Ashlyn Montgomery, Kameron Galm, Nam Jun Lee, Julian Rangel

DATA424 – Online 1

19 February 2023

The data that we received is comprised of approximately 150 PDF files, and a 52.4 MB .CSV file filled with aerospace data. These documents were received from the professor, based on previous years’ research. Each of the PDFs has data based around the number of women versus men in aerospace engineering, or regarding women in technological sectors in general. The .CSV file contains data on respondents that includes gender, job position, and salary. Our proposed method of obtaining data is by scraping each of the PDFs for gendered words and comparing the totals. In terms of storage, we will use local machines and GitLab for sharing documents.

