

Wrangle report

Data wrangling is an essential part of the data analysis process, and it can involve working with various types of data sources. In this report, we will focus on three data resources: CSV files, web scraping, and the Twitter API.

CSV (Comma Separated Values) files are a widely used format for storing and sharing tabular data. They can be easily imported into various data analysis tools and are often used to store and exchange data between different systems. However, CSV files can also be prone to errors, such as missing or incorrectly formatted values, which can require additional effort to correct during the data wrangling process.

Web scraping involves extracting data from websites using specialized software or scripts. This can be a useful way to gather data from websites that do not provide an API or other means of accessing their data. However, web scraping can be a time-consuming and technically challenging process, as it requires understanding HTML and potentially dealing with issues such as CAPTCHAs and website rate limits.

The Twitter API is a service provided by Twitter that allows developers to access and extract data from the social media platform. This can be a useful source of data for analysis, but it also requires some effort to set up and use. Developers must register for an API key and follow Twitter's terms of service, and they may also need to deal with issues such as rate limits and data restrictions.

Overall, data wrangling using these resources can require a significant amount of effort, particularly when working with large or complex datasets. However, with the right tools and techniques, it is possible to effectively clean and structure data for further analysis and visualization.