

# Sociological Summary of State Elections

Thousands of individuals contest various state elections every year in India. Since 2004, it has been made mandatory for contesting candidates to reveal their sociological background when filing their nomination (including their education details, assets etc). These affidavits have been processed to a digital format on [myneta.info](https://myneta.info). Additionally, a collated data table for electoral information for every state election since 1962 is available for download on [lokdhaba.ashoka.edu.in](https://lokdhaba.ashoka.edu.in).

Based on this information, one can answer many questions, such as:

1. What is the sociological composition of a particular state assembly?
2. Does a state have an increasing number of criminal candidates being elected every assembly?
3. Does a particular subregion of a state elect candidates from a typical professional background?

The question can be temporal (trends across years) or spatial (focused on a specific election). Form a relevant question that you can ask based on this data.

Your task would be to scrape this website to extract the data required for your research question using these sources. You can use Scrapy to follow the links on your selected state page (example:

[https://myneta.info/Bihar2020/index.php?action=summary&subAction=candidates\\_analyzed&sort=candidate#summary](https://myneta.info/Bihar2020/index.php?action=summary&subAction=candidates_analyzed&sort=candidate#summary)), visit each candidate page (example:

[https://myneta.info/Bihar2020/candidate.php?candidate\\_id=11423](https://myneta.info/Bihar2020/candidate.php?candidate_id=11423)), and append the information to a table. ALTERNATIVELY, you can use Selenium to emulate clicks on your selected state page to visit each candidate page.

After compiling the scraped data into a tabular format (use CSV writer), you can use either Excel/LibreOffice Calc to generate static visualisation or tools like Matplotlib to generate dynamic visualisation to illustrate and answer your selected question.

## Exercise 5

Submit your tabulated data.

## Project

Submit your sociological analyses of the data.

## Resources

Feel free to use any online resources to help your understanding!

- Scrapy
  - To follow links on a page (you can also just use 2 spiders, one for getting the links to candidate pages and other to scrape them, like we did for newspaper search):
    - <https://docs.scrapy.org/en/latest/intro/tutorial.html#following-links>
    - <https://stackoverflow.com/a/30152967>
- OPTIONAL - Selenium: <https://towardsdatascience.com/web-scraping-using-selenium-python-8a60f4cf40ab>
- [Matplotlib](#)
  - Tutorials: <https://matplotlib.org/stable/tutorials/index.html>
  - Types of plots with examples:  
[https://matplotlib.org/stable/plot\\_types/index.html#basic](https://matplotlib.org/stable/plot_types/index.html#basic)
  - Basic guide: <https://matplotlib.org/stable/tutorials/introductory/pyplot.html>
  - Numerous examples: <https://matplotlib.org/stable/gallery/index.html>
  - In-depth explanation of basic usage:  
<https://matplotlib.org/stable/tutorials/introductory/usage.html#sphx-glr-tutorials-introductory-usage-py>