Data Analytics Case Study

**About the Data**

The data shared herewith is a sample snapshot of research papers published by some publishers from 2023-2024 (2yrs). Each record is an individual research article with its constituent metadata like DOI (persistent identifier link), authors, title, date of publication, journal of publication etc. Some of the metadata fields like authors would be nested to contain detailed information about the same. The structure of an object (research article) is shared in the file “Individual Record Structure.txt”. In case of missing meta-data then that particular object won’t have the field name

**Your Objective**

To analyse the data and find insight/patterns from the sample data which could be of interest to a researcher, publisher, university (affiliation) or general public. Your analysis and conclusions could be submitted as a word document(.docx) or python notebook(. ipynb) with relevant chart/graphs that helped you uncover an insight

Some examples for your reference

1. <https://scholarlykitchen.sspnet.org/2022/11/08/guest-post-publishing-fast-and-slow-a-review-of-publishing-speed-in-the-last-decade/>
2. <https://www.linkedin.com/posts/mtclarke_where-are-all-the-papers-going-i-wrote-activity-7237517444964085760-Eq-h?utm_source=share&utm_medium=member_desktop>
3. <https://www.csescienceeditor.org/article/publishing-in-january-and-impact-on-citations-does-the-data-support-the-strategy/>

Some open questions for your reference

1. Which countries have the most collaboration while publishing research and how has their collaboration effected over the years
2. Publishing trends of a journal/publisher with respect to a country
3. YoY (year on year) publishing trend of a particular journal with respect to a subject area/concept

You are not limited by any of the above questions or examples and are free to choose any other question/insight that you might want to uncover. You are also free to use any additional open-source database to supplement your data analysis. Brownie points if you could support your findings with other articles/news events published on the internet that validate your findings

**General Guidelines**

1**.** Data and this README file can be downloaded from <https://drive.google.com/drive/folders/1RqphzTFi-bMnfJKHntnIdmT2mjuxDB5w?usp=sharing>

2. Please upload your submissions (.docx or . ipynb) in this [form](https://forms.gle/qeoedtQLhUqXmNsx8) by **10:00AM IST 27th January 2025**

3. You may submit multiple time within the submission window and only your latest submission would be considered for evaluation

4. In case of multiple files in a submission, please zip and upload them. **Do not upload the input files again,** only your submission files need to be uploaded