美国2012年总统候选人政治献金数据分析

导入包

In [3]:
```
import numpy as np
import pandas as pd
from pandas import Series,DataFrame
```

方便大家操作，将月份和参选人以及所在政党进行定义

In [4]:
```
months = {'JAN' : 1, 'FEB' : 2, 'MAR' : 3, 'APR' : 4, 'MAY' : 5, 'JUN' : 6,
          'JUL' : 7, 'AUG' : 8, 'SEP' : 9, 'OCT': 10, 'NOV': 11, 'DEC' : 12}
of_interest = ['Obama, Barack', 'Romney, Mitt', 'Santorum, Rick',
               'Paul, Ron', 'Gingrich, Newt']
parties = {
  'Bachmann, Michelle': 'Republican',
  'Romney, Mitt': 'Republican',
  'Obama, Barack': 'Democrat',
  "Roemer, Charles E. 'Buddy' III": 'Reform',
  'Pawlenty, Timothy': 'Republican',
  'Johnson, Gary Earl': 'Libertarian',
  'Paul, Ron': 'Republican',
  'Santorum, Rick': 'Republican',
  'Cain, Herman': 'Republican',
  'Gingrich, Newt': 'Republican',
  'McCotter, Thaddeus G': 'Republican',
  'Huntsman, Jon': 'Republican',
  'Perry, Rick': 'Republican'
 }
```

读取文件

In [6]:
```
data = pd.read_csv('./data/usa_election.txt')
data.head()
```

```
C:\Users\KnightPlan\Anaconda3\lib\site-packages\IPython\core\interactiveshell.py:2785: DtypeWarning: Columns (6) have mixed types.
Specify dtype option on import or set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)
```

Out[6]:

	cmte_id	cand_id	cand_nm	contbr_nm	contbr_city	contbr_st	contbr_zip	contbr_employer	contbr_occupation	contb_receipt_amt
0	C00410118	P20002978	Bachmann, Michelle	HARVEY, WILLIAM	MOBILE	AL	3.6601e+08	RETIRED	RETIRED	250.0
1	C00410118	P20002978	Bachmann, Michelle	HARVEY, WILLIAM	MOBILE	AL	3.6601e+08	RETIRED	RETIRED	50.0
2	C00410118	P20002978	Bachmann, Michelle	SMITH, LANIER	LANETT	AL	3.68633e+08	INFORMATION REQUESTED	INFORMATION REQUESTED	250.0
3	C00410118	P20002978	Bachmann, Michelle	BLEVINS, DARONDA	PIGGOTT	AR	7.24548e+08	NONE	RETIRED	250.0
4	C00410118	P20002978	Bachmann, Michelle	WARDENBURG, HAROLD	HOT SPRINGS NATION	AR	7.19016e+08	NONE	RETIRED	300.0

查看文件样式以及基本信息

【知识点】使用map函数+字典，新建一列各个候选人所在党派party

In [10]:
```
data['party'] = data['cand_nm'].map(parties)
data.head()
```

...

查看单独一行，是否加上了'party'一列

使用np.unique()函数查看colums：party这一列中有哪些元素

In [11]:
```
data['party'].unique()
```

Out[11]:
```
array(['Republican', 'Democrat', 'Reform', 'Libertarian'], dtype=object)
```

使用value_counts()函数，统计party列中各个元素出现次数

In [12]:
```
data['party'].value_counts()
```

Out[12]:
```
Democrat       292400
Republican     237575
Reform           5364
Libertarian       702
Name: party, dtype: int64
```

【知识点】使用groupby()函数，查看各个党派收到的政治献金总数contb_receipt_amt

In [13]:
```
data.groupby(by='party')['contb_receipt_amt'].sum()
```

Out[13]:
```
party
Democrat       8.105758e+07
Libertarian    4.132769e+05
Reform         3.390338e+05
Republican     1.192255e+08
Name: contb_receipt_amt, dtype: float64
```

查看具体每天各个党派收到的政治献金总数contb_receipt_amt
使用groupby([多个分组参数])

In [14]:
```
data.groupby(by=['party','contb_receipt_dt'])['contb_receipt_amt'].sum()
```

...

查看日期格式，并将其转换为'yyyy-mm-dd'日期格式,通过函数加map方式进行转换:months['月份简写']==》mm形式的月份

In [15]:
```
def transform_date(d):  #20-JUN-11
    day,month,year = d.split('-')
    month = str(months[month])

    return '20'+year+'-'+month+'-'+day
```

In [17]:
```
data['contb_receipt_dt'] = data['contb_receipt_dt'].map(transform_date)
data.head()
```

Out[17]:

	cmte_id	cand_id	cand_nm	contbr_nm	contbr_city	contbr_st	contbr_zip	contbr_employer	contbr_occupation	contb_receipt_amt
0	C00410118	P20002978	Bachmann, Michelle	HARVEY, WILLIAM	MOBILE	AL	3.6601e+08	RETIRED	RETIRED	250.0
1	C00410118	P20002978	Bachmann, Michelle	HARVEY, WILLIAM	MOBILE	AL	3.6601e+08	RETIRED	RETIRED	50.0
2	C00410118	P20002978	Bachmann, Michelle	SMITH, LANIER	LANETT	AL	3.68633e+08	INFORMATION REQUESTED	INFORMATION REQUESTED	250.0
3	C00410118	P20002978	Bachmann, Michelle	BLEVINS, DARONDA	PIGGOTT	AR	7.24548e+08	NONE	RETIRED	250.0
4	C00410118	P20002978	Bachmann, Michelle	WARDENBURG, HAROLD	HOT SPRINGS NATION	AR	7.19016e+08	NONE	RETIRED	300.0

查看是否转换成功

查看老兵(捐献者职业)DISABLED VETERAN主要支持谁： 查看老兵们捐赠给谁的钱最多 考察Series索引

In [19]:
```
data['contbr_occupation'] == 'DISABLED VETERAN'
```

...

In [21]:
```
old = data.loc[data['contbr_occupation'] == 'DISABLED VETERAN']
```

In [24]:
```
old.groupby(by='cand_nm')['contb_receipt_amt'].sum()
```

Out[24]:
```
cand_nm
Cain, Herman        300.00
Obama, Barack      4205.00
Paul, Ron          2425.49
Santorum, Rick      250.00
Name: contb_receipt_amt, dtype: float64
```

找出候选人的捐赠者中，捐赠金额最大的人的职业以及捐献额

通过query("查询条件来查找捐献人职业")

In [25]:
```
data['contb_receipt_amt'].max()
```

Out[25]:
```
1944042.43
```

In [26]:
```
data.query('contb_receipt_amt == 1944042.43 ')
```

Out[26]:

	cmte_id	cand_id	cand_nm	contbr_nm	contbr_city	contbr_st	contbr_zip	contbr_employer	contbr_occupation	contb_receipt_amt
176127	C00431445	P80003338	Obama, Barack	OBAMA VICTORY FUND 2012 - UNITEMIZED	CHICAGO	IL	60680	NaN	NaN	1944042.43

In [ ]:
```

```