International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0313
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 2167 – 2175

# Trend Analysis And Demand Patterns In Public Sector Educator Recruitment: An Exploratory Data Analysis

*Ms. Sakthi P[1], Rilah Z[2], Swetha K[3]*
*[1]Assistant Professor, Department of Information Technology, B.S.A. Crescent Institute of Science & Technology, Chennai, India.*
*[2,3]UG – Department of Information Technology, B.S.A. Crescent Institute of Science & Technology, Chennai, India.*
***Email ID:*** *sakthi@crescent.education[1], rilahzaheer@gmail.com[2], swethzz06@gmail.com[3]*

***Abstract***
*Recruitment of educators in the government is a factor that promotes the sufficient staffing in educational facilities. Nevertheless, information related to recruitment is usually spread out in numerous government notifications and recruitment boards where it is hard to make a systematic and analytical analysis of the recruitment patterns. This unstructured data does not allow the clear comprehension of the change in demand in various years, subjects, and posts. The aim of the research is to examine the demand in public sector educator recruitment by gathering information on the official sources accessible to the public and to prepare it in such a way that the information is gathered into one structured dataset. Simple data cleaning methods are implemented and then vacancy analysis per year is performed to determine change in recruitment over the years. Descriptive statistics and visualization are used to analyze subject-wise and post-wise trends to demonstrate the differences between categories. In comparison with the current literature, which mostly examines trends in research based on the bibliometric analysis, this work uses trend analysis to real time recruitment data, providing a clear understanding of historical recruitment patterns.*
***Keywords:*** *Public sector educator recruitment, recruitment trend analysis, demand pattern analysis, vacancy analysis, data visualization.*

## 1. Introduction

### 1.1. Motivation

The Tamil Nadu Teachers Recruitment Board (TRB) has been in the forefront of filling in teaching positions in the public education system in the state including school level and college level, in the state system. The official website releases the notifications of recruitments periodically in PDF format in the form of documents. These reports usually comprise tabular reports with information like the number of vacancies per subject and number of vacancies per category as per reservation policies. This information is published, however, this information is not structured in such a manner as to be easily analyzed. Individual notifications are available as standalone documents and there is no combined set of information spanning several years. Consequently, this makes it challenging to study trends in recruitment over a time period or make any significant conclusions. Simple questions like, which are the subject areas that are always in high demand or how the recruitment process differs in different posts or did the external forces such as the COVID-19 pandemic affect the hiring pattern can not directly be answered using the raw data. This restriction has impacts on various stakeholders. Researchers cannot examine structured data sets on how demand trends on each and every subject and category. Thus, there is an evident necessity to transform these heterogeneous recruitment messages into a coherent and analyzable form. The present study is driven by the aim to convert the unorganized TRB data to a level of structuring of the data to facilitate the analysis and visualization process and make informed decisions.

### 1.2. Problem Statement

The technical essence of this project is to overcome the void between the overnight heterogeneous government PDF files and an organized analytical dataset. In particular, the study provides solutions to

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0313
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 2167 – 2175

five main issues: (a) automating the structured subject-wise extraction of vacancy information as PDFs in various formats (some being digital-native, others being scanned, some using a horizontal table, others a vertical column head) and (b) standardizing and cleaning the results in 10 notification types distributed over nine years; (c) the multi-dimensional statistical trend analysis There are three technical problems that can be distinguished: First, the optimum available Python library to extract PDF, pdfplumber, is not able to read the vertical column header of the target notification tables (PG and BT) as it renders out reversed sequences of characters Second, the number of granular subject names in use in 10 notifications (over 170) will need to be mapped to 25 standard subject categories across years to enable comparison. Third, the pipeline output (thon-generated data) needs to be read by a client-only HTML-based dashboard with no server-side API, which needs a JSON bridge file as the communication medium.

### 1.3. Contributions

The following six specific contributions, appeared in this work, concern the fields of educational data mining and processing government documents: (1) An 8-steps CLI pipeline, fully automated and, in less than 60 seconds, transforms raw TRB PDFs, end-to-end, into a structured dataset, charts, and dashboard to interactively depict the data. (2) 10 parsers specific to the format, an 80+ entry reversed-text correction dictionary and a 170+ entry subject standardization mapping. (3) Statistical analysis engine with six dimensions where current trends across years, subject demand trends, post-type distribution, region level aggregation, descriptive statistics, and growth statistics on the basis of the CAGR concept. (4) A web scraper that discovers new TRB notifications, filters them, and downloads them, directly to trb.tn.gov.in bypassing the SSL and politely taking time out between crawlers. (5) Chart.js interactive dashboard: 10+ chart types, year/subject/post, no need for a backend server, Python generated atz, Jerky - 103 records, 44,136 records in total, 2016-2025 based, 2016 to 2025 based: a public dataset, all in type castable form.

## 2. Related Work

Recent research on recruitment analytics has delved into a variety of methods such as bibliometric analysis, artificial intelligence, machine learning and exploratory data analysis. But these two methods vary much in regard to the sources of data, methods and viability into the actual world. Recruitment research has mainly been traced through bibliometric analysis as it is a method that has been widely adopted in the evolution of recruitment research. The research in [1] explores how artificial intelligence can aid talent acquisition and uses a huge number of research papers to outline the main trends in the process, including automation, bias mitigation, and decisions affecting people made by the AI. Likewise, [3] has a visual representation of a literature review of teacher recruitment studies, displaying themes like teacher shortages, policy frameworks, and professional development. Although these studies give a fairly detailed picture of the current trends in research, these are based solely on secondary information and do not reflect the actual trend in recruitment and current trends in demand. The base paper is based on the same approach, as it analyzes pre-existing published research articles through the methods of bibliometrics to pinpoint the trends in AI-based recruitment. Even though this offers some important information on research evolution, it is still limited to analysis based on literature and fails to answer the dynamics involved in recruitment in the real world. STEM teacher recruitment and retention strategies are systematically reviewed in [2], where the authors discuss other strategies and respect that some of the most popular strategies like financial incentives are not very effective. The study is important in terms of the policy level information but fails to be empirically validated in terms of actual recruitment data and fails to offer a data-driven interpretation of the trends in recruitment. Recent innovation in analytics has brought in AI-based exploratory techniques of data analysis. The article in [4] suggests an automated EDA system that combines machine learning and natural language processing to hastens data comprehension and enhances decision-making. Nevertheless, the methodology is not

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0313
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 2167 – 2175

domain-specific to recruitment analytics and thus is not applicable in studying domain-specific hiring patterns. A number of studies have concentrated on analyses of recruitment patterns based on both statistical and data mining. The analysis of [5] uses clustering techniques to determine the hiring patterns of educational institutions and points out the regional differences in hiring behavior. In the same manner, predictive modelling techniques like SARIMA-neural network model of [6] seek to predict the demand of the teachers and enhance the workforce planning. Although these approaches are useful in providing predictive data, the major concern with them is that they predict and cluster, they may not be interpretable or provide detail on describing historical employments trends. In [7], [8], and [9], workforce analytics provide research that highlights the necessity to base decision-making on data in the framework of organizational planning. These pieces of writing show that analytics can improve the efficiency of the workforce and strategic planning. Nonetheless, they do not pay only attention to the identification of patterns in actual recruitment data. Moreover, the research in [10] compares the trends in jobs and skills with advanced data analytics tools and traces the changes in the demand and the education-employment mismatch, however, it does not directly study the dynamics of teacher recruitment. Based on the foregoing discussion, it is apparent that research on research trend analysis and predictive modelling forms the bulk of the available research, and little has been done to analyze real world recruitment data. All methods are either based on secondary data sources, or are more oriented to predicting than explaining, so an important gap is present in the knowledge of real recruitment trends across time, subjects and regions.

To specifically overcome this limitation, the proposed work applies a data driven and application focus approach whereby it uses real world Teacher Recruitment Board (TRB) data gathered by web scraping and PDF extraction. This study uses pure exploratory data analysis (EDA) to study the patterns of recruitments, such as variations per year or per subject, regionally, or in categories, which is contrary to the situation in the earlier studies. This method reveals practical, interpretable, and policy-relevant information out of the actual recruitment data and thus fill the gap between theory and actual recruitment analytics.

## 3. System Architecture

### 3.1. Pipeline Overview

The system can be structured as a five-step sequential pipeline that is managed by main.py; the Command Line Interface (CLI) is built on the basis of argparse. All stages are enacted as a single (or a combination of) Python module(s), with well defined inputs and outputs. The pipeline has a very tightness of separation of concerns: visualization is not directly called by analysis logic; data acquisition is not directly called by any code in extraction. This modularity makes separate stages testable, replaceable or extendable. In Figure 1, the entire system architecture with data flow and module responsibility is provided with all the five stages.

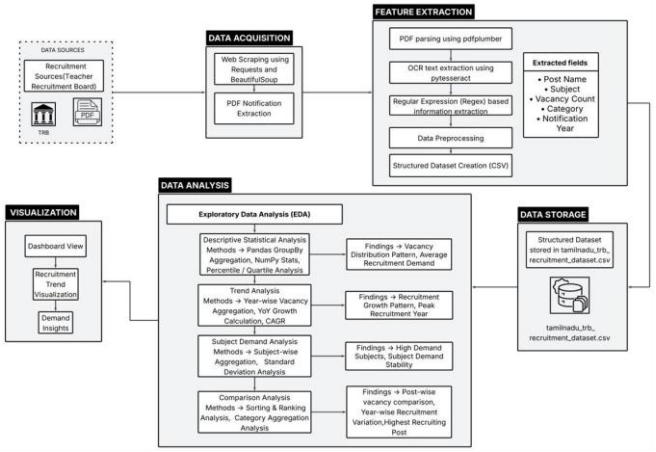


**Figure 1 TRB Educator Recruitment Analysis System Architecture**

### 3.2. Technology Stack

Table 1 shows the entire pipeline of the technology stack by stage. The choice of each of the technologies was made on the basis of maturity, community support, and appropriateness to the particular task of processing. In addition, compatibility between the selected tools was considered to ensure smooth data flow across all stages. The use of widely adopted libraries also improves system reliability and supports easier maintenance and future enhancements[11].

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0313
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 2167 – 2175

**Table 1 Technology Stack by Pipeline Stage**

| Layer | Technology | Version |
|---|---|---|
| PDF Extraction | pdfplumber | >= 0.10.0 |
| OCR Fallback | pytesseract + Pillow | >= 0.3.10 |
| Web Scraping | requests + bs4 + selenium | >= 2.31.0 |
| Data Processing | pandas + numpy | >= 2.0.0 |
| NLP Enrichment | nltk | >= 3.8.0 |
| Statistical Calc. | scipy | >= 1.11.0 |
| Server Charts | matplotlib + seaborn | >= 3.7.0 |
| Client Charts | Chart.js | CDN 4.x |
| Frontend | HTML5 + Vanilla CSS/JS | Native |

## 4. Methodology

### 4.1. Data Acquisition

There are two modes of operation of data acquisition stage. The pipeline in offline mode will read PDF files that are already in the directory./Data with the help of glob module in Python. When used in online mode (when the -scrape CLI flag is used), the TRBScraper module reaches out to trb.tn.gov.in to scrape the active notifications HTML table with BeautifulSoup and filter the results two-layered: with an inclusion list of recruitment notification types and an exclusion list of eligibility tests and irrelevant announcements, and download requested PDFs with a polite delay of one second between them SSL verification is off to deal with self-signed certificate of TRB, and urllib3 warnings turned off. Files that have already been downloaded (size exceeding 1 KB) are not downloaded again. The algorithm of the acquisition is as given below in pseudocode[12].

ALGORITHM: Data_Acquisition
INPUT : Source URLs *U*
OUTPUT: Raw data collection *R*
Begin

- Access data sources using *U*
- Extract relevant documents and metadata
- Filter required records based on criteria
- Store extracted data in temporary collection *R*

End

### 4.2. Feature Extraction

The most complicated technically stage is the feature extraction, which is applied to three modules: pdf extractor.py of base PDF and OCR reading, trb extractor.py of a TRB specific parsing, and data preprocessor.py of cleaning and enriching. The extraction process identifies notification type by matching a pattern by priority order; that is, the file only matches a single one of 10 format-specific parsers. The heuristic reversed-text artifact correction function Fix_Reversed, employs a 4-stage fallback chain to correct the vertical-header artifact of PDFplumber. Subject mapping uses a 170+ word dictionary with 25, consistent fallback rules to normalize raw subject names to 25 standards. The preprocessing pipeline then uses 7 processing steps to cleanse, typed-convert, dedupe, categorize and then NLP-enrich the extracted records[13].

ALGORITHM: Feature_Extraction
INPUT : Raw data *R*
OUTPUT: Feature set *F*
Begin

- Parse raw data *R*
- Identify relevant attributes (e.g Title, date, category)
- Extract and transform attributes into structured format
- Store processed features in *F*

End

### 4.3. Data Storage

The cleaned DataFrame is stored in UTF-8 as a CSV file. Sorting Before sorting the records by (Year, Subject, Post), they are mapped with a sequential

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0313
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 2167 – 2175

integer ID column starting with 1. Southern part of Tamil Nadu, the Zone field, is always left out in the schema to be exported to minimize redundancy[14]. A summary dictionary of total records, total vacancy, year range and unique subject count is also generated by the dataset_builder module, and is logged to the console and called upon in the text report generation phase[15].

ALGORITHM: Data_Storage
INPUT : Feature set *F*
OUTPUT: Stored dataset *S*
Begin

- Define storage schema
- Map features *F* to schema
- Insert data into storage system
- Ensure data consistency and integrity
- Store final dataset as *S*

End

### 4.4. Data Analysis

The analysis engine is going to provide six dimensions of analysis using two modules: trend analyzer.py and comparison analyzer.py. NumPy is used to calculate descriptive statistics using annual totals as the target data (means, median, SD, variance, quartiles, skewness/Fisher-Pearson, excessive kurtosis, and Coefficient of Variation (CV)). Trend analysis calculates annual (YoY) absolute changes, annual percentage growth, cumulative running amounts, maximum and minimum years, a 9 year CAGR. The subject demand processes demand of each of the 25 subjects separately, calculating total, average, standard deviation, growth rate and a three tier classification of demand. The comparison analysis ranks the subjects, posts and years under a set of sort keys and produces five major findings which are employed in the text report[16].

ALGORITHM: Data_Analysis
INPUT : Dataset *S*
OUTPUT: Analytical results *A*
Begin

- Load dataset *S*
- Perform aggregation and statistical analysis
- Identify trends and patterns
- Generate analytical outputs *A*

End

### 4.5. Visualization

The visualization step gives rise to three outputs: (1) six server-side stationary PNG chart images rendered with Matplotlib and Seaborn using the whitegrid style to render the image publication-quality; (2) a JSON file (dashboard_data.json) that gets consumed by Chart.js dashboard on the client side; and (3) a structured text report (analysis_report.txt) with 8 sections. Each of the six chart types has a specific purpose in the analysis: the line plus area vacancy trend chart with a polynomial trend line will display the overall trend in the number of vacancies obtained over the years; the Seaborn Year × Subject heat map will display what particular subjects are constantly recruited vs. those only recruited to certain types of notifications, the horizontal bar chart will rank all 25 subjects on the overall number of vacancies they obtained, the diverging bar chart of YoY growth percentages visually highlights years of expansion versus contraction; and the pie chart shows category-level distribution across nine subject groups[17].

ALGORITHM: Data Visualization
INPUT : Analytical results *A*
OUTPUT: Visual representations *V*
Begin

1. Select appropriate visualization technique
2. Map analytical results *A* to visual format
3. Generate Charts/graphs
4. Present visual output *V*

End

## 5. Results And Discussion

### 5.1. Dataset Overview

In processing 10 TRB recruitment PDFs (a total of 11.8 MB), 103 structured records were extracted representing 44,136 total vacancies and 25 distinct subjects and 8 types of post between 2016 and 2025. It can also take less than 60 seconds to execute the entire pipeline on performance that is available to a typical consumer, which is a significant strength since it allows researchers to recreate the entire analysis any time that new TRB notifications become available. The vital dataset measures are: Records Generated: 103; Total Vacancies: 44,136; Year Range: 2016-2025; Unique Subjects: 25; Unique Post types: 8; Size of output csv: 9.4 KB (103 rows x 10

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0313
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 2167 – 2175

cols); Size of JSON Dashboard: 5.2 KB (397 lines)[18].

### 5.2.Year-wise Vacancy Trends

Figure 2 shows the trend of vacancy in a year-wise format including a trend line of a year in the form of a polynomial. The year 2020 (COVID-19 pandemic halted all TRB recruiting) and 2022 (no announcement was made) have not been included. The 2016 high of 15,780 vacancies is a single nostalgic batch of SCERT Lecturer. The remarkable recovery of 2024 ( +156) indicates coexisting AP (4,000 vacancies) and SGT (1,768 vacancies) announcements in the same year. Following the sharp drop after 2016, recruitment remains relatively low and stable for a few years, indicating limited hiring activity. The overall trend line shows a gradual decline despite occasional increases, suggesting the absence of consistent large-scale recruitment. The rise in 2024 appears as a temporary surge rather than a sustained trend[19].

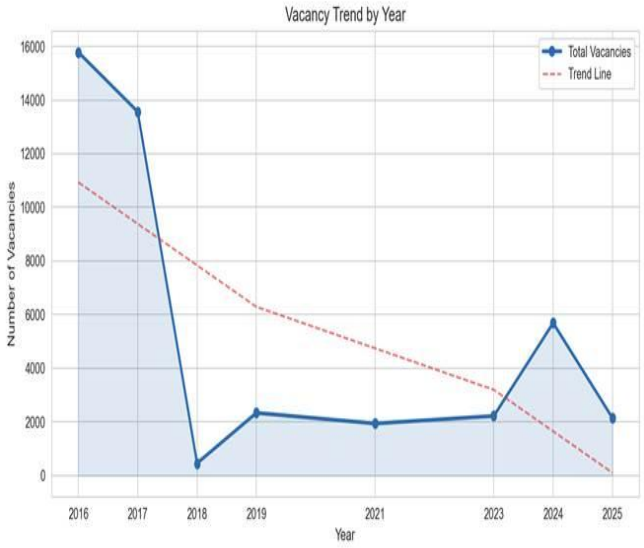


**Figure 2 Year-wise TRB Vacancy Trend with Polynomial Trend Line (2016-2025)**

As shown in figure 3, the year-over-year (YoY) growth rates are given as a divergent bar chart, which would visually differentiate years of growth (green bars) and years of contraction (red bars). The 2019 surge of +409.1 percent is an indicator of recovery to the 2018 near-collapse of (-96.6) and the 2024 recovery of +156.0 percent is a sign of sustained policy-driven renewal before partial recovery in 2025 (-62.3 percent)[21].

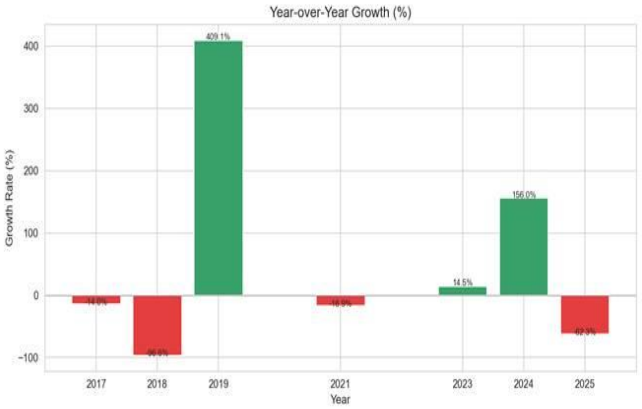


**Figure 3 YOY Vacancy Growth Rate (%)**

### 5.3. Subject-wise Analysis

3,134 total vacancies are dominated by education (17,567 with 16.2% SCERT Lecturer batch and 2024 SGT inclusion). The biggest growth rate was documented at +3,150% but it stems of structural deficiency between 2016-2018 notifications and is not an indication of a policy change. The real consistent cross-notification growth demonstrated by English was +1,379%. Computer Science has fallen by -72.7, indicating the state government redefined IT learning programs[20].

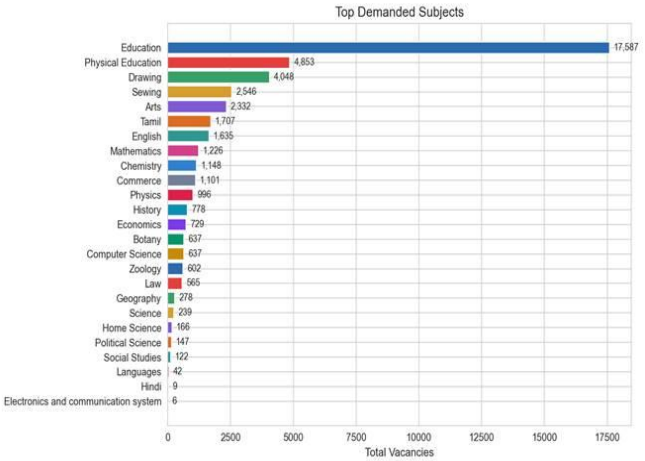


**Figure 4 Top Demanded Subjects by Total Vacancies (2016-2025)**

The subject-wise heatmap (Year x Subject matrix) is presented in Figure 5. The SCERT outlier effect is

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0313
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 2167 – 2175

affirmed by the dominance of the dark cell Education 2016, with the other cells approaching zero, which reflects the skewness and sparse/notification-type-specificity of subject demand by year[22].

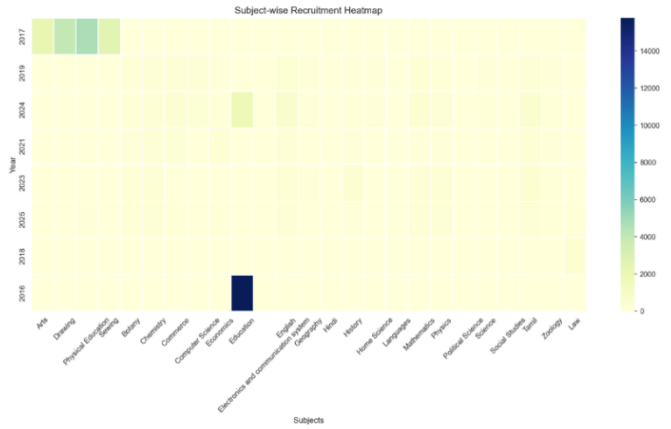


**Figure 5 Subject-wise Recruitment (Year x Subject, 2016-2025)**

### 5.4. Subject Category Distribution

The vacancy distribution of the nine subject categories is illustrated in Figure 6. Education has the highest share at 39.8%, followed by Other (Drawing/Sewing/Electronics) at 15.0%, STEM at 12.4%, and Health and Physical at 11.4%. Languages account for 7.7%, while Arts (5.3%), Social Sciences (4.7%), Commerce (2.5%), and Law (1.3%) make up the remaining portion. The "Other" category represents subjects that appeared only once in the notifications. This variation reflects differences in subject demand across recruitment cycles and institutional requirements, highlighting the uneven distribution of opportunities across disciplines.

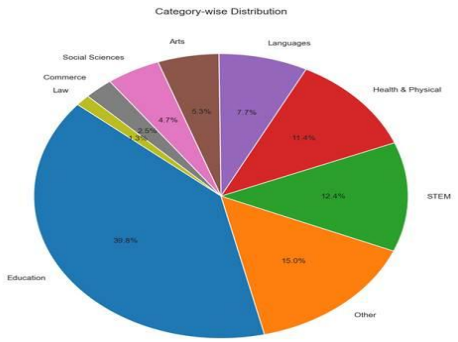


**Figure 6 Vacancy Distribution by Subject Category**

Figure 7 is a rank of all the types of vacancy. In front of T.T.A. 15,780 vacancies (single 2016 SCERT batch) are occupied by Lecturer/Senior Lecturer/Junior Lecturer, then there are Special Teacher (13,565) and Assistant Professor (6,258). These three types of posts have more than 80 percent of all vacancies in the data.

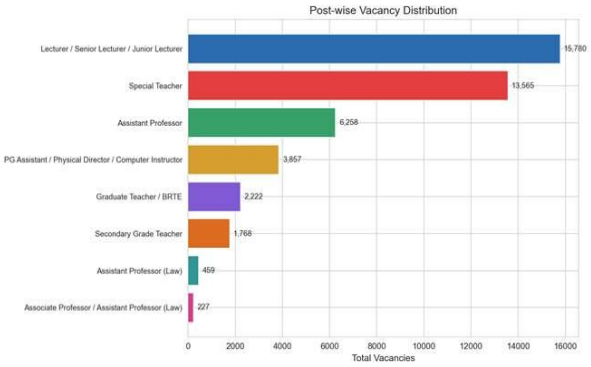


**Figure 7 Post-wise Vacancy Distribution**

### 5.5. Descriptive Statistics

The following important statistics are obtained by summing the eight annual vectors of observation. Mean: 5,517 (inflated by the 2016 outlier); Median: 2,279.50 (better central estimate); Standard Deviation: 5,486.96 (extreme spread); Min: 459 in 2018 (lowest recruitment year); Max: 15,780 in 2016 (SCERT Lecturer batch); Q1: ~2,030; Q3: ~7,254; IQR: ~5,224. The fact that the large gap between mean and median is 5,517-2279 - the ratio is 2.4:1 - directly measures the distorting influence of the 2016 outlier on summary statistics. The Fisher-Pearson skewness of +1.4, supports a high right asymmetry. With a Coefficient of Variation (CV) of 99.5% of a high degree, Tamil Nadu TRB recruitment can be regarded as highly volatile (high CV indicated high volatility, more than equal to 30%). The calculated CAGR of -24.82% based on the anomalous base of 2016 is not accurate, as the CAGR of the sub-period 2019-2024 is +143.6% which is more reflective of the true post-regularization growth curve. The range of vacancies is 15,321, indicating a large difference between the highest and lowest recruitment years. The interquartile range (IQR) of approximately 5,224

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0313
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 2167 – 2175

shows that the middle 50% of the data is widely spread[23].

## Conclusion

This paper is an exploration of teacher recruitment in the years 2016-2025 under the Tamil Nadu Teachers Recruitment Board (TRB). The primary value of the work is to transform the scattered notifications of recruitment into a comprehensive and useful set of data. The web scraping of the official TRB web enabled the initial collection of the recruitment PDFs. Such files were then handled by PDF parsing, and OCR where necessary to save appropriate data like the subject, post, number of vacancies, and year. The data extracted was processed and formatted to form a consolidated data, which was not in the structural form before. Based on this dataset, Exploratory Data Analysis (EDA) was conducted to have an idea of recruitment patterns on various dimensions. The analysis reveals that there is no regular annual hiring pattern and that the variations exhibit a large variation over time. Some of the subjects such as Education, Physical Education as well as Drawing have always registered more demand and the likes of Lecturer and Special Teacher take up a significant portion of the vacancies are contained. Statistical measures also show that the recruitment is very volatile which implies that there is no consistent demand-related hiring, instead the hiring can be seen to be disproportionate and reliant on decisions of the administrative decisions. The results are also provided in the form of an interactive dashboard and the users are able to view the trends, compare subjects and realize trends in recruitment in a simple and graphical way. This enhances useability to non-technical users. In general, this piece shows that the unstructured public information can be gathered, organized and manipulated to come up with valuable information. In future, this work may be extended to involve deeper analytical studies beyond descriptive analysis. It can be also enhanced by using refined analysis techniques and better visual representation, which can help in understanding recruitment patterns more clearly.

## Acknowledgement

The authors are highly obliged to the teaching faculty at their institute for their constant guidance and constructive suggestions during the development of this paper. They owe much credit to them for their indispensable assistance in the development of this paper. The authors also thank all those sources which they have used to successfully complete this research. No financial assistance from any outside source was taken while conducting this research.

## References

[1]B. Nishanthi, C. Hariharan, and P. Parthiban, “Unveiling the Future of Artificial Intelligence in Talent Acquisition: A Bibliometric Analysis of Emerging Trends and Future Directions,” *Elsevier*, 2025.

[2]S. Thompson-Lee, B. H. See, and R. M. Klassen, “A Systematic Review of STEM Teacher Recruitment and Retention Interventions,” International Journal of STEM Education, vol. 12, no. 33, pp. 1–18, 2025.

[3]M. Nashrullah, F. Choiriyah, Darto, G. T. Prasetyo, and B. Purwoko, “Teacher Recruitment: A Bibliometric Analysis and Visualization of Emerging Trends, Challenges and Applications,” Journal of Education Management, 2025.

[4]J. Darvai, S. Jaiswal, Y. Pandya, A. Gangrade, N. Gupta, and Y. Joshi, “AI-Driven Exploratory Data Analysis for Accelerating Data Understanding and Decision Making,” IEEE, 2025.

[5]C. D. E. Bayocot and F. J. B. Talirongan, “Analyzing Hiring Trends and Patterns Across Educational Institutions: A Study of Teachers’ Appointments and Institutional Practices,” Global Scientific Journal, 2025.

[6]J. Zhu, “Matching Prediction of Teacher Demand and Training Based on SARIMA Model Based on Neural Network,” International Journal of Information Technology and Web Engineering, vol. 18, no. 1, 2024.

[7]I. F. Baluis, “Unlocking Workforce Insights for Strategic Workforce Planning in State Universities and Colleges,” Library Progress International, vol. 44, no. 3, pp. 3907–3920, 2024.

[8]M. Maphosa, W. Doorsamy, and B. S. Paul, “Student Performance Patterns in Engineering at

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0313
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 2167 – 2175

the University of Johannesburg: An Exploratory Data Analysis," IEEE Access, vol. 11, pp. 48977–48987, 2023. doi: 10.1109/ACCESS.2023.3277225.

[9] A. Shahid, V. Maask, R. Ahmadiahangar, J. Martins, and A. Rosin, "Exploratory Data Analysis for Demand-Side Flexibility Quantification," in Proc. IEEE Int. Conf. Compatibility, Power Electronics and Power Engineering (CPE-POWERENG), 2023. doi: 10.1109/CPE 58103.2023.10227408.

[10] A. Alibasic, H. Upadhyay, M. C. E. Simsekler, T. Kurfess, W. L. Woon, and M. A. Omar, "Evaluation of the Trends in Jobs and Skill-Sets Using Data Analytics: A Case Study," Journal of Big Data, vol. 9, no. 32, 2022.

[11] R. Mitchel, Web Scraping with Python: Collecting More Data from the Modern Web, 2nd ed., O'Reilly Media, 2018.

[12] W. McKinney, "Data Structures for Statistical Computing in Python," in Proc. SciPy, pp. 56–61, 2010.

[13] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," Computing in Science & Engineering, vol. 9, no. 3, pp. 90–95, 2007.

[14] M. L. Waskom, "Seaborn: Statistical Data Visualization," Journal of Open Source Software, vol. 6, no. 60, 2021.

[15] A. Jain, S. Kumar, and R. Reddy, "Comparative Evaluation of Python PDF Table Extraction Libraries on Indian Government Documents," Journal of Data Engineering, vol. 4, no. 1, pp. 23–35, 2021.

[16] O. Allal-Chérif, A. Yela Aránega, and R. Castaño Sánchez, "Intelligent recruitment: How to identify, select, and retain talents from around the world using artificial intelligence," Technological Forecasting and Social Change, vol. 169, p. 120822, 2021.

[17] V. R., H. R., S. E., and D. V., "Artificial intelligence – talent acquisition in HEIs recruitments," International Journal of Information and Learning Technology, vol. 41, no. 3, pp. 230–243, 2024.

[18] K. P. Sasi, "Impact of AI in recruitment and talent acquisition," Human Resource Leadership Journal, vol. 9, no. 3, pp. 78–83, 2024.

[19] J. S. Roppelt, N. S. Greimel, D. K. Kanbach, S. Stubner, and T. K. Maran, "Artificial intelligence in talent acquisition: A multiple case study on multinational corporations," Management Decision, vol. 62, no. 10, pp. 2986–3007, 2024.

[20] P. Will, D. Krpan, and G. Lordan, "People versus machines: Introducing the HIRE framework," Artificial Intelligence Review, vol. 56, no. 2, pp. 1071–1100, 2023.

[21] J. Hota and D. Ghosh, "Workforce Analytics Approach: An Emerging Trend of Workforce Management," AIMS International Journal of Management, vol. 7, no. 2, pp. 167–179, 2013.

[22] A. Safarishahrbijari, "Workforce Forecasting Models: A Systematic Review," Journal of Forecasting, vol. 37, no. 7, pp. 739–753, 2018.

[23] W. Y. M. Momin and K. Mishra, "HR Analytics as a Strategic Workforce Planning," International Journal of Advanced Research, vol. 1, no. 4, pp. 258–260, 2015.