Limitations

Using Grad Café and similar self-reported websites as data sources comes with some built-in limitations. First, web scraping raises legal and ethical concerns. Not every website allows automated data collection, and scraping without checking the rules can lead to consequences such as having your IP address blocked or losing access to the site altogether. Because of this, it is important to carefully review a website's terms of service and confirm that scraping is permitted before collecting any data. Even if scraping is technically possible, that does not necessarily mean it is allowed.

Second, the technical process of scraping can be slow and challenging. Websites are not designed to make data extraction easy — relevant information is often buried deep within complex HTML elements, making it difficult to reliably locate and organize. In addition, scraping large volumes of data can be time-consuming and memory-intensive. Differences in page structure, inconsistent formatting, and unstructured user-submitted content can also result in messy or incomplete datasets, requiring additional cleaning and processing before analysis.