Recommended Data Sources

Berkeley Library Resources for Text Mining & Computational Text Analysis
https://guides.lib.berkeley.edu/text-mining

Gale Primary Sources on the Digital Scholar Lab (requires Berkeley log in)
https://libproxy.berkeley.edu/login?qurl=https://infotrac.gale.com/itweb/ucberkeley?db=DSLAB

HathiTrust Digital Library (log in with Berkeley credentials for better access)
https://www.hathitrust.org/

Project Gutenberg
https://www.gutenberg.org/

International Consortium for Political and Social Research (ICPSR)
https://www.icpsr.umich.edu/web/pages/

Harvard Dataverse
https://dataverse.harvard.edu/

US Federal Government open data portal
https://catalog.data.gov/dataset

State of California open data portal
https://data.ca.gov/

City of San Francisco open data portal
https://datasf.org/opendata/

Google Dataset Search
https://datasetsearch.research.google.com/

Kaggle
https://www.kaggle.com/datasets

Twitter API
https://developer.twitter.com/en/docs/twitter-api/early-access

Reddit API
https://www.reddit.com/wiki/api