WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal Homepage: www.ijrpr.com ISSN 2582-7421

VISTA: A Municipal Budget Analysis Dashboard for Structured Financial Data Exploration

Mrs. G. Hemaprabha*[1], Reshma M*[2], Rithika G*[3], Sandhiya A*[4], Shivarakkshana S*[5]

[1-5] *Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology , Coimbatore, Tamilnadu, India*
shivarakkshanas25cs@srishakthi.ac.in

A B S T R A C T

Municipal budget documents contain important financial information about public revenue, expenditure, and resource allocation. However, these documents are often presented in lengthy PDF formats, making it difficult to organize, retrieve, and analyse financial information efficiently. This paper presents VISTA (Municipal Budget Analysis Dashboard), a prototype developed to support the structured exploration of municipal budget data, using the Coimbatore City Municipal Corporation Budget Estimate 2026–2027 as its source. The proposed workflow uses PDF text extraction and Optical Character Recognition (OCR), where required, to obtain information from budget documents. The extracted data is reviewed and cleaned before being organized into structured records. PostgreSQL is used for database management, while FastAPI provides backend services and HTML, CSS, and JavaScript are used to develop the user interface. The dashboard is intended to present categorized financial information in a more accessible and organized manner, supporting data retrieval, comparison, and exploration. The project demonstrates how document-processing techniques and database-driven web technologies can assist in making municipal financial data easier to understand and analyse. Data validation and OCR accuracy remain important considerations for improving the reliability of the system.

Keywords: Municipal Budget Analysis, Data Visualization, Optical Character Recognition (OCR), PostgreSQL, FastAPI, Financial Data Management.

1.Introduction

Municipal budgets play a significant role in the financial administration and development of cities. They provide essential information about estimated revenue, planned expenditure, fund allocation, and public development activities. However, these budget details are often published in lengthy PDF documents containing multiple sections and complex financial tables. Extracting and analysing such information manually can be time-consuming and difficult. To address these challenges, VISTA (Municipal Budget Analysis Dashboard) is developed using the Coimbatore City Municipal Corporation Budget Estimate 2026–2027 to organize municipal financial information into a structured and accessible format..The main objectivies of the proposed system:

- To extract relevant financial information from municipal budget PDF documents using text extraction and OCR techniques, where required.
- To clean, organize, and validate the extracted budget data for further processing.
- To store structured financial records efficiently using a PostgreSQL database.
- To categorize municipal revenue and expenditure information for easier exploration.
- To develop a web-based dashboard for accessing and understanding budget information.
- To support efficient retrieval and analysis of municipal financial records through an organized interface.

VISTA integrates document processing, database management, backend services, and web technologies to provide a structured approach to municipal budget analysis. The system aims to reduce the difficulty of manually navigating lengthy budget documents and improve the accessibility of financial information. It demonstrates how digital technologies can be used to organize public financial data and support clearer understanding of municipal budget information.

1.2 Related work

Optical Character Recognition (OCR) and image preprocessing techniques provide the foundation for extracting information from digital and scanned documents. Smith (2007) describes the Tesseract OCR engine for text recognition, while Otsu (1979) introduces an automatic thresholding method for

improving foreground and background separation in grayscale images [1], [2]. VISTA uses PDF text extraction and OCR, where required, to obtain municipal budget information, followed by data cleaning and validation to improve the reliability of extracted records.

Traditional PDF documents present financial information in lengthy pages and tables, making manual retrieval and analysis difficult. Relational database systems such as PostgreSQL provide structured storage, SQL queries, constraints, and views for organizing and accessing financial records [3]. VISTA applies this approach by storing categorized municipal budget information in a PostgreSQL database, providing a structured alternative to manually searching through individual budget documents Modern web technologies support the development of interactive applications for accessing and presenting structured information. FastAPI provides backend services, while PyMuPDF supports PDF text extraction and HTML, CSS, and JavaScript are used to develop the dashboard interface [4], [5]. VISTA integrates these technologies to organize municipal financial data and present it through a web-based dashboard, making budget information easier to retrieve, explore, and understand.

1.3. Problem statement and objectives

Four major challenges motivate this project: (1) municipal budget information is stored in lengthy PDF documents, making financial details difficult to locate and analyse; (2) scanned pages and complex tables create difficulties in accurate data extraction; (3) unstructured financial records make data categorization, storage, and retrieval challenging; and (4) the lack of a centralized platform limits convenient exploration of municipal revenue and expenditure. The objectives are therefore to: (a) extract relevant information from budget PDF documents using text extraction and OCR techniques, where required; (b) clean, organize, and validate financial records; (c) store structured data in a PostgreSQL database; (d) categorize revenue and expenditure information; (e) develop a web-based dashboard for presenting municipal budget details; and (f) improve the accessibility and understanding of municipal financial.

2.System architecture

2.1.Overview

VISTA follows a three-tier architecture comprising the presentation, application, and data layers. The frontend is developed using HTML, CSS, and JavaScript, while FastAPI handles backend operations. PostgreSQL stores structured municipal budget records. PyMuPDF and Tesseract OCR are used for extracting text from PDF documents. The system integrates these technologies to organize and present municipal financial information through a web-based dashboard.

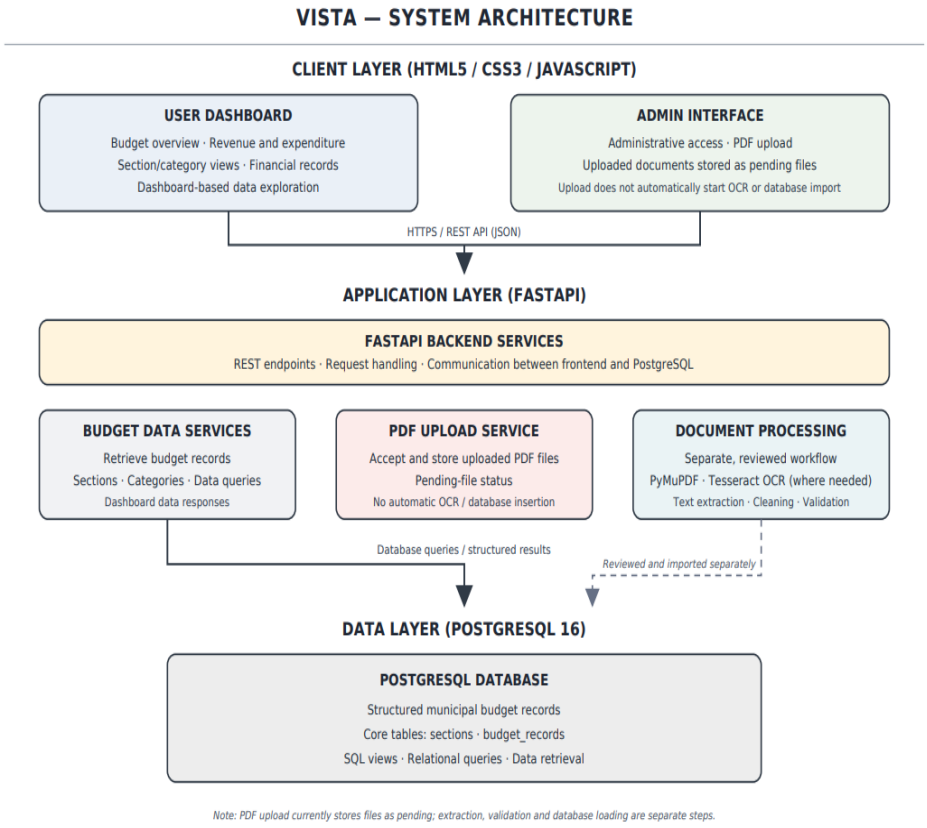


Fig. 1 - Three-tier architecture of VISTA.

2.2.Technology stack

Table 1 lists the technologies used in VISTA. The system combines lightweight frontend technologies, backend services, database management, and document-processing tools to support the structured analysis of municipal budget information

Layer	Technology	Purpose
Client	HTML5, CSS3, JavaScript	Dashboard interface and data presentation
Application	Python, FastAPI	Backened services and API request handling
Database	PostgreSQL 16	Storage and management of structured budget records
PDF Processing	PyMuPDF	Text extraction from PDF documents
OCR	Tesseract OCR	Text recognition from scanned budget pages
Data Processing	Data cleaning and validation	Organizing and checking extracted financial records
Communication	REST API,JSON	Data exchange between frontend and backend

DatabaDesign

VISTA uses PostgreSQL 16 to store and manage structured municipal budget information. The database consists of related tables, including sections and budget_records, to organize financial data systematically. Primary and foreign key relationships maintain data consistency and integrity. SQL queries and views support efficient data retrieval and categorized financial analysis. This design provides a structured foundation for presenting budget information through the dashboard..

2.3Security

VISTA incorporates security measures to protect municipal budget information and administrative operations. Administrator access is designed to use token-based authentication to restrict unauthorized access. API requests are handled through backend validation before accessing protected resources. Database access is managed through controlled backend connections, while sensitive configuration details are maintained separately from the application code. These measures aim to improve data protection and system reliability..

3.OCR-BASED DATA EXTRACTION AND PREPROCESSING

VISTA uses PyMuPDF and Tesseract OCR to extract text from scanned pages of the municipal budget PDF. The extracted data is reviewed and cleaned to reduce recognition errors before being stored in the PostgreSQL database.

3.1.PDF Processing and OCR

The PDF pages were rendered at approximately 300 DPI using PyMuPDF. The resulting images were converted into grayscale and processed using Otsu thresholding to improve text-background separation. Tesseract OCR was applied using page segmentation modes PSM 6 and PSM 4, depending on the page layout. Word-level positions and confidence information were used to inspect extracted content. Since OCR may introduce errors in financial digits, account codes and table boundaries, the extracted values require manual validation before database insertion [1], [2], [5], [6].

3.2. Data Cleaning and Classification

The extracted information passed through multiple preparation stages involving CSV consolidation, duplicate checking, classification and missing-value identification. Records were organized into categories such as Account, Particulars and Summary. Inconsistent descriptions and incomplete financial values were flagged for further review. Separate summary and detailed datasets were prepared to support database loading and dashboard analysis.

The project produced the following intermediate dataset snapshots:

Table 2 - Delay-risk classification rules.

Processing stage	Records	Columns	Description
Intial dataset	114	11	**Extracted working records**
Intial cleaning	76 clean,38 review	11	**Cleaning classification**

Processing stage	Records	Columns	Description
Consolidated master	*369*	*25*	**Combined working dataset**
Analysis ready dataset	***249***	**-**	**232 detailed and 17 summary records**

The 249-record dataset contained 183 Account records, 49 Particulars records and 17 Summary records. These values represent different processing stages and should not be combined as though they belong to one unchanged dataset.

3.4.Data Processing Workflow

The VISTA data-processing workflow consists of the following stages:

1. Identify the municipal budget PDF used as the source.
2. Extract embedded text or apply OCR to scanned pages.
3. Convert extracted information into structured working records.
4. Clean descriptions, classify records and identify duplicates.
5. Review missing or inconsistent financial values.
6. Load prepared records into the PostgreSQL database.
7. Retrieve analytical summaries through SQL views and backend APIs.
8. Display the resulting information through the dashboard.

3.5.Database design and Analytical Processing

VISTA uses PostgreSQL 16 as its relational database management system. The database separates section information from individual budget records using the core tables sections and budget_records. SQL queries and views support aggregation, classification, comparison and retrieval of municipal financial information. This structure allows the backend to obtain reusable analytical results without repeatedly processing the original PDF document [3].

3.6.SQL Analytical Views

Eight analytical views were recorded as created and executed during development. These views support the dashboard's different financial analysis modules.

3.7.Backened and Dashboard Implementation

The application backend is developed using Python and FastAPI, which handles API requests and retrieves structured financial information from PostgreSQL. The recorded analytical endpoints provide budget overview, section summaries, year comparisons, major accounts, financial increases, decreases and data-quality information. The endpoints return structured responses for use by the frontend [3], [4].

3.8.Dashboard Modules

The VISTA dashboard consists of several analytical modules designed to organize municipal budget information:

- **Budget Overview:** Presents high-level financial indicators from the prepared dataset.
- **Section Summary:** Displays categorized information about municipal revenue and expenditure.
- **Year Comparison:** Provides comparisons between the available financial periods.
- **Top Budget Accounts:** Highlights accounts with the highest recorded allocations.
- **Budget Increases and Decreases:** Displays accounts with significant positive and negative changes.
- **Data Quality:** Identifies missing or incomplete financial records.

3.9API Communication

The frontend is developed using HTML5, CSS3 and JavaScript. It communicates with the FastAPI backend through REST endpoints using JSON data exchange. The backend retrieves information using SQL queries and analytical views, processes the requested results and returns them to the dashboard for presentation. This separation supports modular development and easier maintenance

3.9.1Administrative Security and PDF Upload

VISTA includes a separate administrative module that provides login and protected PDF-upload functionality. The backend uses bearer-token authentication to control access to protected administrative operations. The recorded authentication configuration includes a token expiry period of 30 minutes. Uploaded PDF files are stored for pending processing rather than being directly inserted into the analytical database.

1. The current upload workflow does not automatically perform OCR, data cleaning, PostgreSQL insertion or dashboard refreshing. These operations remain separate from the existing upload functionality. The administrative interface and its complete integration should be tested against the final application before publication.

4.Implementation

VISTA is implemented using HTML, CSS, and JavaScript for the frontend, FastAPI for backend services, and PostgreSQL 16 for database management. The system integrates SQL analytical views and REST APIs to retrieve, process, and display municipal budget information through an interactive dashboard.

4.1 User Dashboard

1.1. The User Dashboard presents municipal budget information through an organized interface, including budget overview, section-wise summaries, financial-year comparisons, major budget accounts, significant increases and decreases, and data-quality indicators. Users can explore financial information without manually searching through lengthy PDF documents.

4.2.Admin Portal

1.2. The Admin Portal provides administrative login and protected PDF-upload functionality. Authorized administrators can upload municipal budget documents, which are stored as pending files for further processing. The current upload module does not automatically perform OCR or update the database.

4.3.Budget Analytics Module

The Budget Analytics Module uses FastAPI endpoints and PostgreSQL analytical views to retrieve and organize financial records. It provides summarized budget information, section-level analysis, account rankings and financial comparisons. The retrieved results are delivered to the frontend through REST APIs using JSON responses.

5. Results and discussion

The development of VISTA produced structured datasets, PostgreSQL analytical views and backend services for exploring municipal budget information. The following observations summarize the available development records.

5.1Data Preparation Results

Different dataset snapshots were recorded during the extraction and cleaning process.

Table 4 – Data preparation observations.

OBSERVATION	RECORDED VALUE
Initial extracted records	114
Records classified as clean	76
Records requiring review	38
Consolidated master records	369

OBSERVATION	RECORDED VALUE
Analysis-ready records	249
Summary records	17
Detailed records	232

5.2.Local API Data Verification

A separate historical local API snapshot returned 73 budget records. Among these, 58 records were marked valid, while 15 contained missing financial values. The records were distributed across four municipal budget sections.

Table 5 – Historical local API data snapshot.

SECTIONS	NUMBER OF RECORDS
Capital Payments	9
Capital Receipts	7
Revenue Payments	30
Revenue Receipts	27
TOTAL	73

The 73-record snapshot belongs to a later database/API development stage and is separate from the 249-record analysis-ready dataset. These figures require revalidation against the final submitted database before being presented as final results.

5.3 DATABASE and Functional Observations

The PostgreSQL database connection was tested during local development, and the eight analytical views were recorded as created and executed. The backend development included analytical API endpoints for retrieving summaries and comparisons. The main dashboard was also reported as functioning during local testing. However, previously identified data-display issues, including an undefined largest-allocation value, require rechecking against the final dataset.

The results demonstrate the usefulness of structured database queries and dashboard-based presentation for exploring municipal financial information. However, the reliability of the displayed results depends on the accuracy of OCR extraction, the completeness of the financial records and the validation of the data transferred into the database**.**

LIMITATIONS

The project has several limitations that affect the interpretation of its results. OCR recognition may introduce errors in numeric values, descriptions and table structures. Some summary tables from the source document were not confirmed as completely extracted. Multiple dataset snapshots were generated during development, and some later records contained missing financial values or description inconsistencies. The current PDF upload facility does not automatically perform document extraction or database insertion. Cloud database integration, complete administrative frontend integration and persistent uploaded-file storage also require final verification. Therefore, the available results should be treated as development observations rather than audited municipal financial figures.

CONCLUSIONS AND FUTURE SCOPE

VISTA presents a web-based approach to organizing and exploring municipal budget information through OCR-assisted document processing, data cleaning, PostgreSQL relational storage, FastAPI backend services and dashboard visualization. Using the Coimbatore City Municipal Corporation 2026–2027 budget as its case study, the project demonstrates how financial information from lengthy PDF documents can be organized into structured records and accessed through reusable analytical queries. The development also highlights the importance of data validation and transparency when working with public financial information.

Future enhancements include improving OCR accuracy for financial tables and multilingual content, implementing an automated PDF-to-database processing workflow, adding manual verification for uncertain extracted values, completing cloud database integration, improving dashboard filtering and report exports, and introducing traceability between displayed financial figures and their original source pages.

Acknowledgements

The authors thank Mrs.G.Hemaprabha AP/CSE and Sri Shakthi Institute of Engineering and Technology for their guidance and support during this project.

References:

1. Caamaño-Alegre, J., Lago-Peñas, S., Reyes-Santias, F., & Santiago-Boubeta, A. (2013). Budget Transparency in Local Governments: An Empirical Analysis. Local Government Studies, 39(2), 182–207. https://doi.org/10.1080/03003930.2012.693075

2. Matheus, R., Janssen, M., & Maheshwari, D. (2020). Data Science Empowering the Public: Data-Driven Dashboards for Transparent and Accountable Decision-Making in Smart Cities. Government Information Quarterly, 37(3). https://doi.org/10.1016/j.giq.2018.01.006

3. García-García, J., & Alonso-Magdaleno, M. I. (2022). Deconstructing Government Budgets through Visual Representation Software. International Journal of Public Administration, 45(11), 807–818. https://doi.org/10.1080/01900692.2021.1891549

4. Stanić, B. (2018). Determinants of Subnational Budget/Fiscal Transparency: A Review of Empirical Evidence. Public Sector Economics, 42(4), 449–486.
https://doi.org/10.3326/pse.42.4.4

5. Crossman, H. A., & Fischer, D. (2016). Participatory Budgeting and Transparency in Municipal Finances. Journal of Accounting, Ethics & Public Policy, 17(3), 663. https://doi.org/10.60154/jaepp.2016.v17n3p663

6. ElBerry, N. A., & Goeminne, S. (2021). Fiscal Transparency, Fiscal Forecasting and Budget Credibility in Developing Countries. Journal of Forecasting, 40(1), 144–161. https://doi.org/10.1002/for.2695

7. Does E-Government Initiatives and E-Participation Affect the Level of Budget Transparency? (2022). International Public Management Journal, 25(3), 365–391. https://doi.org/10.1080/10967494.2022.2042437

8. OECD. (2017). OECD Budget Transparency Toolkit: Practical Steps for Supporting Openness, Integrity and Accountability in Public Financial Management. OECD Publishing. https://doi.org/10.1787/9789264282070-en

9. International Monetary Fund. (2002). Manual on Fiscal Transparency. IMF. https://doi.org/10.5089/9781589060555.069

10. World Bank. Financial Management Information Systems and Open Budget Data: Do Governments Report on Where the Money Goes? World Bank original publication

11. World Bank. BOOST Open Budget Portal. https://www.worldbank.org/en/programs/boost-portal/about-the-portal

12. Lockwood, B., Porcelli, F., Redoano, M., & Schiavone, A. (2022). Does Data Disclosure Improve Local Government Performance? Evidence from Italian Municipalities. University of Warwick Working Paper. https://wrap.warwick.ac.uk/id/eprint/171016/

13. Azarin, A., & Supratiwi, W. Implementation of Visual Analytics to Enhance Budget Transparency and Accountability in Surabaya City Government. Indonesian Interdisciplinary Journal of Sharia Economics. https://doi.org/10.31538/iijse.v8i3.7856

14. Olarewaju, T. (2026). From Secrecy to Transparency: Can Publicly Accessible Digital Budgets Deliver Better Value? Journal of Creating Value, 12(1), 96–110. https://doi.org/10.1177/23949643261428217

15. Budget Transparency and Financial Sustainability. Journal of Public Budgeting, Accounting & Financial Management. https://doi.org/10.1108/JPBAFM-02-2022-0025

B. Official technical documentation

16. FastAPI. (n.d.). FastAPI Documentation. https://fastapi.tiangolo.com/

17. PostgreSQL Global Development Group. (2024). PostgreSQL 16 Documentation. https://www.postgresql.org/docs/16/

18. Mozilla Developer Network (MDN). (n.d.). HTML, CSS and JavaScript Documentation. https://developer.mozilla.org/en-US/

19. Tesseract OCR. (n.d.). Tesseract OCR User Manual. https://tesseract-ocr.github.io/tessdoc/

20. PyMuPDF. (n.d.). PyMuPDF Documentation. https://pymupdf.readthedocs.io/en/latest/