Analysis of Academic Research in Turkey

1st Yiğit Kartal
Electrical Engineering MSc.
Özyeğin University
Istanbul, TURKEY
yigit.kartal@ozu.edu.tr

2nd Ceyhun Topal
Computer Science BSc.
Özyeğin University
Istanbul, TURKEY
ceyhun.topal@ozu.edu.tr

***Abstract*—This study aims to comprehensively analyze academic research performance in Turkey by utilizing data science methodologies. Using data collected from the Web of Science database and the official Turkish Higher Education Institution (YÖK), the study presents an extensive complied dataset encompassing academic publications and corresponding academic profiles. The dataset includes key attributes such as publication performance metrics, university affiliations, and personal academic identifiers. To provide a detailed understanding of academic output, statistical analyses were performed to reveal trends in specific topics such as the publication performance by year, number of publications by cities in Turkey, gender-based distribution of publications, university rankings, and publication language preferences. Visualization techniques employing Python's matplotlib library were used to present the statistical results. This paper outlines the methodology and preparatory analyses and lays the groundwork for discussions on the implications of the observed trends in Turkey's academic landscape.**

***Keywords*—academia, data analysis, temporal trends, university rankings.**

I. Introduction

Academic research serves as one of the cornerstones of societal and technological advancement, playing a critical role in shaping economies, driving innovation, and fostering international collaborations. Understanding the dynamics of academic performance at the national level provides valuable insight into the state of research and development (R&D) within a country. Turkey, with its rapidly growing higher education, represents a unique case study to examine academic output and trends over time.

Several studies can be found in recent years that have analyzed the research performance of Turkish higher education institutions. For example, a comprehensive analysis [1] covering the period from 1980 to 2022 provides information on the evolution of research productivity in Turkey's higher education sector. In the reference study, the University Ranking by Academic Performance (URAP) was used to assess and compare the performance of Turkish universities.

Gender disparities in academia have also been the subject of investigation. Research indicates that, unlike persistent trends in Western countries, Turkey exhibits higher proportions of women participating in scientific research and teaching. According to She Figures (2012), 28% of all Grade-A professors in Turkey are female, whereas women only constitute 20% of this category in the EU-27.

This study focuses on evaluating academic research performance in Turkey by analyzing and comparing comprehensive datasets from two key sources: the Web of Science (WoS) database and the Turkish Higher Education Institution (YÖK). The WoS database is a globally recognized repository of scholarly publications, while YÖK provides authoritative information on Turkish academics and their institutional affiliations. By cross-referencing these datasets, we aim to evaluate various aspects of academic productivity, such as temporal trends in research output, gender-based disparities, institutional rankings, and language preferences in scholarly publishing.

Despite the availability of extensive data, several challenges arise when attempting to integrate and analyze such diverse datasets. One notable limitation in the academic profiles dataset from YÖK is the absence of explicit gender identifiers, which are crucial for assessing gender-based trends. To address this, we implemented an SQL-based name-gender classification model, enabling us to assign likely gender identities to academic names with reasonable accuracy. Additionally, the diverse formats and scopes of the data necessitated robust data cleaning and preprocessing methodologies.

The significance of this research lies not only in its immediate findings but also in its potential to inform policy decisions, promote equity, and optimize resource allocation within the Turkish higher education system. By employing data science techniques to analyze and visualize this wealth of information, we aim to provide actionable insights for policymakers, educational administrators, and academic institutions.

II. Methodology

A. Data Collection

The primary data sources for this study were the Web of Science (WoS) database and the YÖK (Turkish Higher Education Institution) website. From WoS, we extracted detailed information on all academic publications in Turkey, including publication titles, author names, publication years and dates, associated institutions, and performance criteria such as the citation count and WoS score.

Concurrently, the YÖK dataset provided a comprehensive list of academics in Turkey, along with their names, surnames, and affiliated universities. Extracting publication data was handled manually from the WoS website and the chunks of data were subsequently merged into one big dataset. For academics data, we employed an algorithm that scrapped the YÖK website for all academics affiliated with each university, one by one, using the Selenium library for Python.

B. Data Integration

To enable meaningful analyses, it was essential to integrate the WoS and YÖK datasets. Cross-matching was conducted using Python, leveraging author names and institutional affiliations as primary matching keys. This process involved:

1) **Standardization:** Ensuring uniformity in data fields such as names and university names by applying string normalization techniques, including lowercasing, removal of special characters, conforming academic names to "Surname, Name" format and handling variations in naming conventions.

2) **Matching Algorithms:** Employing matching techniques using Python's rapidfuzz library to account for minor discrepancies in name spellings. Matching algorithms provided similarity scores between names as strings and allowed us to match names with a similarity higher than a certain threshold.

3) **Validation:** Verifying the accuracy of matches through manual spot-checking and consistency checks for high-confidence matches.

C. Gender Classification

Given the absence of explicit gender information in the YÖK dataset, a name-based gender classification model was implemented. Using an SQL database containing name-gender mappings derived from historical and cultural records, we assigned gender labels to the academic names. The model was calibrated to minimize classification errors, with a focus on ambiguous or gender-neutral names.

D. Statistical Analyses

To derive insights into academic performance trends, various statistical metrics were computed:

- **Temporal Trends:** Annual performance scores were calculated by aggregating publication metrics over time.
- **Gender-Based Distributions:** Publication counts and performance scores were analyzed by gender to identify disparities.
- **Institutional Rankings:** Universities were ranked based on aggregated performance scores and most academics, highlighting the top 20 institutions.
- **Faculty Ranking:** Faculties were ranked based on the number of publications, ratio of the publications by female academics etc., highlighting the top 20 faculties.
- **Nature of Publications:** The distribution of publication languages and types was examined to assess preferences and trends.

E. Visualization

Python's Matplotlib library was employed to create informative visualizations, including bar charts, line graphs, scatter plots and pie charts, to represent the findings. Particular emphasis was placed on clarity and interpretability to ensure the visuals effectively communicate the results to diverse stakeholders.

The results and discussion sections will provide a detailed interpretation of these analyses, supported by comprehensive visualizations and critical commentary on the implications of the findings for Turkey's academic landscape.

III. Results

The statistical analyses mentioned in the previous sections will herein be represented using various visualization methods such as bar charts, tables, scatter plots, and pie charts.

IV. Results & Discussion

A. Gender-Based Publication Trends

Figure 1 highlights the top 20 faculties with the highest ratio of publications by female academics. The Nursing Faculty leads with a ratio exceeding 0.8, followed by Health Sciences and Dentistry faculties. This indicates significant representation and contribution of female academics in these disciplines. Conversely, faculties with traditionally lower female participation, such as Engineering, are ranked lower in the ratio distribution.

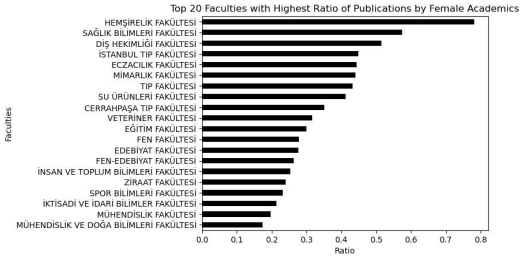

Figure 1: Top 20 faculties with the highest ratio of publications by female academics.

Figure 2 reveals the gender distribution among authors, highlighting the nearly balanced contributions of male (46.5%) and female (37.4%) academics, with a small proportion (16.2%) representing joint authorships or unspecified genders.

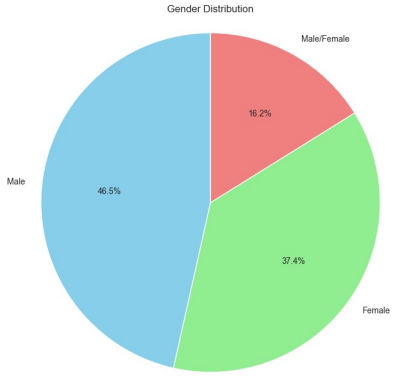

Figure 2: Gender distribution of academic authors.

Figure 3 illustrates the average yearly performance scores of male and female academics in Turkey. The graph highlights several key trends:

- Between 2000 and 2005, both male and female performance scores showed considerable variability, reflecting fluctuations in academic output during this period. Notably, there were instances of gender-based discrepancies in individual years, but the general trend converged by 2005.
- From 2005 onwards, the performance scores for both genders exhibit a steady upward trajectory, signaling consistent improvements in academic productivity. This increase aligns with broader developments in Turkey's higher education sector, such as the expansion of universities and increased funding for research.
- By 2025, the performance gap between genders appears to have minimized, with male and female academics achieving nearly identical average scores. This convergence indicates a progressive narrowing of gender disparities in academic output.
- The overall trends also suggest that female academics have maintained parity with their male counterparts, particularly in the last decade. This observation underscores the increasing contributions of women in the Turkish academic landscape.

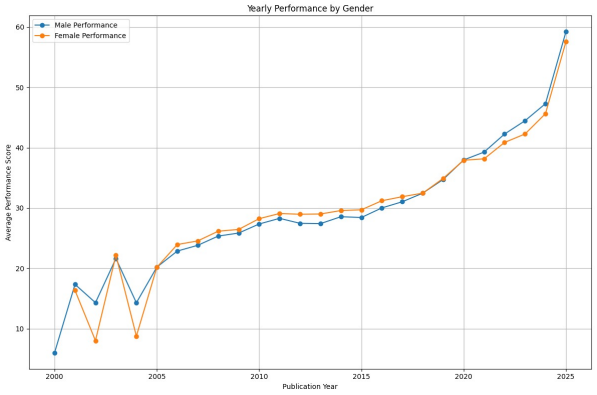

Figure 3: Yearly Performance by Gender: Average performance scores of male and female academics in Turkey from 2000 to 2025.

The trends revealed in Figure 3 suggest an encouraging evolution of gender equity in Turkish academia. While initial years showcased noticeable variability, the alignment in recent years reflects the impact of systemic efforts to promote inclusivity and equality within academic institutions. Further analysis could involve investigating the drivers behind these trends, such as policy interventions, changes in societal attitudes, or targeted programs to enhance female representation in research.

B. Faculty-Wise Publication Volume

The distribution of publication volume across faculties is presented in Figure 4. The Medical Faculty ranks highest with a significant margin, followed by Engineering and Science Faculties. These results underline the dominant contribution of medical and technical disciplines to overall academic output in Turkey.

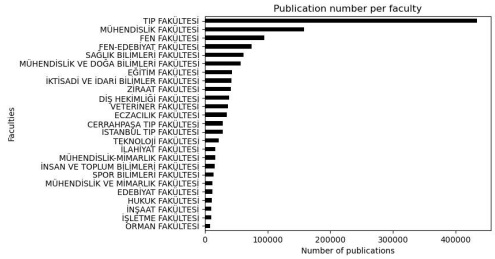

Figure 4: Total publication number per faculty.

C. University Performance and Academic Output

Figure 5 demonstrates the relationship between the average performance score and the number of articles per academic across universities. Notable outliers, such as Koç University, Sabancı University, and Bilkent University, exhibit both high productivity and performance scores, indicating a strong emphasis on research quality and output.

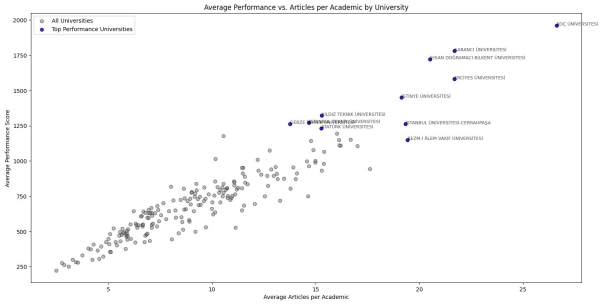

Figure 5: Average performance score vs. average articles per academic by university.

D. Yearly Performance Trends by Institution Type

The comparison of yearly performance trends between state (Devlet) and foundation (Vakıf) universities is illustrated in Figure 6. State universities consistently outperform foundation universities, although the gap narrows in recent years, reflecting increased competitiveness and investment in foundation universities.

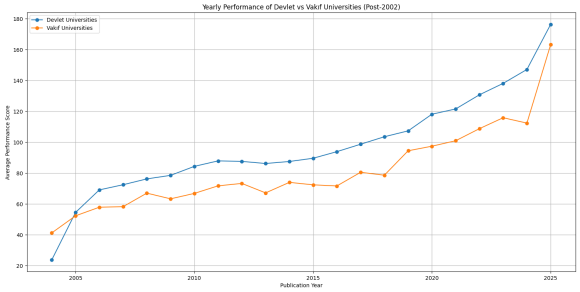

Figure 6: Yearly performance trends of state vs. foundation universities.

E. Publication Characteristics

1) Document Types: The distribution of document types in the Web of Science is shown in Figure 7. Articles dominate, followed by meeting abstracts and reviews, indicating a preference for traditional publishing formats within Turkish academia.

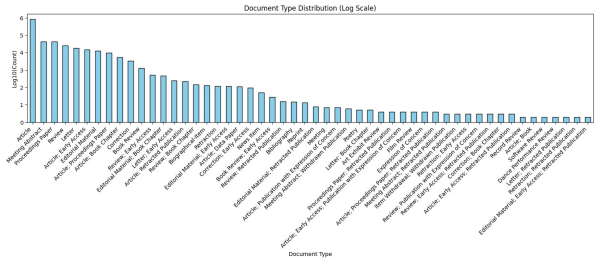

Figure 7: Document type distribution in Web of Science.

2) Publication Volume Over Time: The yearly distribution of publications, depicted in Figure 8, shows a significant increase from 1980 to the present. This growth correlates with the expansion of the Turkish higher education system and increased research funding.

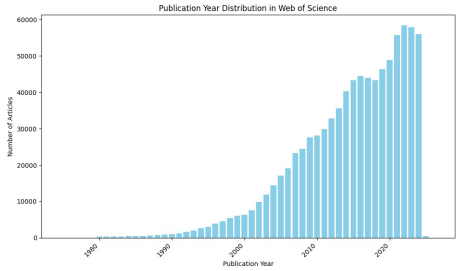

Figure 8: Publication year distribution in Web of Science.

F. Institutions with the Most Academics

The top 20 universities with the largest academic staff are displayed in Figure 9. Health Sciences University leads, followed by Ankara and Istanbul Universities, indicating the scale and diversity of these institutions.

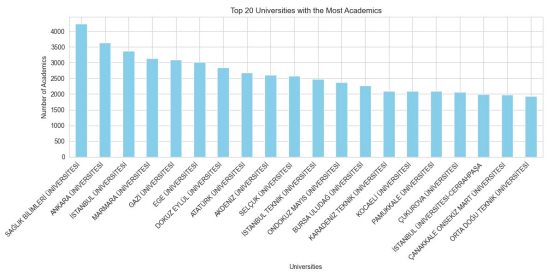

Figure 9: Top 20 universities with the most academics.

G. Regional and Language Trends

1) University Distribution by City: Figure 10 highlights Istanbul as the leading city in terms of the number of universities, followed by Ankara and Izmir. This distribution reflects the concentration of academic institutions in metropolitan areas.

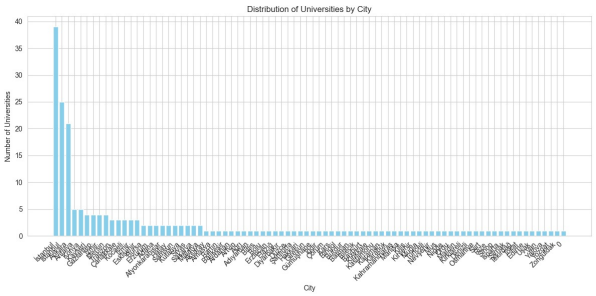

Figure 10: Distribution of universities by city.

2) Language of Publications: English dominates as the primary language of publications, as illustrated in Figure 11, followed by Turkish and German. This trend aligns with global publishing practices, emphasizing the importance of English as the lingua franca of academia.

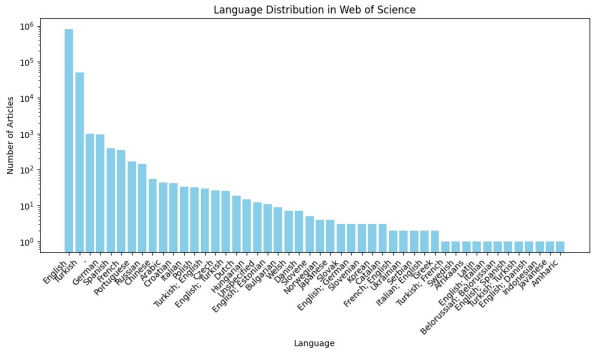

Figure 11: Language distribution in Web of Science publications.

The results presented in this study provide an in-depth overview of the academic research landscape in Turkey, offering valuable insights into publication trends, gender representation, institutional performance, and regional and linguistic patterns. Several key observations emerge from the analysis:

- **Gender Trends:** The significant contribution of female academics, particularly in faculties such as Nursing and Health Sciences, highlights the importance of gender equity in academic research. The narrowing performance gap between male and female academics over time underscores progress in addressing gender disparities. However, some fields, such as engineering, still show underrepresentation of female academics, indicating areas for further improvement.
- **Institutional Performance:** The dominance of state universities in terms of academic output and performance scores reflects their established infrastructure and resources. Foundation universities, while showing significant growth, still lag behind in overall performance. Targeted policies to support research in

foundation universities could further enhance their competitiveness.

- **Publication Characteristics:** The preference for publishing in English aligns with global academic practices, facilitating international collaboration and dissemination. The dominance of journal articles as the primary publication type suggests a focus on traditional academic dissemination methods, which could be complemented by increased emphasis on diverse formats, such as conference papers and book chapters, to foster broader outreach.
- **Regional and Disciplinary Distribution:** The concentration of universities and publications in major cities such as Istanbul and Ankara reflects the centralization of academic resources. Expanding support for institutions in less-represented regions could foster more balanced academic development nationwide.
- **Performance Outliers:** Institutions like Koç University, Sabancı University, and Bilkent University demonstrate exceptional performance, likely attributable to robust funding and research-oriented policies. Lessons from these universities can inform strategies for improving research output across other institutions.

V. Conclusion

This research offers a comprehensive analysis of the academic research landscape in Turkey, integrating data from the Web of Science (WoS) and YÖK databases to examine trends in publication output, gender representation, institutional performance, and other critical dimensions. The findings emphasize the growing contributions of female academics, the pivotal role of state universities, and the dominance of English as the preferred publication language.

References

[1] Maral, M. Research performance of higher education institutions in Türkiye: 1980–2022. Scientometrics 129, 4771–4793 (2024). https://doi.org/10.1007/s11192-024-05097-x

[2] European Commission: Directorate-General for Research and Innovation. (2013). She figures 2012 : gender in research and innovation : statistics and indicators. Publications Office. https://data.europa.eu/doi/10.2777/38520.

[3] Web of Science. "Academic Publications Database." Available: https://www.webofscience.com/

[4] YÖK (Turkish Higher Education Institution). "Database on Turkish Universities." Available: https://yokatlas.yok.gov.tr/universite.php