ROLE AND IMPACT OF THE SCIENTIST IN THE SOCCER ECONOMIC WORLD

Fabio Romero Manrique

J. Sebastián Hernández Aguilera

Andrés D. Pérez, MEng.

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE BOGOTÁ

AGENDA

1. Main & Specific Goals
2. Methodology
3. Dataset
4. Results
5. Conclusions

Main Goal

To show the interdisciplinarity importance of modeling problems in soccer and how it improves its business understanding and behavior.

Specific Goals

To show the relevance of the academic production as a measurement of how developed is the field and how it is impacting the economic world.

To determine if the scientific production around the soccer analysis has increased with the time.

To identify the relation between the article-citation impact and several variables such as the extension of the title and the content of the abstract, among others.

Methodology

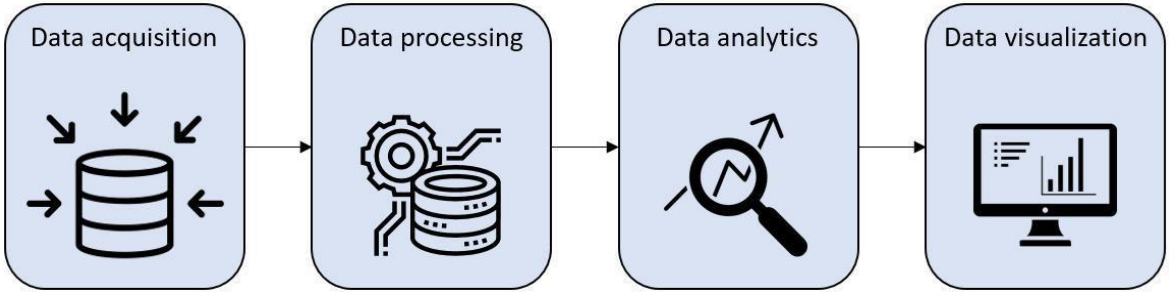

Figure 1. Overall workflow for the project methodology. Contains four stages from left to right which are: Data acquisition, processing, analytics and visualization.

Dataset

The dataset consist in an unstratified sample of 102 registers of scientific production metadata. This dataset contains variables such as the research query (with query terms such as: "soccer", "analytics", "world" and “Machine learning SCIENTIST IN THE SOCCER”, among others) the title, the abstract, the event where the register is presented and the names or the authors (splitted one author per cell).

The data temporal range goes from year 2013 to 2021 by three reviewers.

Results

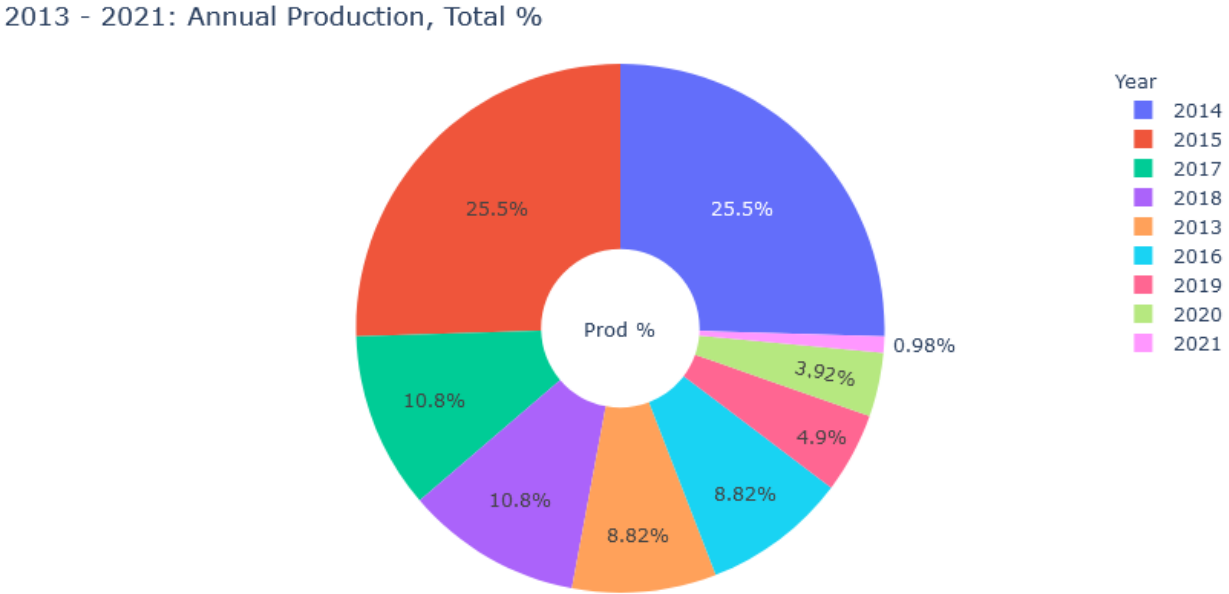

Figure 2. 2013 - 2021 Annual production review at percentage level on the acquired dataset

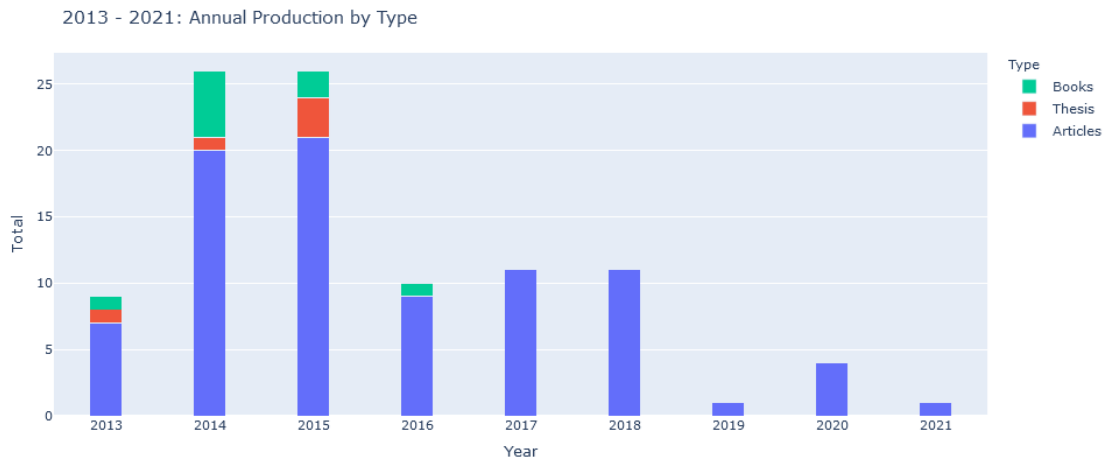

Figure 3. 2013 - 2021 Annual production review desegregated by type of production (books, theses and articles)

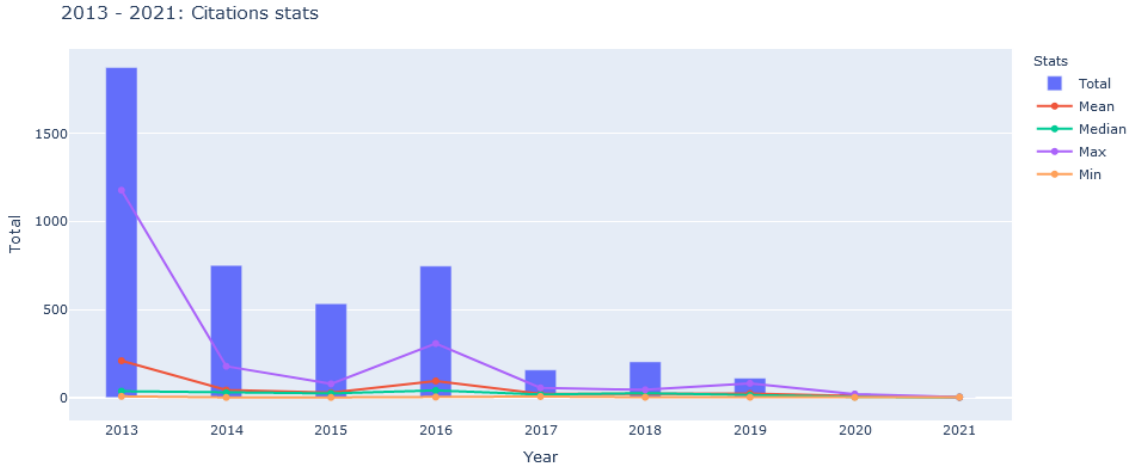

Figure 4. 2013 - 2021 Citation review for annual output along with indicators. Total - Total citations for that year. Max - Maximum number of citations for that year. Min - Minimum number of citations for that year. Mean - Average number of citations for that year. Median - Median number of citations for that year

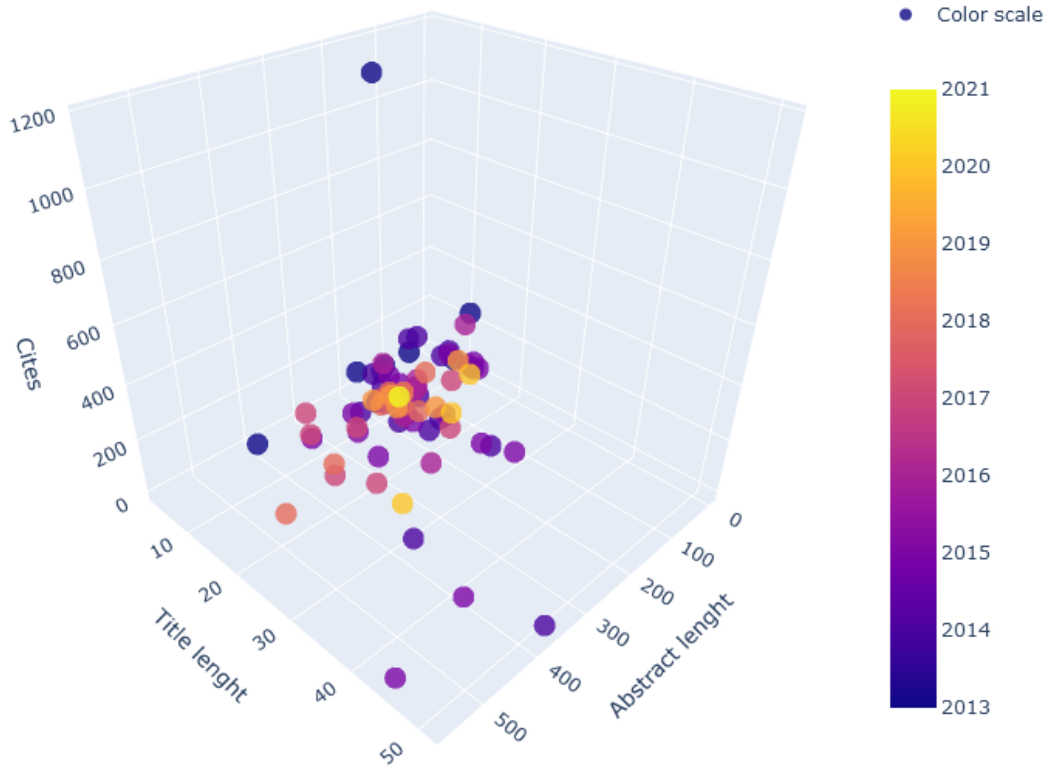

Figure 5. 2013 - 2021 Review of the relationship and distribution between total citations, text length and abstract length for each of the years

Figure 6. Word cloud for the title terms

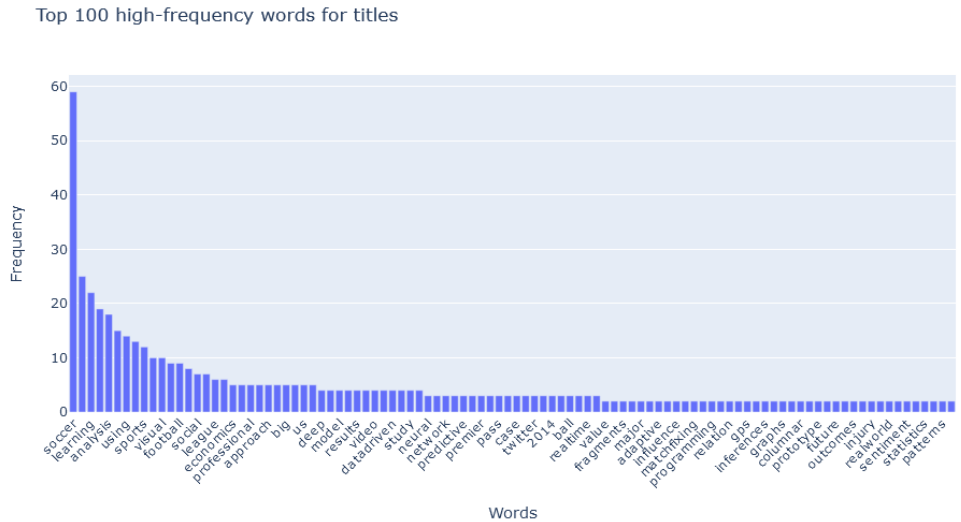

Figure 7. Review of the top 100 most frequent terms in the title word cloud

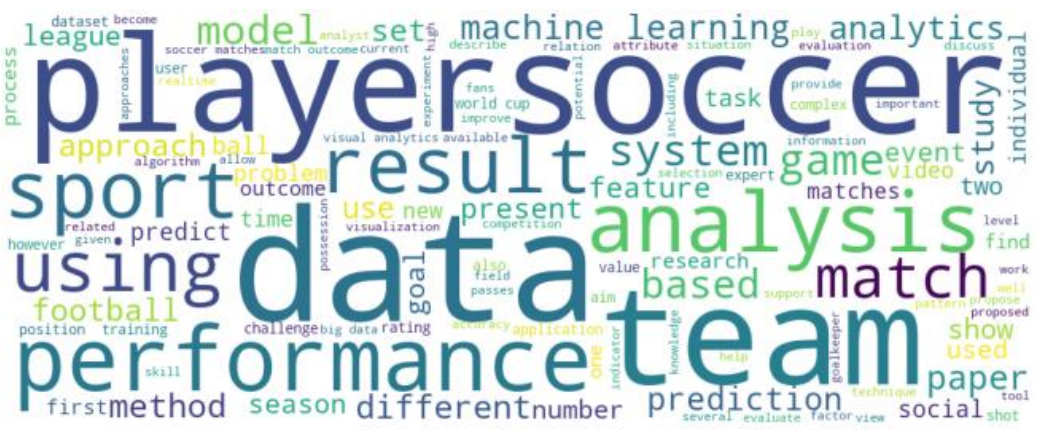

Figure 8. Word cloud for the abstract terms

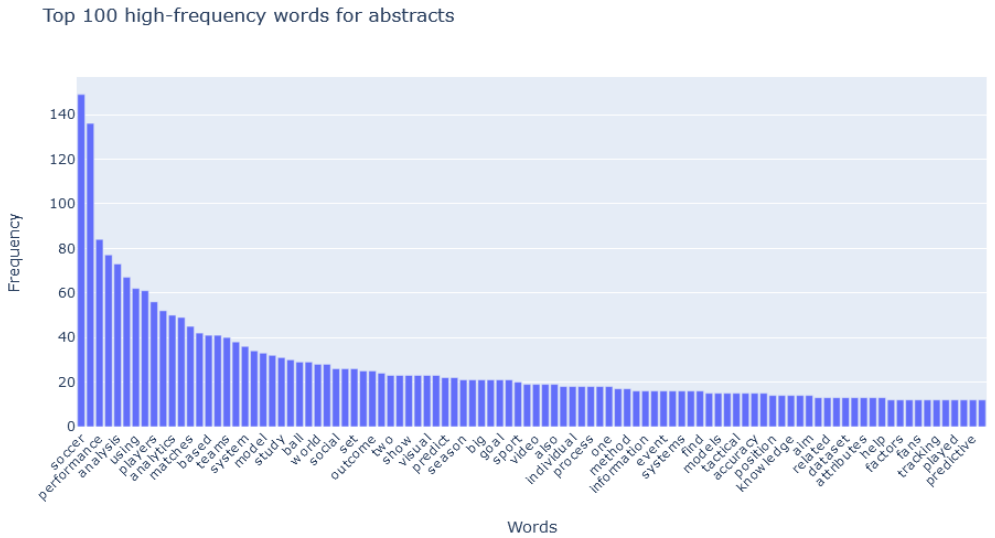

Figure 9. Review of the top 100 most frequent terms in the abstract word cloud

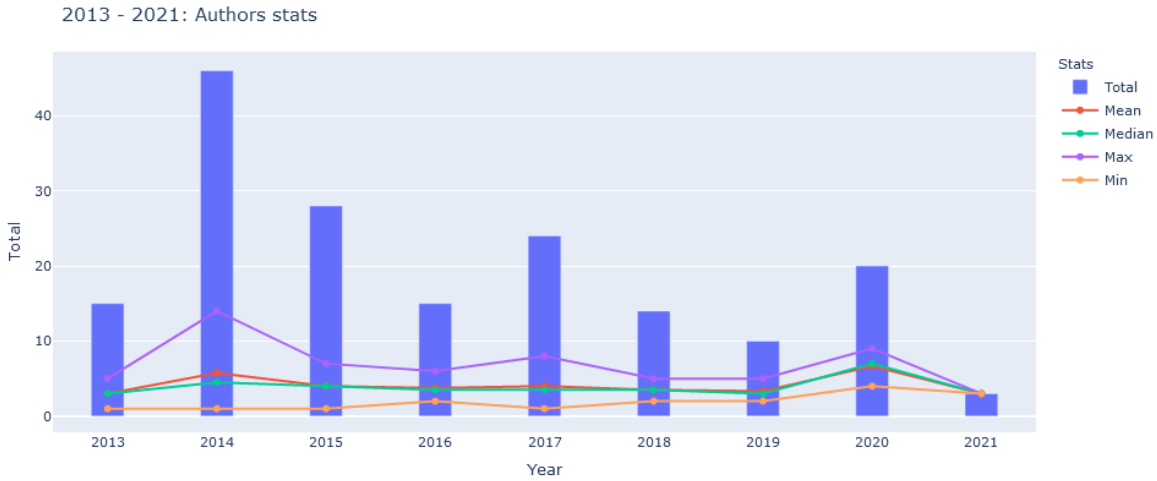

Figure 10. Revision of the number of authors for the annual production together with the indicators. Total - Total number of authors for that year. Max - Maximum number of authors for that year. Min - Minimum number of authors for that year. Mean - Average number of authors for that year. Median - Average number of authors for that year

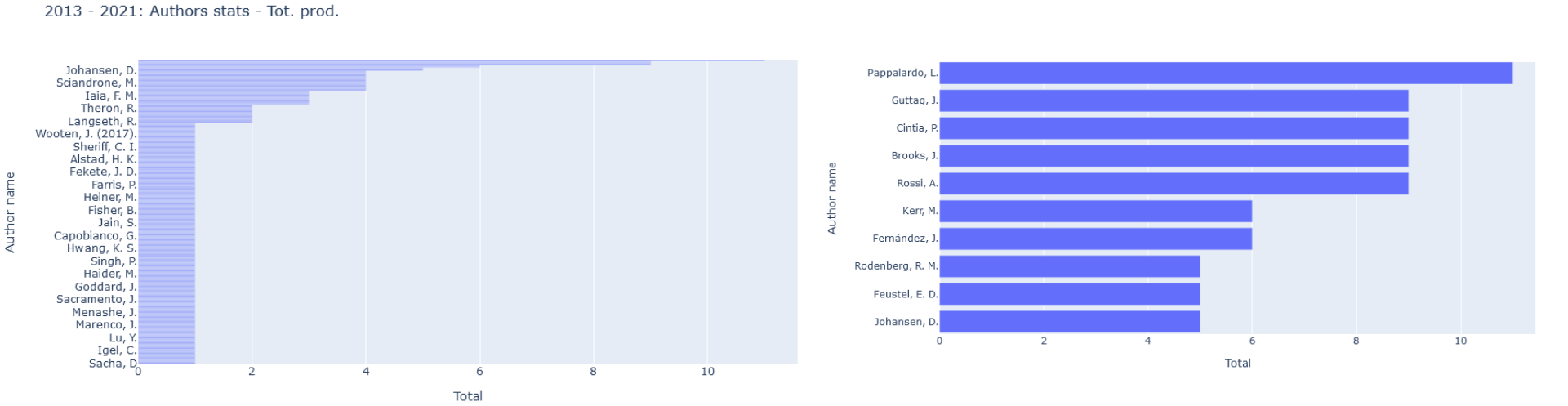

Figure 11. Review of the number of products of each author in the database. Left - Overview. Right - View of authors with 5 or more publications

Conclusions

- The current database shows a significant decrease in scientific production over the last 3 years, which could be related to the current world situation.
- The years of greatest academic production were 2014 and 2015, which shows a particular global interest in this sector with considerable capital injection.
- There is evidence of a decrease in the number of citations of the articles in the database possibly related to the low scientific production and the production's aging.
- It is evident that the length of the title or abstract does not have a considerable impact on the number of citations of an article. However, authors mostly prefer abstracts and short titles.
- Based on the word clouds, it is evident that the titles and their abstracts are largely congruent. On the other hand, the prevalence of terms such as analytics, data, performance and analysis, among others, shows the high impact of scientists in this field.
- There is a direct correlation between the number of authors and the number of publications. However, it could be inferred that quantity does not represent quality and/or interest, when considering the number of authors vs. the number of citations obtained.