

Тест начат	Вторник, 10 октября 2023, 15:44
Состояние	Завершены
Завершен	Вторник, 10 октября 2023, 20:33
Прошло времени	4 час. 48 мин.
Оценка	Еще не оценено

Вопрос **1**

Верно

Баллов: 1,00 из 1,00

Базовые операции с DataFrame

Ваш коллега создал df — объект DataFrame и передал в работу вам:

	I	II	III
0	1.0	5.0	NaN
1	2.0	6.0	NaN
2	3.0	NaN	7.0
3	4.0	NaN	6.0

Выполните следующие операции применительно к данному объекту df:

1. Замените индексы строк на последовательность чисел от 1 до 4, используя соответствующую библиотеки Pandas.
2. Переименуйте названия колонок в последовательность букв A, B, C используя соответствующий метод библиотеки Pandas.
3. Замените пропущенные значения числом 55.

Обратите внимание: никакие результаты работы программы **не требуется** выводить на экран.

Ответ: (штрафной режим: 0%)

Сбросить ответ

```
1 import numpy as np
2 import pandas as pd
3
4 df = pd.DataFrame(np.nan, index=[0, 1, 2, 3], columns=['I', 'II', 'III'])
5 df.loc[0,"I"] = 1
6 df.loc[1,"I"] = 2
7 df.loc[2,"I"] = 3
8 df.loc[3,"I"] = 4
9 df.loc[0,"II"] = 5
10 df.loc[1,"II"] = 6
11 df.loc[2,"III"] = 7
12 df.loc[3,"III"] = 6
13 # далее запишите ваш код
14 df = df.set_axis([x for x in range(1,len(df)+1)], axis = 'index')
15 df = df.set_axis(['ABCDEFGH'[y] for y in range(len(df.columns))], axis = 'columns')
16 df.fillna(55, inplace=True)
```

	Тест	Ожидается	Получил	
✓	print(df.index[0] == 1) print(df.index[1] == 2) print(df.index[2] == 3) print(df.index[3] == 4)	True True True True	True True True True	✓
✓	print(df.columns[0] == 'A') print(df.columns[0] == 'A') print(df.columns[0] == 'A')	True True True	True True True	✓
✓	print(df.iloc[2,1] == 55) print(df.iloc[3,1] == 55) print(df.iloc[0,2] == 55) print(df.iloc[1,2] == 55)	True True True True	True True True True	✓

Все тесты пройдены! ✓

<

Верно
Баллы за эту попытку: 1,00/1,00.

Вопрос **2**

Верно

Баллов: 1,00 из 1,00

Анализ продаж и ассортимента магазина фруктов

Используя имеющийся фрагмент кода (см. окно ввода ответа), выполните последовательно следующие задания:

- Сделайте срез по условию для df. Выберите все строки со значением Q больше 3, для которых значение переменной Shop — это Shop A. Сохраните результат в переменную subset. Получите колонки total для второй строки получившегося среза subset и сохраните его в переменную total2.
- Посчитайте общую выручку в разбивке по всем фруктам (колонка total), результат (объект Series) сохраните в переменную fruit_total.
- Посчитайте общее количество (колонка Q) всех фруктов в разбивке по названию, результат (объект Series) сохраните в переменную fruit_quantity.
- Посчитайте среднюю цену (колонка P) лимонов, результат (объект Series) сохраните в переменную lemon_average_price.

Обратите внимание: в заданиях **не требуется** выполнять вывод полученных результатов на экран.

Ответ: (штрафной режим: 0%)

Сбросить ответ

```
1 import pandas as pd
2 import numpy as np
3
4 fruit = np.array(["lemons", "lemons", "lemons", "lemons",
5                  "apples", "apples", "apples", "apples",
6                  "apples", "apples", "apples"],
7                  dtype=object)
8
9 shop = np.array(["Shop A", "Shop A", "Shop A", "Shop B",
10                 "Shop A", "Shop A", "Shop A", "Shop B",
11                 "Shop B", "Shop B", "Shop A"],
12                 dtype=object)
13
14 pl = np.array(["online", "online", "offline",
15               "online", "online", "offline",
16               "offline", "online", "offline",
17               "offline", "offline"],
18               dtype=object)
19
20 df = pd.DataFrame({'fruit': fruit, 'shop': shop, 'pl': pl,
21                   "Q": [1, 2, 2, 3, 3, 4, 5, 6, 7, 4, 4],
22                   "P": [5, 4, 5, 5, 6, 6, 8, 9, 9, 3, 3]})
23 df['total'] = df['Q']*df['P']
24 subset = df.loc[(df.Q > 3) & (df.shop == 'Shop A')]
25 total2 = subset.reset_index().total[1]
26 fruit_total = df.groupby('fruit')['total'].sum()
27 fruit_quantity = df.groupby('fruit')['Q'].sum()
28 lemon_average_price = df.loc[(df.fruit == 'lemons')].groupby('fruit')['P'].mean()
29
```

	Тест	Ожидается	Получил
✓	print('Задание 1.1') print(total2)	Задание 1.1 40	Задание 1.1 40
✓	print('Задание 1.2') print(fruit_total) print(type(fruit_total))	Задание 1.2 fruit apples 223 lemons 38 Name: total, dtype: int64 <class 'pandas.core.series.Series'>	Задание 1.2 fruit apples 223 lemons 38 Name: total, dtype: int64 <class 'pandas.core.series.Series'>
✓	print('Задание 1.3') print(fruit_quantity) print(type(fruit_quantity))	Задание 1.3 fruit apples 33 lemons 8 Name: Q, dtype: int64 <class 'pandas.core.series.Series'>	Задание 1.3 fruit apples 33 lemons 8 Name: Q, dtype: int64 <class 'pandas.core.series.Series'>

	Тест	Ожидается	Получил
✓	print('Задание 1.4') print(lemon_average_price) print(type(lemon_average_price))	Задание 1.4 fruit lemons 4.75 Name: P, dtype: float64 <class 'pandas.core.series.Series'>	Задание 1.4 fruit lemons 4.75 Name: P, dtype: <class 'pandas.core.ser

Все тесты пройдены! ✓

<

Верно

Баллы за эту попытку: 1,00/1,00.

Вопрос 3

Верно

Баллов: 1,00 из
1,00

Анализ продаж и ассортимента магазина фруктов (продолжение)

Используя имеющийся фрагмент кода (см. окно ввода ответа), выполните следующее задание:

- Создайте сводную таблицу датафрейма df, где будет выведено значение переменной total с разбиением по строкам — категориям переменной shop и колонками по переменной pl. Используйте метод сведения и сохраните таблицу в переменную pivot.
- Выведите значение для второй колонки второй строки таблицы pivot.

Ответ: (штрафной режим: 0%)

Сбросить ответ

```
1 import pandas as pd
2 import numpy as np
3
4 fruit = np.array(["lemons", "lemons", "lemons", "lemons",
5                  "apples", "apples", "apples", "apples",
6                  "apples", "apples", "apples"],
7                  dtype=object)
8
9 shop = np.array(["Shop A", "Shop A", "Shop A", "Shop B",
10                 "Shop A", "Shop A", "Shop A", "Shop B",
11                 "Shop B", "Shop B", "Shop A"],
12                 dtype=object)
13
14 pl = np.array(["online", "online", "offline",
15               "online", "online", "offline",
16               "offline", "online", "offline",
17               "offline", "offline"],
18               dtype=object)
19
20
21
22 df = pd.DataFrame({'fruit': fruit, 'shop': shop, 'pl': pl,
23                   "Q": [1, 2, 2, 3, 3, 4, 5, 6, 7, 4, 4],
24                   "P": [5, 4, 5, 5, 6, 6, 8, 9, 9, 3,3]})
25 df['total'] = df['Q']*df['P']
26 # далее запишите ваш код
27 pivot = pd.pivot_table(df,values = 'total', index = ['shop'], columns = ['pl'], agg
28 print(pivot.iloc[1,1])
```

	Тест	Ожидается	Получил	
✓	print()	69	69	✓
✓	print(pivot.iloc[0,1])	31	31	✓

Все тесты пройдены! ✓

&

Верно

Баллы за эту попытку: 1,00/1,00.

Вопрос **4**
Выполнен
Балл: 1,00

Первичное исследование данных

Это задания не проверяется ЛМС, но выполняя его, вы тренируете навыки работы с реальным набором данных.

Для этого задания вы можете использовать любой доступный вам набор данных. Например, вы загрузите интересный для вас датасет с [сайта Kaggle](#).

Что нужно сделать:

1. Создайте Python-ноутбук, загрузите в него данные из csv-файла.
2. Изучите выбранный набор данных — определите количество столбцов, названия столбцов данных столбцов, использование памяти, индекс диапазона и количество ячеек в каждом столбе (ненулевые значения).
3. Определите, содержатся ли в данных пропущенные значения и строки-дубликаты.
4. Изучите [функции библиотеки Pandas](#) для работы с объектом DataFrame. Подумайте, какие из них могли бы пригодиться при предобработке данных, попробуйте применить их к выбранному набору данных.

По каждому пункту задания сделайте выводы и отразите их в созданном Python-ноутбуке. В качестве решения задания прикрепите ссылку на свой ноутбук.

<https://colab.research.google.com/drive/1PZLqQ7I8bNi29wRb1uGFI87ExVfKtDfw?usp=sharing>

Спасибо! Уверены, что вы хорошо постарались и поработали!

Вот некоторые функции библиотеки Pandas, которые можно использовать для предобработки данных:

- `info()` — выводит краткую сводку о DataFrame;
- `isnull()` — возвращает объект, содержащий булевы значения, которые показывают, какие значения отсутствуют;
- `dropna()` — по умолчанию удаляет все строки, где есть хотя бы одно пропущенное значение в колонках;
- `duplicated()` — возвращает булев объект Series, который для каждой строки показывает, есть ли дубликаты или нет;
- `drop_duplicates()` — удаляет повторяющиеся строки.

Если у вас возникнут вопросы по заданию, пожалуйста, обратитесь к преподавателю на ближайшем семинаре.

Желаем продуктивного обучения!