# General Assembly
## Data Science

acquire.

# nationbuilder

electorate data for
Baltimore's 46$^{th}$ district

67,124
observations

138
features

which included :

occupation
email
party
sex
fundraiser

```
twitter_id',
 'twitter_login',
 'facebook_username',
 'first_name',
 'middle_name',
 'last_name',
 'suffix',
 'full_name',
 'legal_name',
 'employer',
 'occupation',
 'sex',
 'party',
 'website',
 'is_deceased',
 'born_at',
 'email1',
 'email2',
 'email4',
 'email_opt_in',
 'phone_number',
 'work_phone_number',
 'do_not_call',
 'do_not_contact',
 'federal_donotcall',
 'mobile_number',
 'mobile_opt_in',
 'is_mobile_bad',
 'demo',
 'support_level',
 'priority_level',
 'availability',
 'note',
 'parent_id',
 'point_person_name_or_email',
 'recruiter_id',
 'recruiter_name_or_email',
 'tag_list',
 'created_at',
 'unsubscribed_at',
 'is_supporter',
 'is_prospect',
 'is_volunteer',
 'is_donor',
 'is_fundraiser',
 'is_ignore_donation_limits',
 'first_donated_at',
 'last_donated_at',
 'donations_count',
 'donations_amount',
 'donations_amount_in_cents',
 'donations_raised_count',
 'donations_raised_amount',
 'donations_raised_amount_in_cents'
 'donations_pledged_amount',
 'donations_pledged_amount_in_cents',
 'donations_count_this_cycle',
 'donations_amount_this_cycle',
 'donations_amount_this_cycle_in_cents'
 'donations_raised_count_this_cycle',
 'donations_raised_amount_this_cycle',
 'donations_raised_amount_this_cycle_in
_cents',
 'maximum_donation_possible_this_cycl
e',
 'maximum_donation_possible_this_cycl
e_in_cents',
 'nbec_guid',
 'pf_strat_id',
 'state_file_id',
 'van_id',
 'federal_district',
 'state_upper_district',
 'state_lower_district',
 'county_district',
 'village_district',
 'precinct_name',
 'precinct_code',
 'capital_amount_in_cents',
 'spent_capital_amount_in_cents',
 'received_capital_amount_in_cents
 'primary_country_code',
 'primary_country',
 'primary_state',
 'primary_city',
 'primary_county',
 'primary_zip',
 'primary_address1',
 'primary_address2',
 'primary_address3',
 'primary_fips',
 'primary_submitted_address',
 'primary_geocode_accuracy',
 'address_country_code',
 'address_country',
 'address_state',
 'address_city',
 'address_zip',
 'address_address1',
 'address_address2',
 'address_address3',
 'address_geocode_accuracy',
 'mailing_country_code',
 'mailing_country',
 'mailing_state',
 'mailing_city',
 'mailing_county',
 'mailing_zip',
 'mailing_address1',
 'mailing_address2',
 'mailing_geocode_accuracy',
 'mailing_street_number',
 'mailing_street_prefix',
 'mailing_street_name',
 'mailing_street_type',
 'mailing_street_suffix',
 'mailing_unit_number',
 'registered_country_code',
 'registered_country',
 'registered_state',
 'registered_city',
 'registered_county',
 'registered_zip',
 'registered_address1',
 'registered_address2',
 'registered_fips',
 'registered_submitted_address',
 'registered_geocode_accuracy',
 'registered_street_number',
 'registered_street_prefix',
 'registered_street_name',
 'registered_street_type',
 'registered_street_suffix',
 'registered_unit_number',
 'registered_zip5',
 'registered_zip4',
 'registered_sort_sequence',
 'registered_delivery_point',
 'registered_lot',
 'registered_carrier_route',
 'party_member']
```

# poll

2,070 with id's
65,054 without

# id's ranged from 1 to 5

1  strong support

2  leaning towards candidate

3  not sure/unwilling to say

4  leaning against candidate

5  strong dislike

clean.
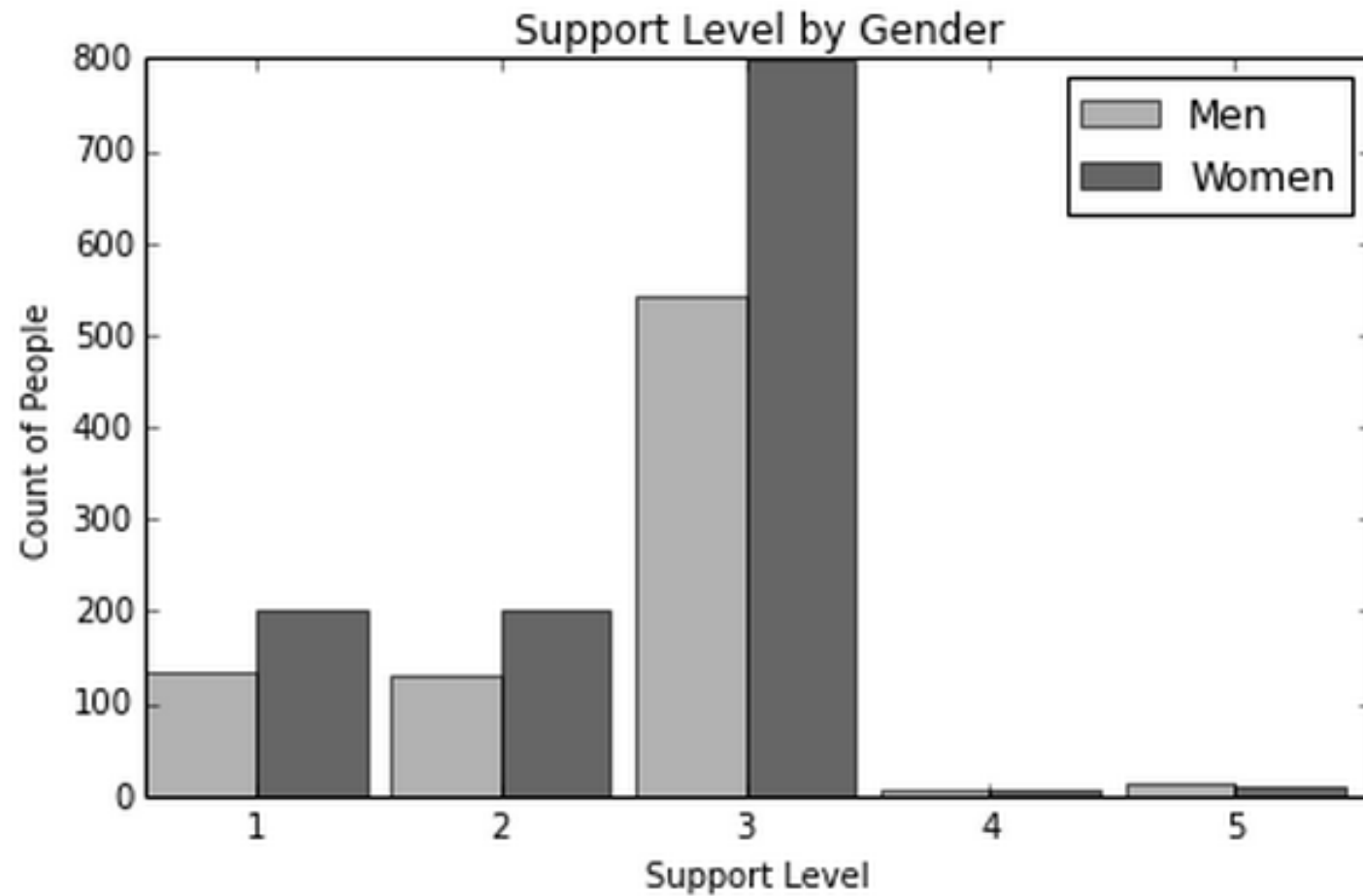
| Features | SupportCounts | NoSupportCounts |
|---|---|---|
| registered_zip4 | 0 | 0 |
| registered_carrier_route | 0 | 0 |
| registered_lot | 0 | 0 |
| registered_delivery_point | 0 | 0 |
| registered_sort_sequence | 0 | 0 |
| legal_name | 0 | 0 |
| priority_level | 0 | 1 |
| pf_strat_id | 0 | 0 |
| website | 0 | 0 |
| recruiter_id | 0 | 0 |
| mailing_county | 0 | 0 |
| recruiter_name_or_email | 0 | 0 |
| registered_street_suffix | 0 | 213 |
| registered_submitted_address | 0 | 0 |
| availability | 0 | 0 |
| email4 | 0 | 0 |
| email2 | 0 | 0 |
| registered_zip5 | 0 | 0 |
| mailing_street_suffix | 1 | 520 |
| primary_submitted_address | 1 | 1 |
| facebook_username | 1 | 1 |
| twitter_login | 1 | 4 |
| twitter_id | 1 | 4 |
| work_phone_number | 2 | 7 |
| unsubscribed_at | 3 | 5 |
| primary_address3 | 4 | 25 |
| first_donated_at | 4 | 10 |
| address_address2 | 4 | 28 |
| last_donated_at | 4 | 10 |
| occupation | 4 | 10 |
| employer | 4 | 10 |
| address_address3 | 4 | 28 |
| parent_id | 10 | 10 |
| point_person_name_or_email | 10 | 10 |

# creating flags

```python
def mobile(x):
    if len(str(x)) < 5:
        return 0
    else:
        return 1


df['mobileflag'] = df['mobile_number'].apply(mobile)
df['phoneflag'] = df['phone_number'].apply(mobile)
df['emailflag'] = df['email1'].apply(mobile)
df['addressflag'] = df['address_zip'].apply(mobile)
```

explore.

Support Level by Gender

```
Women By Party
D       1212
G          1
I          2
dtype: int64

_____

Men By Party
D        818
G          1
I          7
R          2
dtype: int64
```
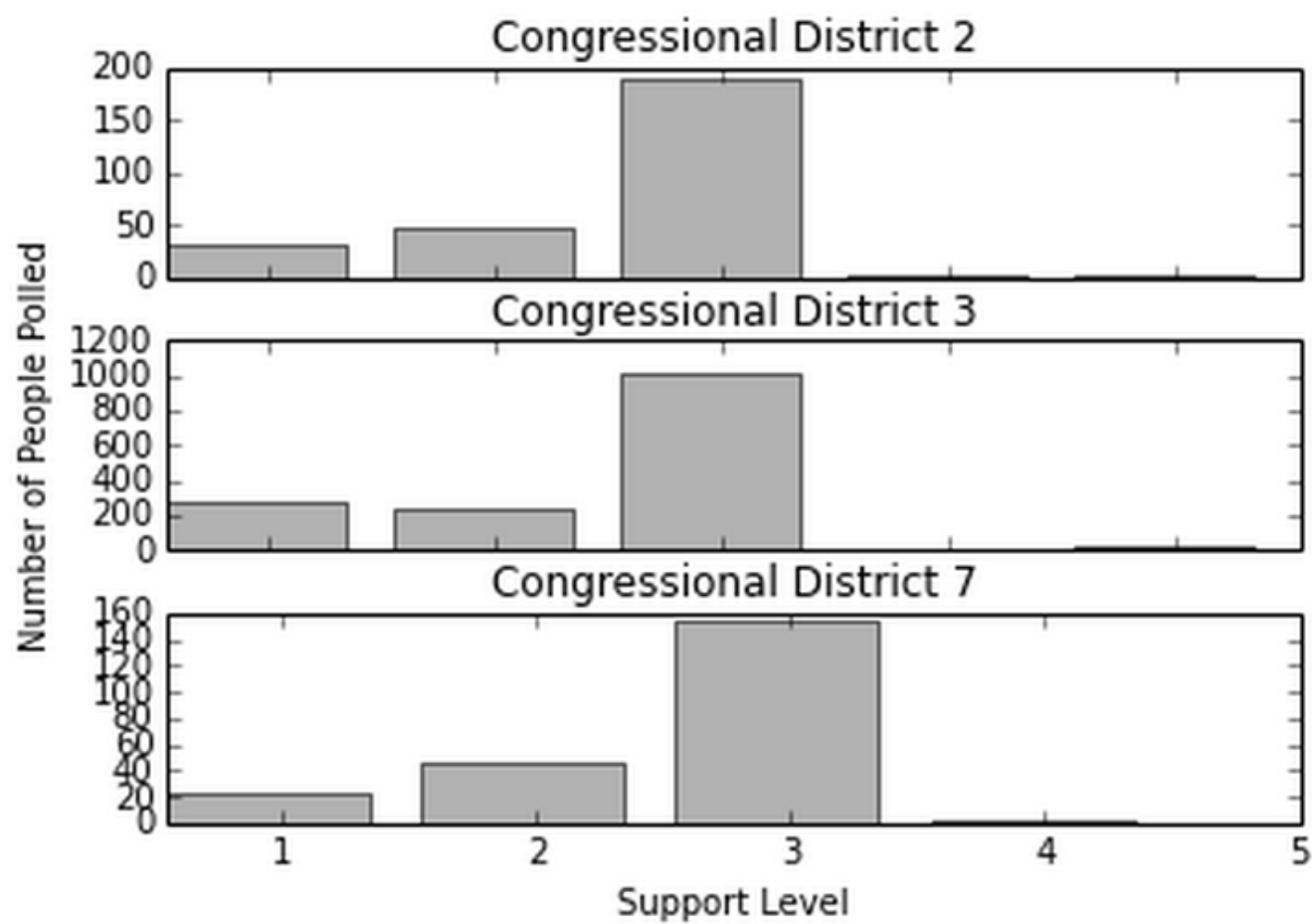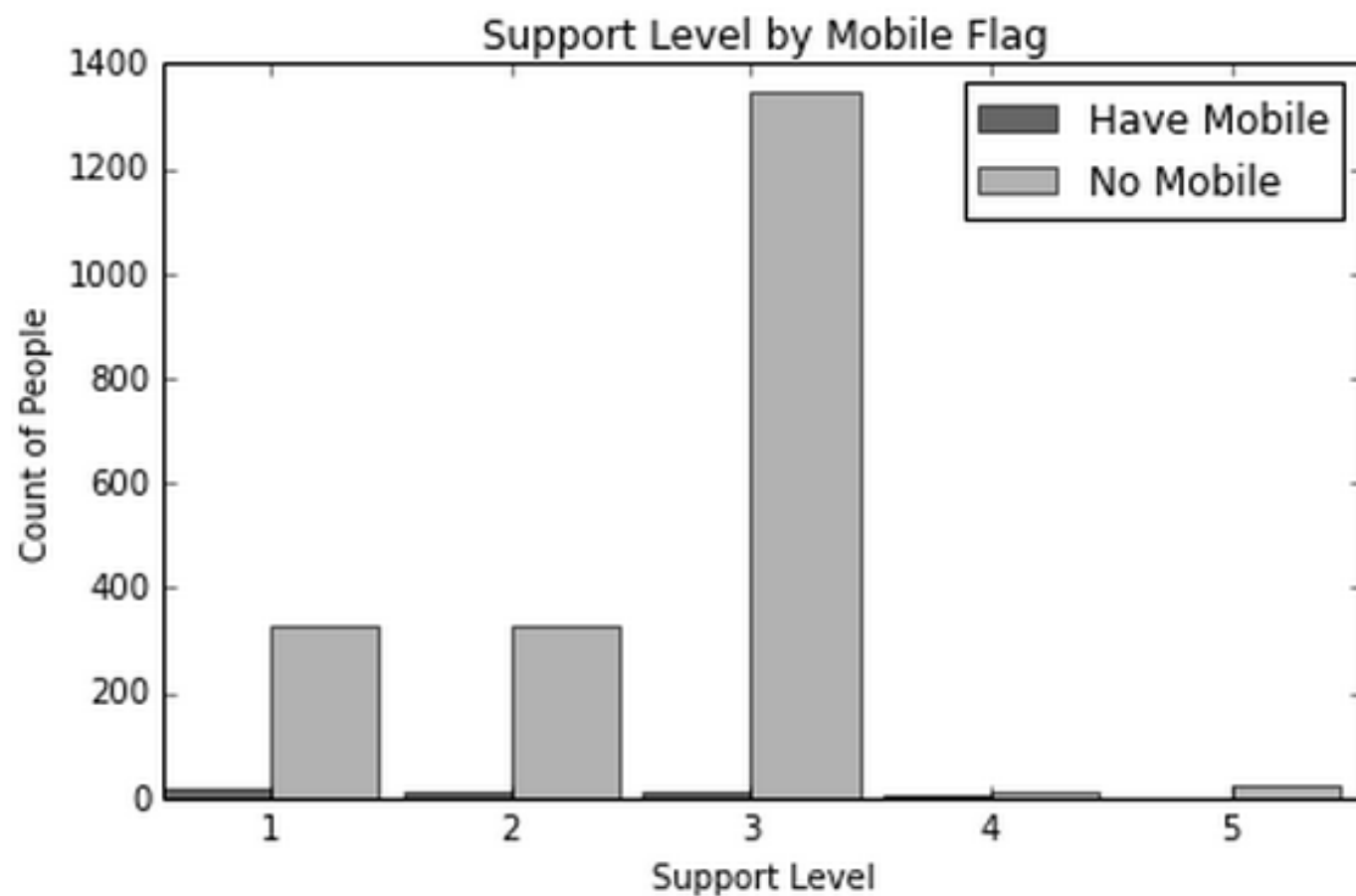
```
In [340]: pd.to_datetime(pd.Series(dfpoll.born_at))

Out[340]: nationbuilder_id
          4206                2039-10-22
          4207                2044-03-04
          4219                2033-11-01
          4225                2053-02-21
          4230                1975-09-04
          4255                1986-09-26
          4260                1984-10-30
          4287                1977-12-07
          4301                2054-12-15
          4311                1971-12-07
          4312                1967-12-10
          4353                1982-11-08
          4361                2033-11-06
          4372                1977-08-31
          4375                1991-06-10
```

Support Level by Mobile Flag

model.

# final features

'donations_raised_amount_this_cycle_in_cents
'donations_raised_count_this_cycle',
'donations_amount_this_cycle_in_cents',
'donations_count_this_cycle',
'donations_pledged_amount_in_cents',
'donations_raised_amount_in_cents',
'donations_raised_count',
'donations_amount_in_cents',
'donations_count',
'federal_district',
'received_capital_amount_in_cents',
'is_fundraiser',
'is_donor',
'is_volunteer',
'is_prospect',

'is_supporter',
'is_mobile_bad',
'federal_donotcall',
'do_not_contact',
'do_not_call',
'mobileflag',
'phoneflag',
'emailflag',
'addressflag',
'districtflag',
'villageflag',
'partyflag',
'sexflag',
'support_level'

# accuracy

| | |
|---|---|
| multinomial nb<br>0.7065 | linear regression<br>0.8089 |
| gaussian nb<br>0.7471 | knn (5)<br>0.8320 |

# next steps

- feature loops
- model visualization
- condense/mesh/simplify
- why?