МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ



ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «ЛИПЕЦКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Институт компьютерных наук

Кафедра автоматизированных систем управления

ЛАБОРАТОРНАЯ РАБОТА №1

по МАШИННОМУ ОБУЧЕНИЮ

на тему «Основы работы в R»

Студент М-РИТ-25-1 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Киселев М.С.

(подпись, дата)

Студент М-РИТ-25-1 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Красиков И.А.

(подпись, дата)

Руководитель

Профессор \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Сараев П.В.

(подпись, дата)

Липецк 2025 г.

**Цель работы**

Изучение основ работы с языком R и базовых методов работы с данными

**Порядок выполнения работы**

1. Подключите таблицу данных состояния качества воздуха в г. Нью-Йорк

airquality следующимм образом:

>library (datasets)

>head (airquality)

2. Посчитайте с помощью средств языка R (без использования циклов):

-Число строк в таблице.

-Число столбцов в таблице.

-Число строк, не имеющих пропусков (NA).

-Число строк, имеющих пропусков одновременно по столбцам Ozone и Solar.R.

-Диапазоны варьирования (минимальное и максимальное значения), а также средние значения по столбцам Ozone, Solar.R, Wind и Temp (без учета пропущенных значений).

-Среднее значение по столбцу Solar.R для 5-го месяца (без учета пропущенных значений).

3. Напишите функцию и сохраните ее в файл meanAir.R(factor, tMin, tMax), которая рассчитывает среднее значение по одному из столбцов factor (Ozone, Solar.R, Wind), когда Temp принимает значения от tMin до t.Max включительно (без учета пропущенных значений). Значения по умолчанию: tMin = 60, tMax = 80.

4. Найдите и выведите на экран с помощью средств языка R средние значения по столбцу Solar.R для каждого месяца (без учета пропущенных значений), используя функции split и sapply/lapply.

5. Реализуйте функцию maxTemp(days = 1), которая возвращает требуемое количество пар месяц/день с максимальной температурой. Например, если параметр days = 3, то должны быть выведены ровно 3 пары значений «месяц/день». Параметр days - количество дней - некоторое натуральное число.

6. Реализуйте функцию testSet(perc = 20), которая возвращает тестовое множество, состоящее из заданного процента строк. Строки должны выбираться случайным образом. Параметр perc - вещественное число от 0 до 100. Если переданное значение параметра выходит за пределы [0; 100], необходимо выдать сообщение об ошибке с помощью функции stop.

7. Сделайте выводы о проделанной работе.

**Ход работы**

Для выполнения лабораторной работы был использован язык программирования Python и библиотека pandas

1. Подключение таблицы данных состояния качества воздуха в г. Нью-Йорк



2. Посчитайте с помощью средств языка Python (без использования циклов):

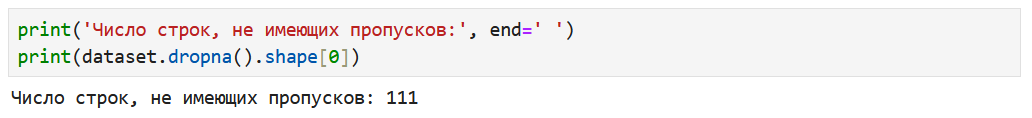
-Число строк в таблице.



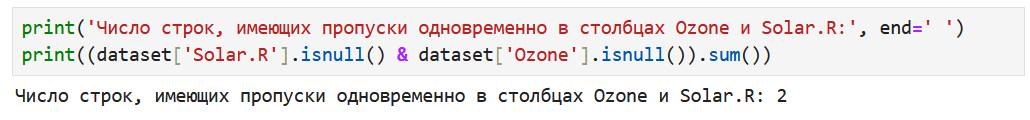
-Число столбцов в таблице.



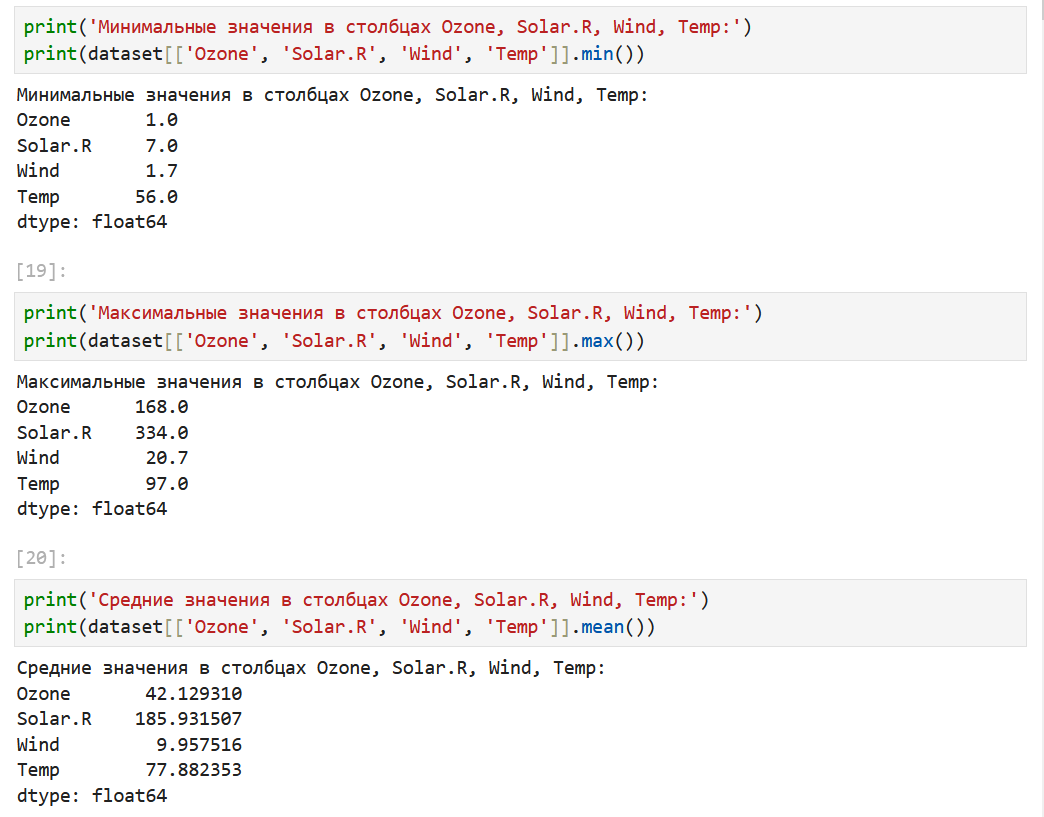
-Число строк, не имеющих пропусков (NA).



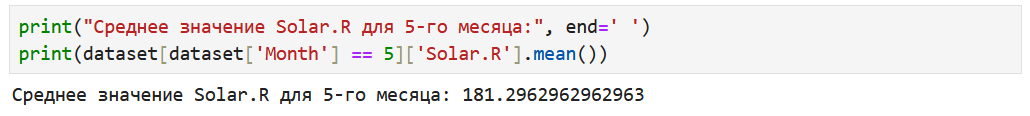
-Число строк, имеющих пропусков одновременно по столбцам Ozone и Solar.R.



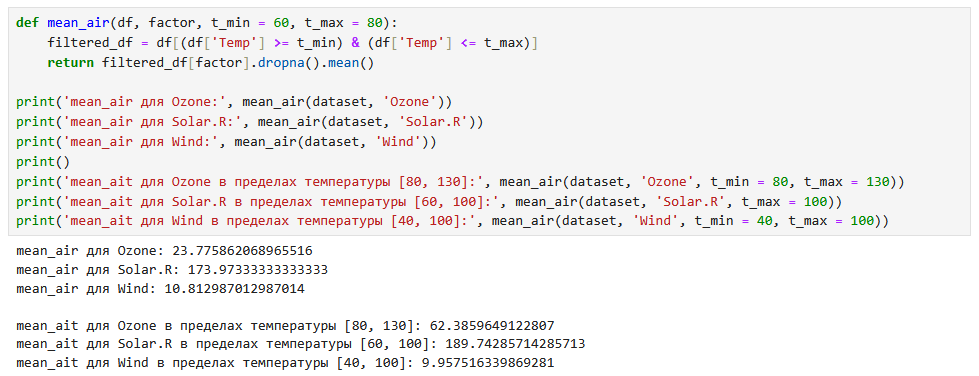
-Диапазоны варьирования (минимальное и максимальное значения), а также средние значения по столбцам Ozone, Solar.R, Wind и Temp (без учета пропущенных значений).



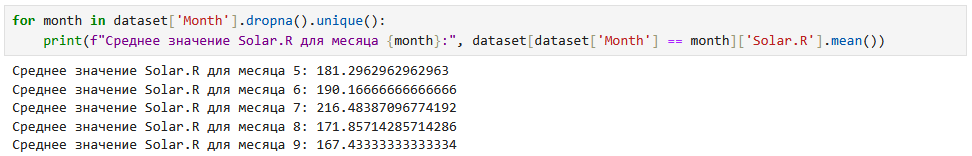
-Среднее значение по столбцу Solar.R для 5-го месяца (без учета пропущенных значений).



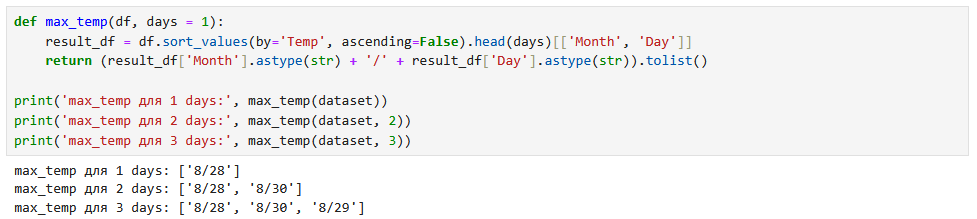
3. Напишите функцию и сохраните ее в файл meanAir.R(factor, tMin, tMax), которая рассчитывает среднее значение по одному из столбцов factor (Ozone, Solar.R, Wind), когда Temp принимает значения от tMin до t.Max включительно (без учета пропущенных значений). Значения по умолчанию: tMin = 60, tMax = 80.



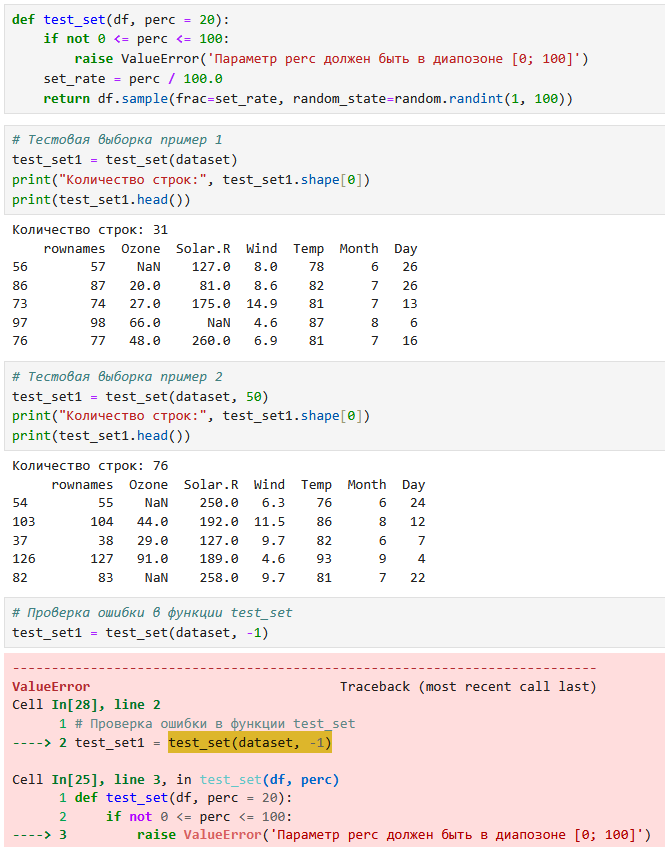
4. Найдите и выведите на экран с помощью средств языка R средние значения по столбцу Solar.R для каждого месяца (без учета пропущенных значений), используя функции split и sapply/lapply.



5. Реализуйте функцию maxTemp(days = 1), которая возвращает требуемое количество пар месяц/день с максимальной температурой. Например, если параметр days = 3, то должны быть выведены ровно 3 пары значений «месяц/день». Параметр days - количество дней - некоторое натуральное число.



6. Реализуйте функцию testSet(perc = 20), которая возвращает тестовое множество, состоящее из заданного процента строк. Строки должны выбираться случайным образом. Параметр perc - вещественное число от 0 до 100. Если переданное значение параметра выходит за пределы [0; 100], необходимо выдать сообщение об ошибке с помощью функции stop.



**Вывод**

Мы изучили основы работы с языком Python и базовыми методами работы с данными. Провели работу над датасетом загрязнений воздуха в городе Нью-Йорк.