Project 2 - Classifiers

Laila El Gohary

acquire.

union data

26k office building janitorial workers

clean.

standardizing
merging
reordering
renaming
adding
clarifying.

explore.

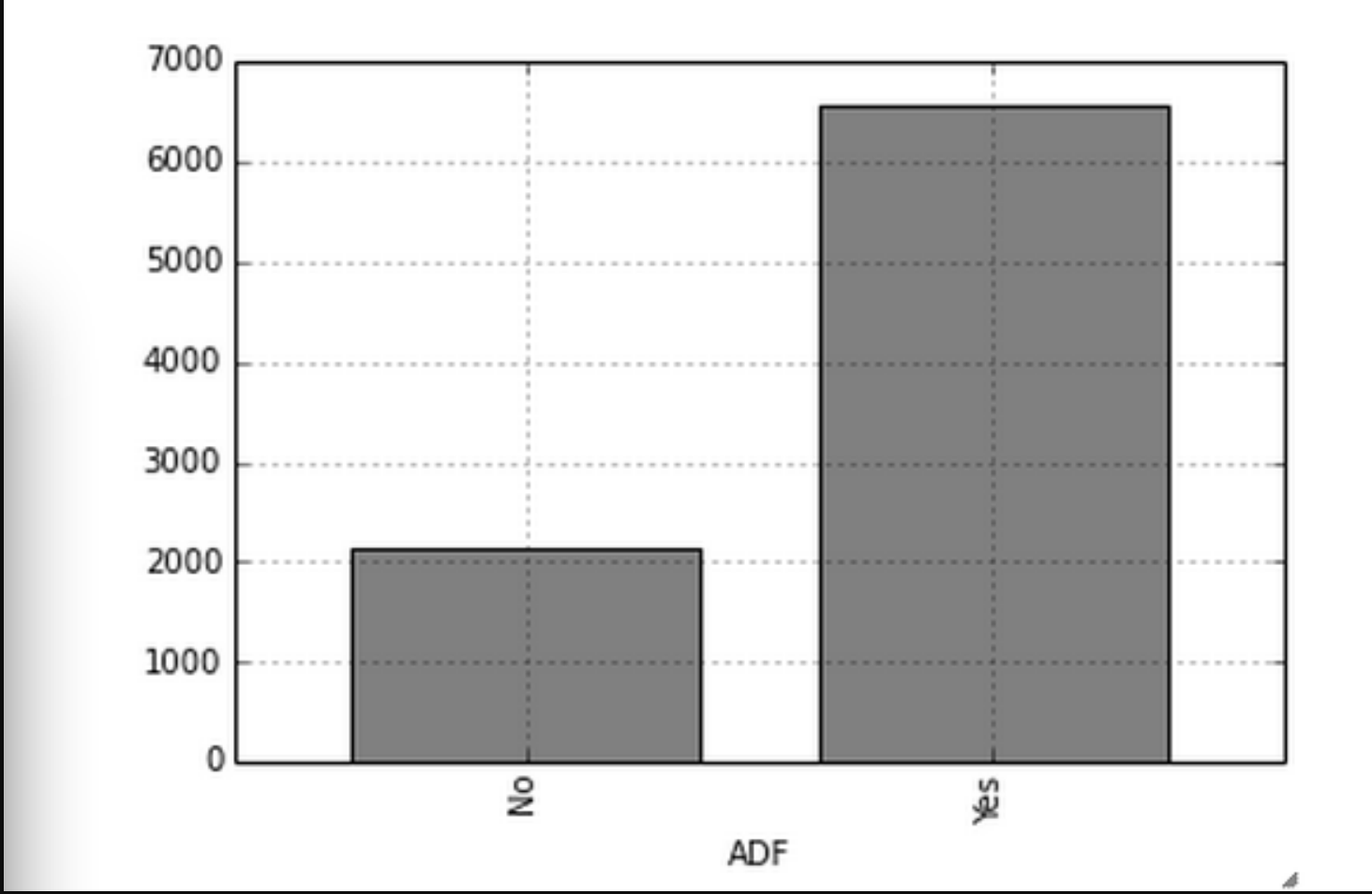
	mnbr	currentyrpoints	currentyrldrpoints	lifetimepoints	uactid
count	27630.000000	27630.000000	27630.000000	27630.000000	27630.000000
mean	238656.318060	0.314007	0.109374	6.631307	26917.992725
std	157687.993195	1.974426	1.351147	28.446562	13857.909232
min	199.000000	0.000000	0.000000	0.000000	12.000000
25%	130888.000000	0.000000	0.000000	0.000000	13941.000000
50%	190165.500000	0.000000	0.000000	0.000000	30500.000000
75%	402923.250000	0.000000	0.000000	5.000000	40452.000000
max	522765.000000	135.000000	129.000000	763.000000	43479.000000

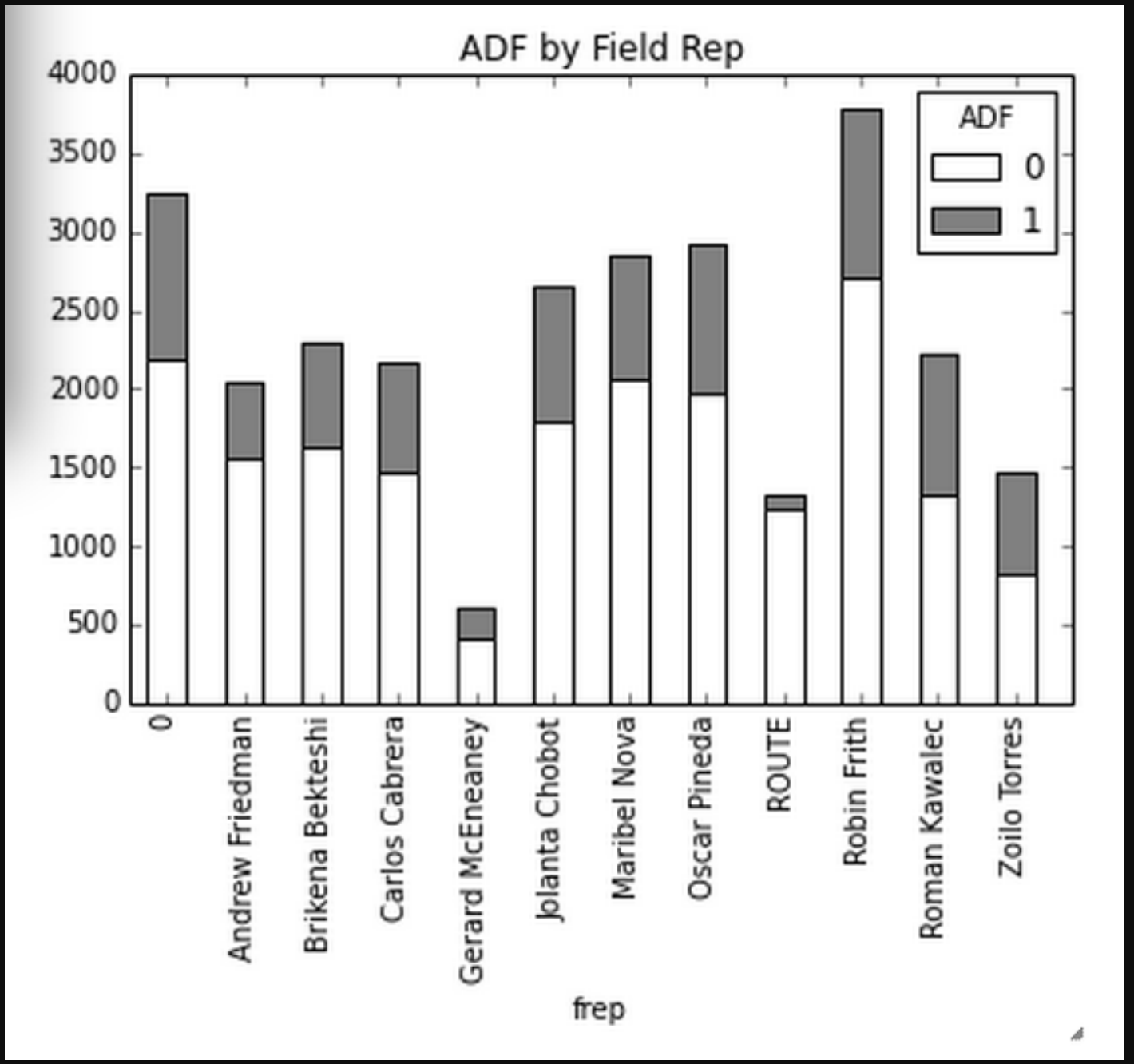
.describe( )

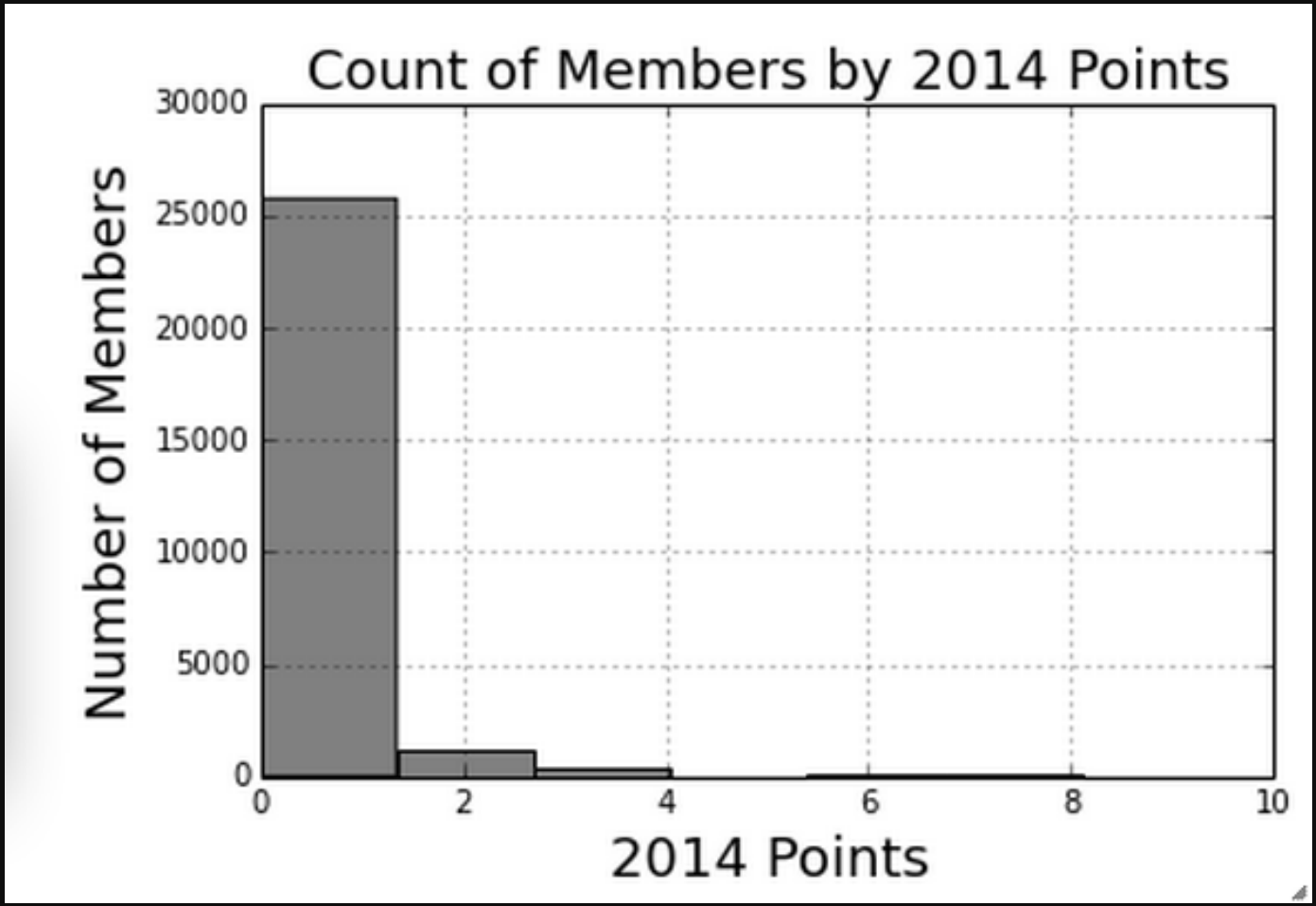
	mnbr	currentyrpoints	currentyrldrpoints	lifetimepoints	uactid
ADF					
No	262508.095832	0.110643	0.017306	2.792173	26681.337647
Yes	183940.544588	0.780520	0.320577	15.438245	27460.877444

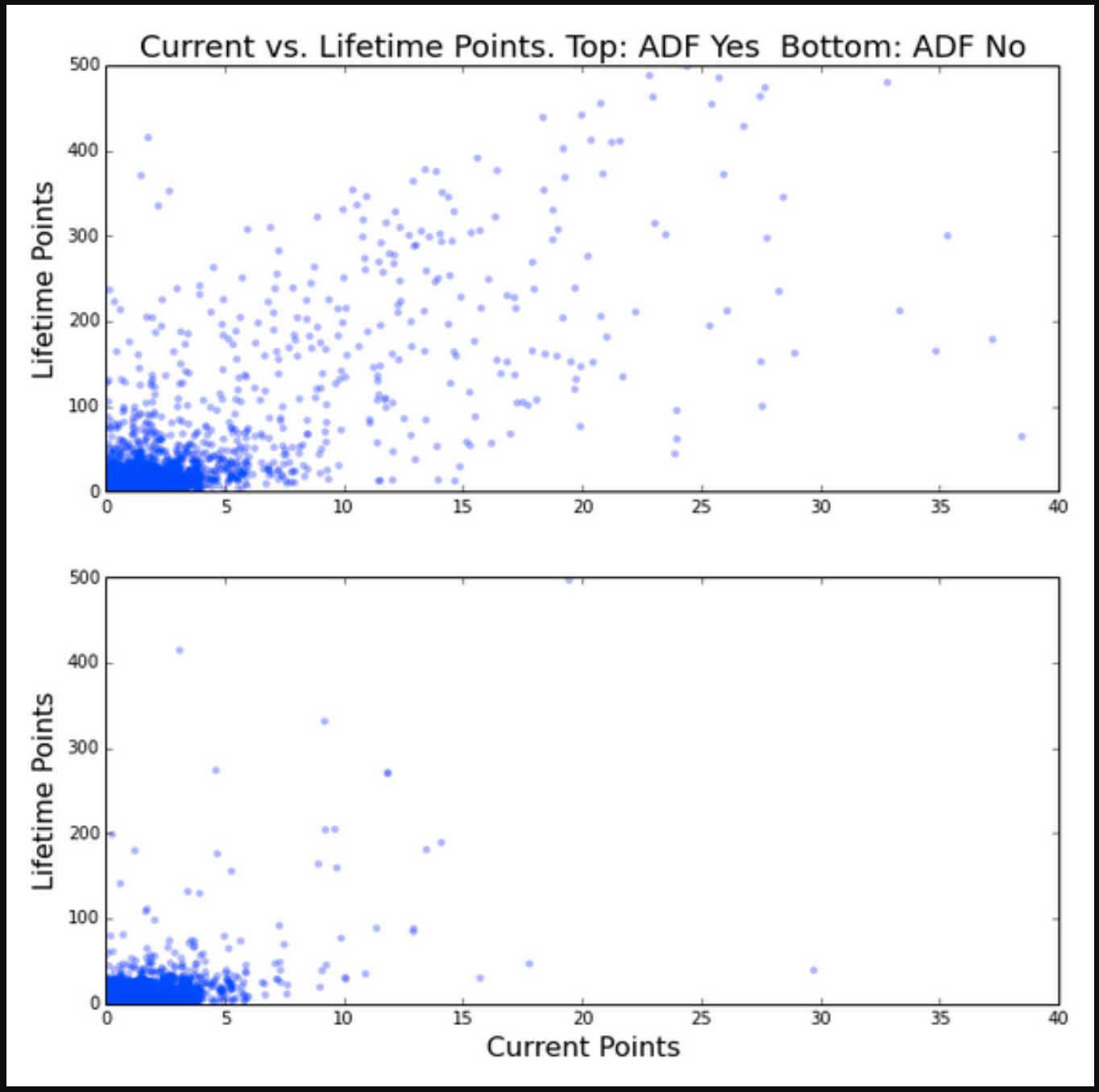
There seems to be a correlation between political contributions (ADF) and points (how many events someone goes to)

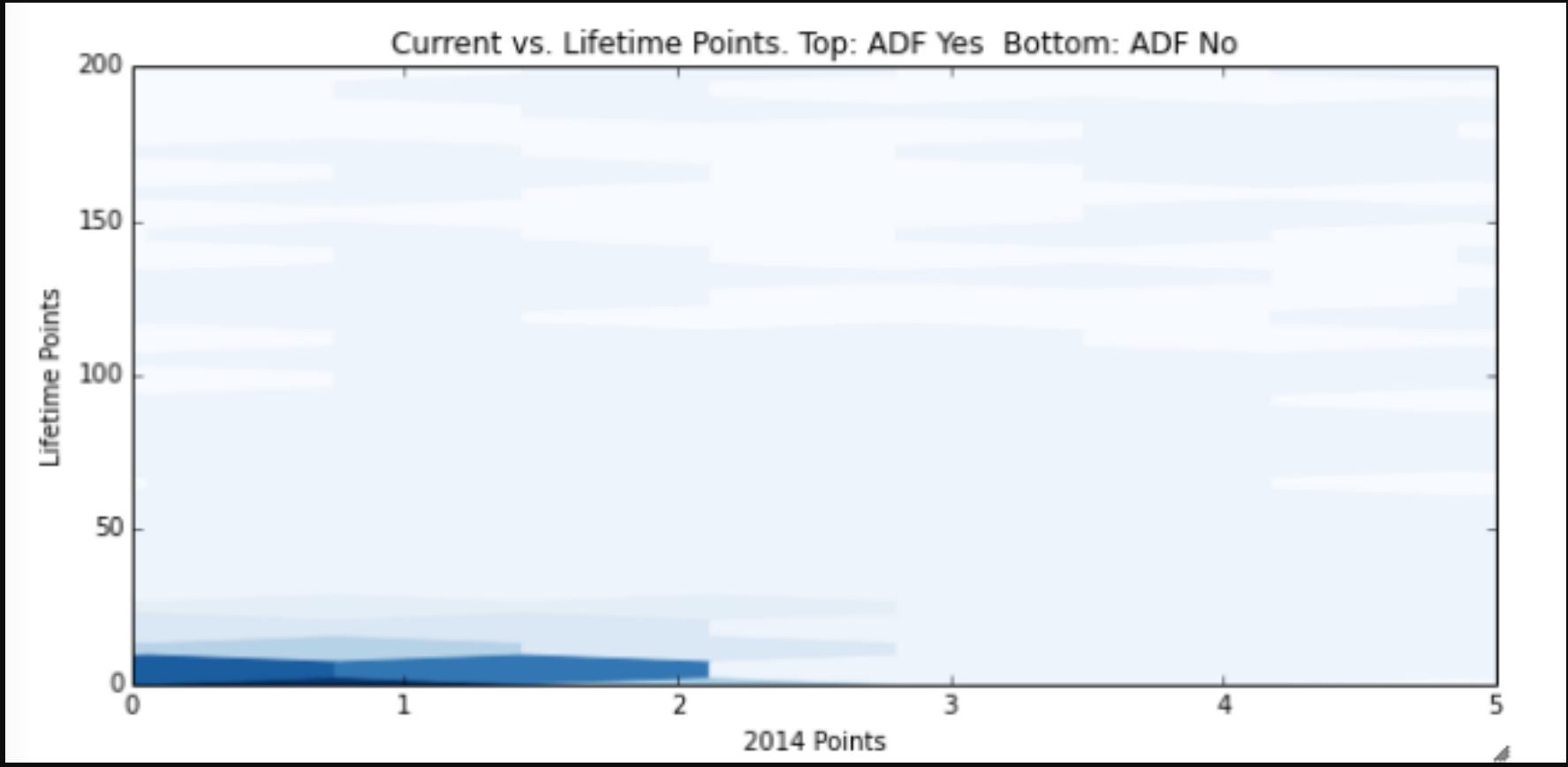
ADF to Points Comparison




ADF by Field Rep
4000
3500
3000
2500
2000
1500
1000
500
0
ADF
0
1
0
Andrew Friedman
Brikena Bekteshi
Carlos Cabrera
Gerard McEneaney
Jolanta Chobot
Maribel Nova
Oscar Pineda
ROUTE
Robin Frith
Roman Kawalec
Zoilo Torres
frep

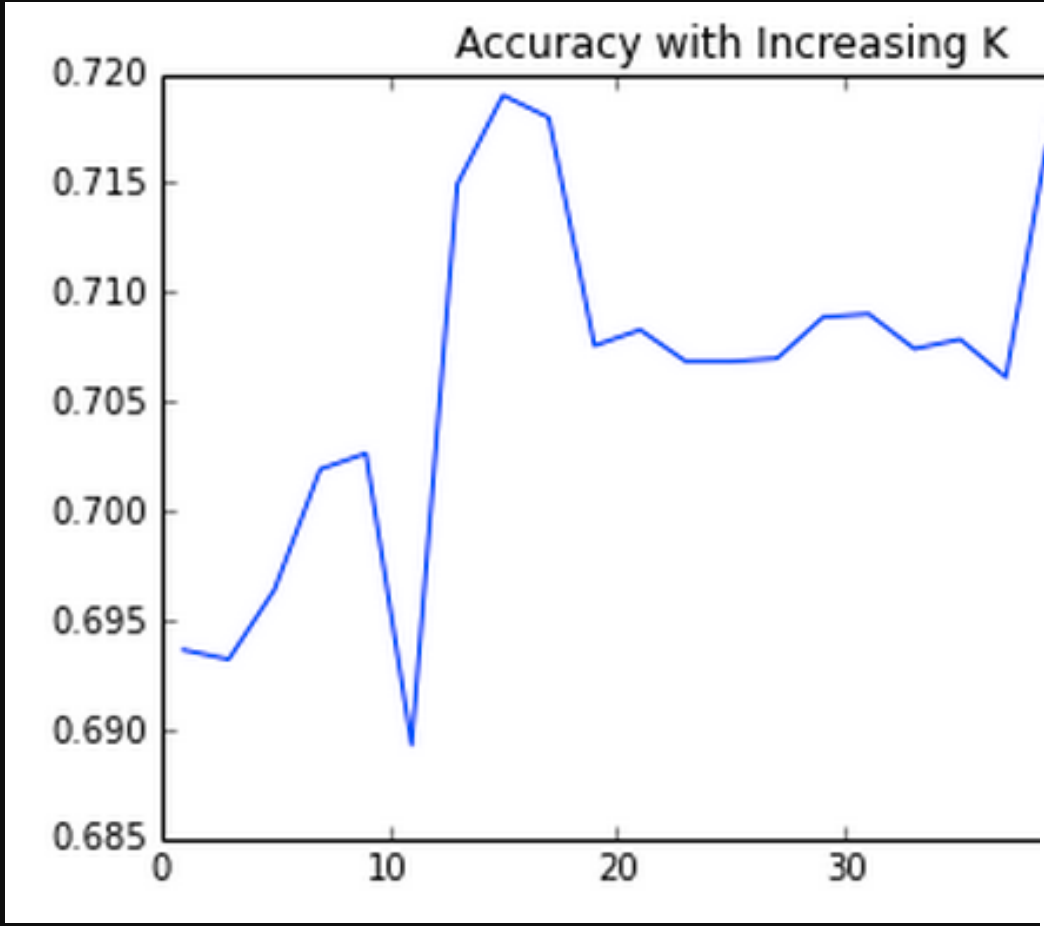

Count of Members by 2014 Points
Number of Members
30000
25000
20000
15000
10000
5000
0
0
2
4
6
8
10
2014 Points


Current vs. Lifetime Points. Top: ADF Yes Bottom: ADF No
Lifetime Points
Lifetime Points
Current Points
0
100
200
300
400
500
0
5
10
15
20
25
30
35
40


Current vs. Lifetime Points. Top: ADF Yes Bottom: ADF No
Lifetime Points
200
150
100
50
0
0
1
2
3
4
5
2014 Points

As my classifier, I ended up using KNN to determine whether someone would be on ADF (political contributions).

My inputs were 2014 Points, Leadership Points, and Lifetime Points.




```
Neighbors: 1, Accuracy: 0.693635
Neighbors: 3, Accuracy: 0.693200
Neighbors: 5, Accuracy: 0.696390
Neighbors: 7, Accuracy: 0.701899
Neighbors: 9, Accuracy: 0.702624
Neighbors: 11, Accuracy: 0.689285
Neighbors: 13, Accuracy: 0.714949
Neighbors: 15, Accuracy: 0.719008
Neighbors: 17, Accuracy: 0.717993
Neighbors: 19, Accuracy: 0.707554
Neighbors: 21, Accuracy: 0.708279
Neighbors: 23, Accuracy: 0.706829
Neighbors: 25, Accuracy: 0.706829
Neighbors: 27, Accuracy: 0.706974
Neighbors: 29, Accuracy: 0.708859
Neighbors: 31, Accuracy: 0.709004
Neighbors: 33, Accuracy: 0.707409
Neighbors: 35, Accuracy: 0.707844
Neighbors: 37, Accuracy: 0.706104
Neighbors: 39, Accuracy: 0.717993
Neighbors: 41, Accuracy: 0.717123
Neighbors: 43, Accuracy: 0.717993
Neighbors: 45, Accuracy: 0.717993
Neighbors: 47, Accuracy: 0.718428
Neighbors: 49, Accuracy: 0.718718
```

Next Steps:

- Learn how to use KNN on non numerical data (i.e. different field rep names)

- Work with different inputs (Home City/Zip, Field Rep)

- Try mapping out KNN using visualizations.