Colab: https://colab.research.google.com/drive/1aFIMcP60C0deh5Q8uoI_gKw1Ud5s9_ms?usp=sharing

```
import pandas as pd
import numpy as np
```

```
!gdown 173A59xh2mnpmljCCB9bhC4C5eP2IS6qZ
```

```
Downloading...
From: https://drive.google.com/uc?id=173A59xh2mnpmljCCB9bhC4C5eP2IS6qZ
To: /content/Pfizer_1.csv
100% 1.51k/1.51k [00:00<00:00, 831kB/s]
```

```
data = pd.read_csv('Pfizer_1.csv')
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18 entries, 0 to 17
Data columns (total 15 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Date        18 non-null     object
 1   Drug_Name   18 non-null     object
                                 object
                                 float64
                                 float64
 5   3:30:00     12 non-null     float64
 6   4:30:00     14 non-null     float64
 7   5:30:00     16 non-null     float64
 8   6:30:00     18 non-null     int64
 9   7:30:00     16 non-null     float64
 10  8:30:00     14 non-null     float64
 11  9:30:00     16 non-null     float64
 12  10:30:00    18 non-null     int64
 13  11:30:00    16 non-null     float64
 14  12:30:00    18 non-null     int64
dtypes: float64(9), int64(3), object(3)
memory usage: 2.2+ KB
```



```
data.head()
```

	Date	Drug_Name	Parameter	1:30:00	2:30:00	3:30:00	4:30:00
0	15-10-2020	diltiazem hydrochloride	Temperature	23.0	22.0	NaN	21.0
1	15-10-2020	diltiazem hydrochloride	Pressure	12.0	13.0	NaN	11.0
2	15-10-2020	docetaxel injection	Temperature	NaN	17.0	18.0	NaN
3	15-10-2020	docetaxel injection	Pressure	NaN	22.0	22.0	NaN
4	15-10-2020	ketamine hydrochloride	Temperature	24.0	NaN	NaN	27.0

```
pd.melt(data, id_vars=["Date", "Parameter", "Drug_Name"])
```

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

216 rows × 5 columns


```
data_melt = pd.melt(data, id_vars=["Date", "Parameter", "Drug_Name"],
        var_name = "time",
        value_name = "reading")
```

```
data_melt.pivot(index=['Date','Drug_Name','Parameter'],
                columns="time",
                values="reading").reset_index()
```

time	Date	Drug_Name	Parameter	10:30:00	11:30:00	12:30:00	1:
0	15-10-2020	diltiazem hydrochloride	Pressure	18.0	19.0	20.0	
1	15-10-2020	diltiazem hydrochloride	Temperature	20.0	20.0	21.0	
2	15-10-2020	docetaxel injection	Pressure	26.0	29.0	28.0	
3	15-10-2020	docetaxel injection	Temperature	23.0	25.0	25.0	
4	15-10-2020	ketamine hydrochloride	Pressure	9.0	9.0	11.0	
5	15-10-2020	ketamine hydrochloride	Temperature	22.0	21.0	20.0	
6	16-10-2020	diltiazem hydrochloride	Pressure	24.0	NaN	27.0	
7	16-10-2020	diltiazem hydrochloride	Temperature	40.0	NaN	42.0	
8	16-10-2020	docetaxel injection	Pressure	28.0	29.0	30.0	
9	16-10-2020	docetaxel injection	Temperature	56.0	57.0	58.0	
10	16-10-2020	ketamine hydrochloride	Pressure	16.0	17.0	18.0	
11	16-10-2020	ketamine hydrochloride	Temperature	13.0	14.0	15.0	
12	17-10-2020	diltiazem hydrochloride	Pressure	11.0	13.0	14.0	
13	17-10-2020	diltiazem hydrochloride	Temperature	14.0	11.0	10.0	
14	17-10-2020	docetaxel injection	Pressure	28.0	29.0	28.0	
15	17-10-2020	docetaxel injection	Temperature	21.0	22.0	23.0	
16		chloride	Pressure	13.0	14.0	15.0	
17	17-10-2020	ketamine hydrochloride	Temperature	22.0	23.0	24.0	

Saving...

```
# pivot table - post read, v v v important


data_tidy = data_melt.pivot(index=['Date','Drug_Name', "time"],
                columns="Parameter",
                values="reading").reset_index()
data_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
		hydrochloride	6:30:00	12.0	18.0
		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

Saving...

```
data_tidy.head()
```

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

```
data_tidy.info()
```

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

```
Data columns (total 5 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     95 non-null     float64
 4   Temperature  95 non-null     float64
dtypes: float64(2), object(3)
memory usage: 4.3+ KB
```

```
# None, NaN (Not a number)
```

```
type(None) # strings datatype
```

```
NoneType
```

```
type(np.nan) # can be used for supoorting both
```

```
float
```

```
pd.Series([1, np.nan, 2, None])
```

```
0    1.0
1    NaN
2    2.0
3    NaN
```



```
pd.Series(["1", "np.nan", "2", None])
```

```
0         1
1    np.nan
2         2
3      None
dtype: object
```

```
pd.Series(["1", "np.nan", "2", np.nan])
```

```
0         1
1    np.nan
2         2
3       NaN
dtype: object
```

```
data.isnull() #isna
```

	Date	Drug_Name	Parameter	1:30:00	2:30:00	3:30:00	4:30:00	5:30:00	6:3
0	False	False	False	False	False	True	False	False	
1	False	False	False	False	False	True	False	False	
2	False	False	False	True	False	False	True	False	
3	False	False	False	True	False	False	True	False	
4	False	False	False	False	True	True	False	True	
5	False	False	False	False	True	True	False	True	
6	False	False	False	False	False	False	False	False	
7	False	False	False	False	False	False	False	False	
8	False	False	False	False	False	True	False	False	
9	False	False	False	False	False	True	False	False	
10	False	False	False	False	False	False	True	False	
11	False	False	False	False	False	False	True	False	
12	False	False	False	False	False	False	False	False	
13	False	False	False	False	False	False	False	False	
14	False	False	False	False	False	False	False	False	
15	False	False	False	False	False	False	False	False	

Saving...

```
data.isnull().sum()
```

```
Date          0
Drug_Name     0
Parameter     0
1:30:00       2
2:30:00       2
3:30:00       6
4:30:00       4
5:30:00       2
6:30:00       0
7:30:00       2
8:30:00       4
9:30:00       2
10:30:00      0
11:30:00      2
12:30:00      0
dtype: int64
```

```
data.isnull().sum(axis=1)
```

```
0     1
1     1
2     4
3     4
4     3
```

```
5     3
6     1
7     1
8     1
9     1
10    2
11    2
12    1
13    1
14    0
15    0
16    0
17    0
dtype: int64
```

```
data.dropna() # all the rows which have missing data
```

	Date	Drug_Name	Parameter	1:30:00	2:30:00	3:30:00	4:30:00	5:30:00	6
14	17-10-2020	docetaxel injection	Temperature	12.0	13.0	14.0	15.0	16.0	
15	17-10-2020	docetaxel injection	Pressure	20.0	22.0	22.0	22.0	22.0	
16	17-10-2020	ketamine hydrochloride	Temperature	13.0	14.0	15.0	16.0	17.0	
17	17-10-2020	ketamine hydrochloride	Pressure	8.0	9.0	10.0	11.0	11.0	

Saving...

```
data.dropna(axis=1)
```

	Date	Drug_Name	Parameter	6:30:00	10:30:00	12:30:00
0	15-10-2020	diltiazem hydrochloride	Temperature	22	20	21
1	15-10-2020	diltiazem hydrochloride	Pressure	14	18	20
2	15-10-2020	docetaxel injection	Temperature	18	23	25
3	15-10-2020	docetaxel injection	Pressure	23	26	28
4	15-10-2020	ketamine hydrochloride	Temperature	26	22	20
5	15-10-2020	ketamine hydrochloride	Pressure	9	9	11
6	16-10-2020	diltiazem hydrochloride	Temperature	38	40	42
7	16-10-2020	diltiazem hydrochloride	Pressure	23	24	27
8	16-10-2020	docetaxel injection	Temperature	49	56	58

```
data.fillna(0)
```

Saving...

	2020	injection						
4	15-10-2020	ketamine hydrochloride	Temperature	24.0	0.0	0.0	27.0	0.0
5	15-10-2020	ketamine hydrochloride	Pressure	8.0	0.0	0.0	7.0	0.0
6	16-10-2020	diltiazem hydrochloride	Temperature	34.0	35.0	36.0	36.0	37.0
7	16-10-2020	diltiazem hydrochloride	Pressure	18.0	19.0	20.0	21.0	22.0
8	16-10-2020	docetaxel injection	Temperature	46.0	47.0	0.0	48.0	48.0
9	16-10-2020	docetaxel injection	Pressure	23.0	24.0	0.0	25.0	26.0
10	16-10-2020	ketamine hydrochloride	Temperature	8.0	9.0	10.0	0.0	11.0
11	16-10-	ketamine hydrochloride	Pressure	12.0	12.0	13.0	0.0	15.0
12	10-2020	diltiazem hydrochloride	Temperature	20.0	19.0	19.0	18.0	17.0

Saving...

```
data.fillna(10000000000)
```

	Date	Drug_Name	Parameter	1:30:00	2:30:00	3:30:00	4:30
0	15-10-2020	diltiazem hydrochloride	Temperature	2.300000e+01	2.200000e+01	1.000000e+10	2.100000e
1	15-10-2020	diltiazem hydrochloride	Pressure	1.200000e+01	1.300000e+01	1.000000e+10	1.100000e
2	15-10-2020	docetaxel injection	Temperature	1.000000e+10	1.700000e+01	1.800000e+01	1.000000e
3	15-10-2020	docetaxel injection	Pressure	1.000000e+10	2.200000e+01	2.200000e+01	1.000000e
4	15-10-2020	ketamine hydrochloride	Temperature	2.400000e+01	1.000000e+10	1.000000e+10	2.700000e
5	15-10-2020	ketamine hydrochloride	Pressure	8.000000e+00	1.000000e+10	1.000000e+10	7.000000e
6	16-10-2020	diltiazem hydrochloride	Temperature	3.400000e+01	3.500000e+01	3.600000e+01	3.600000e
	16-	diltiazem	sure	1.800000e+01	1.900000e+01	2.000000e+01	2.100000e
8	16-10-2020	docetaxel injection	Temperature	4.600000e+01	4.700000e+01	1.000000e+10	4.800000e
9	16-10-2020	docetaxel injection	Pressure	2.300000e+01	2.400000e+01	1.000000e+10	2.500000e
10	16-10-2020	ketamine hydrochloride	Temperature	8.000000e+00	9.000000e+00	1.000000e+01	1.000000e
11	16-10-2020	ketamine hydrochloride	Pressure	1.200000e+01	1.200000e+01	1.300000e+01	1.000000e
12	17-10-2020	diltiazem hydrochloride	Temperature	2.000000e+01	1.900000e+01	1.900000e+01	1.800000e
	17-	diltiazem					

Saving...

```
data["2:30:00"].mean()
```

```
18.8125
```

```
data["2:30:00"].fillna(data["2:30:00"].mean())
```

```
0     22.0000
1     13.0000
2     17.0000
3     22.0000
4     18.8125
5     18.8125
6     35.0000
7     19.0000
8     47.0000
9     24.0000
10     9.0000
11    12.0000
12    19.0000
13     4.0000
14    13.0000
15    22.0000
16    14.0000
17     9.0000
Name: 2:30:00, dtype: float64
```

```
data.isna().sum()
```

```
Date          0
Drug_Name     0
Parameter     0
1:30:00       2
2:30:00       2
3:30:00       6
4:30:00       4
7:30:00       2
8:30:00       4
9:30:00       2
10:30:00      0
11:30:00      2
12:30:00      0
dtype: int64
```



```
data_tidy.isna().sum()
```

```
Parameter
Date            0
Drug_Name       0
time            0
Pressure       13
Temperature    13
dtype: int64
```

```
def temp_mean(x):
  x["Avg_Temperature"] = x["Temperature"].mean()
  return x
data_tidy = data_tidy.groupby("Drug_Name").apply(temp_mean)
```

```
data_tidy["Temperature"].fillna(data_tidy["Avg_Temperature"], inplace=True)

data_tidy.isna().sum()
```

```
Parameter
Date                0
Drug_Name           0
time                0
Pressure           13
Temperature         0
Avg_Temperature     0
dtype: int64
```

```
def pr_mean(x):
  x['Pressure_avg'] = x['Pressure'].mean()
  return x
data_tidy=data_tidy.groupby(["Drug_Name"]).apply(pr_mean)
data_tidy['Pressure'].fillna(data_tidy["Pressure_avg"], inplace=True)
data_tidy
```

	Parameter	Date	Drug_Name	time	Pressure	Temperature	Avg_Temperature	P
0		15-10-2020	diltiazem hydrochloride	10:30:00	18.0	20.0	24.848485	
1		15-10-2020	diltiazem hydrochloride	11:30:00	19.0	20.0	24.848485	
2		15-10-2020	diltiazem hydrochloride	12:30:00	20.0	21.0	24.848485	
3		15-10-2020	diltiazem hydrochloride	1:30:00	12.0	23.0	24.848485	
4		15-10-2020	diltiazem hydrochloride	2:30:00	13.0	22.0	24.848485	
...		...	...	...	...	...	...	...
		17-	ketamine					


```
data_tidy.isna().sum()
```

```
Parameter
Date                0
Drug_Name           0
time                0
Pressure            0
Temperature         0
Avg_Temperature     0
Pressure_avg        0
dtype: int64
```

```
data_tidy["Temperature"].min()
```

```
8.0
```

```
data_tidy["Temperature"].max()
```

```
58.0
```

```
# "low", "medium", "high", "very high"
# 5-20, 20-35, 35-50, 36-60
```

```
temp_points = [5, 20, 35, 50, 60]
temp_labels = ["low", "medium", "high", "very high"]
data_tidy["Cat_Temperature"] = pd.cut(data_tidy["Temperature"],
                                      bins=temp_points,
                                      labels = temp_labels)
```

```
data_tidy
```