

Федеральное государственное бюджетное образовательное учреждение высшего образования

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ
СТРОИТЕЛЬНЫЙ УНИВЕРСИТЕТ»**

Кафедра Информатики и прикладной математики

РАБОЧАЯ ТЕТРАДЬ

для выполнения работ компьютерного практикума по дисциплине

Математические методы работы с большими данными

080401

Студент: _____

Институт: _____

Курс: _____

Группа: _____

Преподаватель: _____

Москва 2023

Результаты сдачи контрольных мероприятий студентом _____			
Контрольное мероприятие	Преподаватель	Отметка о зачете	Подпись
Практическая работа 1			
Практическая работа 2			
Практическая работа 3			
Практическая работа 4			
Практическая работа 5			
Практическая работа 6			
Практическая работа 7			
Практическая работа 8			
Практическая работа 9			
Практическая работа 10			
Практическая работа 11			
Контрольное задание			
Результат текущего контроля			

Рабочая тетрадь предназначена для студентов всех специальностей и направлений подготовки **080401** НИУ МГСУ, изучающих курс «**Математические методы работы с большими данными**». В тетради приведены формы для оформления отчета и для результатов выполнения работы на ЭВМ.

Принятые в заданиях номера институтов МГСУ

Институты	ИСА	ИГЭС	ИИЭСМ	ИЦТМС	ИЭУКСН	МФ (мытищ.филиал)
К	1	2	3	4	5	6

** – Обозначение задач повышенной трудности

Практическая работа № 1

Основы работы с Python

Задание 1.1 Пакет Anaconda. Интерфейс среды разработки.

1. Дистрибутив языка программирования Python -Anaconda.
2. **Jupyter Notebook**: запуск, создание папки, переименование.
3. Создание первого ноутбука. Типы ячеек.
4. Горячие клавиши.

Задание 1.2 Программа на языке Python: типы данных, линейные конструкции.

1. Команда print(), print ("Hello", "World!")
2. Математические действия. Задание: запрограммировать арифметическое выражение А (см. ниже свой вариант задания).
3. Переменные. Имена. Типы int, float и str. Приведение типов
4. Ввод данных с клавиатуры input().

Задание 1.3 Программа, включающая оператор ветвления IF - ELSE

1. Логические операции.
2. Конструкции **IF – ELSE, IF ... ELIF ... ELSE**. Блок-схемы
3. Подготовить отчет работы.
4. Задание: написать программу задачи **В** (см. свой вариант задания).

Варианты заданий

А) $z^{5y} + \sqrt[7]{\frac{z}{y}}$ $y=2; z=2$

1. Напишите программу, которая принимает на вход от пользователя число от 1 до 7 **В)** (переменная number) и выводит на экран соответствующий день недели (на русском языке с маленькой буквы).

А) $x^{y^z} + 0,3y$ $x=2; y=2; z=2$

2. В университете принята следующая шкала оценок перевода десятибалльной системы в четырехбалльную: меньше 4 — 'неудовлетворительно', 4-6 — **В)** 'удовлетворительно', 7-8 — 'хорошо', 9 и выше — 'отлично'. Напишите программу, которая принимает на вход полученную оценку mark и переводит её таким в соответствии с принятой в университете шкалой оценок.

А) $1 + x + \frac{x^2 + \sqrt{x+1}}{2 \cdot 3x}$ $x=1,5$

3. Напишите программу, которая принимает на вход целое число (number) и проверяет его на делимость на 2, 3 и 5. При делимости на 2 без остатка она выведет "Число делится на 2 без остатка.", иначе при делимости на 3 выведет "Число делится на 3 без остатка.", иначе при делимости на 5 выведет "Число делится на 5 без остатка.", а иначе выведет "Число не делится ни на 2, ни на 3, ни на 5 без остатка!".
-

A) $z^{3x} + 3x^z - 0,3$ $x=2; z=3$

4. Вы решили сделать систему, которая может подобрать для людей язык программирования в зависимости от их любимого слова. Напишите программу, которая может вам в этом помочь. Программа принимает на вход слово (`like_word`), и если это "рептилия", "питон" или "змея", выводит "Python", если это "плюс" или "плюсы", выводит "C++", если это "рубин" или "кристалл", выводит "Ruby". Ну, а если что-то другое, то тоже выводит "Python".

A) $\frac{x + 3yt - 4}{0,3xyt}$ $x=1; y=2; t=2$

5. Преподаватель ведёт занятия с 10:10 до 11:50, с 13:30 до 15:00, с 16:50 до 18:20. В университет он приходит в 9 утра, а уходит - в 19 часов. Время в университете, свободное от занятий, он посвящает консультированию студентов. Напишите программу, которая помогла бы студентам подстроиться под свободное время преподавателя: пусть они вводят желаемое время (сначала одно число — часы (hour), затем второе — минуты (minute)), а программа показывает, свободен преподаватель в это время или нет (выводит на экран "Преподаватель свободен." или "Преподаватель занят.", соответственно).

A) $x^{x^x} + (x^x)^x + 0,04$ $x=2$

6. Радуга. Напишите программу, которая принимает на вход от пользователя число от 1 до 7 (переменная `number`) и выводит на экран соответствующее название цвета (на русском языке с маленькой буквы).

A) $\frac{x + 2(x-1)^2}{3xt} - \sqrt{\frac{x}{3,3}}$ $x=3; t=2$

7. Времена года. Напишите программу, которая принимает на вход от пользователя число от 1 до 12 (переменная `number`) и выводит на экран соответствующее название месяца и времени года (на русском языке с маленькой буквы).

A) $\frac{\sqrt{x} + \sqrt[3]{x-1}}{x + 3,5x^2}$
 $x=2$

8. Модели RGB и CMYK. Напишите программу, которая принимает на вход от пользователя букву английского алфавита (переменная `letter`) и выводит на экран соответствующее название цвета и название модели (на русском языке с маленькой буквы), в противном случае выводит сообщение, что «Буква не соответствует цвету моделей».

A) $4x^2 + \frac{3x}{7y} + 0,3$ $x=2; y=3$

9. Классификация зданий. Напишите программу, которая принимает на вход от пользователя число равное высоте здания (переменная `height`) и выводит на экран соответствующую категорию здания. По признаку высоты многоэтажные общественные здания классифицируются следующим образом:
- В) • высотой до 30 м - здания повышенной этажности;
 - до 50 м - I категория многоэтажных зданий;
 - до 75 м - II категория;
 - до 100 м — III категория;
 - более 100 м — высотные здания.

A) $(x^y)^3(y^x)^2 + 0,7$ $x=2; y=2$

Классификация мостов по размеру. Напишите программу, которая принимает на вход от пользователя число равное размеру моста (переменная `lengthbridge`) и длине пролета (переменная `spanlength`)

10. и выводит на экран соответствующую категорию моста.

B) По признаку размера мосты классифицируются следующим образом:

- длиной до 25 м – малые мосты;
 - от 25 м – до 100 м – средние мосты;
 - свыше 100м и мосты длиной менее 100 м, но с пролетами более 60 м - большие мосты.
-

A) $5 - \frac{x + \sqrt[3]{x}}{3 - \frac{x}{1 + \frac{x+1}{x}}}$ $x=2$

11. Классификация городов по численности населения. Напишите программу, которая принимает на вход от пользователя число равное численности населения в тысячах (переменная `population`) и выводит на экран соответствующую категорию города.

По признаку численности населения города классифицируются следующим образом:

- B)
- свыше 1 млн чел – крупнейшие;
 - от 250 тыс чел – до 1 млн чел – крупные;
 - от 100 до 250 тыс чел – большие;
 - от 50 до 100 тыс чел – средние;
 - до 50 тыс чел – малые.
-

A) $(x^y)^3 / (y^x)^2 + 0,7\sqrt[3]{xy}$ $x=2; y=3$

Телефонные номера. Напишите программу, которая принимает на вход от пользователя номер телефона (переменная `number`) и выводит название соответствующей службы при совпадении.

12.

- B)
- 112- единый номер;
 - 101 - пожарная охрана и органы МЧС;
 - 102 – полиция;
 - 103 - скорая медицинская помощь;
 - 104 - аварийная газовая служба.
-

A) $\frac{1-x}{1+x} - 1,6x^3 \sqrt[3]{x+7}$ $x=2$

Автомобильные дороги. Напишите программу, которая принимает на вход от пользователя число автомобилей в сутки (интенсивность) (переменная `number`) и выводит название соответствующей категории обычной дороги.

13.

Используйте следующую классификацию:

- B)
- свыше 14000 – IВ;
 - свыше 6000 – II;
 - свыше 2000 до 6000 – III;
 - свыше 200 до 2000 – IV;
 - до 200 – V.
-

$$A) \frac{x^{2y}}{3y} - 7,9\sqrt[5]{y^3} \quad x=0,3 \quad y=1,4$$

Шум. Напишите программу, которая принимает на вход от пользователя уровень звукового давления (дБ) (переменная number) и выводит категорию источника звуков.

14.

Используйте следующую классификацию:

- В)
- ниже 35 дБ - не вреден для здоровья;
 - 35 до 70 дБ - негативное влияние на нервную систему человека;
 - 70 до 85 дБ - ухудшение слуха, головные боли;
 - 85 до 130 дБ – серьезный вред здоровью;
 - от 130 дБ – разрушение органов.

$$A) \frac{\frac{x}{y} - \sqrt{y^2 + x}}{x^2 + 7xy} \quad x=2; \quad y=3$$

Бетон. Напишите программу, которая принимает на вход от пользователя значение плотности бетона (кг/куб.м) (переменная number) и выводит название категории бетона.

15.

Используйте следующую классификацию:

- В)
- особо лёгкие (менее 500 кг/куб. м);
 - лёгкие (500-1800 кг/куб. м);
 - тяжёлые (1800-2500кг/куб. м);
 - особо тяжелые с плотностью (более 2500 кг/куб. м).

Отчет выполнения практической работы 1

Jupyter Notebook. назначение
Jupyter Notebook. Способы запуска
Ноутбук. Типы ячеек. Назначение. создание:
Горячие клавиши работы в Jupyter Notebook

запуск на выполнение:			
Режим редактирования и командный режим. Особенности			
Решение задачи А			
Математические действия	Оператор в Python	Пример	
Сложить			
Вычесть			
Умножить			
Разделить			
Получить целую часть от деления			
Получить остаток от деления			
Возвести в степень			
Переменные. Имена. правила создания			
Типы int, float и str			
Приведение типов			
Логические операции			
	Обозначение	Оператор в Python	Пример
	больше		
	больше или равно		
	меньше		

Практическая работа № 2
Основы работы с Python. Циклы

Задание 2.1. Цикл с параметром.

1. Общий синтаксис цикла с параметром.
2. Функция range().
3. Функция enumerate().
4. Управление выполнением цикла. Операторы break и continue
5. Написать программу по заданию **A** по вариантам (см. ниже свой вариант задания).

Задание 2.2. Цикл с условием

1. Общий синтаксис цикла с условием.
2. Написать программу по заданию **B** (см. свой вариант задания).
3. Написать отчет.

Отчет выполнения практической работы 2

Общий синтаксис цикла с параметром. Синтаксис. примеры
Функция range() Синтаксис. примеры
Функция enumerate().Синтаксис . примеры
с Общий синтаксис цикла с условием. Синтаксис . примеры

Варианты заданий

1. А) Вы решили написать программу, которая выявляет високосные года в диапазоне лет. Пусть она принимает на вход два числа: нижнюю и верхнюю границы годов (переменные year_1 и year_2), и для каждого года из диапазона выводит информацию: "... год високосный" или "... год не високосный", где вместо многоточия указывается номер года

В) Представьте, что перед вами большая цистерна объемом **N** литров, которую необходимо заполнить водой. У вас есть робот, который доставляет в цистерну сначала **m** литров, а затем каждый раз на **m** литров больше, чем в предыдущий. Изначально в цистерне

нет воды. Через сколько заходов робота вода польётся через край? N , m – вводятся пользователем.

2. А) Напишите программу, которая принимает на вход слово (word) и последовательно выводит все русские гласные из этого слова.

В) На банковской карте хранятся N рублей. Хозяин карты написал программу, в которую вводит последовательно свои расходы (spent), после чего программа показывает баланс (balance). Когда сумма средств на карте становится ниже нуля, программа выводит сообщение: "Слишком большие расходы" и прекращает работу.

Напишите программу, которая выводит на экран начальный баланс карты (число N), баланс после каждого списания и сообщение "Слишком большие расходы" при достижении отрицательного баланса. N – вводятся пользователем.

3. А) Введите своё имя (name) и с помощью цикла выведите каждую букву этого имени на экран, сопроводив порядковым номером с помощью фразы: "Буква ... в этом имени - ..." (например, "Буква 1 в этом имени - С").

В) Вы играете в компьютерную игру, дошли до схватки с финальным боссом, но вот беда: ваш компьютер "заглючил", и вы не можете управлять персонажем в игре. Босс атакует и каждую секунду наносит один удар, который отнимает у вас m единиц здоровья. На схватку с боссом вы вышли с N единицами здоровья. Создайте цикл, позволяющий понять, через сколько секунд босс вас убьёт. N , m – вводятся пользователем

4. А) Напишите программу, которая последовательно выводит нечётные числа, которые делятся на 3 или 5, от первого введённого числа (num1) до второго введённого числа (num2) включительно. Если второе число меньше первого или первое число меньше нуля, выведите "Введён неверный диапазон чисел".

В) Клиент банка открыл вклад под 10% годовых и положил на этот счёт Y рублей. Напишите программу, позволяющую определить, через сколько лет клиент будет иметь на счету сумму не менее Z рублей. На вход программе идут два числа — Y , Z .

5. А) Для ряда целых чисел от a до b нужно вычислить сумму квадратов чисел, делящихся на 2, и кубов чисел, делящихся на 3, после чего вывести эти суммы на экран.

В) Пассажирский лифт имеет ограничение на перевозку: не более 400 кг единовременно. Пусть на вход программе поступает вес (weight) входящих в лифт людей, а в случае, когда возникает перевес, программа выдаёт предупреждение: "Перевес N кг". Например, при перевесе в 15 кг должно быть выведено сообщение: "Перевес 15 кг".

6. А) Напишите программу, которая получает на вход число (number) и последовательно в порядке возрастания выводит все его делители на экран (само число тоже является своим делителем).

В) Вывести все квадраты натуральных чисел, не превосходящие данного числа N .

7. А) Написать программу, которая ищет все общие (не равные 1) делители двух чисел (num_1 и num_2), введённых пользователем, и последовательно в порядке возрастания выводит их на экран. Если общих делителей нет, то выводит соответствующее сообщение.

В) Вводится натуральное число. Найти сумму четных цифр, входящих в его состав.

8. А) Написать программу, которая на вход получает натуральное число N и выводит N -е число ряда Фибоначчи на экран.

Пояснение: последовательность Фибоначчи — это ряд чисел, в котором первые два элемента равны 1, а каждый следующий элемент равен сумме двух предыдущих: 1, 1, 2, 3, 5, 8, 13, 21, ...

В) Числа последовательно вводятся в цикле. Ввод завершается числом 0. Определите индекс наибольшего элемента последовательности. Если наибольших элементов несколько,

выведите индекс первого из них. Нумерация элементов начинается с нуля.

9. А) Напишите программу, которая выводит таблицу умножения от 1 до 10 для выбранного целого числа (`number`). Если `number = 15`, то строка таблицы имеет вид "15 X 3 = 45".

В) Дано целое число, состоящее из разных цифр. Определить, какая из цифр заданного числа больше, т. е. найти наибольшую цифру числа и вывести ее на печать.

10. А) Напишите программу, которая ищет наибольший (не равный 1) общий делитель двух чисел, введенных пользователем (`num1`, `num2`), и выводит его на экран. Если общих делителей нет, выводит на экран "Общих делителей не найдено".

В) По данному натуральному числу N найдите наибольшую целую степень двойки, не превосходящую N . Выведите показатель степени и саму степень. Операцией возведения в степень пользоваться нельзя!

11. А) Напишите программу с использованием оператора `continue`, которая принимает на вход строку (`string`), после чего выводит её посимвольно, но без букв "а", "б", "в". Пример: ввод "арбуз", вывод "руз".

В) Числа последовательно вводятся в цикле. Ввод завершается числом 0. Определите, сколько элементов этой последовательности равны ее наибольшему элементу

12. А) Напишите программу, которая анализирует ответы пользователя на вопрос "Любите ли вы Python?". Если ответ пользователя (`string`) — "Да", то программа отвечает: "Это отлично!", иначе программа отвечает: "Увы, это неправильный ответ". Программа должна "спросить" пользователя максимум 5 раз. При пяти ответах "Нет" выводит: "Это безнадежно!".

В) Вводится число N . Преобразовать его в другое число, цифры которого будут следовать в обратном порядке по сравнению с введенным числом.

13. А) Напишите программу, которая проверяет, является ли введенное число (`number`) простым. Если число простое, то программа выводит "Простое", иначе "Не является простым".

В) Числа последовательно вводятся в цикле. Ввод завершается числом 0. Определите, сколько элементов этой последовательности больше предыдущего элемента.

14. А) Напишите программу, которая проверяет введенную пользователем строку (`string`) на соответствие правилу: у строки нет ни одной гласной русской маленькой буквы на каждой четной позиции. Пример подходящей строки: "пока!" (в Python первый символ является нулевым, поэтому мы считаем первый символ четным). Если строка подходит, программа выводит на экран "Какая хорошая строка!", а иначе — "Строка мне не нравится!".

В) Дано натуральное число A . Определите, каким по счету числом Фибоначчи оно является, то есть выведите такое число n , что $\varphi_n = A$. Если A не является числом Фибоначчи, выведите число -1.

Пояснение: последовательность Фибоначчи — это ряд чисел, в котором первые два элемента равны 1, а каждый следующий элемент равен сумме двух предыдущих: 1, 1, 2, 3, 5, 8, 13, 21, ...

15. А) Напишите программу вычисления факториала числа N .

В) Напишите программу, которая ищет наибольший (не равный 1) общий делитель двух чисел, введенных пользователем (`num1`, `num2`), и выводит его на экран. Если общих делителей нет, выводит на экран "Общих делителей не найдено".

Выполнение практической работы 2

Вариант № _____

Текст программы А

Результаты счета на ЭВМ

Текст программы В

Результаты счета на ЭВМ

<i>Практическая работа № 2.</i>	<i>Фамилия И. О.</i>	<i>Дата</i>	<i>Подпись</i>
<i>Работу выполнил:</i>	<i>Студент</i>		
<i>Выполнение:</i>	<i>Преподаватель</i>		
<i>Защита работы</i>	<i>Преподаватель</i>		

Практическая работа № 3

Основы работы с Python. Списки

Задание 3.1. Работа со списками

1. Создание списка.
2. Метод split.
3. Индексация элементов списка. Обращение к элементу списка.
4. Работа со срезами. Границы срезов.
5. Срез с шагом.

Задание 3.2. Методы для работы со списками

1. Метод Append.
2. Сортировка элементов списка. Метод sort.
3. Статистические показатели списка: функции min, max, метод count
4. Генераторы списков.
5. Написать отчет.
6. Написать программу.

Отчет выполнения практической работы 3

Что такое список. Примеры
Создание списка. Синтаксис. Примеры
Метод split Синтаксис. Примеры
Индексация элементов списка. Обращение к элементу списка. Примеры
Работа со срезами. Границы срезов Примеры
Метод Append Синтаксис. Примеры
Метод sort Синтаксис. Примеры

функции min, max, метод count. Синтаксис. Примеры
Генераторы списков. Примеры
Что делает данный код <code>my_list = [x for x in range(10, 100) if x%3 == 0 and x%5 == 0]</code>

Варианты заданий

1. Напишите программу, на вход которой подаётся список чисел одной строкой (через генератор). Программа должна для каждого элемента этого списка вывести сумму двух его соседей. Для элементов списка, являющихся крайними, одним из соседей считается элемент, находящийся на противоположном конце этого списка. Например, если на вход подаётся список «1 3 5 6 10», то на выход ожидается список «13 6 9 15 7».

Если на вход пришло только одно число, надо вывести его же. Вывод должен содержать одну строку с числами нового списка, разделёнными пробелом.

2. Напишите программу, которая принимает на вход список чисел в одной строке (через генератор) и выводит на экран в одну строку значения, которые повторяются в нём более одного раза.

Выводимые числа не должны повторяться, порядок их вывода может быть произвольным. Например: 4 7 0 2 3 4 0 3 8. Результат: 4 0 3

3. Напишите программу, которая принимает на вход список чисел в одной строке (через генератор). Поменять местами самый большой и самый маленький элементы списка.

4. Напишите программу, которая принимает на вход список чисел в одной строке (через генератор). Сформировать новый список из возрастающих чётных чисел.

5. Напишите программу, которая принимает на вход список чисел в одной строке (через генератор). Проверить, является ли введенный список симметричным (первый элемент равен последнему, второй - предпоследнему и т.д.).

6. Напишите программу, которая принимает на вход список чисел в одной строке (через генератор). Подсчитать среднее арифметическое всех положительных элементов списка, введенного пользователем.

7. Напишите программу, которая принимает на вход список чисел в одной строке (через генератор) размера N и целые числа K и L ($1 < K \leq L \leq N$). Найти сумму всех элементов списка, кроме элементов с номерами от K до L включительно.

8. Напишите программу, которая принимает на вход список букв в одной строке (через генератор). Сформируйте список из гласных букв.

Практическая работа № 4

Основы работы с Python. Словари.

Задание 4.1. Работа со словарями.

1. Создание словаря.
2. Обращение к ключам словаря.
3. Метод setdefault().
4. Перебор элементов словаря: по ключам, по значениям, по ключам и значениям одновременно.

Задание 4.2. Вложенные структуры

1. Вложенные списки.
2. Проход по вложенному списку. Фильтрация.
3. Вложенные словари.
4. Фильтрация вложенных словарей.

Задание 4.3. Структуры для хранения данных в файлах

1. JSON-файл. Скачать файл **data.json**
2. Открытие JSON -файла в Python. Библиотека json и функция open
3. Работа с данными. библиотека Collections метод Counter().
4. Анализ данных.
5. Написать отчет.
6. Написать программу.
7. Выполнить задание.

Примечание. Обозначения, используемые в файле **data.json**:

- идентификатор события (id);
- идентификатор клиента (client_id);
- идентификатор пользователя (user_id);
- категория действия (category);
- действие, которое совершил пользователь (action);
- дополнительные параметры события (options).

Отчет выполнения практической работы 4

Что такое словарь. Примеры
Создание словаря. Синтаксис. Примеры
Обращение к ключам словаря. Примеры
Перебор элементов словаря: по ключам, по значениям, по ключам и значениям одновременно. Примеры

Вложенные списки. Примеры
Фильтрация вложенного списка. Примеры
Вложенные словари Примеры
Фильтрация вложенных словарей. Примеры
JSON-файл. Назначение
Открытие JSON -файла в Python. Пример
Метод Counter. Синтаксис. Пример

Варианты заданий

1. **А)** Дан текст. Создайте словарь, в котором перечислены все слова текста и какова их длина.
 В) Выполните на основе набора данных data.json задание:
 Подсчитайте количество клиентов (client_id), которые совершали какие-либо действия. _____
2. **А)** Дан текст. Создайте словарь, в котором перечислены все слова текста и сколько раз они встречаются.
 В) Выполните на основе набора данных data.json задание:
 Сколько действий совершил клиент (client_id) под номером 60459? _____
3. **А)** Дан список стран и городов каждой страны. Затем даны названия городов. Для

каждого города укажите, в какой стране он находится.

В) Выполните на основе набора данных `data.json` задание:

Сколько уникальных клиентов совершили действия с категорией (`category`) = `report`?

4. А) Напишите программу, которая принимает на вход список чисел в одной строке (через генератор). Сформировать словарь с информацией сколько среди чисел четных/нечётных, положительных/отрицательных и нулевых значений.

В) Выполните на основе набора данных `data.json` задание: Сколько действий с категорией (`category`) = `page` совершил клиент под номером 62602?

5. А) Напишите программу, которая на основе списков городов и стран из произвольно текста формирует словарь с перечнем стран и городов, встретившихся в нем.

В) Выполните на основе набора данных `data.json` задание: Сколько клиентов не совершали действий с категориями (`category`) = `datepicker` и `table`?

6. А) Напишите программу, которая принимает на вход список чисел и букв в одной строке (через генератор). Выполните фильтрацию этого списка с распределением чисел на: четные/нечётные, положительные/отрицательные и буквы.

В) Выполните на основе набора данных `data.json` задание:

Напишите идентификатор клиента, который совершил максимальное количество действий с категорией (`category`) = `page`

7. А) Напишите программу, которая принимает на вход список букв русских и английских в одной строке (через генератор). Распределите их по категориям: русские (гласные и согласные) и английский (гласные и согласные).

В) Выполните на основе набора данных `data.json` задание: Сколько всего действий совершили клиенты (`client_id`) под номерами 18923 и 52492?

8. А) Напишите программу, которая анализирует текст на основе словаря, содержащий название города и количество жителей и сообщает к какой категории этот город относится.

По признаку численности населения города классифицируются следующим образом:

- свыше 1 млн чел – крупнейшие;
- от 250 тыс чел – до 1 млн чел – крупные;
- от 100 до 250 тыс чел – большие;
- от 50 до 100 тыс чел – средние;
- до 50 тыс чел – малые.

В) Выполните на основе набора данных `data.json` задание: Какой клиент (`client_id`) совершил наибольшее количество действий с (`category`) = `datepicker`

9. А) Напишите программу, которая анализирует текст на основе словаря, содержащий название моста и его длину, и длину пролета и сообщает к какой категории этот мост относится.

По признаку размера мосты классифицируются следующим образом:

- длиной до 25 м – малые мосты;
- от 25 м – до 100 м – средние мосты;
- свыше 100м и мосты длиной менее 100 м, но с пролетами более 60 м - большие мосты.

В) Выполните на основе набора данных `data.json` задание:

Сколько клиентов совершили 5 действий с категорией (`category`) = `page`?

10. А) Напишите программу, которая анализирует текст на основе словаря, содержащий название дороги и интенсивность движения в сутки и сообщает к какой категории эта дорога относится.

Используйте следующую классификацию:

- выше 14000 – IV;
- выше 6000 – III;
- выше 2000 до 6000 – II;
- выше 200 до 2000 – I;
- до 200 – V.

В) Выполните на основе набора данных data.json задание:

Сколько клиентов совершили действий с категорией (category) = datepicker?

11. А) Напишите программу, которая анализирует текст на основе словаря, содержащий название события и интенсивность шума и сообщает о влиянии этого шума на здоровье.

Используйте следующую классификацию:

- ниже 35 дБ - не вреден для здоровья;
- 35 до 70 дБ - негативное влияние на нервную систему человека;
- 70 до 85 дБ - ухудшение слуха, головные боли;
- 85 до 130 дБ – серьезный вред здоровью;
- от 130 дБ – угроза жизни.

В) Выполните на основе набора данных data.json задание:

Сколько клиентов совершили только одно действие?

12. А) Напишите программу, которая анализирует текст на основе словаря, содержащий вопрос о номере телефона в случае соответствующего события.

Используйте следующие данные:

- 112- единый номер;
- 101 - пожарная охрана и органы МЧС;
- 102 – полиция;
- 103 - скорая медицинская помощь;
- 104 - аварийная газовая служба.

В) Выполните на основе набора данных data.json задание: Сколько клиентов совершили только одно действие с категорией (category) = page?

13. А) Напишите программу, которая анализирует текст на основе словаря, содержащий название объекта и его высоту и сообщает о категории этого здания.

Используйте следующую классификацию

- высотой до 30 м - здания повышенной этажности;
- до 50 м - I категория многоэтажных зданий;
- до 75 м - II категория;
- до 100 м — III категория;
- более 100 м — высотные здания.

В) Выполните на основе набора данных data.json задание: Сколько клиентов совершили действия с категориями (category) = page или report?

14. А) Напишите программу, которая анализирует текст на основе словаря, содержащий имя студента и количество баллов, которые он получил по 10-балльной системе, и сообщает оценку по 4-х балльной системе.

Используйте следующую систему перевода: меньше 4 — 'неудовлетворительно'; 4-6 — 'удовлетворительно'; 7-8 — 'хорошо'; 9 и выше — 'отлично'.

В) Выполните на основе набора данных data.json задание:

Практическая работа № 5

Библиотека NumPy для работы с данными

Задание 5.1. Библиотека NumPy. Массивы.

1. Библиотека NumPy. Импорт библиотеки.
2. Массивы. Структура.
3. Создание массива. Функции `array()` и `values()`.
4. Функции создания массива заданного вида.
5. Индексирование массивов NumPy. Индекс и срез.
6. Операции с массивами. Универсальные (поэлементные) функции.

Задание 5.2. Математические и статистические операции.

1. Функция `mean()`
2. Функция `median()`
3. Функция `corrcoef()`
4. Дисперсия `var()`.
5. Стандартное отклонение `std()`
6. Выполнить задания части 1 и части 2.
7. Написать отчет.

Отчет выполнения практической работы 5

Библиотека NumPy. Импорт библиотеки
Массивы. Структура. Примеры
Создание массива. Функции <code>array()</code> и <code>values()</code>. Синтаксис. Примеры
Функции создания массива заданного вида. Примеры
Индексирование массивов NumPy. Индекс и срез. Примеры

Варианты заданий Часть 2

1. Сформировать матрицу $A[n,n]$ случайным образом (random). Найти столбец с наименьшей суммой элементов.

Вычислить значение медианы этого столбца. Вычисление медианы выполнить двумя способами: через стандартную функцию и через программирование формулы.

2. Сформировать целочисленную матрицу $A[n,m]$ с помощью генератора случайных чисел (random). Посчитать количество элементов матрицы, превосходящих среднее арифметическое значение элементов матрицы.

Вычислить стандартное отклонение для этих значений. Ответ округлите до сотых. Вычисление стандартного отклонения выполнить двумя способами: через стандартную функцию и через программирование формулы.

3. Сформировать целочисленную матрицу $A[n,m]$ с помощью генератора случайных чисел (random). Найти минимальное значение среди сумм элементов всех ее строк.

Чему равен коэффициент корреляции между элементами матрицы с чётными и нечётными индексами?

4. Сформировать целочисленную матрицу $A[n,m]$ с помощью генератора случайных чисел (random). Найти количество четных и нечетных чисел в матрице.

Чему равен коэффициент корреляции между чётными и нечётными элементами матрицы?

5. Сформировать целочисленную матрицу $A[n,m]$ с помощью генератора случайных чисел (random). Отсортировать по возрастанию элементы последней строки матрицы.

Вычислить значение медианы этой строки матрицы. Вычисление медианы выполнить двумя способами: через стандартную функцию и через программирование формулы.

6. Сформировать целочисленную матрицу $A[n,m]$ с помощью генератора случайных чисел (random). Найти сумму модулей отрицательных нечетных элементов.

Вычислить стандартное отклонение для этих значений. Ответ округлите до сотых. Вычисление стандартного отклонения выполнить двумя способами: через стандартную функцию и через программирование формулы.

7. Для вещественной матрицы $A[n,n]$, сформированной с помощью генератора случайных чисел (random), найти наименьший элемент на побочной диагонали.

Вычислить дисперсию элементов побочной диагонали. Ответ округлите до сотых. Вычисление дисперсии выполнить двумя способами: через стандартную функцию и через программирование формулы.

8. Сформировать целочисленную матрицу $A[n,m]$ с помощью генератора случайных чисел (random). Отсортировать матрицу по убыванию элементов последнего столбца.

Вычислить среднее значение элементов последнего столбца. Ответ округлите до сотых. Вычисление среднего значения выполнить двумя способами: через стандартную функцию и через программирование формулы.

9. Сформировать целочисленную матрицу $A[n,m]$ с помощью генератора случайных чисел (random). В матрице поменять местами наибольшие элементы в первом и последнем столбцах.

Вычислить значение коэффициента корреляции между элементами первого и последнего столбца. Ответ округлите до сотых.

10. Сформировать целочисленную матрицу $A[n,m]$ с помощью генератора случайных чисел (random). Определить, сколько элементов среди всех элементов матрицы равны минимальному значению. Вывести их индексы.

Вычислить стандартное отклонение для всех значений матрицы. Ответ округлите до сотых. Вычисление стандартного отклонения выполнить двумя способами: через стандартную функцию и через программирование формулы.

11. Сформировать целочисленную матрицу $A[n,m]$ с помощью генератора случайных чисел (random). Найти все элементы, превышающие по абсолютной величине заданное число B . Подсчитать число таких элементов и записать их в массив C .

Вычислить значение медианы для этого массива C . Вычисление медианы выполнить двумя способами: через стандартную функцию и через программирование формулы.

12. Сформировать целочисленную матрицу $A[n,m]$ с помощью генератора случайных чисел (random). Вставьте первую строку после строки, в которой находится первый встреченный минимальный элемент.

Вычислить значение медианы первой строки. Вычисление медианы выполнить двумя способами: через стандартную функцию и через программирование формулы.

13. Сформировать целочисленную матрицу $A[n,m]$ с помощью генератора случайных чисел (random). Получить новую матрицу путем деления всех элементов исходной матрицы на ее наибольший по модулю элемент

Вычислить дисперсию элементов новой матрицы. Ответ округлите до сотых. Вычисление дисперсии выполнить двумя способами: через стандартную функцию и через программирование формулы

14. Сформировать целочисленную матрицу $A[n,n]$ с помощью генератора случайных чисел (random). Вычислить сумму элементов матрицы, расположенных ниже главной диагонали.

Вычислить стандартное отклонение для элементов главной диагонали матрицы. Ответ округлите до сотых. Вычисление стандартного отклонения выполнить двумя способами: через стандартную функцию и через программирование формулы

15. Сформировать целочисленную матрицу $A[n,n]$ с помощью генератора случайных чисел (random). Найти в каждой строке наибольший элемент и поменять его местами с элементом главной диагонали.

Вычислить значение медианы для элементов главной диагонали преобразованной матрицы. Вычисление медианы выполнить двумя способами: через стандартную функцию и через программирование формулы.

Выполнение практической работы 5 Часть 2

Вариант № _____

Текст программы

Практическая работа № 6

Библиотека Pandas. Основные объекты. Функции

Задание 6.1. Библиотека Pandas. Структуры Series и DataFrame

1. Библиотека Pandas. Импорт библиотеки.
2. Структура Series.
3. Создание Series .
4. Функция display.
5. Доступ к элементам Series с использованием .loc или .iloc.
6. Объект DataFrame. Создание.

Задание 6.2. Основные операции

1. Функция read_csv. Загрузить файл **football.csv**
2. Получение информации о датафрейме: head и tail
3. Получение информации о датафрейме: info
4. Получение информации о датафрейме: describe
5. Индексация и извлечение данных: статистические методы
6. Извлечение данные по условиям
7. Выполнить задание.
8. Написать отчет.

Отчет выполнения практической работы 6

Библиотека Pandas. Импорт библиотеки
Структура Series. Примеры
Создание Series . Примеры
Функция display. Синтаксис. Примеры
Доступ к элементам Series с использованием .loc или .iloc. Примеры
Объект DataFrame. Создание. Синтаксис. Примеры

5. Выполните задание на основе набора данных **football.csv**:
Какова средняя скорость (SprintSpeed) футболистов, зарплата (Wage) которых выше среднего? Ответ округлите до сотых.
-
6. Выполните задание на основе набора данных **football.csv**:
Укажите возраст (Age) самого молодого игрока, у которого хладнокровие (Composure) и реакция (Reactions) превышают 90% от максимального значения, представленного в датасете.
-
7. Выполните задание на основе набора данных **football.csv**:
Сколько пенальти (Penalties) забили бразильские (Nationality, Brazil) футболисты за период, данные о котором представлены в датасете.
-
8. Выполните задание на основе набора данных **football.csv**:
Какую позицию (Position) занимает футболист с самой высокой зарплатой (Wage).
-
9. Выполните задание на основе набора данных **football.csv**:
Какова средняя скорость (SprintSpeed) футболистов, зарплата (Wage) которых ниже среднего? Ответ округлите до сотых.
-
10. Выполните задание на основе набора данных **football.csv**:
Укажите средний возраст (Age) игроков, у которых точность удара головой (HeadingAccuracy) > 50. Ответ округлите до сотых.
-
11. Выполните задание на основе набора данных **football.csv**:
Из какой страны (Nationality) происходит больше всего игроков, чья агрессивность (Aggression) достигает максимума?
-
12. Выполните задание на основе набора данных **football.csv**:
Во сколько раз средняя зарплата (Wage) игроков забивающих максимально число пенальти (Penalties) выше средней зарплате игроков, имеющих максимальную точность удара головой (HeadingAccuracy)?
-
13. Выполните задание на основе набора данных **football.csv**:
Игроки какой страны (Nationality) обладают уровнем хладнокровием (Composure) выше среднего. И какое число пенальти они забили?
-
14. Выполните задание на основе набора данных **football.csv**:
Во сколько раз больше игроки с максимальным уровнем хладнокровия (Composure) забивали больше число пенальти (Penalties) по сравнению с игроками с высоким уровнем агрессии (Aggression). Ответ округлите до сотых.
-
15. Выполните задание на основе набора данных **football.csv**:
Определите разницу между средним возрастом (Age) игроков, обладающих реакцией (Reactions) выше средней и забивших максимальное число пенальти (Penalties), и средним возрастом игроков с уровнем агрессии (Aggression) выше среднего и максимальной скоростью (SprintSpeed)
-

Практическая работа № 7

Анализ датасета. Функции. Строковые данные

Задание 7.1 Создание своих функций и их применение

1. Определение функции. Аргументы функции.
2. Вызов функции.
3. Функция как объект**.
4. Функция `map()`**.
5. Лямбда-функции**.
6. Функция `filter()`**
7. Выполнить задание.

Задание 7.2. Операции над строками

1. Строка - итерируемый объект. Индексация элементов строки.
2. Поиск подстроки в строке. Срез.
3. Методом `find()`.
4. Метод `count()`.
5. Методы `lower()` и `upper()`
6. Метод `replace()`.
7. Регулярные выражения. Модуль `re`**

Задание 7.3. Анализ данных текстового файла.

1. Загрузить файл `StudentsPerformance.csv`
2. Преобразование данных файла в список.
3. Преобразование данных файла в словарь.
4. Общий алгоритм анализа данных.
5. Выполнить задание.
6. Написать отчет.

Примечание. Информация в файле `StudentsPerformance.csv` содержит категории:

- *gender* — пол абитуриента;
- *race/ethnicity* — национальность (зашифрована в виде значений "Group A", "Group B" и т.д.);
- *parental level of education* — уровень образования родителей;
- *lunch* — насколько качественно пообедал абитуриент перед тестом;
- *test preparation course* — закончил ли абитуриент подготовительные курсы;
- *math score* — оценка по математике;
- *reading score* — оценка по чтению;
- *writing score* — оценка по письму.

Отчет выполнения практической работы 7

Определение функции. Структура. Аргументы функции. Пример

Вызов функции. Примеры
Функция как объект**. Синтаксис. Примеры
Функция map()**.Синтаксис. Пример
Лямбда-функции**. Пример
Функция filter()**.Пример.
Строка - итерируемый объект. Индексация элементов строки. Синтаксис. Примеры
Поиск подстроки в строке. Срез. Синтаксис. Примеры
Методом find().Синтаксис. Примеры.
Метод count(). Синтаксис. Примеры

В. Создать функцию-шифровальщик, которая все английские буквы а,е заменяет на *, а все буквы i,o на +. Применить функцию к столбцу “parental level of education“ через лямбда функцию и результат шифровки записать в новый столбец “crypted”.

3. Выполните задание на основе набора данных **StudentsPerformance.csv**:

А. Сколько процентов абитуриентов, получивших на экзамене по письму (writing score) оценку более 90, хорошо пообедали перед экзаменом (lunch = standard)? Ответ округлите до одного знака после запятой.

В. Создать функцию, которая на вход получает целую строку датафрейма, далее смотрит на сумму всех числовых столбцов и если их сумма больше 150 то результат - “good”, а если меньше, то результат - “bad”. Применить эту функцию ко всему датафрейму через лямдба функцию и результат записать в новый столбец.

4. Выполните задание на основе набора данных **StudentsPerformance.csv**:

А. Сколько абитуриентов, относящихся к этнической группе С, закончили подготовительные курсы?

В. Создать функцию, которая принимает строку, сначала переводит ее всю в верхний регистр, а потом считает, сколько в этой строке букв “S”. И количество возвращает как результат. Применить эту функцию к столбцу “parental level of education“ через лямбда функцию и результат записать в новый столбец.

5. Выполните задание на основе набора данных **StudentsPerformance.csv**:

А. Сколько абитуриентов получили на экзамене по письму (writing score) оценку выше 90?

В. Написать функцию, которая принимает строку и ищет в ней индекс буквы “h”. Если индекс больше 1, то как результат вернуть подстроку, начиная с этого индекса и до конца исходной строки. Применить эту функцию к столбцу “parental level of education“ через лямбда функцию и результат записать в новый столбец.

6. Выполните задание на основе набора данных **StudentsPerformance.csv**:

А. Сколько девочек, родители которых имеют степень магистра, набрали по математике больше 90 баллов?

В. Создать функцию, которая на вход получает целую строку датафрейма, далее если пол мужской и writing score больше 50 то результат - “good”, а если меньше, то результат - “bad”. В случае с женским полом - тоже самое, только смотреть столбец reading score. Применить эту функцию ко всему датафрейму через лямдба функцию и результат записать в новый столбец.

7. Выполните задание на основе набора данных **StudentsPerformance.csv**:

А. Какой средний балл экзамене по чтению (reading score) получили девочки? Ответ округлите до трёх знаков после запятой.

В. Создать функцию, которая на вход получает целую строку датафрейма, далее если количество букв “r” в столбце lunch больше 1 а пол женский, то результат - “1 cat”. Если количество букв “r” в столбце lunch больше 1, а пол мужской, то результат - “2 cat”. Если количество букв “r” в столбце lunch меньше 1, то результат - “3 cat”. Применить эту функцию ко всему датафрейму через лямдба функцию и результат записать в новый столбец.

8. Выполните задание на основе набора данных **StudentsPerformance.csv**:

А. Сколько мальчиков закончили подготовительные курсы (test preparation course = completed)?

В. Написать функцию, которая принимает строку и сначала понижает регистр всей строки. Затем ищет, есть ли буквы a, b, c. Если да, то результат - "subgroup1", если нет, то результат - "subgroup2". Применить эту функцию к столбцу "race/ethnicity" через лямбда функцию и результат записать в новый столбец.

9. Выполните задание на основе набора данных **StudentsPerformance.csv**:

А. Какой средний балл, полученный мальчиками на экзамене по чтению? Ответ округлите до трёх цифр после запятой.

В. Написать функцию, которая принимает строку, удаляет пробелы и с помощью регулярных выражений возвращает только строку, в которой есть только буквы. Если там есть другие символы, нужно возвращать "error". Применить эту функцию к столбцу "lunch" через лямбда функцию и результат записать в новый столбец.

10. Выполните задание на основе набора данных **StudentsPerformance.csv**:

А. Сколько мальчиков хорошо пообедали перед экзаменом (lunch = standard)?

В. Написать функцию, которая принимает строку и считает сколько в ней пробелов. Если 2 и больше, нужно возвращать "good", в другом случае нужно найти индекс буквы "s" и вернуть подстроку от начала исходной, до этого индекса. Применить эту функцию к столбцу "parental level of education" через лямбда функцию и результат записать в новый столбец.

11. Выполните задание на основе набора данных **StudentsPerformance.csv**:

А. Сколько мальчиков хорошо пообедали перед экзаменом (lunch = standard)?

В. Написать функцию, которая принимает строку, удаляет пробелы, переводит в нижний регистр и с помощью регулярных выражений смотрит, там все буквы от d до z? Если да, нужно возвращать "gr1", в другом случае - "gr2". Применить эту функцию к столбцу "race/ethnicity" через лямбда функцию и результат записать в новый столбец.

12. Выполните задание на основе набора данных **StudentsPerformance.csv**:

А. Сколько абитуриентов получили на экзамене по чтению (reading score) оценку ниже среднего?

В. Создать функцию-шифровальщик, которая меняет местами первую половину строки и вторую половину. Применить функцию к столбцу "test preparation course" через лямбда функцию и результат шифровки записать в новый столбец "crypted".

13. Выполните задание на основе набора данных **StudentsPerformance.csv**:

А. Сколько разных вариантов значений встречается в столбце "parental level of education"? Введите ответ в виде целого числа.

В. Создать функцию, которая на вход получает целую строку датафрейма, далее берет первые 5 букв из столбца "parental level of education" и последние 5 букв из столбца "lunch" и отдает полученную строку как результат. Применить эту функцию ко всему датафрейму через лямбда функцию и результат записать в новый столбец

14. Выполните задание на основе набора данных **StudentsPerformance.csv**:

А. Сколько процентов абитуриентов полноценно пообедали перед экзаменом? Иными словами, у скольких процентов абитуриентов значение столбца "lunch" = "standard"? Ответ округлите до одной цифры после запятой.

В. Написать функцию, которая принимает число, преобразует в строку и с помощью регулярных выражений смотрит, там все отдельные цифры от 0 до 5? Если да, нужно

Практическая работа № 8.

Предобработка данных. Очистка данных

Задание 8.1. Предобработка данных.

1. Чтение файла. Функция `read_csv()`. Загрузить файл **sample.csv**.
2. Редактирование названия колонок. Функция `columns()`.
3. Очистка данных. Валидность данных. Функции `unique()` и `info()`.
4. Поиск значений с ошибками в файле **sample.csv**. Удаление столбца функция `drop()`.
5. Фильтрация ошибок. Функция `query()`. Функции `str.match()` и `str.contains()`.
6. Преобразование данных. Функция `apply()`.

Задание 8.2. Анализ данных файла *films.csv*.

1. Загрузить файл **films.csv**.
2. Написать отчет.
3. Выполнить задание.

Отчет выполнения практической работы 8

Чтение файла. Функция <code>read_csv()</code>. Аргументы функции. Пример
Функция <code>columns()</code>. Синтаксис. Примеры
Очистка данных. Валидность данных.
Функции <code>unique()</code> и <code>info()</code>. Синтаксис. Пример
Перечислите ошибки в данных в файле <i>sample.csv</i>.
Удаление столбца функция <code>drop()</code>. Синтаксис. Пример
Фильтрация ошибок. Функция <code>query()</code>. Синтаксис. Примеры

5. Выполните задание с применением собственных функций на основе набора данных **films.csv**:

Какой режиссер выпускает (суммарно по годам) больше всего фильмов зимой?

6. Выполните задание с применением собственных функций на основе набора данных **films.csv**:

Сколько суммарно вышло фильмов летом (за июнь, июль, август)?

7. Выполните задание с применением собственных функций на основе набора данных **films.csv**:

Какая студия заработала больше всего денег в жанре комедий за все время?

8. Выполните задание с применением собственных функций на основе набора данных **films.csv**:

В фильмах какого жанра больше всего снимался Nicolas Cage?

9. Выполните задание с применением собственных функций на основе набора данных **films.csv**:

Какой актер снялся в большем количестве высокобюджетных фильмов? *Примечание:* в фильмах, где бюджет выше среднего по данной выборке.

10. Выполните задание с применением собственных функций на основе набора данных **films.csv**:

Какой актер принес меньше всего прибыли в 2012 году?

Примечание. Под «прибылью» понимается разность между сборами и бюджетом фильма.

11. Выполните задание с применением собственных функций на основе набора данных **films.csv**:

Кто из режиссеров снял больше всего прибыльных фильмов?

Примечание. Под «прибылью» понимается разность между сборами и бюджетом фильма.

12. Выполните задание с применением собственных функций на основе набора данных **films.csv**:

Какого жанра среди прибыльных фильмов больше всего?

Примечание. Под «прибылью» понимается разность между сборами и бюджетом фильма.

13. Выполните задание с применением собственных функций на основе набора данных **films.csv**:

Самый убыточный фильм за период с 2012 по 2014 гг. (включительно)?

Примечание. Под «убытками» понимается разность между сборами и бюджетом фильма.

14. Выполните задание с применением собственных функций на основе набора данных **films.csv**:

Кто из режиссеров принес больше всего прибыли?

Примечание. Под «прибылью» понимается разность между сборами и бюджетом фильма.

15. Выполните задание с применением собственных функций на основе набора данных **films.csv**:

Кто из режиссеров снял больше всего фильмов?

Практическая работа № 9.

Описательные статистики и графики на Python.

Задание 9.1. Описательные статистики.

1. Основные понятия: случайная величина, наблюдение, генеральная совокупность и выборка.
2. Меры центра: выборочное среднее, истинное среднее, медиана, мода.
3. Квартили. Эксклюзивный метод подсчета.
4. Меры разброса: межквартильный размах, стандартное отклонение.

Задание 9.2. Визуализация данных файла *tips.csv*.

1. Метод `plot`. Настройка параметров метода. Применение метода ко всему датафрейму, к отдельному показателю (гистограмма распределение признака), к категориальными (нечисловыми) переменными. Отображение двух показателей на графике.
2. Библиотека `Matplotlib` модуль `pyplot`. Методы `axes()` и `hist()`.
3. Форматирование графика: заголовок диаграммы, подписи осей, легенда.
4. Создание линейных графиков. Функция `matplotlib.pyplot.plot()`.
5. Графическая библиотека `Seaborn`. Гистограммы распределения признаков. Метод `distplot()`. Метод `countplot()`. Метод `boxplots()`. Метод `heatmap()`.
6. Напишите отчет.

Примечание. Файл **tips.csv** содержит информацию о ресторанах:

- `total_bill` — общая сумма, уплаченная за заказ;
- `tip` — размер чаевых;
- `sex` — пол клиента;
- `smoker` — является ли клиент курильщиком;
- `day` — день недели;
- `time` — время (обед или ужин);
- `size` — количество посетителей, обедавших за столом.

Отчет выполнения практической работы 9

Основные понятия: случайная величина, наблюдение, генеральная совокупность и выборка Примеры
Выборочное среднее, истинное среднее. Примеры. Как среднее реагирует на выбросы?

Метод plot с категориальными (нечисловыми) переменными	
Постройте графики и опишите их характеристики	
Отображение двух показателей на графике.	
Постройте графики и опишите их характеристики	
Библиотека Matplotlib модуль pyplot . Методы axes() и hist()	
Форматирование графика: заголовок диаграммы, подписи осей, легенда.	
Постройте графики и опишите их характеристики	

Описание	
Графическая библиотека Seaborn	
Метод distplot(). Синтаксис. Пример	
Описание	
Метод countplot().Синтаксис. Пример	
Описание	
Метод boxplots().Синтаксис. Пример	

N	24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
x_i	13 14 14 8 5 7 12 2 1 8 15 12 12 7 5 9 6 7 7 5
y_i	12 13 14 9 6 7 11 1 1 7 15 13 12 6.5 5 8 6 6.5 8 4

Решение практической работы 10 Часть 1

Практическая работа № 11.

Разведывательный анализ данных (EDA)

Задание 11.1. Анализ данных файла `cacao_flavors.xls`

- Основные этапы:
 - Посмотреть на данные.
 - Проверить данные на пустые значения.
 - Проверить данные на дублированные/полностью скоррелированные значения.
 - Проверить данные на наличие выбросов.
 - Отобразить данные, пригодные для дальнейшего построения модели.
 - Проверить модель.
- Выполнить алгоритм с данными `cacao_flavors.xls`
- Написать отчет.
- Выполнить задание.

Примечания. Данные в файле `cacao_flavors.xls`:

- `Bar id` — ID батончика в базе данных
- `Company (Maker-if known)` — название компании, изготовившей батончик.
- `Specific Bean Origin or Bar Name` — точное место происхождения какао-бобов.
- `Broad Bean Origin` — название региона, где были выращены какао-бобы.
- `Company Location` — страна, в которой произведён шоколадный батончик.
- `Bean Type` — вид какао-бобов, используемых в батончике.
- `REF` — числовое значение, показывающее, как давно ревью было добавлено в базу. Чем выше значение, тем более свежее ревью.
- `Review Date` — год публикации рецензии.
- `Cocoa Percent` — процент какао в шоколаде.
- `Rating` — оценка батончика фокус-группой

Отчет выполнения практической работы 11

Ответьте на вопросы по данным из датасета.

1. Первичный осмотр данных	
1.1 Сколько в датасете числовых столбцов?	
1.2 Сколько в датасете строковых столбцов?	
1.3 Переименуйте столбцы	
Новые названия:	
<code>title_col=['bar_id', 'company', 'specific_origin', 'ref', 'review_date', 'cocoa_percent', 'company_location', 'rating', 'bean_type', 'broad_origin']</code>	
1.4. Анализ данных столбцов	
1.4.1. Данные <code>var_id</code> . Сколько уникальных значений? График	

1.4.2. Данные company .	
Сколько компаний?	
Отфильтровать лишние данные и вывести только те компании, у которых в рейтинге больше 10 продуктов. Сколько таких?	
1.4.3 Данные specific_origin .	
Сколько уникальных регионов в датасете?	
Сколько значений, встретившихся в столбце более 10 раз?	
1.4.4 Данные ref . Построить график распределения. Выводы	
Есть аномалии в данных? пропуски?	
1.4.5 Данные review_date . Построить график распределения. Выводы	
Есть аномалии в данных? пропуски?	
Содержит ли этот столбец выбросы?	
1.4.6 Данные cocoa_percent . Предварительно преобразовать данные в вещественные числа.	
Есть аномалии в данных? пропуски?	
Содержит ли этот столбец выбросы?	
Построить график распределения. Выводы	

Отфильтруйте данные руководствуясь здравым смыслом	
1.4.7 Данные company_location .	
Сколько стран?	
Сколько стран, у которых больше 10 ревью?	
1.4.8 Данные rating - будет учиться предсказывать будущая модель	
Есть выбросы?	
Постройте график распределения оценок после выявления выбросов	
Отфильтруйте данные	
1.4.9 Данные bean_type .	
Есть пропуски?	
Сколько значений, встретившихся в столбце более 10 раз	
Отфильтруйте данные.	

1.4.10 Данные broad_origin.	
Есть пропуски?	
Отфильтруйте данные.	
Сколько значений, встретившихся в столбце более 10 раз	
Сколько уникальных значений?	
В столбце много повторяющихся значений с различным написанием. Напишите функцию, исправляющую эту ситуацию.	
Как изменятся данные	
Сколько значений, встретившихся в столбце более 10 раз	
Сколько уникальных значений?	
2. Корреляционный анализ числовых данных	
Построить матрицу корреляций. Сделать выводы	
3. Анализ номинативных переменных.	
Построить графики распределения	
4. Общие выводы	

<i>Практическая работа № 11</i>	<i>Фамилия И. О.</i>	<i>Дата</i>	<i>Подпись</i>
<i>Работу выполнил:</i>	<i>Студент</i>		
<i>Выполнение на ЭВМ:</i>	<i>Преподаватель</i>		
<i>Ручной счет:</i>	<i>Преподаватель</i>		
<i>Защита работы</i>	<i>Преподаватель</i>		

Тема Курсового проекта

Состав типового задания для **курсового проекта**:

1. Выбрать область данных (например, программное обеспечение, компьютеры, строительные объекты, музыкальные треки и т.п.)
2. Найти данные об объектах выбранной области
3. Провести первичный анализ данных выбранного файла.
4. Выявить наличие ошибочных данных.
5. Выявить наличие выбросов.
6. Отфильтровать данные.
7. Определить характеристики для построения модели заданного типа.
8. Построить модель.
9. Оценить ее качество.
10. Представить полученные результаты.
11. Оформить выполнение работы и ее результаты в виде текстового документа. Правила оформления размещены на сайте МГСУ, журнал **Строительство: наука и образование** в разделе **Авторам**.