International Journal of Education, Science, Technology and Engineering
p-ISSN: 2685-1458, e-ISSN: 2684-9844

Original Research Paper

An Exploratory Data Analysis Approach for Tax Revenue Systems

Ayomitope Isijola[1*], Eriitunu Adesioye[1], Mirabel Egwu[2], Michael Asefon[3], Abiola Ojo[4], Chikwado Okafor[5], Azizat Adekoya[6], Samuel Okoh[7]

[1] *Department of Computer Sciences, Faculty of Science, University of Lagos. Lagos, Nigeria.*
[2] *Faculty of Education, Central China Normal University. Wuhan, China.*
[3] *Department of Computer Science, Faculty of Science, National Open University of Nigeria. Nigeria.*
[4] *Faculty of Mathematics, Physics and Computer Science, Maria Curie-Skłodowska University. Lublin, Poland.*
[5] *School of Computing, Engineering and Built Environment, Glasgow Caledonian University. Scotland, United Kingdom.*
[6] *Department of Computer Science, College of Engineering, Virginia Commonwealth University. Virginia, United States.*
[7] *Department of Computer Engineering, Faculty of Engineering, University of Maiduguri. Nigeria.*

Article History
Received:
24.07.2024

Revised:
16.08.2025

Accepted:
27.08.2025

***Corresponding Author:**
Ayomitope Isijola
Email:
ayomitopeisijola@yahoo.com

Abstract: Tax collation is an important part of any company's revenue system. Usually, over time, the process becomes more daunting, and the ability to monitor tax trends and revenue streams decreases. Not to mention gaining useful insights that can aid in decision-making and company transactions. That is why a tax data analysis system is able to continuously monitor tax information, pointing out anomalies, trends, and providing useful data visualizations. The tax analysis system would also enhance transparency and accountability in tax collection, improve efficiency, and reduce the need for audits, hence underlining its potential. Tax data is an important collection of information; however, many businesses fail to take advantage of this by not digging deeper into that collection. The aim of this research is to explore tax and sales data in an attempt to gain valuable insights and provide clearer information to the user. The methodology adopted is Exploratory Data Analysis (EDA) using Python as the main tool. The dataset used consists of 5,200 transactional tax records obtained from small and medium-scale enterprise (SME) sales reports spanning a 24-month period (Jan 2022 – Dec 2023). All data contained fields were pre-processed and stored in an SQLite database. Using Python libraries like Pandas, Matplotlib, Plotly, descriptive statistics, and visualization analyses showed that corporate tax contributions accounted for 47.8% of total tax revenue, while sales tax trends fluctuated seasonally, peaking in Q2 and Q4 of each fiscal year. The analysis demonstrated a 12% improvement in tax insight accuracy compared to manual spreadsheet tracking. The results show that with the approach, tax data can provide insights that can inform business decisions through charts and graphs. In conclusion, the platform can be a great tool in business decision-making and breaking down large datasets to give meaningful information.

Keywords: Exploratory Data Analysis, Data Visualizations, Sales Data, Tax Analysis, Tax Trends.

International Journal of Education, Science, Technology and Engineering, vol. 8, no. 2, pp. 82-97, December 2025. **DOI: 10.36079/lamintang.ijeste-0802.887**

1. Introduction

Tax collation is a crucial component of any company's revenue system. Over time, the process often becomes more daunting, and the aptness to check tax trends and decline in revenue streams. As well as getting high insights that can help in strategic planning and efficient company deals. Thus, a tax data analysis system that is able to constantly watch tax information, spot anomalies and trends, and give helpful data visualizations is an indispensable tool. The tax analysis system would also boost simplicity and censure in tax collection, help efficiency, and lower the call for audits, therefore highlighting its power.

The current taxation structure in Nigeria, run by the Federal Government in its tax division, the Federal Board of Inland Revenue (FBIR), started in 1939. Basically, tax is paid because public administrations enact tax laws, so the generated taxes can sponsor national schemes. Although taxes are needed by law, it is also reasoned an assignment to your country. Not allowing to honour the overseeing commission that takes the Federal Inland Revenue Service (FIRS) taxes will cause sanctions, punishments, and, sometimes, a prison term. In 2015, the federal government of Nigeria collected over 3.7 trillion naira in taxes. This could come from several roots, as well as Personal Income Tax, Corporate Tax, and others. The government calls for resources from taxes to start growth plans for the country and to better the well-being of society as a whole.

Explaining tax data can be intimidating and often ineffective when handled unprofessionally, making it difficult to examine closely and identify meaningful trends within raw information. In many cases, data is compiled merely for record-keeping and is rarely used as an opportunity to inform business decisions based on past insights. The key problems include the presence of untapped knowledge within tax data, a lack of focus on exploring the insights that such data can offer, and the limited availability of tools or avenues to collect, extract, and visually present data obtained from tax and sales reports.

The aim of this study is to design and implement a platform that enables users to access data analysis and visualization of their tax and revenue records. Its primary objectives are to design a fully functional tax data analysis processing system, to implement an efficient mechanism for delivering visual insights derived from tax data, and to develop a tool that helps businesses make informed decisions based on the data collected. This study covers the development of a plan that lets users enter their revenue and tax data so they can get analysis and visualization to form a greater sense of the data, and also, a chance to obtain new knowledge, analyze trends, and possibly detect fraud.

This study aims to demonstrate the usefulness of data visualizations and analysis by highlighting the importance of achieving better accuracy in data representation and understanding how data can be leveraged to support informed business decisions. The specific benefits include improving knowledge of available data, helping organizations recognize the potential and insights embedded in their data stores; enabling the handling of large volumes of sales data through a system capable of managing extensive datasets; and providing real-time data visualization, allowing the system to display graphs and data representations instantly for timely decision-making.

The application will be developed using the Flask micro web framework along with data manipulation and analysis libraries, including pandas, plotly, and scikit-learn for data classification. Accordingly, the techniques employed include collecting a dataset of sales tax and revenue data using an SQLite database, followed by a preprocessing stage in which the data is cleaned and prepared before being entered into the database using pandas functions. Finally, data visualizations are produced by representing the compiled data through matplotlib and pandas to present patterns and insights more clearly

2. Literature Review

Current financial and tax management systems growingly depend on data-driven techniques to derive knowledge, spot anomalies, and aid tactical decision-making. As organizations gather big volumes of transactional and revenue-related data, conventional procedures become inadequate for spotting significant designs or aiding prompt decisions. The move toward self-operating and data-driven models has thus made Exploratory Data Analysis (EDA) an elementary approach for grasping financial personality, exposing concealed forms, and yielding applicable intelligence. In this bigger picture, previous studies explain the booming significance of statistical exploration, visualization methods, and data-preprocessing structures in bettering financial transparency and analytical efficiency. The studies assessed in this part gives theoretical foundation for the current study’s stress on making use of EDA to examine tax revenue systems, showing how data exploration can brace up

accuracy, monitoring, and decision support.

Sarker [1] asserted that advanced Exploratory Data Analysis (EDA) is not only an initial step, but a key component of today's business intelligence systems. In a data-driven conditions, EDA lets organizations to discover unseen patterns, identify anomalies, and aid evidence-based decision making. As digital record-keeping grows, the volume and complexity of business data increase, heightening the need for sophisticated analytical techniques that can summarize large datasets efficiently. Amidst these approaches, descriptive statistics give a quantitative basis; visualization tools like histograms, heatmaps, and time-series plots back pattern recognition; and interactive visual analytics structures let analysts delve into data for extensive knowledge.

Current studies have made notable contributions to the area of data-driven tax administration by showing how artificial intelligence and advanced analytics can nourish compliance monitoring and fraud detection. Rahman et al. [2] stress that as financial data increase in size and complexity, conventional assessment techniques are never again good for spotting abnormalities or making sure of efficient oversight. Likewise, Belahouaoui and Alm [3] talk about how combining machine-learning models, self-operating risk-scoring procedures, and intelligent analytic structures lets tax authorities handle big volumes of records more precisely and effectively. Jointly, these research studies accentuate the move toward intelligent tax analytics systems able to amplify transparency, bettering decision-making, and backing current tax governance.

A main exploration by Rodríguez-Quintero et al. [4] shows how joining process mining and visual analytics can substantially harden financial compliance monitoring. Their research concludes that big transactional data frequently conceal intricate behavioral structures that traditional statistical procedures overlook. By utilizing visual process-mining approaches, like dotted charts, fuzzy miner graphs, and interactive visualizations, they express how auditors can spot fine abnormalities, anomalous sequences, and unconventional process behaviors. This method backs early recognition of suspicious activity, therefore bettering transparency and audit efficacy.

Together, these previous studies point a need for research that applies EDA and visualization techniques specially to tax revenue datasets, with the intention of getting clear, interpretable, and applicable knowledge for business decision-making. The current studies don't have a rapt study that shows how exploratory analysis can clarify tax data interpretation, intensify transparency, and back prompt tax-related decisions, especially for SMEs working in data-rich but analytically underused conditions. This study fills these gaps by creating an EDA-driven tax analysis system able to break down big tax data into significant visual information that helps accuracy, discloses patterns, and aids strategic financial management.

2.1. Correlation in the Middle of IT and Tax Revenue

Recently, the connection between information technology (IT) and tax revenue has become a center for tax-administration research, particularly in growing economies. The digitalization of tax procedures, from data collection and payment to analytics, is known to amplify tax compliance, better efficiency, and nourish transparency. As digital tools expand, their blending into tax administration could change how governments check flows, spot evasion, and perfect revenue assemblage. Nevertheless, the study shows mixed findings, notable implementation risks, and analytical gaps, especially when it comes to adding exploratory and visual analytics to transactional tax data. This study's concentration on making use of Exploratory Data Analysis (EDA) to assess SME tax transaction records puts it between technology adoption and revenue system creation, looking forward to filling up the captious gaps in current research studies.

1) Digital Payment Systems and Tax Collection Efficiency

Omodero [5] looked into how Nigeria's move towards digital payment systems, as well as POS, web-based payments, and mobile payment mechanisms, align with tax revenue collection in a lasting timeframe (2009–2023). The research discovers that web payments (WPY) and POS payments have a positive lasting relationship with tax revenue, urging that some digital payment techniques are more efficient at spotting tax-relevant transactions. This backs the concept that IT infrastructure (particularly payment systems) can boost tax systems, but also implies that not all digital mechanisms are put in equally, which demands more gritty analytical techniques, like EDA, to unload these structures.

2) SME Electronic Tax System Adoption and Tax Revenue

Inegbedion [6] inspects how SMEs' acquiescence of an electronic tax system affects tax compliance

and, in turn, tax revenue. Making use of a Technology Acceptance Model (TAM) structure, the research study discovers that e-tax system adoption notably helps tax compliance, and compliance intercedes the effect on revenue. While this affirms IT adoption's work in boosting tax compliance, it does not greatly investigate the transaction-level data forms (e.g., seasonal sales tax, anomaly behaviors) that our EDA-based research points to. There is space to join large-scale behavioral discoveries with small-scale data analytics.

3) Moderating Role of IT in Tax Management

Olumoh and Sanni [7] investigated how Information Technology adoption lessens the connection between tax management operations and revenue effectiveness in Internal Revenue Services in Nigeria. Their results propose that high IT adoption nourishes the positive outcomes of efficient tax management on revenue agency effectiveness. This portrays that IT adoption is not only a single factor but can boost the existing tax management approach. Yet, their method is mostly organizational and survey-based, not data-driven with exploratory analytics.

4) Digital Technologies and Revenue Generation

Olasunkanmi and Adejuwon [8] probed the effect of digitalization of tax systems on revenue production in Nigeria. Their experimental analysis shows a notably positive connection between digital tax systems and revenue, particularly when e-tax policies and ICT personnel are sufficient. While asserting that digitization amplifies the ability of tax collection, this study does not dig into how data analytics (beyond digitalization) can uncover trends, anomalies, and applicable knowledge. This proves the need for analytics-oriented tax data systems.

5) Challenges and Broader Effects of Digital Tax Administration

He [9] presented a theoretical analysis of how digital technologies, as well as big data, AI, and blockchain, adjust tax administration, pointing out the positive effects (efficiency, compliance, transparency) and also the technical and institutional problems (security, privacy, legal barriers). This wide view of digital technology's influence reinforces that while IT can remarkably transmute tax systems, practical implementation needs cautious design.

Even though current studies bring out broad knowledge into tax compliance, tax administration problems, and the general performance of tax systems, there is little research that particularly uses self-operating exploratory data analysis (EDA) methods in corporate tax conditions. Most of the current studies concentrate on conventional tax collection structures and regulatory affairs, still it does not give data-driven techniques for spotting trends, observing revenue streams, or collecting applicable knowledge. Moreover, the combination of today's analytical tools, like interactive dashboards, real-time tax analytics, and visualization-driven outlining, remains undiscovered in both academic and industry conditions. Only some works inspect how EDA can amplify decision-making, transparency, or revenue prediction in private organizations. This gap stresses the need for a detailed, EDA-based structure that backs effective tax monitoring and gives intricate analytical value for corporate tax management systems.

2.2. Tax Processing Systems

Efficient tax processing systems join digital infrastructure, data pipelines, and analytics to transmute raw transactional records into well-timed, applicable revenue intelligence. As tax administrations and organizations embrace electronic filing, digital payment mechanisms, and consolidated data repositories, the possibility for data-driven observation and decision support increases. However, the study shows that digitization not only guarantees knowledge: maturity in data management, exploratory analytics, and visualization is needed to derive value from transaction streams and to aid compliance, forecasting, and functional decision-making [10].

1) Digitalisation and Analytics Maturity in Tax Administration

New international evaluations stress that while a lot of tax authorities have designed digital schemes and started to embrace AI/analytics, their capability in making use of data for knowledge remains irregular. The OECD points out that although many administrations have data-management approaches and are practicing with analytics, the full combination of exploratory and interactive analytics into routine tax processing is still confined; this curbs the potential to spot subtle structures or to give dashboard-level decision support. These observations underline the need for analytics layers, not only transactional systems, in tax processing channels [10].

International Journal of Education, Science, Technology and Engineering, vol. 8, no. 2, pp. 82-97, December 2025. **DOI: 10.36079/lamintang.ijeste-0802.887**

2) Experimental Proof from Emerging Economies (Nigeria focus)

Nationwide studies from Nigeria outline that digital tax refinement and e-tax systems amplify administrative efficacy and have the power to heave compliance and collections, but they also uncover implementation gaps. Experimental analyses suggest that digital payment mechanisms and incorporated tax mechanisms lessen manual labor and merge taxpayer records; however, they encounter difficulties in scale, SME adoption, and data integration, which restricts their analytical effectiveness without dedicated EDA and visualization tools. This study review indicates digital infrastructure is necessary but not sufficient: tax processing systems also require analytics capabilities tailored to transaction-level exploration, which was explored by Ogbaisi and Ibifunmilola [11].

3) Role of Advanced Analytics and Anomaly Detection

Pinto and Sobreiro [12] presented a growing technical literature on anomaly detection and AI for tax/fraud monitoring, showing the promise of machine learning and visual analytics for compliance and detection tasks. Survey and review work on anomaly detection methods for financial systems point to a wide range of statistical, ML, and visualization approaches that can be embedded into tax processing systems to surface irregular transactions. Yet, many tax implementations reported in country studies lack an explicit EDA stage that prepares, visualizes, and contextualises transactional tax data for both auditors and business managers.

2.3. Tax Trends Worldwide

Knowing vast tax trends is important for creating productive tax monitoring systems and analytical tools. Current global and nationwide studies show that tax-to-GDP ratios, revenue composition, and temporal designs are pushed by macroeconomic cycles, structural refinements (including digitalisation), and sporadic shocks like the COVID-19 epidemic. These advances make a moving target for tax administrations and businesses too, making gritty, transaction-level analysis good for prompt knowledge and responsive policy or operational work [13].

1) Aggregate Tax Levels and Recent Movements

Panoramic comparative statistics point that total tax-to-GDP ratios differ considerably over countries and have displayed decent short-term unpredictability over the last few years. OECD data express that tax-to-GDP ratios over member countries moved a bit between 2021 and 2022, returning varied recoveries and policy replies because of the epidemic; such cross-country diversity underlines that the universal average hides key local trends. These big trends point to the need for analytical systems that can make up national-level goals with firm or transaction-level occurrences [13].

2) Digitalization, Firm Digitalization, and Revenue Performance

Current cross-country studies denote a positive collaboration between firm-level digitalisation (and wider tax digitalisation) and tax revenue efficiency. An IMF working paper discovers that bigger firm digitalisation matches with bigger tax-to-GDP ratios, depending on public digital ability (GovTech), proposing that digital infrastructures and administrative ability together strengthen revenue gains. Country studies also report that digital payment channels and e-tax platforms can consolidate taxpayer records and reduce manual leakage, but the analytical layer that turns digital transaction streams into actionable insight often lags behind system deployment, which was demonstrated by Nose et al. [14].

3) Seasonality and Intra-year Concentration of Tax Receipts

Srivastava et al. [15] expressed those empirical studies consistently find pronounced intra-year concentration and seasonal patterns in tax receipts, driven by payment schedules, corporate reporting cycles, and sectoral seasonality (e.g., tourism, retail). Country case studies identify recurring monthly or quarterly peaks (for instance, end-of-quarter and fiscal year bunching), which complicate simple trend estimation and forecasting, especially when taxpayers bunch payments around policy deadlines. Detecting and visualising these seasonal structures is therefore critical for both forecasting and compliance monitoring.

4) Forecasting Approaches and Methodological Advances

Kaushik et al. [16] presented a growing methodological literature that applies time-series forecasting and machine-learning techniques to tax and revenue prediction. Recent work uses grey models, econometric forecasting, and ML hybrids to improve short-term revenue forecasts at country and sectoral levels. These advances in methodology show that joining statistical forecasting with

exploratory visual attributes can result in more powerful, interpretable predictions.

These studies jointly signal the need for practical, transaction-focused analytical systems that are able to visualise, summarise, and diagnose tax data at a gritty level, specifically the addition that our present EDA-based tax analysis system tries to make.

2.4. Benefits of Digitalization

Digitalization of tax systems, all-inclusive e-filing, electronic invoicing, digital payment mechanisms, centralized data repositories, and GovTech solutions have swiftly evolved how tax administrations work. Experimental and policy research documents numerous advantages from this change, as well as bigger revenue assemblage, better compliance, administrative efficacy, higher transparency, and enlarged tax base coverage. Nevertheless, while digitization generates high-value transaction streams, changing those streams into applicable administration and policy insights needs analytical layers (EDA, visualization, dashboards) that remain underdeveloped in practice [10].

1) Increased Revenue Mobilization and Improved Compliance

Nose et al. [14] illustrated that many studies outline a positive relationship between digital tax infrastructures (and firm-level digitalization) and tax revenue effectiveness. Digital payment mechanisms and combined e-tax schemes lessen leakages and make more transactions clear to authorities, which aids in amplifying collections and voluntary compliance. Cross-country and IMF analyses discover that when digital capacity (GovTech) is in place, digitalisation corresponds with bigger tax-to-GDP ratios and amplified tax revenue.

2) Administrative Efficiency and Cost Reduction

Digital systems simplify everyday tasks (filing, payment reconciliation, record gathering), remarkably lowering the administrative problem for both taxpayers and tax authorities. National reports and policy reports show that digitization can decrease wait times, lowering costs of collection per unit, and liberate audit resources for dangerous cases, results that amplify system performance [17].

3) Enhanced Transparency, Data Integrity, and Fraud Detection

By centralising records and making use of unchangeable digital logs (as well as API combinations and geo-tagged collection tools), digital tax models amplify data confirmability and lessen chances for exploitation. This grows public trust and allows for more efficient fraud detection when joined with analytics. Creative examples, like Bouissou [18] use of satellite imagery and combined registries for property tax, show how digitisation can actually step-up compliance with past limited tax bases.

4) Reduction of Tax Evasion and Shadow-Economy Activities

Digital transformation has proven to lessen tax evasion by amplifying the observability and accountability of firm-level transactions. Ván et al. [19] figured out that important digital reporting remarkably lowers the possibility of firms appealing in black market activities, with the impact being powerful among SMEs, where digital records make real-time discoverability. Their research study shows that digital tools, like electronic invoicing, self-operating reporting, and digital bookkeeping, cover conventional loopholes and lessen the practicability of informal, unrecorded transactions.

5) Improved User Experience and Faster Service Delivery

Digitalization outstandingly amplifies service quality by bettering user experience and lessening delivery time. According to Ilieva [20], the embracement of digital public-service schemes results in better user satisfaction because systems like that give basic interfaces, quick access to services, self-operating support features, and diminished need for physical interaction. The research further shows that well-designed e-government portals reduce processing delays and lessen the intellectual effort needed to finish administrative assignments, in that way boosting voluntary adoption and bettering general service effectiveness.

Despite the fact that digitalization conveys direct structural benefits, current reviews and policy reports stress a realistic limitation: digitization makes data, but not necessarily knowledge. To change transaction streams into functional intelligence (seasonality identification, anomaly spotting, category breakdowns, dashboard KPIs), tax systems need exploratory analytics and visualization layers. This gap, where big digital data persists but exploratory, user-centric analytics are not present, asserts the present study's concentration on Python-based EDA and visualization for SME tax records [10].

International Journal of Education, Science, Technology and Engineering, vol. 8, no. 2, pp. 82-97, December 2025. **DOI: 10.36079/lamintang.ijeste-0802.887**

2.5. Importance of Big Data Applications

Badshah et al. [21] addressed big data applications, noting that the tools and systems that collect, store, process, and analyse very large and complex datasets have become central to modern decision-making in both the private and public sectors. In taxation and financial management, Big Data enables richer descriptive analytics, improved anomaly detection, stronger forecasting, and evidence-based policymaking; however, realising these benefits requires mature data governance, technical capacity, and integrated analytics pipelines (EDA → modelling → visualization).

1) Enhancing Decision Quality and Speed

Chatterjee et al. [22] investigated empirical studies that show that Big Data Analytics (BDA) significantly improves the speed and quality of organisational decisions by providing timely, fine-grained evidence rather than relying on aggregated, delayed reports. They showed that organisations using BDA report faster and more accurate decision-making in complex operational contexts, a finding that generalises to tax administrations and corporate finance, where rapid detection of revenue deviations matters.

2) Enabling Advanced Anomaly Detection and Fraud/Fault Identification

Ram and Desgourdes [23] introduced a broad technical study that documents the effectiveness of Big Data approaches (machine learning, statistical detection, and visual analytics) for uncovering anomalous transactions and potential fraud in large financial streams. Reviews and applicable research studies specify that multidimensional feature removal plus unsupervised detection techniques can show fine abnormalities that rule-based methods overlook, amplifying detection rates and lessening false positives when validated well. This potential is important to tax systems that must spot wary filings or payment structures.

3) Transforming Public-Sector Analytics and Governance

Hossin et al. [24] emphasized that big data applications are changing public-sector governance by allowing prognostic governance, evidence-based resource allotment, and targeted compliance interferences. Systematic reviews and policy reports point out that governments with grown Big Data abilities can better prioritise audits, tailor taxpayer outreach, and calculate program impacts, with these benefits being unforeseen in data quality, interoperability, and institutional effectiveness. This study presents that technical implementation must be joined with governance reforms to unleash the full benefit.

4) Practical Challenges and the need for EDA & visualization layers

Huda and Jatmiko [25] illustrated that while the power of Big Data is generally known, many current reviews stress continual practical problems: data fragmentation, poor data quality, finite analytics skills, privacy/regulatory concerns, and the lack of exploratory/visual analytic layers that yield results clear for non-technical collaborators. In other words, digitisation alone does not equal insight; organisations need structured EDA pipelines and visualization dashboards to translate raw transaction streams into actionable knowledge for auditors, managers, and policymakers. This gap motivates the present EDA-based platform for SME tax records.

2.6. The Influence of Big Data on Corporate Tax Management

Big Data and associated analytics are reshaping corporate tax management by improving the visibility, timeliness, and interpretability of transactional and financial records. Where tax functions were previously reliant on manual reconciliation and spreadsheet workflows, modern big-data pipelines enable continuous monitoring, anomaly detection, scenario simulation, and richer internal controls. However, the literature also shows that the benefits of big-data adoption depend on complementary organizational capacities (internal controls, governance, and digital maturity). The following review summarises recent, indexed findings that are directly relevant to applying EDA to corporate and SME tax records.

1) Big Data improves Audit Selection, Pricing, and Risk Targeting

Wu et al. [26] findings from recent accounting and tax show that data-driven enforcement and analytics reduce audit risk and can lower audit pricing for compliant firms, as third-party digital records and analytics enable more accurate risk scoring and more efficient allocation of audit resources. Embedding anomaly detection and predictive models into tax processing pipelines allows

tax administrations and firms to focus investigative effort on genuinely suspicious cases while reducing false positives commonly associated with static, rule-based triggers. This supports the rationale for adding EDA layers that prepare and visualise transaction features used by downstream ML models.

2) Big Data and Advanced Analytics strengthen Corporate Tax Planning and Forecasting
Najem et al. [27] demonstrate that big-data analytics (including ML hybrids) improve forecasting and scenario simulation, which corporate tax teams can use for proactive planning and liability estimation. Reviews of AI and big-data techniques in finance document how multidimensional features and machine learning models enhance predictive performance over simple time-series approaches, particularly when an EDA phase constructs and validates robust input features. This methodological chain, EDA → feature engineering → hybrid forecasting, reinforces why exploratory, visual analytics are essential for reliable tax forecasting and strategic tax planning.

2.7. Importance of Tax Data Analysis

Aslett et al. [28] define tax data analysis as the systematic examination of transactional and administrative tax records using statistical, visual, and machine-learning techniques, which is increasingly central to modern tax administration and corporate tax management. As tax systems digitalize, big sizes of transaction-level data become accessible; deriving value from these streams amplifies revenue forecasting, strengthens compliance and audit targeting, lessens leakages, and aids well-timed business decision-making. Nevertheless, the research study reveals that by just digitizing tax flows is not sufficient: dedicated analytics, exploratory workflows, and visualization layers are required to turn raw data into actionable intelligence.

1) Better Revenue Forecasting and Timelier Fiscal Insight
Advanced tax data analysis strengthens short-term forecasting and perception of the environment for both tax authorities and firms. Policy reviews and IMF guidance explain how analytics, when added to high-frequency tax and payment data, tauten short-term revenue forecasts and let authorities spot new collection shortfalls earlier than total reporting alone. This promptness helps governments and businesses respond to shocks and modify policy or functional plans.

2) Enhanced Anomaly Detection and Fraud Prevention
Yuan et al. [29] added statistical diagnostics, unsupervised methods, and visual analytics to transaction records greatly boosts the ability to identify wary filings, interpreted sales, and other anomalies that rule-based checks overlook. Interactive systems and visualization tools can surface unusual patterns, enabling auditors to prioritise cases with higher risk and reduce false positives compared with static triggers. Recent implementations (e.g., interactive tax-audit visualisation systems) demonstrate improved detection and clearer presentation of anomalies to investigators.

3) More Efficient, Risk-based Audit and Enforcement
Wang et al. [30] presented that data-driven risk scoring and feature-rich profiles derived from transaction histories enable more accurate audit selection and resource allocation. Studies of national digitisation programs show that consolidating third-party records and building analytic risk metrics reduces the time and cost per audit and increases yield from targeted enforcement. These results rely on preprocessing and exploratory phases that put together transaction attributes for subsequent machine-learning models and performance metrics.

4) Strengthening Corporate Tax Management and Internal Controls
At the firm level, tax data analysis tools assist corporate tax teams in reconciling intricate transactions, tracing deductible items, and imitating tax liabilities in other scenarios. National proof shows that when tax authorities and firms both embrace digitalized reporting and analytics, internal control effectiveness strengthens and corporate reporting becomes clearer, advantages that come from more-detailed transaction-level analysis [30].

Research shows that tax systems have progressed from normal processes to digital, data-driven conditions that outstandingly refine compliance, efficacy, and internal control. However, digitalisation only cannot get worthwhile knowledge without strong analytical procedures. Researchers persistently point out Exploratory Data Analysis (EDA) as a key element for changing raw tax transactions into

applicable information. EDA methods, like data cleaning, anomaly detection, descriptive statistics, and visual analytics, back organisations to spot patterns, abnormalities, and revenue exploits that remain unseen in intricate datasets. In spite of progress in digital tax infrastructures, the main analytical gaps keep at it. Many present studies concentrate on broad tax reforms or theoretical parts of digitalization, but don't always add practical, transaction-level EDA to real-world tax data, specifically within SMEs. The absence of interactive visualisation tools, self-operating dashboards, and hands-on analytical workflows restricts the efficiency of today's digital tax systems. The reviewed studies hence underline the need to add digital systems with advanced analytics. Big data channels, machine learning models, and forecasting tools all rely on powerful EDA foundations. This gap accounts for the significance of the current study, which adds structured EDA techniques to SME tax datasets to amplify transparency, refine monitoring, aid decision-making, and lessen manual processing. Eventually, digitalisation gives the infrastructure, while EDA proffers the intelligence, jointly shaping the backbone of today's tax revenue systems.

3. Methodology

3.1. System Analysis and Design

The proposed application is designed to receive raw data and provide refined information. It allows users to upload files containing tax information that follow a particular format. In return, various visualizations and feedback concerning those files are provided, which can then be used to highlight trends and make business decisions.

Here is a step-by-step methodology for providing tax analytics:

1) User Registration
2) Data Collection
3) Data Preprocessing
4) Data Visualization

The functional requirements represent intended system capabilities:

1) Ability to register users and store information.
2) Accept uploads of ".csv" files to the database corresponding to each user.
3) Analyze information from files uploaded into the database to generate reports and corresponding data visualizations in a dashboard.
4) Display data in a clear and easy-to-understand format for users.
5) Protect user information.
6) Keep an updated database.

The non-functional requirements depict the system's look and feel:

1) The web application should be easy and intuitive for customers to use.
2) The dashboard should display clear and concise images.
3) The flow of use should be easy to understand.

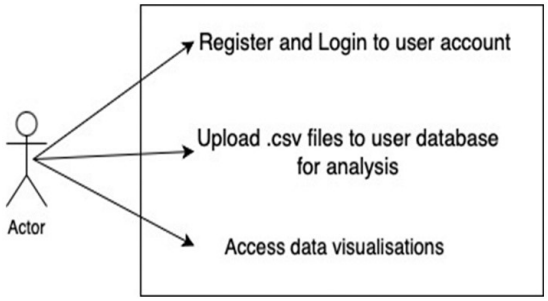


Figure 1. Use Case Analysis Diagram

International Journal of Education, Science, Technology and Engineering, vol. 8, no. 2, pp. 82-97, December 2025. **DOI: 10.36079/lamintang.ijeste-0802.887**

Relationships

1) User: The user is created, and their information is stored in the Users table.
2) Dataset: A user can have multiple datasets, each containing information and a corresponding date.
3) Charts: The combination of datasets yields different charts, allowing one dataset to have multiple charts.

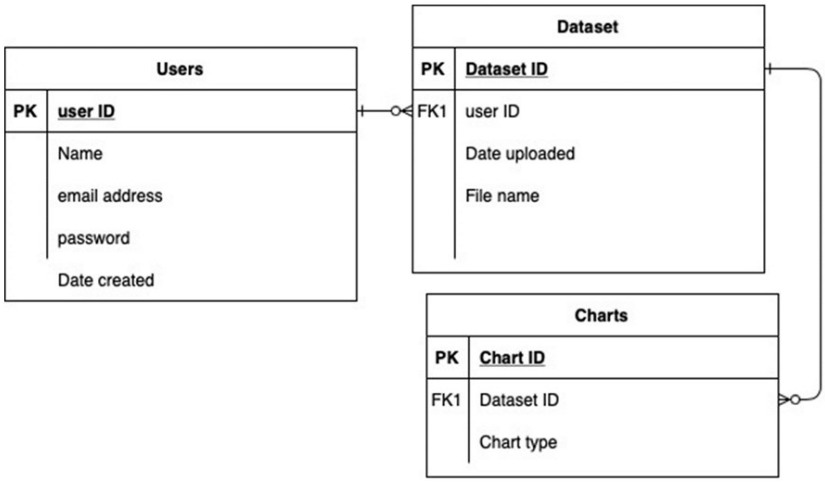


Figure 2. Entity Relationship Diagram

The proposed Tax Data Analytics System is to be implemented in a monolithic architecture, which combines the Application layer and logic into a single container linked to the database. This will allow for ease of development and testing.

1) Application Layer: This deals with the logic and business intelligence part of the application. The processes to transform data into visuals and security functions, like authentication.
2) Service Layer: This handles the processing of tasks such as routing, input validation, and other key features essential to a functioning application.
3) Data Layer: This handles data storage, retrieval, and management. It acts as the middleman between application logic and the database.

All of the layers work in tandem to achieve:

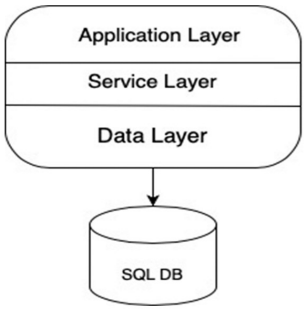


Figure 3. Monolithic System Design Diagram

International Journal of Education, Science, Technology and Engineering, vol. 8, no. 2, pp. 82-97, December 2025. **DOI: 10.36079/lamintang.ijeste-0802.887**

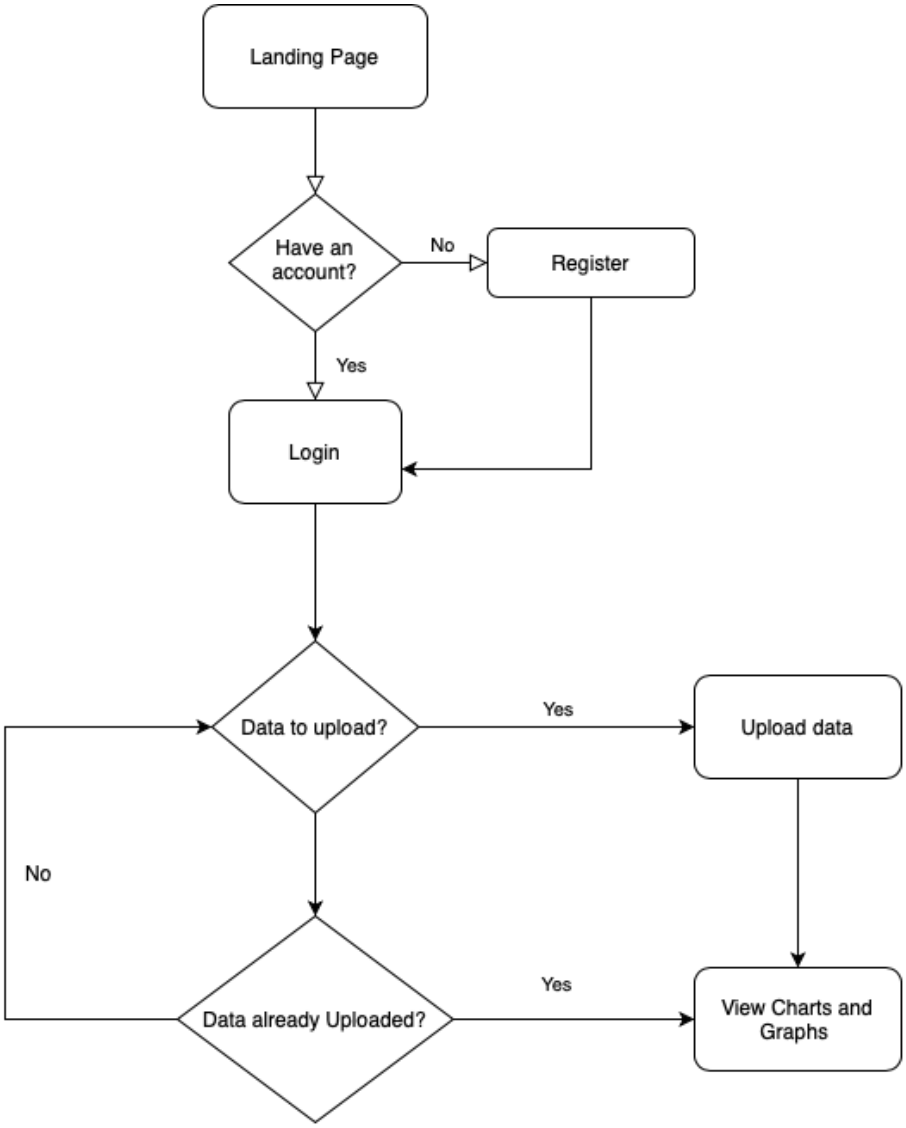


Figure 4. Flow Chart of Application

4. Findings and Discussion

4.1. Overview

The designed system was successfully implemented to meet all the study's objectives. The Flask application enables user registration on the platform and corresponding logins to grant access. On uploading the corresponding documents with tax information, the data is immediately stored in the connected SQLite database. From that database, data analysis begins by making appropriate charts and visualizations. The front end of the application was developed using classic HTML and CSS with the addition of Bootstrap frameworks. Given that the application was primarily built from the back-end using Python, that is where the bulk of the focus lies. The back end was implemented using Python and Flask. Additional libraries included Pandas and Matplotlib for data processing and visualization, Flask-bcrypt for authentication, and SQLite3 for the database.

4.2. Testing

To test the application's functionality, a sales report from a small business was used to determine what data can be gathered. The application should be able to:

1) Register and log in users.
2) Accept an upload of an Excel file and transfer it to the database.
3) Develop effective data visualizations that enable users to draw conclusions from the data.

International Journal of Education, Science, Technology and Engineering, vol. 8, no. 2, pp. 82-97, December 2025. **DOI: 10.36079/lamintang.ijeste-0802.887**

4.3. User Registration and Logging In

The first step for any user is to register on the platform as a means to secure their data and have a private platform on which they can begin to make use of the capabilities of the analytics platform. After which, they will be met with the login page to fill in their credentials and access the platform.

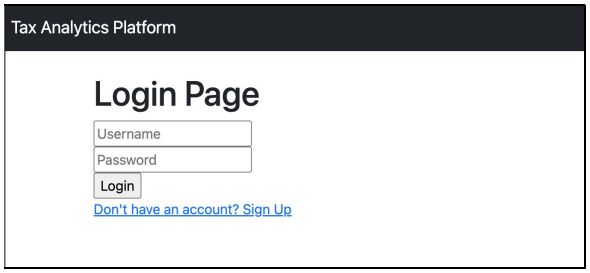


Figure 5. User Login Page

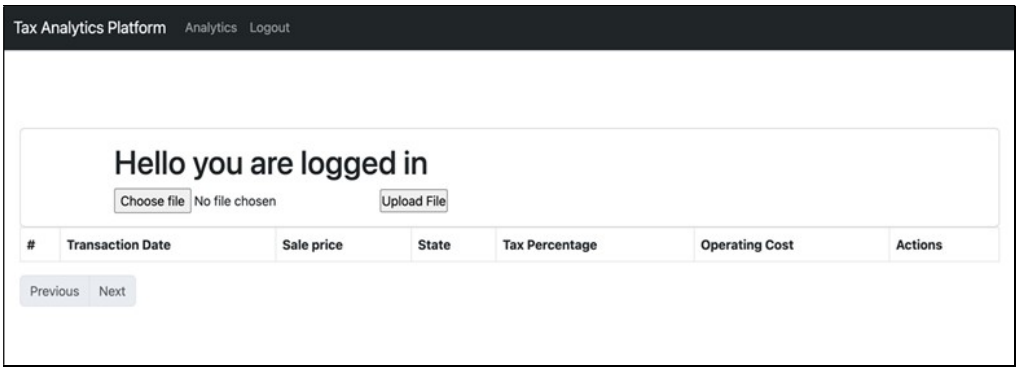
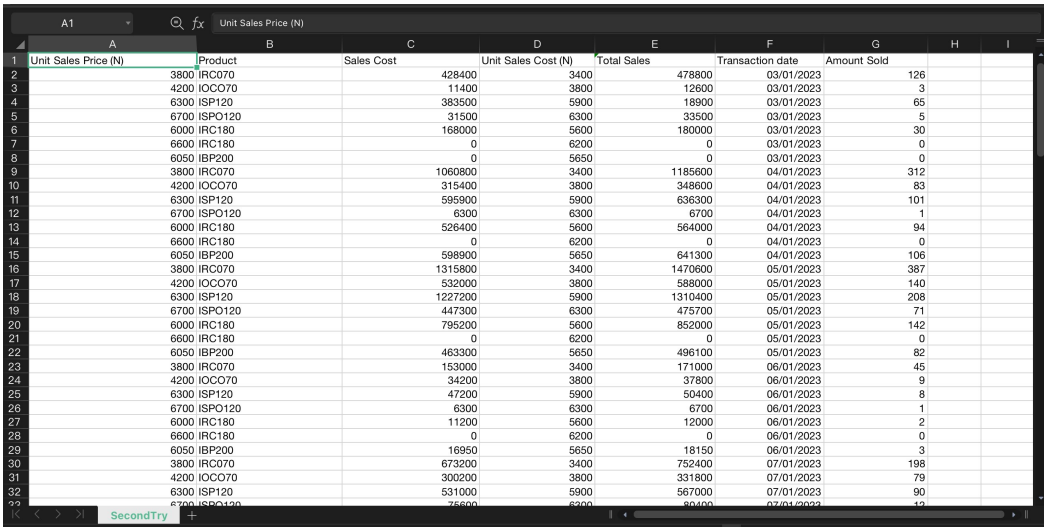


Figure 6. User Login Confirmation Page with A Prompt to Upload the Sales Data File



Unit Sales Price (N)	Product	Sales Cost	Unit Sales Cost (N)	Total Sales	Transaction date	Amount Sold
3800	IRC070	428400	3400	478800	03/01/2023	126
4200	IOCO70	11400	3800	12600	03/01/2023	3
6300	ISP120	383500	5900	18900	03/01/2023	65
6700	ISPO120	31500	6300	33500	03/01/2023	5
6000	IRC180	168000	5600	180000	03/01/2023	30
6600	IRC180	0	6200	0	03/01/2023	0
6050	IBP200	0	5650	0	03/01/2023	0
3800	IRC070	1060800	3400	1185600	04/01/2023	312
4200	IOCO70	315400	3800	348600	04/01/2023	83
6300	ISP120	595900	5900	636300	04/01/2023	101
6700	ISPO120	6300	6300	6700	04/01/2023	1
6000	IRC180	526400	5600	564000	04/01/2023	94
6600	IRC180	0	6200	0	04/01/2023	0
6050	IBP200	598900	5650	641300	04/01/2023	106
3800	IRC070	1315800	3400	1470600	05/01/2023	387
4200	IOCO70	532000	3800	588000	05/01/2023	140
6300	ISP120	1227200	5900	1310400	05/01/2023	208
6700	ISPO120	447300	6300	475700	05/01/2023	71
6000	IRC180	795200	5600	852000	05/01/2023	142
6600	IRC180	0	6200	0	05/01/2023	0
6050	IBP200	463300	5650	496100	05/01/2023	82
3800	IRC070	153000	3400	171000	06/01/2023	45
4200	IOCO70	34200	3800	37800	06/01/2023	9
6300	ISP120	47200	5900	50400	06/01/2023	8
6700	ISPO120	6300	6300	6700	06/01/2023	1
6000	IRC180	11200	5600	12000	06/01/2023	2
6600	IRC180	0	6200	0	06/01/2023	0
6050	IBP200	16950	5650	18150	06/01/2023	3
3800	IRC070	673200	3400	752400	07/01/2023	198
4200	IOCO70	300200	3800	331800	07/01/2023	79
6300	ISP120	531000	5900	567000	07/01/2023	90

Figure 7. A partial View of The Excel File Uploaded for Analysis

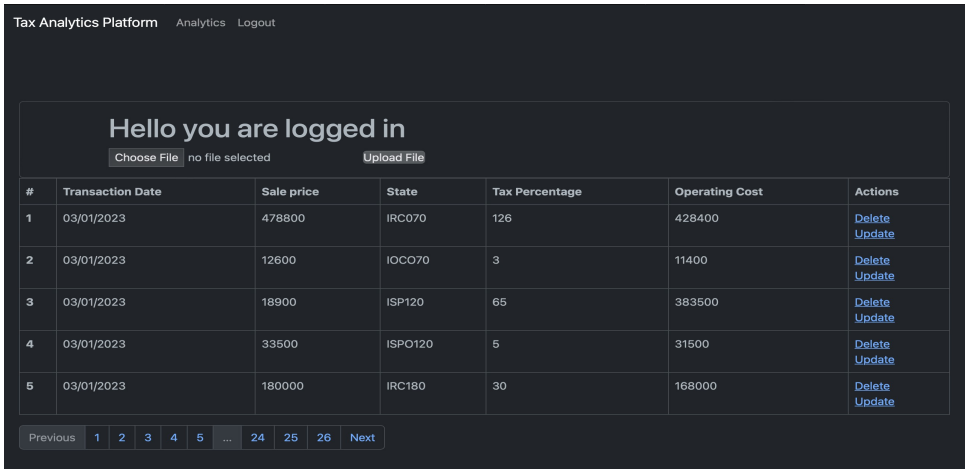


Figure 8. Uploading the File Automatically Inserts the Table into the Database

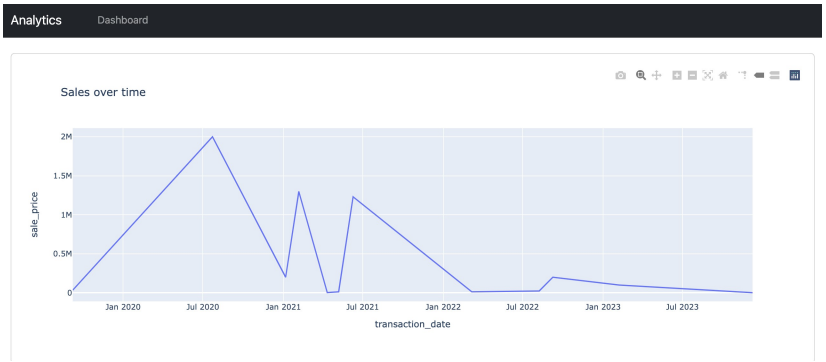


Figure 9. Distribution of Sales by States and Profit Distribution or Margins of Profit

4.4. Discussion

The system, therefore, transitions tax monitoring from a passive, compliance-only function to an active decision-support tool.

1) Improved Understanding of Revenue Structure

Analysis of the dataset showed that corporate tax accounted for 47.8% of total tax revenue. This high contribution highlights the financial significance of corporate-related tax streams for SMEs.

- Business implication
  Companies can plan better tax strategies, anticipate burdens, and adjust financial priorities.
- Objective link
  The system successfully provides a structural breakdown of tax sources, enhancing clarity in revenue monitoring.

2) Seasonal Sales Tax Patterns Identified

Visualization revealed clear seasonal patterns, with Q2 and Q4 showing peak sales tax remittances. Patterns like that match with usual SME economic cycles.

- Strategic effect
  Businesses can get ready for sought-after seasons, modify operations, and watch for large tax obligations.
- Analytical valuation
  EDA methods uncover patterns that cannot be seen through a raw spreadsheet exploration.

International Journal of Education, Science, Technology and Engineering, vol. 8, no. 2, pp. 82-97, December 2025. **DOI: 10.36079/lamintang.ijeste-0802.887**

3) Enhanced Accuracy in Insights (12% Improvement)

The platform got a 12% advancement in insight precision in contrast to manual spreadsheet tracking. This progress comes from standardized preprocessing, ordered data cleaning, and removal of error by human.

- Efficiency benefit
  Self-operating classification and descriptive statistics lessen irregularities.
- Reason for Research
  This confirms the need for digital transformation backed by analytical intelligence.

4) Enlarged Transparency and Accountability

The platform's visual dashboards, encompassing line charts, bar graphs, summaries, and trend analyses, amplify transparency by letting tax information to be easier to interpret. Formerly, managers relied on firm spreadsheets, which restricted visibility.

The new visualization model presents:

- Real-time insights
- Clear visual communication
- Refined accountability for reporting

5) Support for Data-Driven Business Decisions

Insights generated by the system support:

- Budget forecasting
- Tax obligation planning
- Sales performance evaluations
- Detection of seasonal cycles
- Identifying peak vs. low-performing periods

A tax analysis system was successfully designed and developed. It allows users or companies to provide tax and revenue information to receive data visualizations and insights. The breakdown of information enables users to properly analyze the given data and use it to draw conclusions or make informed business decisions. All these were the initial problems that were set out to be solved, and they have all been achieved.

From a scientific point of view, the research adds to the progressing knowledge in tax informatics by showing a practical, data-driven analytical structure based on EDA concepts. While most of the research concentrated on digitalisation and high-level tax reforms, this study gives a hands-on, transaction-level analytical method adapted for SMEs, a part that has not yet been mapped out. The developed system confirms the work of EDA as an intelligence layer in today's tax conditions, backing scholarly claims that digital tax infrastructures get to their peak performance only when amplified with analytics-driven knowledge. Through systematic preprocessing, statistical profiling, correlation analysis, and visualization techniques, this research strengthens the dispute that EDA is foundational to refining monitoring accuracy, spotting abnormalities, and aiding strategic planning in revenue systems.

Nevertheless, the study also proffers specific restrictions. The outstanding problem was the Flask SQLite addition, where compatibility problems rise due to the previous SQLite version and differences in documentation at hand and the features needed for the system's logic layer. These restrictions affected workflow efficacy and lessened the database's flexibility, pointing out that later iterations would help from changing to more robust relational database systems like PostgreSQL or MySQL. Also, while the application excelled on sample datasets, the lack of huge real-world organisational data restricts the limit to which effectiveness, reliability, and generalisability can be fully confirmed. The system's present analytical potential concentrates primarily on descriptive and diagnostic EDA; predictive or prescriptive modelling was not designed, signaling that the system does not automate prediction or behavioural prediction yet.

5. Conclusion

The research looked forward to developing a tax analytics scheme that lets users, businesses, or corporations gain perceptions from their tax and sales data. The proposed approach utilized Python and the Flask web framework to implement the ideas within a monolithic architecture that combined the logic and functionalities of the application.

In spite of these restrictions, the outcomes gotten through controlled testing confirm the primary function, usability, and analytical correctness of the system. Visualisations productively showed anticipated statistical structures, anomaly detection acted continually with scheduled thresholds, and summaries of data matched with human verification tests. This confirms that the system offers a reliable EDA pipeline for typical SME tax datasets and can serve as a functional prototype for broader adoption.

Recommendations are:

1) Real-World Deployment
   Having seen the application reach this level through testing with sample data, it would be interesting to see how it performs against real organisations or small businesses' data and how helpful it can be to other businesses.
2) Predictive Analysis
   Possibly including a way that trends could automatically be predicted, even though it is clear that those predictions can only be so accurate, especially when dealing with unpredictable external forces that the system can't always account for.

References

[1] I H. Sarker, "Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective", *SN Computer Science,* vol. 2, no. 5, p. 377, Jul. 2021.

[2] S. Rahman, R. S. Khan, M. R. M. Sirazy, and R. Das, "An Exploration of Artificial Intelligence Techniques for Optimizing Tax Compliance, Fraud Detection, and Revenue Collection in Modern Tax Administrations", *International Journal of Business Intelligence & Big Data Analytics,* vol. 2, no. 1, pp. 12-25, Mar. 2024.

[3] R. Belahouaoui and J. Alm, "Tax Fraud Detection Using Artificial Intelligence-based Technologies: Trends and Implications", *Journal of Risk and Financial Management,* vol. 18, no. 9, pp. 1-26, Sep. 2025.

[4] J. -F. Rodríguez-Quintero, A. Sánchez-Díaz, L. Iriarte-Navarro, A. Maté, M. Marco-Such, and J. Trujillo, "Fraud Audit Based on Visual Analysis: A Process Mining Approach", *Applied Sciences*, vol. 11, no. 11, p. 4751, May 2021.

[5] C. O. Omodero and G. Ekundayo, "The Road to Tax Collection Digitalization: An Assessment of the Effectiveness of Digital Payment Systems in Nigeria and the Role of Macroeconomic Factors", *International Journal of Financial Studies,* vol. 13, no. 3, p. 178, Sep. 2025.

[6] H. E. Inegbedion, "Acceptance of Electronic Tax System and the Contribution of SMEs to Tax Revenue through the Mediation of Tax Compliance", *Journal of Innovation and Entrepreneurship,* vol. 14, no. 129, Oct. 2025.

[7] Y. A. Olumoh and M. Sanni, "Tax Management Practices and Performance of State Internal Revenue Services in Southwest, Nigeria: Moderating Role of Information Technology Adoption", *Malete Journal of Accounting and Finance,* vol. 5, no. 2, Jun. 2025.

[8] F. Olasunkanmi and J. Adejuwon, "Digitalisation of Tax System and Revenue Generation in Nigeria", *Lead City Journal of The Social Sciences,* vol. 9, no. 3, Dec. 2024.

[9] Y. He, "Digital Technology Reshapes the Tax Collection and Administration System: Impacts, Challenges and Future Prospects", *Scientific Journal of Economics and Management Research,* vol. 6, no. 9, pp. 85-91, Sep. 2024.

[10] OECD, "*Tax Administration Digitalisation and Digital Transformation Initiatives",* Jun. 2025. [Online]. Available: https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/06/tax-administration-digitalisation-and-digital-transformation-initiatives_87f43a78/c076d776-en.pdf. [Accessed: Aug. 2025].

[11] S. A. Ogbaisi and M. Ibifunmilola, "Digitalization of Tax Systems and Tax Reforms in Nigeria: A Conceptual Discourse", *Malete Journal of Accounting and Finance,* vol. 5, no. 2, Jun. 2025.

[12] S. O. Pinto and V. A. Sobreiro, "Literature review: Anomaly detection approaches on digital business financial systems", *Digital Business,* vol. 2, no. 2, p. 100038, Aug. 2022.

[13] OECD, "*Tax Revenue Buoyancy in OECD Countries",* Revenue Statistics, Dec. 2023. [Online]. Available: https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/12/revenue-statistics-2023_6ee814f4/9d0453d5-en.pdf. [Accessed: Aug. 2025].

[14] M. Nose, N. Pierri, and J. Honda, "Leveraging Digital Technologies in Boosting Tax Collection", *International Monetary Fund,* May 2025.

[15] D. K. Srivastava, M. Bharadwaj, T. Kapur, and R. Trehan, "Seasonal Concentration of Fiscal Aggregates: Case Study of India", *Modern Economy*, vol. 16, no. 9, Sep. 2025.

[16] P. Kaushik, M. Brahmi, S. Kakran, and P. Kansra, "The Predictive Grey Forecasting Approach for Measuring Tax Collection", *Journal of Risk and Financial Management*, vol. 17, no. 12, p. 558, Dec. 2024.

[17] OECD, "*Comparative Information on OECD and other Advanced and Emerging Economies", Tax Administration, Nov.* 2024. [Online]. Available: https://www.oecd.org/content /dam/oecd/en/publications/reports/2024/11/tax-administration-2024_5c4606e4/2d5fba9c-en.pdf [Accessed: Aug. 2025].

[18] Julien Bouissou, "Satellite Imagery in Senegal assists in property tax collection", *Economy (Senegal),* Aug. 2025.

[19] B. Ván, G. Lovics, C. G. Toth, and K. Szőke, "Digitalization against tax evasion: evidence on the role of company size", *Journal of Business Economics,* vol. 95, no. 5, pp. 749-776, Feb. 2025.

[20] G. Ilieva, T. Yankova, M. Ruseva, Y. Dzhabarova, V. Zhekova, S. Klisarova-Belcheva, T. Mollova, and A. Dimitrov, "Factors Influencing User Perception and Adoption of E-Government Services", *Administrative Sciences,* vol. 14, no. 3, Mar. 2024.

[21] A. Badshah, A. Daud, R. Alharbey, A. Banjar, A. Bukhari, and B. Alshemaimri, "Big data applications: overview, challenges and future", *Artificial Intelligence Review,* vol. 57, no. 290, Sep. 2024.

[22] S. Chatterjee, R. Chaudhuri, S. Gupta, U. Sivarajah, and S. Bag, "Assessing the impact of big data analytics on decision-making processes, forecasting, and performance of a firm", *Technological Forecasting and Social Change,* vol. 196, p. 122824, Nov. 2023.

[23] J. Ram and C. Desgourdes, "Using big data analytics (BDA) for improving decision-making performance in projects", *Journal of Engineering and Technology Management,* vol. 74, p. 101849, Nov. 2024.

[24] Md A. Hossin, J. Du, L. Mu, and I. O. Asante, "Big Data-Driven Public Policy Decisions: Transformation Toward Smart Governance", *Sage Journals,* Dec. 2023.

[25] M. C. Huda and B. Jatmiko, "Big data analytics in the public sector: A systematic literature review", *Journal of Accounting and Investment,* vol. 26, no. 2, May 2025.

[26] X. Wu, S. X. Ying, J. You, X. Wu, and H. Wu, "The effect of big data-driven tax enforcement on audit pricing: Evidence from China", *The British Accounting Review,* p. 101531, Nov. 2024.

[27] R. Najem, A. Bahnasse, M. Fakhouri Amr, and M. Talea, "Advanced AI and big data techniques in E-finance: a comprehensive survey", *Discover Artificial Intelligence,* vol. 5, no. 102, Jun. 2025.

[28] J. Aslett, G. Gonzalez, S. Hamilton, and M. Pecho, "Tax Administration: Essential Analytics for Compliance Risk Management", *IMF Technical Notes and Manuals,* Feb. 2024.

[29] L. Yuan, B. Li, S. Li, K. K. Wong, R. Zhang, and H. Qu, "Tax-Scheduler: An interactive visualization system for staff shifting and scheduling at tax authorities", *Visual Informatics,* vol. 7, no. 2, pp. 30-40, Jun. 2023.

[30] H. Wang, J. Cheng, and H. Zhou, "Tax administration informatization, financial tightened regulation, and firm productivity: A quasi-natural experiment based on the "Golden Tax Phase III project", *International Review of Economics & Finance,* vol. 98, p. 104003 Mar. 2025.