interpreted and is used, among other things, for writing scripts. The disadvantages of the language are often lower speed and higher memory consumption of programs written in it compared to similar code written in C or C++.

### **Linear algebra in Python, Numpy**

Mathematical algorithms implemented in interpreted languages (Python) often work much slower than the same algorithms implemented in compiled languages (for example, Fortran, C, Java). The NumPy library provides implementations of computational algorithms (in the form of functions and operators) optimized for working with multidimensional arrays. As a result, any algorithm that can be expressed as a sequence of operations on arrays (matrices) and implemented using NumPy works as fast as the equivalent code executed in MATLAB.

### **Time series in Python, Pandas**

Pandas is a Python software library for data processing and analysis. pandas' work with data is built on top of the NumPy library, which is a lower-level tool. Provides special data structures and operations for manipulating numeric tables and time series.

### **Visualization, matplotlib**

Matplotlib is a library in the Python programming language for visualizing data with two—dimensional (2D) graphics (3D graphics are also supported). The resulting images can be used as illustrations in publications

### **Data science and machine learning, scikit-learn**

The Scikit-learn library is the most common choice for solving classical machine learning problems. It provides a wide range of learning algorithms with and without a teacher. Learning with a teacher assumes the presence of a marked-up dataset in which the value of the target attribute is known. While teaching without a teacher does not imply the presence of markup in the dataset — you need to learn how to extract useful information from arbitrary data. One of the main advantages of the library is that it works on the basis of several common mathematical libraries, and easily integrates them with each other.

### **Neural networks, tensorflow and keras**

Keras is an open library written in Python and provides interaction with artificial neural networks. It is an add-on to the TensorFlow framework. Prior to version 2.3, different versions of neural network libraries were supported, such as TensorFlow, Microsoft Cognitive Toolkit, DeepLearning, and Theano. It is aimed at operational work with deep learning networks, while designed to be compact, modular and extensible.

### **GUI in Python, Tkinter**

Tkinter – это пакет для Python, предназначенный для работы с библиотекой Tk. Библиотека Tk содержит

компоненты графического интерфейса пользователя (graphical user interface – GUI). Эта библиотека написана на языке программирования Tcl.

### **NLP in Python, gensim**

Gensim is an open source Python library written by Radim Rehurek, which is used for unsupervised thematic modeling and natural language processing (NLP). It is designed to extract semantic topics from documents. Gensim can work with large text collections.

### **Final control**

Exam in final individual projects

## **6. Методические указания для обучающихся по освоению дисциплины**

Освоение дисциплины требует систематического изучения всех тем в той последовательности, в какой они указаны в рабочей программе.

Основными видами учебной работы являются аудиторские занятия. Их цель - расширить базовые знания обучающихся по осваиваемой дисциплине и систему теоретических ориентиров для последующего более глубокого освоения программного материала в ходе самостоятельной работы. Обучающемуся важно помнить, что контактная работа с преподавателем эффективно помогает ему овладеть программным материалом благодаря расстановке необходимых акцентов и удержанию внимания интонационными модуляциями голоса, а также подключением аудио-визуального механизма восприятия информации.

Самостоятельная работа преследует следующие цели:

- закрепление и совершенствование теоретических знаний, полученных на лекционных занятиях;
- формирование навыков подготовки текстовой составляющей информации учебного и научного назначения для размещения в различных информационных системах;
- совершенствование навыков поиска научных публикаций и образовательных ресурсов, размещенных в сети Интернет;
- самоконтроль освоения программного материала.

Обучающемуся необходимо помнить, что результаты самостоятельной работы контролируются преподавателем во время проведения мероприятий текущего контроля и учитываются при промежуточной аттестации.

Обучающимся с ОВЗ и инвалидов предоставляется возможность выбора форм проведения мероприятий текущего контроля, альтернативных формам, предусмотренным рабочей программой дисциплины. Предусматривается возможность увеличения в пределах 1 академического часа времени, отводимого на выполнение контрольных мероприятий.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации.

При проведении текущего контроля применяются оценочные средства, обеспечивающие передачу информации, от обучающегося к преподавателю, с учетом психофизиологических особенностей здоровья обучающихся.

## **7. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине**

При самостоятельной работе обучающимся следует использовать:

- конспекты лекций;
- литературу из перечня основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля);
- текст лекций на электронных носителях;
- ресурсы информационно-телекоммуникационной сети "Интернет", необходимые для освоения дисциплины;
- лицензионное и свободно распространяемое программное обеспечение из перечня информационных технологий, используемых при осуществлении образовательного процесса по дисциплине;
- методические указания для обучающихся по освоению дисциплины.

## 8. Перечень основной и дополнительной учебной литературы

### Основная:

1. Antonio Badia. SQL for Data Science / Antonio Badia // Publisher Name: Springer, Cham. — 2020. — 285 p. — ISBN 978-3-030-57592-2. [Электронный ресурс]. <https://link.springer.com/book/10.1007/978-3-030-57592-2>

### Дополнительная:

1. Connor P. Milliken. Python Projects for Beginners / Connor P. Milliken // Publisher Name: Apress, Berkeley, CA. — 2020. — 332 p. — ISBN 978-1-4842-5355-7. [Электронный ресурс]. <https://link.springer.com/book/10.1007/978-1-4842-5355-7>



## 9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

<https://numpy.org> The fundamental package for scientific computing with Python

<https://pandas.pydata.org> pandas official

<https://matplotlib.org> Matplotlib: Visualization with Python

<https://scikit-learn.org/stable/> Machine Learning in Python

<https://keras.io> Machine Learning in Python

## 10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

Образовательный процесс по дисциплине **Analyzing Data in Python** предполагает использование следующего программного обеспечения и информационных справочных систем:

The educational process in the discipline involves:

- presentation materials (slides on the topics of lectures and practical classes);
- on-line access to the Electronic Library System (EBS);
- access to the electronic information and educational environment of the university.

List of necessary licensed and (or) freely distributed software:

1. Python programming language development environment (free software).

При освоении материала и выполнения заданий по дисциплине рекомендуется использование материалов, размещенных в Личных кабинетах обучающихся ЕТИС ПГНИУ (**student.psu.ru**).

При организации дистанционной работы и проведении занятий в режиме онлайн могут использоваться:

система видеоконференцсвязи на основе платформы BigBlueButton (<https://bigbluebutton.org/>).

система LMS Moodle (<http://e-learn.psu.ru/>), которая поддерживает возможность использования текстовых материалов и презентаций, аудио- и видеоконтент, а так же тесты, проверяемые задания, задания для совместной работы.

система тестирования Indigo (<https://indigotech.ru/>).

## 11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Laboratory classes require a computer class equipped with personal computers and appropriate software. The composition of the equipment is defined in the Passport of the computer class.

To carry out independent work, a laboratory of information technologies in forecasting and managing the processes of socio-economic development is needed, equipped with specialized equipment, premises of the Scientific Library of the PSNIU.

For the current control, a computer class equipped with personal computers and appropriate software is required. The composition of the equipment is defined in the Passport of the computer class.

Individual and group consultations - an audience equipped with presentation equipment (projector, screen, computer/laptop) with appropriate software, chalk(s) or marker board, or an audience equipped with chalk (s) or marker board

Помещения научной библиотеки ПГНИУ для обеспечения самостоятельной работы обучающихся:

1. Научно-библиографический отдел, корп.1, ауд. 142. Оборудован 3 персональными компьютера с доступом к локальной и глобальной компьютерным сетям.

2. Читальный зал гуманитарной литературы, корп. 2, ауд. 418. Оборудован 7 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

3. Читальный зал естественной литературы, корп.6, ауд. 107а. Оборудован 5 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

4. Отдел иностранной литературы, корп.2 ауд. 207. Оборудован 1 персональным компьютером с доступом к локальной и глобальной компьютерным сетям.

5. Библиотека юридического факультета, корп.9, ауд. 4. Оборудована 11 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

6. Читальный зал географического факультета, корп.8, ауд. 419. Оборудован 6 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

Все компьютеры, установленные в помещениях научной библиотеки, оснащены следующим программным обеспечением:

Операционная система ALT Linux;

Офисный пакет Libreoffice.

Справочно-правовая система «КонсультантПлюс»

**Фонды оценочных средств для аттестации по дисциплине  
Analyzing Data in Python**

**Планируемые результаты обучения по дисциплине для формирования компетенции.  
Индикаторы и критерии их оценивания**

**ПК.3**

**Способен к планированию и организации аналитических работ, в том числе, с использованием технологий больших данных**

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
<p><b>ПК.3.3</b> Определяет необходимые ресурсы для проведения аналитических работ</p>	<p>Know • the main types of data and program constructions in Python; • methods of generating random variables with a given distribution law; • types of simulation models; • support for working in Python with regression models and time series models; • support in Python for working with objects of a non-numeric nature. Be able to • create functions and scripts in Python; • implement deterministic and stochastic simulation models with discrete and continuous time, as well as hybrid models in Python; • plan an experiment; • process the results of calculations; • work in Python with regression models — evaluate parameters, test statistical hypotheses, perform forecasting; • work in Python with time series models; • work with models containing objects of a non-numeric nature. Possess • basic techniques of working in the Python statistical package; • methods of solving equations and mathematical programming problems; • basic methods of statistical processing of the results of a computational experiment.</p>	<p align="center"><b>Неудовлетворител</b></p> <p>Does not know how to create functions and scripts in Python languages; Does not have data processing skills in Python; Does not know how to import and export data to Python; Does not know how to process calculation results; Does not know how to visualize calculation results; Does not know how to work with regression models in Python — evaluate parameters, check statistical hypotheses, perform forecasting; Does not know how to work in Python with time series models; Cannot work with stochastic processes and evaluate their parameters; Does not know how to build ensemble models and carry out their validation on a training and test sample;</p> <p align="center"><b>Удовлетворительн</b></p> <p>Can create functions and scripts in Python; Has data processing skills in Python; Can import and export data in Python; Cannot process calculation results; Cannot visualize calculation results; Cannot work in Python with regression models — evaluate parameters, check statistical hypotheses, perform forecasting; Cannot work in Python with does not know how to work with stochastic processes and evaluate their parameters; Does not know how to build ensemble models and carry out their validation on a training and test sample;</p> <p align="center"><b>Хорошо</b></p> <p>Can create functions and scripts in Python; Has data processing skills in Python; Can import and export data to Python; Can process calculation results; Can visualize calculation results; Can work in Python with regression models — evaluate parameters, check statistical</p>

Компетенция (индикатор)	Планируемые результаты обучения	Критерии оценивания результатов обучения
		<p style="text-align: center;"><b>Хорошо</b></p> <p>hypotheses, perform forecasting; Cannot work in Python with time series models; Does not know how to work with stochastic processes and evaluate their parameters; Does not know how to build ensemble models and carry out their validation on a training and test sample;</p> <p style="text-align: center;"><b>Отлично</b></p> <p>Can create functions and scripts in Python; Has data processing skills in Python; Can import and export data to Python; Can process calculation results; Can visualize calculation results; Can work in Python with regression models — evaluate parameters, check statistical hypotheses, perform forecasting; Can work in Python with time series models; Is able to work with stochastic processes and evaluate their parameters; Is able to build ensemble models and carry out their validation on a training and test sample;</p>

## Оценочные средства текущего контроля и промежуточной аттестации

Схема доставки : Базовая

**Вид мероприятия промежуточной аттестации :** Зачет

**Способ проведения мероприятия промежуточной аттестации :** Оценка по дисциплине в рамках промежуточной аттестации определяется на основе баллов, набранных обучающимся на контрольных мероприятиях, проводимых в течение учебного периода.

**Максимальное количество баллов :** 100

### Конвертация баллов в отметки

«отлично» - от 81 до 100

«хорошо» - от 61 до 80

«удовлетворительно» - от 50 до 60

«неудовлетворительно» / «незачтено» менее 50 балла

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
<b>Входной контроль</b>	Input control <b>Входное тестирование</b>	Input testing is a test consisting of 20 questions and tasks on general topics related to algorithmization and programming: - variables, parameters, functions; - cycles; - object classes and their properties; - code compilation; - algorithmization of calculations; - vectors, arrays of data;
<b>ПК.3.3</b> Определяет необходимые ресурсы для проведения аналитических работ	Time series in Python, Pandas <b>Письменное контрольное мероприятие</b>	The basics of syntax, the basic constructs of the Python programming language. Methods of using specialized Python packages for solving problems of processing and analyzing economic data. Methods for solving non-standard problems, analysis of errors occurring in the code, and methods for finding errors and solutions to eliminate these errors.
<b>ПК.3.3</b> Определяет необходимые ресурсы для проведения аналитических работ	Neural networks, tensorflow and keras <b>Письменное контрольное мероприятие</b>	Methods of using specialized Python packages for solving problems of processing and analyzing economic data. Methods for solving non-standard problems, analysis of errors occurring in the code, and methods for finding errors and solutions to eliminate these errors.

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
<b>ПК.3.3</b> Определяет необходимые ресурсы для проведения аналитических работ	Final control <b>Итоговое контрольное мероприятие</b>	Basics of Python programming language syntax, basic language structures. Features of the Python language, specialized language packages for automating the processing and analysis of economic data, complex economic calculations. Solving non-standard tasks related to calculations, analyzing errors that occur in the code, independently finding a solution to eliminate these errors within the Python language. The use of advanced methods for solving problems of processing and analyzing economic data, allowing to minimize the execution time of the code.

### Спецификация мероприятий текущего контроля

#### Input control

Продолжительность проведения мероприятия промежуточной аттестации: **1 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **0**

Проходной балл: **0**

Показатели оценивания	Баллы
Knows the basics of programming language syntax. Knows the specifics of using specialized packages and is able to solve and automate calculations of complex tasks.	20
Knows the basics of programming language syntax. Knows the specifics of using specialized packages and is able to solve, automate calculations of complex tasks. He is able to solve non-standard tasks related to calculations, analyze errors that occur in the code and independently find solutions to eliminate these errors. He is able to use advanced methods of solving problems that minimize the execution time of the code.	15
Knows the basics of programming language syntax. Knows the specifics of using specialized packages and is able to solve, automate calculations of complex tasks. He is able to solve non-standard tasks related to calculations, analyze errors that occur in the code and independently find solutions to eliminate these errors.	10
Does not know the basics of programming language syntax and is not able to solve the simplest tasks in this topic	5

#### Time series in Python, Pandas

Продолжительность проведения мероприятия промежуточной аттестации: **1 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **15**

<b>Показатели оценивания</b>	<b>Баллы</b>
Knows the basics of programming language syntax. Knows the specifics of using specialized packages and is able to solve, automate calculations of complex tasks. He is able to solve non-standard tasks related to calculations, analyze errors that occur in the code and independently find solutions to eliminate these errors. He is able to use advanced methods of solving problems that minimize the execution time of the code.	30
Knows the basics of programming language syntax. Knows the specifics of using specialized packages and is able to solve and automate calculations of complex tasks.	20
Knows the basics of programming language syntax and is able to solve the simplest tasks in this topic	10
Does not know the basics of programming language syntax and is not able to solve the simplest tasks in this topic	0

### **Neural networks, tensorflow and keras**

Продолжительность проведения мероприятия промежуточной аттестации: **1 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **15**

<b>Показатели оценивания</b>	<b>Баллы</b>
Knows the basics of programming language syntax. Knows the specifics of using specialized packages and is able to solve, automate calculations of complex tasks. He is able to solve non-standard tasks related to calculations, analyze errors that occur in the code and independently find solutions to eliminate these errors. He is able to use advanced methods of solving problems that minimize the execution time of the code.	30
Knows the basics of programming language syntax. Knows the specifics of using specialized packages and is able to solve and automate calculations of complex tasks.	20
Knows the basics of programming language syntax and is able to solve the simplest tasks in this topic	10
Does not know the basics of programming language syntax and is not able to solve the simplest tasks in this topic	0

### **Final control**

Продолжительность проведения мероприятия промежуточной аттестации: **1 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **40**

Проходной балл: **20**

<b>Показатели оценивания</b>	<b>Баллы</b>
Knows the basics of programming language syntax. Knows the specifics of using specialized packages and is able to solve, automate calculations of complex tasks. He is able to solve non-standard tasks related to calculations, analyze errors that occur in the code and independently	40

find solutions to eliminate these errors. He is able to use advanced methods of solving problems that minimize the execution time of the code.	
Knows the basics of programming language syntax. Knows the specifics of using specialized packages and is able to solve and automate calculations of complex tasks.	30
Knows the basics of programming language syntax and is able to solve the simplest tasks in this topic	20
Does not know the basics of programming language syntax and is not able to solve the simplest tasks in this topic	10