Lena Marvin

Business Problem

As a librarian at the University of Missouri-St. Louis (UMSL), my role involves enhancing access and discoverability of faculty scholarly works through the institutional repository. Each year, UMSL faculty members publish numerous articles in peer-reviewed journals. Unfortunately, many of these articles remain behind paywalls despite their potential for open access. Identifying which works can be made freely accessible is a highly time-consuming process.

To address this issue, I possess metadata pertaining to the articles published by the faculty. This metadata includes DOIs (Digital Object Identifiers), and there exists a tool (with a free API - https://openaccessbutton.org/api) that allows querying to determine whether an article can be liberated from a paywall and shared as an open access article. However, utilizing this tool on an individual basis is highly inefficient. With my progress in the LaunchCode program, I believe I am nearing the ability to use the API effectively. I possess a large CSV file containing comprehensive article metadata and have access to an API that can assist in identifying articles suitable for sharing. Automating this labor-intensive project is a key objective of mine, and I would greatly appreciate it if this could be considered as my final project for this class.

In addition to Open Access information this data provides opportunities for:

- Subject Area Analysis:
  - Social Sciences / Medicine / Psychology / Business
- Document Type Analysis:
  - Articles / Reviews / Book Chapters / Books / Data Papers
- Identifying Funding Sponsors:
  - National Science Foundation / National Institute of Medical Sciences / ect
  - *Visualization Idea: funding sponsors*
- Keyword Analysis:
  - *Visualization Idea: keyword analysis.*
- Analysis of Collaborating Universities:
  - *Visualization Idea: illustration of collaborations between UMSL faculty and faculty from other universities*
- Identifying Individuals Who Most Frequently Publish:
  - Determining which UMSL faculty member has published the most.

Data:

The dataset comprises 2,892 rows and 248 columns. Access it at the following GitHub link: https://github.com/e3la/EDAExercises/blob/master/umsl-scholarship.csv