

PrimaryDataAnalysis

October 3, 2019

1 [REDACTED] Democratic Primary Analysis

Summary:

- 75 Responses
- Collected from 9/29/2019 to 10/3/2019 (5 days)

```
[1]: #Datacol names:

cols = ["Timestamp", "[REDACTED] status", "Gender", "Ethnicity", "Age Range",
        →"Household Income",
        "Region", "Voting Registration", "Party Registration", "Primary/Caucus",
        →"Ideology",
        "First Candidate", "Second Candidate", "GE Odds", "Biden Fav", "Warren Fav",
        →"Sanders Fav",
        "Buttigieg Fav", "Harris Fav", "Yang Fav", "First Republican", "Trump
        →Approval",
        "Impeachment Approval", "Biden V Trump", "Warren V Trump", "Sanders V Trump",
        →"Buttigieg V Trump",
        "Harris V Trump", "Yang V Trump", "2016 Primary", "2016 GE", "Electoral
        →College Approval",
        "Top Issue", "Gun Reform Approval", "Climate Change Approval"]
```

```
[2]: #Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from math import sqrt
```

```
[3]: poll_data = pd.read_csv("data/[REDACTED]DemPoll.csv")
poll_data.columns = cols
poll_data.shape
```

```
[3]: (75, 35)
```

```
[4]: poll_data.head()
```

[4]:

	Timestamp	status	Gender	\
0	2019/09/28 10:58:19 AM AST	Current Student	Male	
1	2019/09/28 10:58:47 AM AST	Current Student	Male	
2	2019/09/28 11:00:51 AM AST	Current Student	Male	
3	2019/09/28 11:14:00 AM AST	Current Student	Male	
4	2019/09/28 11:17:58 AM AST	Current Student	Male	

	Ethnicity	Age Range	Household Income	Region	\
0	Hispanic or Latino	26-29	Under \$50,000 a year	NorthEast	
1	White or Caucasian	18-21	\$50,000 to \$100,000 a year	NorthEast	
2	Other or Mixed Race	18-21	\$50,000 to \$100,000 a year	MidWest	
3	Black or African American	18-21	\$50,000 to \$100,000 a year	NorthEast	
4	White or Caucasian	18-21	Under \$50,000 a year	NorthEast	

	Voting Registration	Party Registration	Primary/Caucus	...	\
0	Yes	Democratic Party	Democratic Primary or Caucus	...	
1	Yes	Democratic Party	Democratic Primary or Caucus	...	
2	Yes	Democratic Party	Democratic Primary or Caucus	...	
3	No	Democratic Party	Unsure	...	
4	Yes	Democratic Party	Democratic Primary or Caucus	...	

	Sanders V Trump	Buttigieg V Trump	Harris V Trump	Yang V Trump	\
0	Bernie Sanders	Pete Buttigieg	Kamala Harris	Andrew Yang	
1	Bernie Sanders	Pete Buttigieg	Kamala Harris	Andrew Yang	
2	Bernie Sanders	3rd Party/Write in	Kamala Harris	3rd Party/Write in	
3	Bernie Sanders	Pete Buttigieg	3rd Party/Write in	Andrew Yang	
4	Bernie Sanders	Pete Buttigieg	3rd Party/Write in	Andrew Yang	

	2016 Primary	2016 GE	\
0	Did not vote/ineligible	Hillary Clinton	
1	Did not vote/ineligible	Did not vote/ineligible	
2	Hillary Clinton	Hillary Clinton	
3	Did not vote/ineligible	Did not vote/ineligible	
4	Did not vote/ineligible	Hillary Clinton	

	Electoral College Approval	Top Issue	Gun Reform Approval	\
0	No	Health Care	More Likely	
1	No	Climate Change	More Likely	
2	No	Climate Change	More Likely	
3	No	Climate Change	More Likely	
4	No	Climate Change	More Likely	

	Climate Change Approval
0	Very Important
1	Very Important
2	Very Important
3	Very Important

4 Very Important

[5 rows x 35 columns]

```
[5]: last_name = lambda x: x.split()[1]
```

```
[6]: poll_data["First Candidate"] = poll_data["First Candidate"].apply(last_name)
poll_data["First Candidate"]
```

```
[6]: 0        Sanders
1        Warren
2        Warren
3        Biden
4        Sanders
...
70       Harris
71       Sanders
72    Buttigieg
73       Yang
74       Warren
```

Name: First Candidate, Length: 75, dtype: object

```
[7]: poll_data["Second Candidate"] = poll_data["Second Candidate"].apply(last_name)
poll_data["GE Odds"] = poll_data["GE Odds"].apply(last_name)
poll_data["First Republican"] = poll_data["First Republican"].apply(last_name)
poll_data["Biden V Trump"] = poll_data["Biden V Trump"].apply(last_name)
poll_data["Warren V Trump"] = poll_data["Warren V Trump"].apply(last_name)
poll_data["Sanders V Trump"] = poll_data["Sanders V Trump"].apply(last_name)
poll_data["Buttigieg V Trump"] = poll_data["Buttigieg V Trump"].apply(last_name)
poll_data["Harris V Trump"] = poll_data["Harris V Trump"].apply(last_name)
poll_data["Yang V Trump"] = poll_data["Yang V Trump"].apply(last_name)
poll_data.head()
```

```
[7]:                    Timestamp                    status Gender \
0  2019/09/28 10:58:19 AM AST  Current Student    Male
1  2019/09/28 10:58:47 AM AST  Current Student    Male
2  2019/09/28 11:00:51 AM AST  Current Student    Male
3  2019/09/28 11:14:00 AM AST  Current Student    Male
4  2019/09/28 11:17:58 AM AST  Current Student    Male
```

```
                  Ethnicity Age Range                    Household Income                    Region \
0        Hispanic or Latino    26-29                    Under $50,000 a year    NorthEast
1        White or Caucasian    18-21                    $50,000 to $100,000 a year    NorthEast
2        Other or Mixed Race    18-21                    $50,000 to $100,000 a year    Midwest
3  Black or African American    18-21                    $50,000 to $100,000 a year    NorthEast
4        White or Caucasian    18-21                    Under $50,000 a year    NorthEast
```

```
                  Voting Registration                    Party Registration                    Primary/Caucus    ... \
0                    Yes    Democratic Party    Democratic Primary or Caucus    ...
```

1	Yes	Democratic Party	Democratic Primary or Caucus	...
2	Yes	Democratic Party	Democratic Primary or Caucus	...
3	No	Democratic Party	Unsure	...
4	Yes	Democratic Party	Democratic Primary or Caucus	...

	Sanders V Trump	Buttigieg V Trump	Harris V Trump	Yang V Trump	\
0	Sanders	Buttigieg	Harris	Yang	
1	Sanders	Buttigieg	Harris	Yang	
2	Sanders	Party/Write	Harris	Party/Write	
3	Sanders	Buttigieg	Party/Write	Yang	
4	Sanders	Buttigieg	Party/Write	Yang	

	2016 Primary	2016 GE	\
0	Did not vote/ineligible	Hillary Clinton	
1	Did not vote/ineligible	Did not vote/ineligible	
2	Hillary Clinton	Hillary Clinton	
3	Did not vote/ineligible	Did not vote/ineligible	
4	Did not vote/ineligible	Hillary Clinton	

	Electoral College Approval	Top Issue	Gun Reform Approval	\
0	No	Health Care	More Likely	
1	No	Climate Change	More Likely	
2	No	Climate Change	More Likely	
3	No	Climate Change	More Likely	
4	No	Climate Change	More Likely	

	Climate Change Approval
0	Very Important
1	Very Important
2	Very Important
3	Very Important
4	Very Important

[5 rows x 35 columns]

```
[8]: main_poll = poll_data[["XXXXXXXXXX status", "Gender", "Ethnicity", "Age Range",  
→ "Household Income",  
"Region", "Voting Registration", "Party Registration", "Primary/Caucus",  
→ "Ideology",  
"First Candidate", "Second Candidate"]]
```

```
[9]: main_poll.head()
```

	XXXXXXXXXX status	Gender	Ethnicity	Age Range	\
0	Current Student	Male	Hispanic or Latino	26-29	
1	Current Student	Male	White or Caucasian	18-21	
2	Current Student	Male	Other or Mixed Race	18-21	
3	Current Student	Male	Black or African American	18-21	

4	Current Student	Male	White or Caucasian	18-21
---	-----------------	------	--------------------	-------

	Household Income	Region	Voting Registration	\
0	Under \$50,000 a year	NorthEast	Yes	
1	\$50,000 to \$100,000 a year	NorthEast	Yes	
2	\$50,000 to \$100,000 a year	MidWest	Yes	
3	\$50,000 to \$100,000 a year	NorthEast	No	
4	Under \$50,000 a year	NorthEast	Yes	

	Party Registration	Primary/Caucus	Ideology	\
0	Democratic Party	Democratic Primary or Caucus	Very Liberal	
1	Democratic Party	Democratic Primary or Caucus	Very Liberal	
2	Democratic Party	Democratic Primary or Caucus	Very Liberal	
3	Democratic Party	Unsure	Somewhat Liberal	
4	Democratic Party	Democratic Primary or Caucus	Very Liberal	

	First Candidate	Second Candidate
0	Sanders	Warren
1	Warren	Sanders
2	Warren	Harris
3	Biden	Sanders
4	Sanders	Warren

```
[10]: main_poll.shape
```

```
[10]: (75, 12)
```

```
[12]: # Create a dataframe of aggregate candidate choices
```

```
first_candidate = main_poll["First Candidate"].value_counts()
first_candidate = (first_candidate * 100) / first_candidate.sum()
first_candidate = pd.DataFrame(first_candidate.items())
first_candidate.columns = ["Candidates", "Percentage"]
first_candidate["Percentage"] = first_candidate["Percentage"].apply(lambda x:
    round(x))
first_candidate["labels"] = first_candidate["Percentage"].apply(lambda x: "{}%".
    format(x))

# Margin of Error
s_e = sqrt((first_candidate.loc[0, "Percentage"] / 100) * (1 - (first_candidate.
    loc[0, "Percentage"] / 100)) / main_poll.shape[0])
cv = 1.64
ME = s_e * cv

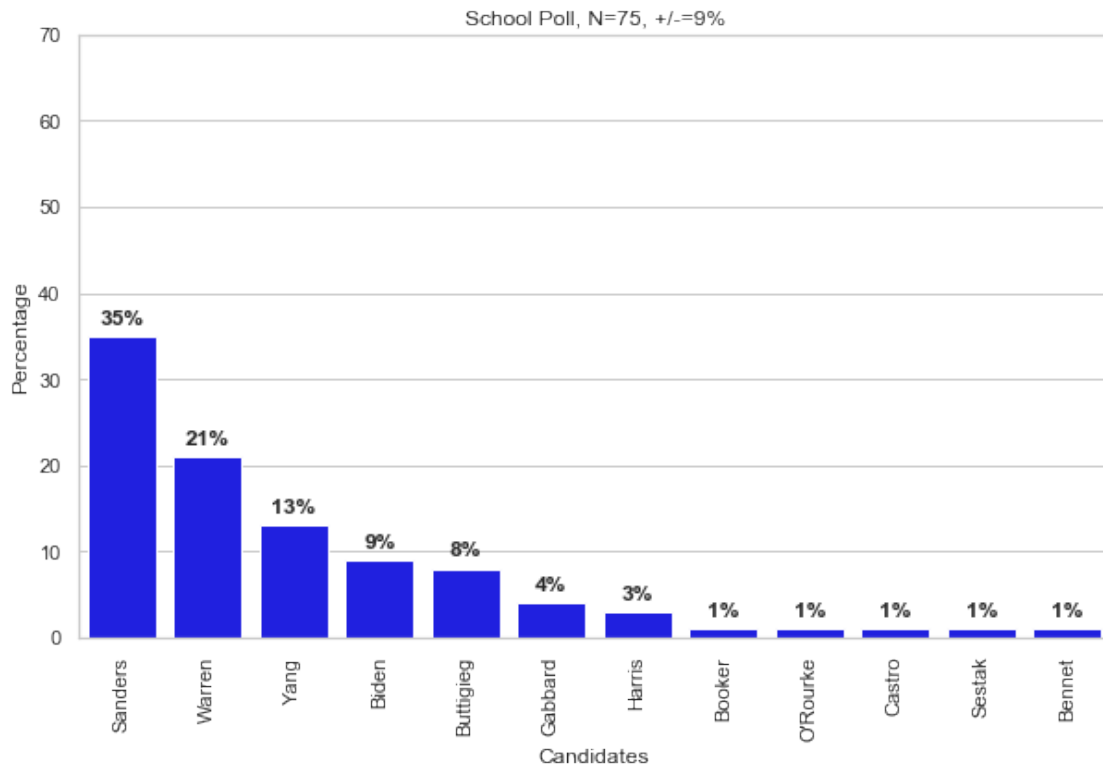
# Create plot
plt.figure(figsize=(10, 6))
sns.set(style = "whitegrid")
```

```

ax = sns.barplot(x=first_candidate["Candidates"],
    →y=first_candidate["Percentage"], color="Blue")
plt.xticks(range(first_candidate["Candidates"].shape[0]),
    →first_candidate["Candidates"], rotation=90)
ax.set_title("School Poll, N=75, +/-={}%"
    .format(round(ME*100)))
ax.set_ylim(top=first_candidate["Percentage"].max()*2)

bars = ax.patches
for rect, label in zip(bars, first_candidate["labels"]):
    height = rect.get_height()
    ax.text(rect.get_x() + rect.get_width() / 2, height + 1, label,
        ha='center', va='bottom', fontweight="bold", fontsize=12)
plt.show()

```



```
[13]: first_candidate[["Candidates", "labels"]]
```

```

[13]: Candidates labels
0     Sanders    35%
1     Warren    21%
2      Yang    13%
3     Biden     9%
4  Buttigieg    8%
5     Gabbard    4%

```

6	Harris	3%
7	Booker	1%
8	O'Rourke	1%
9	Castro	1%
10	Sestak	1%
11	Bennet	1%

```
[14]: # Second Candidates
second_choices = main_poll[["First Candidate", "Second Candidate"]]
second_choices = second_choices[second_choices["First Candidate"] != _]
    ↳ second_choices["Second Candidate"]]
second_choices["Count"] = 1
second_choices.groupby(["First Candidate", "Second Candidate"]).count()
```

```
[14]:
```

		Count
First Candidate	Second Candidate	
Bennet	Biden	1
Biden	Buttigieg	1
	Delaney	1
	Gabbard	1
	Harris	1
	Sanders	1
	Warren	1
	Yang	1
Booker	Yang	1
Buttigieg	Biden	2
	Harris	3
	Warren	1
Castro	Warren	1
Gabbard	Booker	1
Harris	Delaney	1
	Sanders	1
O'Rourke	Sanders	1
Sanders	Biden	2
	Castro	1
	Gabbard	1
	Warren	16
	Williamson	3
	Yang	3
Warren	Bennet	1
	Biden	4
	Booker	1
	Buttigieg	1
	Harris	2
	O'Rourke	1
	Sanders	6
Yang	Gabbard	2
	Sanders	5

[: