

# Day\_39\_051223

January 23, 2024

```
[85]: import pandas as pd
import numpy as np
data = pd.read_csv("Pfizer_1.csv")
```

```
[86]: data.head()
```

```
[86]:
```

	Date	Drug_Name	Parameter	1:30:00	2:30:00	\
0	15-10-2020	diltiazem hydrochloride	Temperature	23.0	22.0	
1	15-10-2020	diltiazem hydrochloride	Pressure	12.0	13.0	
2	15-10-2020	docetaxel injection	Temperature	NaN	17.0	
3	15-10-2020	docetaxel injection	Pressure	NaN	22.0	
4	15-10-2020	ketamine hydrochloride	Temperature	24.0	NaN	

  

	3:30:00	4:30:00	5:30:00	6:30:00	7:30:00	8:30:00	9:30:00	10:30:00	\
0	NaN	21.0	21.0	22	23.0	21.0	22.0	20	
1	NaN	11.0	13.0	14	16.0	16.0	24.0	18	
2	18.0	NaN	17.0	18	NaN	NaN	23.0	23	
3	22.0	NaN	22.0	23	NaN	NaN	27.0	26	
4	NaN	27.0	NaN	26	25.0	24.0	23.0	22	

  

	11:30:00	12:30:00
0	20.0	21
1	19.0	20
2	25.0	25
3	29.0	28
4	21.0	20

```
[87]: data_melt = pd.melt(data,
↳ id_vars=['Date', 'Drug_Name', 'Parameter'], var_name='Time', value_name='Reading')
```

```
[88]: data_melt.head()
```

```
[88]:
```

	Date	Drug_Name	Parameter	Time	Reading
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

```
[89]: data_tidy = data_melt.  
      ↪pivot(index=['Date', 'Drug_Name', 'Time'], columns='Parameter', values='Reading').  
      ↪reset_index()
```

```
[90]: data_tidy.head()
```

```
[90]: Parameter      Date      Drug_Name      Time  Pressure  \  
0      15-10-2020  diltiazem hydrochloride  10:30:00      18.0  
1      15-10-2020  diltiazem hydrochloride  11:30:00      19.0  
2      15-10-2020  diltiazem hydrochloride  12:30:00      20.0  
3      15-10-2020  diltiazem hydrochloride   1:30:00      12.0  
4      15-10-2020  diltiazem hydrochloride   2:30:00      13.0  
  
Parameter  Temperature  
0           20.0  
1           20.0  
2           21.0  
3           23.0  
4           22.0
```

## 1 Grouping using drug name and apply function

```
[91]: def temp_mean(x):  
      x['Average temperature'] = x['Temperature'].mean()  
      return x  
  
data_tidy = data_tidy.groupby('Drug_Name').apply(temp_mean)
```

```
[92]: data_tidy
```

```
[92]: Parameter      Date      Drug_Name      Time  \  
Drug_Name  
diltiazem hydrochloride 0      15-10-2020  diltiazem hydrochloride  10:30:00  
                        1      15-10-2020  diltiazem hydrochloride  11:30:00  
                        2      15-10-2020  diltiazem hydrochloride  12:30:00  
                        3      15-10-2020  diltiazem hydrochloride   1:30:00  
                        4      15-10-2020  diltiazem hydrochloride   2:30:00  
...  
ketamine hydrochloride 103     17-10-2020  ketamine hydrochloride   5:30:00  
                        104     17-10-2020  ketamine hydrochloride   6:30:00  
                        105     17-10-2020  ketamine hydrochloride   7:30:00  
                        106     17-10-2020  ketamine hydrochloride   8:30:00  
                        107     17-10-2020  ketamine hydrochloride   9:30:00  
  
Parameter      Pressure  Temperature  Average temperature  
Drug_Name  
diltiazem hydrochloride 0           18.0           20.0           24.848485
```

	1	19.0	20.0	24.848485
	2	20.0	21.0	24.848485
	3	12.0	23.0	24.848485
	4	13.0	22.0	24.848485
...		...	...	...
ketamine hydrochloride	103	11.0	17.0	17.709677
	104	12.0	18.0	17.709677
	105	12.0	19.0	17.709677
	106	11.0	20.0	17.709677
	107	12.0	21.0	17.709677

[108 rows x 6 columns]

```
[93]: data_tidy[:20]
```

```
[93]: Parameter          Date          Drug_Name      Time \
Drug_Name
diltiazem hydrochloride 0   15-10-2020  diltiazem hydrochloride  10:30:00
1   15-10-2020  diltiazem hydrochloride  11:30:00
2   15-10-2020  diltiazem hydrochloride  12:30:00
3   15-10-2020  diltiazem hydrochloride   1:30:00
4   15-10-2020  diltiazem hydrochloride   2:30:00
5   15-10-2020  diltiazem hydrochloride   3:30:00
6   15-10-2020  diltiazem hydrochloride   4:30:00
7   15-10-2020  diltiazem hydrochloride   5:30:00
8   15-10-2020  diltiazem hydrochloride   6:30:00
9   15-10-2020  diltiazem hydrochloride   7:30:00
10  15-10-2020  diltiazem hydrochloride   8:30:00
11  15-10-2020  diltiazem hydrochloride   9:30:00
36  16-10-2020  diltiazem hydrochloride  10:30:00
37  16-10-2020  diltiazem hydrochloride  11:30:00
38  16-10-2020  diltiazem hydrochloride  12:30:00
39  16-10-2020  diltiazem hydrochloride   1:30:00
40  16-10-2020  diltiazem hydrochloride   2:30:00
41  16-10-2020  diltiazem hydrochloride   3:30:00
42  16-10-2020  diltiazem hydrochloride   4:30:00
43  16-10-2020  diltiazem hydrochloride   5:30:00
```

Parameter		Pressure	Temperature	Average temperature
Drug_Name				
diltiazem hydrochloride	0	18.0	20.0	24.848485
	1	19.0	20.0	24.848485
	2	20.0	21.0	24.848485
	3	12.0	23.0	24.848485
	4	13.0	22.0	24.848485
	5	NaN	NaN	24.848485
	6	11.0	21.0	24.848485

7	13.0	21.0	24.848485
8	14.0	22.0	24.848485
9	16.0	23.0	24.848485
10	16.0	21.0	24.848485
11	24.0	22.0	24.848485
36	24.0	40.0	24.848485
37	NaN	NaN	24.848485
38	27.0	42.0	24.848485
39	18.0	34.0	24.848485
40	19.0	35.0	24.848485
41	20.0	36.0	24.848485
42	21.0	36.0	24.848485
43	22.0	37.0	24.848485

## 2 Filling the null values of Temperature and pressure using mean

```
[94]: data_tidy.Temperature.fillna(data_tidy.Temperature.mean(),inplace=True)
```

```
[95]: data_tidy.Pressure.fillna(data_tidy.Pressure.mean(),inplace=True)
```

```
[96]: data_tidy.isna().sum()
```

```
[96]: Parameter
Date                0
Drug_Name           0
Time                0
Pressure            0
Temperature          0
Average temperature  0
dtype: int64
```

## 3 Binning the data using cut function in pandas

```
[97]: data_tidy.Temperature.min()
```

```
[97]: 8.0
```

```
[98]: data_tidy.Temperature.max()
```

```
[98]: 58.0
```

```
[99]: data_tidy.Pressure.min()
```

```
[99]: 3.0
```

```
[100]: data_tidy.Pressure.max()
```

```
[100]: 30.0
```

```
[101]: temp_points = [5,20,35,50,65]
temp_labels = ['low','medium','high','very_high']
data_tidy['Temperature category']= pd.cut(data_tidy.
↳Temperature,bins=temp_points,labels=temp_labels)
```

```
[102]: data_tidy
```

```
[102]: Parameter                                Date                                Drug_Name                                Time \
Drug_Name
diltiazem hydrochloride 0      15-10-2020  diltiazem hydrochloride  10:30:00
                        1      15-10-2020  diltiazem hydrochloride  11:30:00
                        2      15-10-2020  diltiazem hydrochloride  12:30:00
                        3      15-10-2020  diltiazem hydrochloride   1:30:00
                        4      15-10-2020  diltiazem hydrochloride   2:30:00
...
ketamine hydrochloride 103    17-10-2020  ketamine hydrochloride   5:30:00
                        104    17-10-2020  ketamine hydrochloride   6:30:00
                        105    17-10-2020  ketamine hydrochloride   7:30:00
                        106    17-10-2020  ketamine hydrochloride   8:30:00
                        107    17-10-2020  ketamine hydrochloride   9:30:00
```

```
Parameter                                Pressure  Temperature  Average temperature \
Drug_Name
diltiazem hydrochloride 0              18.0          20.0          24.848485
                        1              19.0          20.0          24.848485
                        2              20.0          21.0          24.848485
                        3              12.0          23.0          24.848485
                        4              13.0          22.0          24.848485
...
ketamine hydrochloride 103             11.0          17.0          17.709677
                        104             12.0          18.0          17.709677
                        105             12.0          19.0          17.709677
                        106             11.0          20.0          17.709677
                        107             12.0          21.0          17.709677
```

```
Parameter                                Temperature category
Drug_Name
diltiazem hydrochloride 0                      low
                        1                      low
                        2                    medium
                        3                    medium
                        4                    medium
...
ketamine hydrochloride 103                      low
                        104                      low
```

```

105          low
106          low
107        medium

```

[108 rows x 7 columns]

```

[103]: press_points = [5,15,16,25]
       press_labels = ['Below_average', 'Average', 'Above_average']
       data_tidy['Pressure category'] = pd.cut(data_tidy.
       ↪Pressure, bins=press_points, labels=press_labels)

```

```

[104]: data_tidy['Pressure category'].value_counts()

```

```

[104]: Pressure category
       Above_average    43
       Below_average    40
       Average         3
       Name: count, dtype: int64

```

```

[105]: data_tidy['Temperature category'].value_counts()

```

```

[105]: Temperature category
       low          45
       medium       43
       high         15
       very_high     5
       Name: count, dtype: int64

```

## 4 Retrieving the data contains certain string using Contains function

```

[106]: data_tidy.loc[data_tidy.Drug_Name.str.contains('hydrochloride', case=False)] #_
       ↪Case will ignore whether it is lower or upper case

```

```

[106]: Parameter          Date          Drug_Name      Time \
       Drug_Name
diltiazem hydrochloride  0    15-10-2020  diltiazem hydrochloride  10:30:00
                               1    15-10-2020  diltiazem hydrochloride  11:30:00
                               2    15-10-2020  diltiazem hydrochloride  12:30:00
                               3    15-10-2020  diltiazem hydrochloride   1:30:00
                               4    15-10-2020  diltiazem hydrochloride   2:30:00
...
ketamine hydrochloride  103   17-10-2020  ketamine hydrochloride   5:30:00
                               104   17-10-2020  ketamine hydrochloride   6:30:00
                               105   17-10-2020  ketamine hydrochloride   7:30:00
                               106   17-10-2020  ketamine hydrochloride   8:30:00
                               107   17-10-2020  ketamine hydrochloride   9:30:00

```

Parameter		Pressure	Temperature	Average temperature \
Drug_Name				
diltiazem hydrochloride	0	18.0	20.0	24.848485
	1	19.0	20.0	24.848485
	2	20.0	21.0	24.848485
	3	12.0	23.0	24.848485
	4	13.0	22.0	24.848485
...		...		...
ketamine hydrochloride	103	11.0	17.0	17.709677
	104	12.0	18.0	17.709677
	105	12.0	19.0	17.709677
	106	11.0	20.0	17.709677
	107	12.0	21.0	17.709677

Parameter		Temperature category	Pressure category
Drug_Name			
diltiazem hydrochloride	0	low	Above_average
	1	low	Above_average
	2	medium	Above_average
	3	medium	Below_average
	4	medium	Below_average
...		...	...
ketamine hydrochloride	103	low	Below_average
	104	low	Below_average
	105	low	Below_average
	106	low	Below_average
	107	medium	Below_average

[72 rows x 8 columns]

## 5 Date and Time Functions in Pandas

```
[107]: data_tidy[['Date', 'Time']]
```

```
[107]: Parameter      Date      Time
Drug_Name
diltiazem hydrochloride 0    15-10-2020  10:30:00
                        1    15-10-2020  11:30:00
                        2    15-10-2020  12:30:00
                        3    15-10-2020   1:30:00
                        4    15-10-2020   2:30:00
...
ketamine hydrochloride 103   17-10-2020   5:30:00
                        104   17-10-2020   6:30:00
                        105   17-10-2020   7:30:00
```

```

106 17-10-2020 8:30:00
107 17-10-2020 9:30:00

```

[108 rows x 2 columns]

## 6 Getting year from Date column

```

[108]: def get_year(x):
        return x[2]
data_tidy['Year'] = data_tidy['Date'].str.split('-').apply(get_year)

```

```
[109]: data_tidy
```

```

[109]: Parameter          Date          Drug_Name      Time \
Drug_Name
diltiazem hydrochloride 0    15-10-2020  diltiazem hydrochloride  10:30:00
                                1    15-10-2020  diltiazem hydrochloride  11:30:00
                                2    15-10-2020  diltiazem hydrochloride  12:30:00
                                3    15-10-2020  diltiazem hydrochloride   1:30:00
                                4    15-10-2020  diltiazem hydrochloride   2:30:00
...
ketamine hydrochloride 103   17-10-2020  ketamine hydrochloride   5:30:00
                                104  17-10-2020  ketamine hydrochloride   6:30:00
                                105  17-10-2020  ketamine hydrochloride   7:30:00
                                106  17-10-2020  ketamine hydrochloride   8:30:00
                                107  17-10-2020  ketamine hydrochloride   9:30:00

```

```

Parameter          Pressure  Temperature  Average temperature \
Drug_Name
diltiazem hydrochloride 0         18.0         20.0         24.848485
                                1         19.0         20.0         24.848485
                                2         20.0         21.0         24.848485
                                3         12.0         23.0         24.848485
                                4         13.0         22.0         24.848485
...
ketamine hydrochloride 103         11.0         17.0         17.709677
                                104         12.0         18.0         17.709677
                                105         12.0         19.0         17.709677
                                106         11.0         20.0         17.709677
                                107         12.0         21.0         17.709677

```

```

Parameter          Temperature category  Pressure category  Year
Drug_Name
diltiazem hydrochloride 0                low      Above_average  2020
                                1                low      Above_average  2020
                                2              medium      Above_average  2020

```



	3	medium	Below_average	2020
	4	medium	Below_average	2020
...		...	...	...
ketamine hydrochloride	103	low	Below_average	2020
	104	low	Below_average	2020
	105	low	Below_average	2020
	106	low	Below_average	2020
	107	medium	Below_average	2020

[108 rows x 9 columns]

```
[110]: data_tidy['Time stamp'] = data_tidy['Date'] + ' ' + data_tidy['Time']
```

```
[111]: data_tidy
```

```
[111]: Parameter          Date          Drug_Name      Time \
Drug_Name
diltiazem hydrochloride 0    15-10-2020  diltiazem hydrochloride  10:30:00
1    15-10-2020  diltiazem hydrochloride  11:30:00
2    15-10-2020  diltiazem hydrochloride  12:30:00
3    15-10-2020  diltiazem hydrochloride   1:30:00
4    15-10-2020  diltiazem hydrochloride   2:30:00
...
ketamine hydrochloride 103   17-10-2020  ketamine hydrochloride    5:30:00
104   17-10-2020  ketamine hydrochloride    6:30:00
105   17-10-2020  ketamine hydrochloride    7:30:00
106   17-10-2020  ketamine hydrochloride    8:30:00
107   17-10-2020  ketamine hydrochloride    9:30:00
```

Parameter		Pressure	Temperature	Average temperature	\
Drug_Name					
diltiazem hydrochloride	0	18.0	20.0	24.848485	
	1	19.0	20.0	24.848485	
	2	20.0	21.0	24.848485	
	3	12.0	23.0	24.848485	
	4	13.0	22.0	24.848485	
...		...	...	...	
ketamine hydrochloride	103	11.0	17.0	17.709677	
	104	12.0	18.0	17.709677	
	105	12.0	19.0	17.709677	
	106	11.0	20.0	17.709677	
	107	12.0	21.0	17.709677	

Parameter		Temperature category	Pressure category	Year	\
Drug_Name					
diltiazem hydrochloride	0	low	Above_average	2020	
	1	low	Above_average	2020	

	2	medium	Above_average	2020
	3	medium	Below_average	2020
	4	medium	Below_average	2020
...		...	...	...
ketamine hydrochloride	103	low	Below_average	2020
	104	low	Below_average	2020
	105	low	Below_average	2020
	106	low	Below_average	2020
	107	medium	Below_average	2020

Parameter		Time stamp
Drug_Name		
diltiazem hydrochloride	0	15-10-2020 10:30:00
	1	15-10-2020 11:30:00
	2	15-10-2020 12:30:00
	3	15-10-2020 1:30:00
	4	15-10-2020 2:30:00
...		...
ketamine hydrochloride	103	17-10-2020 5:30:00
	104	17-10-2020 6:30:00
	105	17-10-2020 7:30:00
	106	17-10-2020 8:30:00
	107	17-10-2020 9:30:00

[108 rows x 10 columns]

## 7 Converting string into day format

```
[112]: data_tidy['Time stamp'] = pd.to_datetime(data_tidy['Time stamp'])
```

C:\Users\saita\AppData\Local\Temp\ipykernel\_2660\3180829823.py:1: UserWarning:  
Parsing dates in %d-%m-%Y %H:%M:%S format when dayfirst=False (the default) was  
specified. Pass `dayfirst=True` or specify a format to silence this warning.

```
data_tidy['Time stamp'] = pd.to_datetime(data_tidy['Time stamp'])
```

```
[113]: data_tidy
```

```
[113]: Parameter          Date          Drug_Name      Time \
Drug_Name
diltiazem hydrochloride 0    15-10-2020  diltiazem hydrochloride  10:30:00
                        1    15-10-2020  diltiazem hydrochloride  11:30:00
                        2    15-10-2020  diltiazem hydrochloride  12:30:00
                        3    15-10-2020  diltiazem hydrochloride   1:30:00
                        4    15-10-2020  diltiazem hydrochloride   2:30:00
...
ketamine hydrochloride 103  17-10-2020  ketamine hydrochloride   5:30:00
                        104  17-10-2020  ketamine hydrochloride   6:30:00
```

105	17-10-2020	ketamine hydrochloride	7:30:00
106	17-10-2020	ketamine hydrochloride	8:30:00
107	17-10-2020	ketamine hydrochloride	9:30:00

Parameter		Pressure	Temperature	Average temperature	\
Drug_Name					
diltiazem hydrochloride	0	18.0	20.0	24.848485	
	1	19.0	20.0	24.848485	
	2	20.0	21.0	24.848485	
	3	12.0	23.0	24.848485	
	4	13.0	22.0	24.848485	
...		...	...	...	
ketamine hydrochloride	103	11.0	17.0	17.709677	
	104	12.0	18.0	17.709677	
	105	12.0	19.0	17.709677	
	106	11.0	20.0	17.709677	
	107	12.0	21.0	17.709677	

Parameter		Temperature category	Pressure category	Year	\
Drug_Name					
diltiazem hydrochloride	0	low	Above_average	2020	
	1	low	Above_average	2020	
	2	medium	Above_average	2020	
	3	medium	Below_average	2020	
	4	medium	Below_average	2020	
...		...	...	...	
ketamine hydrochloride	103	low	Below_average	2020	
	104	low	Below_average	2020	
	105	low	Below_average	2020	
	106	low	Below_average	2020	
	107	medium	Below_average	2020	

Parameter		Time stamp
Drug_Name		
diltiazem hydrochloride	0	2020-10-15 10:30:00
	1	2020-10-15 11:30:00
	2	2020-10-15 12:30:00
	3	2020-10-15 01:30:00
	4	2020-10-15 02:30:00
...		...
ketamine hydrochloride	103	2020-10-17 05:30:00
	104	2020-10-17 06:30:00
	105	2020-10-17 07:30:00
	106	2020-10-17 08:30:00
	107	2020-10-17 09:30:00

[108 rows x 10 columns]

```
[114]: type(data_tidy['Time stamp'][1])
```

```
C:\Users\saita\AppData\Local\Temp\ipykernel_2660\352291693.py:1: FutureWarning:
Series.__getitem__ treating keys as positions is deprecated. In a future
version, integer keys will always be treated as labels (consistent with
DataFrame behavior). To access a value by position, use `ser.iloc[pos]`
type(data_tidy['Time stamp'][1])
```

```
[114]: pandas._libs.tslibs.timestamps.Timestamp
```

## 8 Accessing the Day, Month, Year from a time stamp

```
[115]: Date = data_tidy['Time stamp'][0]
```

```
C:\Users\saita\AppData\Local\Temp\ipykernel_2660\2824729332.py:1: FutureWarning:
Series.__getitem__ treating keys as positions is deprecated. In a future
version, integer keys will always be treated as labels (consistent with
DataFrame behavior). To access a value by position, use `ser.iloc[pos]`
Date = data_tidy['Time stamp'][0]
```

```
[116]: Date.year
```

```
[116]: 2020
```

```
[117]: Date.day
```

```
[117]: 15
```

```
[118]: Date.month
```

```
[118]: 10
```

```
[119]: Date.month_name()
```

```
[119]: 'October'
```

```
[120]: Date.day_name()
```

```
[120]: 'Thursday'
```

```
[121]: data_tidy['Time stamp'].dt.month_name()
```

```
[121]: Drug_Name
diltiazem hydrochloride  0      October
                        1      October
                        2      October
                        3      October
                        4      October
```

```

ketamine hydrochloride    103    ...
                           104    October
                           105    October
                           106    October
                           107    October
Name: Time stamp, Length: 108, dtype: object

```

## 9 If you want to get the date in specified format

```
[122]: Date
```

```
[122]: Timestamp('2020-10-15 10:30:00')
```

### Printing only year

```
[125]: Date.strftime('%Y') #Or we can use lower y which give year in this format YY
```

```
[125]: '2020'
```

```
[127]: Date.strftime('%M')
```

```
[127]: '30'
```

```
[128]: Date.strftime('%h')
```

```
[128]: 'Oct'
```

```
[129]: Date.strftime("%d-%m-%y")
```

```
[129]: '15-10-20'
```

## 10 Saving the CSV File

```
[130]: data_tidy.to_csv("Pfizer_tidy.csv")
```

```
[ ]:
```