Untitled5

August 30, 2022

```python
[2]: # import data science package
     import pandas as pd
     import seaborn as sns
```

```python
[4]: # read the data
     data = pd.read_excel( r'C:\Users\zhuoxun.yang001\Desktop\ -    DLP   2020.1.
       ↪1-2022.8.23 .xlsx', sheet_name='  ')
```

```python
[5]: # check the shape
     data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 42297 entries, 0 to 42296
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   ID            42297 non-null  int64
 1                 42297 non-null  object
 2                 42297 non-null  object
 3                 42297 non-null  object
 4                 42297 non-null  object
 5                 1 non-null      object
 6   Destination   42297 non-null  object
 7   Severity      42297 non-null  object
 8   Action        42297 non-null  object
 9                 42297 non-null  object
 10  File Name     42109 non-null  object
dtypes: int64(1), object(10)
memory usage: 3.5+ MB
```

```python
[6]: # check the data head
     data.head(5)
```

```
[6]:          ID                                                      \
     0  1186978063  22 Aug. 2020, 12:54:50 PM GMT+0800    ---
     1  1186978147  22 Aug. 2020, 12:54:48 PM GMT+0800    ---
     2  1186970770  22 Aug. 2020, 12:54:08 PM GMT+0800    ---
     3  1186977831  22 Aug. 2020, 12:54:06 PM GMT+0800    ---
```

```
4   1185066658   21 Aug. 2020, 10:22:02 AM GMT+0800      ---

                                              Destination Severity  \
0   01-114              Endpoint LAN  NaN  \\192.168.4.234\        Medium
1   01-114              Endpoint LAN  NaN  \\192.168.4.234\        Medium
2   01-114              Endpoint LAN  NaN  \\192.168.4.234\        Medium
3   01-114              Endpoint LAN  NaN  \\192.168.4.234\        Medium
4   01-110-         Endpoint printing  NaN                        Medium

        Action                                          File Name
0  Permitted        \\192.168.4.234\     \04.     \12.2020    \…
1  Permitted        \\192.168.4.234\     \04.     \12.2020    \…
2  Permitted        \\192.168.4.234\     \04.     \12.2020    \…
3  Permitted        \\192.168.4.234\     \04.     \12.2020    \…
4  Permitted        Microsoft Word –                   …
```

[7]:
```python
# sum the null data
data.isnull().sum()
```

[7]:
```
ID                  0
                    0
                    0
                    0
                    0
                42296
Destination         0
Severity            0
Action              0
                    0
File Name         188
dtype: int64
```

[8]:
```python
# drop the null value based on column
data.dropna(inplace=True, axis=1)
```

[9]:
```python
# tranform the date format, unify the format
data['  ']= pd.to_datetime(data['  '], errors='coerce')
```

[10]:
```python
# check the shape of data
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 42297 entries, 0 to 42296
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   ID            42297 non-null  int64
 1                 27810 non-null  datetime64[ns, pytz.FixedOffset(-480)]
```

```
  2                 42297 non-null  object
  3                 42297 non-null  object
  4                 42297 non-null  object
  5   Destination  42297 non-null  object
  6   Severity      42297 non-null  object
  7   Action         42297 non-null  object
  8                 42297 non-null  object
dtypes: datetime64[ns, pytz.FixedOffset(-480)](1), int64(1), object(7)
memory usage: 2.9+ MB
```

[11]:
```python
# check the data head
data.head(5)
```

[11]:
```
          ID                                                              \
0  1186978063 2020-08-22 12:54:50-08:00      ---          01-114
1  1186978147 2020-08-22 12:54:48-08:00      ---          01-114
2  1186970770 2020-08-22 12:54:08-08:00      ---          01-114
3  1186977831 2020-08-22 12:54:06-08:00      ---          01-114
4  1185066658 2020-08-21 10:22:02-08:00      ---          01-110-


                            Destination Severity       Action
0         Endpoint LAN  \\192.168.4.234\    Medium  Permitted
1         Endpoint LAN  \\192.168.4.234\    Medium  Permitted
2         Endpoint LAN  \\192.168.4.234\    Medium  Permitted
3         Endpoint LAN  \\192.168.4.234\    Medium  Permitted
4  Endpoint printing                        Medium  Permitted
```

[12]:
```python
# subset the data columns
data1 = data[['   ', '   ']]
```

[13]:
```python
# applying lambda function to unify date format
data1['   '] = data1['   '].apply(lambda a: pd.to_datetime(a).date())
```

```
C:\Users\zhuoxun.yang001\AppData\Local\Temp\ipykernel_2692\3533790785.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  data1['   '] = data1['   '].apply(lambda a: pd.to_datetime(a).date())
```

[14]:
```python
# save the featurers into excel
output = 'C://Users//zhuoxun.yang001//Documents//fude//   //baokong.xlsx'
data1.to_excel(output)
```

[20]:
```python
# read the excel file using pandas
```

```
data2 = pd.read_excel( r'C:\Users\zhuoxun.yang001\Documents\fude\  \baokong.
  ↪xlsx', sheet_name='Sheet1')
```

[21]:
```
# check the shape
data2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 42297 entries, 0 to 42296
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Unnamed: 0  42297 non-null  int64
 1               27810 non-null  datetime64[ns]
 2               42297 non-null  object
dtypes: datetime64[ns](1), int64(1), object(1)
memory usage: 991.5+ KB
```

[22]:
```
# drop null values based on rows
data2.dropna(axis=0, inplace=True)
```

[23]:
```
# extract years from datetime & save to column
data2['  2'] = pd.DatetimeIndex(data2['  ']).year
```

[25]:
```
# extract months from datetime & save to column
data2['  3'] = pd.DatetimeIndex(data2['  ']).month
```

[28]:
```
# concate month and year and set formate to string
data2["year_month"] = data2['  2'].astype(str) +"."+ data2["  3"].astype(str)
```

[29]:
```
# check data2 head
data2.head()
```

[29]:
```
   Unnamed: 0               2     3 year_month
0           0 2020-08-22  2020     8    2020.8
1           1 2020-08-22  2020     8    2020.8
2           2 2020-08-22  2020     8    2020.8
3           3 2020-08-22  2020     8    2020.8
4           4 2020-08-21  2020     8    2020.8
```

[30]:
```
# output file save to the path
output = 'C://Users//zhuoxun.yang001//Documents//fude//  //baokong1.xlsx'
data2.to_excel(output)
```

[ ]:
```
# count several categories
```

[42]:
```
# using conditional sum to sum up values from different categories
cgsf_count = data2.query('  ==" " & year_month =="2020.10"')['  '].count()
gl_count = data2.query('  ==" " & year_month == "2020.10"')['  '].count()
```

```
lj_count = data2.query('  ==" " & year_month == "2020.10"')['  '].count()
sf_count = data2.query('  ==" " & year_month == "2020.10"')['  '].count()
yxsf_count = data2.query('   =="  " & year_month == "2020.10"')['  '].count()
print(cgsf_count, gl_count, lj_count, sf_count, yxsf_count)
```

201 0 7 0 0

```
[43]: # count several categories
cgsf_count1 = data2.query('  =="  " & year_month =="2020.11"')['  '].count()
gl_count1 = data2.query('  ==" " & year_month == "2020.11"')['  '].count()
lj_count1 = data2.query('  ==" " & year_month == "2020.11"')['  '].count()
sf_count1 = data2.query('  ==" " & year_month == "2020.11"')['  '].count()
yxsf_count1 = data2.query('   =="  " & year_month == "2020.11"')['  '].count()
print(cgsf_count1, gl_count1, lj_count1, sf_count1, yxsf_count1)
```

122 0 19 1 0

```
[44]: cgsf_count3 = data2.query('  =="  " & year_month =="2020.12"')['  '].count()
gl_count3 = data2.query('  ==" " & year_month == "2020.12"')['  '].count()
lj_count3 = data2.query('  ==" " & year_month == "2020.12"')['  '].count()
sf_count3 = data2.query('  ==" " & year_month == "2020.12"')['  '].count()
yxsf_count3 = data2.query('   =="  " & year_month == "2020.12"')['  '].count()
print(cgsf_count3, gl_count3, lj_count3, sf_count3, yxsf_count3)
```

816 3 21 0 0

```
[45]: cgsf_count4 = data2.query('  =="  " & year_month =="2020.2"')['  '].count()
gl_count4 = data2.query('  ==" " & year_month == "2020.2"')['  '].count()
lj_count4 = data2.query('  ==" " & year_month == "2020.2"')['  '].count()
sf_count4 = data2.query('  ==" " & year_month == "2020.2"')['  '].count()
yxsf_count4 = data2.query('   =="  " & year_month == "2020.2"')['  '].count()
print(cgsf_count4, gl_count4, lj_count4, sf_count4, yxsf_count4)
```

200 0 0 0 0

```
[46]: cgsf_count5 = data2.query('  =="  " & year_month =="2020.3"')['  '].count()
gl_count5 = data2.query('  ==" " & year_month == "2020.3"')['  '].count()
lj_count5 = data2.query('  ==" " & year_month == "2020.3"')['  '].count()
sf_count5 = data2.query('  ==" " & year_month == "2020.3"')['  '].count()
yxsf_count5 = data2.query('   =="  " & year_month == "2020.3"')['  '].count()
print(cgsf_count5, gl_count5, lj_count5, sf_count5, yxsf_count5)
```

214 0 0 0 0

```
[47]: cgsf_count6 = data2.query('  =="  " & year_month =="2020.4"')['  '].count()
gl_count6 = data2.query('  ==" " & year_month == "2020.4"')['  '].count()
lj_count6 = data2.query('  ==" " & year_month == "2020.4"')['  '].count()
sf_count6 = data2.query('  ==" " & year_month == "2020.4"')['  '].count()
yxsf_count6 = data2.query('   =="  " & year_month == "2020.4"')['  '].count()
print(cgsf_count6, gl_count6, lj_count6, sf_count6, yxsf_count6)
```

```
1799 0 23 0 0
```

```
[48]: cgsf_count7 = data2.query('  =="  " & year_month =="2020.5"')['  '].count()
      gl_count7 = data2.query('  ==" " & year_month == "2020.5"')['  '].count()
      lj_count7 = data2.query('  ==" " & year_month == "2020.5"')['  '].count()
      sf_count7 = data2.query('  ==" " & year_month == "2020.5"')['  '].count()
      yxsf_count7 = data2.query('  =="  " & year_month == "2020.5"')['  '].count()
      print(cgsf_count7, gl_count7, lj_count7, sf_count7, yxsf_count7)
```

```
140 0 55 0 0
```

```
[49]: cgsf_count8 = data2.query('  =="  " & year_month =="2020.6"')['  '].count()
      gl_count8 = data2.query('  ==" " & year_month == "2020.6"')['  '].count()
      lj_count8 = data2.query('  ==" " & year_month == "2020.6"')['  '].count()
      sf_count8 = data2.query('  ==" " & year_month == "2020.6"')['  '].count()
      yxsf_count8 = data2.query('  =="  " & year_month == "2020.6"')['  '].count()
      print(cgsf_count8, gl_count8, lj_count8, sf_count8, yxsf_count8)
```

```
242 0 43 0 0
```

```
[50]: cgsf_count9 = data2.query('  =="  " & year_month =="2020.7"')['  '].count()
      gl_count9 = data2.query('  ==" " & year_month == "2020.7"')['  '].count()
      lj_count9 = data2.query('  ==" " & year_month == "2020.7"')['  '].count()
      sf_count9 = data2.query('  ==" " & year_month == "2020.7"')['  '].count()
      yxsf_count9 = data2.query('  =="  " & year_month == "2020.7"')['  '].count()
      print(cgsf_count9, gl_count9, lj_count9, sf_count9, yxsf_count9)
```

```
509 0 26 0 0
```

```
[51]: cgsf_count10 = data2.query('  =="  " & year_month =="2020.8"')['  '].count()
      gl_count10 = data2.query('  ==" " & year_month == "2020.8"')['  '].count()
      lj_count10 = data2.query('  ==" " & year_month == "2020.8"')['  '].count()
      sf_count10 = data2.query('  ==" " & year_month == "2020.8"')['  '].count()
      yxsf_count10 = data2.query('  =="  " & year_month == "2020.8"')['  '].count()
      print(cgsf_count10, gl_count10, lj_count10, sf_count10, yxsf_count10)
```

```
940 0 25 0 0
```

```
[52]: cgsf_count11 = data2.query('  =="  " & year_month =="2020.9"')['  '].count()
      gl_count11 = data2.query('  ==" " & year_month == "2020.9"')['  '].count()
      lj_count11 = data2.query('  ==" " & year_month == "2020.9"')['  '].count()
      sf_count11 = data2.query('  ==" " & year_month == "2020.9"')['  '].count()
      yxsf_count11 = data2.query('  =="  " & year_month == "2020.9"')['  '].count()
      print(cgsf_count11, gl_count11, lj_count11, sf_count11, yxsf_count11)
```

```
427 0 30 0 0
```

```
[53]: cgsf_count12 = data2.query('  =="  " & year_month =="2021.1"')['  '].count()
      gl_count12 = data2.query('  ==" " & year_month == "2021.1"')['  '].count()
      lj_count12 = data2.query('  ==" " & year_month == "2021.1"')['  '].count()
```

```python
sf_count12 = data2.query('   ==" " & year_month == "2021.1"')['   '].count()
yxsf_count12 = data2.query('    ==" " & year_month == "2021.1"')['   '].count()
print(cgsf_count12, gl_count12, lj_count12, sf_count12, yxsf_count12)
```

134 1 26 0 0

```python
cgsf_count13 = data2.query('  =="  " & year_month =="2021.10"')['   '].count()
gl_count13 = data2.query('  ==" " & year_month == "2021.10"')['   '].count()
lj_count13 = data2.query('  ==" " & year_month == "2021.10"')['   '].count()
sf_count13 = data2.query('  ==" " & year_month == "2021.10"')['   '].count()
yxsf_count13 = data2.query('   ==" " & year_month == "2021.10"')['   '].count()
print(cgsf_count13, gl_count13, lj_count13, sf_count13, yxsf_count13)
```

40 0 15 0 0

```python
cgsf_count14 = data2.query('  =="  " & year_month =="2021.11"')['   '].count()
gl_count14 = data2.query('  ==" " & year_month == "2021.11"')['   '].count()
lj_count14 = data2.query('  ==" " & year_month == "2021.11"')['   '].count()
sf_count14 = data2.query('  ==" " & year_month == "2021.11"')['   '].count()
yxsf_count14 = data2.query('   ==" " & year_month == "2021.11"')['   '].count()
print(cgsf_count14, gl_count14, lj_count14, sf_count14, yxsf_count14)
```

368 0 27 0 0

```python
cgsf_count15 = data2.query('  =="  " & year_month =="2021.12"')['   '].count()
gl_count15 = data2.query('  ==" " & year_month == "2021.12"')['   '].count()
lj_count15 = data2.query('  ==" " & year_month == "2021.12"')['   '].count()
sf_count15 = data2.query('  ==" " & year_month == "2021.12"')['   '].count()
yxsf_count15 = data2.query('   ==" " & year_month == "2021.12"')['   '].count()
print(cgsf_count15, gl_count15, lj_count15, sf_count15, yxsf_count15)
```

266 0 11 0 0

```python
cgsf_count16 = data2.query('   =="  " & year_month =="2021.2"')['   '].count()
gl_count16 = data2.query('  ==" " & year_month == "2021.2"')['   '].count()
lj_count16 = data2.query('  ==" " & year_month == "2021.2"')['   '].count()
sf_count16 = data2.query('  ==" " & year_month == "2021.2"')['   '].count()
yxsf_count16 = data2.query('   ==" " & year_month == "2021.2"')['   '].count()
print(cgsf_count16, gl_count16, lj_count16, sf_count16, yxsf_count16)
```

103 1 18 0 0

```python
cgsf_count17 = data2.query('   =="  " & year_month =="2021.3"')['   '].count()
gl_count17 = data2.query('  ==" " & year_month == "2021.3"')['   '].count()
lj_count17 = data2.query('  ==" " & year_month == "2021.3"')['   '].count()
sf_count17 = data2.query('  ==" " & year_month == "2021.3"')['   '].count()
yxsf_count17 = data2.query('   ==" " & year_month == "2021.3"')['   '].count()
print(cgsf_count17, gl_count17, lj_count17, sf_count17, yxsf_count17)
```

3754 2 37 0 0

```
[59]: cgsf_count18 = data2.query('    =="   " & year_month =="2021.4"')['   '].count()
      gl_count18 = data2.query('   ==" " & year_month == "2021.4"')['   '].count()
      lj_count18 = data2.query('   ==" " & year_month == "2021.4"')['   '].count()
      sf_count18 = data2.query('   ==" " & year_month == "2021.4"')['   '].count()
      yxsf_count18 = data2.query('    =="   " & year_month == "2021.4"')['   '].count()
      print(cgsf_count18, gl_count18, lj_count18, sf_count18, yxsf_count18)
```

1100 1 30 0 0

```
[60]: cgsf_count19 = data2.query('    =="   " & year_month =="2021.5"')['   '].count()
      gl_count19 = data2.query('   ==" " & year_month == "2021.5"')['   '].count()
      lj_count19 = data2.query('   ==" " & year_month == "2021.5"')['   '].count()
      sf_count19 = data2.query('   ==" " & year_month == "2021.5"')['   '].count()
      yxsf_count19 = data2.query('    =="   " & year_month == "2021.5"')['   '].count()
      print(cgsf_count19, gl_count19, lj_count19, sf_count19, yxsf_count19)
```

3442 0 35 0 0

```
[62]: cgsf_count20 = data2.query('    =="   " & year_month =="2021.6"')['   '].count()
      gl_count20 = data2.query('   ==" " & year_month == "2021.6"')['   '].count()
      lj_count20 = data2.query('   ==" " & year_month == "2021.6"')['   '].count()
      sf_count20 = data2.query('   ==" " & year_month == "2021.6"')['   '].count()
      yxsf_count20 = data2.query('    =="   " & year_month == "2021.6"')['   '].count()
      print(cgsf_count20, gl_count20, lj_count20, sf_count20, yxsf_count20)
```

1983 0 52 0 0

```
[63]: cgsf_count21 = data2.query('    =="   " & year_month =="2021.7"')['   '].count()
      gl_count21 = data2.query('   ==" " & year_month == "2021.7"')['   '].count()
      lj_count21 = data2.query('   ==" " & year_month == "2021.7"')['   '].count()
      sf_count21 = data2.query('   ==" " & year_month == "2021.7"')['   '].count()
      yxsf_count21 = data2.query('    =="   " & year_month == "2021.7"')['   '].count()
      print(cgsf_count21, gl_count21, lj_count21, sf_count21, yxsf_count21)
```

726 0 14 1 0

```
[64]: cgsf_count22 = data2.query('    =="   " & year_month =="2021.8"')['   '].count()
      gl_count22 = data2.query('   ==" " & year_month == "2021.8"')['   '].count()
      lj_count22 = data2.query('   ==" " & year_month == "2021.8"')['   '].count()
      sf_count22 = data2.query('   ==" " & year_month == "2021.8"')['   '].count()
      yxsf_count22 = data2.query('    =="   " & year_month == "2021.8"')['   '].count()
      print(cgsf_count22, gl_count22, lj_count22, sf_count22, yxsf_count22)
```

1405 0 26 0 0

```
[65]: cgsf_count23 = data2.query('    =="   " & year_month =="2021.9"')['   '].count()
      gl_count23 = data2.query('   ==" " & year_month == "2021.9"')['   '].count()
      lj_count23 = data2.query('   ==" " & year_month == "2021.9"')['   '].count()
      sf_count23 = data2.query('   ==" " & year_month == "2021.9"')['   '].count()
      yxsf_count23 = data2.query('    =="   " & year_month == "2021.9"')['   '].count()
```

```
print(cgsf_count23, gl_count23, lj_count23, sf_count23, yxsf_count23)
```

909 1 26 0 0

```
[66]: cgsf_count24 = data2.query('  =="  " & year_month =="2022.1"')['  '].count()
      gl_count24 = data2.query('  =="  " & year_month == "2022.1"')['  '].count()
      lj_count24 = data2.query('  =="  " & year_month == "2022.1"')['  '].count()
      sf_count24 = data2.query('  =="  " & year_month == "2022.1"')['  '].count()
      yxsf_count24 = data2.query('   =="  " & year_month == "2022.1"')['  '].count()
      print(cgsf_count24, gl_count24, lj_count24, sf_count24, yxsf_count24)
```

1164 1 29 0 0

```
[67]: cgsf_count25 = data2.query('  =="  " & year_month =="2022.2"')['  '].count()
      gl_count25 = data2.query('  =="  " & year_month == "2022.2"')['  '].count()
      lj_count25 = data2.query('  =="  " & year_month == "2022.2"')['  '].count()
      sf_count25 = data2.query('  =="  " & year_month == "2022.2"')['  '].count()
      yxsf_count25 = data2.query('   =="  " & year_month == "2022.2"')['  '].count()
      print(cgsf_count25, gl_count25, lj_count25, sf_count25, yxsf_count25)
```

3532 2 28 0 0

```
[69]: cgsf_count26 = data2.query('  =="  " & year_month =="2022.3"')['  '].count()
      gl_count26 = data2.query('  =="  " & year_month == "2022.3"')['  '].count()
      lj_count26 = data2.query('  =="  " & year_month == "2022.3"')['  '].count()
      sf_count26 = data2.query('  =="  " & year_month == "2022.3"')['  '].count()
      yxsf_count26 = data2.query('   =="  " & year_month == "2022.3"')['  '].count()
      print(cgsf_count26, gl_count26, lj_count26, sf_count26, yxsf_count26)
```

12 8 26 0 0

```
[70]: cgsf_count27 = data2.query('  =="  " & year_month =="2022.4"')['  '].count()
      gl_count27 = data2.query('  =="  " & year_month == "2022.4"')['  '].count()
      lj_count27 = data2.query('  =="  " & year_month == "2022.4"')['  '].count()
      sf_count27 = data2.query('  =="  " & year_month == "2022.4"')['  '].count()
      yxsf_count27 = data2.query('   =="  " & year_month == "2022.4"')['  '].count()
      print(cgsf_count27, gl_count27, lj_count27, sf_count27, yxsf_count27)
```

85 3 75 0 0

```
[71]: cgsf_count28 = data2.query('  =="  " & year_month =="2022.5"')['  '].count()
      gl_count28 = data2.query('  =="  " & year_month == "2022.5"')['  '].count()
      lj_count28 = data2.query('  =="  " & year_month == "2022.5"')['  '].count()
      sf_count28 = data2.query('  =="  " & year_month == "2022.5"')['  '].count()
      yxsf_count28 = data2.query('   =="  " & year_month == "2022.5"')['  '].count()
      print(cgsf_count28, gl_count28, lj_count28, sf_count28, yxsf_count28)
```

50 0 18 1 0

```
[72]: cgsf_count29 = data2.query('  =="  " & year_month =="2022.6"')['  '].count()
      gl_count29 = data2.query('  ==" " & year_month == "2022.6"')['  '].count()
      lj_count29 = data2.query('  ==" " & year_month == "2022.6"')['  '].count()
      sf_count29 = data2.query('  ==" " & year_month == "2022.6"')['  '].count()
      yxsf_count29 = data2.query('   =="  " & year_month == "2022.6"')['  '].count()
      print(cgsf_count29, gl_count29, lj_count29, sf_count29, yxsf_count29)
```

1153 1 25 2 0

```
[73]: cgsf_count30 = data2.query('  =="  " & year_month =="2022.7"')['  '].count()
      gl_count30 = data2.query('  ==" " & year_month == "2022.7"')['  '].count()
      lj_count30 = data2.query('  ==" " & year_month == "2022.7"')['  '].count()
      sf_count30 = data2.query('  ==" " & year_month == "2022.7"')['  '].count()
      yxsf_count30 = data2.query('   =="  " & year_month == "2022.7"')['  '].count()
      print(cgsf_count30, gl_count30, lj_count30, sf_count30, yxsf_count30)
```

746 0 20 0 0

```
[74]: cgsf_count31 = data2.query('  =="  " & year_month =="2022.8"')['  '].count()
      gl_count31 = data2.query('  ==" " & year_month == "2022.8"')['  '].count()
      lj_count31 = data2.query('  ==" " & year_month == "2022.8"')['  '].count()
      sf_count31 = data2.query('  ==" " & year_month == "2022.8"')['  '].count()
      yxsf_count31 = data2.query('   =="  " & year_month == "2022.8"')['  '].count()
      print(cgsf_count31, gl_count31, lj_count31, sf_count31, yxsf_count31)
```

379 1 14 0 0

```
[146]: # use pandas to read excel
       data3 = pd.read_excel(r'C://Users//zhuoxun.yang001//Documents//fude//  //
         baokong1.xlsx', sheet_name='Sheet2')
```

```
[147]: # check the data head
       data3.head(5)
```

```
[147]:         date date_year  cgsf  gl  lj  sf  yxsf
       0  2020-02-01   2020/02    200   0   0   0     0
       1  2020-03-01   2020/03    214   0   0   0     0
       2  2020-04-01   2020/04   1799   0  23   0     0
       3  2020-05-01   2020/05    140   0  55   0     0
       4  2020-06-01   2020/06    242   0  43   0     0
```

```
[246]: # subset data3 several categories
       data4 = data3[['date_year', 'cgsf', 'gl', 'lj', 'sf', 'yxsf']]
```

```
[247]: # check data4 head
       data4.head(5)
```
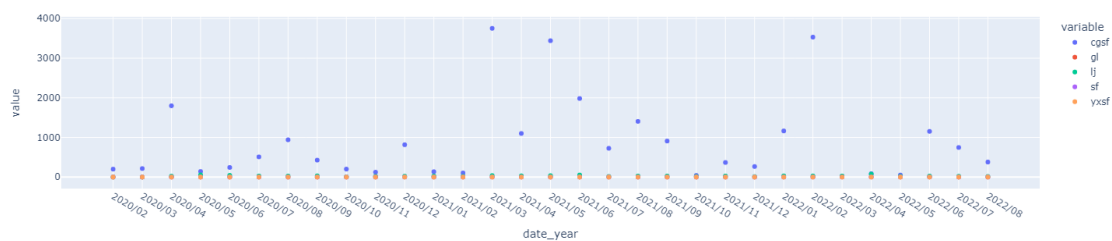
```
[247]:   date_year  cgsf  gl  lj  sf  yxsf
       0   2020/02    200   0   0   0     0
```

```
1    2020/03     214    0    0    0       0
2    2020/04    1799    0   23    0       0
3    2020/05     140    0   55    0       0
4    2020/06     242    0   43    0       0
```

[159]:
```python
# transform the date_year str to datetime
data4['date_year_converted'] = pd.to_datetime(data4['date_year'])
```

[248]:
```python
# import plotly and applying scatter
import plotly.express as pl
fig = pl.scatter(data4, x='date_year', y=['cgsf', 'gl', 'lj', 'sf', 'yxsf'])
fig.show()
```


[265]:
```python
# import plotly to set axis and legend
import plotly.graph_objects as go

fig = go.Figure()

fig.add_trace(go.Scatter(
    x=data4['date_year'],
    y=data4['cgsf'],
    name="   "
))


fig.add_trace(go.Scatter(
    x=data4['date_year'],
    y=data4['gl'],
    name=" ",
    yaxis="y2"
))

fig.add_trace(go.Scatter(
    x=data4['date_year'],
    y=data4['lj'],
    name=" ",
```

```python
        yaxis="y3"
))

fig.add_trace(go.Scatter(
    x=data4['date_year'],
    y=data4['sf'],
    name=" ",
    yaxis="y4"
))

fig.add_trace(go.Scatter(
    x=data4['date_year'],
    y=data4['yxsf'],
    name="   ",
    yaxis="y5"
))


# Create axis objects
fig.update_layout(
    xaxis=dict(
        domain=[0.25, 0.75]
    ),
    yaxis=dict(
        title="  ",
        titlefont=dict(
            color="#1f77b4"
        ),
        tickfont=dict(
            color="#1f77b4"
        )
    ),
    yaxis2=dict(
        title=" ",
        titlefont=dict(
            color="#ff7f0e"
        ),
        tickfont=dict(
            color="#ff7f0e"
        ),
        anchor="free",
        overlaying="y",
        side="left",
        position=0.15
    ),
    yaxis3=dict(
        title=" ",
```

```python
            titlefont=dict(
                color="#d62728"
            ),
            tickfont=dict(
                color="#d62728"
            ),
            anchor="x",
            overlaying="y",
            side="right"
        ),
        yaxis4=dict(
            title=" ",
            titlefont=dict(
                color="#9467bd"
            ),
            tickfont=dict(
                color="#9467bd"
            ),
            anchor="free",
            overlaying="y",
            side="right"
        ),
        yaxis5=dict(
            title="   ",
            titlefont=dict(
                color="#9467bd"
            ),
            tickfont=dict(
                color="#9467bd"
            ),
            anchor="free",
            overlaying="y",
            side="right",
            position=0.85
        )
)

# Update layout properties
fig.
↪update_layout(title_text="    ",width=800,autosize=True,title_font_family='yahei',␣
↪xaxis=dict(tickangle=45))

fig.show()
```
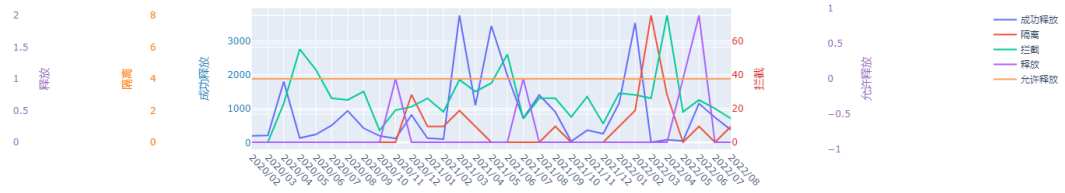
防火墙行为拦截

[ ]: