

Pandas

✓ Чтение и запись данных, часть 1

Домашнее задание

Описание домашнего задания и формат сдачи

Выполните предложенные задачи — впишите свой код в ячейки после комментария `### YOUR CODE HERE ###` и сохраните изменения, используя опцию *Save and Checkpoint* из вкладки меню *File* или кнопку *Save and Checkpoint* на панели инструментов. Итоговый файл в формате `.ipynb` (файл Jupyter Notebook) загрузите в личный кабинет и отправьте на проверку.

```
[ ] from google.colab import drive
    drive.mount('/content/drive')
```

↻ Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

✓ Задача 1

Загрузите содержимое файла `Homework1.csv` в `DataFrame` и выведите первые пять строк на экран. В следующей ячейке посчитайте среднее значение температуры (столбец `Temperature`) и выведите её на экран.

Чтобы изменить содержимое ячейки, дважды нажмите на нее (или выберите "Ввод")

```
[ ] import pandas as pd
    from matplotlib import pyplot as plt
```

```
[ ] df=pd.read_csv('Homework1.csv')
    df.head()
```

↻

	id	Date	City	Temperature
0	32208	01.01.2012	New York City	47.4
1	32209	02.01.2012	New York City	42.9
2	32210	03.01.2012	New York City	29.8
3	32211	04.01.2012	New York City	19.1
4	32212	05.01.2012	New York City	31.7

```
[ ] print('Среднее значение температуры', df['Temperature'].mean())
```

↻ Среднее значение температуры 58.068579234972674

▼ Задача 2

Сохраните DataFrame из предыдущей задачи в файл `result1.csv`, в качестве разделителя используйте запятую `,`, а в качестве названия столбца-индекса — `index`.

```
[ ] df.to_csv('result1.csv', sep=',', index_label = 'index')
```

▼ Задача 3

Загрузите содержимое файла `Homework1.csv` в DataFrame, при этом отформатируйте столбец с датой (`Date`) и приведите его к стандартизированному виду. Выведите первые пять строк DataFrame на экран.

```
df=pd.read_csv('Homework1.csv',parse_dates=['Date'],dayfirst=True)
df.head()
```



	id	Date	City	Temperature
0	32208	2012-01-01	New York City	47.4
1	32209	2012-01-02	New York City	42.9
2	32210	2012-01-03	New York City	29.8
3	32211	2012-01-04	New York City	19.1
4	32212	2012-01-05	New York City	31.7

Задача 4

Сохраните DataFrame из предыдущей задачи в файл `result2.xlsx`.

▼ Новый раздел

```
[ ] df.to_excel('result2.xlsx')
```

✓ Задача 5

Загрузите содержимое файла `Homework2.csv` в `DataFrame`, укажите правильный разделитель столбцов, отформатируйте столбец с датой (`Date`) и приведите его к стандартизированному виду. Выведите первые пять строк `DataFrame` на экран.

```
[ ] df=pd.read_csv('Homework 2.csv',sep=';', parse_dates=['Date'],dayfirst=True)
    df.head()
```



	id	Date	City	Temperature
0	32208	2012-01-01	New York City	47.4
1	32209	2012-01-02	New York City	42.9
2	32210	2012-01-03	New York City	29.авг
3	32211	2012-01-04	New York City	19.январ
4	32212	2012-01-05	New York City	31.июл

✓ Задача 6

Сохраните `DataFrame` из предыдущей задачи в файл `result3.xlsx`, при сохранении задайте листу имя `MySheet`.

(Примечание: для указания имени листа в файле Excel как при чтении, так и при записи используется параметр `sheet_name`.)

```
[ ] df.to_excel('result3.xlsx', sheet_name = 'MySheet')
```

✓ Задача 7

Загрузите содержимое файла `Homework3.xlsx` в `DataFrame`, отформатируйте столбец с датой (`Date`) и приведите его к стандартизированному виду. Выведите первые пять строк `DataFrame` на экран.


После этого в следующей ячейке сохраните полученный `DataFrame` в этот же файл, в лист с названием `MySheet`. При сохранении передайте в функцию сохранения параметр `index` с булевым значением `False`, чтобы не сохранять столбец-индекс.

Затем откройте файл `Homework3.xlsx` с помощью Excel или аналога и посмотрите, сколько листов сейчас содержит этот файл.

Ответ запишите в секции `### YOUR COMMENTS HERE ###` под блоками для кода.

```
[ ] df=pd.read_excel('Homework3.xlsx', parse_dates=['Date'])
    df.head()
```

```
[ ] df=pd.read_excel('Homework3.xlsx', parse_dates=['Date'])
df.head()
```



	id	Date	City	Temperature
0	32208	2012-01-01	New York City	47.4
1	32209	2012-01-02	New York City	42.9
2	32210	2012-01-03	New York City	29.8
3	32211	2012-01-04	New York City	19.1
4	32212	2012-01-05	New York City	31.7


```
[ ] df.to_excel('Homework3.xlsx', sheet_name='MySheet', index=False)
```

Один лист

▼ Задача 8

Загрузите содержимое файлов `Homework1.csv` и `Homework4.csv` в разные датафреймы (DataFrame), выведите первые пять строк каждого датафрейма на экран. После этого объедините датафреймы по столбцу `id` с использованием метода `merge`, выведите первые пять строк объединённого датафрейма на экран и запишите объединённый датафрейм в файл `result4.csv`, указав в качестве разделителя точку с запятой `;`.

```
[ ] df1=pd.read_csv('Homework1.csv')
df1.head()
```



	id	Date	City	Temperature
0	32208	01.01.2012	New York City	47.4
1	32209	02.01.2012	New York City	42.9
2	32210	03.01.2012	New York City	29.8
3	32211	04.01.2012	New York City	19.1
4	32212	05.01.2012	New York City	31.7

```
[ ] df2=pd.read_csv('Homework4.csv')
df2.head()
```

4 32212 05.01.2012 New York City 31.7

```
[ ] df2=pd.read_csv('Homework4.csv')
df2.head()
```

id Sales

0	32208	36
1	32209	34
2	32210	4
3	32211	35
4	32212	48

```
df_merged=pd.merge(df1,df2,on='id',how='inner')
df_merged.head()
```

id Date City Temperature Sales

0	32208	01.01.2012	New York City	47.4	36
1	32209	02.01.2012	New York City	42.9	34
2	32210	03.01.2012	New York City	29.8	4
3	32211	04.01.2012	New York City	19.1	35
4	32212	05.01.2012	New York City	31.7	48

Дополнительные (необязательные) задачи

Задача 9

Загрузите содержимое файлов `Homework2.csv` и `Homework5.xlsx` в разные датафреймы (DataFrame), при этом в каждом датафрейме отформатируйте столбец с датой (`date`) и приведите его к стандартизированному виду. После этого выведите первые пять строк каждого датафрейма на экран. Затем объедините датафреймы по столбцу `date` с использованием метода `merge`, выведите первые пять строк объединённого датафрейма на экран и запишите объединённый датафрейм в файл `result5.csv`, указав в качестве разделителя символ табуляции `\t`.

```
[ ] df_excel=pd.read_excel('Homework5.xlsx', parse_dates=['Date'])
df_excel.head()
```

```
[ ] df_excel=pd.read_excel('Homework5.xlsx', parse_dates=['Date'])
df_excel.head()
```



	Date	Sales
0	2012-01-01	36
1	2012-01-02	34
2	2012-01-03	40
3	2012-01-04	35
4	2012-01-05	48

```
[ ] df_csv = pd.read_csv('Homework 2.csv',sep = ';', parse_dates=['Date'], dayfirst = True)
df_csv.head()
```



	id	Date	City	Temperature
0	32208	2012-01-01	New York City	47.4
1	32209	2012-01-02	New York City	42.9
2	32213	2012-01-06	New York City	41.5
3	32214	2012-01-07	New York City	51.6
4	32215	2012-01-08	New York City	45.1

```
[ ] df_merged = pd.merge(df_excel,df_csv,on='Date')
df_merged.head()
```



	Date	Sales	id	City	Temperature
0	2012-01-01	36	32208	New York City	47.4
1	2012-01-02	34	32209	New York City	42.9
2	2012-01-06	44	32213	New York City	41.5
3	2012-01-07	51	32214	New York City	51.6
4	2012-01-08	36	32215	New York City	45.1

```
df_merged = pd.merge(df_excel,df_csv,on='Date')
df_merged.head()
```



	Date	Sales	id	City	Temperature
0	2012-01-01	36	32208	New York City	47.4
1	2012-01-02	34	32209	New York City	42.9
2	2012-01-06	44	32213	New York City	41.5
3	2012-01-07	51	32214	New York City	51.6
4	2012-01-08	36	32215	New York City	45.1

```
[ ] df_merged.to_csv("result5.csv",sep='\t')
```