Business Analytics Programming

Lab 1b (Pandas, NY 2016 Fundraising Data)

Dr. Wajahat Gilani

Rutgers Business School

February 7, 2019

Lab 1 - Analyze the NY Fund-Raising Data

You are a data-scientist hired by a political candidate to analyze any possible trends of NY donors. The following questions the campaign wants to know.

1. Whether its possible to identify the 'Party' for each candidate (data wrangling)
2. Convert the contb_receipt_dt column into an actual date object (data wrangling)
3. Using group by, show the number (count) of donations given to each party
4. Using group by, show the number of donations given to each party, over time
5. Using group by, show the total dollar amount of donations given to each party
6. Using group by, show the total dollar amount of donations given to each party, over time
7. Which occupations donated the top 5 most money?

Lab 1 - Analyze the NY Fund-Raising Data (Continued)

8. Which occupations donated the least 5 amount of money?
9. Which employer's employees gave the most money, give the top 5.
10. For each candidate, what were the top 5 occupations that donated to their election
11. For the 5 candidates that raised the most money, graph their donations by time, in a line graph

Question 1: Identify Party

Political candidates belong to a political party, but that column is missing from the table. Given a list of associations we need to create a column that has each candidates party affiliation.

Candidate Name	**Politial Party**
Clinton, Hillary Rodham	Democrat
Sanders, Bernard	Democrat
Trump, Donald J.	Republican
Cruz, Rafael Edward 'Ted'	Republican
Carson, Benjamin S.	Republican
Rubio, Marco	Republican
Bush, Jeb	Republican
Kasich, John R.	Republican
Fiorina, Carly	Republican
Paul, Rand	Republican
Stein, Jill	Green
Johnson, Gary	Libertarian

Question 1: Identify Party (Continued)

Candidate Name	**Politial Party**
Christie, Christopher J.	Republican
Graham, Lindsey O.	Republican
O'Malley, Martin Joseph	Democrat
Walker, Scott	Republican
Huckabee, Mike	Republican
Pataki, George E.	Republican
Lessig, Lawrence	Democrat
McMullin, Evan	Independent
Santorum, Richard J.	Republican
Webb, James Henry Jr.	Democrat
Perry, James R. (Rick)	Republican
Jindal, Bobby	Republican
Gilmore, James S III	Republican

Question 1: Identify Party - Solution

- Create a table of the politicians names (unique names) and their political party
- Convert the table to a dictionary
- Map that dictionary into a new column

```
dfc = nyc.cand_nm.value_counts()

type(dfc)
ucm = dfc.index.values
dfc2 = pd.DataFrame({'cand_nm':ucm})
dfc2.loc[[0,1,14,21,18],'Party']='Democrat'
```

	cand_nm	Party
0	Clinton, Hillary Rodham	Democrat
1	Sanders, Bernard	Democrat
2	Trump, Donald J.	NaN
3	Cruz, Rafael Edward 'Ted'	NaN
4	Carson, Benjamin S.	NaN
5	Rubio, Marco	NaN

Question 1: Identify Party - Solution (Continued)

```
dfc2.loc[10,'Party']='Green'
dfc2.loc[11,'Party']='Libertarian'
dfc2.loc[19,'Party']='Independent'
dfc2.loc[dfc2['Party'].isnull(),'Party']='Republican'
```

	cand_nm	Party
0	Clinton, Hillary Rodham	Democrat
1	Sanders, Bernard	Democrat
2	Trump, Donald J.	Republican
3	Cruz, Rafael Edward 'Ted'	Republican
4	Carson, Benjamin S.	Republican
5	Rubio, Marco	Republican
6	Bush, Jeb	Republican
7	Kasich, John R.	Republican
8	Fiorina, Carly	Republican
9	Paul, Rand	Republican
10	Stein, Jill	Green
11	Johnson, Gary	Libertarian

Question 1: Identify Party - Solution

- Create a table of the politicians names (unique names) and their political party
- Convert the table to a dictionary
- Map that dictionary into a new column

```
dfc = nyc.cand_nm.value_counts()

type(dfc)
ucm = dfc.index.values
dfc2 = pd.DataFrame({'cand_nm':ucm})
dfc2.loc[[0,1,14,21,18],'Party']='Democrat'
```

	cand_nm	Party
0	Clinton, Hillary Rodham	Democrat
1	Sanders, Bernard	Democrat
2	Trump, Donald J.	NaN
3	Cruz, Rafael Edward 'Ted'	NaN
4	Carson, Benjamin S.	NaN
5	Rubio, Marco	NaN

Question 1: Identify Party - Solution (Continued)

```
dfc2.loc[10,'Party']='Green'
dfc2.loc[11,'Party']='Libertarian'
dfc2.loc[19,'Party']='Independent'
dfc2.loc[dfc2['Party'].isnull(),'Party']='Republican'
```

	cand_nm	Party
0	Clinton, Hillary Rodham	Democrat
1	Sanders, Bernard	Democrat
2	Trump, Donald J.	Republican
3	Cruz, Rafael Edward 'Ted'	Republican
4	Carson, Benjamin S.	Republican
5	Rubio, Marco	Republican
6	Bush, Jeb	Republican
7	Kasich, John R.	Republican
8	Fiorina, Carly	Republican
9	Paul, Rand	Republican
10	Stein, Jill	Green
11	Johnson, Gary	Libertarian

Question 1: Identify Party - Solution (Dictionary)

```
1 cand_dict = dict(zip(dfc2.cand_nm, dfc2.Party))
```

{'Clinton, Hillary Rodham': 'Democrat', 'Sanders, Bernard': 'Democrat', 'Trump, Donald J.': 'Republican', "Cruz, Rafael Edward 'Ted'": 'Republican', 'Carson, Benjamin S.': 'Republican', 'Rubio, Marco': 'Republican', 'Bush, Jeb': 'Republican', 'Kasich, John R.': 'Republican', 'Fiorina, Carly': 'Republican', 'Paul, Rand': 'Republican', 'Stein, Jill': 'Green', 'Johnson, Gary': 'Libertarian', 'Christie, Christopher J.': 'Republican', 'Graham, Lindsey O.': 'Republican', "O'Malley, Martin Joseph": 'Democrat', 'Walker, Scott': 'Republican', 'Huckabee, Mike': 'Republican', 'Pataki, George E.': 'Republican', 'Lessig, Lawrence': 'Democrat', 'McMullin, Evan': 'Independent', 'Santorum, Richard J.': 'Republican', 'Webb, James Henry Jr.': 'Democrat', 'Perry, James R. (Rick)': 'Republican', 'Jindal, Bobby': 'Republican', 'Gilmore, James S III': 'Republican'}

Question 1: Identify Party - Solution (Mapping)

```
1 nyc['Party'] = nyc['cand_nm'].map(cand_dict)
2 print(nyc[['cand_nm','Party']].head(10))
```

	cand_nm	Party
0	Clinton, Hillary Rodham	Democrat
1	Clinton, Hillary Rodham	Democrat
2	Sanders, Bernard	Democrat
3	Sanders, Bernard	Democrat
4	Clinton, Hillary Rodham	Democrat
5	Sanders, Bernard	Democrat
6	Sanders, Bernard	Democrat
7	Clinton, Hillary Rodham	Democrat
8	Clinton, Hillary Rodham	Democrat
9	Trump, Donald J.	Republican

Question 2: Convert the contb_receipt_dt to Date type

```
1 nyc['Date'] = pd.to_datetime(nyc['contb_receipt_dt'])
2 print(nyc[['contb_receipt_dt','Date']].head(10))
```

	contb_receipt_dt	Date
0	15-APR-16	2016-04-15
1	24-APR-16	2016-04-24
2	06-MAR-16	2016-03-06
3	04-MAR-16	2016-03-04
4	12-APR-16	2016-04-12
5	05-MAR-16	2016-03-05
6	05-MAR-16	2016-03-05
7	19-APR-16	2016-04-19
8	08-APR-16	2016-04-08
9	01-SEP-16	2016-09-01

Group By Example

```
A=pd.Series(['red','blue','yellow','orange','red','blue','
    yellow','orange'])
B=pd.Series([1,1,1,1,2,2,2,2])
Price = pd.Series(np.arange(1,9))
dfexmp=pd.DataFrame({'A':A,'B':B,'Price':Price})
```

	A	B	Price
0	red	1	1
1	blue	1	2
2	yellow	1	3
3	orange	1	4
4	red	2	5
5	blue	2	6
6	yellow	2	7
7	orange	2	8

Group By Example - Groups

```
dfexmp.groupby('A')
```

	A	B	Price
0	red	1	1
4	red	2	5

	A	B	Price
1	blue	1	2
5	blue	2	6

	A	B	Price
2	yellow	1	3
6	yellow	2	7

	A	B	Price
3	orange	1	4
7	orange	2	8

```
dfexmp.groupby('A')['Price']
```

	A	Price
0	red	1
4	red	5

	A	Price
1	blue	2
5	blue	6

	A	Price
2	yellow	3
6	yellow	7

	A	Price
3	orange	4
7	orange	8

Group By Example - Group Computation

```
1 dfexmp.groupby('A')['Price'].sum()
```

A	Price
red	6

A	Price
blue	8

A	Price
yellow	10

A	Price
orange	12

A	Price
blue	8
orange	12
red	6
yellow	10

Question 3: Party's Frequency of donations

```
1 nyc.groupby('Party')['contb_receipt_amt'].count()
```

Party	contb_receipt_amt
Democrat	574591
Green	1001
Independent	103
Libertarian	782
Republican	72983

Question 4: Party's Frequency of donations by Date

```
1 nyc.groupby(['Party', 'Date'])['contb_receipt_amt'].count()
```

		contb_receipt_amt
Party	Date	
Democrat	2014-11-20	3
	2014-11-21	1
	2014-12-24	1
	2015-02-26	1
	2015-03-15	1
	2015-03-29	1
	2015-04-05	1
	2015-04-12	365
	2015-04-13	156
	2015-04-14	118
	2015-04-15	86
	2015-04-16	100
	2015-04-17	94
	2015-04-18	29

Question 5: Party's Sum of donations

```
nyc.groupby('Party')['contb_receipt_amt'].sum()
```

Party	contb_receipt_amt
Democrat	1.536526e+08
Green	2.636641e+05
Independent	3.018750e+04
Libertarian	2.468605e+05
Republican	1.745958e+07

```
pd.options.display.float_format = '{:,.2f}'.format
nyc.groupby('Party')['contb_receipt_amt'].sum()
```

Party	contb_receipt_amt
Democrat	153,652,597.45
Green	263,664.10
Independent	30,187.50
Libertarian	246,860.47
Republican	17,459,578.06

Question 6: Party's Sum of donations, by Date

```
1 nyc.groupby(['Party', 'Date'])['contb_receipt_amt'].sum()
```

		contb_receipt_amt
Party	Date	
Democrat	2014-11-20	7,800.00
	2014-11-21	500.00
	2014-12-24	250.00
	2015-02-26	2,600.00
	2015-03-15	500.00
	2015-03-29	1,000.00
	2015-04-05	50.00
	2015-04-12	245,296.27
	2015-04-13	150,179.16
	2015-04-14	212,985.00
	2015-04-15	128,199.00
	2015-04-16	144,311.85
	2015-04-17	127,645.16
	2015-04-18	39,035.00

Question 7: Top 5 Donors, by Occupation

```
df7 = nyc.groupby('contbr_occupation')['contb_receipt_amt'].
    sum().reset_index()
```

	contbr_occupation	contb_receipt_amt
0	ADMINISTRATIVE ASSISTANT	150.00
1	ATTORNEY	290.00
2	CHARITY CONSULTANT	250.00
3	EDUCATOR	67.50
4	HEALTHCARE MANAGER	34.00
5	LIBRARIAN	100.00
6	SMALL BUSINESS CONSULTANT	100.00
7	& PROFFESSOR	110.00
8	"RETIRED"	134.00
9	–	400.00

Question 7: Top 5 Donors, by Occupation (Continued)

```
df7.sort_values('contb_receipt_amt', ascending=False, inplace
    =True)
df7.head(5)

#or

df7.nlargest(5,'contb_receipt_amt')
```

	contbr_occupation	contb_receipt_amt
13284	RETIRED	10,094,554.11
1246	ATTORNEY	7,094,876.09
7893	INFORMATION REQUESTED	3,756,408.91
8616	LAWYER	3,202,850.66
10328	NOT EMPLOYED	2,856,233.71

Question 8: Bottom 5 Donors, by Occupation

```
df8 = nyc.groupby('contbr_occupation')['contb_receipt_amt'].
    sum().reset_index()
df8.sort_values(by='contb_receipt_amt',inplace=True)
df8.head(5)

# OR

df7.tail(5)

#OR
df8.nsmallest(5,'contb_receipt_amt')
```

	contbr_occupation	contb_receipt_amt
7524	HOMECARE DIRECTOR	-1,000.00
2957	COACHING	-469.36
16304	UTILITIES ENGINEER	-355.00
15717	TENANT RELATIONS	-340.00
9335	MARKETING DIRECTO	-320.00

Question 8: Bottom 5 Donors, by Occupation (Continued)

```
1 df8[df8.contb_receipt_amt>0].head(5)
```

contbr_occupation	contb_receipt_amt
FINANCIAL ENGINEER	0.80
HEDGE FUND - INVESTMENT PROFESSIONAL	1.00
FREELANCE TV & FILM PRODUCER	1.00
DRUG REP	1.00
CORPORATE FINANCIAL ADVISORY	1.00

Question 9: Top 5 Donors, by Employer

```
df9 = nyc.groupby('contbr_employer')['contb_receipt_amt'].sum().reset_index()
df9.sort_values('contb_receipt_amt', ascending=False, inplace=True)
df9.iloc[0:5]
```

contbr_employer	contb_receipt_amt
SELF-EMPLOYED	12,060,418.91
RETIRED	4,748,780.15
INFORMATION REQUESTED	3,611,890.31
NOT EMPLOYED	1,804,569.37
NONE	1,337,244.12

Question 10: Top 5 Occupations that donated to Each Candidate

```
1 df10 = nyc.groupby(['cand_nm', 'contbr_occupation'])['
      contb_receipt_amt'].sum().reset_index()
2 df10.sort_values(['cand_nm', 'contb_receipt_amt'], ascending
      =[True, False],inplace=True)
3 df10.groupby('cand_nm').head(5)
```

cand_nm	contbr_occupation	contb_receipt_amt
Bush, Jeb	RETIRED	399174
Bush, Jeb	HOMEMAKER	286522
Bush, Jeb	ATTORNEY	228157
Bush, Jeb	FINANCE	144400
Bush, Jeb	INFORMATION REQUESTED P…	132795
Carson, Benjamin S.	RETIRED	257267
Carson, Benjamin S.	INFORMATION REQUESTED P…	92623
Carson, Benjamin S.	PHYSICIAN	22513.2
Carson, Benjamin S.	HOMEMAKER	19304
Carson, Benjamin S.	ADMINISTRAT…	13300

Question 11: Top 5 Fundraising Candidates Line Graph

```
df11 = nyc.groupby('cand_nm')['contb_receipt_amt'].sum().
    reset_index()
df11_p = df11.nlargest(5,'contb_receipt_amt')
df11_g = nyc[nyc.cand_nm.isin(df11_p.cand_nm)][['cand_nm','
    Date','contb_receipt_amt']]

dfpiv=pd.pivot_table(df11_g , values='contb_receipt_amt',
    index=['Date'],columns=['cand_nm'], aggfunc=np.sum)
```

cand_nm	Bush, Jeb	Clinton, Hillary Rodham	Rubio, Marco	Sander
Date				
2016-01-29	3,250.00	80,599.35	13,380.00	
2016-01-30	-10.00	44,427.81	1,765.00	

Question 11: Top 5 Fundraising Candidates Line Graph (Continued)

```
1 dfpiv.loc['2016-01-01':'2016-01-30'].plot.line()
```