

Задачи по pandas

Задача 1

Как найти евклидово расстояние между двумя Series (точками) a и b, не используя встроенную формулу?

Ввод:

```
a = pd.Series([2, 4, 6, 8])
b = pd.Series([1, 3, 5, 7])
```

Ожидаемый вывод:

2.0

```
# Решение (с использованием функции из NumPy)
np.linalg.norm(a-b)
```

Задача 2

Как найти максимально возможное абсолютное значение корреляции каждого столбца с другими столбцами в df?

Ввод:

```
df = pd.DataFrame(np.random.randint(1, 100, 16).reshape(4, 4),
                  columns=list('efgh'), index=list('abcd'))
```

Задача 3

Как нормализовать все столбцы в DataFrame?

1. Нормализуйте все столбцы `df` путём вычитания среднего значения столбца и деления на стандартное отклонение.
2. Сделайте так, чтобы все значения в `df` находились в диапазоне от 0 до 1.

Ввод:

```
df = pd.DataFrame(np.random.randint(1, 100, 16).reshape(4, 4))
```

Задача 4

Как объединить два `DataFrame` по двум столбцам так, чтобы остались только общие строки?

Объедините `df1` и `df2` по столбцам `fruit-frukt` и `weight-ves`.

Ввод:

```
df1 = pd.DataFrame({'fruit': ['apple', 'banana', 'orange'] * 3,  
                    'weight': ['low', 'medium', 'high'] * 3,  
                    'price': np.random.randint(0, 100, 9)})  
  
df2 = pd.DataFrame({'frukt': ['apple', 'banana', 'melon'] * 2,  
                    'ves': ['low', 'high'] * 3,  
                    'price': np.random.randint(0, 100, 6)})
```

Задача 5

Как узнать частоту уникальных значений во всём DataFrame?

Ввод:

```
df = pd.DataFrame(np.random.randint(1, 10, 16).reshape(4, 4),
columns=list('abcd'))
```

Задача 6

Как создать новый столбец, который содержит номера ближайших по евклидовому расстоянию столбцов?

Ввод:

```
df = pd.DataFrame(np.random.randint(1, 100, 16).reshape(4, 4),
columns=list('efgh'), index=list('abcd'))
```

```
df
#      e    f    g    h
# a  51  57  11  15
# b  63   5  81  48
# c   7  63  98  88
# d   6  31  29  37
```

Ожидаемый вывод:

```
df
#      e    f    g    h nearest_row  dist
# a  51  57  11  15              c  122.0
# b  63   5  81  48              a   94.0
# c   7  63  98  88              a  122.0
# d   6  31  29  37              c   92.0
```

Задачи по NumPy

Задача 7

Как поменять местами две строки в двумерном массиве NumPy? Поменяйте местами строки 1 и 3 массива `a`.

Ввод:

```
a = np.arange(9).reshape(3,3)
```

Задача 8

Как найти количество уникальных значений в массиве NumPy? Найдите уникальные значения и их количество в столбце `species` таблицы `iris`.

Ввод:

```
url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'  
iris = np.genfromtxt(url, delimiter=',', dtype='object')
```

Задача 9

Как найти второе максимальное значение в массиве, который сгруппирован по другому массиву? Найдите значение второго самого длинного petallength вида setosa в таблице iris.

Ввод:

```
url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'  
iris = np.genfromtxt(url, delimiter=',', dtype='object')
```

Задача 10

Как отранжировать элементы массива NumPy?

Ввод:

```
a = np.random.randint(100, size=10)  
print(a)
```

```
# [9 15 64 28 89 93 29 8 73 0]
```

Ожидаемый вывод:

```
[2 3 6 4 8 9 5 1 7 0]
```

Задача 11

Как найти результат деления минимального значения на максимальное в каждой строке двумерного массива?

Ввод:

```
np.random.seed(10)
a = np.random.randint(1, 10, [3, 3])
a
# array([[5, 1, 2],
#        [1, 2, 9],
#        [1, 9, 7]])
```

Задача 12

Как найти повторяющиеся значения в массиве NumPy? Найдите повторяющиеся значения (начиная со второго вхождения) в заданном массиве и отметьте их как True. Первое вхождение отмечайте как False.

Ввод:

```
np.random.seed(10)
a = np.random.randint(0, 7, 10)
print(a)
# [1 5 4 0 1 3 4 1 5 0]
```

Ожидаемый вывод:

```
[False False False False True False True True True True]
```

Задача 13

Как удалить из массива NumPy строки, которые содержат nan?

Ввод:

```
url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
iris = np.genfromtxt(url, delimiter=',', dtype='float', usecols=[0, 1, 2, 3])
iris[np.random.randint(150, size=20), np.random.randint(4, size=20)] = np.nan
```

Задачи по DataTable

Задача 14

Как сделать left join двух объектов `datatable.Frame`? Объедините заданные объекты по ключу A.

Ввод:

```
import datatable as dt
df1 = dt.Frame(A=[1, 2, 3, 4], B=['a', 'b', 'c', 'd'])
df2 = dt.Frame(A=[1, 2, 3, 4, 5], C=['a2', 'b2', 'c2', 'd2', 'e2'])
```

Задача 15

Как преобразовать `datatable.Frame` в формат Pandas, NumPy, словаря, списка, кортежа?

[Датасет CSV-файла для формирования](https://raw.githubusercontent.com/selva86/datasets/master/BostonHousing.csv) `datatable.Frame` находится по адресу:

<https://raw.githubusercontent.com/selva86/datasets/master/BostonHousing.csv>

Задача 16

Как узнать типы данных всех столбцов в `datatable.Frame`? [Датасет из примера.](#)

Ожидаемый вывод:

```
crim : stype.float64
zn : stype.float64
indus : stype.float64
chas : stype.bool8
nox : stype.float64
rm : stype.float64
```



```
age : stype.float64
dis : stype.float64
rad : stype.int32
tax : stype.int32
ptratio : stype.float64
b : stype.float64
lstat : stype.float64
medv : stype.float64
```

Задание 17:

Решите без использования циклов средствами NumPy (каждый пункт решается в 1-2 строчки)

3. Создайте вектор с элементами от 12 до 42
4. Создайте вектор из нулей длины 12, но его пятый элемент должен быть равен 1
5. Создайте матрицу (3, 3), заполненную от 0 до 8
6. Найдите все положительные числа в `np.array([1,2,0,0,4,0])`
7. Умножьте матрицу размерности (5, 3) на (3, 2)
8. Создайте матрицу (10, 10) так, чтобы на границе были 0, а внутри 1
9. Создайте случайный вектор и отсортируйте его
10. Каков эквивалент функции `enumerate` для numpy массивов?
11. Создайте случайный вектор и выполните нормализацию столбцов (из каждого столбца вычтите среднее этого столбца, из каждого столбца вычтите `sd` этого столбца)
12. Для заданного числа найдите ближайший к нему элемент в векторе. Найдите `N` наибольших значений в векторе

