Agenda

- **Restructuring data** ✅
  - `pd.melt()` ✅
  - `pd.pivot()` ✅
  - `pd.pivot_table()` ✅
  - `pd.cut()`
- **Dealing with Missing Values**✅
  - `None and nan values` ✅
  - `isna() and isnull()` ✅
  - `dropna()` ✅
  - `fillna()` ✅
- **String method in pandas**
- **Handling datetime**
- **Writing to a file**

Importing Dataset

In [2]:
```python
import numpy as np
import pandas as pd
```

In [3]:
```python
df=pd.read_csv("/Users/nikhilsanghi/Downloads/01_dsml-course-main-live/batches/2_Sept_Beg_Tue_Oct_Beg_Tue/12_Pandas_5/Pfizer_1.csv")
df
```

Out[3]:

	Date	Drug_Name	Parameter	1:30:00	2:30:00	3:30:00	4:30:00	5:30:00	6:30:00	7:30:00	8:30:00	9:30:00	10:30:00	11:30:00	12:30:00
0	15-10-2020	diltiazem hydrochloride	Temperature	23.0	22.0	NaN	21.0	21.0	22	23.0	21.0	22.0	20	20.0	21
1	15-10-2020	diltiazem hydrochloride	Pressure	12.0	13.0	NaN	11.0	13.0	14	16.0	16.0	24.0	18	19.0	20
2	15-10-2020	docetaxel injection	Temperature	NaN	17.0	18.0	NaN	17.0	18	NaN	NaN	23.0	23	25.0	25
3	15-10-2020	docetaxel injection	Pressure	NaN	22.0	22.0	NaN	22.0	23	NaN	NaN	27.0	26	29.0	28
4	15-10-2020	ketamine hydrochloride	Temperature	24.0	NaN	NaN	27.0	NaN	26	25.0	24.0	23.0	22	21.0	20
5	15-10-2020	ketamine hydrochloride	Pressure	8.0	NaN	NaN	7.0	NaN	9	10.0	11.0	10.0	9	9.0	11
6	16-10-2020	diltiazem hydrochloride	Temperature	34.0	35.0	36.0	36.0	37.0	38	37.0	38.0	39.0	40	NaN	42
7	16-10-2020	diltiazem hydrochloride	Pressure	18.0	19.0	20.0	21.0	22.0	23	24.0	25.0	25.0	24	NaN	27
8	16-10-2020	docetaxel injection	Temperature	46.0	47.0	NaN	48.0	48.0	49	50.0	52.0	55.0	56	57.0	58
9	16-10-2020	docetaxel injection	Pressure	23.0	24.0	NaN	25.0	26.0	27	28.0	29.0	28.0	28	29.0	30
10	16-10-2020	ketamine hydrochloride	Temperature	8.0	9.0	10.0	NaN	11.0	12	12.0	11.0	NaN	13	14.0	15
11	16-10-2020	ketamine hydrochloride	Pressure	12.0	12.0	13.0	NaN	15.0	15	15.0	15.0	NaN	16	17.0	18
12	17-10-2020	diltiazem hydrochloride	Temperature	20.0	19.0	19.0	18.0	17.0	16	15.0	NaN	13.0	14	11.0	10
13	17-10-2020	diltiazem hydrochloride	Pressure	3.0	4.0	4.0	4.0	6.0	8	9.0	NaN	9.0	11	13.0	14
14	17-10-2020	docetaxel injection	Temperature	12.0	13.0	14.0	15.0	16.0	17	18.0	19.0	20.0	21	22.0	23
15	17-10-2020	docetaxel injection	Pressure	20.0	22.0	22.0	22.0	22.0	23	25.0	26.0	27.0	28	29.0	28
16	17-10-2020	ketamine hydrochloride	Temperature	13.0	14.0	15.0	16.0	17.0	18	19.0	20.0	21.0	22	23.0	24
17	17-10-2020	ketamine hydrochloride	Pressure	8.0	9.0	10.0	11.0	11.0	12	12.0	11.0	12.0	13	14.0	15

In [4]:
```python
df.shape
```

Out[4]:
```
(18, 15)
```

Restructuring data

`pd.melt()`

In [12]:
```python
pd.melt(df,
       id_vars=["Date","Drug_Name","Parameter"])
```

Out[12]:

	Date	Drug_Name	Parameter	variable	value
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

In [13]:
```python
df_melt=pd.melt(df,
       id_vars=["Date","Drug_Name","Parameter"],
       var_name="Time",
       value_name="Reading")
df_melt
```

Out[13]:

	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
...	...	...	...	...	...
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

`pd.pivot()`

In [16]:
```python
df_pivoted_raw=pd.pivot(df_melt,
       index=["Date","Drug_Name","Time"],
       columns="Parameter",
       values="Reading")
df_pivoted_raw
```

Out[16]:

Date	Drug_Name	Time	Pressure	Temperature
15-10-2020	diltiazem hydrochloride	10:30:00	18.0	20.0
		11:30:00	19.0	20.0
		12:30:00	20.0	21.0
		1:30:00	12.0	23.0
		2:30:00	13.0	22.0
...	...	...	...	...
17-10-2020	ketamine hydrochloride	5:30:00	11.0	17.0
		6:30:00	12.0	18.0
		7:30:00	12.0	19.0
		8:30:00	11.0	20.0
		9:30:00	12.0	21.0

108 rows × 2 columns

In [17]:
```python
df_pivoted_raw.index
```

Out[17]:
```
MultiIndex([('15-10-2020', 'diltiazem hydrochloride', '10:30:00'),
            ('15-10-2020', 'diltiazem hydrochloride', '11:30:00'),
            ('15-10-2020', 'diltiazem hydrochloride', '12:30:00'),
            ('15-10-2020', 'diltiazem hydrochloride',  '1:30:00'),
            ('15-10-2020', 'diltiazem hydrochloride',  '2:30:00'),
            ('15-10-2020', 'diltiazem hydrochloride',  '3:30:00'),
            ('15-10-2020', 'diltiazem hydrochloride',  '4:30:00'),
            ('15-10-2020', 'diltiazem hydrochloride',  '5:30:00'),
            ('15-10-2020', 'diltiazem hydrochloride',  '6:30:00'),
            ('15-10-2020', 'diltiazem hydrochloride',  '7:30:00'),
            ...
            ('17-10-2020',  'ketamine hydrochloride', '12:30:00'),
            ('17-10-2020',  'ketamine hydrochloride',  '1:30:00'),
            ('17-10-2020',  'ketamine hydrochloride',  '2:30:00'),
            ('17-10-2020',  'ketamine hydrochloride',  '3:30:00'),
            ('17-10-2020',  'ketamine hydrochloride',  '4:30:00'),
            ('17-10-2020',  'ketamine hydrochloride',  '5:30:00'),
            ('17-10-2020',  'ketamine hydrochloride',  '6:30:00'),
            ('17-10-2020',  'ketamine hydrochloride',  '7:30:00'),
            ('17-10-2020',  'ketamine hydrochloride',  '8:30:00'),
            ('17-10-2020',  'ketamine hydrochloride',  '9:30:00')],
           names=['Date', 'Drug_Name', 'Time'], length=108)
```

In [18]:
```python
df_pivot=df_pivoted_raw.reset_index()
df_pivot
```

Out[18]:

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

In [22]:
```python
df_pivot.columns
```

Out[22]:
```
Index(['Date', 'Drug_Name', 'Time', 'Pressure', 'Temperature'], dtype='object', name='Parameter')
```

In [24]:
```python
df_pivot.columns.name=None
```

In [25]:
```python
df_pivot
```

Out[25]:

	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

In [28]:
```python
df_pivot.index.name
```

In [29]:
```python
df["Date"]
```

Out[29]:
```
0     15-10-2020
1     15-10-2020
2     15-10-2020
3     15-10-2020
4     15-10-2020
5     15-10-2020
6     16-10-2020
7     16-10-2020
8     16-10-2020
9     16-10-2020
10    16-10-2020
11    16-10-2020
12    17-10-2020
13    17-10-2020
14    17-10-2020
15    17-10-2020
16    17-10-2020
17    17-10-2020
Name: Date, dtype: object
```

In [ ]:

`pd.pivot_table()`

In [30]:
```python
pd.pivot_table(df_pivot,
              index="Drug_Name",
              columns="Date",
              values="Temperature")
```

Out[30]:

Date	15-10-2020	16-10-2020	17-10-2020
Drug_Name			
diltiazem hydrochloride	21.454545	37.454545	15.636364
docetaxel injection	20.750000	51.454545	17.500000
ketamine hydrochloride	23.555556	11.500000	18.500000

In [32]:
```python
pd.pivot_table(df_pivot,
              index="Date",
              columns="Drug_Name",
              values="Temperature")
```

Out[32]:

Drug_Name	diltiazem hydrochloride	docetaxel injection	ketamine hydrochloride
Date			
15-10-2020	21.454545	20.750000	23.555556
16-10-2020	37.454545	51.454545	11.500000
17-10-2020	15.636364	17.500000	18.500000

In [33]:
```python
pd.pivot_table(df_pivot,
              index="Drug_Name",
              columns="Date",
              values=["Temperature","Pressure"])
```

Out[33]:

	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	16.000000	22.545455	7.727273	21.454545	37.454545	15.636364
docetaxel injection	24.875000	27.000000	24.500000	20.750000	51.454545	17.500000
ketamine hydrochloride	9.333333	14.800000	11.500000	23.555556	11.500000	18.500000

In [34]:
```python
pd.pivot_table(df_pivot,
              index="Drug_Name",
              columns="Date",
              values=["Temperature","Pressure"],
              aggfunc="max")
```

Out[34]:

	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	24.0	27.0	14.0	23.0	42.0	20.0
docetaxel injection	29.0	30.0	29.0	25.0	58.0	23.0
ketamine hydrochloride	11.0	18.0	15.0	27.0	15.0	24.0

In [35]:
```python
pd.pivot_table(df_pivot,
              index="Drug_Name",
              columns="Date",
              values=["Temperature","Pressure"],
              aggfunc=["max","min"])
```

Out[35]:

	max						min					
	Pressure			Temperature			Pressure			Temperature		
Date	15-10-2020	16-10-2020	17-10-2020	15-10-2020	16-10-2020	17-10-2020	15-10-2020	16-10-2020	17-10-2020	15-10-2020	16-10-2020	17-10-2020
Drug_Name												
diltiazem hydrochloride	24.0	27.0	14.0	23.0	42.0	20.0	11.0	18.0	3.0	20.0	34.0	10.0
docetaxel injection	29.0	30.0	29.0	25.0	58.0	23.0	22.0	23.0	20.0	17.0	46.0	12.0
ketamine hydrochloride	11.0	18.0	15.0	27.0	15.0	24.0	7.0	12.0	8.0	20.0	8.0	13.0

In [36]:
```python
pd.pivot_table(df_pivot,
              index="Drug_Name",
              columns="Date",
              values="Temperature",
              aggfunc="max",
              margins=True)
```

Out[36]:

Date	15-10-2020	16-10-2020	17-10-2020	All
Drug_Name				
diltiazem hydrochloride	23.0	42.0	20.0	42.0
docetaxel injection	25.0	58.0	23.0	58.0
ketamine hydrochloride	27.0	15.0	24.0	27.0
All	27.0	58.0	24.0	58.0

In [ ]:

In [ ]:

`pd.cut()`

In [ ]:

Dealing with Missing Values

In [37]:
```python
None
```

In [38]:
```python
type(None)
```

Out[38]:
```
NoneType
```

In [39]:
```python
np.nan
```

Out[39]:
```
nan
```

In [40]:
```python
type(np.nan)
```

Out[40]:
```
float
```

In [41]:
```python
pd.Series([1,np.nan,None])
```

Out[41]:
```
0    1.0
1    NaN
2    NaN
dtype: float64
```

In [42]:
```python
pd.Series([1,2.0,np.nan,None])
```

Out[42]:
```
0    1.0
1    2.0
2    NaN
3    NaN
dtype: float64
```

In [43]:
```python
pd.Series(["1","2.0",np.nan,None])
```

Out[43]:
```
0       1
1     2.0
2     NaN
3    None
dtype: object
```

In [44]:
```python
pd.Series(["1","2.0","np.nan",None])
```

Out[44]:
```
0         1
1       2.0
2    np.nan
3      None
dtype: object
```

In [45]:
```python
df.isna().sum()
```

Out[45]:
```
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
```

In [49]:
```python
df.isna().sum(axis=1)
```

Out[49]:
```
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
```

In [50]:
```python
df.isnull().sum()
```

Out[50]:
```
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
```

In [51]:
```python
pd.isna
```

Out[51]:
```
<function pandas.core.dtypes.missing.isna(obj)>
```

In [52]:
```python
pd.isnull
```

Out[52]:
```
<function pandas.core.dtypes.missing.isna(obj)>
```

In [53]:
```python
df1=pd.DataFrame([[np.nan, 2, np.nan, 0],
                 [3, 4, np.nan, 1],
                 [np.nan, np.nan, np.nan, np.nan],
                 [np.nan, 3, np.nan, 4]],
                columns=["A","B","C","D"])
df1
```

Out[53]:

	A	B	C	D
0	NaN	2.0	NaN	0.0
1	3.0	4.0	NaN	1.0
2	NaN	NaN	NaN	NaN
3	NaN	3.0	NaN	4.0

In [55]:
```python
df1["B"].mean()
```

Out[55]:
```
3.0
```

In [56]:
```python
df1["C"].mean()
```

Out[56]:
```
nan
```

In [57]:
```python
df1["A"].mean()
```

Out[57]:
```
3.0
```

In [59]:
```python
df1
```

Out[59]:

	A	B	C	D
0	NaN	2.0	NaN	0.0
1	3.0	4.0	NaN	1.0
2	NaN	NaN	NaN	NaN
3	NaN	3.0	NaN	4.0

In [58]:
```python
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4 entries, 0 to 3
Data columns (total 4 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   A       1 non-null      float64
 1   B       3 non-null      float64
 2   C       0 non-null      float64
 3   D       3 non-null      float64
dtypes: float64(4)
memory usage: 256.0 bytes
```

In [60]:
```python
df1.fillna(0)
```

Out[60]:

	A	B	C	D
0	0.0	2.0	0.0	0.0
1	3.0	4.0	0.0	1.0
2	0.0	0.0	0.0	0.0
3	0.0	3.0	0.0	4.0

In [61]:
```python
df1.fillna(100)
```

Out[61]:

	A	B	C	D
0	100.0	2.0	100.0	0.0
1	3.0	4.0	100.0	1.0
2	100.0	100.0	100.0	100.0
3	100.0	3.0	100.0	4.0

In [62]:
```python
df1["A"].fillna(100)
```

Out[62]:
```
0    100.0
1      3.0
2    100.0
3    100.0
Name: A, dtype: float64
```

In [63]:
```python
df1
```

Out[63]:

	A	B	C	D
0	NaN	2.0	NaN	0.0
1	3.0	4.0	NaN	1.0
2	NaN	NaN	NaN	NaN
3	NaN	3.0	NaN	4.0

In [64]:
```python
df1["D"].mean()
```

Out[64]:
```
1.6666666666666667
```

In [65]:
```python
df1["A"].fillna(df1["D"].mean())
```

Out[65]:
```
0    1.666667
1    3.000000
2    1.666667
3    1.666667
Name: A, dtype: float64
```

In [67]:
```python
df1
```

Out[67]:

	A	B	C	D
0	NaN	2.0	NaN	0.0
1	3.0	4.0	NaN	1.0
2	NaN	NaN	NaN	NaN
3	NaN	3.0	NaN	4.0

In [68]:
```python
df1.fillna(method="ffill")
```

Out[68]:

	A	B	C	D
0	NaN	2.0	NaN	0.0
1	3.0	4.0	NaN	1.0
2	3.0	4.0	NaN	1.0
3	3.0	3.0	NaN	4.0

In [70]:
```python
df1
```

Out[70]:

	A	B	C	D
0	NaN	2.0	NaN	0.0
1	3.0	4.0	NaN	1.0
2	NaN	NaN	NaN	NaN
3	NaN	3.0	NaN	4.0

In [69]:
```python
df1.fillna(method="bfill")
```

Out[69]:

	A	B	C	D
0	3.0	2.0	NaN	0.0
1	3.0	4.0	NaN	1.0
2	NaN	3.0	NaN	4.0
3	NaN	3.0	NaN	4.0

In [72]:
```python
df1.interpolate()
```

Out[72]:

	A	B	C	D
0	NaN	2.0	NaN	0.0
1	3.0	4.0	NaN	1.0
2	3.0	3.5	NaN	2.5
3	3.0	3.0	NaN	4.0

In [75]:
```python
df1["D"].fillna(df1["D"].mean())
```

Out[75]:
```
0    0.000000
1    1.000000
2    1.666667
3    4.000000
Name: D, dtype: float64
```

In [78]:
```python
df1
```

Out[78]:

	A	B	C	D
0	NaN	2.0	NaN	0.0
1	3.0	4.0	NaN	1.0
2	NaN	NaN	NaN	NaN
3	NaN	3.0	NaN	4.0

In [77]:
```python
df1.dropna()
```

Out[77]:

A	B	C	D

In [79]:
```python
df1.dropna(how="any")
```

Out[79]:

A	B	C	D

In [80]:
```python
df1.dropna(how="all")
```

Out[80]:

	A	B	C	D
0	NaN	2.0	NaN	0.0
1	3.0	4.0	NaN	1.0
3	NaN	3.0	NaN	4.0

In [82]:
```python
df1.dropna(how="all",axis=1)
```

Out[82]:

	A	B	D
0	NaN	2.0	0.0
1	3.0	4.0	1.0
2	NaN	NaN	NaN
3	NaN	3.0	4.0

In [85]:
```python
df1.dropna(thresh=3)
```

Out[85]:

	A	B	C	D
1	3.0	4.0	NaN	1.0

In [86]:
```python
df1.dropna(thresh=2)
```

Out[86]:

	A	B	C	D
0	NaN	2.0	NaN	0.0
1	3.0	4.0	NaN	1.0
3	NaN	3.0	NaN	4.0

In [89]:
```python
df1
```

Out[89]:

	A	B	C	D
0	NaN	2.0	NaN	0.0
1	3.0	4.0	NaN	1.0
2	NaN	NaN	NaN	NaN
3	NaN	3.0	NaN	4.0

In [87]:
```python
df1.dropna(thresh=2,axis=1)
```

Out[87]:

	B	D
0	2.0	0.0
1	4.0	1.0
2	NaN	NaN
3	3.0	4.0

In [90]:
```python
df_pivot.isna().sum()
```

Out[90]:
```
Date            0
Drug_Name       0
Time            0
Pressure       13
Temperature    13
dtype: int64
```

In [91]:
```python
df_pivot
```

Out[91]:

	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

In [ ]:
```python
# Find way to replace values
# Create dataframe groupby --> Date and DrugName
```

String method in pandas

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

Handling datetime

In [ ]:

In [ ]:

In [ ]:

Writing to a file

Doubts

In [ ]:
```python
def custom(x):
    x['tt'] = x['tt'] + 1
    return x['tt']
df.groupby(['M']).apply(custom)
```