Predict the Flip!

A U.S. Senate Election Analysis

By Darius Fuller

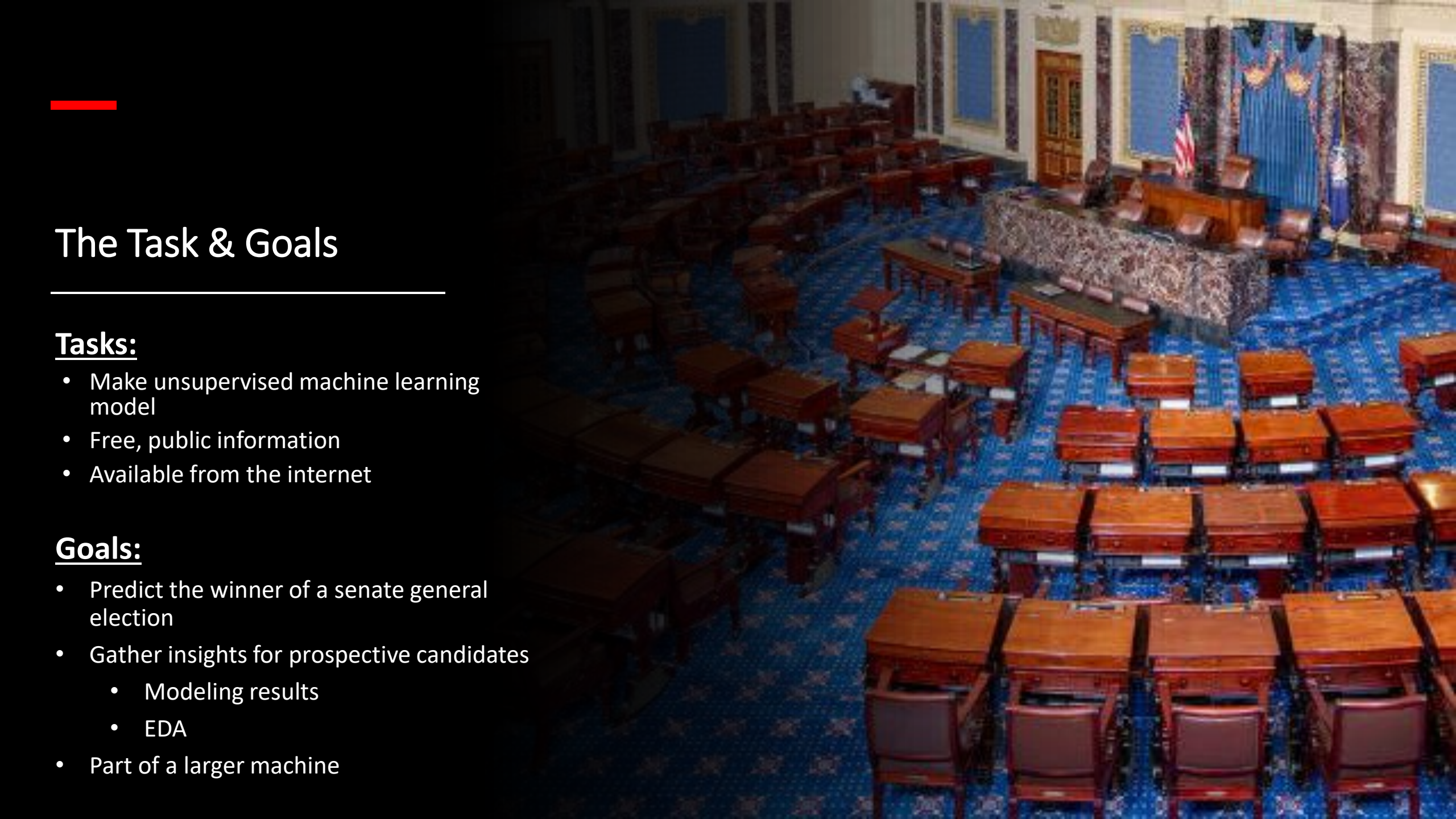
The Task & Goals

Tasks:

- Make unsupervised machine learning model
- Free, public information
- Available from the internet

Goals:

- Predict the winner of a senate general election
- Gather insights for prospective candidates
  - Modeling results
  - EDA
- Part of a larger machine



The Dataset

What:

- Candidates in senate general elections
- 1920-2016
- 5,588 datapoints
- 8 total features

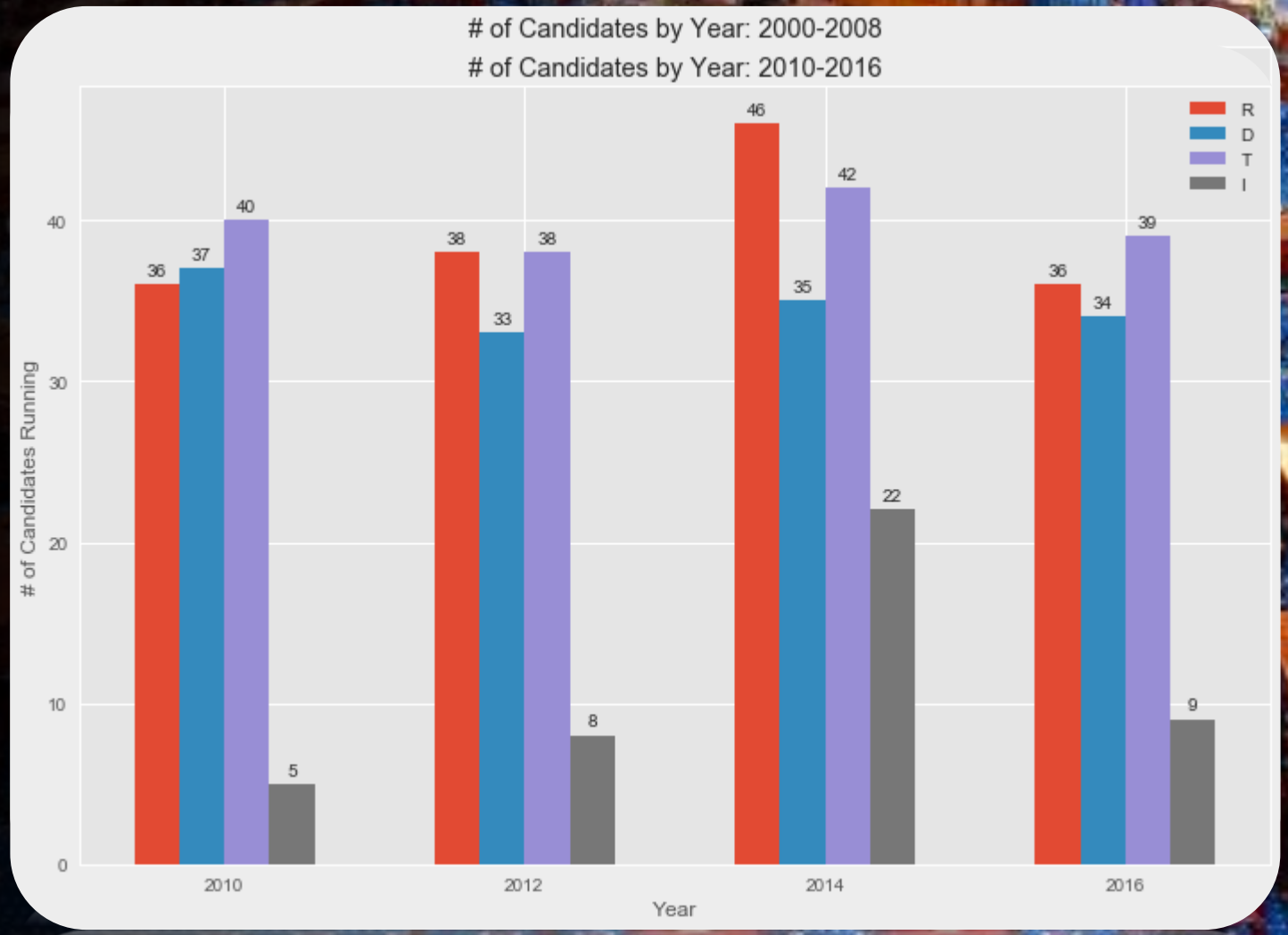
How:

- ‘Web scraping’
  - BeautifulSoup package
  - Pandas package
- Genderize.io

The Dataset - Explored

Stiff Competition:

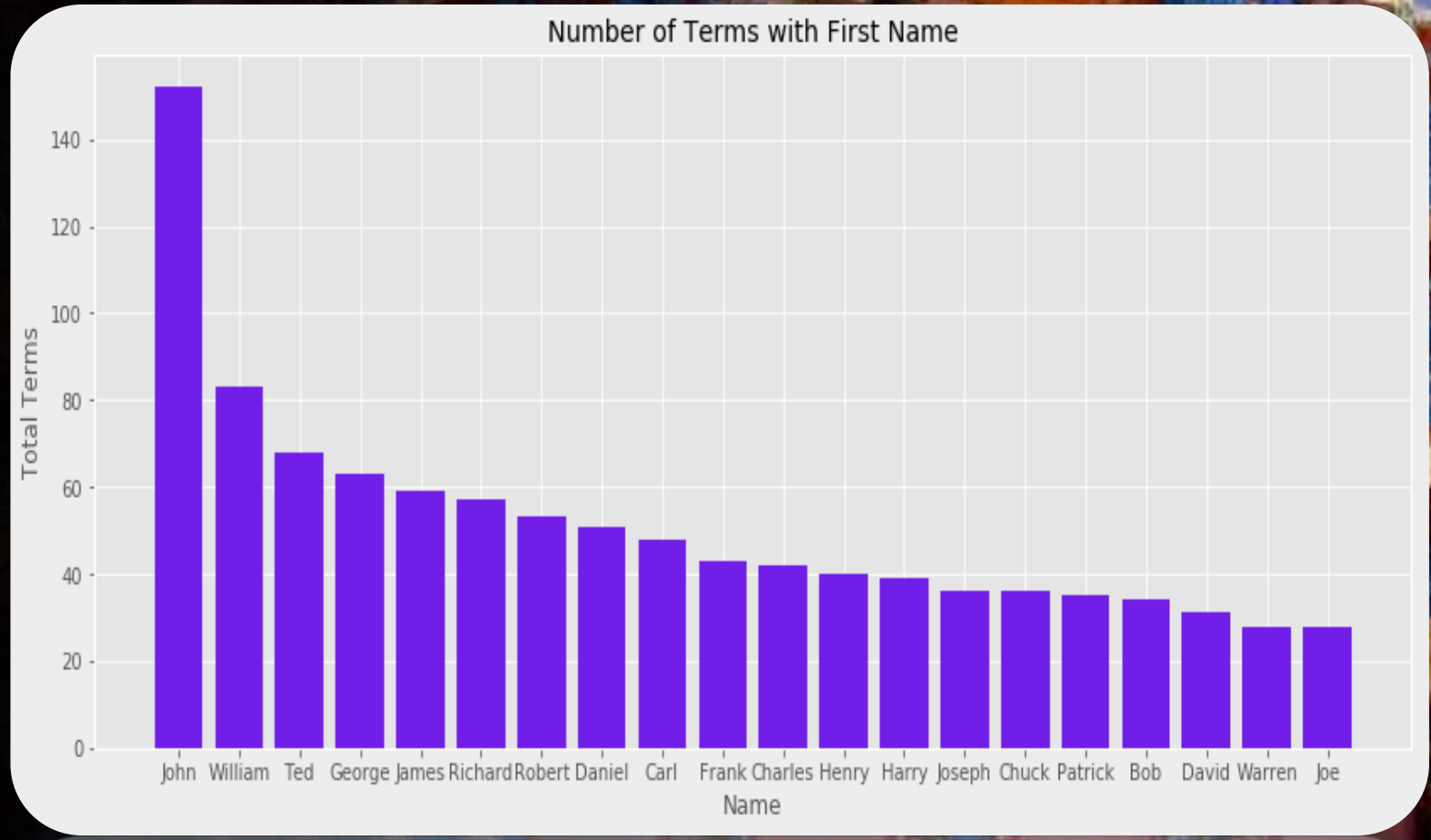
- Number of participants
  - Per party
- Trends & Oddities
  - Party relevance



The Dataset - Explored

No surprises:

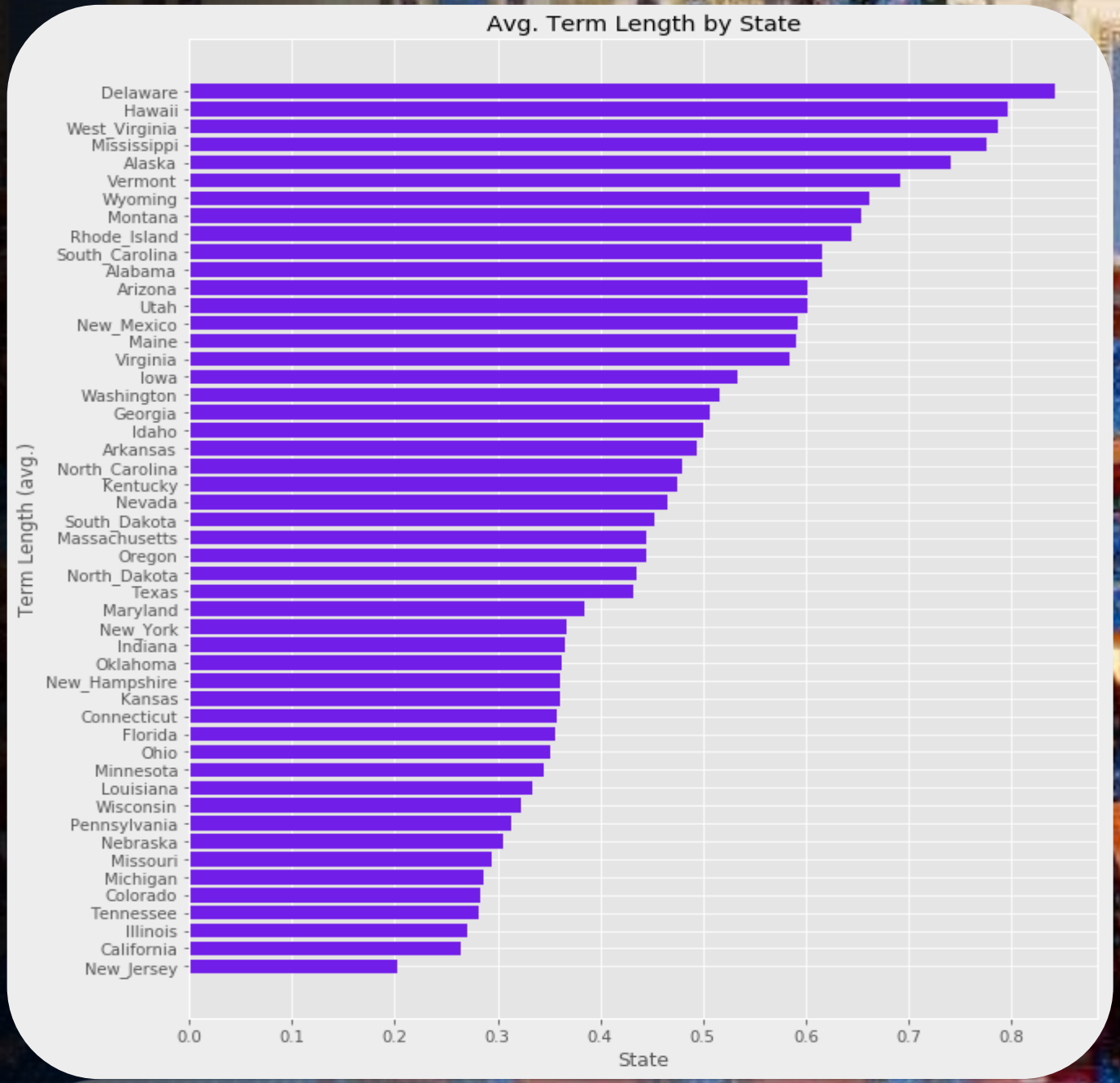
- Traditional names
- Barbara was 1st female name
  - 34th most terms (16 appearances)
- More potential
  - Last names?



The Dataset - Explored

Big surprises:

- Accounts for all candidates each year
- Higher number means:
  - Little to no competition
  - More consecutive terms
  - Vice-versa



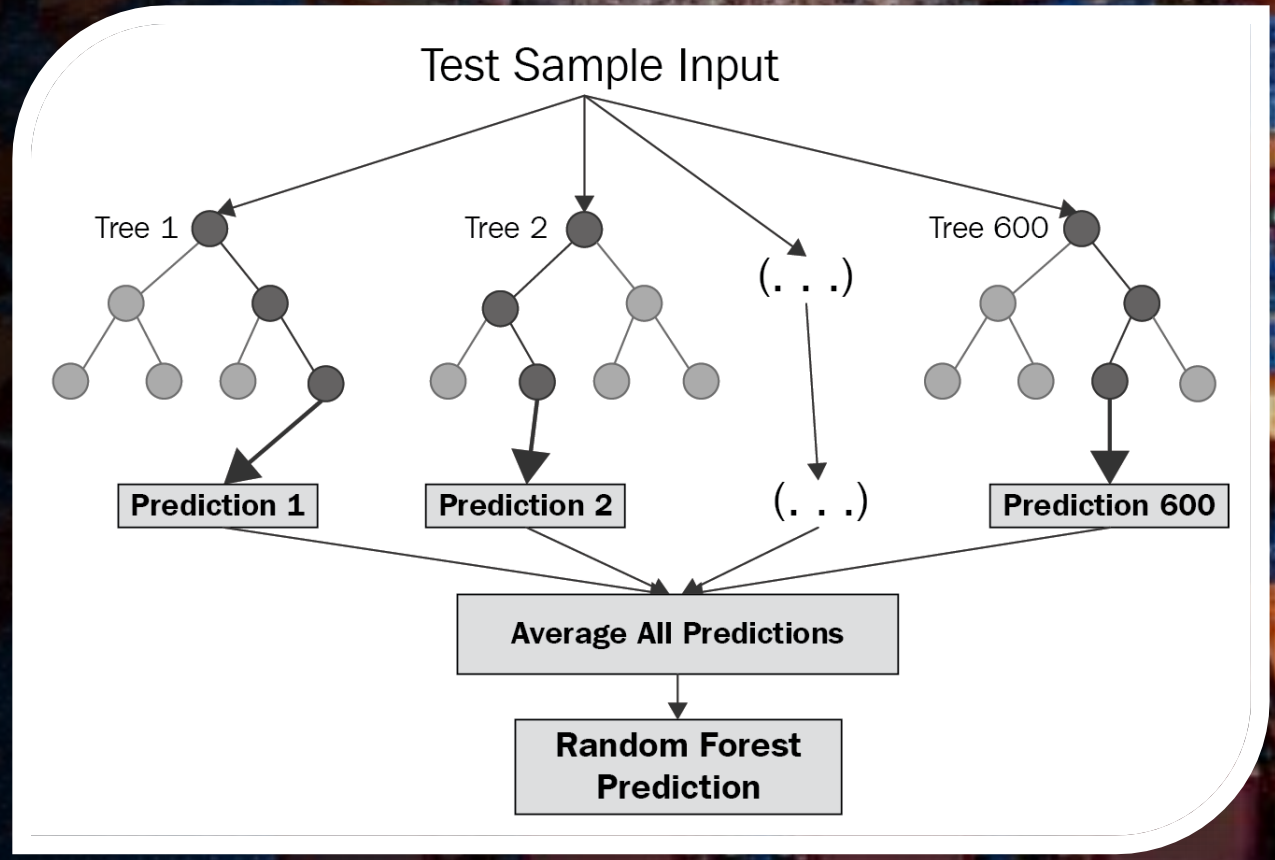
Modeling Process

What:

- Linear regression
  - Predict the percentage of votes
- Random Forest from Sci-Kit Learn
  - ‘Decision trees’

How:

- Learns from training data
  - ‘Make decisions’ based on learned underlying patterns
- Predict using unseen data



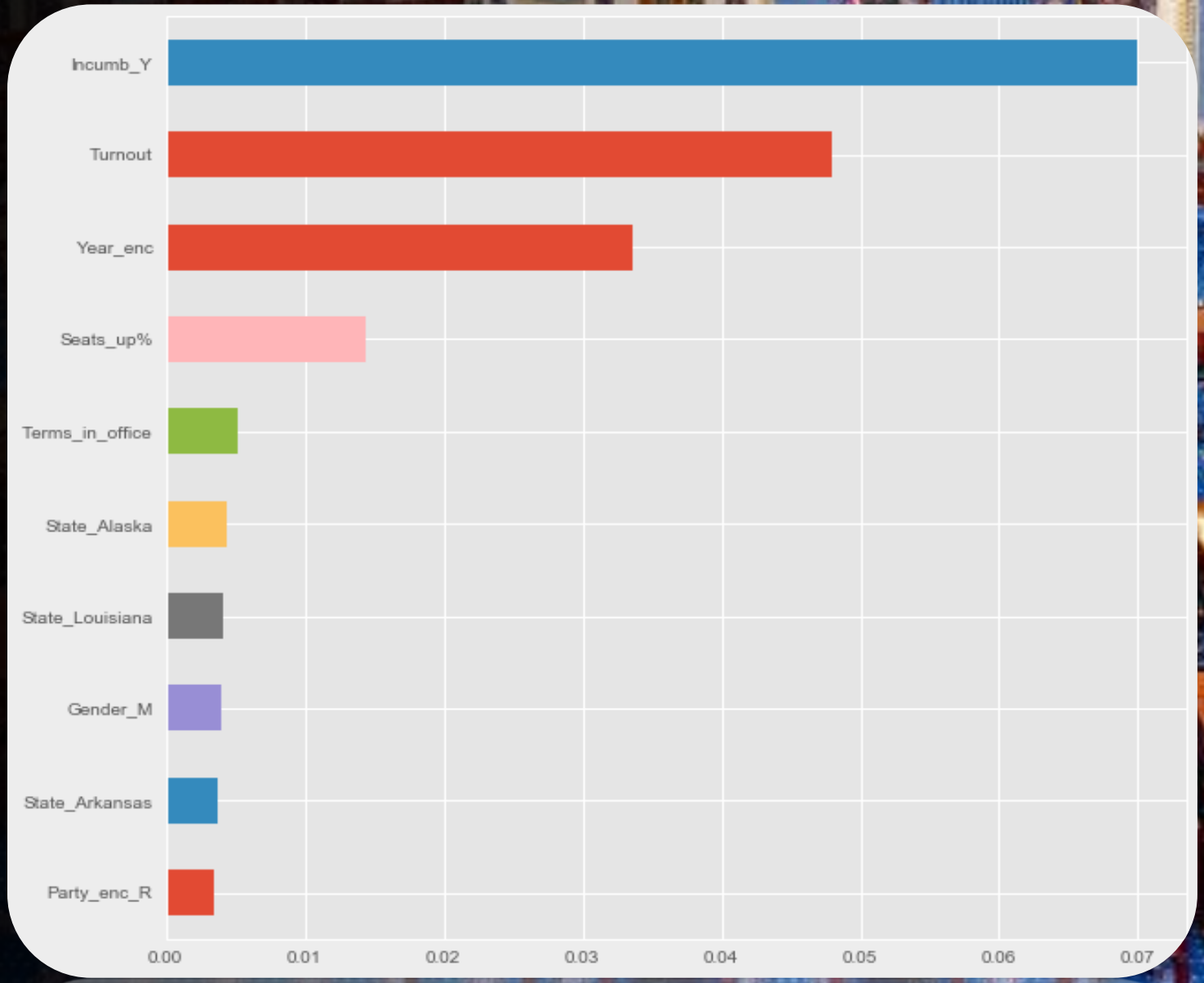
Results

Feature Importance:

- Represents the weight each feature has
  - Does not indicate direction
  - Relative sense
- Seats_before% = 0.75
- Top 10 (w/o #1 feature)

Insights:

- Timing is critical
  - Prior year is most important
- Experience is crucial
- Party does matter



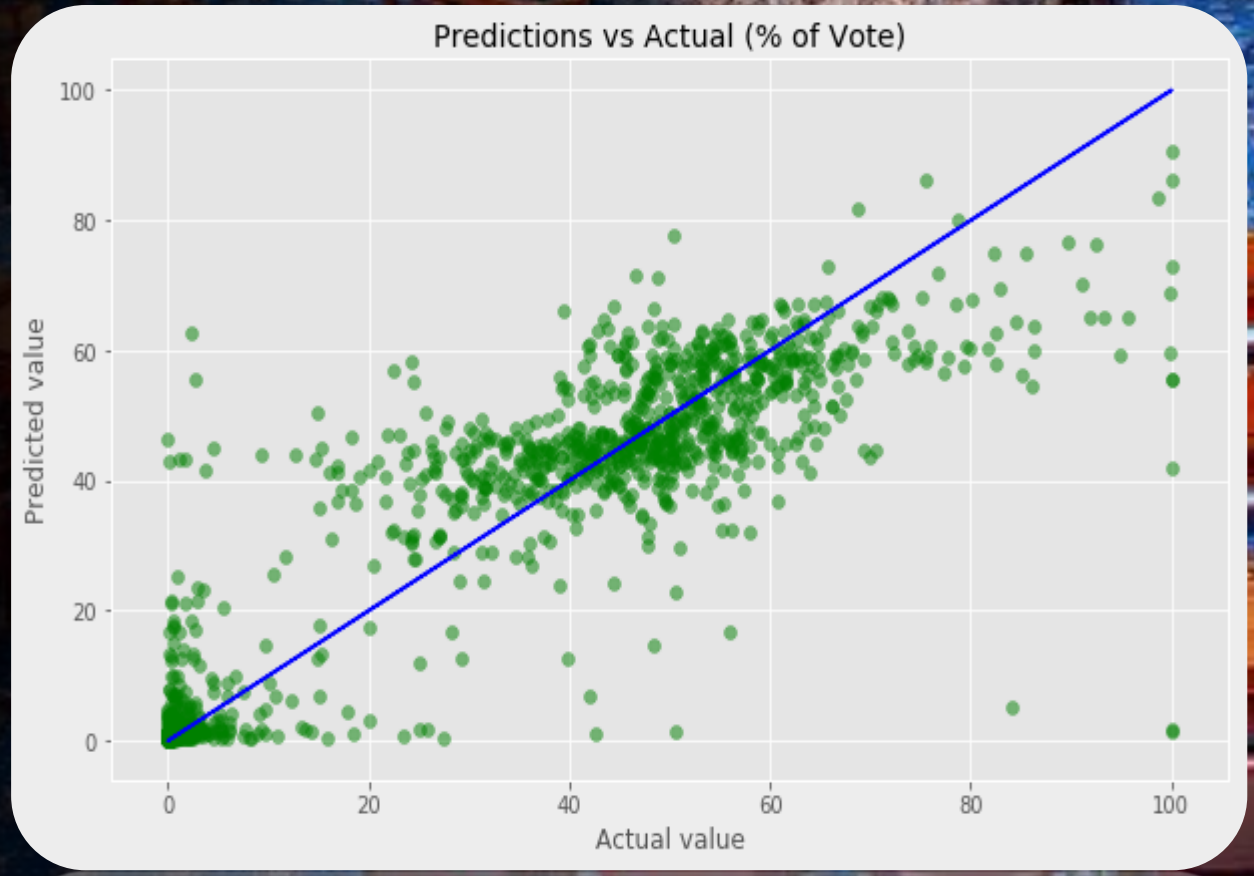
Results

Model Performance:

- Root Mean Squared Error (RMSE)
- Best model
  - RMSE = 10.863349
- % of votes

Observations:

- ‘Mirrored’ low end
- Bottom-heavy upper bound
- Strong cluster in center



Recommendations

- **<u>Controllable:</u>**
- Decide to be either Democratic or Republican
  - Being different doesn’t help
- Get Out the Vote (GOTV)
  - May be party dependent
- Incumbency is key

- **<u>Environmental:</u>**
- Last cycle’s performance sets the tone
  - Ride the wave
- State elections, national stage

Future Works

More Data:

- Cycles happen every 2 years
  - Take some time
- Increase breadth of features
  - More controllable variables
  - Age, birthplace, number of attempts, military service, etc.

Building Bigger:

- Implement model to predict state elections
  - Given a set of candidates
- Aggregate on a national level

Thank You!