

Задание 2.2. Объединение датафреймов в Pandas. Дата и время в Pandas

Подготовка к выполнению работы:

1. Посмотрите видеоурок 4 «Обработка и анализ данных на Python. Объединение датафреймов в Pandas», размещенный по ссылке <https://www.youtube.com/watch?v=lsCBbWOOE18>
2. Посмотрите видеоурок 5 «Обработка и анализ данных на Python. Библиотека Pandas. Дата и время в Pandas», размещенный по ссылке <https://www.youtube.com/watch?v=f5FKjFcMsEI>

Выполните следующее:

1. Осуществите краткое описание Вашего набора данных. Приведите электронную ссылку на Ваши данные. Перечислите названия полей и их расшифровку (что они означают).
2. Импортируйте внешнюю библиотеку Pandas: `Import pandas as pd`
3. Прочитайте файл с Вашими данными.
4. Определите количество наблюдений в Ваших данных.
5. Разбейте Ваш файл с данными на части по количеству наблюдений.
6. Приведите скриншот папки с файлами являющимися частями исходного файла с данными.
7. Объедините данные из Ваших частей в один файл и выведите его содержимое.
8. Переиндексируйте итоговый датафрейм и выведите результат.
9. Сгруппируйте данные по значениям в выбранном Вами столбце в Ваших данных. Затем агрегируйте сгруппированные данные по некоторым другим колонкам (см. функцию `aggregate`) и отсортируйте свои данные по возрастанию в каком-нибудь из Ваших столбцов (см. функцию `sort_values`), затем по убыванию значений в этом столбце. При написании кода используйте знак переноса строки (см. `\`) для каждого из метода. Сохраните результат в файл. Пример см. ниже:

```
In [40]: df \
        .query("status == 'Завершен'") \
        .groupby(['title'], as_index=False) \
        .aggregate({'money': 'sum', 'number': 'count'}) \
        .sort_values('money', ascending=False) \
        .rename(columns={'number': 'success_orders'})
```

Out[40]:

	title	money	success_orders
5	Подписка «ОНЛАЙН ДИЕТОЛОГ» с ежемесячным автос...	366947.20	125
0	Курс обучения «Консультант»	208163.49	31
1	Курс обучения «Специалист»	160862.64	7
2	Курс обучения «Эксперт»	148992.80	5
4	Курс от Школы Диетологов. Повышение квалификац...	88384.92	9
3	Курс от Школы Диетологов. Бизнес	18752.54	3

```
In [31]: money_by_city = df \
        .groupby(['title', 'city'], as_index=False) \
        .aggregate({'money': 'sum'}) \
        .sort_values('money', ascending=False)
```

```
In [34]: money_by_city.to_csv('money_by_city.csv', index=False)
```

10. Выведите информацию о Вашем датафрейме. Каковы типы колонок в ваших данных? Есть ли в Ваших данных колонка с датой и временем? Если такой колонки нет, то создайте ее, заполнив ее некоторыми сформированными Вами в коде на Python значениями.
11. Создайте колонку, в которую будут для каждого наблюдения записываться дни недели для соответствующих дат. Выведите результат.
12. Создайте группы (метод `groupby`) для Ваших данных по дням недели. Выведите информацию о некоторой колонке в Ваших данных по дням недели.

13. Импортируйте библиотеку `matplotlib`. Постройте столбчатую диаграмму: по горизонтальной оси – день недели, по вертикальной – значение выводимого Вами признака.
14. Оформите отчёт о результатах исследования и прикрепите его в Teams к заданию 2.2 до или во время своей пары.
15. К этому же заданию прикрепите файл с данными и ru-файл с программой.
16. На паре надо с Вашего ноутбука продемонстрировать работу программы и word-отчет.

Требования к оформлению отчета. Отчет по заданию надо выполнять в строгом соответствии с нумерацией задач. Т.е., переписать задачу с ее номером, скриншот кода на Python для этой задачи, скриншот вывода результата по этому коду.

В конце отчета приведите скриншоты для полного кода, так чтобы были видны номера строк.

Требования к оформлению кода. Код надо писать в строгом соответствии с нумерацией задач. Перед каждым из участков кода, соответствующим задаче задания, пишите соответствующий задаче номер и сопровождайте комментариями.