Voter Fraud Detection - **NON PARTISAN**

- No ID required to submit mail-in ballots, vulnerable to exploitation.
- Ballots have been found belonging to deceased or ineligible voters.

Look at voter registration data for PA(GA was $250 with 1-2 week delay)

- fields: voter ID number, name, sex, date of birth, date registered, status (i.e., active or inactive), date status last changed, party, residential address, mailing address, polling place, date last voted, all districts in which the voter votes (i.e., congressional, legislative, school district, etc.), voter history, and date the voter's record was last changed.

Use CloudyCluster web request job to check ballot status and authenticity in parallel, then analyze data to detect and model voter fraud.

Will impact the integrity of democracy.

Fraud Finders

Members:

- Carlos Miranda
- Frederick Morris
- Ronesha Shaw
- Cole McKnight(mentor)

Look at voter registration data for Pennsylvania.

Use CloudyCluster web request job to check ballot status and authenticity of suspected fraudulent voters in parallel, then analyze data to detect and model voter fraud in PA.

Deliverables:

- MVP: Analytics on suspected 2020 voter fraud in PA, model to find suspected fraudulent registrations.
- Visualization of 2020 voter fraud in PA based on ballot data.
- ML classification model to detect voter fraud based on readily available registration data(state independent).

Fraud Finders - Challenge 1

Members:

- Carlos Miranda
- Ronesha Shaw
- Frederick Morris
- Cole McKnight

Dataset: Pennsylvania Full Voter Export - 5.8Gi

- https://www.pavoterservices.pa.gov/pages/PurchasePAFULLVoterExport.aspx
- Statewide voter registration data.
- 153 features: Names, Active/Inactive, Voting History, DOB, County, Address, Registration Date, etc

We are using the voter registration data to detect suspected fraud.

We will inspect the features and define cases that indicate suspected fraud. We will process registrations suspected to be fraudulent through the PA Ballot Tracker to see if, when, and where a ballot was mailed in and if it was accepted.

We may use the ballot data with the registration data to train a classification model to detect suspected fraud.

```
df = pandas.read_csv("/content/drive/My Drive/Colab Notebooks/HPC in the City/ADAMS FVE 20201102.txt", sep='\t', lineterminator='\r', names=result)
df
```

```
/usr/local/lib/python3.6/dist-packages/IPython/core/interactiveshell.py:2718: DtypeWarning: Columns (12,21,24,29,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,9
  interactivity=interactivity, compiler=compiler, result=result)
```

	ID Number	Title	Last Name	First Name	Middle Name	Suffix	Gender	DOB	Registration Date	Voter Status	Status Change Date	Party Code	House Number	House Number Suffix	Street Name	Apartment Number	Address Line 2	C
0	001003958-01	NaN	WEIKERT	MARK	K	NaN	M	11/20/1932	01/01/1955	A	01/11/2017	R	13	NaN	FULTON DR	NaN	NaN	NEW OXFO
1	\n"001006201-01"	NaN	STOUFFER	JOHN	C	NaN	M	02/03/1927	01/01/1959	A	09/06/2017	D	6375	NaN	CHAMBERSBURG RD	#105	NaN	FAYETTEVI
2	\n"001006466-01"	NaN	ADKINS	ELLEN	E	NaN	F	09/06/1920	01/01/1960	I	07/10/2018	R	2990	NaN	CARLISLE PIKE	NaN	NaN	NEW OXFO
3	\n"001008846-01"	NaN	SMITH	JACK	A	NaN	M	04/08/1934	01/01/1964	I	07/31/2020	R	1639	NaN	COON RD	NaN	NaN	ASPI
4	\n"001010053-01"	NaN	RISS	JEAN	A	NaN	F	10/06/1940	01/01/1965	A	09/01/2015	R	2304	NaN	CHAMBERSBURG RD	NaN	NaN	BIGLERVI
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
72334	\n"111157206-01"	NaN	WARREN	FREDERICK	S	III	M	05/04/1972	10/28/2020	A	10/28/2020	NF	191	B	GREENBRIAR RD	NaN	NaN	YO SPRIN
72335	\n"111158380-01"	NaN	DECAVAIGNAC	LINDSEY	PAYTON	NaN	F	10/12/2001	10/29/2020	A	10/29/2020	D	300	NaN	N WASHINGTON ST	0567	NaN	GETTYSBU
72336	\n"111158407-01"	NaN	DIPRETE	CHARLOTTE	CALAT	NaN	F	03/09/1999	10/29/2020	A	10/29/2020	NO	42	NaN	E LINCOLN AVE	NaN	NaN	GETTYSBU
72337	\n"111160304-01"	NaN	WASSUTA	MICHAEL	NaN	NaN	M	02/22/1952	10/30/2020	A	10/30/2020	NF	2490	NaN	CHAMBERSBURG RD	NaN	PO BOX 4865	GETTYSBU
72338	\n	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

72339 rows × 153 columns

Fraud Finders - Challenge 2

- Goal: Query PA ballot tracking website with registrations suspected of fraud

    ~5000 suspected entries

1. Get input elements from entries suspected of fraud(Last, First, DOB, County)
2. Create Slurm batch job to query website and get html output.
3. Clean html output and organize into dataframe.

-
- Bonus Task: Check number of registrations against eligible voting population for each county

New Judicial Watch Study Finds 353 U.S. Counties in 29 States with Voter Registration Rates Exceeding 100%

https://www.judicialwatch.org/press-releases/new-jw-study-voter-registration/

Fraud Finders - Progress

1. Get input elements from entries suspected of fraud(Last, First, DOB, County)
    - Used awk to get input fields
2. Create Slurm batch job to query website and get html output.
    - Built single sample working model pulling HTML data using curl
    - Parallelizing the model to be run on Cloudy Cluster
    - 67 batch jobs processing 2500 queries
3. Clean html output and organize into dataframe.
    - Obtained single sample HTML file output
    - Designing job to clean and aggregate into files by county

Bonus Task: Check number of registrations against eligible voting population for each county

- Compared statewide voter registrations to number of eligible voters