

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Методические указания к практическим занятиям по дисциплине «Анализ данных на Python»

для студентов направления подготовки

02.03.01 «Математика и компьютерные науки»

направленность (профиль)

«Анализ данных и искусственный интеллект»

Введение

Практические занятия - это целенаправленная форма организации педагогического процесса, направленная на углубление научно-теоретических знаний и овладение определенными методами работы, в процессе которых вырабатываются умения и навыки выполнения тех или иных учебных действий в данной сфере науки.

Практические занятия предназначены для углубленного изучения учебных дисциплин и играют важную роль в выработке у студентов умений и навыков применения полученных знаний для решения практических задач совместно с педагогом. Кроме того, они развивают научное мышление и речь, позволяют проверить знания студентов и выступают как средства оперативной обратной связи.

Цель практических занятий - углублять, расширять, детализировать знания, полученные на лекции, в обобщенной форме и содействовать выработке навыков профессиональной деятельности.

Данные методические указания к практическим занятиям разработаны в соответствии с рабочей программой по дисциплине **«Анализ данных на Python»** для студентов направления подготовки 02.03.01 «Математика и компьютерные науки».

Практическое занятие 1

Тема: «Ввод информации».

Цель: познакомиться с работой функции Input().

Краткая теоретическая справка.

Для ввода данных Python 3 предназначена функция input().

Функция input() принимает **текст**, который пользователь вводит с клавиатуры.

Обратите внимание: **input() принимает данные, которые мы вводим с клавиатуры, как текст!**

Задание №1. Запустите IDLEPython 3. Перейдите в режим инструкций. Откройте **File-NewFile**. Откроется новое окно (Режим сценариев). Наберите текст, который выделен.

Комментарий не набирать:

```
# Это комментарий его не нужно набирать.
```

```
Name= input('Привет. Как тебя зовут?')
```

```
# Name ,будет указывать на информацию, которую мы сейчас введем с клавиатуры.
```

```
# текст -(‘Привет. Как тебя зовут?’)- это сообщение которое выводится при выполнении.
```

```
print(“Привет ”, name)
```

Сохраните файл: File – Save as. (Ctrl+Shift+S). Запустите программу: RUN-Run Module (F5).

Задание №2. При выполнении функции input, в предыдущем примере, можно видеть, что функция input() задала вопрос пользователю и стала ожидать, пока пользователь что-нибудь введет.

После нажатия Enter все, что ввел пользователь, функция input возвратит как строку в точности.

А переменная name, будет указывать на эту строку.

Наберем следующую программу. Предыдущий текст можно удалить:

```
nm=input('Введите число < 10=')
```

```
print('Ответ ',nm+nm)
```

Сохраните файл: File – Save as. (Ctrl+Shift+S). Запустите программу: RUN-Run Module (F5).

Если вы ожидали, что программа выдаст сумму вашего числа, то вы ошиблись. Все дело в том, что:

input() принимает данные, которые мы вводим с клавиатуры, как текст!

Поэтому nm+nm- это не сложение, а сцепление, склеивание, двух строк.

Если мы хотим после ввода получить число, то нужно преобразовать введенную строку в число.

Это возможно при использовании функции int().

Функция int() преобразует строковое выражение в целое число.

Можно так: int(input('Введите число='))

Задание 3:

Наберите текст:

```
nm=int(input('Введите число='))
```

```
print("Результат сложения=", nm+nm)
```

```
input()
```

```
# Допустим введем 3
```

```
# Получим: Результат сложения=6
```

```
#Ожидаем нажатия Enter.
```

Сохраните файл: File – Saveas. (Ctrl+Shift+S). Запустите программу: RUN-Run Module (F5).

Задание 4.

Найдем среднеарифметическое двух чисел a , b .

	мента х в список
Spis.sort()	Сортирует элементы списка по возрастанию
Spis.reverse()	Обращает порядок следования элементов.

Ход работы.

Задание 1. Исправьте ошибки.

```
sp1=['March', 'April', 'May']
sp2=[Spring, Summer, Winter]
print('sp1=', sp1) #вывод списка
print('sp1 elements=', len(sp1)) # длина списка sp1
sp1.append('June') #Добавляем элемент June в конец списка sp1
print("Элемент добавлен нажмите Enter")
input()
print("Расширяем список")
sp1.extend(sp2) # добавляем элементы второго списка в конец sp1.
print("список обновлен\n", sp1)
print('sp1 elements=', len(sp1)) # длина списка sp1
print("Удаляем пятый элемент. Нажмите Enter")
Input()
delsp1(4)
print("Удаление завершено\n", sp1)
print("Удаляем последний элемент\n")
sp1.pop()
print("All list sp1\n", sp1)
```

Задание на повторение:

Создайте список дней недели и список дел на эту неделю.

Напишите инструкции:

1. для добавления элемента в список
2. отображения всех элементов
3. удаления последнего и повторного вывода всех элементов.
4. добавьте инструкции для расширения первого списка элементами второго
5. вывода всех элементов первого списка
6. затем удаление элементов
7. вывод элементов первого списка.

Практическое занятие 3

Тема: Работа с текстовыми файлами

Цель работы: закрепление знаний, умений и навыков по работе с файлами в Python.

Все задания требуется выполнить с использованием языка Python. В качестве справочного материала, рекомендуется использовать лекционный материал и источники [1, 2]

Задание 1

Дан текст на языке племени Мумба-Юмба. Выведите все слова, встречающиеся в тексте, разделяя их пробелом. Слова должны быть отсортированы по убыванию их количества появления в тексте, а при одинаковой частоте появления в алфавитном порядке.

Подсказка. После того, как вы создадите словарь всех слов, вам захочется отсортировать его по частоте встречаемости слова. Желаемого можно добиться, если создать список, элементами которого будут кортежи из двух элементов: частота встречаемости слова и само слово. Например, [(2, 'hi'), (1, 'what'), (3, 'is')]. Тогда стандартная сортировка будет сортировать список кортежей, при этом кортежи сравниваются по первому элементу, а если они равны – то по второму. Это почти то, что требуется в задаче, а чтобы сделать то что нужно – вспомните про параметр `key` в сортировке.

Задание 2

Дан файл, каждая строка которого может содержать одно или несколько целых чисел, разделенных одним или несколькими пробелами.

Вычислите сумму чисел в каждой строке и выведите эти суммы через пробел (для каждой строки выводится сумма чисел в этой строке).

Задание 3

Выведите все строки данного входного файла в обратном порядке. Для этого считайте список всех строк при помощи метода `readlines()`.

Последняя строка входного файла заканчивается символом `\n`.

Задание 4

Дан файл. Определите сколько в нем букв (латинского алфавита), слов, строк. Выведите три найденных числа в формате, приведенном в примере. Словом считается последовательность больших и маленьких латинских букв (для проверки того, состоит ли строка только из латинских букв удобно пользоваться методом `isalpha()`). Все остальные символы считаются разделителями слов.

Задание 5

В качестве ответа введите все строки наибольшей длины из входного файла, не меняя их порядок.

В данной задаче удобно считать список строк входного файла целиком при помощи метода `readlines()`.

Задание 6

Выведите в обратном порядке содержимое всего файла полностью. Для этого считайте файл целиком при помощи метода `read()`.

Содержание отчета

В качестве отчета прикладываются скрипты выполнения заданий и выходные файлы с результатами работы

Практическое занятие 4

Тема: Работа с csv-файлами

Цель работы: закрепление знаний, умений и навыков по работе с csv-файлами в Python.

Все задания требуется выполнить с использованием языка Python. В качестве справочного материала, рекомендуется использовать лекционный материал и источники [3]

Задание 1

В csv-файле (разделитель – точка с запятой, кавычки не используются) содержится анонимизированная информация о зарплатах сотрудников в различных компаниях. В первом столбце записано название компании, а во втором - зарплата. Упорядочите компании по возрастанию средней зарплаты. В случае одинаковых средних зарплат в нескольких компаниях их следует упорядочить по алфавиту.

Названия компаний следует выводить по одному в строке.

Совет: используйте словарь, где ключом будет служить название компании, а в качестве значения будет храниться суммарная зарплата и количество сотрудников.

Задание 2

В csv-файле (разделитель – точка с запятой, кавычки не используются) содержится анонимизированная информация о зарплатах сотрудников в различных компаниях. В первом столбце записано название компании, а во втором – зарплата.

Поменяйте местами первый и второй столбцы, по-прежнему разделяя значения в ячейках одной строки точкой с запятой. Сохраняйте порядок строк.

Задание 3

Вася составил таблицу с ценами на продукты в разных магазинах. В первой строке таблицы (кроме первой ячейки) записаны названия продуктов. Во всех строках, начиная со второй, записана информация о ценах в магазине. В первой ячейке написано название магазина, а в ячейках, начиная со второй - цена на товар, название которого записано в первой строке соответствующего столбца.

Таблица задана как csv-файл, разделителем ячеек выступает точка с запятой, а строковые константы не окружаются кавычками.

Вася очень хочет поесть, но денег у него мало. Поэтому помогите ему определить цену самого дешевого продукта. В качестве ответа введите одно число.

Задание 4

Вася составил таблицу с ценами на продукты в разных магазинах. В первой строке таблицы (кроме первой ячейки) записаны названия продуктов. Во всех строках, начиная со второй, записана информация о ценах в магазине. В первой ячейке написано название магазина, а в ячейках, начиная со второй - цена на товар, название которого записано в первой строке соответствующего столбца.

Таблица задана как csv-файл, разделителем ячеек выступает точка с запятой, а строковые константы не окружаются кавычками.

Вася очень хочет поесть, но денег у него мало. Поэтому помогите ему определить самый дешевый продукт и в каком магазине он продается. Название продукта следует записать в первой строке, а название магазина - во второй. Если несколько товаров стоят одинаково, то выведите то название, которое раньше в алфавитном порядке. Если этот товар продается в нескольких магазинах по одной минимальной цене, то выведите минимальное в алфавитном порядке название магазина.

Содержание отчета

В качестве отчета прикладываются скрипты выполнения заданий и выходные файлы с результатами работы

Практическое занятие 5

Тема: Работа с текстовыми xlsx-файлами

Цель работы: закрепление знаний, умений и навыков по работе с файлами Excel в Python.

Все задания требуется выполнить с использованием языка Python. В качестве справочного материала, рекомендуется использовать лекционный материал и источники [5-7]

Задание 1

Главный бухгалтер компании "Рога и копыта" случайно удалил ведомость с начисленной зарплатой. К счастью, у него сохранились расчётные листки всех сотрудников. Помогите по этим расчётным листкам восстановить зарплатную ведомость. Архив с расчётными листками доступен по ссылке

<https://stepik.org/media/attachments/lesson/245299/rogaikopyta.zip> (вы можете скачать и распаковать его вручную или самостоятельно научиться делать это с помощью скрипта на Питоне).

Ведомость должна содержать 1000 строк, в каждой строке должно быть указано ФИО сотрудника и, через пробел, его зарплата. Сотрудники должны быть упорядочены по алфавиту.

Задание 2

Вася планирует карьеру и переезд. Для это составил таблицу, в которой для каждого региона записал зарплаты для разных интересных ему профессий. Таблица доступна по ссылке <https://stepik.org/media/attachments/lesson/245267/salaries.xlsx>. Выведите название региона с самой высокой медианной зарплатой (медианой называется элемент, стоящий в середине массива после его упорядочивания) и, через пробел, название профессии с самой высокой средней зарплатой по всем регионам.

Задание 3

Васю назначили завхозом в туристической группе и он подошёл к подготовке ответственно, составив справочник продуктов с указанием калорийности на 100 грамм, а также содержание белков, жиров и углеводов на 100 грамм продукта. Ему не удалось найти всю информацию, поэтому некоторые ячейки остались незаполненными (можно считать их значение равным нулю). Также он использовал какой-то странный офисный пакет и разделял целую и дробную часть чисел запятой. Таблица доступа по ссылке <https://stepik.org/media/attachments/lesson/245290/trekking1.xlsx>

Вася хочет минимизировать вес продуктов и для этого брать самые калорийные продукты. Помогите ему и упорядочите продукты по убыванию калорийности. В случае, если про-

дукты имеют одинаковую калорийность - упорядочите их по названию. В качестве ответа необходимо сдать названия продуктов, по одному в строке.

Задание 4

Васю назначили завхозом в туристической группе и он подошёл к подготовке ответственно, составив справочник продуктов с указанием калорийности на 100 грамм, а также содержание белков, жиров и углеводов на 100 грамм продукта. Ему не удалось найти всю информацию, поэтому некоторые ячейки остались незаполненными (можно считать их значение равным нулю). Также он использовал какой-то странный офисный пакет и разделял целую и дробную часть чисел запятой. Таблица доступна по ссылке <https://stepik.org/media/attachments/lesson/245290/trekking2.xlsx>

Вася составил раскладку по продуктам на один день (она на листе "Раскладка") с указанием названия продукта и его количества в граммах. Посчитайте 4 числа: суммарную калорийность и граммы белков, жиров и углеводов. Числа округлите до целых вниз и введите через пробел.

Содержание отчета

В качестве отчета прикладываются скрипты выполнения заданий и выходные файлы с результатами работы

Практическое занятие 6

Тема: Работа с html страницами 1

Цель работы: закрепление знаний, умений и навыков по работе с html страницами в Python.

Все задания требуется выполнить с использованием языка Python. В качестве справочного материала, рекомендуется использовать лекционный материал и источники [8-11]

Задание 1

Мы сохранили страницу с википедии про языки программирования и сохранили по адресу <https://stepik.org/media/attachments/lesson/209717/1.html>

Скачайте её с помощью скрипта на Питоне и посчитайте, какой язык упоминается чаще Python или C++ (ответ должен быть одной из этих двух строк). Необходимо просто подсчитать количество вхождений слова Python или C++ как подстроки.

Задание 2

Мы сохранили статью с википедии, она доступна по ссылке

<https://stepik.org/media/attachments/lesson/258939/webpage.html>

Вам необходимо обработать ее с помощью BeautifulSoup и вывести все ссылки, которые есть на этой странице, в том порядке как они встречались по одной в строке.

Под ссылкой понимается содержимое атрибута href тега a.

Вам могут быть полезны методы find_all для супа (он позволяет найти все теги на странице), метод has_attr для тега (проверяет есть ли такой атрибут у тега) и доступ к атрибуту тега по аналогии со словарем.

Практическое занятие 7

Тема: Работа с html страницами 2

Задание 1

Мы сохранили статью с википедии, она доступна по ссылке <https://stepik.org/media/attachments/lesson/258939/webpage.html>

Вам необходимо обработать ее с помощью BeautifulSoup и вывести все внешние ссылки, которые есть на этой странице, в том порядке как они встречались по одной в строке.

Под ссылкой понимается содержимое атрибута href тега a. Ссылка называется внешней, если она ведет на другой сайт (т.е. начинается с http:// или https://).

Вам могут быть полезны методы find_all для супа (он позволяет найти все теги на странице), метод has_attr для тега (проверяет есть ли такой атрибут у тега) и доступ к атрибуту тега по аналогии со словарем.

Задание 2

Файл <https://stepik.org/media/attachments/lesson/209719/2.html> содержит статью с Википедии про язык Python. В этой статье есть теги code, которыми выделяются конструкции на языке Python. Вам нужно найти все строки, содержащиеся между тегами <code> и </code> и найти те строки, которые встречаются чаще всего и вывести их в алфавитном порядке, разделяя пробелами.

Например, если исходный текст страницы выглядел бы так:

```
<code class="language-language-html"><code>a</code>
<a>bracadabr</a>
<code>c</code>
<code>b</code>
<code>b</code>
<code>c</code>
```

то в ответ надо было бы ввести строку "b c".

Практическое занятие 8

Тема: Работа с html страницами 3

Задание 1

Мы сохранили статью с википедии, она доступна по ссылке <https://stepik.org/media/attachments/lesson/258943/Moscow.html>

Вам необходимо обработать ее с помощью BeautifulSoup и подсчитать все внутренние ссылки, которые не содержат в себе двоеточия (не являются ссылкой на техническую статью в википедии) и не начинаются с символа #.

Под ссылкой понимается содержимое атрибута href тега a. Ссылка называется внешней, если она ведет на другой сайт (т.е. начинается с http:// или https://). Все остальные ссылки являются внутренними.

Вам могут быть полезны методы find_all для супа (он позволяет найти все теги на странице), метод has_attr для тега (проверяет есть ли такой атрибут у тега) и доступ к атрибуту тега по аналогии со словарем.

В качестве ответа выведите количество внутренних ссылок, удовлетворяющих условию.

Задание 2

В этой задаче достаточно обернуть в функцию то, что делает предыдущая и вызвать подсчет числа внутренних ссылок для двух статей:

<https://stepik.org/media/attachments/lesson/258943/Moscow.html>

и <https://stepik.org/media/attachments/lesson/258944/New-York.html>

В качестве ответа на задачу введите два числа через пробел.

Практическое занятие 9

Тема: Работа с html страницами 4

Задание 1

В этой задаче достаточно вам необходимо посчитать количество внутренних ссылок, которые есть в обеих статьях: <https://stepik.org/media/attachments/lesson/258943/Moscow.html> и <https://stepik.org/media/attachments/lesson/258944/New-York.html>

Обратите внимание, что если ссылка встречается в статье несколько раз, то учитывать ее нужно лишь однажды, т.е. необходимо найти количество различных страниц википедии, на которых есть ссылка как из первой, так и из второй статьи.

Задание 2

В этой задаче вам необходимо найти все внутренние ссылки, которые есть в обеих статьях: <https://stepik.org/media/attachments/lesson/258943/Moscow.html> и <https://stepik.org/media/attachments/lesson/258944/New-York.html> и вывести их в алфавитном порядке по одной в строке.

Обратите внимание, что если ссылка встречается в статье несколько раз, то учитывать ее нужно лишь однажды.

Содержание отчета

В качестве отчета прикладываются скрипты выполнения заданий и выходные файлы с результатами работы

Практическое занятие 10

Тема: XML и обработка геоданных 1

Цель работы: закрепление знаний, умений и навыков по работе с xml и обработкой геоданных в Python.

Все задания требуется выполнить с использованием языка Python. В качестве справочного материала, рекомендуется использовать лекционный материал и источники [8-11]

Задание 1

В OpenStreetMap XML встречаются теги `node`, которые соответствуют некоторым точкам на карте. Ноды могут не только обозначать какой-то точечный объект, но и входить в состав `way` (некоторой линии, возможно замкнутой) и не иметь собственных тегов. Для доступного по ссылке <https://stepik.org/media/attachments/lesson/245681/map2.osm> фрагмента карты посчитайте, сколько всего тегов `node` не содержат в себе ни одного тега `tag` (первое число в ответе), а сколько содержит хотя бы один тег `tag` (второе число в ответе). Числа введите через пробел.

Содержание отчета

В качестве отчета прикладываются скрипты выполнения заданий и выходные файлы с результатами работы

Задание 2

В OpenStreetMap XML встречаются теги way, которые соответствуют некоторым линиям и многоугольникам на карте. Way состоит из списка нодов (точек), которые задаются тегами nd вложенными в тег way. Для доступного по ссылке <https://stepik.org/media/attachments/lesson/245681/map2.osm> определите id того way, который содержит в себе наибольшее количество нодов. Если таких несколько - выведите тот id, который встречается в файле раньше.

Содержание отчета

В качестве отчета прикладываются скрипты выполнения заданий и выходные файлы с результатами работы

Задание 3

Вася решил открыть АЗС (заправку). Чтобы оценить уровень конкуренции он хочет изучить количество заправок в интересующем его районе. Вася скачал интересующий его кусок карты OSM <https://stepik.org/media/attachments/lesson/245681/map2.osm> и хочет посчитать, сколько на нём отмечено точечных объектов (node), являющихся заправкой. В качестве ответа вам необходимо вывести одно число - количество АЗС.

Заправка в OSM обозначается парой ключ-значение amenity=fuel. Эта пара находится среди тегов tag внутри node.

Практическое занятие 11

Тема: XML и обработка геоданных 2

Содержание отчета

В качестве отчета прикладываются скрипты выполнения заданий и выходные файлы с результатами работы

Задание 1

Вася, открывший заправку в прошлой задаче, разорился. Конкуренция оказалась слишком большой. Вася предполагает, что это произошло от того, что теги заправки могут быть не только на точке, но и на каком-то контуре. Определите, сколько заправок на самом деле (не только обозначенных node, но и way) есть на фрагменте карты <https://stepik.org/media/attachments/lesson/245681/map2.osm>

Содержание отчета

В качестве отчета прикладываются скрипты выполнения заданий и выходные файлы с результатами работы

Задание 2

Напишите программу, которая будет разрезать большую прямоугольную область на $N * N$ одинаковых прямоугольных областей. Области задаются четырьмя координатами: минимальной широтой, минимальной долготой, максимальной широтой, максимальной долготой.

При выводе области должны быть упорядочены по возрастанию минимальной широты, а в случае равных широт - по возрастанию минимальной долготы.
Гарантируется, что все числа во входных данных положительны.

Содержание отчета

В качестве отчета прикладываются скрипты выполнения заданий и выходные файлы с результатами работы

Задание 3

Вам дана область карты <https://stepik.org/media/attachments/lesson/266078/mapcity.osm>
Пройдите по первым ста node в этой области и выведите для каждого три числа: id, широту (атрибут lat) и долготу (атрибут lon).
Числа для каждого node выводите в отдельной строке, разделяя одним пробелом. Обрабатывать node нужно в том же порядке, в котором они встречаются во входном файле.
Если вы все делаете правильно, то ваш ответ должен начинаться со строк:
60276311 55.5695795 37.5764692
60276312 55.5682232 37.5754573

Практическое занятие 12

Тема: XML и обработка геоданных 3

Содержание отчета

В качестве отчета прикладываются скрипты выполнения заданий и выходные файлы с результатами работы

Задание 1

Вам дана область карты <https://stepik.org/media/attachments/lesson/266078/mapcity.osm>
Пройдите по всем way в этой области, выделите среди них замкнутые (те, у которых совпадает ссылка на первый и последний node), среди всех замкнутых выделите те, у которых установлен tag с ключом building и произвольным значением.
Для каждого подходящего под условия way выведите его id по одному в строке.
Если вы все делаете правильно, то ваш ответ должен начинаться со строк
28889642
28911067

Содержание отчета

В качестве отчета прикладываются скрипты выполнения заданий и выходные файлы с результатами работы

Задание 2

Вам дана область карты <https://stepik.org/media/attachments/lesson/266078/mapcity.osm>
Пройдите по всем way в этой области, выделите среди них замкнутые (те, у которых совпадает ссылка на первый и последний node), среди всех замкнутых выделите те, у которых установлен tag с ключом building и произвольным значением.
Для каждого подходящего под условия way выведите две строки. В первой укажите одно число - id этого way. Во второй перечислите через пробел id (ref) всех nd, входящих в этот

way в том же порядке, в котором они перечисляются в файле.

Выводить ответы для подходящих way нужно в том порядке, в котором они встречаются во входном файле

Если вы все делаете правильно, то ваш ответ должен начинаться со строк

28889642

317555544 317555545 317555546 317555547 317555544

28911067

317558736 529559432 317558559 317558560 317558561 317558562 317558290 529559420

529559416 529559414 529559412 529559410 529559426 529559424 529559422 317558289

317558288 317558736

Содержание отчета

В качестве отчета прикладываются скрипты выполнения заданий и выходные файлы с результатами работы

Задание 3

Вам дана область карты <https://stepik.org/media/attachments/lesson/266078/mapcity.osm>

Пройдите по всем way в этой области, выделите среди них замкнутые (те, у которых совпадает ссылка на первый и последний node), среди всех замкнутых выделите те, у которых установлен tag с ключом building и произвольным значением.

Для каждого подходящего под условия way выведите две строки. В первой укажите одно число - id этого way. Во второй выведите список кортежей, содержащих координаты (широту и долготу) всех node, входящих в этот way. Для этого удобно в первой задаче к предыдущему видео сохранить все координаты в словарь, где ключом является id нода, и совместить ее с последней задачей к предыдущему видео.

Выводить ответы для подходящих way нужно в том порядке, в котором они встречаются во входном файле

Если вы все делаете правильно, то ваш ответ должен начинаться со строк

28889642

[(55.5652795, 37.5695507), (55.5651145, 37.5702288), (55.5648475, 37.5700314), (55.5650147, 37.5693509), (55.5652795, 37.5695507)]

28911067

[(55.5683532, 37.5644676), (55.5682987, 37.5644271), (55.5679549, 37.5641683), (55.5679974, 37.5639919), (55.5683976, 37.5642929), (55.5686577, 37.5632112), (55.5687302, 37.5632692), (55.5687094, 37.5633574), (55.5687319, 37.5633741), (55.5686567, 37.5636906), (55.5686342, 37.5636738), (55.5685984, 37.5638247), (55.5686198, 37.5638406), (55.5684996, 37.5643462), (55.5684605, 37.5643171), (55.5684327, 37.5644347), (55.5683718, 37.5643896), (55.5683532, 37.5644676)] |

Содержание отчета

В качестве отчета прикладываются скрипты выполнения заданий и выходные файлы с результатами работы

Учебно-методическое дисциплины «Анализ данных на Python»

Перечень основной литературы:

1. Груздев А. В. Изучаем Pandas / А. В. Груздев, М. Хейдт, перевод с английского А. В. Груздева. – 2-ое изд., испр. и доп. – Москва: ДМК Пресс, 2019. – 700 с. – ISBN 978-5-97060-670-4. – Текст: электронный // Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/131693>. – Режим доступа: для авториз. пользователей.
2. Маккинли, У. Python и анализ данных / У. Маккинли; перевод А. Слинкина. - 3-е изд. - Саратов: Профобразование, 2024. - 482 с. - ISBN 978-5-4488-0046-7. - Текст: электронный// Цифровой образовательный ресурс IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/145897.html>

Перечень дополнительной литературы:

1. Федоров Д. Ю. Программирование на языке высокого уровня Python: учебное пособие для среднего профессионального образования / Д. Ю. Федоров. – 2-е изд. – Москва: Издательство Юрайт, 2021. – 161 с. – (Профессиональное образование). – ISBN 978-5-534-11961-9. –Текст: электронный // Образовательная платформа Юрайт [сайт]. – URL: <https://urait.ru/bcode/472986>.
2. Титов А. Н. Визуализация распределения данных и ядерная оценка плотности в Python: учебно-методическое пособие / А. Н. Титов, Р. Ф. Тазиева. - Казань: Издательство КНИТУ, 2024. - 96 с. - ISBN 978-5-7882-3479-3. - Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/147881.html>
3. Конкина В. В. Введение в большие данные и анализ информации: учебное пособие / В. В. Конкина, А. Б. Борисенко, И. Л. Коробова. - Тамбов: Тамбовский государственный технический университет, ЭБС АСВ, 2024. - 81 с. - ISBN 978-5-8265-2749-8. - Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/145326.html>

Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. <http://www.intuit.ru> – Интернет-Университет Компьютерных технологий.
2. <https://pandas.pydata.org/>– документация библиотеки pandas
3. <https://scikit-learn.org/stable/index.html>- документация библиотеки sklearn
4. <https://www.kaggle.com/>- репозиторий проектов по анализу данных

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Методические указания
по организации и проведению самостоятельной работы
по дисциплине
«Анализ данных на Python»
для студентов направления подготовки
02.03.01 Математика и компьютерные науки
направленность (профиль)
«Анализ данных и искусственный интеллект»

Общие положения

Самостоятельная работа - планируемая учебная, учебно-исследовательская, научно-исследовательская работа студентов, выполняемая во внеаудиторное (аудиторное) время по заданию и при методическом руководстве преподавателя, но без его непосредственного участия (при частичном непосредственном участии преподавателя, оставляющем ведущую роль за работой студентов).

Самостоятельная работа студентов (СРС) в ВУЗе является важным видом учебной и научной деятельности студента. Самостоятельная работа студентов играет значительную роль в рейтинговой технологии обучения.

К основным видам самостоятельной работы студентов относятся:

- формирование и усвоение содержания конспекта лекций на базе рекомендованной лектором учебной литературы, включая информационные образовательные ресурсы (электронные учебники, электронные библиотеки и др.);
- написание докладов;
- подготовка к семинарам, практическим и лабораторным работам, их оформление;
- составление аннотированного списка статей из соответствующих журналов по отраслям знаний (педагогических, психологических, методических и др.);
- выполнение учебно-исследовательских работ, проектная деятельность;
- подготовка практических разработок и рекомендаций по решению проблемной ситуации;
- выполнение домашних заданий в виде решения отдельных задач, проведения типовых расчетов, расчетно-компьютерных и индивидуальных работ по отдельным разделам содержания дисциплин и т.д.;
- компьютерный текущий самоконтроль и контроль успеваемости на базе электронных обучающих и аттестующих тестов;
- выполнение курсовых работ (проектов) в рамках дисциплин;
- выполнение выпускной квалификационной работы и др.

Методика организации самостоятельной работы студентов зависит от структуры, характера и особенностей изучаемой дисциплины, объема часов на ее изучение, вида заданий для самостоятельной работы студентов, индивидуальных качеств студентов и условий учебной деятельности.

Процесс организации самостоятельной работы студентов включает в себя следующие этапы:

- подготовительный (определение целей, составление программы, подготовка методического обеспечения, подготовка оборудования);
- основной (реализация программы, использование приемов поиска информации, усвоения, переработки, применения, передачи знаний, фиксирование результатов, самоорганизация процесса работы);
- заключительный (оценка значимости и анализ результатов, их систематизация, оценка эффективности программы и приемов работы, выводы о направлениях оптимизации труда).

Самостоятельная работа по дисциплине **«Анализ данных на Python»** направлена на формирование следующих **компетенций**:

ПК-7. Способен подготавливать данные и проводить аналитические работы по исследованию больших данных

1. Цель и задачи самостоятельной работы

Ведущая цель организации и осуществления СРС совпадает с целью обучения студента – формирование набора профессиональных компетенций будущего бакалавра по направлению подготовки «Математика и компьютерные науки».

При организации СРС важным и необходимым условием становятся формирование умения самостоятельной работы для приобретения знаний, навыков и возможности организации учебной и научной деятельности. Целью самостоятельной работы студентов является овладение фундаментальными знаниями, профессиональными умениями и навыками деятельности по профилю, опытом творческой, исследовательской деятельности. Самостоятельная работа студентов способствует развитию самостоятельности, ответственности и организованности, творческого подхода к решению проблем учебного и профессионального уровня.

Задачами СРС являются:

- систематизация и закрепление полученных теоретических знаний и практических умений студентов;
- углубление и расширение теоретических знаний;
- формирование умений использовать нормативную, правовую, справочную документацию и специальную литературу;
- развитие познавательных способностей и активности студентов: творческой инициативы, самостоятельности, ответственности и организованности;
- формирование самостоятельности мышления, способностей к саморазвитию, самосовершенствованию и самореализации;
- развитие исследовательских умений;
- использование материала, собранного и полученного в ходе самостоятельных занятий на семинарах, на практических и лабораторных занятиях, при написании курсовых и выпускной квалификационной работ, для эффективной подготовки к итоговым зачетам и экзаменам.

2. Технологическая карта самостоятельной работы студента

Коды реализуемых компетенций	Вид деятельности студентов	Средства и технологии оценки	Объем часов, в том числе (акад.)		
			СРС	Контактная работа с преподавателем	Всего
8 семестр					
ПК-7	Самостоятельное изучение литературы	Собеседование Защита реферата	28	2	30
ПК-7	Решение практических задач	Конспект	28	2	30
Итого за 8 семестр			56	4	60
Итого			56	4	60

3. Порядок выполнения самостоятельной работы студентом

3.1. Методические рекомендации по работе с учебной литературой

При работе с книгой необходимо подобрать литературу, научиться правильно ее читать, вести записи. Для подбора литературы в библиотеке используются алфавитный и систематический каталоги.

Важно помнить, что рациональные навыки работы с книгой - это всегда большая экономия времени и сил.

Правильный подбор учебников рекомендуется преподавателем, читающим лекционный курс. Необходимая литература может быть также указана в методических разработках по данному курсу.

Изучая материал по учебнику, следует переходить к следующему вопросу только после правильного уяснения предыдущего, описывая на бумаге все выкладки и вычисления (в том числе те, которые в учебнике опущены или на лекции даны для самостоятельного вывода).

При изучении любой дисциплины большую и важную роль играет самостоятельная индивидуальная работа.

Особое внимание следует обратить на определение основных понятий курса. Студент должен подробно разбирать примеры, которые поясняют такие определения, и уметь строить аналогичные примеры самостоятельно. Нужно добиваться точного представления о том, что изучаешь. Полезно составлять опорные конспекты. При изучении материала по учебнику полезно в тетради (на специально отведенных полях) дополнять конспект лекций. Там же следует отмечать вопросы, выделенные студентом для консультации с преподавателем.

Выводы, полученные в результате изучения, рекомендуется в конспекте выделять, чтобы они при перечитывании записей лучше запоминались.

Опыт показывает, что многим студентам помогает составление листа опорных сигналов, содержащего важнейшие и наиболее часто употребляемые формулы и понятия. Такой лист помогает запомнить формулы, основные положения лекции, а также может служить постоянным справочником для студента.

Чтение научного текста является частью познавательной деятельности. Ее цель – извлечение из текста необходимой информации. От того на сколько осознанна читающим собственная внутренняя установка при обращении к печатному слову (найти нужные сведения, усвоить информацию полностью или частично, критически проанализировать материал и т.п.) во многом зависит эффективность осуществляемого действия.

Выделяют **четыре основные установки в чтении научного текста**:

информационно-поисковый (задача – найти, выделить искомую информацию)

усваивающая (усилия читателя направлены на то, чтобы как можно полнее осознать и запомнить как сами сведения излагаемые автором, так и всю логику его рассуждений)

аналитико-критическая (читатель стремится критически осмыслить материал, проанализировав его, определив свое отношение к нему)

творческая (создает у читателя готовность в том или ином виде – как отправной пункт для своих рассуждений, как образ для действия по аналогии и т.п. – использовать суждения автора, ход его мыслей, результат наблюдения, разработанную методику, дополнить их, подвергнуть новой проверке).

Основные виды систематизированной записи прочитанного:

Аннотирование – предельно краткое связное описание просмотренной или прочитанной книги (статьи), ее содержания, источников, характера и назначения;

Планирование – краткая логическая организация текста, раскрывающая содержание и структуру изучаемого материала;

Тезирование – лаконичное воспроизведение основных утверждений автора без привлечения фактического материала;

Цитирование – дословное выписывание из текста выдержек, извлечений, наиболее существенно отражающих ту или иную мысль автора;

Конспектирование – краткое и последовательное изложение содержания прочитанного.

Конспект – сложный способ изложения содержания книги или статьи в логической последовательности. Конспект аккумулирует в себе предыдущие виды записи, позволяет всесторонне охватить содержание книги, статьи. Поэтому умение составлять план, тезисы, делать выписки и другие записи определяет и технологию составления конспекта.

Методические рекомендации по составлению конспекта:

1. Внимательно прочитайте текст. Уточните в справочной литературе непонятные слова. При записи не забудьте вынести справочные данные на поля конспекта;
2. Выделите главное, составьте план;
3. Кратко сформулируйте основные положения текста, отметьте аргументацию автора;
4. Законспектируйте материал, четко следуя пунктам плана. При конспектировании старайтесь выразить мысль своими словами. Записи следует вести четко, ясно.
5. Грамотно записывайте цитаты. Цитируя, учитывайте лаконичность, значимость мысли.

В тексте конспекта желательно приводить не только тезисные положения, но и их доказательства. При оформлении конспекта необходимо стремиться к емкости каждого предложения. Мысли автора книги следует излагать кратко, заботясь о стиле и выразительности написанного. Число дополнительных элементов конспекта должно быть логически обоснованным, записи должны распределяться в определенной последовательности, отвечающей логической структуре произведения. Для уточнения и дополнения необходимо оставлять поля.

Овладение навыками конспектирования требует от студента целеустремленности, повседневной самостоятельной работы.

3.2. Методические рекомендации по подготовке к практическим и лабораторным занятиям

Для того чтобы практические и лабораторные занятия приносили максимальную пользу, необходимо помнить, что упражнение и решение задач проводятся по вычитанному на лекциях материалу и связаны, как правило, с детальным разбором отдельных вопросов лекционного курса. Следует подчеркнуть, что только после усвоения лекционного материала с определенной точки зрения (а именно с той, с которой он излагается на лекциях) он будет закрепляться на практических занятиях как в результате обсуждения и анализа лекционного материала, так и с помощью решения проблемных ситуаций, задач. При этих условиях студент не только хорошо усвоит материал, но и научится применять его на практике, а также получит дополнительный стимул (и это очень важно) для активной проработки лекции.

При самостоятельном решении задач нужно обосновывать каждый этап решения, исходя из теоретических положений курса. Если студент видит несколько путей решения проблемы (задачи), то нужно сравнить их и выбрать самый рациональный. Полезно до начала вычислений составить краткий план решения проблемы (задачи). Решение проблемных задач или примеров следует излагать подробно, вычисления располагать в строгом порядке, отделяя вспомогательные вычисления от основных. Решения при необходимости нужно сопровождать комментариями, схемами, чертежами и рисунками.

Следует помнить, что решение каждой учебной задачи должно доводиться до окончательного логического ответа, которого требует условие, и по возможности с выводом. Полученный ответ следует проверить способами, вытекающими из существа данной задачи. Полезно также (если возможно) решать несколькими способами и сравнить получен-

ные результаты. Решение задач данного типа нужно продолжать до приобретения твердых навыков в их решении.

3.3. Методические рекомендации по самопроверке знаний

После изучения определенной темы по записям в конспекте и учебнику, а также решения достаточного количества соответствующих задач на практических занятиях и самостоятельно студенту рекомендуется, провести самопроверку усвоенных знаний, ответив на контрольные вопросы по изученной теме.

В случае необходимости нужно еще раз внимательно разобраться в материале.

Иногда недостаточность усвоения того или иного вопроса выясняется только при изучении дальнейшего материала. В этом случае надо вернуться назад и повторить плохо усвоенный материал. Важный критерий усвоения теоретического материала - умение решать задачи или пройти тестирование по пройденному материалу. Однако следует помнить, что правильное решение задачи может получиться в результате применения механически заученных формул без понимания сущности теоретических положений.

3.4. Методические рекомендации по написанию научных текстов (докладов, рефератов, эссе, научных статей и т.д.)

Перед тем, как приступить к написанию научного текста, важно разобраться, какова истинная цель вашего научного текста - это поможет вам разумно распределить свои силы и время.

Во-первых, сначала нужно определиться с идеей научного текста, а для этого необходимо научиться либо относиться к разным явлениям и фактам несколько критически (своя идея – как иная точка зрения), либо научиться увлекаться какими-то известными идеями, которые нуждаются в доработке (идея – как оптимистическая позиция и направленность на дальнейшее совершенствование уже известного). Во-вторых, научиться организовывать свое время, ведь, как известно, свободное (от всяких глупостей) время – важнейшее условие настоящего творчества, для него наконец-то появляется время. Иногда именно на организацию такого времени уходит немалая часть сил и талантов.

Писать следует ясно и понятно, стараясь основные положения формулировать четко и недвусмысленно (чтобы и самому понятно было), а также стремясь структурировать свой текст. Каждый раз надо представлять, что ваш текст будет кто-то читать и ему захочется сориентироваться в нем, быстро находить ответы на интересующие вопросы (заодно представьте себя на месте такого человека). Понятно, что работа, написанная «сплошным текстом» (без заголовков, без выделения крупным шрифтом наиболее важным мест и т. п.), у культурного читателя должна вызывать брезгливость и даже жалость к автору (исключения составляют некоторые древние тексты, когда и жанр был иной и к текстам относились иначе, да и самих текстов было гораздо меньше – не то, что в эпоху «информационного взрыва» и соответствующего «информационного мусора»).

Объем текста и различные оформительские требования во многом зависят от принятых в конкретном учебном заведении порядков.

Реферат (доклад) - это самостоятельное исследование студентом определенной проблемы, комплекса взаимосвязанных вопросов.

Реферат не должна составляться из фрагментов статей, монографий, пособий. Кроме простого изложения фактов и цитат, в реферате должно проявляться авторское видение проблемы и ее решения.

Рассмотрим основные этапы подготовки реферата студентом.

Выполнение реферата начинается с выбора темы.

Затем студент приходит на первую консультацию к руководителю, которая предусматривает:

- обсуждение цели и задач работы, основных моментов избранной темы;

- консультирование по вопросам подбора литературы;
- составление предварительного плана.

Следующим этапом является работа с литературой. Необходимая литература подбирается студентом самостоятельно.

После подбора литературы целесообразно сделать рабочий вариант плана работы. В нем нужно выделить основные вопросы темы и параграфы, раскрывающие их содержание.

Составленный список литературы и предварительный вариант плана уточняются, согласуются на очередной консультации с руководителем.

Затем начинается следующий этап работы - изучение литературы. Только внимательно читая и конспектируя литературу, можно разобраться в основных вопросах темы и подготовиться к самостоятельному (авторскому) изложению содержания реферата. Конспектируя первоисточники, необходимо отразить основную идею автора и его позицию по исследуемому вопросу, выявить проблемы и наметить задачи для дальнейшего изучения данных проблем.

Систематизация и анализ изученной литературы по проблеме исследования позволяют студенту написать работу.

Рабочий вариант текста реферата предоставляется руководителю на проверку. На основе рабочего варианта текста руководитель вместе со студентом обсуждает возможности доработки текста, его оформление. После доработки реферат сдается на кафедру для его оценивания руководителем.

Требования к написанию реферата

Написание 1 реферата является обязательным условием выполнения плана СРС по любой дисциплине профессионального цикла.

Тема реферата может быть выбрана студентом из предложенных в рабочей программе или фонде оценочных средств дисциплины, либо определена самостоятельно, исходя из интересов студента (в рамках изучаемой дисциплины). Выбранную тему необходимо согласовать с преподавателем.

Реферат должен быть написан научным языком.

Объем реферата должен составлять 20-25 стр.

Структура реферата:

- Введение (не более 3-4 страниц). Во введении необходимо обосновать выбор темы, ее актуальность, очертить область исследования, объект исследования, основные цели и задачи исследования.
- Основная часть состоит из 2-3 разделов. В них раскрывается суть исследуемой проблемы, проводится обзор мировой литературы и источников Интернет по предмету исследования, в котором дается характеристика степени разработанности проблемы и авторская аналитическая оценка основных теоретических подходов к ее решению. Изложение материала не должно ограничиваться лишь описательным подходом к раскрытию выбранной темы. Оно также должно содержать собственное видение рассматриваемой проблемы и изложение собственной точки зрения на возможные пути ее решения.
- Заключение (1-2 страницы). В заключении кратко излагаются достигнутые при изучении проблемы цели, перспективы развития исследуемого вопроса
- Список использованной литературы (не меньше 10 источников), в алфавитном порядке, оформленный в соответствии с принятыми правилами. В список использованной литературы рекомендуется включать работы отечественных и зарубежных авторов, в том числе статьи, опубликованные в научных журналах в течение последних 3-х лет и ссылки на ресурсы сети Интернет.
- Приложение (при необходимости).

Требования к оформлению:

- текст с одной стороны листа;
- шрифт Times New Roman;

- кегль шрифта 14;
- межстрочное расстояние 1,5;
- поля: сверху 2,5 см, снизу – 2,5 см, слева - 3 см, справа 1,5 см;
- реферат должен быть представлен в сброшюрованном виде.

Порядок защиты реферата:

Защита реферата проводится на практических занятиях, после окончания работы студента над ним и исправления всех недочетов, выявленных преподавателем в ходе консультаций. На защиту реферата отводится 5-7 минут времени, в ходе которого студент должен показать свободное владение материалом по заявленной теме. При защите реферата приветствуется использование мультимедиа-презентации.

Оценка реферата

Реферат оценивается по следующим критериям:

- соблюдение требований к его оформлению;
- необходимость и достаточность для раскрытия темы приведенной в тексте реферата информации;
- умение студента свободно излагать основные идеи, отраженные в реферате;
- способность студента понять суть задаваемых преподавателем и сокурсниками вопросов и сформулировать точные ответы на них.

Критерии оценки:

Оценка «отлично» выставляется студенту, если в докладе студент исчерпывающе, последовательно, четко и логически стройно излагает материал; свободно справляется с задачами, вопросами и другими видами применения знаний; использует для написания доклада современные научные материалы; анализирует полученную информацию; проявляет самостоятельность при написании доклада.

Оценка «хорошо» выставляется студенту, если качество выполнения доклада достаточно высокое. Студент твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопросы по теме доклада.

Оценка «удовлетворительно» выставляется студенту, если материал доклада излагается частично, но пробелы не носят существенного характера, студент допускает неточности и ошибки при защите доклада, дает недостаточно правильные формулировки, наблюдаются нарушения логической последовательности в изложении материала.

Оценка «неудовлетворительно» выставляется студенту, если он не подготовил доклад или допустил существенные ошибки. Студент неуверенно излагает материал доклада, не отвечает на вопросы преподавателя.

Описание шкалы оценивания

Максимально возможный балл за весь текущий контроль устанавливается равным 55. Текущее контрольное мероприятие считается сданным, если студент получил за него не менее 60% от установленного для этого контроля максимального балла. Рейтинговый балл, выставляемый студенту за текущее контрольное мероприятие, сданное студентом в установленные графиком контрольных мероприятий сроки, определяется следующим образом:

Уровень выполнения контрольного задания	Рейтинговый балл (в % от максимального балла за контрольное задание)
Отличный	100
Хороший	80
Удовлетворительный	60
Неудовлетворительный	0

Темы рефератов

1. Python – стандарт для работы с большими данными
2. Прикладные задачи бизнес-аналитиков, для решения которых применяется язык Python
3. Установка сторонних библиотек в Python
4. Преобразования между типами в Python
5. Работа с последовательностями в Python
6. Работа с текстовыми файлами в Python
7. Работа с модулями в Python
8. Формат иерархических данных JSON
9. Формат иерархических данных XML
10. Модуль для работы с числовыми данными NumPy
11. Особенные типы данных в NumPy
12. Работа с векторами и матрицами
13. Вычисление главных статистических метрик с помощью NumPy
14. Модуль для работы с табличным представлением данных Pandas
15. Преобразование словарей в табличный формат Pandas
16. Инструменты визуализации данных для Python
17. Американский и европейский стандарт записи дат
18. Регулярные выражения regex
19. Сбор данных с произвольных веб-страниц
20. Формирование файла CSV средствами Python
21. Загрузка сформированных в Python данных в программу
22. Прямая интеграция с документами Excel
23. Преимущества использования Python для анализа данных по сравнению с другими инструментами.
24. Прикладные задачи бизнес-аналитики, для решения которых подходит инструментарий Python.
25. Обзор библиотек и инструментов.
26. Программирование на Python: типы данных, основные структуры, методы и функции, условные операторы, циклы, списковые включения, функция map() и анонимные функции, отладка кода (try/except).
27. Алгоритмы: оптимизация и сложность на примере алгоритмов сортировок.
28. Библиотека numpy: векторы и массивы, специальные типы данных.
29. Библиотека pandas.
30. Основы работы с датафреймами: загрузка, очистка, фильтрация, группировка и агрегация.
31. Описательные статистики, распределения, разведывательный анализ, работа с пропущенными значениями, постановка гипотезы.
32. Сбор данных из открытых источников: web-scraping, работа с API, парсинг текста и таблиц, сохранение файлов.
33. Виды графиков, их корректное использование и интерпретация.
34. Принципы хорошей визуализации, основные ошибки при визуализации данных.
35. Основы визуализации в matplotlib.
36. Оформление и кастомизация графиков.
37. Визуализации отфильтрованных и сгруппированных данных.
38. Создание интерактивных визуализаций в Plotly.
39. Нейронные сети. Области применения.
40. Использование существующих решений для собственных задач
41. Американский и европейский стандарт записи дат
42. Формат записи дат ISO
43. Формат хранения дат Unix Time
44. Регулярные выражения regex

45. Сбор данных с произвольных веб-страниц
46. Автоматизация передачи входных данных через аргументы командной строки
47. Функции построения графиков в pandas
48. Недостатки Python
49. Модуль для работы с числовыми данными NumPy
50. Особенные типы данных в NumPy
51. Работа с векторами и матрицами
52. Вычисление главных статистических метрик с помощью NumPy
53. Модуль для работы с табличным представлением данных Pandas
54. Преобразование словарей в табличный формат Pandas
55. Загрузка данных из внешних источников
56. Инструменты визуализации данных для Python

Список литературы для выполнения СРС

Перечень основной литературы:

1. Груздев А. В. Изучаем Pandas / А. В. Груздев, М. Хейдт, перевод с английского А. В. Груздева. – 2-ое изд., испр. и доп. – Москва: ДМК Пресс, 2019. – 700 с. – ISBN 978-5-97060-670-4. – Текст: электронный // Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/131693>. – Режим доступа: для авториз. пользователей.
2. Маккинли, У. Python и анализ данных / У. Маккинли; перевод А. Слинкина. - 3-е изд. - Саратов: Профобразование, 2024. - 482 с. - ISBN 978-5-4488-0046-7. - Текст: электронный// Цифровой образовательный ресурс IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/145897.html>

Перечень дополнительной литературы:

1. Федоров Д. Ю. Программирование на языке высокого уровня Python: учебное пособие для среднего профессионального образования / Д. Ю. Федоров. – 2-е изд. – Москва: Издательство Юрайт, 2021. – 161 с. – (Профессиональное образование). – ISBN 978-5-534-11961-9. – Текст: электронный // Образовательная платформа Юрайт [сайт]. – URL: <https://urait.ru/bcode/472986>.
2. Титов А. Н. Визуализация распределения данных и ядерная оценка плотности в Python: учебно-методическое пособие / А. Н. Титов, Р. Ф. Тагиева. - Казань: Издательство КНИТУ, 2024. - 96 с. - ISBN 978-5-7882-3479-3. - Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/147881.html>
3. Конкина В. В. Введение в большие данные и анализ информации: учебное пособие / В. В. Конкина, А. Б. Борисенко, И. Л. Коробова. - Тамбов: Тамбовский государственный технический университет, ЭБС АСВ, 2024. - 81 с. - ISBN 978-5-8265-2749-8. - Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. - URL: <https://www.iprbookshop.ru/145326.html>

Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. <http://www.intuit.ru> – Интернет-Университет Компьютерных технологий.
2. <https://pandas.pydata.org/>– документация библиотеки pandas
3. <https://scikit-learn.org/stable/index.html>- документация библиотеки sklearn
4. <https://www.kaggle.com/>- репозиторий проектов по анализу данных