Campaign Finance

Data Science Immersive Remote

Andrea Pascale

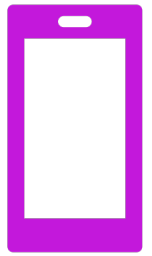
Overview

Data collection and storage

Defining a data science experiment

Data wrangling

Creating predictive models

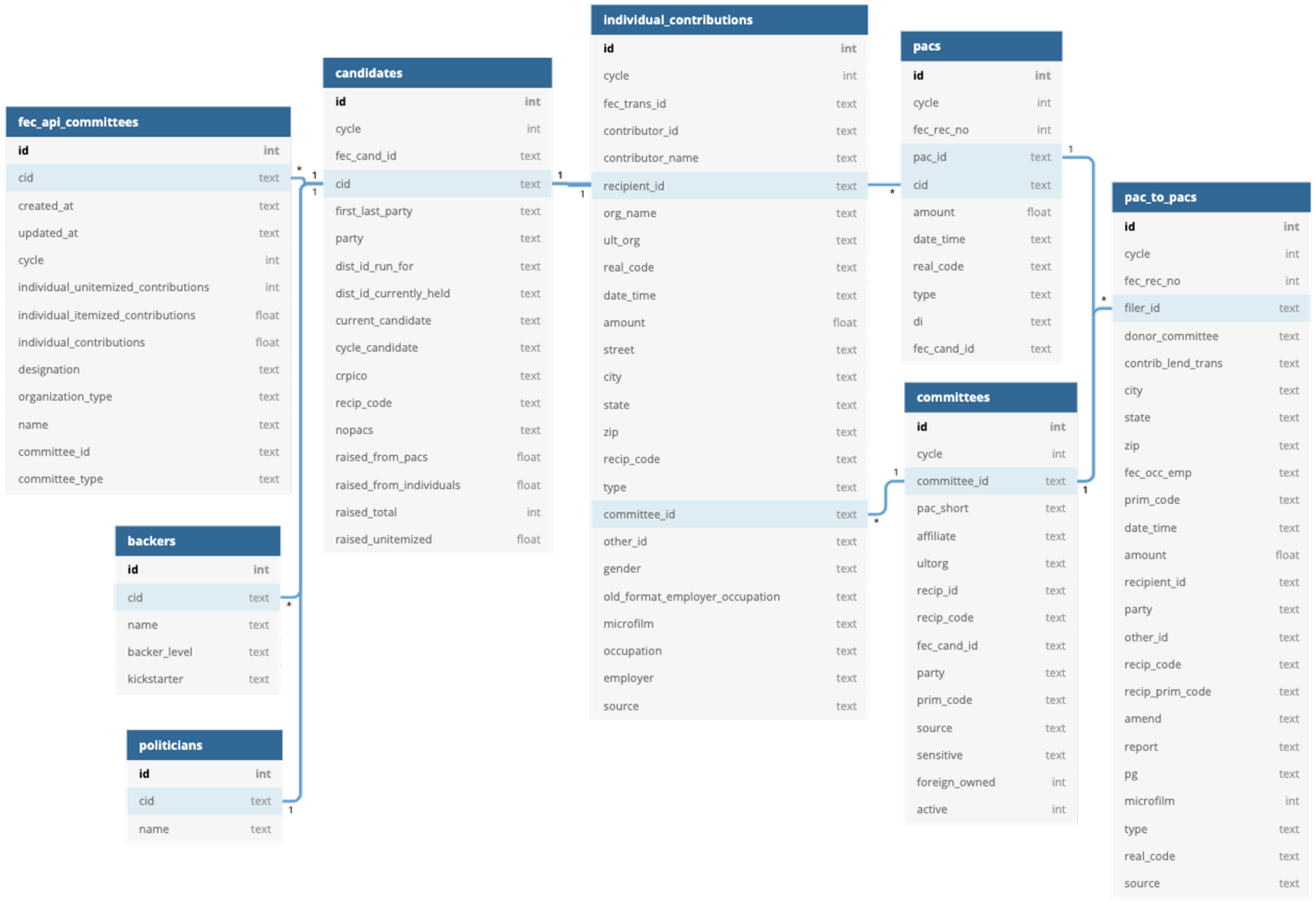


Web App

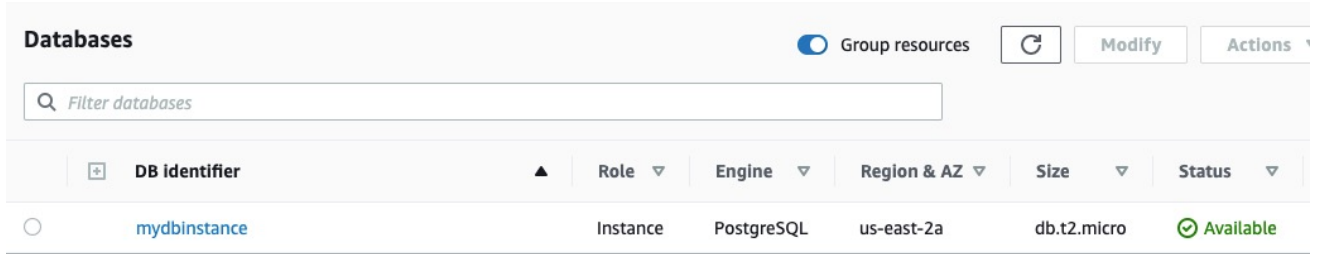
Further exploration

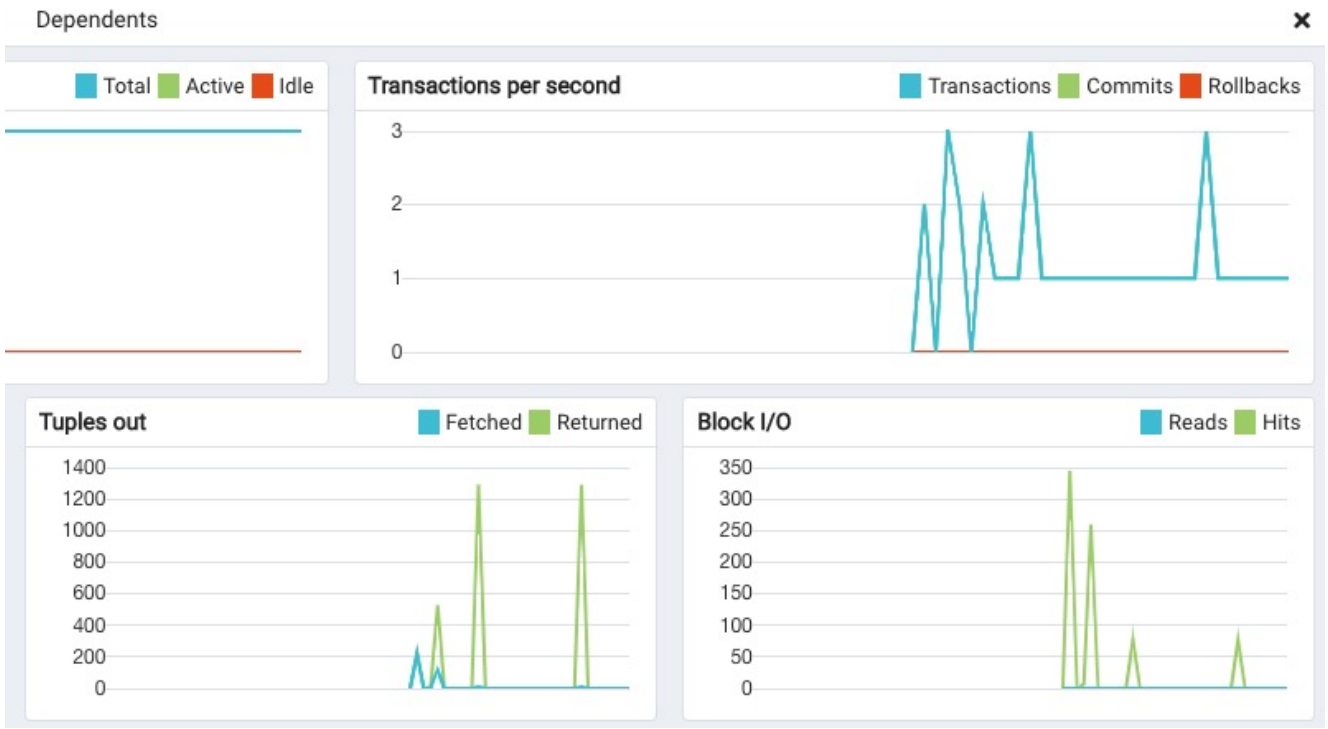
Original Dataset

- 30+ million records
- From 1990-2016



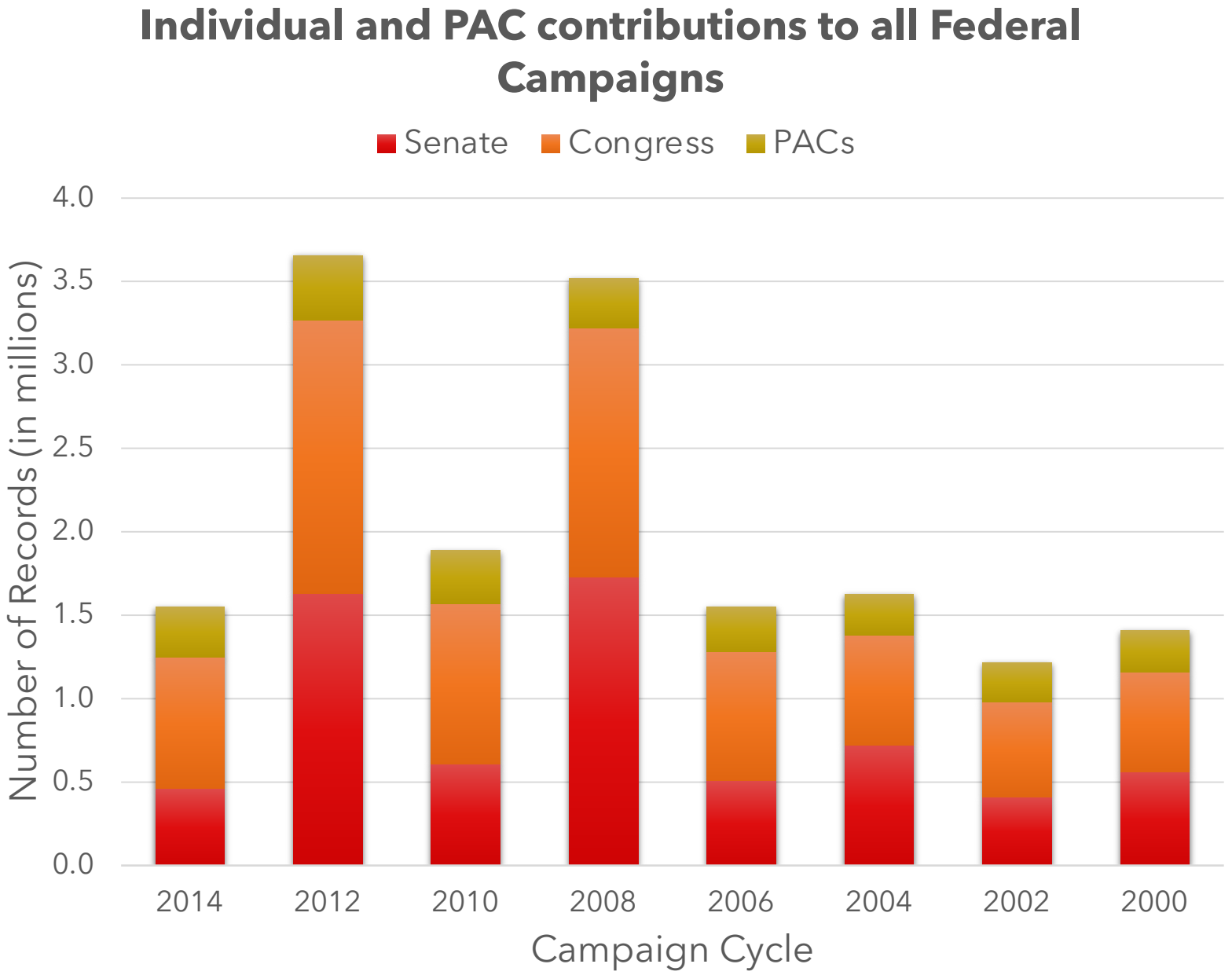
Database Configuration





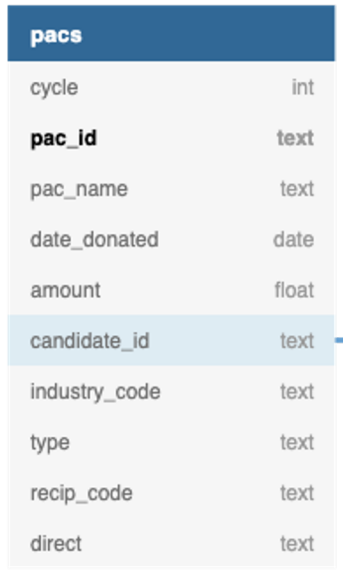
Aggregating by Cycle

- From 2000-2014
- Greater number of records in presidential election years

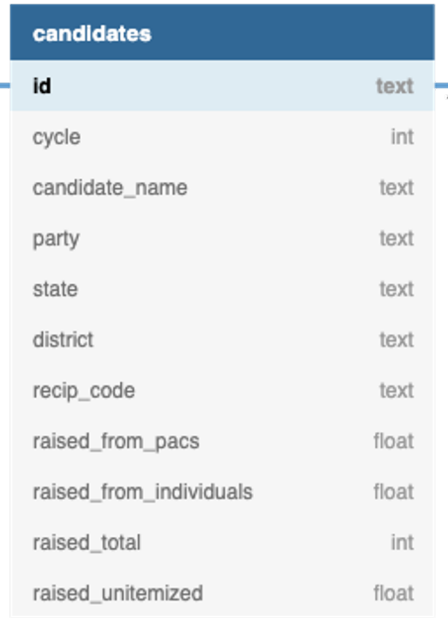


Final Schema

2.5 million records



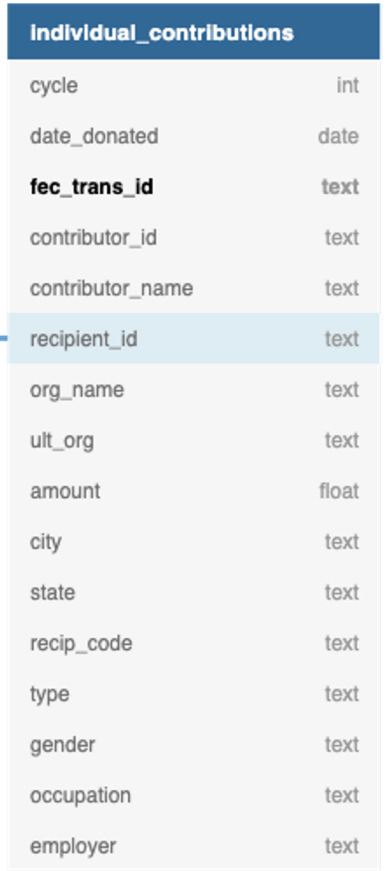
pacs	
cycle	int
pac_id	**text**
pac_name	text
date_donated	date
amount	float
candidate_id	text
industry_code	text
type	text
recip_code	text
direct	text



candidates	
id	**text**
cycle	int
candidate_name	text
party	text
state	text
district	text
recip_code	text
raised_from_pacs	float
raised_from_individuals	float
raised_total	int
raised_unitemized	float

68,000 records

14 million records

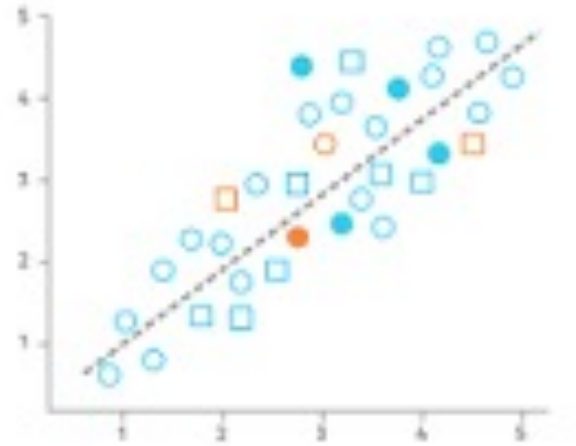


individual_contributions	
cycle	int
date_donated	date
fec_trans_id	**text**
contributor_id	text
contributor_name	text
recipient_id	text
org_name	text
ult_org	text
amount	float
city	text
state	text
recip_code	text
type	text
gender	text
occupation	text
employer	text

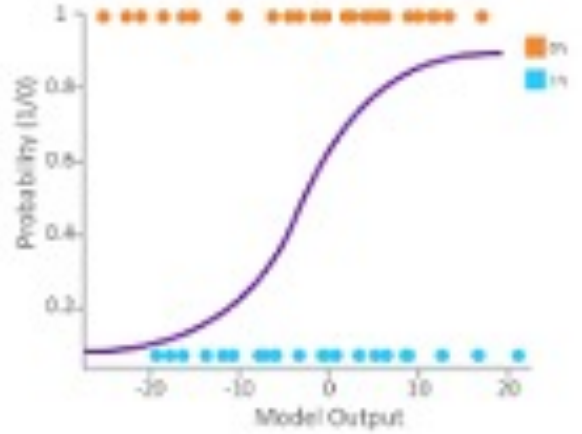
Data Science Experiment

- Linear Regression
  - To predict the amount of money that needs to be raised to win a congressional or senate race
- Logistic Regression
  - To predict the probability of winning based on target fundraising goals

Linear Regression



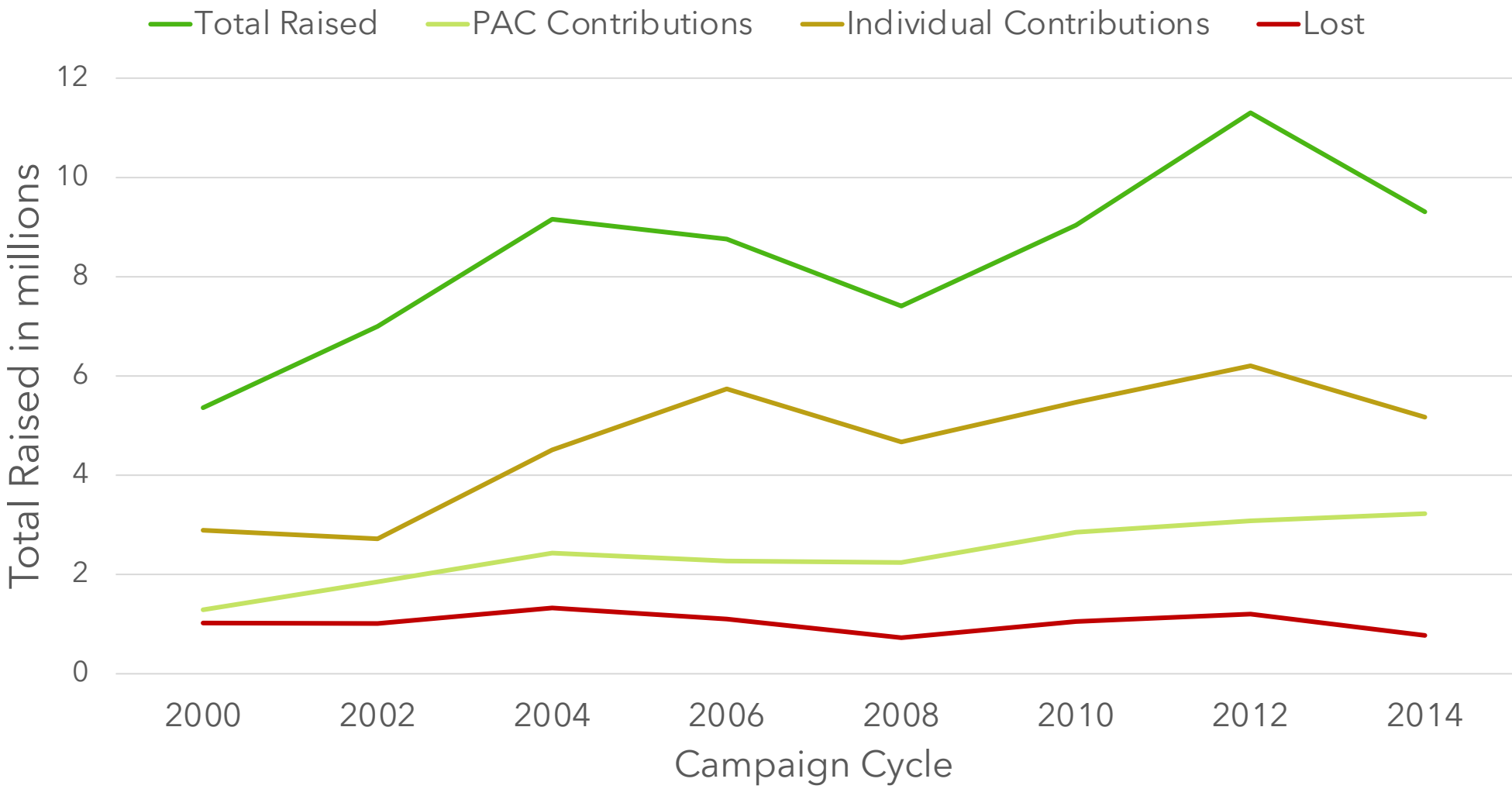
Logistic Regression

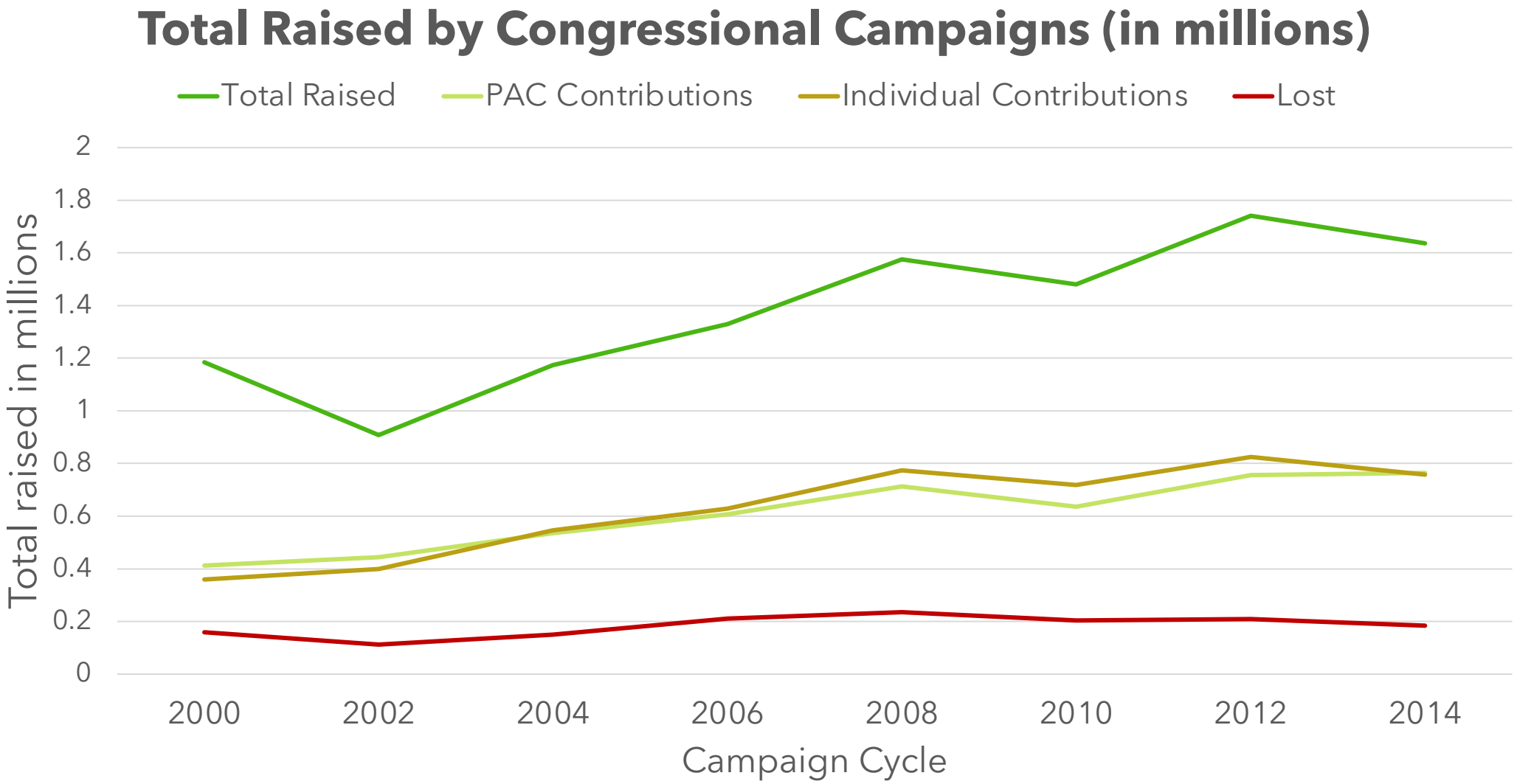


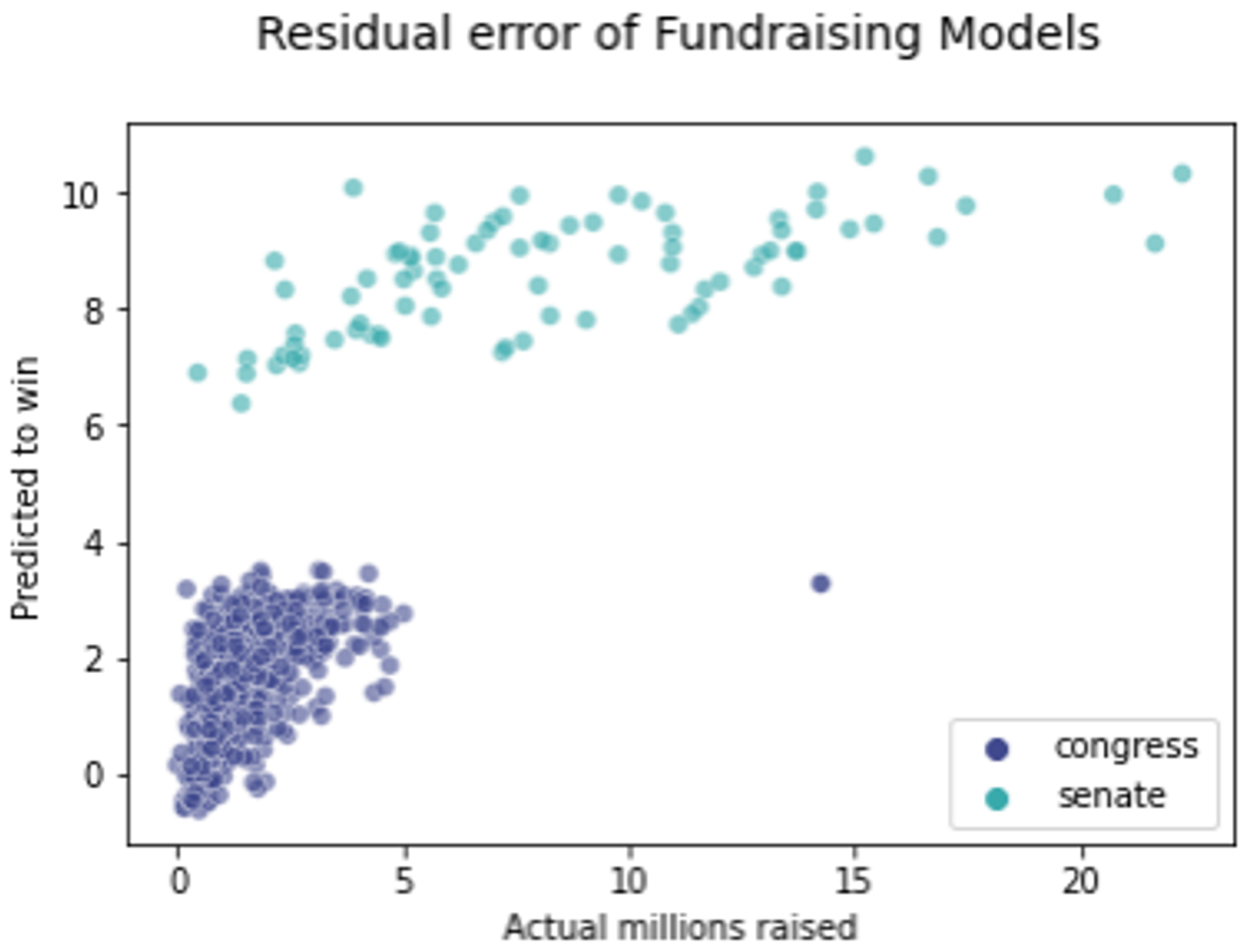
Data Wrangling

Delete	Separate	Reassign	Extract
U.S. territories & D.C. ▪ Guam ▪ Virgin Islands ▪ American Samoa ▪ Puerto Rico Any districts matching ‘00’	Congressional and Senate races Calculate ranking per state/district	Third-party candidates ▪ Bernie Sanders ▪ Angus King ▪ Lisa Murkowski	Recip Codes ▪ W = win ▪ L = lost ▪ I = incumbent ▪ C = challenge ▪ O = open seat

Total Raised by Senate Campaigns (in millions)

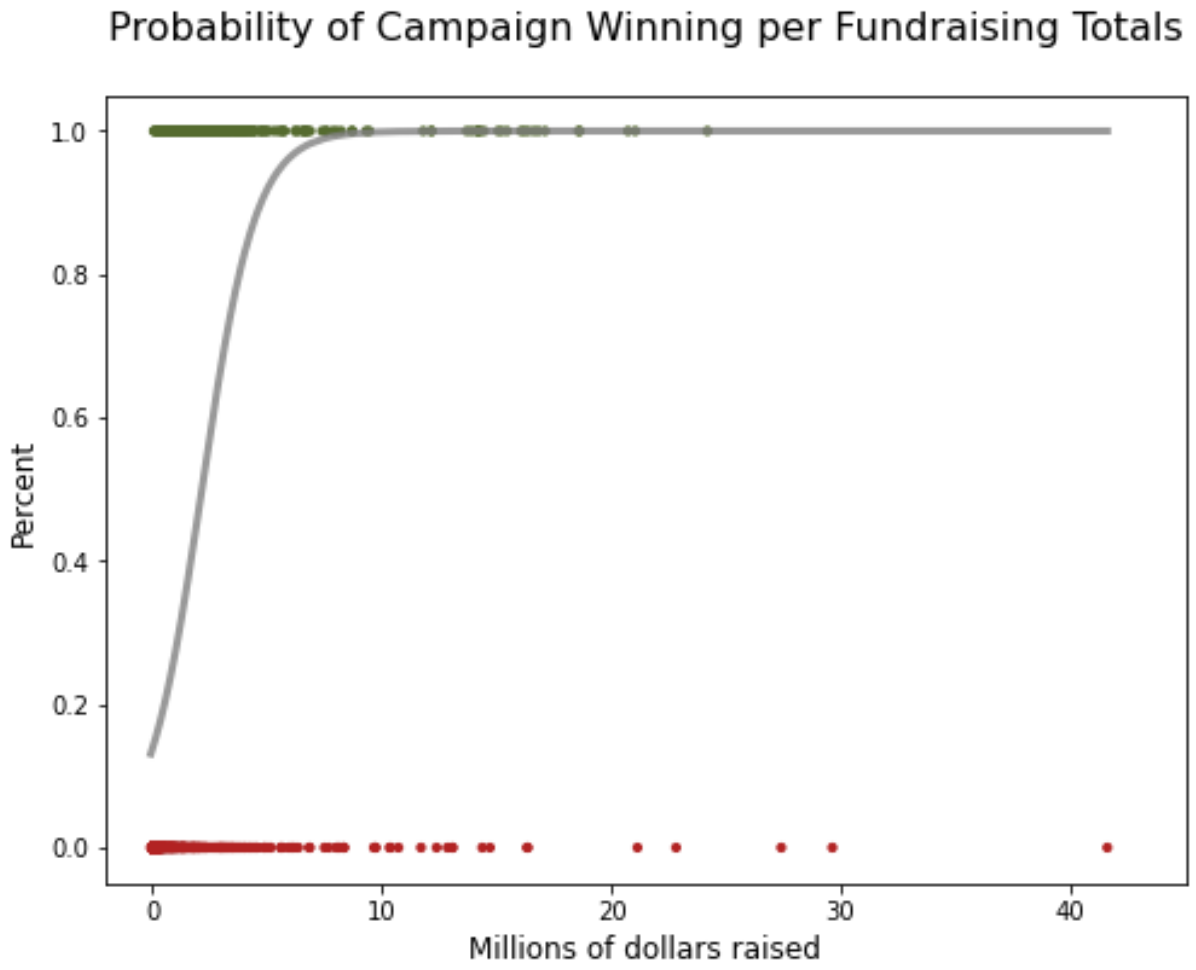



Total Raised by Congressional Campaigns (in millions)
Total Raised
PAC Contributions
Individual Contributions
Lost
Total raised in millions
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
2000
2002
2004
2006
2008
2010
2012
2014
Campaign Cycle



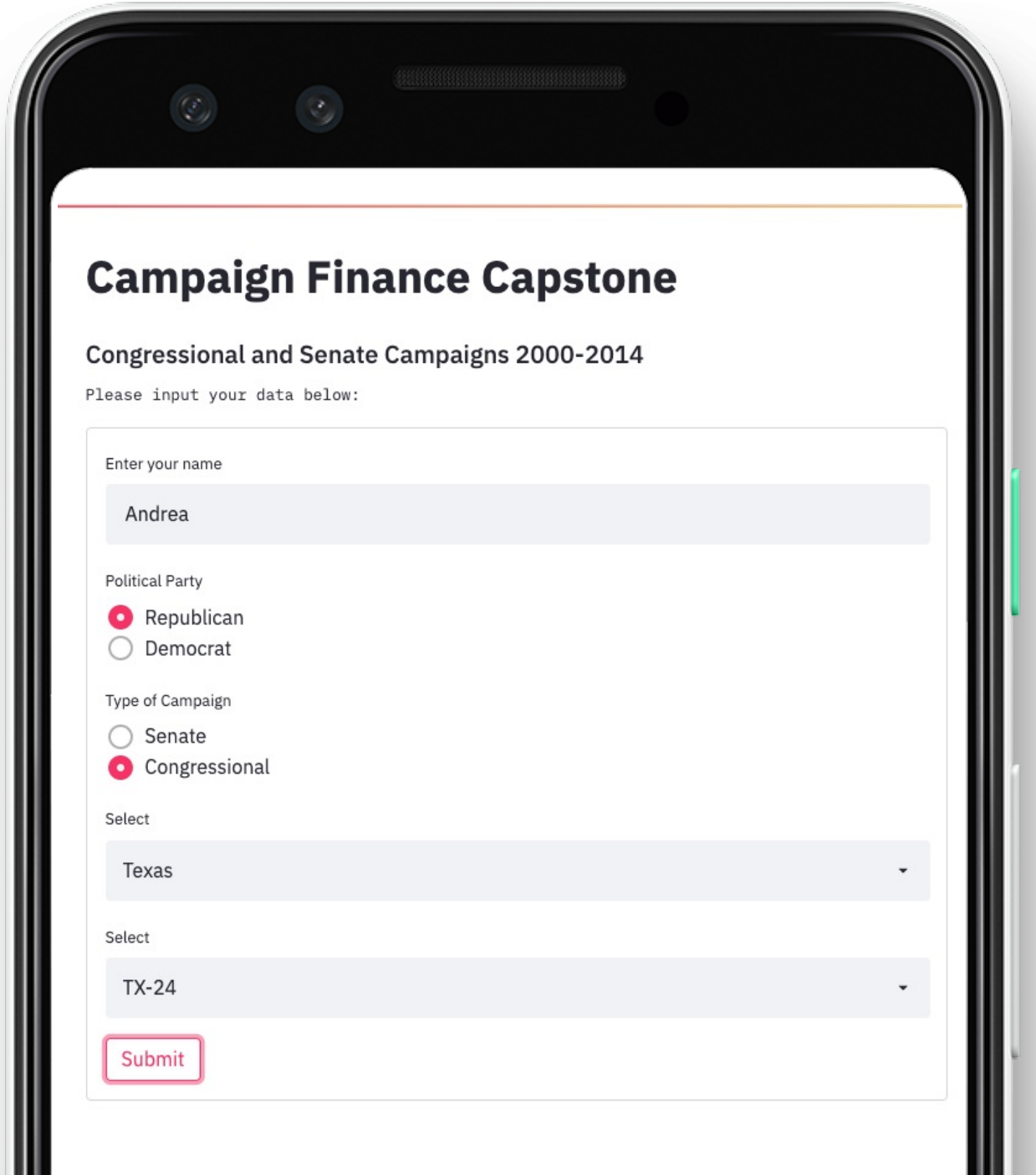
Linear Regression Model

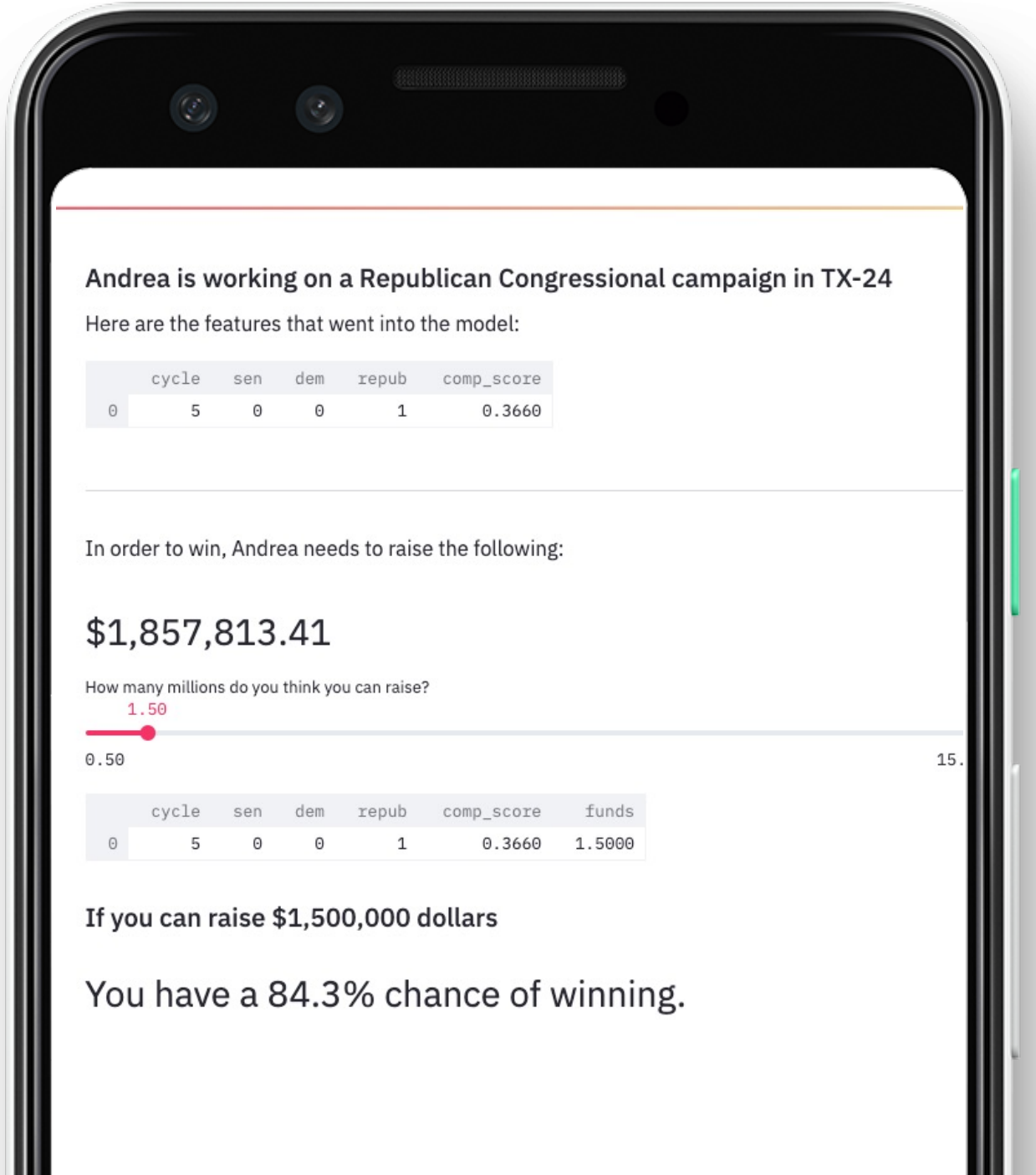
- Features
  - cycle, party affiliation, senate/congressional campaign, and competitive ranking
- Accuracy
  - 56% on training data
  - 64% on test data



Logistic Regression Model

- Features
  - cycle, party affiliation, senate/congressional campaign, competitive ranking, and total raised
- Accuracy
  - 82% on training data
  - 81% on test data


Campaign Finance Capstone
Congressional and Senate Campaigns 2000-2014
Please input your data below:
Enter your name
Andrea
Political Party
Republican
Democrat
Type of Campaign
Senate
Congressional
Select
Texas
Select
TX-24
Submit


Andrea is working on a Republican Congressional campaign in TX-24
Here are the features that went into the model:
cycle sen dem repub comp_score
0 5 0 0 1 0.3660
In order to win, Andrea needs to raise the following:
$1,857,813.41
How many millions do you think you can raise?
1.50
0.50
15.
cycle sen dem repub comp_score funds
0 5 0 0 1 0.3660 1.5000
If you can raise $1,500,000 dollars
You have a 84.3% chance of winning.

Next Steps

Clustering of candidate data (candidate profiles)

- Generic senate candidate in the Midwest region should earn X dollars and target Y donors

PAC and individual donor recommender system

- Correlate based on issue set or political affiliation

Seasonality in corporate and individual giving

- Develop separate models for midterm and presidential years