

pandas その2

①Excelで以下のようなデータを作成し、Jupyterで使用するフォルダに保存
ファイルの種類を「csv UTF-8(コンマ区切り) (*.csv)」で保存

教科	氏名	点数
国語	山田	64
数学	山田	97
英語	山田	65
理科	山田	97
社会	山田	14
国語	佐藤	59
数学	佐藤	5
英語	佐藤	67
理科	佐藤	35
社会	佐藤	45
国語	佐藤	89
数学	佐藤	15
英語	佐藤	47
理科	佐藤	87
社会	佐藤	36

②csvファイルの読み込み
変数=[DataFrame],read_csv(ファイルパス)

```
import pandas as pd
csv_df=pd.read_csv('pandas.csv')
csv_df[:5]
```

	教科	氏名	点数
0	国語	山田	64
1	数学	山田	97
2	英語	山田	65
3	理科	山田	97
4	社会	山田	14

ヘッダーがない場合

測定器の計測データのような場合

```
csv_df2=pd.read_csv('pandas.csv', header=None, names=['A', 'B', 'C'])  
csv_df2[:5]
```

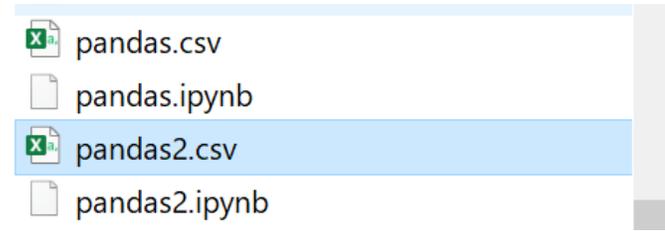
	A	B	C
0	教科	氏名	点数
1	国語	山田	64
2	数学	山田	97
3	英語	山田	65
4	理科	山田	97

CSVファイル保存

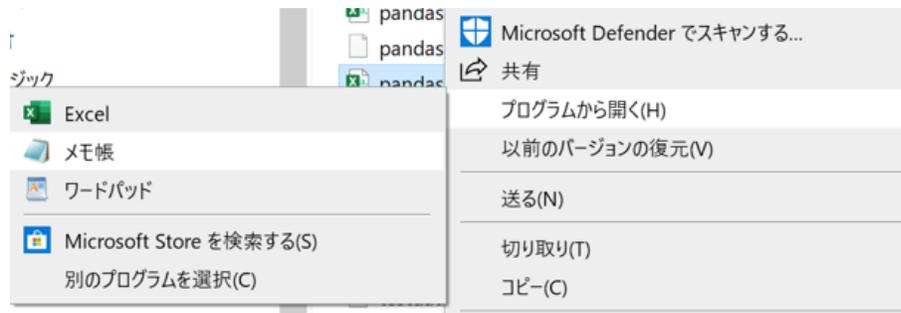
```
sorted_df = csv_df.sort_values('教科')  
sorted_df.to_csv('pandas2.csv')
```

[DataFrame].to_csv(ファイルパス)

①ファイル名選んで右クリック



②プログラムから開く→メモ帳クリック



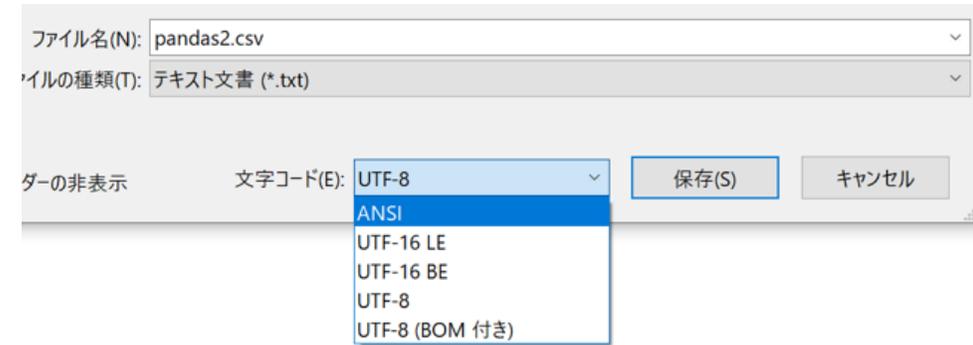
③文字化けしてないことを確認

pandas2.csv - メモ帳

ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)

```
教科, 氏名, 点数  
0, 国語, 山田, 64  
5, 国語, 佐藤, 59  
10, 国語, 佐藤, 89  
1, 数学, 山田, 97  
6, 数学, 佐藤, 5  
11, 数学, 佐藤, 15  
3, 理科, 山田, 97  
8, 理科, 佐藤, 35  
13, 理科, 佐藤, 87  
4, 社会, 山田, 14  
9, 社会, 佐藤, 45  
14, 社会, 佐藤, 36  
2, 英語, 山田, 65  
7, 英語, 佐藤, 67  
12, 英語, 佐藤, 47
```

④文字コードの「ANSI」を選択して保存



⑤Excelを開いて確認

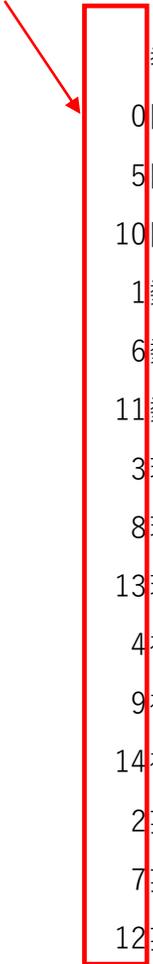
	教科	氏名	点数
0	国語	山田	64
5	国語	佐藤	59
10	国語	佐藤	89
1	数学	山田	97
6	数学	佐藤	5
11	数学	佐藤	15
3	理科	山田	97
8	理科	佐藤	35
13	理科	佐藤	87
4	社会	山田	14
9	社会	佐藤	45
14	社会	佐藤	36
2	英語	山田	65
7	英語	佐藤	67
12	英語	佐藤	47

文字化けの場合

0	謨ア・腕	蠶ケ謨-	64
5	蝗ス隠・菴	占陸	59
10	蝗ス隠・菴	占陸	89
1	謨-蟄ヲ	蠶ア逕-	97
6	謨-蟄ヲ	菴占陸	5
11	謨-蟄ヲ	菴占陸	15
3	逅・ア・蠶	ア逕-	97
8	逅・ア・菴	占陸	35
13	逅・ア・菴	占陸	87
4	遠セ苺・蠶	ア逕-	14
9	遠セ苺・菴	占陸	45
14	遠セ苺・菴	占陸	36
2	闇ア隠・蠶	ア逕-	65
7	闇ア隠・菴	占陸	67
12	闇ア隠・菴	占陸	47

```
sorted_df=csv_df.sort_values('教科')
sorted_df.to_csv('pandas2.csv')
```

インデックスあり



	教科	氏名	点数
0	国語	山田	64
5	国語	佐藤	59
10	国語	佐藤	89
1	数学	山田	97
6	数学	佐藤	5
11	数学	佐藤	15
3	理科	山田	97
8	理科	佐藤	35
13	理科	佐藤	87
4	社会	山田	14
9	社会	佐藤	45
14	社会	佐藤	36
2	英語	山田	65
7	英語	佐藤	67
12	英語	佐藤	47

```
sorted_df=csv_df.sort_values('教科')
sorted_df.to_csv('pandas3.csv', index=False)
```

インデックスなし

教科	氏名	点数
国語	山田	64
国語	佐藤	59
国語	佐藤	89
数学	山田	97
数学	佐藤	5
数学	佐藤	15
理科	山田	97
理科	佐藤	35
理科	佐藤	87
社会	山田	14
社会	佐藤	45
社会	佐藤	36
英語	山田	65
英語	佐藤	67
英語	佐藤	47

Excelファイルの場合

① Windows Powershellあるいはコマンドプロンプトを用いて
`pip install openpyxl`を入力して実行

② Excelファイル読み込み

変数=[DataFrame],`read_excel(ファイルパス)`

```
import pandas as pd
xl_df=pd.read_excel('pandas.xlsx')
xl_df[:10]
```

	教科	氏名	点数
0	国語	山田	64
1	数学	山田	97
2	英語	山田	65
3	理科	山田	97
4	社会	山田	14
5	国語	佐藤	59
6	数学	佐藤	5
7	英語	佐藤	67
8	理科	佐藤	35
9	社会	佐藤	45

インデックス不要の場合

```
sorted_df=xl_df.sort_values('教科')
sorted_df.to_excel('pandas3.xlsx', index=False)
```

Excelファイルに保存

[DataFrame],`to_excel(ファイルパス)`

```
sorted_df=xl_df.sort_values('教科')
sorted_df.to_excel('pandas2.xlsx')
```

インデックス	教科	氏名	点数
0	国語	山田	64
5	国語	佐藤	59
10	国語	佐藤	89
1	数学	山田	97
6	数学	佐藤	5
11	数学	佐藤	15
3	理科	山田	97
8	理科	佐藤	35
13	理科	佐藤	87
4	社会	山田	14
9	社会	佐藤	45
14	社会	佐藤	36
2	英語	山田	65
7	英語	佐藤	67
12	英語	佐藤	47