DE 1 Solution.docx

by Habeeb Omotunde

DATA EXPLORATION Project Part1

Submitted by

Habeeb Omotunde, 200584861

1. **Introduction**

   The data set of interest is the Scimago journal ranking for the year 2022 which was sourced from here.

   Curated from Scopus database, It captures data about journal distribution across countries and different fields of science including various metrics, such as H-Index, Scimago Journal Rank, References per document etc. These insights aid ranking of every journal over a period of time.

   The CSV file has a total of 27, 995 observations with 21 variables with details provided in the table (data dictionary) below. A full description of the data dictionary is provided in the attached excel spreadsheet.

#	Variable	Description	Type	Dependency
1	Rank	This variable represents the hierarchy of the journal relative to other journals in the observation depending on the SJR Value	Categorical, Ordinal.	Dependent Variable
2	Sourceid	This is a unique ID for identifying every entry in the dataset provided by Scimago	Numerical, Discrete	Independent Variable
3	Title	This is practically the unique name of the journal provided by the publisher. In many cases, the journal title lends a clear hint regarding the field and category of a journal	Categorical, Nominal.	Independent Variable
4	Type	The kind of publication accepted by the publisher. It could be a journal, conference proceedings or book series.	Categorical, Nominal.	Dependent Variable
5	ISSN	This is a unique value globally recognized as the standard means for identifying books, journals, magazines etc. around the world. It stands for International Standard serial Number	Categorical, Nominal.	Independent Variable
6	SJR	A measure of the scientific influence of scholarly journals based on the number of citations received by articles in the journal within a period of time.	Numerical, Continuous	Dependent Variable

7	SJR_Best_Quartile	This variable measures the grade, also known as quartile ranking, of every journal in the database. The quartile values ranges between Q1, Q2, Q3 & Q4 where Q1 is the highest and Q4 Is the lowest quartile.	Categorical, Nominal.	Dependent Variable
8	H_index	This variable measures the number of published work relative to the highest citation count of each publication of a scholar.	Numerical, Discrete	Dependent Variable
9	Total_Docs_2022	This is the total number of documents published as at 2022	Numerical, Discrete	Dependent Variable
10	Total_Docs_3years	This is the total number of documents published in the last 3 years	Numerical, Discrete	Dependent Variable
11	Total_Refs	This is the total number of references included in the journal's publications as at 2022	Numerical, Discrete	Dependent Variable
12	Total_Cites_3years	This variable measures the total number of citations in the past 3 years. i.e. citations in 2022 received by all published articles in 2019, 2020 & 2021	Numerical, Discrete	Dependent Variable
13	Citable_Docs_3years	This is the total number of the journal's citable documents in the last 3 years i.e. 2019, 2020 & 2021.	Numerical, Discrete	Dependent Variable
14	Cites_per_Doc_2years	This metric measures the average citation per document in a 2 year period.	Numerical, Continuous	Dependent Variable
15	Ref_per_Doc	Average number of references for each document published by the journal in 2022	Numerical, Continuous	Dependent Variable
16	Country	Location where journal is based	Categorical, Nominal.	Independent Variable
17	Region	Region where the country falls	Categorical, Nominal.	Independent Variable
18	Publisher	The organization or company responsible for publishing the journal.	Categorical, Nominal.	Independent Variable
19	Coverage	A range of years the journal has been active.	Date/Time	Dependent Variable
20	Categories	Field of Science the journal specializes in.	Categorical, Nominal.	Independent Variable
21	Areas	Classification of the journal into specific subject by category	Categorical, Nominal.	Dependent Variable

2. **Dataset Properties**
- **Number of Variables**: 21
- **Source**: Scimago journal & country rank: https://www.scimagojr.com/journalrank.php
- **Assumptions about dataset**: This dataset is believed to be authoritative as it includes current scientific journals developed from the information contained in the Scopus database managed by Elsevier Publishing company. However, a critical number of scholars are of the

opinion that the distribution of high impact journals visa viz mechanism of inclusion into this database is highly biased and money -driven. Hence, the observations are technically considered to be a sample of a larger unregistered population of high-quality journals.

By no means is this dataset exhaustive but it covers regions and countries interesting enough to derive insights and information about how research and developments is documented around the world and the consequences of poor research effort and documentation.

3. **Motivation:**
   - **Summary:** My interest is driven by the need to ascertain if the classification of journals into certain categories is solely based on merit and contribution to the body of knowledge or largely money driven. Furthermore, to understand the extent of influence of research outputs in improving the quality of service in the respective research areas of each country. Wrong Form

   - **Backstory:** During my postgraduate studies in Malaysia, there was immense pressure on all research students to meet a quota in publications to certain publishing houses like Elsevier, Springer, ACM, etc. The pressure was so intense lots of student had to quit while many of those who stayed, though succeeded in publishing few articles in requisite journals, were inundated with different medical and psychological issues.

   The cost of publishing wasn't cheap either. If students wanted timely release of their articles in open access journals, they had to bear the cost which in many cases get cascaded to the supervisor. As for closed journals, access to their repositories were secured behind pay walls for which universities had to subscribe and budget huge amounts for.

   This wasn't sustainable as the government began to cut funding for research given low return on investments, fewer patents, poor translation of research finding to service offerings & tight competition for meagre grants. As these events unfolded, scholars in Malysia and Europe began to expose the fraudulent extortion of researchers by publishing houses in return for recognition and access to their intellectual property. A direct response was the emergence of ResearchGate and Sci-Hub.

4. **Research Questions (FINER)**
   a. How does the distribution of research output vary across different scientific disciplines and areas by country or region within the Scimago database affect participation in research and development   Sp.   Confused
   b. Does the distribution of publishing houses across regions support or inhibit participation in high quality research and development by countries with low ranking?
   c. To what degree does a country's investment in research and development (R&D), research output influence what gets published where and when, if any?
   d. Can metrics such as citation per doc, reference per doc, total cites or a combination or these metrics influence investment in certain research areas or categories in any country or region.
      Missing "?"

e. Can metrics such as citation per doc, reference per doc, total cites or a combination or these metrics indicate potential journals across lower ranking countries for inclusion or upgrade to higher quality ranking in the Scimago database.

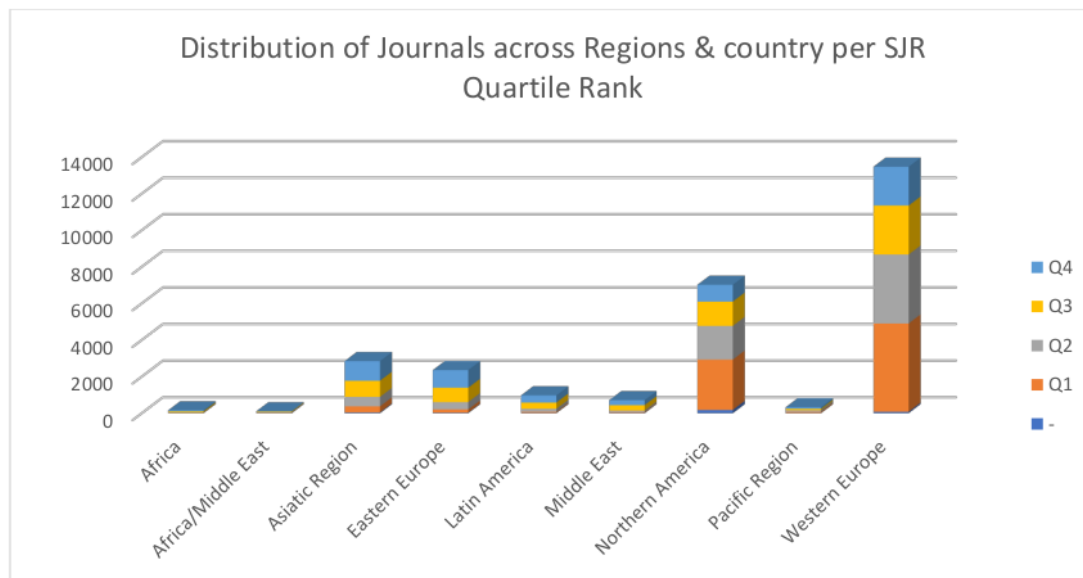5. **Additional Data Requirement**

The available data is presumed sufficient for the stated research questions. As the exploration continues, need for more data might ensue but this is not anticipated.
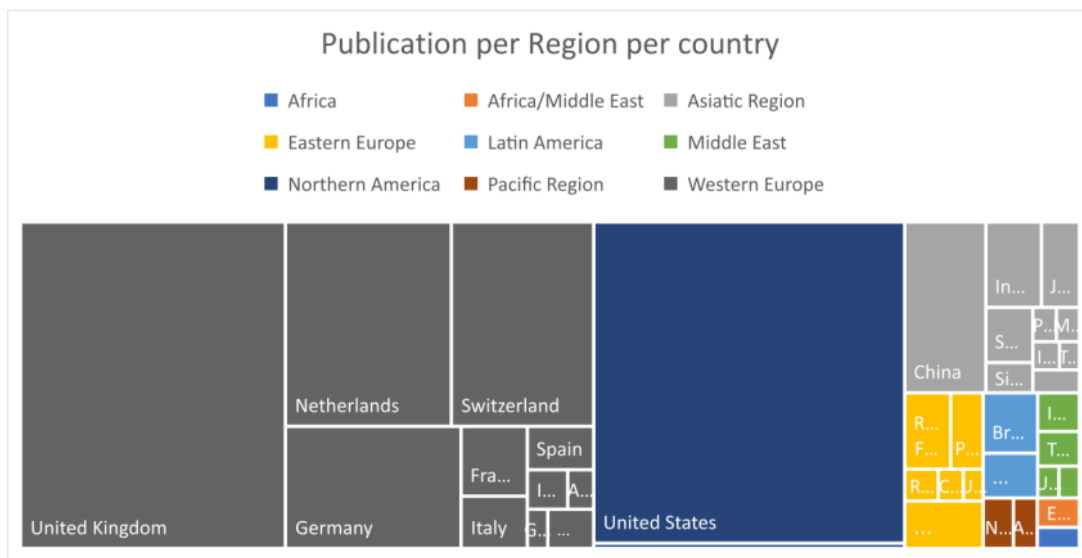
6. **Tracking Analysis.**

Distribution of Journals across Regions & country per SJR Quartile Rank

Legend: Q4, Q3, Q2, Q1, -

Regions: Africa, Africa/Middle East, Asiatic Region, Eastern Europe, Latin America, Middle East, Northern America, Pacific Region, Western Europe
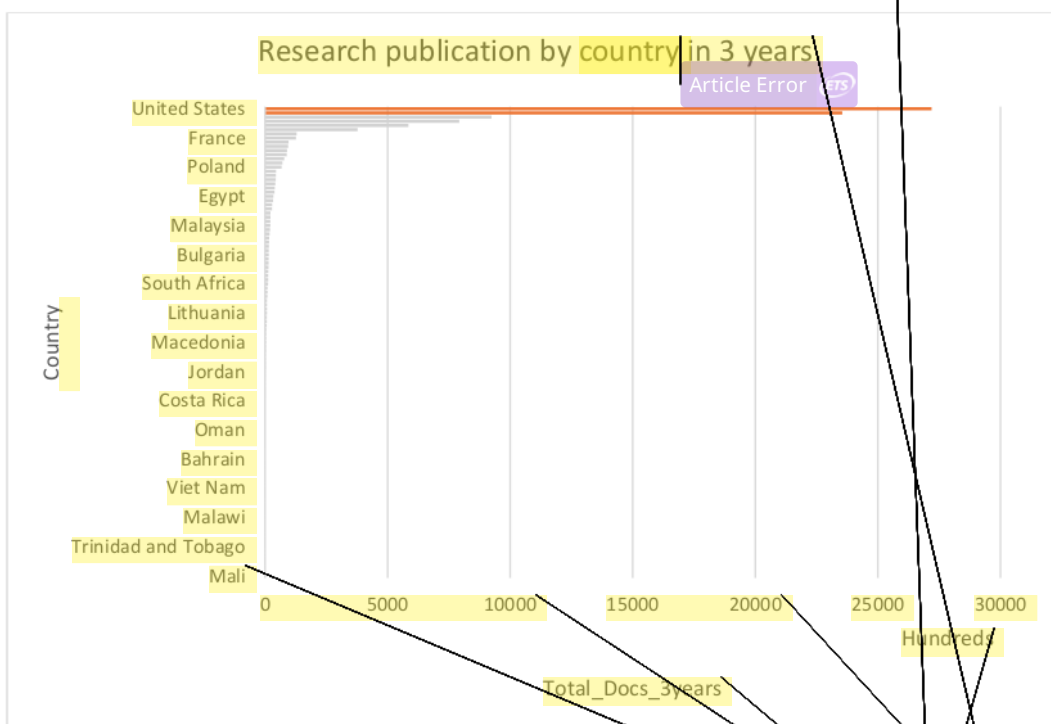
Making sense of journal distribution across the world

A slicer named region and country is provided on this pivot chart to understand the participation of each country in research & Development effort via journal publication. Kindly refer to Tracking and analysis 1 for details.

Publication per Region per country

- Africa
- Africa/Middle East
- Asiatic Region
- Eastern Europe
- Latin America
- Middle East
- Northern America
- Pacific Region
- Western Europe

United Kingdom · Netherlands · Switzerland · Germany · Italy · France · Spain · United States · China · Japan · India · Singapore · Russia · Brazil · etc.

Research contribution per region and country in a 3 year-window

Article Error

Research publication by country in 3 years

Article Error

Country (y-axis): United States, France, Poland, Egypt, Malaysia, Bulgaria, South Africa, Lithuania, Macedonia, Jordan, Costa Rica, Oman, Bahrain, Viet Nam, Malawi, Trinidad and Tobago, Mali

x-axis: 0, 5000, 10000, 15000, 20000, 25000, 30000

Hundreds

Total_Docs_3years

Ranking of countries by highest number of publications in 3 years

Confused   Article Error

In total, there are five tracking and analysis sheets provided in the attached spread sheet. Kindly refer to the spreadsheet for detailed analysis.

Garbled

DE 1 Solution.docx

7% SIMILARITY INDEX **2**% INTERNET SOURCES **7**% PUBLICATIONS **1**% STUDENT PAPERS

PRIMARY SOURCES

1   Roberto Todeschini, Alberto Baccini. "Handbook of Bibliometric Indicators", Wiley, 2016
    Publication
    4%

2   Gouri Ginde, Snehanshu Saha, Archana Mathur, Sukrit Venkatagiri, Sujith Vadakkepat, Anand Narasimhamurthy, B. S. Daya Sagar. "ScientoBASE: a framework and model for computing scholastic indicators of non-local influence of journals via native data acquisition algorithms", Scientometrics, 2016
    Publication
    1%

3   www.tandfonline.com
    Internet Source
    1%

4   cdn.shopify.com
    Internet Source
    1%

5   V. R. Renjith, I. Shihab. "h-index of geology journals: A statistical analysis based on scimago journal and country ranking", COLLNET Journal of Scientometrics and Information Management, 2018
    1%

Exclude quotes   Off   Exclude matches   Off
Exclude bibliography   Off

DE 1 Solution.docx

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Missing "," Review the rules for using punctuation marks.

Missing "," Review the rules for using punctuation marks.

Hyph. Review the rules for using punctuation marks.

Article Error You may need to use an article before this word. Consider using the article **the**.

Missing "," Review the rules for using punctuation marks.

Article Error You may need to use an article before this word.

Article Error You may need to use an article before this word.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

P/V You have used the passive voice in this sentence. You may want to revise it using the active voice.

Article Error You may need to remove this article.

P/V You have used the passive voice in this sentence. You may want to revise it using the active voice.

Wrong Form You may have used the wrong form of this word.

Missing ","

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Confused You have used either an imprecise word or an incorrect word.

Article Error You may need to use an article before this word.

Missing "?" Review the rules for using punctuation marks.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Missing "?" Review the rules for using punctuation marks.

Article Error You may need to remove this article.

Missing "," Review the rules for using punctuation marks.

Article Error You may need to remove this article.

Article Error You may need to use an article before this word.

Confused You have used either an imprecise word or an incorrect word.

Article Error You may need to use an article before this word. Consider using the article **the**.

Wrong Article You may have used the wrong article or pronoun. Proofread the sentence to make sure that the article or pronoun agrees with the word it describes.

Compound These two words should be a compound word.

Garbled This sentence contains several grammatical or spelling errors that make your meaning unclear. Proofread the sentence to identify and fix the mistakes.