# アメダス気象データ分析チャレンジ!入門

5. データ分析の例

主催:気象ビジネス推進コンソーシアム

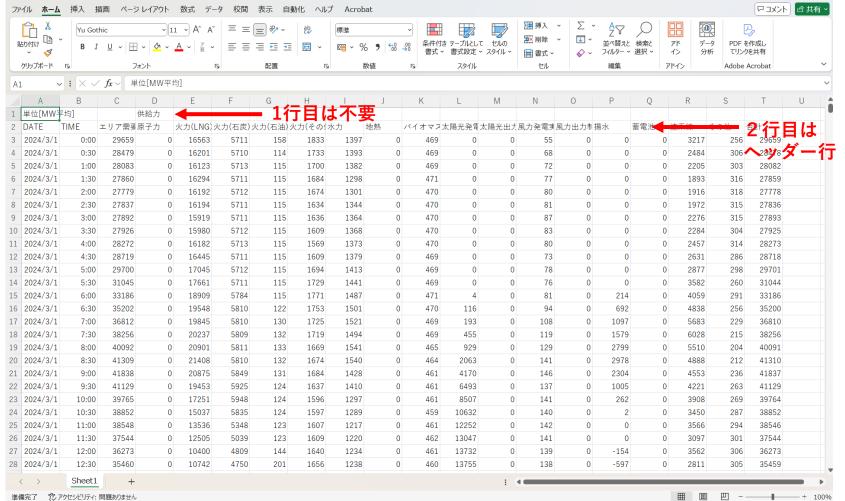
共催:岐阜大学工学部附属応用気象研究センター

資料作成:吉野 純(岐阜大学)





まずは、東京電力の1ヶ月ごとのcsv形式ファイルを、Power Query(Excel 2016以降)を使って、1つのエクセルファイルに連結してみましょう。次のようにエクセルファイルが12個存在する場合を考えます。

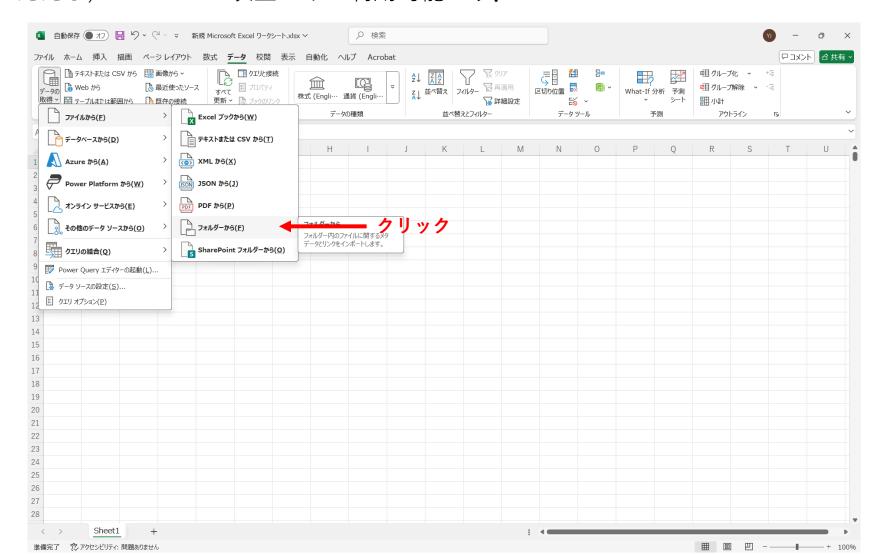


□ 名前	更新日時	種類	サイズ
a eria_jukyu_202403_03.csv	2025/05/22 22:49	Microsoft Excel CS	125 KB
aria_jukyu_202404_03.csv	2025/05/22 22:43	Microsoft Excel CS	120 KB
aria_jukyu_202405_03.csv	2025/05/22 22:49	Microsoft Excel CS	125 KB
aria_jukyu_202406_03.csv	2025/05/22 22:49	Microsoft Excel CS	121 KB
aria_jukyu_202407_03.csv	2025/05/22 22:49	Microsoft Excel CS	127 KB
aria_jukyu_202408_03.csv	2025/05/22 22:49	Microsoft Excel CS	126 KB
aria_jukyu_202409_03.csv	2025/05/22 22:49	Microsoft Excel CS	122 KB
aria_jukyu_202410_03.csv	2025/05/22 22:49	Microsoft Excel CS	127 KB
aria_jukyu_202411_03.csv	2025/05/22 22:49	Microsoft Excel CS	122 KB
aria_jukyu_202412_03.csv	2025/05/22 22:49	Microsoft Excel CS	127 KB
aria_jukyu_202501_03.csv	2025/05/22 22:49	Microsoft Excel CS	126 KB
ai eria_jukyu_202502_03.csv	2025/05/22 22:49	Microsoft Excel CS	114 KB

2024年3月~2025年2月までの計12個のcsvファイルを結合します.

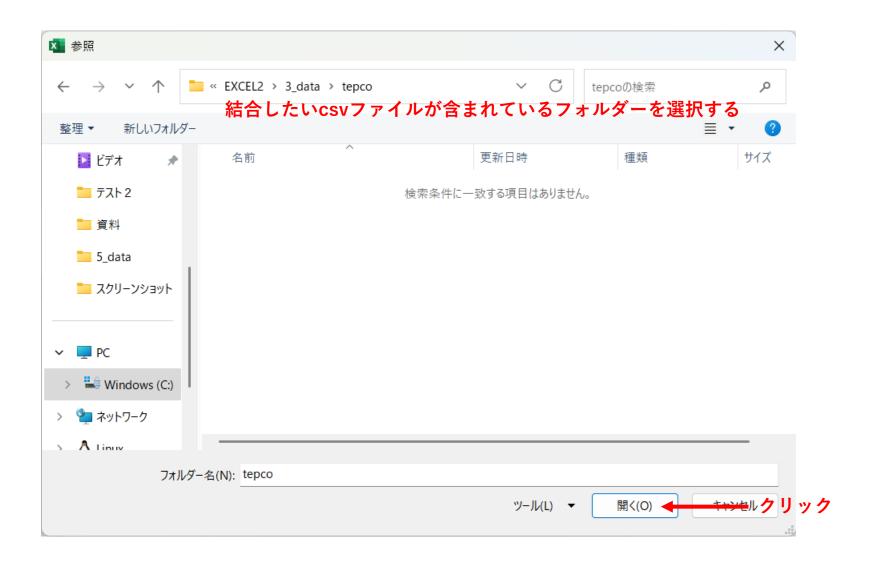


タブ「データ」の「データの取得」から、「ファイルから(F)」「フォルダーから(F)」の順にたどり、クリックします。ただし、Excel 2016以上のみで利用可能です。





結合したいcsvファイルが保存されているフォルダを選択して,「開く(O)」ボタンをクリックします.



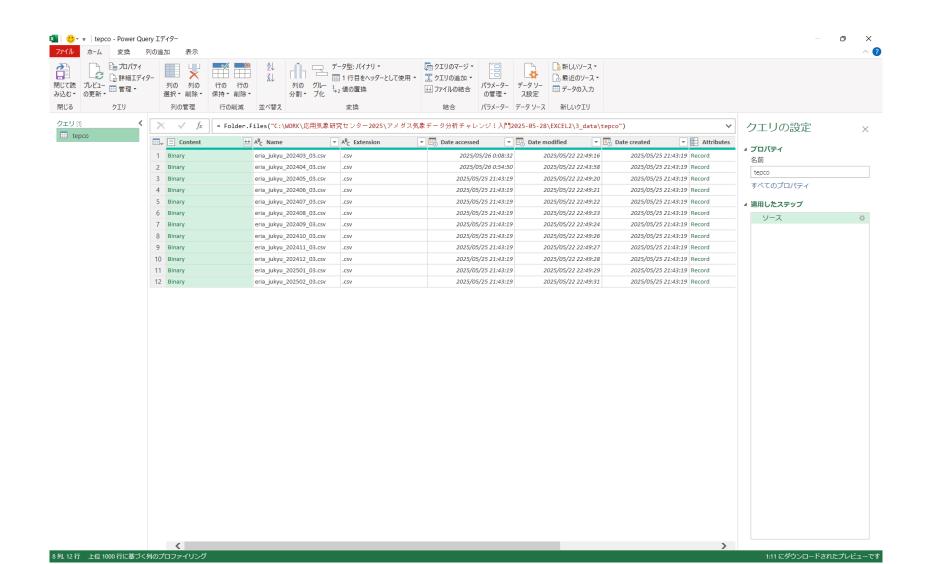


結合したい12個のcsvファイルの一覧が表示されますので、次に「データの変換」をクリックします。



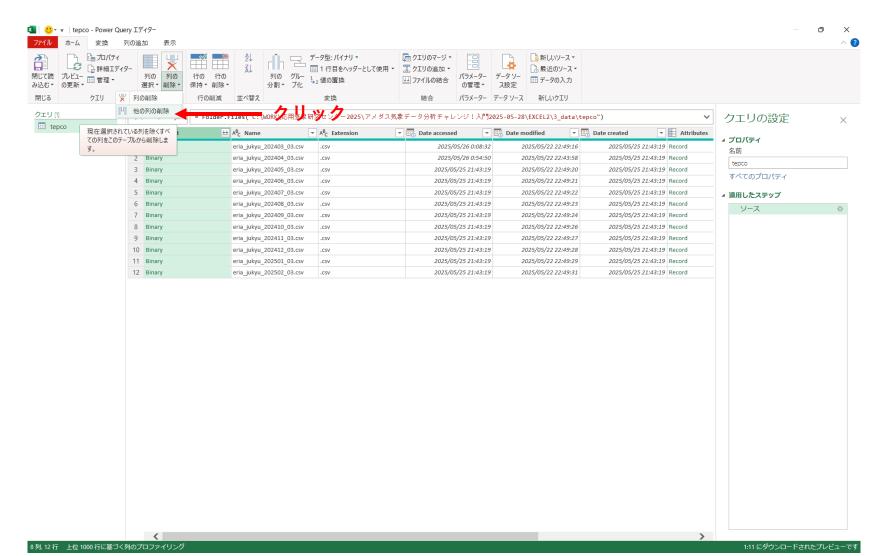


Power Queryエディターが表示されます。このエディターでデータの結合や不要行の削除などを行います。



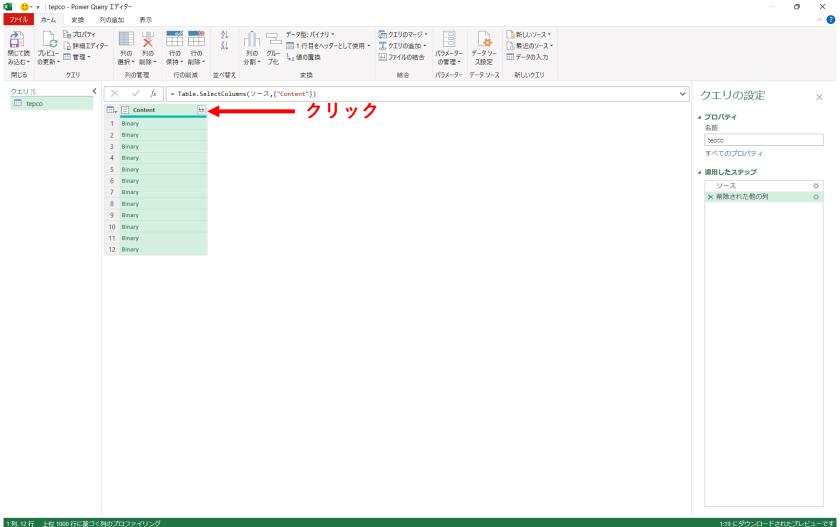


まずは不要な情報を削除します。 "Content"の列を選択した状態で、「列の削除」「他の列の削除」を選択して削除します。



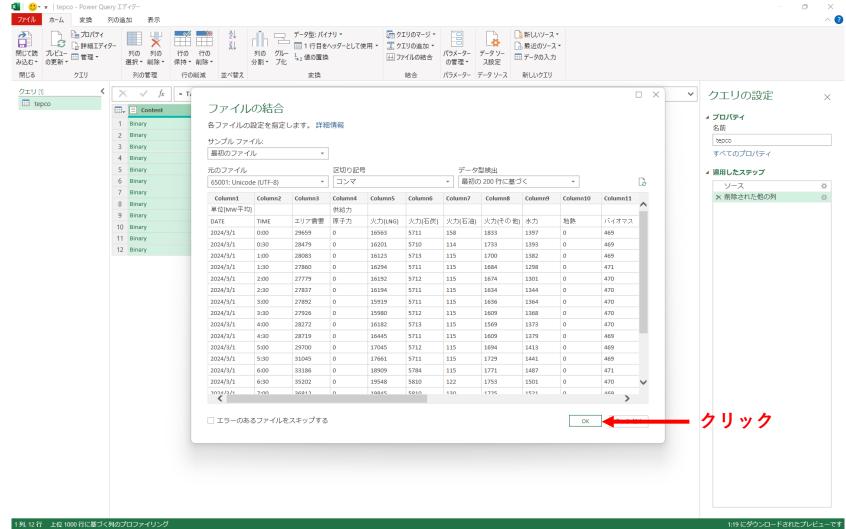


これで、必要な情報のみが残りました.次に、csvファイルを結合させる操作になります. "Content"の右にあ るボタンをクリックします.



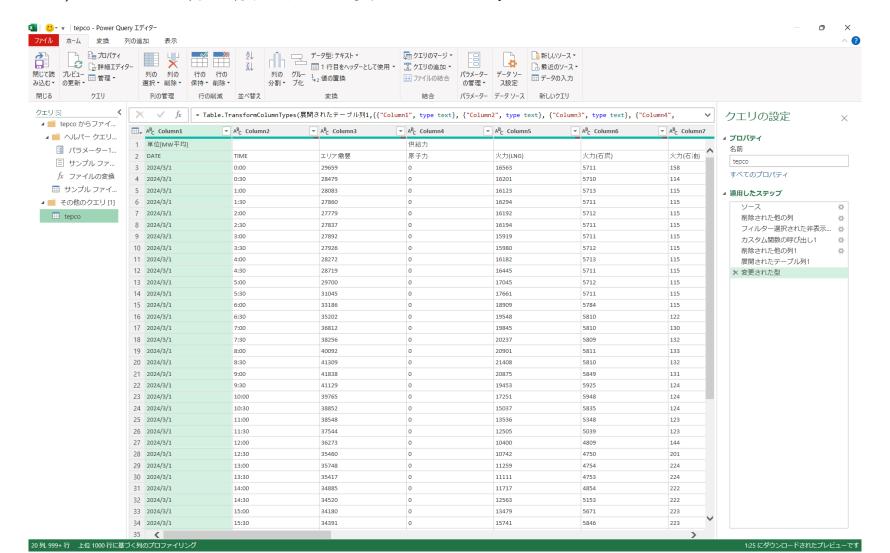


最初のファイルの中身が表示されますので、「OK」をクリックします.



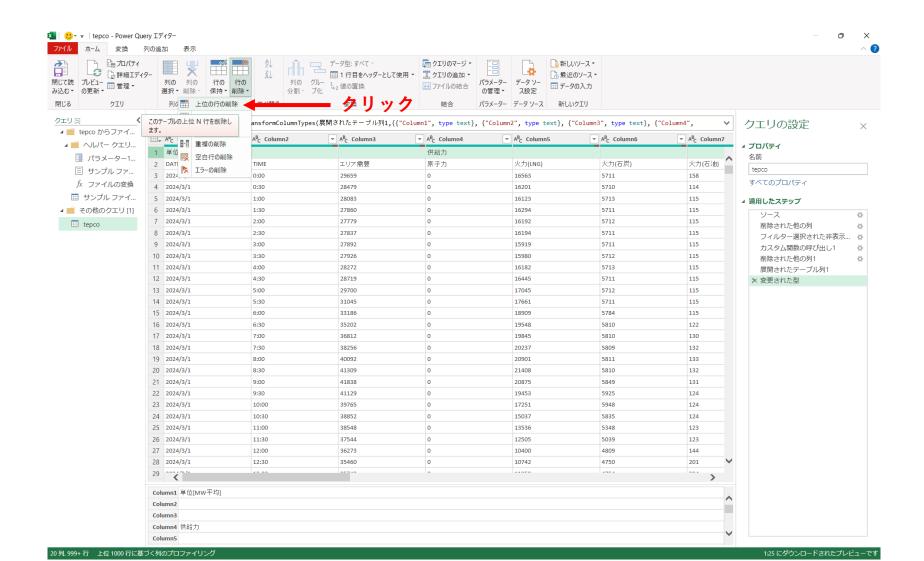


**結合されたファイルの内容が表示されます**. ただし、まだ不要な行(1行目)が含まれているため削除する必要があります。また、ヘッダー行を設定する必要があります。



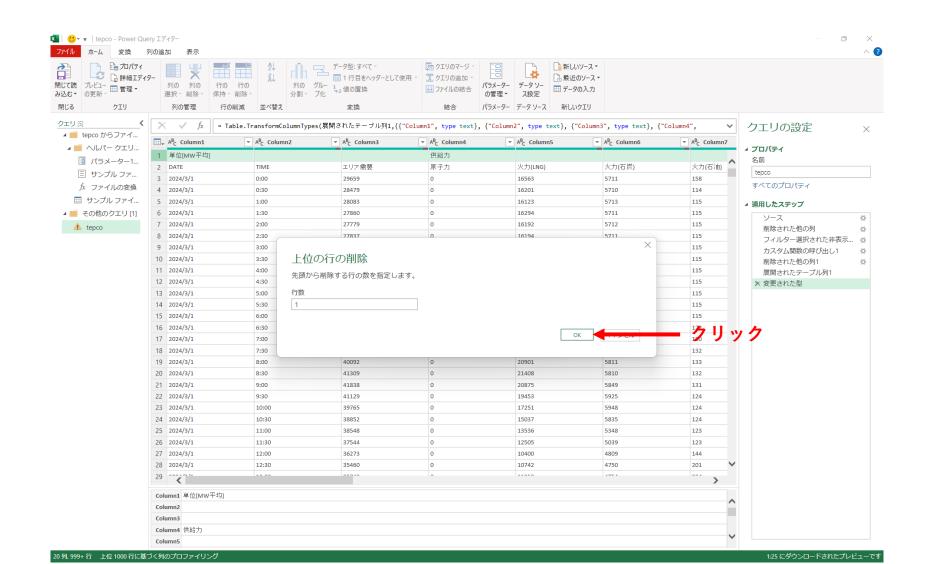


1行目の不要な行を削除します. 「行の削除」「上位の行の削除」をクリックします.



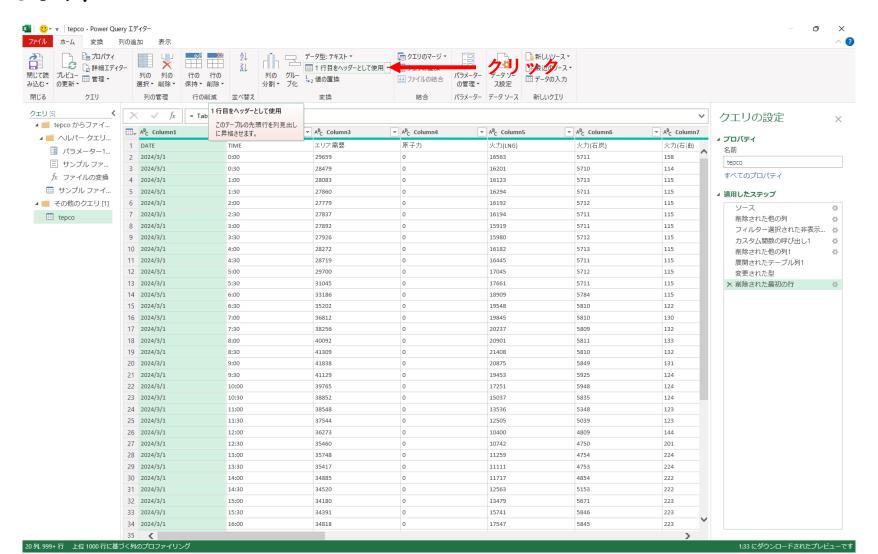


先頭から削除する行の数を「1」と入力して、「OK」をクリックします.



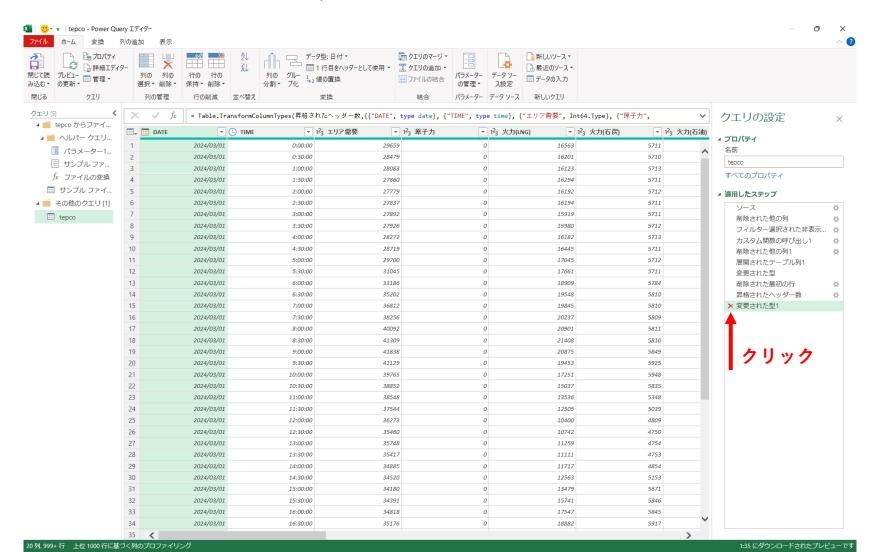


1行目が削除されます。削除された後の先頭行をヘッダーとして使用します。「1行目をヘッダーとして使用」をクリックします。



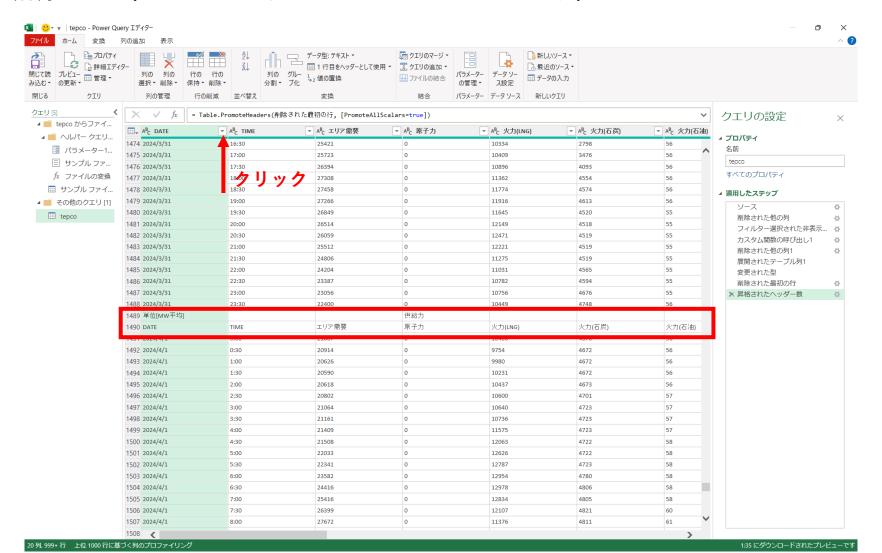


先頭がヘッダー行となりました. 望まない型変更が行われますので, 右の「クエリの設定」から「変更された型1」の左のX印をクリックして削除します.



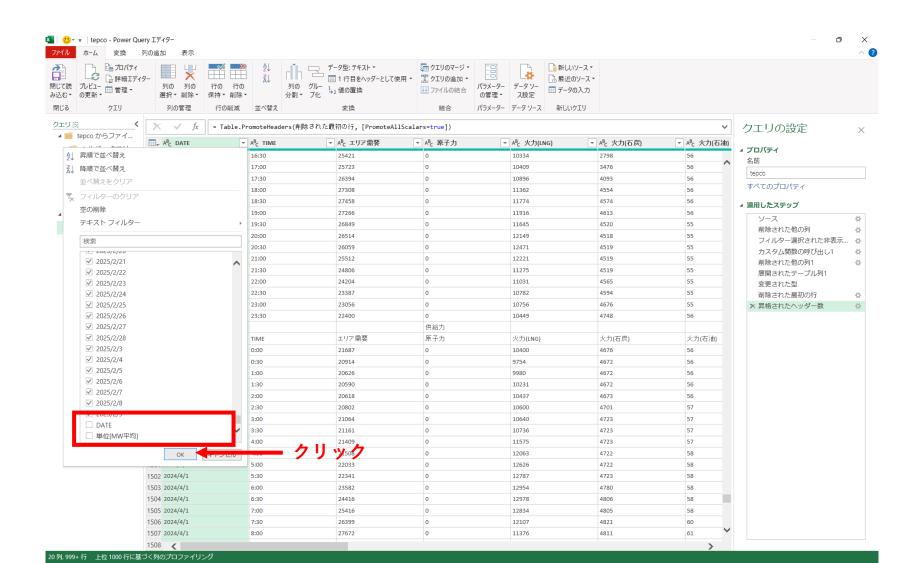


下の方へとスクロールしてみると、3月と4月のファイルの境目のところでまだ不要な情報が含まれています。次に、これらを削除します。"DATE"の右のボタンをクリックします。



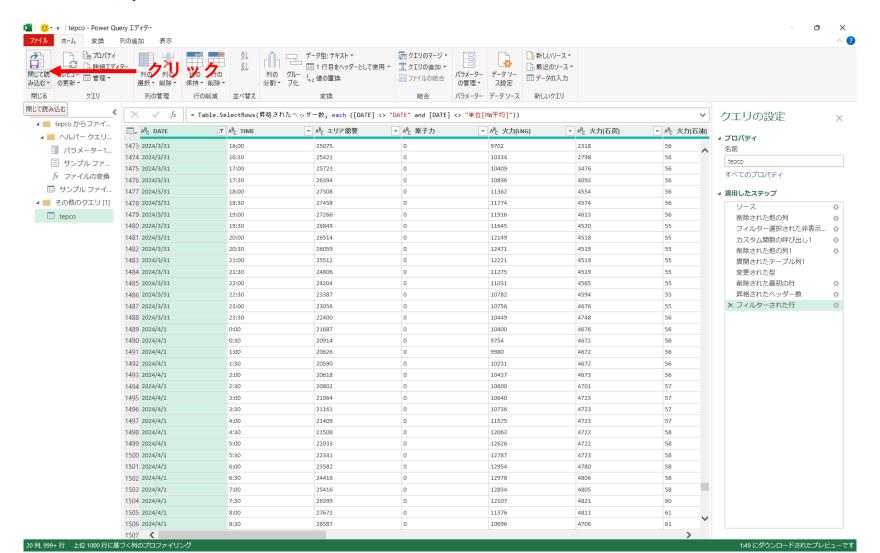


"DATE"と"単位[MW平均]"という行は不要ですので、チェックを外して、「OK」をクリックします.





**きれいにデータ部分のみが残りました!**これでPower Queryエディターによる作業は終わりです。「閉じて読み込む」をクリックします。

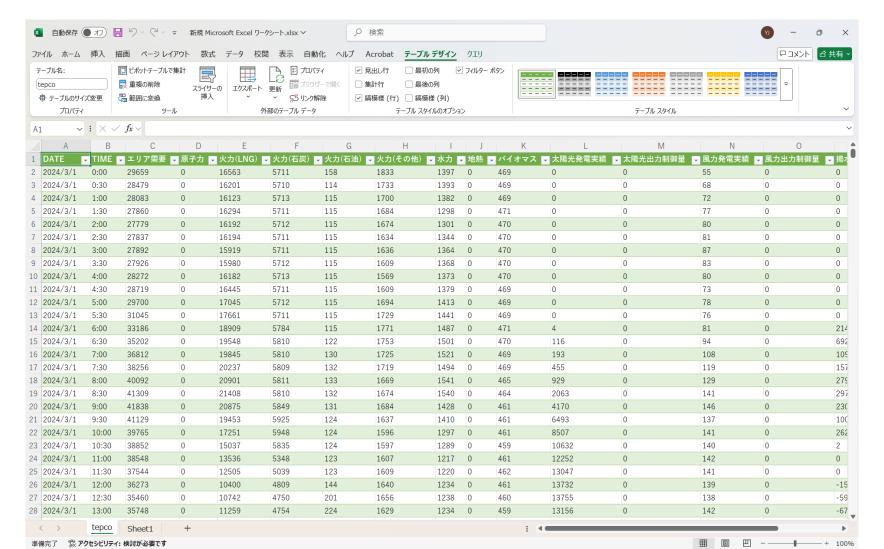


#### データの完成



2024年3月から2025年2月までの1年間分のデータが完成しました。このデータを使って分析しましょう!

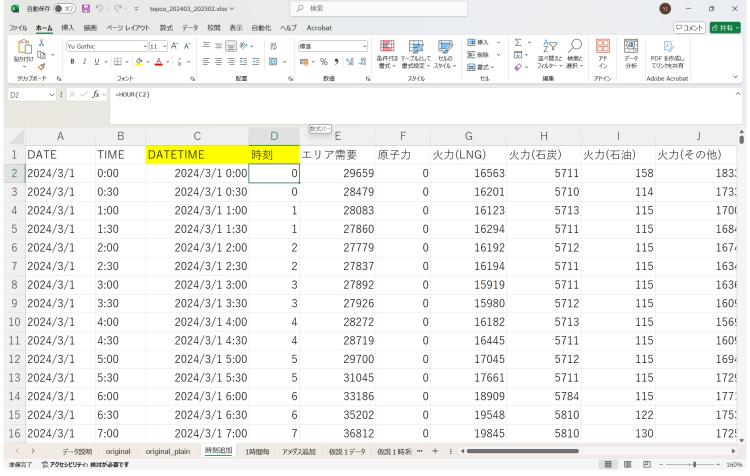
まず,電力データ「tepco\_202403\_202502.xlsx」を開いてみましょう.



## データの前処理



まず、このデータは30分ごとの時間間隔のデータになっていますが、1時間ごとのデータに変換しましょう。 そのために、年月日時分(DATETIME)と時刻の列を作ります



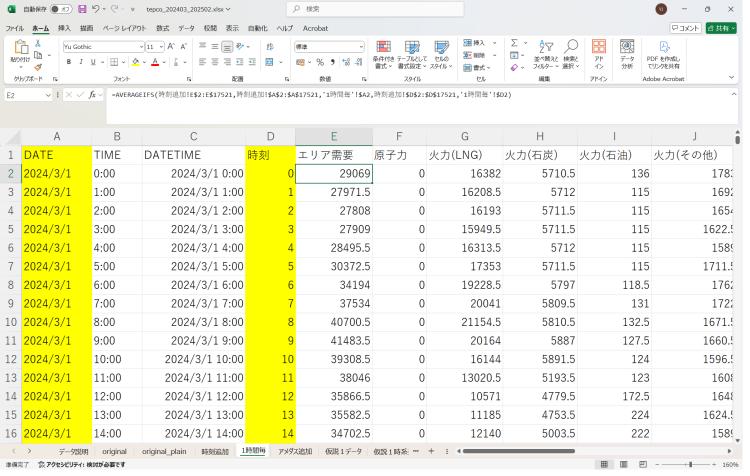
5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています

## データの前処理



日付と時刻が一致する2つのデータの平均を取ります. これにより1時間毎のデータになりました.

=AVERAGEIFS(時刻追加!E\$2:E\$17521,時刻追加!\$A\$2:\$A\$17521,'1時間毎'!\$A2,時刻追加!\$D\$2:\$D\$17521,'1時間毎'!\$D2)
平均対象(各変数) 条件1の範囲(日付) 条件1(日付) 条件2の範囲(時刻) 条件2(時刻)

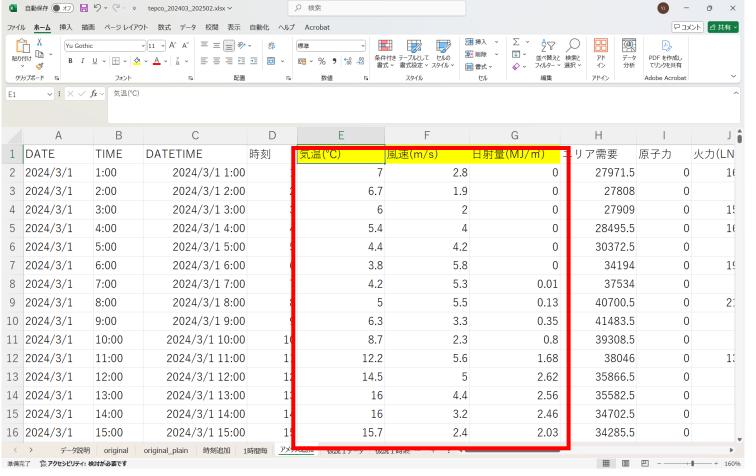


5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています

## データの前処理



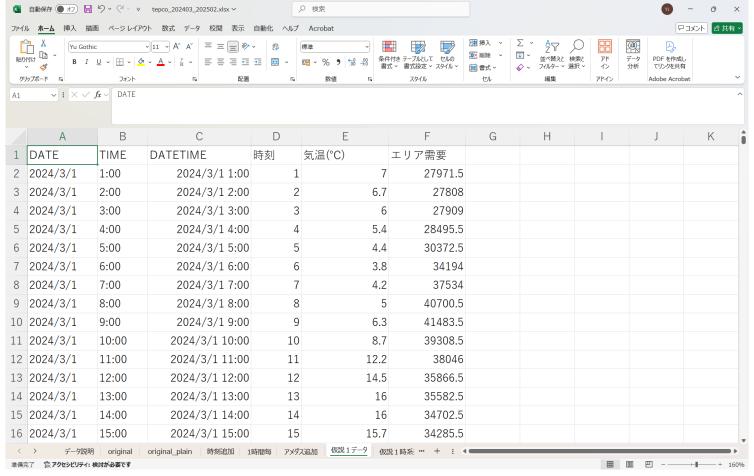
次に、気象庁アメダス(東京)の同じ期間のデータから気温、風速、日射量のデータをコピーします。気象庁アメダスのデータは3/1 1:00から、東京電力のデータは3/1 0:00からはじまります。データの範囲が1時間ずれていることに気を付けながらコピーします。



5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



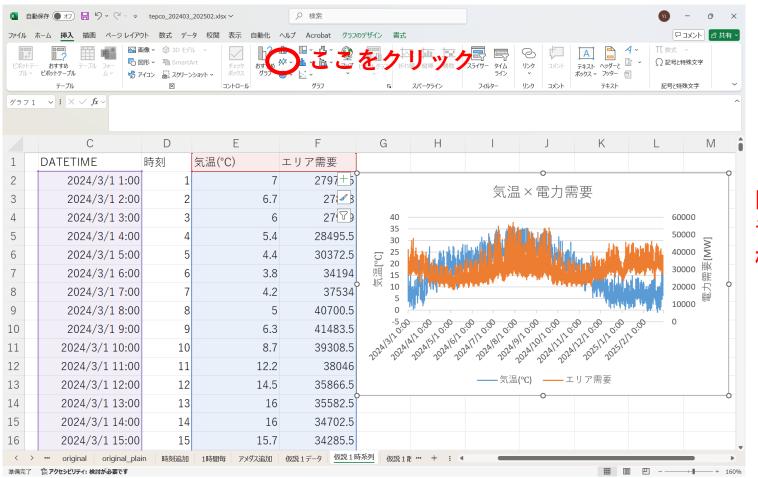
ここで、東京電力の電力消費量と東京の気温との間には関係性があるのではないかと仮説を立てて分析をして みましょう、気温とエリア需要以外は削除します。



5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



日付(C列) 気温(E列) 電力(F列) を選択して、2つの時系列を重ねて表示しましょう。 電力の時系列図には第2軸を表示させましょう。



時系列図が完成した 青色が気温 橙色が電力の時系列

5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています

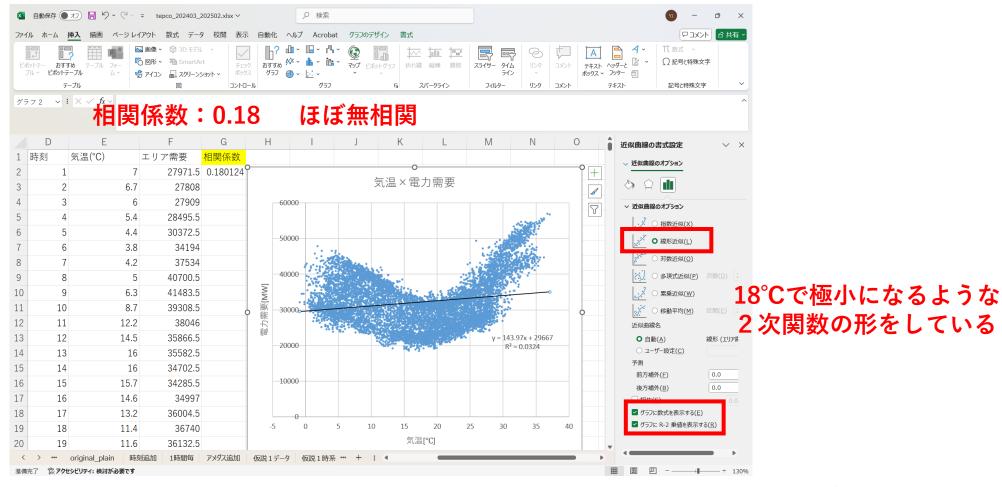


気温(E列)電力(F列)を選択して、相関係数を計算し、散布図を作成しましょう.

=CORREL(E2:E8760,F2:F8760)

気温

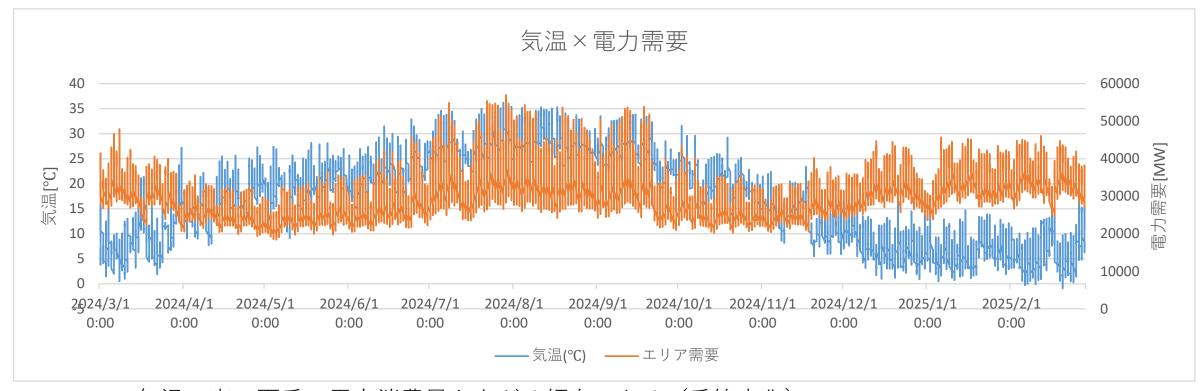
電力



5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



#### 時系列図から読み取れることをぼんやりと考えてみましょう.

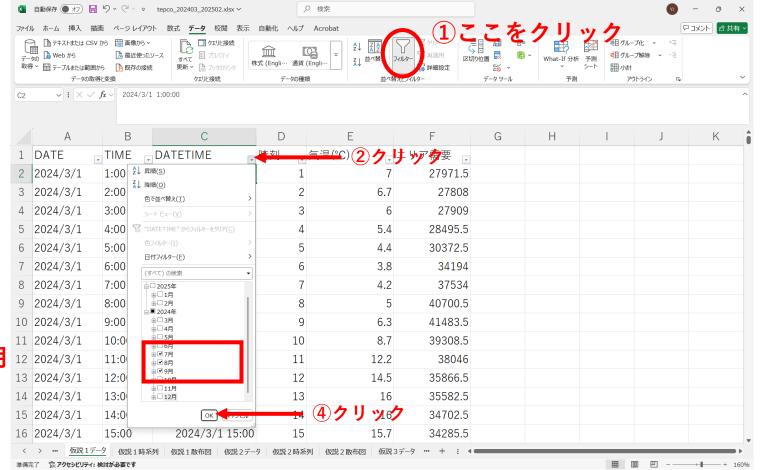


- ・気温の高い夏季に電力消費量も上がる傾向にある(季節変化)
- ・気温の低い冬季になると電力消費量が再び上昇する傾向にある(季節変化)
- ・気温の高い昼間と気温の低い夜間の電力消費量の差が大きい(特に夏季に)(日変化)
- ・気温では見られないが、電力消費量は1週間スケールでも周期的に変化している(週変化)

ここでの気づきが、以降のデータ分析の仮説を立てるうえで重要です.



次に、先の考察を踏まえて、「夏季と冬季に区別して分析すればより明瞭な関係性が得られる」と新たな仮説を立てます。フィルターを使ってデータを7月~9月の3ヶ月間に限定します。



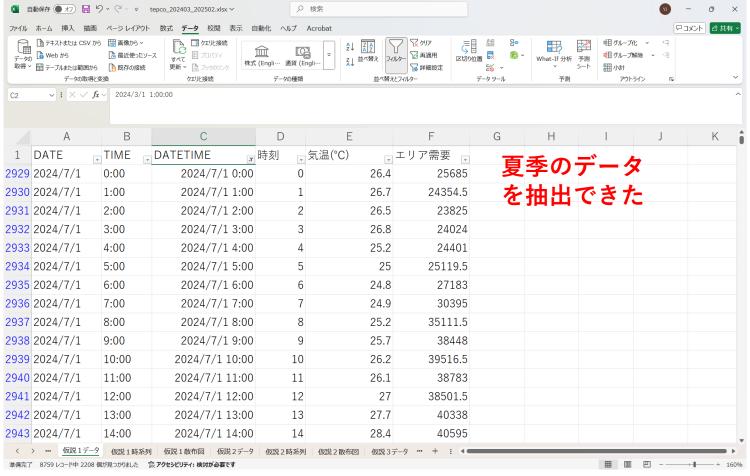
③ 7月,8月,9月 の3ヶ月間に チェック

5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



#### 7~9月のデータのみを抽出することができた.

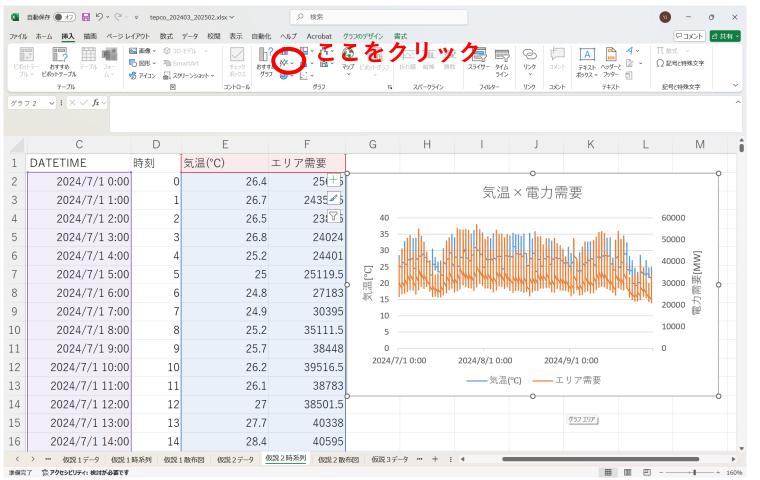
その他データは非表示になっただけで消えたわけではないことに注意. 別のワークシートにコピー&ペーストしましょう.



5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



日付(C列) 気温(E列) 電力(F列) を選択して、2つの時系列を重ねて表示しましょう。 電力の時系列図には第2軸を表示させましょう。



**2つの時系列を表示できた**.

5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています

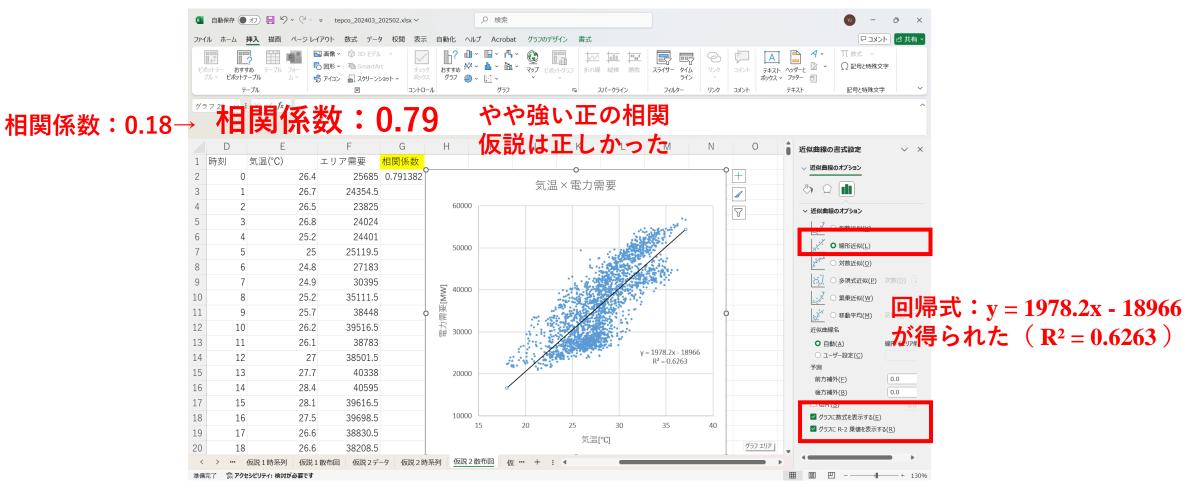


気温(E列)電力(F列)を選択して、相関係数を計算し、散布図を作成しましょう.

=CORREL(E2:E2209,F2:F2209)

気温

電力



5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



次に、先の考察を踏まえて、「土曜日と日曜日のデータを取り除いて平日のみを分析すればより明瞭な関係性が得られる」と新たな仮説を立てます。まず、WEEKDAY関数で行ごとに曜日を判定します。

=WEEKDAY(A2,

1)

DATE 1:日曜日~7:土曜日

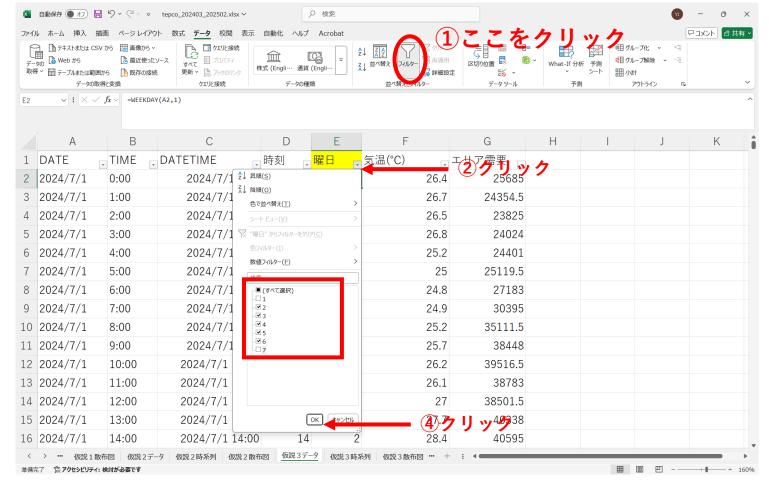
DATE(A列)から 曜日を判定できた

	ル 〜 ビボットテーブル <b>テーブル</b>	A > 3	アイコン 🖥 スクリーンショット v ボックス 図 コントロール	Ø57 <b>●</b> ~	<u>ド</u> ・	スパークラ	ライン <b>フィルター</b>	リンク コメント	テキスト ヘッダーと ボックス ~ フッター デキスト	민	と特殊文字
JN	1 v : X v	$f_{x}$ =WEEKDA	Y(A2,1)								
	A	В	С	D	Е	F	G	Н		J	K
							エリア需要			-	
2	2024/7/1	0:00	2024/7/1 0:00		=WEEKD	_	25685				
3	2024/7/1	1:00	2024/7/1 1:00	1	WEEKDAY(>	リアル値, [種類]) 26.7	24354.5				
1	2024/7/1	2:00	2024/7/1 2:00	2	2	26.5	23825				
)	2024/7/1	3:00	2024/7/1 3:00	3	2	26.8	24024	Ŋ:	ラフ エリア		
ò	2024/7/1	4:00	2024/7/1 4:00	4	2	25.2	24401				
7	2024/7/1	5:00	2024/7/1 5:00	5	2	25	25119.5				
3	2024/7/1	6:00	2024/7/1 6:00	6	2	24.8	27183				
)	2024/7/1	7:00	2024/7/1 7:00	7	2	24.9	30395				
0	2024/7/1	8:00	2024/7/1 8:00	8	2	25.2	35111.5				
1	2024/7/1	9:00	2024/7/1 9:00	9	2	25.7	38448				
2	2024/7/1	10:00	2024/7/1 10:00	10	2	26.2	39516.5				
3	2024/7/1	11:00	2024/7/1 11:00	11	2	26.1	38783				
4	2024/7/1	12:00	2024/7/1 12:00	12	2	27	38501.5				
5	2024/7/1	13:00	2024/7/1 13:00	13	2	27.7	40338				
6	2024/7/1	14:00	2024/7/1 14:00	14	2	28.4	40595				

5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



フィルターを使ってデータを日曜日(1)と土曜日(7)以外の平日のデータに限定します.



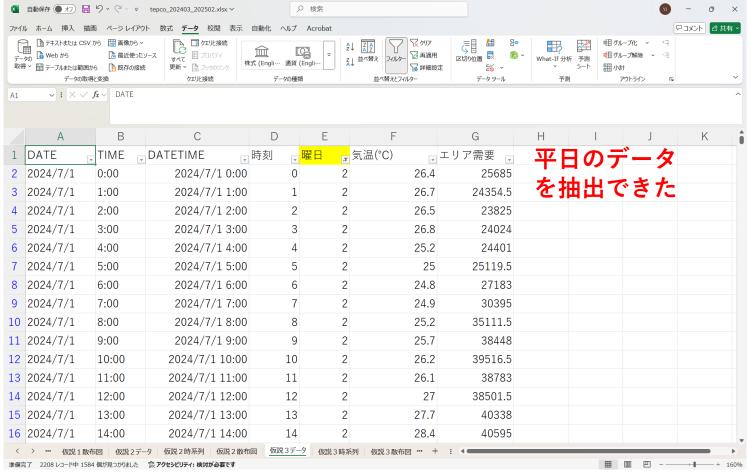
③ 月(2),火(3), 水(4),木(5), 金(6) にチェック

5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



#### 「7~9月」「平日」のデータのみを抽出することができた.

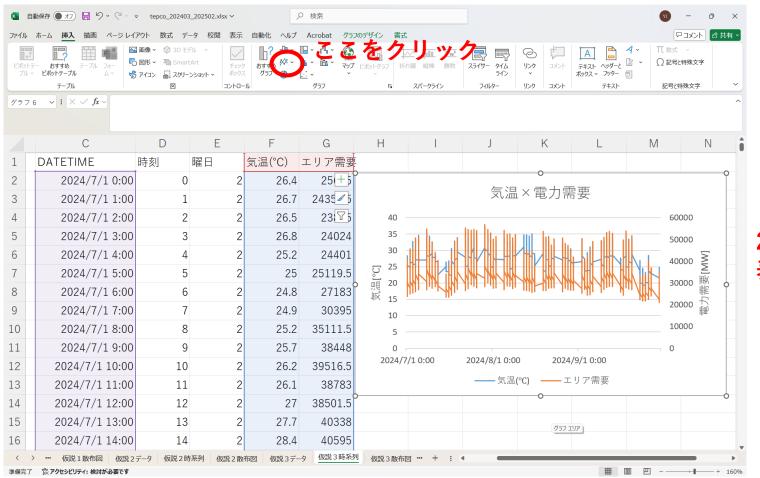
その他データは非表示になっただけで消えたわけではないことに注意. 別のワークシートにコピー&ペーストしましょう.



5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



日付(C列) 気温(F列) 電力(G列) を選択して、2つの時系列を重ねて表示しましょう。 電力の時系列図には第2軸を表示させましょう。



2つの時系列を 表示できた.

5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています

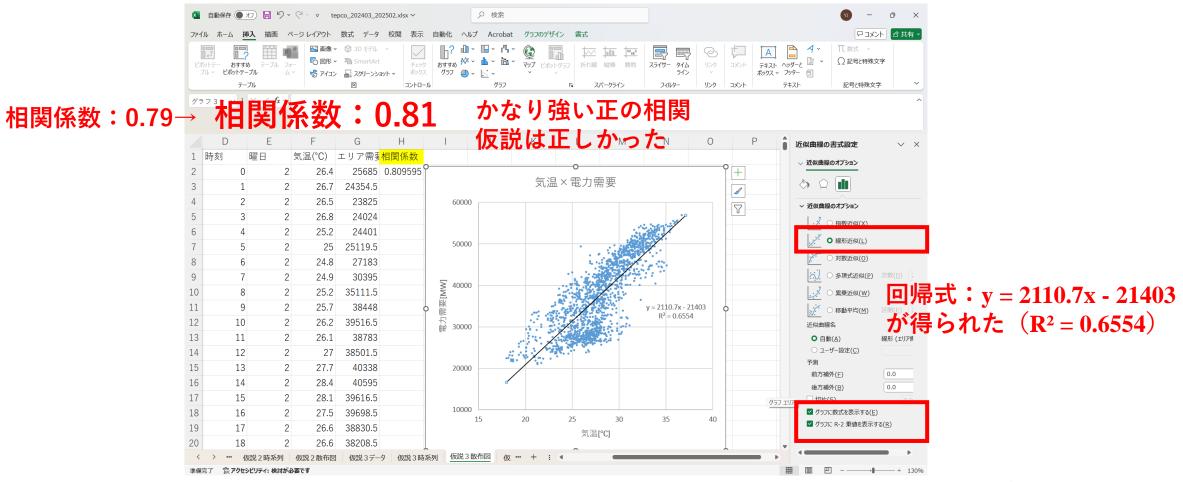


気温(F列)電力(G列)を選択して、相関係数を計算し、散布図を作成しましょう.

=CORREL(F2:F261,G2:G261)

気温

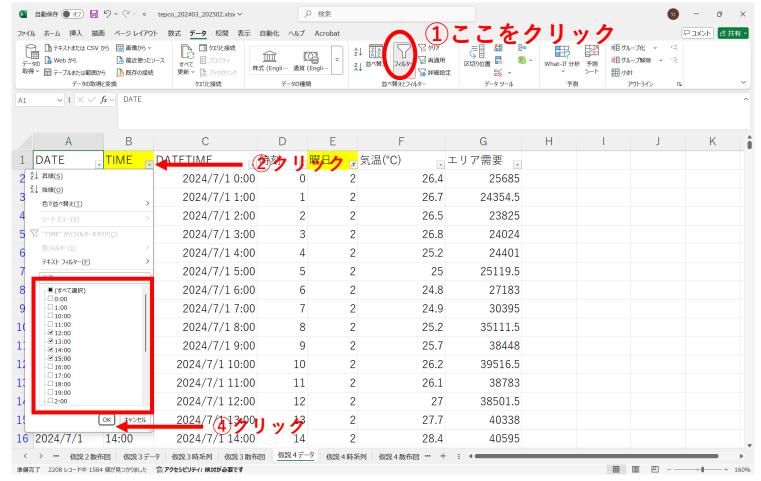
電力



5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



次に、先の考察を踏まえて、「気温が最も高くなる12~15時のみを抽出して分析すればより明瞭な関係性が得られる」と新たな仮説を立てます。フィルターを使ってデータを12~15時の4時間に限定します。



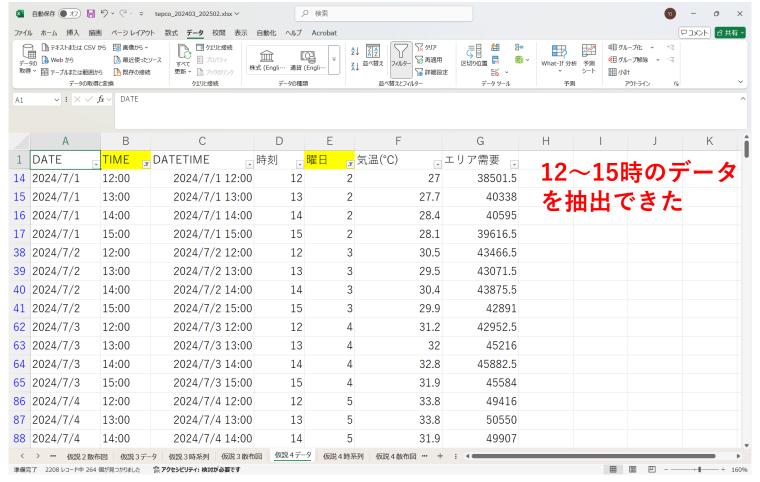
③ 12~15時 の 4 時間分に チェック

5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



#### 「 $7\sim9$ 月」「平日」「 $12\sim15$ 時」のデータのみを抽出することができた。

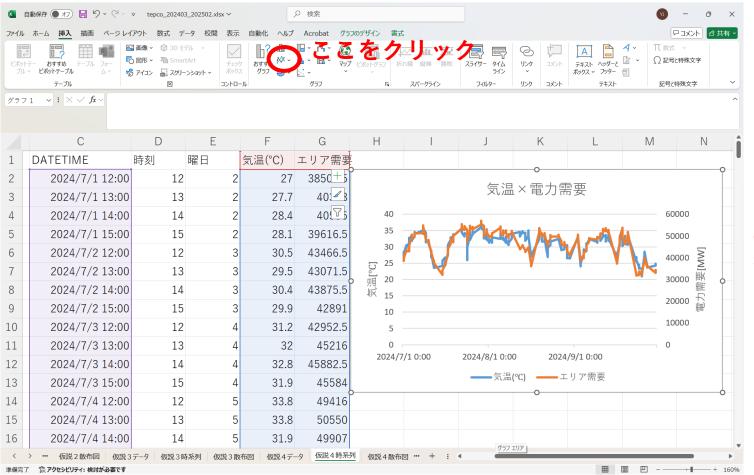
その他データは非表示になっただけで消えたわけではないことに注意. 別のワークシートにコピー&ペーストしましょう.



5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



日付(C列) 気温(F列) 電力(G列) を選択して、2つの時系列を重ねて表示しましょう。 電力の時系列図には第2軸を表示させましょう。



2つの時系列を 表示できた.

5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています

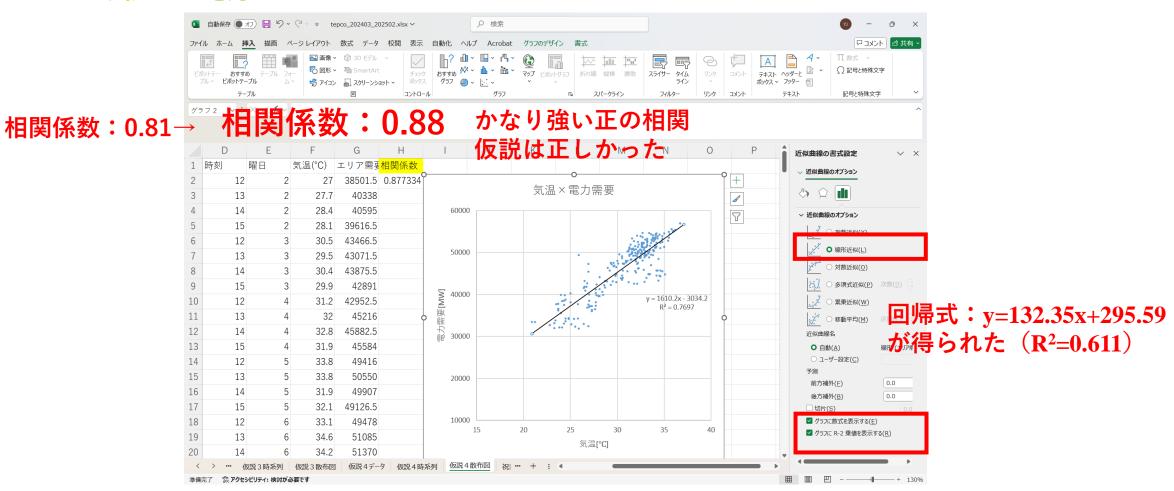


気温(F列)電力(G列)を選択して,相関係数を計算し,散布図を作成しましょう.

=CORREL(F2:F369,G2:G369)

気温

電力

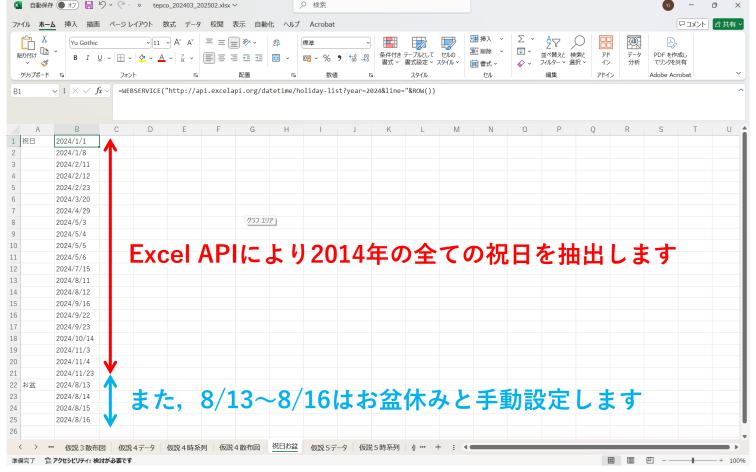


5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



次に、先の考察を踏まえて、「祝日(7/15, 8/11, 8/12, 9/16, 9/23)と夏季休暇( $8/13\sim8/16$ )のデータを取り除いて分析すればより明瞭な関係性が得られる」と新たな仮説を立てます。

=WEBSERVICE("http://api.excelapi.org/datetime/holiday-list?year=2024&line="&ROW())

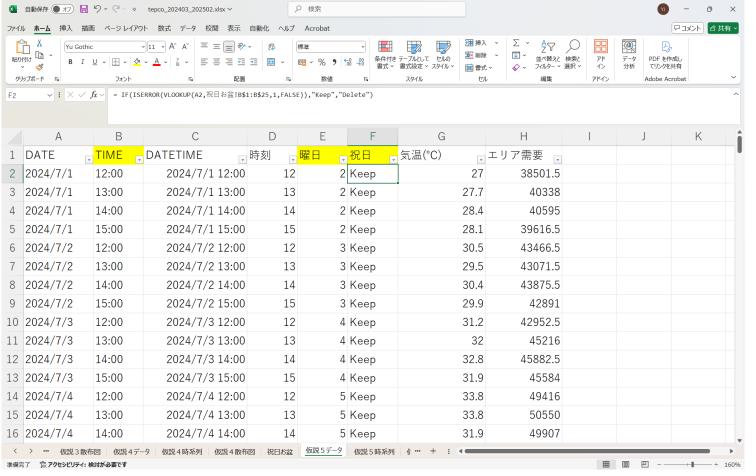


5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



祝日やお盆の日付(前のシートのB列)と日付(このシートのA列)がvlookup関数により合致しなかった場合 (検索エラーが発生)は"Keep"とし、合致した場合(検索エラーが非発生)は"Delete"とします。

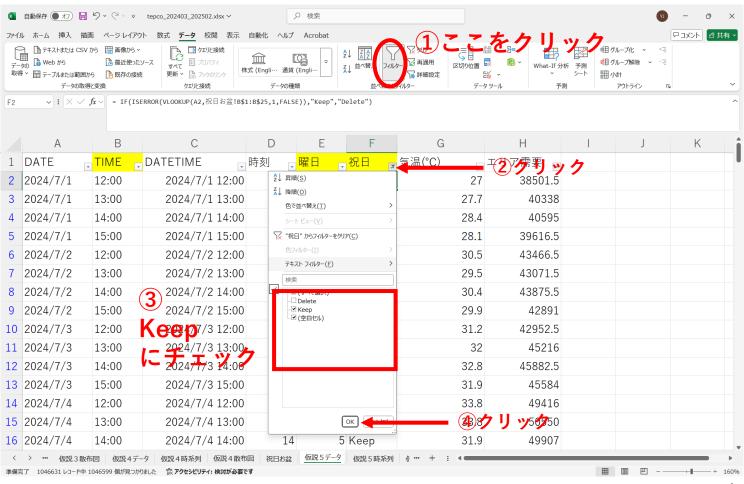
= IF(ISERROR(VLOOKUP(A2,祝日お盆!B\$1:B\$25,1,FALSE)),"Keep","Delete")



5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



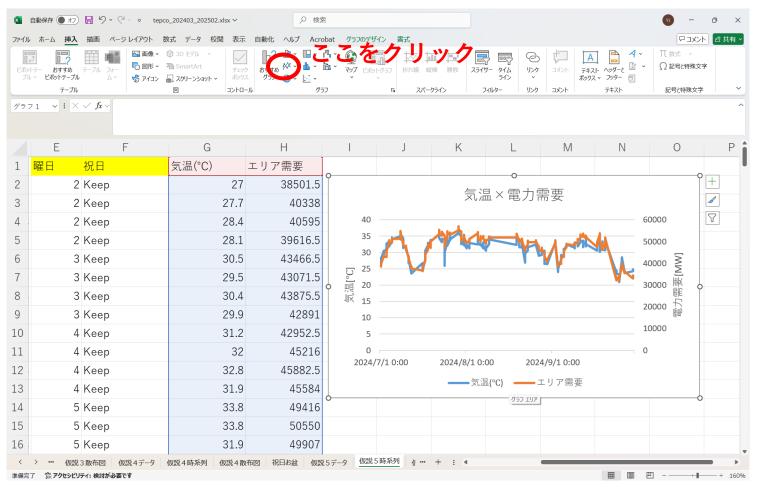
フィルターを使ってデータをDeleteの行(祝日、お盆休暇)を削除します.



5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



日付(C列) 気温(G列) 電力(H列) を選択して、2つの時系列を重ねて表示しましょう。 電力の時系列図には第2軸を表示させましょう。



2つの時系列を 表示できた. 2つの時系列は よく似た変化傾 向を示している.

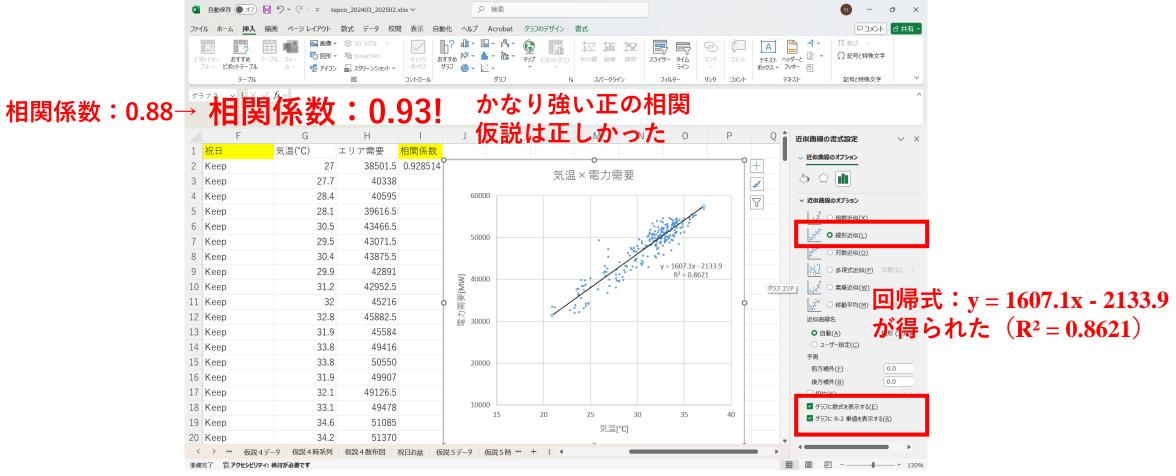
5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています



気温(G列)電力(H列)を選択して、相関係数を計算し、散布図を作成しましょう.

=CORREL(G2:G213,H2:H213)

気温 電力



5\_data/tepco\_202403\_202502.xlsxに一連の処理が記録されています

## やってみよう2



- 気温と電力消費量との間には関係があり、正確に電力消費量を 予測するためには、さらに、季節、時間、曜日等を考慮する必 要があることがわかりました。
- 冬季のデータを使って気温と電力消費量との関係性を調べてみましょう。
- 風力発電と風速,太陽光発電と日射量の関係性を調べてみましょう.

