

## Программа учебной дисциплины «Основы программирования в Python»

Утверждена

Академическим советом ОП

Протокол № 2.6.-02/01от

29.08.2019

Разработчик	Тамбовцева Алла Андреевна, ассистент кафедры высшей математики
Число кредитов	4
Контактная работа (час.)	40
Самостоятельная работа (час.)	112
Курс, Образовательная программа	3 курс бакалавриата, «Политология»
Формат изучения дисциплины	С использованием онлайн курса

### 1. Цель, результаты освоения дисциплины и пререквизиты

Целями освоения дисциплины «Основы программирования в Python» являются овладение навыками программирования на языке Python, овладение методами сбора и обработки данных для решения политологических и социально-экономических задач.

В результате освоения дисциплины студент должен:

**знать:**

- основные методы сбора и обработки данных в Python;

**уметь:**

- находить необходимые для работы на языке программирования данные;

**владеть:**

- навыками программирования в Python;
- навыками работы с разными форматами файлов с данными.

Изучение дисциплины «Основы программирования в Python» базируется на следующих дисциплинах:

- Математика и статистика;
- Научно-исследовательский семинар.

Для освоения учебной дисциплины студенты должны владеть следующими знаниями и компетенциями:

- Базовые навыки работы с персональным компьютером.
- Знания математики и статистики в рамках обязательного курса «Математика и статистика».
- Знание английского языка, достаточное для чтения учебной литературы и документации.

Изучение дисциплины предполагает использование онлайн-курса "Introduction to Python" на образовательной платформе DataCamp (URL: <https://www.datacamp.com/courses/intro-to-python-for-data-science>), который посвящен базовым элементам программирования в Python для анализа данных. Перед семинаром студентам предлагается ознакомиться с материалами по соответствующей теме: прослушать видео-лекции и выполнить ряд простых заданий. На семинаре преподаватель отвечает на вопросы студентов, предлагает задачи для самостоятельного решения и осуществляет их разбор. Выполнение блоков данного онлайн-курса учитывается в формуле оценивания.

Основные положения дисциплины должны быть использованы в дальнейшем при изучении следующих дисциплин:

- Анализ данных в Python;
- Введение в сетевой анализ.

## 2. Содержание учебной дисциплины

Тема (раздел дисциплины)	Объем в часах	Планируемые результаты обучения (ПРО), подлежащие контролю	Формы контроля
Тема 1. Начало работы в Python	см: 4	Умеет работать в среде Jupyter Notebook, умеет создавать переменные разных типов	Домашнее задание; тест на 15 минут
	онл: 2		
	ср: 2		
Тема 2. Управляющие конструкции в Python	см: 2	Умеет решать прикладные задачи с применением условных конструкций и циклов	Тест на 15 минут; контрольная работа на 80 минут
	онл: 4		
	ср: 4		
Тема 3. Структуры данных в Python	см: 4	Знает особенности изменяемых и неизменяемых типов данных; умеет создавать разные структуры данных	Домашнее задание; тест на 15 минут; контрольная работа на 80 минут
	онл: 2		
	ср: 6		
Тема 4. Циклы в Python	см: 2	Умеет решать прикладные задачи с применением условных конструкций и циклов	Тест на 15 минут; контрольная работа на 80 минут
	онл: 4		
	ср: 4		
Тема 5. Функции в Python	см: 2	Умеет писать функции на Python и тестировать их	Домашнее задание; контрольная работа на 80 минут
	онл: 4		
	ср: 6		
Тема 6. Работа с файлами	см: 4	Умеет загружать файлы с данными разных форматов	Домашнее задание; проект
	ср: 6		

Тема 7. Работа с текстами	см: 2	Знает этапы предварительной обработки текста и умеет реализовывать их на практике	Домашнее задание
	ср: 8		
Тема 8. Массивы NumPy	онл: 4	Умеет выполнять операции с массивами NumPy	Домашнее задание
	ср: 6		
Тема 9. Работа с таблицами	см: 4	Умеет загружать файлы с данными; выполняет обработку табличных данных средствами Python	Домашнее задание; проект
	ср: 12		
Тема 10. Визуализация данных	см: 4	Умеет строить графики математических функций и визуализировать данные в Python	Домашнее задание; проект
	ср: 6		
Тема 11. Работа с html-файлами	см: 4	Умеет выгружать информацию из веб-страниц	Домашнее задание; проект
	ср: 10		
Тема 12. Управление браузером из Python	см: 4	Умеет выгружать информацию из веб-страниц, используя средства для управления браузера	Проект
	ср: 10		
Тема 13. Работа с API	см: 4	Умеет выгружать информацию из API	Проект
	ср: 12		
<b>Часов по видам учебных занятий:</b>	см: 40		
	онл: 20		
	ср: 92		
<b>Итого часов:</b>	<b>152</b>		

### *Содержание разделов дисциплины:*

#### *Тема 1. Начало работы в Python.*

Установка Python 3 (дистрибутив Anaconda). Знакомство с интерфейсом Jupyter Notebook. Элементы языка разметки Markdown. Установка и импортирование библиотек. Элементарные вычисления в Python. Переменные в Python. Типы данных в Python: числовой, целочисленный, логический, строковый. Преобразование типов.

#### *Тема 2. Управляющие конструкции в Python.*

Понятие управляющих конструкций. Построение блок-схемы для понимания структуры кода. Условные конструкции и операторы if-else. Особенности использования оператора elif.

#### *Тема 3. Структуры данных в Python*

Списки, кортежи, множества, словари. Изменяемые и неизменяемые типы данных.

#### *Тема 4. Циклы в Python*

Циклы for и while. Применение циклов к строкам, спискам, кортежам и словарям. Альтернатива коду с циклами: списковые включения.

### *Тема 5. Функции в Python*

Устройство функций в Python. Написание простейших функций. Lambda-функции. Исключения. Поиск ошибок в коде и отладка.

### *Тема 6. Работа с файлами*

Работа с файлами в Python: открытие, изменение, сохранение. Разные форматы хранения данных: csv-файлы, json-файлы, txt-файлы.

### *Тема 7. Работа с текстами*

Предварительная обработка текста. Токенизация, лемматизация, стемминг. Обзор возможностей библиотек `rumorphy2` и `rumystem3`.

### *Тема 8. Массивы NumPy*

Массивы vs списки в Python. Работа с массивами NumPy. Применение массивов в анализе данных.

### *Тема 9. Работа с таблицами*

Работа с файлами Excel: открытие и сохранение файлов. Обзор возможностей библиотеки `pandas`. Преобразование датафреймов `pandas`: добавление строк и столбцов в таблицу, фильтрация строк по условиям. Группировка и агрегирование данных. Объединение таблиц.

### *Тема 10. Визуализация данных*

Построение графиков функций средствами библиотеки `matplotlib`. Визуализация качественных и количественных данных с помощью библиотек `matplotlib` и `seaborn`. Облака слов и возможности библиотеки `wordcloud`.

### *Тема 11. Работа с html-файлами*

Введение в HTML и элементы веб-дизайна. Извлечение данных из веб-страниц. Парсинг html-файлов в Python с помощью библиотек `requests` и `BeautifulSoup`.

### *Тема 12. Управление браузером из Python*

Знакомство с инструментом CSS-selector. Установка библиотеки `Selenium` и драйверов для браузера. Управление браузером с помощью библиотеки `Selenium`.

### *Тема 13. Работа с API*

Знакомство с API. API как источник данных. API Wikipedia. Работа с API социальных сетей на примере ВКонтакте.

### 3. Оценивание

Курс предусматривает следующие контрольные мероприятия: контрольная работа, домашние задания, тесты, выполнение блоков онлайн-курса, проект. Все перечисленные виды работ оцениваются в 10-балльной шкале.

- ***Контрольная работа (КР)***

Контрольная работа проводится в конце первого модуля и рассчитана на 80 минут. Контрольная работа состоит из двух частей: теоретической (40 минут) и практической (40 минут). Теоретическая часть включает несколько тестовых вопросов и вопросов с развернутым ответом по базовым фактам о языке программирования Python (типы данных, изменяемость объектов, особенности работы с разными объектами Python), а также задания с небольшими отрывками кода, в которых необходимо найти ошибки или пояснить, что будет выведено на экран при исполнении кода. Практическая часть включает три задачи по программированию (по темам первого модуля), выполняемые на компьютере.

Во время выполнения теоретической части пользоваться какими-либо материалами нельзя, во время выполнения практической части можно использовать материалы курса, а также интернет. Во время контрольной запрещается общаться с другими студентами, использовать мобильные телефоны, открывать социальные сети и создавать новые вопросы на компьютерных форумах и ресурсах по типу stackoverflow.

В случае нарушения правил проведения контрольной работы студент удаляется с контрольной работы с оценкой «0». К нарушениям правил проведения контрольной работы относятся: коммуникация с другими студентами во время выполнения работы, использование социальных сетей/телефона во время работы (с любой целью), списывание.

- ***Домашнее задание (ДЗ)***

Домашние задания предполагают решение задач по программированию по пройденным темам. С учетом разного уровня подготовки студентов домашние задания представляют собой набор задач разной сложности, которые весят разное количество баллов. В начале каждого домашнего задания указано общее число баллов, которое нужно набрать, чтобы получить оценку 10. Таким образом, студент может выбирать, какие задачи ему решать. Например, если за все домашнее задание нужно набрать 8 баллов, чтобы получить оценку 10, студент может решить 8 задач по одному баллу или 4 задачи по одному баллу и одну задачу на 4 балла. Кроме того, студент может решить больше задач, чтобы потренироваться. Все задачи будут проверены и оценены, но получить оценку выше 10 за домашнее задание невозможно.

Работа студентов оценивается по следующим критериям: работоспособность и корректность кода программы (код должен запускаться без ошибок и выдавать ожидаемый результат), использование рассмотренных в курсе средств и методов, соответствие требованиям преподавателя (формат и срок сдачи заданий). Задачи, для решения которых предоставлен неработающий код (код, который невозможно запустить из-за наличия грубых ошибок/опечаток), не засчитываются, даже если при этом зафиксирован верный результат.

Оценка за домашние задания, которая подставляется в формулу оценивания, определяется как округленное до целого среднее арифметическое оценок, полученных за домашние работы (способ округления – стандартный арифметический). Домашние задания, сданные после срока, оцениваются с использованием понижающих коэффициентов: опоздание в пределах часа – штраф 10% от полученной оценки, в пределах суток – штраф 20%, в пределах недели – штраф 50%. Домашние задания, сданные через неделю после указанного срока и позже, не принимаются и не оцениваются.

Если при проверке работ установлен факт нарушения академической этики, студент получает оценку «0» за данную работу. Работа студента, предоставившего свою работу для списывания, также аннулируется.

- ***Тесты***

Тесты представляют собой набор открытых вопросов и вопросов с выбором ответа по пройденным темам (разобраным очно или онлайн на платформе DataCamp). Тесты направлены на проверку теоретических знаний студентов по программированию в Python: назначение и особенности работы операторов, устройство управляющих конструкций и циклов, особенности типов и структур данных, типы ошибок. Тесты проводятся в начале семинара и рассчитаны на 10-15 минут. Во время теста пользоваться какими-либо материалами нельзя.

Оценка за тесты, которая подставляется в формулу оценивания, представляет собой округленное среднее арифметическое за все тесты (способ округления – стандартный арифметический).

- ***Онлайн-курс (Онлайн)***

Выполнение блоков из онлайн курса "Introduction to Python" на платформе DataCamp по разбираемым темам. Перед семинаром на платформе DataCamp вывешивается задание с фиксированным дедлайном – просмотреть материалы по определенной теме и выполнить предложенные в блоке задачи.

Оценка за онлайн часть представляет собой долю выполненных студентом блоков от общего числа заданных блоков, домноженную на 10 и округленную (способ округления – стандартный арифметический).

- ***Проект***

Проект представляет собой написание корректно работающей программы прикладного назначения с последующей устной защитой в виде презентации. Проект включает в себя следующие части: программа Python (файл с расширением .py или файл Jupyter Notebook с расширением .ipynb), pdf-файл с описанием назначения и принципов работы программы (документация) и презентация, включающая публичную демонстрацию работы программы. Проект обязательно должен включать использование навыков, полученных в результате прохождения тем второго модуля, то есть выгрузку данных с веб-страницы или с помощью API, сохранение данных в датафрейм pandas с последующей их обработкой, визуализация данных средствами matplotlib или других библиотек.

Проект выполняется индивидуально или в группе не более 3 человек. При коллективной защите проекта все члены группы должны быть подготовлены и готовы ответить на вопросы по проекту.

Оценка за проект определяется по следующей формуле (способ округления – стандартный арифметический):

$$\text{Проект} = \text{Округление}(0.15 * \text{План} + 0.4 * \text{Программа} + 0.15 * \text{Документация} + 0.3 * \text{Презентация}).$$

- **План:** описание проекта; включает назначение проекта, принцип работы программы, источник используемых данных, планируемые средства визуализации и первичного анализа данных.
- **Программа:** сама программа (файл .py или .ipynb); оценивается по следующим критериям: работоспособность кода, соответствие используемых инструментов заявленной цели, применение при ее написании навыков, изученных в течение курса, корректное оформление кода.
- **Документация:** pdf-файл с описанием функционала программы, требований к данным, подаваемых на вход, источников данных, результата выполнения программы, формата данных на выходе.
- **Презентация:** устная защита проекта в любом формате; обязательно включает демонстрацию работы программы; оценивается по следующим критериям: наличие объяснения цели программы, ее принципа работы и получаемых результатов, способность корректно ответить на вопросы по работе программы в рамках тем, изученных на курсе, объяснить ограничения работы программы и потенциальные проблемы.

Итоговая оценка по курсу определяется по следующей формуле (способ округления – стандартный арифметический):

$$\text{Итоговая оценка} = \text{Округление}(0.2 * \text{КР} + 0.15 * \text{ДЗ} + 0.15 * \text{Тесты} + 0.1 * \text{Онлайн} + 0.4 * \text{Проект}).$$

На первой и второй передаче студенту предлагается сдать проект в формате, предусмотренном программой курса. Все компоненты проекта (план, программа, документация, презентация) и критерии оценивания сохраняются. В случае передачи проект выполняется индивидуально, выполнение в группах не допускается.

#### 4. Примеры оценочных средств

Блокирующие элементы не предусмотрены.

##### *Примеры вопросов контрольной работы*

1. Соотнесите код с результатами вывода (одной строке кода соответствует ровно один результат):

i. <code>int(float('4.5'))</code>	a) 4.0
ii. <code>int(5.1)</code>	b) 5
iii. <code>float('4')</code>	c) 4
iv. <code>int('4.0')</code>	d) '4.0'
v. <code>str(float(4))</code>	e) ValueError.

2. Словарь (dictionary) в Python:

- a) Может хранить повторяющиеся ключи.
- b) Может хранить повторяющиеся значения.
- c) Представляет из себя кортеж кортежей.

3. Дан код:

```
def my_prod(a, b):  
    return a * b
```

Какие из следующих примеров вызова пользовательской функции будут исполнены без ошибок?

- a) `my_prod(1, 2)`
- b) `my_prod(1.0, 2.0)`
- c) `my_prod('1', '2')`
- d) `my_prod([1,2], [3,4])`
- e) `my_prod('1', 2)`
- f) `my_prod(1).`

4. Напишите, в чем заключается принципиальная разница между списком (list) и кортежем (tuple).



5. Какие из перечисленных операций могут быть успешно (без ошибок) выполнены со списком  $t = [10, 100, 1000]$ :

- a) `t.append(10)`
- b) `t = t + (100, 10)`
- c) `t[2] = 100`
- d) `t = tuple(t) + [10, 11]`?

## 5. Ресурсы

### 5.1. Рекомендуемая основная литература

№п/п	Наименование
1	Федоров Д. Ю. Программирование на языке высокого уровня Python. 2-е изд. – Москва: Издательство Юрайт, 2019. – 161 с. <i>Электронный ресурс по подписке ВШЭ</i>
2	Nelli F. Python Data Analytics: With Pandas, NumPy, and Matplotlib. 2nd ed. – Apress, 2018. – 576 с. – На англ.яз. <i>Электронный ресурс по подписке ВШЭ</i>

### 5.2. Рекомендуемая дополнительная литература

№п/п	Наименование
1	Nair V. G. Getting Started with Beautiful Soup. Packt Publishing Limited, 2010. – На англ.яз. <i>Электронный ресурс по подписке ВШЭ</i>

### 5.3. Программное обеспечение

№п/п	Наименование	Условия доступа/скачивания
1	Microsoft Windows 7 Professional RUS	<i>Из внутренней сети университета (договор)</i>
2	Microsoft Office Professional Plus 2010	<i>Из внутренней сети университета (договор)</i>
3	Python Software Foundation Python	<i>Свободное лицензионное соглашение</i>

5.4. Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)

№п/п	Наименование	Условия доступа/скачивания
1	Электронно-библиотечная система Юрайт	URL: <a href="https://biblio-online.ru/">https://biblio-online.ru/</a>
2	Образовательная платформа DataCamp	URL: <a href="https://www.datacamp.com/">https://www.datacamp.com/</a>

## **6. Материально-техническое обеспечение дисциплины**

Учебные аудитории для лекционных занятий по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:

- ПЭВМ с доступом в Интернет (операционная система, офисные программы, антивирусные программы);
- мультимедийный проектор с дистанционным управлением.

Учебные аудитории для лабораторных и самостоятельных занятий по дисциплине оснащены ПЭВМ, с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.

## **7. Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов**

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося) а для инвалидов также в соответствии с индивидуальной программой реабилитации инвалида, могут предлагаться следующие варианты восприятия учебной информации с учетом их индивидуальных психофизических особенностей, в том числе с применением электронного обучения и дистанционных технологий:

*7.1. для лиц с нарушениями зрения:* в печатной форме увеличенным шрифтом; в форме электронного документа; в форме аудиофайла (перевод учебных материалов в аудиоформат); в печатной форме на языке Брайля; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания и консультации.

*7.2. для лиц с нарушениями слуха:* в печатной форме; в форме электронного документа; видеоматериалы с субтитрами; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания и консультации.

*7.3. для лиц с нарушениями опорно-двигательного аппарата:* в печатной форме; в форме электронного документа; в форме аудиофайла; индивидуальные задания и консультации.

## **8. Дополнительные сведения**

Материалы курса публикуются на странице <http://math-info.hse.ru>.