

**Автономная некоммерческая организация дополнительного профессионального образования «Центральный многопрофильный институт профессиональной переподготовки и повышения квалификации» АНО ДПО «ЦМИ»**



**Дополнительная общеобразовательная  
общеразвивающая программа  
«Анализ данных на Python»**

*Направленность: техническая  
Уровень сложности: «Базовый»*

*144 часа*

**Общие данные о Дополнительной общеобразовательной программе  
«Анализ данных на Python»**

**Об организации**

Наименование поля	Допустимые значения поля	Значение поля
ИНН организации, осуществляющей образовательную деятельность	10 арабских цифр	2632108764
Наименование организации	строка	Автономная некоммерческая организация дополнительного профессионального образования «Центральный многопрофильный институт профессиональной переподготовки и повышения квалификации» АНО ДПО «ЦМИ»
Логотип организации	изображение в формате jpeg разрешением не менее 100x100 пиксель	
Ссылка на логотип организации	URL на изображение, находящееся в сети интернет	<a href="https://drive.google.com/file/d/13efbea8lxUgQBuaRVzhZd55yOcl-W1TX/view?usp=sharing">https://drive.google.com/file/d/13efbea8lxUgQBuaRVzhZd55yOcl-W1TX/view?usp=sharing</a>
Контакты ответственного за программу (с указанием фамилии, имени, отчества).	строка от 5 до 255 символов	Пташкина Евгения Александровна
Контакты ответственного за программу. Должность	строка от 5 до 255 символов	Старший методист
Контакты ответственного за программу. Телефон	Формат +7(XXX)XXXXXXX	+7(495) 1523955
Контакты ответственного за программу. E-mail	строка	kod-sigma@edu-sigma.ru

## Информация о программе

Наименование поля	Допустимые значения поля	Значение поля (примеры)
Название программы (курса)	строка	Анализ данных на Python
Описание программы	строка не менее 1000 не более 5000 символов	<p>Данная программа предусмотрена для обучающихся 8-11 классов и обучающихся по программам среднего профессионального образования. В процессе освоения образовательной программы обучающийся получит опыт проведения исследовательского анализа данных для поиска зависимостей (EDA), обучится методам очистки данных, алгоритмам кластеризации и методу ближайших соседей, а также обучающийся научится собирать данные через API, выбирать тот или иной алгоритм для поставленной задачи, искать и обрабатывать ошибки в коде, разбивать решение задачи на подзадачи, использовать математические и логические способности, грамотно и эффективно решать практические задачи для Data Science. Содержание программы способствует повышению интереса обучающихся к программированию на языке Python, формированию самостоятельности и творческого подхода к решению задач с помощью средств современной вычислительной техники.</p> <p><b>Задачи программы:</b>  <b>Обучающие:</b></p> <ul style="list-style-type: none"> <li>• обучить языку программирования Python и созданию программ на его основе;</li> <li>• познакомить с методами обработки больших данных;</li> <li>• расширить кругозор обучающихся в области программирования.</li> </ul> <p><b>Развивающие:</b></p> <ul style="list-style-type: none"> <li>• развивать память и внимание, познавательную и творческую активность;</li> <li>• развивать логическое мышление.</li> </ul> <p><b>Воспитательные:</b></p>

		<ul style="list-style-type: none"> <li>• прививать интерес к активному творческому самовыражению, культуре труда;</li> <li>• воспитывать упорство в достижении желаемого результата;</li> <li>• воспитывать чувство взаимопомощи, доверия, коллективизма.</li> </ul>
<p>Аннотация (для размещения на маркетплейсе, понятное и привлекательное для Потенциальных получателей поддержки, включающее полное и содержательное описание Дополнительной общеобразовательной программы:</p> <ol style="list-style-type: none"> <li>1) краткое описание Дополнительной общеобразовательной программы;</li> <li>2) описание требований и рекомендаций для обучения по Дополнительной общеобразовательной программе;</li> <li>3) краткое описание результатов обучения в свободной форме, включая описание практикоориентированного характера Дополнительной общеобразовательной программы)</li> </ol>	<p>строка до 1000 символов</p>	<p>Программа «Анализ данных на Python» направлена на изучение языка программирования Python и на развитие навыков анализа данных и машинного обучения с целью дальнейшего профессионального самоопределения обучающихся, развитие инженерного мышления, воспитание конкурентоспособной личности.</p> <p>В рамках программы обучающиеся изучат основы Python, включая синтаксис языка, работу с переменными, функциями и циклами. Они также научатся работать с библиотеками NumPy и Pandas для обработки и анализа данных.</p> <p>Кроме того, обучающиеся будут изучать основные концепции машинного обучения, такие как классификация, регрессия и кластеризация. Они научатся использовать библиотеки Scikit-learn для построения моделей машинного обучения.</p> <p>Программа также включает практические задания, которые помогут обучающимся применить полученные знания на практике. Обучающиеся получат опыт работы с реальными данными из различных источников, таких как социальные сети, финансовые отчеты и метеорологические данные.</p> <p>Обучаться по программе могут школьники, обучающиеся с 8 по 11 классы и обучающиеся по программам среднего профессионального образования, являющиеся гражданами Российской Федерации и успешно сдавшие вступительное испытание.</p> <p><b>В результате обучения</b> выпускник программы будет способен:</p> <ul style="list-style-type: none"> <li>• использовать базовые типы данных, арифметические операции, условные логические конструкции, итерационные и условные циклы, методы и функции на языке программирования Python;</li> </ul>

		<ul style="list-style-type: none"> <li>● использовать библиотеки Pandas, Matplotlib, NumPy, Sklearn;</li> <li>● проводить анализ и визуализацию данных;</li> <li>● использовать алгоритмы обработки больших данных, в том числе и алгоритмы машинного обучения с учителем и без учителя;</li> <li>● использовать методы NLP.</li> </ul>
Цель программы	строка не менее 100 символов	<b>Цель программы</b> - изучение языка программирования Python с целью дальнейшего профессионального самоопределения обучающихся, развитие инженерного мышления, воспитание конкурентоспособной личности, а также получение практических навыков использования Python для Data Science
Актуальность	строка не менее 500 символов	<b>Актуальность программы</b> обусловлена необходимостью профориентации обучающихся в области информационных технологий, а также повышения их мотивации к обучению современным языкам программирования. Содержание программы способствует формированию необходимых компетенций для применения информационных технологий в России. Важнейшей характеристикой подрастающего поколения является активность в информационном пространстве, интернет-коммуникации. Обучающиеся приобретают необходимые навыки по использованию Python для Data Science, что им помогает определиться с профессиональной сферой деятельности на будущее.
Дополнительная информация	строка	
Формат обучения	значение из: очная форма без применения дистанционных образовательных	очная форма с применением дистанционных образовательных технологий, в том числе, с применением средств электронного обучения

	технологий; очная форма с применением дистанционных образовательных технологий, в том числе, с применением средств электронного обучения	
Уровень сложности	значение из: «Начальный» «Базовый» «Продвинутый»	«Базовый»
Срок освоения образовательной программы	строка, значение в ак.ч.	144 ак.ч.
Объем каждого модуля в ак.ч.	целое число	36 ак.ч.
Объем часов в неделю в ак.ч.	целое число	4 ак.ч.
Количество занятий	целое число	55
Направленность программы	строка	техническая
Язык программирования	строка	Python
Дополнительная общеобразовательная программа не представлена для участия в иных федеральных проектах, направленных на дополнительное образование граждан, кроме федерального проекта «Развитие кадрового потенциала ИТ- отрасли»	строка, значения: «Не представлена»	«Не представлена»
Дополнительная общеобразовательная программа не была реализована до начала отбора и/или не реализовывается в период отбора на безвозмездной основе	строка, значения «Не реализована»	«Не реализована»

Категория обучающихся по программе	строка не менее 10 символов	Учащиеся 8 класса, Учащиеся 9 класса, Учащиеся 10 класса, Учащиеся 11 класса Обучающиеся по программам среднего профессионального образования
Описание планируемых результатов обучения	строка не менее 10 символов	<b>В результате обучения</b> выпускник программы будет способен: <ul style="list-style-type: none"> <li>● использовать базовые типы данных, арифметические операции, условные логические конструкции, итерационные и условные циклы, методы и функции на языке программирования Python;</li> <li>● использовать библиотеки Pandas, Matplotlib, NumPy, Sklearn;</li> <li>● проводить анализ и визуализацию данных;</li> <li>● использовать алгоритмы обработки больших данных, в том числе и алгоритмы машинного обучения с учителем и без учителя;</li> <li>● использовать методы NLP.</li> </ul>
Ссылка на лендинг Образовательной программы	строка не менее 10 символов	<a href="https://school.edu-sigma.ru/pythonbase">https://school.edu-sigma.ru/pythonbase</a>
Ссылка на LMS	строка не менее 10 символов	<a href="https://odin.study/ru/">https://odin.study/ru/</a>
Страница обучения на курсе	строка не менее 10 символов	<a href="https://www.odin.study/ru/EducationalProgram/Info/7536">https://www.odin.study/ru/EducationalProgram/Info/7536</a>

## Аттестация

Промежуточная аттестация		
Модуль 1. Применение основных конструкций Python для задач анализа данных		
Количество академических часов	строка не менее 10 символов	1 академический час
Формы контроля	строка не менее 10 символов	Тестирование с практическими заданиями
Диагностические инструменты	строка не менее 10 символов	Тестирование с практическими заданиями
Показатели и критерии оценивания	строка не менее 10 символов	Оценка "5" выставляется, если обучающийся набрал 5 баллов за тестирование. Оценка "4" ставится, если обучающийся набрал 4 балла за тестирование. Оценка "3" ставится, если обучающийся набрал 3 балла за тестирование. Оценка "2" ставится, если обучающийся набрал менее 3 баллов за тестирование.
Примеры заданий	строка не менее 10 символов	<p>1. Какой индекс у числа 17 в списке numbers?  <code>numbers = [1, 100, 7, 20, 17, 37, 22]</code>                      Выберите один вариант из списка</p> <ol style="list-style-type: none"> <li>1. 3</li> <li>2. 5</li> <li>3. 4</li> <li>4. 6</li> </ol> <p>2. Что будет выведено в результате выполнения следующего программного кода?  <code>numbers = [0, 1, 3, 14, 2, 7, 9, 8, 10]</code>  <code>print(numbers)</code>                      Выберите один вариант из списка</p> <ol style="list-style-type: none"> <li>1. [0, 1, 3, 14, 2, 7, 9, 8, 10]</li> <li>2. 0, 1, 3, 14, 2, 7, 9, 8, 10</li> <li>3. ['0', '1', '3', '14', '2', '7', '9', '8', '10']</li> <li>4. 0 1 3 14 2 7 9 8 10</li> </ol>

		<p>3. Что будет выведено в результате выполнения следующего программного кода?</p> <pre>colors = ['Orange'] colors.append('Red') colors.append('Blue') colors.append('Green') colors.insert(0, 'Violet') colors.insert(2, 'Purple') print(colors)</pre> <p>Выберите один вариант из списка</p> <ol style="list-style-type: none"> <li>1. ['Orange', 'Violet', 'Purple', 'Red', 'Blue', 'Green']</li> <li>2. ['Violet', 'Orange', 'Purple', 'Red', 'Blue', 'Green']</li> <li>3. ['Violet', 'Purple', 'Orange', 'Red', 'Blue', 'Green']</li> <li>4. ['Violet', 'Red', 'Purple', 'Blue', 'Green']</li> </ol> <p>4. Напишите код программы, который запрашивает ввод с клавиатуры натурального числа и выводит на экран сумму его цифр.</p> <p>Формат входных данных: На вход программе подается натуральное число, например 5678</p> <p>Формат выходных данных: Программа должна вывести одно число в соответствии с условием задачи, например 26</p> <p>5. Дан код: <math>T = (4, 2, 3)</math></p> <p>Дополните код программы, который приведет к тому, что переменная T будет ссылаться на кортеж (1, 2, 3).</p>
Шкала оценивания, нижнее значение	строка не менее 10 символов	"2"
Шкала оценивания, верхнее значение	строка не менее 10 символов	"5"

Шкала оценивания, минимальный проходной балл	строка не менее 10 символов	"3"
<b>Модуль 2. Big Data: Основы работы с большими массивами данных</b>		
Количество академических часов	строка не менее 10 символов	1 академический час
Формы контроля	строка не менее 10 символов	Тестирование с практическими заданиями
Диагностические инструменты	строка не менее 10 символов	Тестирование с практическими заданиями
Показатели и критерии оценивания	строка не менее 10 символов	Оценка "5" выставляется, если обучающийся набрал 5 баллов за тестирование. Оценка "4" ставится, если обучающийся набрал 4 балла за тестирование. Оценка "3" ставится, если обучающийся набрал 3 балла за тестирование. Оценка "2" ставится, если обучающийся набрал менее 3 баллов за тестирование.
Примеры заданий	строка не менее 10 символов	<ol style="list-style-type: none"> <li>1. Какую функцию нужно использовать для чтения данных, хранящихся в виде CSV-файла, и преобразования их в DataFrame?</li> </ol> <p>Выбери верный вариант ответа.</p> <ol style="list-style-type: none"> <li>1. read_excel()</li> <li>2. read_csv()</li> <li>3. to_excel()</li> <li>4. to_csv()</li> </ol> <ol style="list-style-type: none"> <li>2. Какой код выполнит фильтрацию определенных строк из DataFrame?</li> </ol> <p>Выбери верный вариант ответа.</p> <ol style="list-style-type: none"> <li>1. df["Age"]</li> <li>2. df[df["Age"] &gt; 18]</li> <li>3. df["Age"] &gt; 18</li> </ol>

4. df.describe()

3. Дан код:

```
import json
str_js = """
{ "values": [
  { "time": 15642456455,
    "score": 56.8 }
  ]
}
"""
tek_data = json.loads(str_js)
print(tek_data['values'][0]['score'])
```

Что выведет данный код?

Выбери верный вариант ответа.

1. score
2. Ошибка
3. 15642456455
4. 56.8

4. Дан код:

```
from bs4 import BeautifulSoup
a = "<div class='col-lg-4'>
<div class='card'>
<a href='/list/9/'><img alt=''' class='card' src='/img/9.jpg'/></a>
<div class='card-body'>
<h4 class='card-title'>
<a href='/list/9/'>1 кг. яблок</a>
</h4> <h5>240 р.</h5>
</div> </div> </div>'"
bs = BeautifulSoup(a, 'html.parser')
itemName = #Тут добавить свой код
itemPrice = #Тут добавить свой код
print(f {itemPrice} за {itemName}')
```

		<p>Дополните данный код программы, чтобы результатом вывода на экран было следующее: 240 р. за 1 кг. яблок.</p> <p>5. Напишите код программы, который выводит информацию о минимальном и максимальном возрасте пассажиров “Титаника”, о количестве пассажиров младше 18 лет. Обязательно использование библиотеки pandas (датасет выдает преподаватель).</p>
Шкала оценивания, нижнее значение	строка не менее 10 символов	"2"
Шкала оценивания, верхнее значение	строка не менее 10 символов	"5"
Шкала оценивания, минимальный проходной балл	строка не менее 10 символов	"3"
<b>Модуль 3. Exploratory Data Analysis</b>		
Количество академических часов	строка не менее 10 символов	1 академический час
Формы контроля	строка не менее 10 символов	Тестирование с практическими заданиями
Диагностические инструменты	строка не менее 10 символов	Тестирование с практическими заданиями
Показатели и критерии оценивания	строка не менее 10 символов	<p>Оценка "5" выставляется, если обучающийся набрал 5 баллов за тестирование.</p> <p>Оценка "4" ставится, если обучающийся набрал 4 балла за тестирование.</p> <p>Оценка "3" ставится, если обучающийся набрал 3 балла за тестирование.</p> <p>Оценка "2" ставится, если обучающийся набрал менее 3 баллов за тестирование.</p>
Примеры заданий	строка не менее 10 символов	<p>1. Как выбрать первые 10 записей столбца Name в DataFrame?</p> <p>Выбери верный вариант ответа.</p> <p>1. <code>df.Name[:11]</code></p>

		<p>2. <code>df.Name[:10]</code>  3. <code>df.Name[10:]</code>  4. <code>df.Name[:11:]</code></p> <p>2. Какой метод применяется для получения визуального представления данных из DataFrame:</p> <p>Выбери верный вариант ответа.</p> <p>1. <code>df.plot()</code>  2. <code>df.info()</code>  3. <code>df.describe()</code>  4. <code>df.read_csv()</code></p> <p>3. Что будет выведено в результате выполнения следующего программного кода?</p> <pre>import pandas as pd  data = pd.Series([11, 20, 22, 31, 32, 33, 41, 42, 43, 44, 51, 52, 53, 54, 55, 61, 62, 63, 64, 65, 66, 71, 72, 73, 74, 75, 76, 77, 81, 82, 83, 84, 85, 86, 87, 88, 91, 92, 93, 94, 95, 96, 97, 98, 99])  data.hist(bins=4, alpha=0.5)</pre> <p>Выбери верный вариант ответа.</p> <p>1. гистограмма частот  2. гистограмма плотностей  3. линейный график  4. диаграмма рассеяния</p> <p>4. Напишите код программы, который выполняет очистку и преобразования данных по заданию (датасет выдает преподаватель).</p> <p>5. Напишите код программы, который выполняет выборку и анализ данных по заданию (датасет выдает преподаватель).</p>
Шкала оценивания, нижнее значение	строка не менее 10 символов	"2"

Шкала оценивания, верхнее значение	строка не менее 10 символов	"5"
Шкала оценивания, минимальный проходной балл	строка не менее 10 символов	"3"
<b>Модуль 4. Базовые алгоритмы машинного обучения в Sklearn</b>		
Количество академических часов	строка не менее 10 символов	1 академический час
Формы контроля	строка не менее 10 символов	Практическое задание
Диагностические инструменты	строка не менее 10 символов	Оценка полноты выполнения работы. Оценка способности оперировать полученными знаниями и умениями при решении практической задачи. Проверка результата запуска программного решения на языке Python.
Показатели и критерии оценивания	строка не менее 10 символов	Оценка "5" ставится, если верно выполнены все пять заданий промежуточной аттестации. Оценка "4" ставится, если верно выполнены четыре задания промежуточной аттестации. Оценка "3" ставится, если верно выполнены три задания промежуточной аттестации. Оценка "2" ставится, если верно выполнены два и менее задания промежуточной аттестации.
Примеры заданий	строка не менее 10 символов	Выполните следующие 5 заданий по предложенному датасету: Задание 1. Загрузить данные из файла (Датасет предоставляет выдает преподаватель) Задание 2. Заполнить пропуски и отсортировать данные по заданию (Вариант задания выдает преподаватель) Задание 3. Разделить их на 2 части: обучающую и тестовую выборки (Вариант задания выдает преподаватель) Задание 4. Сделать предсказание по тестовой выборке по заданию (Вариант задания выдает преподаватель) Задание 5. Сравнить и оценить реальные значения с предсказанием по заданию (Вариант задания выдает преподаватель)

Шкала оценивания, нижнее значение	строка не менее 10 символов	"2"
Шкала оценивания, верхнее значение	строка не менее 10 символов	"5"
Шкала оценивания, минимальный проходной балл	строка не менее 10 символов	"3"

## Преподаватели

ФИО	Наименование основного места работы	Должность	Высшее образование или среднее профессиональное образование по направлению «Образование и педагогические науки»	Высшее образование или среднее профессиональное образование по иному направлению соответствующим направленно сти ДОП	Ссылка на веб-страницы с портфолио	Информация о курсах повышения квалификации по профилю преподаваемой дисциплины (за последние 3 года)	Пройдена промежуточная аттестация не менее чем за два года обучения по образовательным программам высшего образования по специальностям и направлениям подготовки, соответствующим направленности ДОП	Отметка о полученном согласии на обработку персональных данных
строка от 2 до 100 символов	строка от 2 до 255 символов.	строка от 2 до 255 символов	да/нет	да/нет	строка		да/нет	да/нет
Строев Сергей Павлович	ФГБОУ ВО «Орловский государственный университет имени И.С. Тургенева»	заведующий кафедрой алгебры и математических методов в экономике	нет	да	<a href="https://docs.google.com/document/d/1DFJZV5S4STXgH_JTT8yZ8Ch9UM5oBXme/edit?usp=sharing&amp;ouid=118411873647758244068&amp;rtpof=true&amp;sd=true">https://docs.google.com/document/d/1DFJZV5S4STXgH_JTT8yZ8Ch9UM5oBXme/edit?usp=sharing&amp;ouid=118411873647758244068&amp;rtpof=true&amp;sd=true</a>	Профессиональная переподготовка "Машинное обучение для преподавателей вузов " НИУ "Высшая школа экономики"; Повышение квалификации "Практико-ориентированные подходы в преподавании профильных ИТ дисциплин" Университет Иннополис; Профессиональная переподготовка "Аналитик данных" АНО Дпо "ШАД"; Повышение квалификации "Искусственный интеллект" Санкт-Петербургский государственный университет; Повышение квалификации "Организация	да	да

						образовательных программ в области математики, компьютерных наук и цифровых технологий" НИУ "Высшая школа экономики";		
Астапов Павел Евгеньевич	Компьютерная академия "ТОР"	преподаватель, заведующий кафедры «разработки программного обеспечения»	нет	да	<a href="https://docs.google.com/document/d/1IcA01BmI8bDymvtTT1av2Fa6jdHd0jq9/edit?usp=sharing&amp;ouid=118411873647758244068&amp;rtpof=true&amp;sd=true">https://docs.google.com/document/d/1IcA01BmI8bDymvtTT1av2Fa6jdHd0jq9/edit?usp=sharing&amp;ouid=118411873647758244068&amp;rtpof=true&amp;sd=true</a>		да	да

Рабочая программа с описанием каждого модуля

Модуль (описание)	Тема	Содержание	Вид учебных занятий	Объем в ак.ч.
<b>Модуль 1.</b> <b>Применение</b> <b>основных</b> <b>конструкций Python</b> <b>для задач анализа</b> <b>данных</b> Базовые структуры и функции на языке программирования Python	<b>Тема 1.</b> Задачи анализа данных. Настройка IDE.	Введение в анализ данных. Причины выбора языка программирования Python для анализа данных. Среды программирования Google Colab, Jupyter Notebook. Типы данных. Условные операторы if, else, elif.	теоретические занятия	1
		Знакомство с Google Colab. Режимы работы в Google Colab. Решение практических задач по обработки данных на языке программирования Python. Работа с условными конструкциями.	практические занятия	2
		Работа по закреплению материала. Упражнения по использованию тернарного условного оператора. Написать программу на Python, которая запрашивает у пользователя два числа и выводит на экран, являются ли они взаимно простыми.	самостоятельная работа	2
	<b>Тема 2.</b> Решение задач на основные конструкции	Фильтрация наборов данных с помощью циклической конструкции. Группировка наборов данных с помощью циклов. Сортировка наборов данных с помощью циклов. Поиск данных с помощью циклов.	теоретические занятия	1
		Решение практических задач на исполнение алгоритмов с циклами. Написать программу для подсчета количества уникальных слов в заданном тексте. Написать программу для нахождения количества слов в заданном файле, которые имеют длину больше 5 символов. Классические алгоритмы перебора данных с использованием цикла.	практические занятия	3

		Работа по закреплению материала. Упражнения по использованию циклических конструкции при решении задач на языке Python.	самостоятельная работа	2
<b>Тема 3.</b> Структуры данных		Изучение основных структур данных в Python.  Использование структур данных для обработки больших объемов данных. Создание собственных структур данных на основе существующих. Оптимизация работы со структурами данных для повышения производительности. Удаление дубликатов данных.	теоретические занятия	1
		Решение практических задач по методам хранения и представления данных. Создать структуру данных "дерево" для хранения информации об объектах. Написать функции для добавления и удаления объектов, а также для поиска объектов по ключу. Написать программу, которая использует структуры данных для решения задач обработки данных.	практические занятия	4
		Работа по закреплению материала. Упражнения по применению методов работы со строками при решении задач на языке Python.  Упражнения по использованию методов работы со списками при решении задач на языке Python.	самостоятельная работа	3
	<b>Тема 4.</b> Применение функционального программирования для		Lambda функции в Python для работы с данными. Map и filter функции для обработки данных Reduce функции для агрегирования данных. Инструкции def, return.	теоретические занятия
		Решение практических задач по применению функций. Создать функцию для сортировки слов по длине в	практические занятия	3

	обработки данных	заданном списке строк. Написать программу для вывода на экран всех слов, которые встречаются более одного раза в заданном тексте.		
		Работа по закреплению материала. Упражнения по применению функций, подключению библиотек при решении задач на языке Python.	самостоятельная работа	2
	<b>Тема 5.</b> Классы и методы в задачах анализа данных	Создание классов в Python для представления объектов в анализе данных. Методы класса в Python, их использование для работы с объектами. Наследование классов в Python, создание иерархии классов. Полиморфизм в Python, реализация разных методов для разных типов объектов. Абстракция в Python, использование абстрактных классов и интерфейсов.	теоретические занятия	1
		Решение практических задач по работе с классами и методами. Написать класс с наборами атрибутов. Написать методы класса, которые обрабатывают данные.	практические занятия	5
		Работа по закреплению материала.  Создать класс "Магазин", который будет объединять информацию о пользователях, заказах и товарах. Написать метод "найти пользователя", который будет искать пользователя по его имени и возвращать его заказы.	самостоятельная работа	3
	<b>Тема 6.</b> Отработка навыков решения задач по модулю.	Отработка навыков решения задач на языке Python по материалам модуля 1.	практические занятия	1

	Промежуточная аттестация - выполнение тестирования с практическими заданиями по модулю 1.	промежуточная аттестация	1	
			<b>Объем в ак.ч.</b>	<b>Объем в %</b>
<b>ИТОГО:</b>			теоретические занятия	5
			практические занятия	18   44%
			самостоятельная работа	12   33%
			аттестация	1
			<b>Всего:</b>	<b>36</b>

Модуль (описание)	Тема	Содержание	Вид учебных занятий	Объем в ак.ч.
<b>Модуль 2. Big Data: Основы работы с большими массивами данных</b> Работа с библиотеками Pandas, NumPy. Получение данных через API	<b>Тема 1.</b> Введение в Big Data. Основы работы с большими массивами данных	Введение в Big Data. Цель анализа данных. Источники данных. Типы данных. Принципы работы с данными. Этапы работы с данными.	теоретические занятия	1
		Работа по закреплению материала. Упражнения по работе по определению типов данных разных источников.	самостоятельная работа	2
	<b>Тема 2.</b> Введение в Pandas. Обработка файлов данных	Библиотеки NumPy, Pandas. Применение библиотеки pandas для анализа данных. Series. DataFrame. Работа с внешними источниками данных. Операции над данными. Сортировка. Группировка. Поиск данных.	теоретические занятия	1
		Решение практических задач по получению данных из файла. Манипуляции с данными с помощью библиотеки pandas.	практические занятия	5
		Работа по закреплению материала. Упражнения по работе с структурами данных Series и DataFrame	самостоятельная работа	3
		Предобработка данных.	теоретические занятия	1

	<b>Тема 3.</b> Дескриптивный анализ в Pandas. Очистка данных	Инструменты для работы с типами. Очистка данных. Поиск дубликатов. Отсутствующие данные. Замена отсутствующих данных. Категоризация в данных		
		Решение практических задач по работе с преобразованием типов данных, проверки дубликатов, работа с пропусками в данных. Работа с категориальными признаками.	практические занятия	4
		Работа по закреплению материала. Упражнения по работе с категориальными признаками	самостоятельная работа	3
	<b>Тема 4.</b> Получение данных через API. Формат JSON..	Сбор данных через API. Распространенные форматы текстовых файлов: JSON. Понятие исключения, обработка исключений. Собственные исключения.	теоретические занятия	1
		Решение практических задач по сбору данных через API. Разбор формата JSON.	практические занятия	3
		Работа по закреплению материала. Упражнения по работе с API на языке Python	самостоятельная работа	2
	<b>Тема 5.</b> Веб-скрапинг на Python. Библиотека BeautifulSoup	Библиотека Requests. Углубленный анализ HTML кода. HTML. Атрибуты http-запросы, html-страницы. Метод requests.get(). Статус коды. Получаем содержимое response объекта. Парсинг AJAX. Веб-краулеры. Модели веб-краулинга. Сбор данных из форм Скрапинг данных JS.	теоретические занятия	1
Решение практических задач по работе с библиотеками requests и BeautifulSoup. Работа с протоколом HTTP. Углубленный анализ HTML кода. Поиск элементов на странице. Поиск узлов и элементов.		практические занятия	5	

		Пагинация. Скрапинг табличных данных. Сохранение результата в Excel.			
		Работа по закреплению материала. Упражнения по работе по анализу HTML кода на языке Python	самостоятельная работа	2	
	<b>Тема 6.</b> Отработка навыков решения задач.	Отработка навыков решения задач на языке Python по материалам модуля 2.	практические занятия	1	
		Промежуточная аттестация - выполнение тестирования с практическими заданиями по модулю 2.	промежуточная аттестация	1	
				<b>Объем в ак.ч.</b>	<b>Объем в %</b>
<b>ИТОГО:</b>			теоретические занятия	5	
			практические занятия	18	50%
			самостоятельная работа	12	36%
			аттестация	1	
			<b>Всего:</b>	<b>36</b>	

<b>Модуль (описание)</b>	<b>Тема</b>	<b>Содержание</b>	<b>Вид учебных занятий</b>	<b>Объем в ак.ч.</b>
<b>Модуль 3.</b> <b>Exploratory Data Analysis</b> Анализ и визуализация данных	<b>Тема 1.</b> Работа с датами. Временные ряды	Библиотеки для работы с датой и временем. Типы временных рядов. Предсказания на основе временных рядов. Проектирование признаков для потоков событий	теоретические занятия	1
		Решение практических задач по работе с функциями конвертирования даты. Разбиение дат на несколько признаков. Работа с временными интервалами. Работа с временными метками.	практические занятия	2
		Работа по закреплению материала. Упражнения по работе с преобразованием даты и времени.	самостоятельная работа	2

	<b>Тема 2.</b> Хранение данных в БД	Базы данных. Архитектура "клиент/ сервер". Модели данных. Отношения. Ключи. Нормализация. Объекты и процессы баз данных. Выборка данных	теоретические занятия	1
		Решение практических задач по работе с созданием базы данных в DBeaver	практические занятия	2
		Работа по закреплению материала. Упражнения по работе с созданием базы данных в DBeaver	самостоятельная работа	2
	<b>Тема 3.</b> Простая работа с базами данных с ORM	SQL. Работа с сортировкой данных, группировкой данных. Функции агрегации . Объединение таблиц с использованием агрегатных функций	теоретические занятия	1
		Решение практических задач по работе с извлечением строк из таблицы. Условие. Поиск значения NULL. Объединение значений из столбцов. Ограничение на количество возвращаемых строк. Сортировка результата запроса. Работа с разными типами данных	практические занятия	4
		Работа по закреплению материала. Упражнения по работе с составлением SQL запросов к БД в DBeaver	самостоятельная работа	2
	<b>Тема 4.</b> Построение графиков. Практика по визуализации данных	Визуализация данных. Виды графиков. Библиотеки matplotlib, seaborn, plotly. Настройка внешнего вида диаграммы. Визуализация категориальных данных..	теоретические занятия	1
		Решение практических задач по работе с построением графиков с использованием библиотеки matplotlib. seaborn, plotly	практические занятия	5
		Работа по закреплению материала. Упражнения по работе с построением графиков на языке программирования Python	самостоятельная работа	3
	<b>Тема 5.</b> EDA. Поиск закономерностей в данных	Исследовательский анализ данных. Поиск закономерностей в данных. Выявление наиболее важных переменных для анализа Анализ графиков. Обнаружение выбросов и аномалий	теоретические занятия	1

		Решение практических задач по работе с выявлением выбросов в данных. Оценка корреляции между признаками. Парное сравнение признаков.	практические занятия	4	
		Работа по закреплению материала. Упражнения по работе с анализом больших объемов данных.	самостоятельная работа	3	
	<b>Тема 6.</b> Отработка навыков решения задач.	Отработка навыков решения задач на языке Python по материалам модуля 3.	практические занятия	1	
		Промежуточная аттестация - выполнение тестирования с практическими заданиями по модулю 3	промежуточная аттестация	1	
				<b>Объем в ак.ч.</b>	<b>Объем в %</b>
<b>ИТОГО:</b>			теоретические занятия	5	
			практические занятия	18	50%
			самостоятельная работа	12	33%
			аттестация	1	
			<b>Всего:</b>	<b>36</b>	

Модуль (описание)	Тема	Содержание	Вид учебных занятий	Объем в ак.ч.
<b>Модуль 4. Базовые алгоритмы машинного обучения в Sklearn</b> Алгоритмы машинного обучения с учителем и без учителя, метод NLP, построение рекомендательных систем	<b>Тема 1.</b> Введение в машинное обучение. Библиотеки для Data Science	Введение в машинное обучение. Моделирование. Поиск связи между входными данными и целевой переменной. Обучение с учителем и без. Виды моделей. Основы проектирования признаков. Переобучение. Оценка моделей. Отбор модели.	теоретические занятия	1
		Решение практических задач по работе с определением набора входных признаков, разделение набора данных на тестовую и обучающую выборку. Определение репрезентативности обучающей выборки.	практические занятия	2

		Работа по закреплению материала. Упражнения по работе с отбором модели из нескольких обучающих алгоритмов	самостоятельная работа	1
	<b>Тема 2.</b> KNN, деревья решений, Gradient Boosting	Классификация. Модель KNN. Деревья решений. Случайные леса. Энтропия.	теоретические занятия	1
		Решение практических задач по созданию дерева решений на языке Python. Тренировка классификационного дерева принятия решений. Визуализация модели дерева принятия решений. Управление размером дерева. Тренировка классификационного случайного леса. Построение модели KNN. Идентификация наилучшего размера окрестности. Отбор признаков, Оценка моделей.	практические занятия	4
		Работа по закреплению материала. Упражнения по работе с визуализацией модели дерева принятия решений.	самостоятельная работа	3
	<b>Тема 3.</b> Кластеризация	Кластеризация. Модель. Виды. Алгоритм k-means.	теоретические занятия	1
		Решение практических задач по работе с кластеризацией методом k-средних	практические занятия	4
		Работа по закреплению материала. Упражнения по работе с построением модели кластеризации методом DBSCAN	самостоятельная работа	2
	<b>Тема 4.</b> NLP. Мешок слов	Генерация базовых NLP-признаков и построение первого варианта модели. Модель “мешок слов”. Нормализация признаков, полученных из «мешка слов», алгоритмом tf-idf.	теоретические занятия	1
		Решение практических задач по работе с текстом: удаление знаков препинания, лемматизация текста, удаление стоп-слов, кодирование текста в качестве мешка слов.	практические занятия	4

		Работа по закреплению материала. Упражнения по работе с текстом	самостоятельная работа	2	
	<b>Тема 5.</b> Создание рекомендательной системы для киносайта	Рекомендательные системы на языке Python. Рекомендация популярных тем. Коллаборативная фильтрация по схожести пользователей, предметов	теоретические занятия	1	
		Решение практических задач по аналитической обработке данных. Генерация рекомендаций на языке Python.	практические занятия	2	
		Работа по закреплению материала. Упражнения по работе с рекомендательными системами.	самостоятельная работа	3	
	<b>Тема 6.</b> Отработка навыков решения задач.	Отработка навыков решения задач на языке Python по материалам модуля 4.	практические занятия	3	
		Промежуточная аттестация - выполнение практического задания по модулю 4	промежуточная аттестация	1	
				<b>Объем в ак.ч.</b>	<b>Объем в %</b>
<b>ИТОГО:</b>			теоретические занятия	5	
			практические занятия	19	53%
			самостоятельная работа	11	31%
			аттестация	1	
			<b>Всего:</b>	<b>36</b>	

### Календарно-тематическое планирование

№	Тема и № модуля	Тема занятия	Кол-во занятий*	Кол-во часов	Дата
1	Модуль 1. Применение основных конструкций Python для задач анализа данных	Тема 1. Задачи анализа данных. Настройка IDE.	2	5	02.10.23 05.10.23

2		Тема 2. Решение задач на основные конструкции	2	6	09.10.23 12.10.23
3		Тема 3. Структуры данных	3	8	16.10.23 19.10.23 23.10.23
4		Тема 4. Применение функционального программирования для обработки данных	2	6	26.10.23 30.10.23
5		Тема 5. Классы и методы в задачах анализа данных	3	9	02.11.23 06.11.23 09.11.23
6		Тема 6. Отработка навыков решения задач по модулю	1	1	13.11.23
	<b>Аттестация по модулю 1</b>			1	16.11.23
1	<b>Модуль 2. Big Data: Основы работы с большими массивами данных</b>	Тема 1. Введение в Big Data. Основы работы с большими массивами данных	1	3	21.11.23
2		Тема 2. Введение в Pandas. Обработка файлов данных	3	9	23.11.23 27.11.23 30.11.23
3		Тема 3. Дескриптивный анализ в Pandas. Очистка данных	3	8	04.12.23 07.12.23 11.12.23
4		Тема 4. Получение данных через API. Формат JSON	2	6	14.12.23 18.12.23
5		Тема 5. Веб-скрапинг на Python. Библиотека BeautifulSoup	3	8	09.01.24 11.01.24 15.01.24
6		Тема 6. Отработка навыков решения задач	1	1	17.01.24
	<b>Аттестация по модулю 2</b>			1	18.01.24

1	<b>Модуль 3. Exploratory Data Analysis</b>	Тема 1 Работа с датами. Временные ряды	2	5	22.01.24 25.01.24
2		Тема 2 Хранение данных в БД	2	5	29.01.24 01.02.24
3		Тема 3 Простая работа с базами данных с ORM	3	7	05.02.24 08.02.24 12.02.24
4		Тема 4 Построение графиков. Практика по визуализации данных	3	9	15.02.24 19.02.24 22.02.24
5		Тема 5 EDA. Поиск закономерностей в данных	3	8	26.02.24 29.02.24 04.03.24
6		Тема 6 Отработка навыков решения задач	1	1	11.03.24
<b>Аттестация по модулю 3</b>				1	14.03.24
1	<b>Модуль 4. Базовые алгоритмы машинного обучения в Sklearn</b>	Тема 1 Введение в машинное обучение. Библиотеки для Data Science	2	4	25.03.24 28.03.24
2		Тема 2 KNN, деревья решений, Gradient Boosting	3	8	01.04.24 04.04.24 08.04.24
3		Тема 3 Кластеризация	3	7	11.04.24 15.04.24 18.04.24
4		Тема 4 NLP. Мешок слов	3	7	22.04.24 25.04.24 29.04.24
5		Тема 5 Создание рекомендательной системы для киносайта	2	6	06.05.24 08.05.24

6		Тема 6 Отработка навыков решения задач	2	3	13.05.24 16.05.24
	Аттестация по модулю 4			1	20.05.24

\*количество занятий не включают часы, отведенные на самостоятельное изучение, и часы, отведенные на прохождение аттестации

#### Учебно-методические материалы

Наименование поля	Допустимые значения полей	Значение полей	Значение полей	Значение полей	Значение полей
Порядковый номер модуля	строка не менее 10 символов	1	2	3	4
Методы, формы и технологии	строка не менее 10 символов	<p>Лекционные занятия реализуются на базе инструментов видеоконференцсвязи и включают в себя установочные интерактивные лекции, с обязательным использованием инструмента обратной связи. Для проведения онлайн занятий применяется формат вебинаров. В качестве площадок для совместной синхронной работы будут использованы виртуальные доски и Google-сервисы.</p> <p>Практические занятия реализуется в режиме реального времени в формате индивидуальной или групповой работы обучающихся. Преподаватель разбивает задание на несколько частей и дает обучающимся по 10–15 минут на выполнение каждой части (либо это могут быть несколько отдельных заданий). Обучающиеся в течение практики самостоятельно выполняют задания. После завершения времени на выполнение каждого задания происходит его разбор и совместное обсуждение в эфире полученных результатов. Преподаватель может выборочно попросить слушателей продемонстрировать свой экран с программным кодом либо прокомментировать результат в микрофон. Для организации групповой работы необходимо предварительно разбить обучающихся на подгруппы и распределить их по сессионным залам.</p> <p>Изучение теоретического материала (СРС) предполагается до и после онлайн занятия.</p>			
Методические разработки	строка не менее 10 символов	<p>Методические рекомендации и пособия по изучению программы. Программа реализуется в формате онлайн обучения, с применением активных технологий обучения и общения с аудиторией в электронной среде. Обучение в очной части реализуется в виде лекционных и практических занятий, в</p>			

		которых применяются интерактивные формы обучения, семинары, мастер–классы, активные и ситуативные методы обучения, решения кейсов и обсуждения.
Материалы модуля	строка не менее 10 символов	По данной программе имеется электронный учебно-методический комплекс в LMS Odin. УМК содержит: систему навигации по программе (учебно-тематический план, интерактивный график работы по программе, сведения о результатах обучения, о преподавателях программы, чат для объявлений и вопросов преподавателям), набор видеолекций, презентации к лекциям, набор ссылок на внешние образовательные ресурсы и инструменты, систему заданий с подробных установок и инструкциями, списки основной и дополнительной литературы. В электронном курсе реализована система обратной связи, а также онлайн–площадки для взаимного обучения.
Учебная литература	строка не менее 10 символов	<ol style="list-style-type: none"> <li>1. Доусен М. Прографируем на Python / М. Доусен - СПб.: Питер, 2016. - 416 с.</li> <li>2. Банкрашков, А.В. Програмирование для детей на языке Python / А.В. Банкрашков. - М.: АСТ, 2018. - 288 с.</li> <li>3. Райан Митчелл, Современный скрапинг веб-сайтов с помощью Python. 2-е межд. издание . — СПб.: Питер, 2021.</li> <li>4. Грофф Дж., Вайнберг П., Оппель Э. SQL. Полное руководство. 3-е изд. М. : Диалектика, 2019. 960 с.</li> <li>5. Силен Дэви, Мейсман Арно, Али Мохамед Основы Data Science и Big Data. Python и наука о данных. — СПб.: Питер, 2017. — 336 с.</li> <li>6. Скиена, Стивен С. Наука о данных: учебный курс. : Пер. с англ. - СПб. : ООО "Диалектика", 2020. - 544 с.</li> <li>7. Плас Дж., Вандер. Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 576 с.</li> <li>8. Марц, Натан. Большие данные: принципы и практика построения масштабируемых систем обработки данных в реальном времени / Натан Марц, Джеймс Уоррен ; [пер. с англ. и ред. И. В. Берштейна]. - Москва [и др.] : Вильямс, 2017. - 368 с.</li> <li>9. Свейгарт, Эл. Автоматизация рутинных задач с помощью Python: практическое руководство для начинающих. Пер. с англ. — М.: Вильямс, 2016. — 592 с.</li> </ol>

### Материально-технические условия реализации программы

Наименование поля	Допустимые значения полей	Значение полей	Значение полей	Значение полей	Значение полей
Порядковый номер модуля	строка не менее 10 символов	1	2	3	4
Наименование требуемого оборудования	строка не менее 2 символов	Manufacturer - Acer Model - Veriton M6660G Size - 17x42x37 cm Extra details - CPU i7 8700 / RAM 32 GB DDR4 2400 GHz / HDD 1Tb / SSD 256 / nVidia GeForce GTX1050Ti GPU 4 GB / Win10 и выше мышь, клавиатура, монитор			
Наименование требуемого программного обеспечения	строка не менее 2 символов	ОС - MS Windows 10 и выше Браузер			
Электронные информационные ресурсы	строка не менее 10 символов	<ol style="list-style-type: none"> <li>1. <a href="https://www.python.org/">https://www.python.org/</a></li> <li>2. <a href="https://pythonworld.ru/samouchitel-python">https://pythonworld.ru/samouchitel-python</a></li> <li>3. <a href="https://pythoner.name/">https://pythoner.name/</a></li> <li>4. <a href="https://colab.research.google.com/">https://colab.research.google.com/</a></li> </ol>			
Электронные образовательные ресурсы	строка не менее 10 символов	<ol style="list-style-type: none"> <li>1. Тренажер по Python <a href="https://www.w3schools.com/python/default.asp">https://www.w3schools.com/python/default.asp</a></li> <li>2. Тренажер по Python <a href="https://pythontutor.ru/">https://pythontutor.ru/</a></li> <li>3. Тренажер по SQL <a href="http://sql-tutorial.ru/ru/book_database_ships.html">http://sql-tutorial.ru/ru/book_database_ships.html</a></li> </ol>			



