**DATA EXPLORATION Project Part1**

**Submitted by**

**Habeeb Omotunde, 200584861**

The data set of interest is the Scimago journal ranking for the year 2022 which was sourced from [here](https://www.scimagojr.com/journalrank.php).

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 below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **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 |

Dataset properties

* Number of Variables: 21
* Variable classes
  + Categorical variables:
  + Numeric variables: