## import pandas as pd

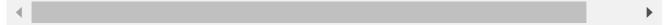
import numpy as np

!gdown 173A59xh2mnpmljCCB9bhC4C5eP2IS6qZ

## Downloading...

From: <a href="https://drive.google.com/uc?id=173A59xh2mnpmljCCB9bhC4C5eP2IS6qZ">https://drive.google.com/uc?id=173A59xh2mnpmljCCB9bhC4C5eP2IS6qZ</a>

To: /Users/nikhilsanghi/Downloads/dsml-course-main-live/batches/May-Beg-Aug-Adv/Pfize 100%| 1.51k/1.51k [00:00<00:00, 1.83MB/s]



df=pd.read\_csv("Pfizer\_1.csv")
df

		Date	Drug_Name	Parameter	1:30:00	2:30:00	3:30:00	4:30:00	5:30:00	6:30:
	0	15- 10- 2020	diltiazem hydrochloride	Temperature	23.0	22.0	NaN	21.0	21.0	
	1	15- 10- 2020	diltiazem hydrochloride	Pressure	12.0	13.0	NaN	11.0	13.0	
	•	15-	docetaxel	÷ .	A.I. A.I.	47.0	400	A.1 A.1	470	
df_me	lt=p	d.melt	(df,							
		_	["Date","Drug	g_Name","Para	meter"],					
		var_name="Time",								
٠		i_ue_na	me="Result")							
df_me	ΤT									

	Date	Drug_Name	Parameter	Time	Result
0	15-10-2020	diltiazem hydrochloride	Temperature	1:30:00	23.0
1	15-10-2020	diltiazem hydrochloride	Pressure	1:30:00	12.0
2	15-10-2020	docetaxel injection	Temperature	1:30:00	NaN
3	15-10-2020	docetaxel injection	Pressure	1:30:00	NaN
4	15-10-2020	ketamine hydrochloride	Temperature	1:30:00	24.0
211	17-10-2020	diltiazem hydrochloride	Pressure	12:30:00	14.0
212	17-10-2020	docetaxel injection	Temperature	12:30:00	23.0
213	17-10-2020	docetaxel injection	Pressure	12:30:00	28.0
214	17-10-2020	ketamine hydrochloride	Temperature	12:30:00	24.0
215	17-10-2020	ketamine hydrochloride	Pressure	12:30:00	15.0

216 rows × 5 columns

Time 10:30:00 11:30:00 12:30:00 1:30:00 2:30:00 3:30

Date	Drug_Name	Parameter						
15-	diltiazem	Pressure	18.0	19.0	20.0	12.0	13.0	
10- 2020	hydrochloride	Temperature	20.0	20.0	21.0	23.0	22.0	
	docetaxel	Pressure	26.0	29.0	28.0	NaN	22.0	
	injection	Temperature	23.0	25.0	25.0	NaN	17.0	
	ketamine	Pressure	9.0	9.0	11.0	8.0	NaN	
	hydrochloride	Temperature	22.0	21.0	20.0	24.0	NaN	
16- 10-	diltiazem hydrochloride	Pressure	24.0	NaN	27.0	18.0	19.0	
2020		Temperature	40.0	NaN	42.0	34.0	35.0	
	docetaxel	Pressure	28.0	29.0	30.0	23.0	24.0	
	injection	Temperature	56.0	57.0	58.0	46.0	47.0	
	ketamine	Pressure	16.0	17.0	18.0	12.0	12.0	
	hydrochloride	Temperature	13.0	14.0	15.0	8.0	9.0	
17-	diltiazem	Pressure	11.0	13.0	14.0	3.0	4.0	

df\_pivoted.reset\_index()

Time	Date	Drug_Name	Parameter	10:30:00	11:30:00	12:30:00	1:30:00	2:30:00
0	15- 10- 2020	diltiazem hydrochloride	Pressure	18.0	19.0	20.0	12.0	13.0
1	15- 10- 2020	diltiazem hydrochloride	Temperature	20.0	20.0	21.0	23.0	22.0
2	15- 10- 2020	docetaxel injection	Pressure	26.0	29.0	28.0	NaN	22.0
3	15- 10- 2020	docetaxel injection	Temperature	23.0	25.0	25.0	NaN	17.0
4	15- 10- 2020	ketamine hydrochloride	Pressure	9.0	9.0	11.0	8.0	NaN
	15-	katamina						

df\_melt

	Date	Drug_Name	Parameter	Time	Result
0	15-10-2020	diltiazem hydrochloride	Temperature	1:30:00	23.0
1	15-10-2020	diltiazem hydrochloride	Pressure	1:30:00	12.0
2	15-10-2020	docetaxel injection	Temperature	1:30:00	NaN
3	15-10-2020	docetaxel injection	Pressure	1:30:00	NaN
4	15-10-2020	ketamine hydrochloride	Temperature	1:30:00	24.0
211	17-10-2020	diltiazem hydrochloride	Pressure	12:30:00	14.0
212	17-10-2020	docetaxel injection	Temperature	12:30:00	23.0
213	17-10-2020	docetaxel injection	Pressure	12:30:00	28.0
214	17-10-2020	ketamine hydrochloride	Temperature	12:30:00	24.0
215	17-10-2020	ketamine hydrochloride	Pressure	12:30:00	15.0

216 rows × 5 columns

df\_tidy

Parameter	Date	Drug_Name	Time	Pressure	Temperature
0	15-10-2020	diltiazem hydrochloride	10:30:00	18.0	20.0
1	15-10-2020	diltiazem hydrochloride	11:30:00	19.0	20.0
2	15-10-2020	diltiazem hydrochloride	12:30:00	20.0	21.0
3	15-10-2020	diltiazem hydrochloride	1:30:00	12.0	23.0
4	15-10-2020	diltiazem hydrochloride	2:30:00	13.0	22.0
103	17-10-2020	ketamine hydrochloride	5:30:00	11.0	17.0
104	17-10-2020	ketamine hydrochloride	6:30:00	12.0	18.0
105	17-10-2020	ketamine hydrochloride	7:30:00	12.0	19.0
106	17-10-2020	ketamine hydrochloride	8:30:00	11.0	20.0
107	17-10-2020	ketamine hydrochloride	9:30:00	12.0	21.0

108 rows × 5 columns

df\_tidy.isna().sum()

Parameter
Date 0
Drug\_Name 0
Time 0
Pressure 13
Temperature 13
dtype: int64

df\_tidy[df\_tidy["Drug\_Name"]=="diltiazem hydrochloride"]

P	Parameter	Date	Drug_Name	Time	Pressure	Temperature
	0	15-10-2020	diltiazem hydrochloride	10:30:00	18.0	20.0
	1	15-10-2020	diltiazem hydrochloride	11:30:00	19.0	20.0
	2	15-10-2020	diltiazem hydrochloride	12:30:00	20.0	21.0
	3	15-10-2020	diltiazem hydrochloride	1:30:00	12.0	23.0
	4	15-10-2020	diltiazem hydrochloride	2:30:00	13.0	22.0
	5	15-10-2020	diltiazem hydrochloride	3:30:00	NaN	NaN
	6	15-10-2020	diltiazem hydrochloride	4:30:00	11.0	21.0
	7	15-10-2020	diltiazem hydrochloride	5:30:00	13.0	21.0
	8	15-10-2020	diltiazem hydrochloride	6:30:00	14.0	22.0
	9	15-10-2020	diltiazem hydrochloride	7:30:00	16.0	23.0
	10	15-10-2020	diltiazem hydrochloride	8:30:00	16.0	21.0
	11	15-10-2020	diltiazem hydrochloride	9:30:00	24.0	22.0
	36	16-10-2020	diltiazem hydrochloride	10:30:00	24.0	40.0
	37	16-10-2020	diltiazem hydrochloride	11:30:00	NaN	NaN
	38	16-10-2020	diltiazem hydrochloride	12:30:00	27.0	42.0
	39	16-10-2020	diltiazem hydrochloride	1:30:00	18.0	34.0
	40	16-10-2020	diltiazem hydrochloride	2:30:00	19.0	35.0
	41	16-10-2020	diltiazem hydrochloride	3:30:00	20.0	36.0
	42	16-10-2020	diltiazem hydrochloride	4:30:00	21.0	36.0
	43	16-10-2020	diltiazem hydrochloride	5:30:00	22.0	37.0
	44	16-10-2020	diltiazem hydrochloride	6:30:00	23.0	38.0
	45	16-10-2020	diltiazem hydrochloride	7:30:00	24.0	37.0
	46	16-10-2020	diltiazem hydrochloride	8:30:00	25.0	38.0
	47	16-10-2020	diltiazem hydrochloride	9:30:00	25.0	39.0
	72	17-10-2020	diltiazem hydrochloride	10:30:00	11.0	14.0
	73	17-10-2020	diltiazem hydrochloride	11:30:00	13.0	11.0
	74	17-10-2020	diltiazem hydrochloride	12:30:00	14.0	10.0

df\_tidy.groupby("Drug\_Name")["Temperature"].mean()

Drug\_Name

diltiazem hydrochloride 24.848485 docetaxel injection 30.387097 ketamine hydrochloride 17.709677 Name: Temperature, dtype: float64

```
df tidy.groupby("Drug Name")["Pressure"].mean()
     Drug Name
     diltiazem hydrochloride
                                15.424242
     docetaxel injection
                                25.483871
     ketamine hydrochloride
                                11.935484
     Name: Pressure, dtype: float64
df1=pd.DataFrame([[np.nan, 2, np.nan, 0],
                   [3, 4, np.nan, 1],
                   [np.nan, np.nan, np.nan],
                   [np.nan, 3, np.nan, 4]],
                  columns=["A","B","C","D"])
df1
                       C
           Α
                 В
                             D
                2.0
                           0.0
      0
         NaN
                    NaN
      1
          3.0
               4.0
                    NaN
                           1.0
         NaN
              NaN
                    NaN
                          NaN
      3
         NaN
                3.0 NaN
                           4.0
df1["C"].fillna(df1["C"].mean())
     0
         NaN
     1
         NaN
     2
         NaN
         NaN
     Name: C, dtype: float64
df_tidy["Pressure"]=df_tidy.groupby("Drug_Name")["Pressure"].transform(lambda x:x.fillna(x
df_tidy["Temperature"]=df_tidy.groupby("Drug_Name")["Temperature"].transform(lambda x:x.fi
df tidy.isna().sum()
     Parameter
                    0
     Date
     Drug Name
                    0
     Time
     Pressure
                    0
     Temperature
     dtype: int64
df_tidy
```

Par	ameter	Date	Drug_Name	Time	Pressure	Temperature
	0	15-10-2020	diltiazem hydrochloride	10:30:00	18.0	20.0
	1	15-10-2020	diltiazem hydrochloride	11:30:00	19.0	20.0
	2	15-10-2020	diltiazem hydrochloride	12:30:00	20.0	21.0
	3	15-10-2020	diltiazem hydrochloride	1:30:00	12.0	23.0
	4	15-10-2020	diltiazem hydrochloride	2:30:00	13.0	22.0
	103	17-10-2020	ketamine hydrochloride	5:30:00	11.0	17.0
	104	17-10-2020	ketamine hydrochloride	6:30:00	12.0	18.0
	105	17-10-2020	ketamine hydrochloride	7:30:00	12.0	19.0
	106	17-10-2020	ketamine hydrochloride	8:30:00	11.0	20.0
	107	17-10-2020	ketamine hydrochloride	9:30:00	12.0	21.0

Parameter	Pressure			Temperature		
Date	15-10- 2020	16-10- 2020	17-10- 2020	15-10- 2020	16-10- 2020	17-10- 2020
Drug_Name						
diltiazem hydrochloride	11.0	15.424242	3.0	20.000000	24.848485	
docetaxel injection	22.0	23.000000	20.0	17.000000	30.387097	

df\_tidy.pivot\_table?

Parameter	Pressure			Temperature			
Date	15-10- 2020	16-10- 2020	17-10- 2020	15-10- 2020	16-10- 2020	17-10- 2020	
Drug_Name							
diltiazem hydrochloride	24.000000	27.0	15.424242	24.848485	42.000000	24.84{	
docetaxel injection	29.000000	30.0	29.000000	30.387097	58.000000	23.000	

	Parameter Pressure				Temperature			
Drug_Name		diltiazem hydrochloride	docetaxel injection	ketamine hydrochloride	diltiazem hydrochloride	docetaxel injection	ketami hydroc	
	Date	Date						
	15-10- 2020	11.000000	22.0	7.000000	20.000000	17.000000	17	
	16-10- 2020	15.424242	23.0	11.935484	24.848485	30.387097	{	
	17-10- 2020	3.000000	20.0	8.000000	10.000000	12.000000	10	

	Parameter	Pressure				Temperature		
Drug_Name		diltiazem hydrochloride	docetaxel injection	ketamine hydrochloride	All	diltiazem hydrochloride		
	Date							
	15-10- 2020	15.952020	25.077957	9.983871	17.004616	21.737374		
	16-10- 2020	21.952020	26.873656	14.322581	21.049419	36.404040		
	17-10- 2020	8.368687	24.500000	11.500000	14.789562	16.404040		
	AII	15.424242	25.483871	11.935484	17.614532	24.848485		

df\_tidy[(df\_tidy["Drug\_Name"]=="diltiazem hydrochloride")&(df\_tidy["Date"]=="15-10-2020")]

	Parameter	Date	Drug_Name	Time	Pressure	Temperature
	0	15-10-2020	diltiazem hydrochloride	10:30:00	18.000000	20.000000
	1	15-10-2020	diltiazem hydrochloride	11:30:00	19.000000	20.000000
	2	15-10-2020	diltiazem hydrochloride	12:30:00	20.000000	21.000000
	3	15-10-2020	diltiazem hydrochloride	1:30:00	12.000000	23.000000
	4	15-10-2020	diltiazem hydrochloride	2:30:00	13.000000	22.000000
	5	15-10-2020	diltiazem hydrochloride	3:30:00	15.424242	24.848485
- <b>:</b>	dv					

df\_tidy

Parameter	Date	Drug_Name	Time	Pressure	Temperature
0	15-10-2020	diltiazem hydrochloride	10:30:00	18.0	20.0
1	15-10-2020	diltiazem hydrochloride	11:30:00	19.0	20.0
2	15-10-2020	diltiazem hydrochloride	12:30:00	20.0	21.0
3	15-10-2020	diltiazem hydrochloride	1:30:00	12.0	23.0
4	15-10-2020	diltiazem hydrochloride	2:30:00	13.0	22.0
103	17-10-2020	ketamine hydrochloride	5:30:00	11.0	17.0
104	17-10-2020	ketamine hydrochloride	6:30:00	12.0	18.0
105	17-10-2020	ketamine hydrochloride	7:30:00	12.0	19.0
106	17-10-2020	ketamine hydrochloride	8:30:00	11.0	20.0
107	17-10-2020	ketamine hydrochloride	9:30:00	12.0	21.0
103 104 105 106	17-10-2020 17-10-2020 17-10-2020 17-10-2020	ketamine hydrochloride ketamine hydrochloride ketamine hydrochloride ketamine hydrochloride	5:30:00 6:30:00 7:30:00 8:30:00	11.0 12.0 12.0 11.0	17. 18. 19. 20.

108 rows × 5 columns

labels=temp\_labels,
right=False)

df\_tidy

Parameter	Date	Drug_Name	Time	Pressure	Temperature	Temp_cat
0	15-10-2020	diltiazem hydrochloride	10:30:00	18.0	20.0	medium
1	15-10-2020	diltiazem hydrochloride	11:30:00	19.0	20.0	medium
2	15-10-2020	diltiazem hydrochloride	12:30:00	20.0	21.0	medium
3	15-10-2020	diltiazem hydrochloride	1:30:00	12.0	23.0	medium
4	15-10-2020	diltiazem hydrochloride	2:30:00	13.0	22.0	medium
103	17-10-2020	ketamine hydrochloride	5:30:00	11.0	17.0	low
104	17-10-2020	ketamine hydrochloride	6:30:00	12.0	18.0	low
105	17-10-2020	ketamine hydrochloride	7:30:00	12.0	19.0	low
106	17-10-2020	ketamine hydrochloride	8:30:00	11.0	20.0	medium
107	17-10-2020	ketamine hydrochloride	9:30:00	12.0	21.0	medium

108 rows × 6 columns

df\_tidy["Temp\_cat"].value\_counts()

low 44 medium 43 high 15 very\_high 6

Name: Temp\_cat, dtype: int64

df\_tidy.loc[df\_tidy["Drug\_Name"].str.contains("hydrochloride")]

Parameter	Date	Drug_Name	Time	Pressure	Temperature	Temp_cat
0	15-10-2020	diltiazem hydrochloride	10:30:00	18.0	20.0	medium
1	15_10_2020	diltiazem hydrochloride	11.30.00	10 N	20.0	medium

df\_tidy["Timestamp"]=df\_tidy["Date"]+" "+df\_tidy["Time"]
df\_tidy

Parameter	Date	Drug_Name	Time	Pressure	Temperature	Temp_cat
0	15-10-2020	diltiazem hydrochloride	10:30:00	18.0	20.0	medium
1	15-10-2020	diltiazem hydrochloride	11:30:00	19.0	20.0	medium
2	15-10-2020	diltiazem hydrochloride	12:30:00	20.0	21.0	medium
3	15-10-2020	diltiazem hydrochloride	1:30:00	12.0	23.0	medium
4	15-10-2020	diltiazem hydrochloride	2:30:00	13.0	22.0	medium
103	17-10-2020	ketamine hydrochloride	5:30:00	11.0	17.0	low
104	17-10-2020	ketamine hydrochloride	6:30:00	12.0	18.0	low
105	17-10-2020	ketamine hydrochloride	7:30:00	12.0	19.0	low
106	17-10-2020	ketamine hydrochloride	8:30:00	11.0	20.0	medium
107	17-10-2020	ketamine hydrochloride	9:30:00	12.0	21.0	medium

108 rows × 7 columns

memory usage: 5.5+ KB

df\_tidy.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 108 entries, 0 to 107
Data columns (total 7 columns):

_ 0. 0 0.	00-0		
#	Column	Non-Null Count	Dtype
0	Date	108 non-null	object
1	Drug_Name	108 non-null	object
2	Time	108 non-null	object
3	Pressure	108 non-null	float64
4	Temperature	108 non-null	float64
5	Temp_cat	108 non-null	category
6	Timestamp	108 non-null	object
dtype	es: category(	1), float64(2),	object(4)

df\_tidy["Timestamp"]=pd.to\_datetime(df\_tidy["Timestamp"])
df\_tidy

Parameter	Date	Drug_Name	Time	Pressure	Temperature	Temp_cat
0	15-10-2020	diltiazem hydrochloride	10:30:00	18.0	20.0	medium
1	15-10-2020	diltiazem hydrochloride	11:30:00	19.0	20.0	medium
2	15-10-2020	diltiazem hydrochloride	12:30:00	20.0	21.0	medium
3	15-10-2020	diltiazem hydrochloride	1:30:00	12.0	23.0	medium
4	15-10-2020	diltiazem hydrochloride	2:30:00	13.0	22.0	medium
103	17-10-2020	ketamine hydrochloride	5:30:00	11.0	17.0	low
104	17-10-2020	ketamine hydrochloride	6:30:00	12.0	18.0	low
105	17-10-2020	ketamine hydrochloride	7:30:00	12.0	19.0	low
106	17-10-2020	ketamine hydrochloride	8:30:00	11.0	20.0	medium
107	17-10-2020	ketamine hydrochloride	9:30:00	12.0	21.0	medium

100 rouge v 7 columns

df\_tidy.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 108 entries, 0 to 107

Data columns (total 7 columns):

#	Column	Non-Null Count	Dtype
0	Date	108 non-null	object
1	Drug_Name	108 non-null	object
2	Time	108 non-null	object
3	Pressure	108 non-null	float64
4	Temperature	108 non-null	float64
5	Temp_cat	108 non-null	category
6	Timestamp	108 non-null	datetime64[ns]

dtypes: category(1), datetime64[ns](1), float64(2), object(3)

memory usage: 5.5+ KB

```
ts=df_tidy["Timestamp"][0]
ts
```

Timestamp('2020-10-15 10:30:00')

ts.year

2020

ts.month

10

```
ts.day
     15
ts.month_name()
     'October'
ts.day_name()
     'Thursday'
ts.dayofweek
     3
ts.hour
     10
ts.minute
     30
ts.second
     0
df_tidy["Timestamp"].dt
     <pandas.core.indexes.accessors.DatetimeProperties object at 0x7f9fe81b7340>
df_tidy["Timestamp"].dt.year
     0
            2020
     1
            2020
     2
            2020
     3
            2020
            2020
            . . .
     103
            2020
     104
            2020
     105
            2020
     106
            2020
     107
            2020
     Name: Timestamp, Length: 108, dtype: int64
# Homework
```

# strftime

df\_tidy["year"]=df\_tidy["Timestamp"].dt.year

df\_tidy

Parameter	Date	Drug_Name	Time	Pressure	Temperature	Temp_cat
0	15-10-2020	diltiazem hydrochloride	10:30:00	18.0	20.0	medium
1	15-10-2020	diltiazem hydrochloride	11:30:00	19.0	20.0	medium
2	15-10-2020	diltiazem hydrochloride	12:30:00	20.0	21.0	medium
3	15-10-2020	diltiazem hydrochloride	1:30:00	12.0	23.0	medium
4	15-10-2020	diltiazem hydrochloride	2:30:00	13.0	22.0	medium
103	17-10-2020	ketamine hydrochloride	5:30:00	11.0	17.0	low
104	17-10-2020	ketamine hydrochloride	6:30:00	12.0	18.0	low
105	17-10-2020	ketamine hydrochloride	7:30:00	12.0	19.0	low
106	17-10-2020	ketamine hydrochloride	8:30:00	11.0	20.0	medium
107	17-10-2020	ketamine hydrochloride	9:30:00	12.0	21.0	medium

108 rows × 8 columns

```
df_tidy.to_csv("pfizer_result.csv",sep=",")
df_tidy.to_csv("pfizer_result3.csv")
```

!1s

Numpy Assessments.ipynb	Pandas-5.ipynb
Numpy-1	Pandas-6.ipynb
Numpy-2 Lecture Notes.ipynb	Pfizer_1.csv
Numpy-2.ipynb	directors.csv
Numpy-3.ipynb	dog.jpeg
Numpy-4.ipynb	fitness.txt
Numpy-5.ipynb	<pre>gapminder.csv</pre>
Pandas-1.ipynb	movies.csv
Pandas-2.ipynb	<pre>pfizer_result.csv</pre>
Pandas-3.ipynb	<pre>pfizer_result1.csv</pre>
Pandas-4.ipynb	<pre>pfizer_result2.csv</pre>

df\_tidy.to\_csv?

df\_tidy

Parameter	Date	Drug_Name	Time	Pressure	Temperature	Temp_cat
0	15-10-2020	diltiazem hydrochloride	10:30:00	18.0	20.0	medium
1	15-10-2020	diltiazem hydrochloride	11:30:00	19.0	20.0	medium
2	15-10-2020	diltiazem hydrochloride	12:30:00	20.0	21.0	medium
3	15-10-2020	diltiazem hydrochloride	1:30:00	12.0	23.0	medium
4	15-10-2020	diltiazem hydrochloride	2:30:00	13.0	22.0	medium
103	17-10-2020	ketamine hydrochloride	5:30:00	11.0	17.0	low
104	17-10-2020	ketamine hydrochloride	6:30:00	12.0	18.0	low
105	17-10-2020	ketamine hydrochloride	7:30:00	12.0	19.0	low
106	17-10-2020	ketamine hydrochloride	8:30:00	11.0	20.0	medium
107	17-10-2020	ketamine hydrochloride	9:30:00	12.0	21.0	medium

108 rows × 8 columns

df\_tidy["Timestamp"].max()

Timestamp('2020-10-17 12:30:00')

Colab paid products - Cancel contracts here