November 16, 2020

[1]:
```
import os, re
import pandas as pd
import numpy as np
```

[2]:
```
#     keywords
data_path = os.path.join('data', '    .csv')
data = pd.read_csv(data_path)
pattern = r'[^\u4e00-\u9fa5\d]'
data['  '] = data['  '].apply(lambda x: re.sub(pattern, '', x))
data.to_csv(data_path, index=False, encoding="utf_8_sig")
```

[3]:
```
#
result = pd.DataFrame()
for wj in os.listdir('data'):
    data_path = os.path.join('data', wj)
    data = pd.read_csv(data_path)
    result = result.append(data)
result.to_csv("data/     .csv", index=False, encoding="utf_8_sig")
```

[4]:
```
#
data = pd.read_csv("data/     .csv")

#
data = data[data.   <= '2020-11-09']
print(f" 2020-11-09            {len(data)}  ")
```

```
 2020-11-09            2108
```

[5]:
```
#
_data = data.groupby('  ').count().reset_index()[['  ', '  ']]
_data.rename(columns={"  ": "  "}, inplace=True)
```

[6]:
```
# 2020-01-30
num1 = _data[_data.   <= '2020-01-30'].   .sum()
data0 = pd.DataFrame([['2020-01-30 ', num1]], columns=['  ', '  '])
# 2020-02-01 ~ 2020-02-21
data1 = _data[(_data.   >= '2020-02-01') & (_data.   <= '2020-02-21')]
```

[7]:
```
# 2020-02-21 ~ 2020-11-05
num2 = _data[(_data.    >= '2020-02-21')
             & (_data.    <= '2020-11-05')].   .sum()

# 2020-11-06 ~ 2020-11-09      2020-11-09
print(f"2020-11-06   {_data[_data.   =='2020-11-06'].iloc[0,1]} ")

data2 = _data[(_data.    > '2020-11-06') & (_data.    <= '2020-11-09')]
```

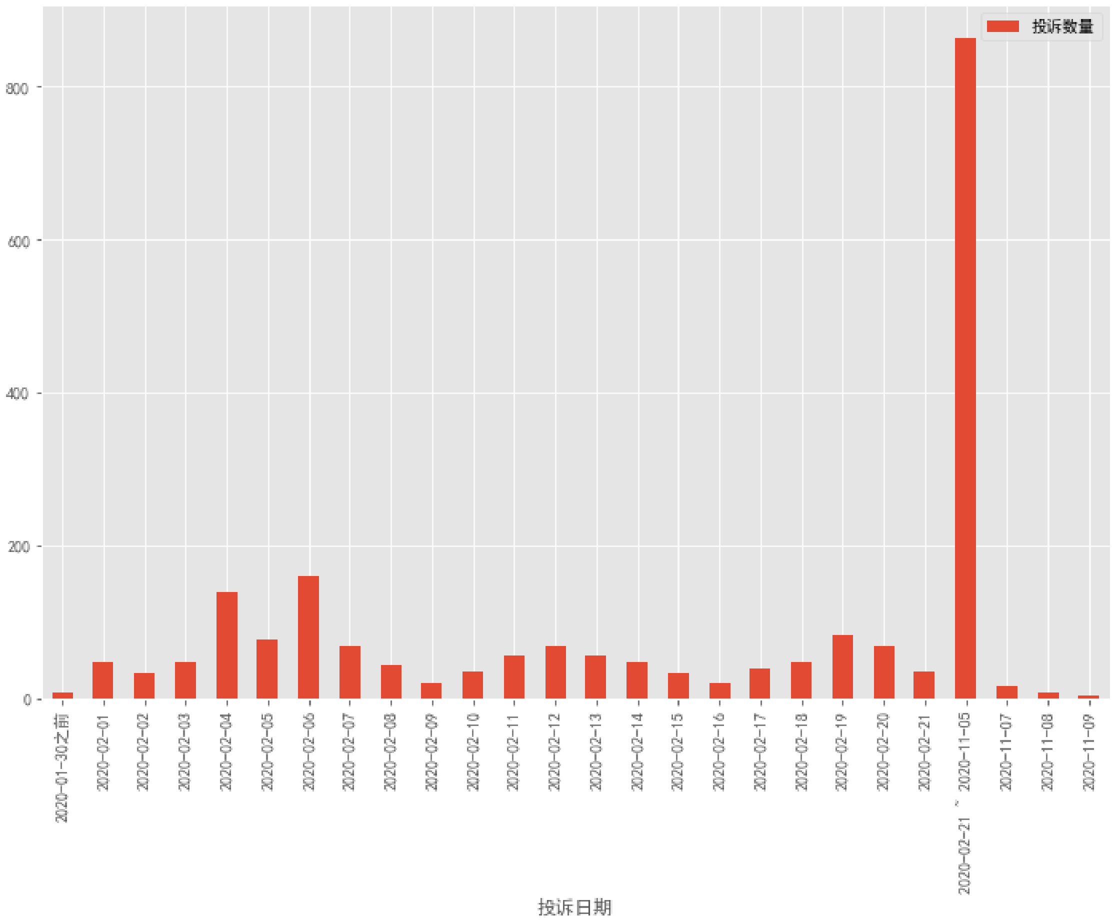
```
2020-11-06   16
```

[8]:
```
data3 = pd.DataFrame([['2020-02-21 ~ 2020-11-05', num2]],
                     columns=['   ', '   '])
```

[9]:
```
new_data = pd.concat([data0, data1, data3, data2])
```

[10]:
```
'''   '''
import matplotlib.pyplot as plt
%matplotlib inline
plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['font.size'] = 18
plt.rcParams['figure.figsize'] = (12, 8)
plt.style.use("ggplot")
new_data.set_index('   ').plot(kind='bar')  #   2020-11-06   24093
```

[10]:
```
<AxesSubplot:xlabel='   '>
```



[11]:
```
import jieba
import re
import collections
import PIL.Image as img
from wordcloud import WordCloud
import PIL.Image as img
from wordcloud import WordCloud
```

[12]:
```
#
all_word = ''
for line in data.values:
    word = line[4]
    all_word = all_word + word

# jieba
result = list(jieba.cut(all_word))
```

```
Building prefix dict from the default dictionary …
Loading model from cache C:\Users\Steven\AppData\Local\Temp\jieba.cache
Loading model cost 0.810 seconds.
Prefix dict has been built successfully.
```

[13]:

```
#
wordcloud = WordCloud(
    width=800,
    height=600,
    background_color='white',
    font_path='MSYH.TTC',   #
    max_font_size=500,
    min_font_size=20).generate(' '.join(result))
image = wordcloud.to_image()
image.show()   #
wordcloud.to_file('    .png')   #
```

[13]:

```
<wordcloud.wordcloud.WordCloud at 0x1e73ceec708>
```

[14]:

```
#
all_word = ''
for line in data.values:
    word = line[2]
    all_word = all_word + word

# jieba
result = list(jieba.cut(all_word))
```

[15]:

```
#
#
wordcloud = WordCloud(
    width=800,
    height=600,
    background_color='white',
    font_path='MSYH.TTC',   #
    max_font_size=500,
    min_font_size=20).generate(' '.join(result))
image = wordcloud.to_image()
image.show()   #
wordcloud.to_file('    .png')   #
```

[15]:

```
<wordcloud.wordcloud.WordCloud at 0x1e74062b608>
```

[16]:

```
#
all_word = ''
for line in data.values:
    word = line[3]
    all_word = all_word + word

# jieba
```

```
result = list(jieba.cut(all_word))

#
#
wordcloud = WordCloud(
    width=800,
    height=600,
    background_color='white',
    font_path='MSYH.TTC',  #
    max_font_size=500,
    min_font_size=20).generate(' '.join(result))
image = wordcloud.to_image()
image.show()  #
wordcloud.to_file('    .png')  #
```

```
[16]: <wordcloud.wordcloud.WordCloud at 0x1e74062bf48>
```