Lindsey Mcallister (*lindseym*) & Erica Zhou (*ezhou*)
18 September 2019
Repo: *https://github.com/ericazhou7/6.s080-labs*
Commit Hash: *ffd02165ad6ee503e58570551c682a707205f1cc*

6.S080 Problem Set 1

1

1961 candidates ran for president in 2016.

```
Pandas Output
      0
0  1961

SQLite Output
   COUNT(CAND_ID)
0            1961
```

Lindsey Mcallister (*lindseym*) & Erica Zhou (*ezhou*) 18 September 2019
Repo: *https://github.com/ericazhou7/6.s080-labs*
Commit Hash: *ffd02165ad6ee503e58570551c682a707205f1cc*

6.S080 Problem Set 1

2

Libertarians yielded the most senate candidates (35) in 2016.

```
Pandas Output
CAND_PTY_AFFILIATION COUNT
LIB                  35
OTH                  33
NPA                  13
UN                   12
UNK                  10
GRE                  9
CON                  4
AM                   2
NP                   1
NNE                  1
IAP                  1
I                    1
FSP                  1

Name: CAND_ID, dtype: int64

SQLite Output
   CAND_PTY_AFFILIATION  COUNT(CAND_ID)
0                   LIB              35
1                   OTH              33
2                   NPA              13
3                    UN              12
4                   UNK              10
5                   GRE               9
6                   CON               4
7                    AM               2
8                                     1
9                   FSP               1
10                    I               1
11                  IAP               1
12                   NA               1
13                  NNE               1
14                   NP               1
```

Lindsey Mcallister (*lindseym*) & Erica Zhou (*ezhou*) 18 September 2019
Repo: *https://github.com/ericazhou7/6.s080-labs*
Commit Hash: *ffd02165ad6ee503e58570551c682a707205f1cc*

6.S080 Problem Set 1

3

The following Super PACs had the most total receipts in 2016:

Pandas Output

	CMTE_NM	TTL_RECEIPTS
4374	PRIORITIES USA ACTION	1.920658e+08
6516	RIGHT TO RISE USA	1.216952e+08
6540	SENATE LEADERSHIP FUND	1.169120e+08
5535	NEXTGEN CLIMATE ACTION COMMITTEE	9.301024e+07
4053	SENATE MAJORITY PAC	9.282108e+07
5315	CONSERVATIVE SOLUTIONS PAC	6.056422e+07
4360	HOUSE MAJORITY PAC	5.587207e+07
4543	CONGRESSIONAL LEADERSHIP FUND	5.105302e+07
9082	GET OUR JOBS BACK, INC	5.032603e+07
9362	FOR OUR FUTURE	4.894757e+07

SQLite Output

	CMTE_NM	sum(TTL_RECEIPTS)
0	PRIORITIES USA ACTION	1.920658e+08
1	RIGHT TO RISE USA	1.216952e+08
2	SENATE LEADERSHIP FUND	1.169120e+08
3	NEXTGEN CLIMATE ACTION COMMITTEE	9.301024e+07
4	SENATE MAJORITY PAC	9.282108e+07
5	CONSERVATIVE SOLUTIONS PAC	6.056422e+07
6	HOUSE MAJORITY PAC	5.587207e+07
7	CONGRESSIONAL LEADERSHIP FUND	5.105302e+07
8	GET OUR JOBS BACK, INC	5.032603e+07
9	FOR OUR FUTURE	4.894757e+07

Lindsey Mcallister (*lindseym*) & Erica Zhou (*ezhou*) 18 September 2019
Repo: *https://github.com/ericazhou7/6.s080-labs*
Commit Hash: *ffd02165ad6ee503e58570551c682a707205f1cc*

6.S080 Problem Set 1

4

Mike Huckabee was the only presidential candidate with a name containing `HUCK`.

```
Pandas Output
        CAND_NAME                        CMTE_NM                      CMTE_ST1
0  HUCKABEE, MIKE  HUCKABEE FOR PRESIDENT, INC.  10800 FINANCIAL CENTER PKWY

SQLite Output
        CAND_NAME                        CMTE_NM                      CMTE_ST1
0  HUCKABEE, MIKE  HUCKABEE FOR PRESIDENT, INC.  10800 FINANCIAL CENTER PKWY
```

Lindsey Mcallister (*lindseym*) & Erica Zhou (*ezhou*) 18 September 2019
Repo: *https://github.com/ericazhou7/6.s080-labs*
Commit Hash: *ffd02165ad6ee503e58570551c682a707205f1cc*

6.S080 Problem Set 1

5

The following senate campaign committees raised the most funds per capita:

Pandas Output

	CMTE_NM	CMTE_ST	TTL_RECEIPTS	RECEIPTS_PER_CAPITA
131	MAGGIE FOR NH	NH	18698223.00	13.924853
129	FRIENDS OF KELLY AYOTTE INC	NH	13980279.89	10.411329
135	CATHERINE CORTEZ MASTO FOR SENATE	NV	19968577.25	6.660546
211	LEAHY FOR U.S. SENATOR COMMITTEE	VT	3930892.11	6.302971
1	LISA MURKOWSKI FOR US SENATE	AK	4431094.00	5.989624
191	FRIENDS OF JOHN THUNE	SD	5170324.00	5.945184
170	RUSS FOR WISCONSIN	WI	24857171.00	4.289059
136	FRIENDS OF JOE HECK	NV	12356469.00	4.121517
127	HOEVEN FOR SENATE	ND	2943354.00	3.896454
168	RON JOHNSON FOR SENATE INC	WI	17683385.71	3.051236
75	GRASSLEY COMMITTEE INC	IA	8954942.00	2.846715
35	BENNET FOR COLORADO	CO	15575583.00	2.777805
73	SCHATZ FOR SENATE	HI	3842453.03	2.691664
189	FRIENDS OF PAT TOOMEY	VA	22084253.00	2.607344
227	MIKE CRAPO FOR US SENATE	ID	4289036.00	2.498065
125	MISSOURIANS FOR KANDER	MO	13213897.00	2.161418
152	WYDEN FOR SENATE	OR	8685642.00	2.096575
48	BLUMENTHAL FOR CONNECTICUT	CT	7362368.00	2.051837
173	VAN HOLLEN FOR SENATE	MD	11960703.00	1.976265
124	FRIENDS OF ROY BLUNT	MO	12077039.00	1.975460

SQLite Output

	CMTE_NM	CMTE_ST	TTL_RECEIPTS	RECEIPTS_PER_CAPITA
0	MAGGIE FOR NH	NH	18698223.00	13.924853
1	FRIENDS OF KELLY AYOTTE INC	NH	13980279.89	10.411329
2	CATHERINE CORTEZ MASTO FOR SENATE	NV	19968577.25	6.660546
3	LEAHY FOR U.S. SENATOR COMMITTEE	VT	3930892.11	6.302971
4	LISA MURKOWSKI FOR US SENATE	AK	4431094.00	5.989624
5	FRIENDS OF JOHN THUNE	SD	5170324.00	5.945184
6	RUSS FOR WISCONSIN	WI	24857171.00	4.289059
7	FRIENDS OF JOE HECK	NV	12356469.00	4.121517
8	HOEVEN FOR SENATE	ND	2943354.00	3.896454
9	RON JOHNSON FOR SENATE INC	WI	17683385.71	3.051236
10	GRASSLEY COMMITTEE INC	IA	8954942.00	2.846715
11	BENNET FOR COLORADO	CO	15575583.00	2.777805
12	SCHATZ FOR SENATE	HI	3842453.03	2.691664
13	FRIENDS OF PAT TOOMEY	VA	22084253.00	2.607344
14	MIKE CRAPO FOR US SENATE	ID	4289036.00	2.498065
15	MISSOURIANS FOR KANDER	MO	13213897.00	2.161418
16	WYDEN FOR SENATE	OR	8685642.00	2.096575
17	BLUMENTHAL FOR CONNECTICUT	CT	7362368.00	2.051837

18	VAN HOLLEN FOR SENATE	MD	11960703.00	1.976265
19	FRIENDS OF ROY BLUNT	MO	12077039.00	1.975460

Lindsey Mcallister (*lindseym*) & Erica Zhou (*ezhou*) 18 September 2019
Repo: *https://github.com/ericazhou7/6.s080-labs*
Commit Hash: *ffd02165ad6ee503e58570551c682a707205f1cc*

6.S080 Problem Set 1

6

The following House candidates raised more than $100,000 and had the smallest ratios of individual contributions:

Pandas Output

	CAND_NAME	CAND_PTY_AFFILIATION	TTL_INDIV_CONTRIB	TTL_RECEIPTS	RATIO_INDIV
159	HANEY, PERRY	DEM	0.00	1001000.00	0.000000
481	FITZPATRICK, MICHAEL G.	REP	0.00	112473.57	0.000000
120	MODICA, JULIEN	DEM	25.00	550386.26	0.000045
933	TRONE, DAVID	DEM	6726.45	13421964.36	0.000501
1225	FLINN, GEORGE S DR JR	REP	1748.00	3027559.25	0.000577
781	SCOTT, BOBBY L.	REP	250.00	135250.00	0.001848
182	LLOP, WILLIAM	REP	1690.00	407940.00	0.004143
39	FLANEGAN, CHARLES EDWARD JR	REP	750.00	153750.00	0.004878
1131	BABINEC, MARTIN	IND	22273.57	3015593.95	0.007386
563	JONES, CHRISTINE	REP	50915.94	3774123.42	0.013491

SQLite Output

	CAND_NAME	CAND_PTY_AFFILIATION	TTL_INDIV_CONTRIB	TTL_RECEIPTS	RATIO_INDIV
0	HANEY, PERRY	DEM	0.00	1001000.00	0.000000
1	FITZPATRICK, MICHAEL G.	REP	0.00	112473.57	0.000000
2	MODICA, JULIEN	DEM	25.00	550386.26	0.000045
3	TRONE, DAVID	DEM	6726.45	13421964.36	0.000501
4	FLINN, GEORGE S DR JR	REP	1748.00	3027559.25	0.000577
5	SCOTT, BOBBY L.	REP	250.00	135250.00	0.001848
6	LLOP, WILLIAM	REP	1690.00	407940.00	0.004143
7	FLANEGAN, CHARLES EDWARD JR	REP	750.00	153750.00	0.004878
8	BABINEC, MARTIN	IND	22273.57	3015593.95	0.007386
9	JONES, CHRISTINE	REP	50915.94	3774123.42	0.013491

Lindsey Mcallister (*lindseym*) & Erica Zhou (*ezhou*) 18 September 2019
Repo: *https://github.com/ericazhou7/6.s080-labs*
Commit Hash: *ffd02165ad6ee503e58570551c682a707205f1cc*

6.S080 Problem Set 1

7

The following parties had the highest ratios of individual contributions in the 2016 senate race:

Pandas Output

CAND_PTY_AFFILIATION	TTL_INDIV_CONTRIB	TTL_RECEIPTS	RATIO_INDIV
UN	1.293100e+04	1.293520e+04	0.999675
GRE	1.197860e+05	1.289450e+05	0.928970
LIB	4.547360e+05	5.071100e+05	0.896721
DEM	2.433290e+08	3.120217e+08	0.779846
REP	1.718998e+08	2.654628e+08	0.647548
IND	6.800790e+05	1.199491e+06	0.566973
NPA	5.700000e+03	1.070900e+04	0.532263
NNE	1.696000e+03	5.196000e+03	0.326405
OTH	6.562680e+04	2.556507e+05	0.256705
UNK	2.690000e+02	1.763100e+04	0.015257

SQLite Output

	CAND_PTY_AFFILIATION	SUM(TTL_INDIV_CONTRIB)	SUM(TTL_RECEIPTS)	RATIO_INDIV
0	UN	1.293100e+04	1.293520e+04	0.999675
1	GRE	1.197860e+05	1.289450e+05	0.928970
2	LIB	4.547360e+05	5.071100e+05	0.896721
3	DEM	2.433290e+08	3.120217e+08	0.779846
4	REP	1.718998e+08	2.654628e+08	0.647548
5	IND	6.800790e+05	1.199491e+06	0.566973
6	NPA	5.700000e+03	1.070900e+04	0.532263
7	NNE	1.696000e+03	5.196000e+03	0.326405
8	OTH	6.562680e+04	2.556507e+05	0.256705
9	UNK	2.690000e+02	1.763100e+04	0.015257

Lindsey Mcallister (*lindseym*) & Erica Zhou (*ezhou*) 18 September 2019
Repo: *https://github.com/ericazhou7/6.s080-labs*
Commit Hash: *ffd02165ad6ee503e58570551c682a707205f1cc*

6.S080 Problem Set 1

8

California had the highest sum of total contributions (883903955) from individuals.

```
SQLite Output
    STATE  TOTAL_CONTRIBUTIONS
0      CA          883903955.0
1      NY          684083669.0
2      TX          412118374.0
3      FL          368252262.0
4      IL          298131396.0
..    ...                  ...
124    DF                30.0
125    FS                25.0
126    N/                25.0
127    FR                15.0
128    HO                 5.0
```

Lindsey Mcallister (*lindseym*) & Erica Zhou (*ezhou*) 18 September 2019
Repo: *https://github.com/ericazhou7/6.s080-labs*
Commit Hash: *ffd02165ad6ee503e58570551c682a707205f1cc*

6.S080 Problem Set 1

9

The following states had the highest individual contributions (per capita) to super PACs:

```
SQLite Output
  STATE  PER_CAPITA_CONTRIB
0    NV           28.504724
1    NY           10.840999
2    CT            9.678276
3    IL            9.357326
4    NE            9.017235
```

Lindsey Mcallister (*lindseym*) & Erica Zhou (*ezhou*) 18 September 2019
Repo: *https://github.com/ericazhou7/6.s080-labs*
Commit Hash: *ffd02165ad6ee503e58570551c682a707205f1cc*

6.S080 Problem Set 1

10

The following Senate candidates had the highest out-of-state individual contributions:

```
SQLite Output
                CAND_NAME CAND_OFFICE_ST  TOTAL_AMOUNT
0   FEINGOLD, RUSSELL DANA             WI    10540346.0
1  MCGINTY, KATHLEEN ALANA             PA     9517756.0
2            PORTMAN, ROB              OH     9338598.0
3    HASSAN, MARGARET WOOD             NH     9098834.0
4            RUBIO, MARCO              FL     9001949.0
```

Lindsey Mcallister (*lindseym*) & Erica Zhou (*ezhou*) 18 September 2019
Repo: *https://github.com/ericazhou7/6.s080-labs*
Commit Hash: *ffd02165ad6ee503e58570551c682a707205f1cc*

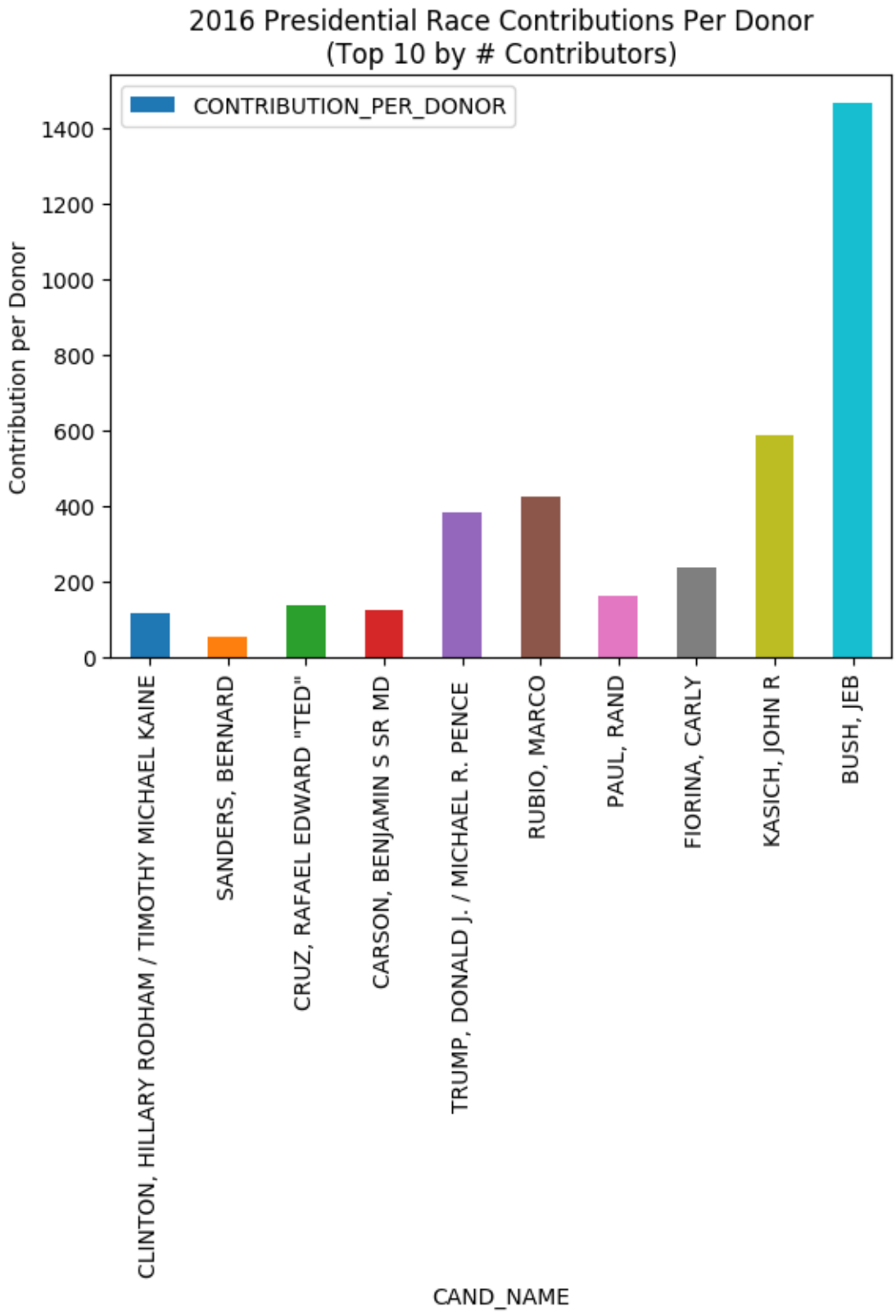
6.S080 Problem Set 1

11

Question: Of the 2016 Presidential candidates with the most donors, who received the highest contribution per donor? List top 10 with *cand_name* and *contribution per donor*. Use `indiv_contrib` and `candidate` tables.

SQLite Output

	CAND_NAME	TOTAL_IND_CONTRIBUTIONS	TOTAL_IND_CONTRIBUTORS	CONTRIBUTION_PER_DONOR
0	CLINTON, HILLARY RODHAM / TIMOTHY MICHAEL KAINE	292132031.0	2516261	116.097667
1	SANDERS, BERNARD	84358086.0	1532148	55.058706
2	CRUZ, RAFAEL EDWARD "TED"	54006390.0	398631	135.479654
3	CARSON, BENJAMIN S SR MD	27688948.0	219963	125.880025
4	TRUMP, DONALD J. / MICHAEL R. PENCE	57042047.0	148218	384.852359
5	RUBIO, MARCO	39977109.0	94063	425.003551
6	PAUL, RAND	5045895.0	30737	164.163549
7	FIORINA, CARLY	6577176.0	27744	237.066609
8	KASICH, JOHN R	15137450.0	25672	589.648255
9	BUSH, JEB	32424600.0	22099	1467.242862


2016 Presidential Race Contributions Per Donor
(Top 10 by # Contributors)
CONTRIBUTION_PER_DONOR
Contribution per Donor
0
200
400
600
800
1000
1200
1400
CLINTON, HILLARY RODHAM / TIMOTHY MICHAEL KAINE
SANDERS, BERNARD
CRUZ, RAFAEL EDWARD "TED"
CARSON, BENJAMIN S SR MD
TRUMP, DONALD J. / MICHAEL R. PENCE
RUBIO, MARCO
PAUL, RAND
FIORINA, CARLY
KASICH, JOHN R
BUSH, JEB
CAND_NAME

Lindsey Mcallister (*lindseym*) & Erica Zhou (*ezhou*) 18 September 2019
Repo: *https://github.com/ericazhou7/6.s080-labs*
Commit Hash: *ffd02165ad6ee503e58570551c682a707205f1cc*

6.S080 Problem Set 1

12

Reflection:

- Sequential thinking: Sequential thinking was easier in pandas - you could have some filter and make a new dataframe with it, and then go forward and perform further operations on that object. In SQL, it could be more challenging to think about writing the query in a top-to-bottom manner because the ordering didn't feel as natural.
- Aside from ordering, SQL felt more intuitive to me, especially for simple steps. For example, if you only want valid 2016 senate candidates, you can very quickly and easily express that with conditions under "WHERE," while in Pandas, even simple filters could get cluttered very quickly.