M.SC. DATA ANALYTICS & TECHNOLOGIES

DAT7301- DATA ANALYSIS AND VISUALISATION

PORTFOLIO 1 - ASSIGNMENT 3

COMPREHENSIVE REPORT OF NIGERIAN STATES' INTERNALLY GENERATED REVENUE (2019 - 2023)

BY: OKIKE U. J.

Student ID: 2423983

INSTRUCTOR: HARINDA EUGEN

DATE. 10-05-2025

ABSTRACT

This report delivers a strategic analysis of Internally Generated Revenue (IGR) across Nigerian states from 2019 to 2023, with the aim of informing policy and financial planning decisions. The study implemented a combination of exploratory data analysis, visualisation, clustering, statistical comparison and time series forecasting to find patterns of growth, difference and future potential within Nigeria's financial setting. The data was sourced from the National Bureau of Statistics of Nigeria and analysed in R using reproducible and transparent methods. Key insights include the identification of high-performing and underperforming state clusters, statistically significant differences in average revenue and a projected national IGR exceeding ₦3 trillion by 2025. These findings provide a practical foundation for targeted financial interventions and long-term planning. Key insights include Lagos State's continued dominance in IGR generation, statistically confirmed revenue clusters and a projected increase in national IGR to over ₦3 trillion by 2025. The report demonstrates how data-driven methods can inform differentiated financial strategies and future planning.

Keywords;

IGR, Nigerian States, Revenue Forecasting, Clustering, ETS Model, ANOVA, Financial Analysis, R Programming.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. EXECUTIVE SUMMARY

This report presents a concentrated and evidence-based analysis of Internally Generated Revenue (IGR) trends and projections across Nigeria's 37 states from 2019 to 2023. Developed with the needs of financial policymakers, planners and institutional stakeholders in mind, the findings offer an objective view of both performance differences and growth potential within Nigeria's sub-national revenue landscape.

With tools such as clustering, ANOVA and ETS forecasting, the report identifies systemic inequalities, emerging high-growth regions and reliable predictors for future revenue performance. Designed to support data-driven planning and resource allocation, the analysis concludes in strategic recommendations that are practical, validated and designed for different levels of state performance.

Key findings reveal that while national IGR has steadily increased, surpassing ₦2.3 trillion in 2023, revenue remains highly concentrated, with Lagos State consistently outperforming all others. K-means clustering grouped states into three distinct categories based on revenue performance and growth. Subsequent statistical tests confirmed these clusters to be significantly different in average IGR. Time series forecasting using an ETS model projects that Nigeria's IGR will exceed ₦3 trillion by 2025, assuming continued growth momentum. These insights are vital for informing differentiated financial policy strategies, improving state-level revenue mobilisation and addressing structural disproportions in Nigeria's financial condition.

2. METHODOLOGY

The dataset used in this analysis was sourced from the National Bureau of Statistics' publicly available IGR reports, which compile yearly state-level revenue figures across five years (2019 - 2023). Each year's data was extracted from Excel spreadsheets and programmatically loaded using the `readxl` package. Preprocessing steps included standardising inconsistent state names (e.g., "Federal Capital Territory" to "fct", Nassarrawa to Nasarawa) and combining yearly records into a final unified tidy dataset (`gr_total`). The subsequent structure contained three columns; `state`, `year` and `total` (total IGR in Nigerian naira).

2.1 Exploratory Data Analysis

Series of visualisations were created to uncover broad trends and regional disparities including:

- Line charts were used to examine the national IGR trajectory.
- Faceted bar plots and histograms highlighted the distribution and inequality of IGR at the state level.
- A correlation heatmap visualised pairwise similarity in state IGR over time.

2.2 Clustering and Statistical Testing

K-means clustering was applied to group states based on two metrics:

- Average IGR from 2019 to 2023
- Percentage growth over the same period

The elbow method confirmed k = 3 as the best and optimal number of clusters. To statistically validate the clusters, a one-way ANOVA was performed, followed by Tukey's HSD post-hoc test to identify which pairs of clusters significantly differed in their average IGR levels.

2.3 Time Series Forecasting

The national IGR was aggregated yearly and transformed into a time series object using the `ts()` function. An Exponential Smoothing (ETS) model was fitted using the `forecast` package. Forecasts were generated for 2024 and 2025 with 95% confidence intervals to estimate Nigeria's expected IGR under current growth trends. All analysis was carried out in RStudio using RMarkdown for full reproducibility.

3. KEY FINDINGS AND INSIGHTS

3.1 National Revenue Growth

Figure 1: Line plot showing total national IGR from 2019 to 2023.

From 2019 to 2023, Nigeria's total IGR increased from ₦1.6 trillion to over ₦2.3 trillion, indicating an aggregate growth rate of over 40%. A slight decline in 2020 is likely attributable to the COVID-19 pandemic, followed by a strong recovery in 2021 - 2023. Visual analysis confirmed this trend through smooth upward lines, supporting the notion of macroeconomic resilience.

3.2 Inter-State Revenue Inequality

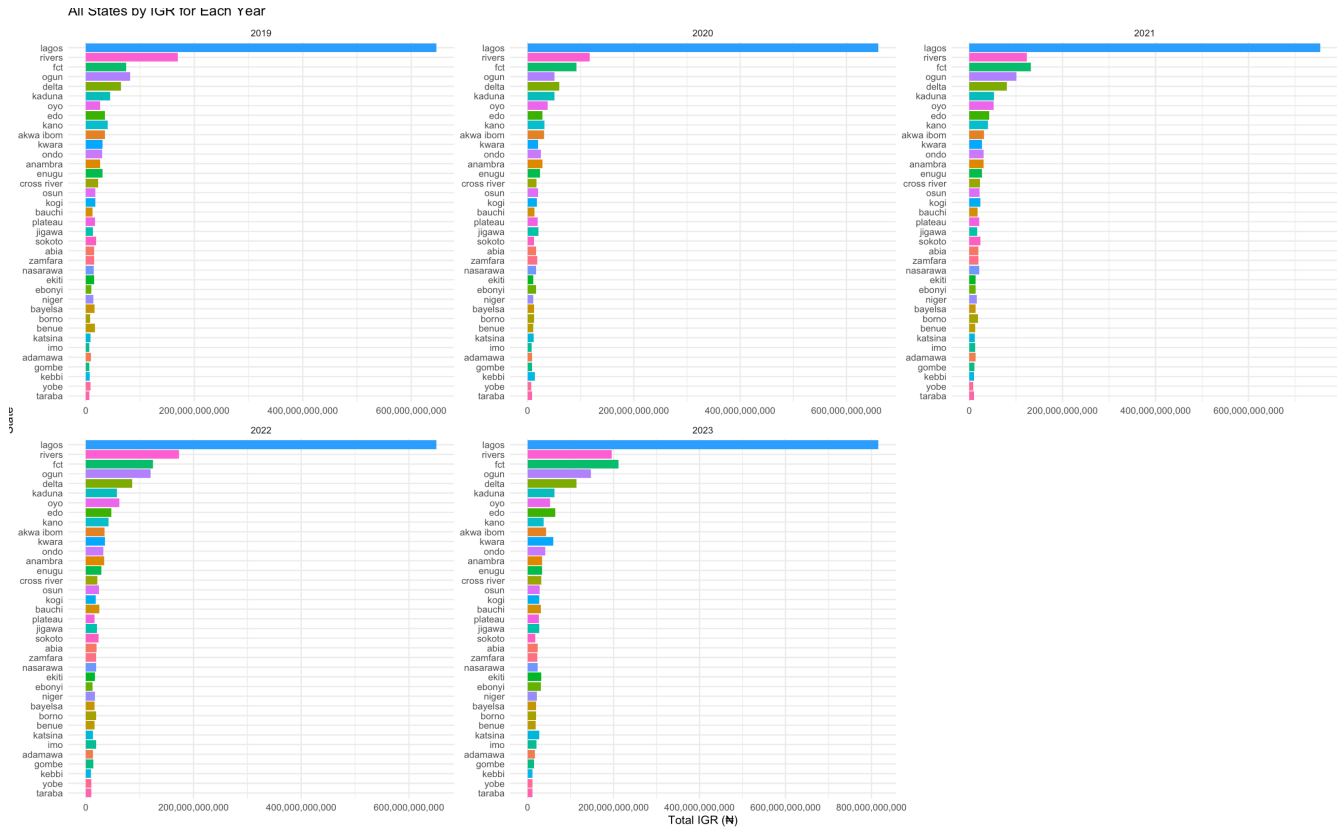

Figure 2: Faceted bar chart of state-by-state IGR values for each year (2019 - 2023).

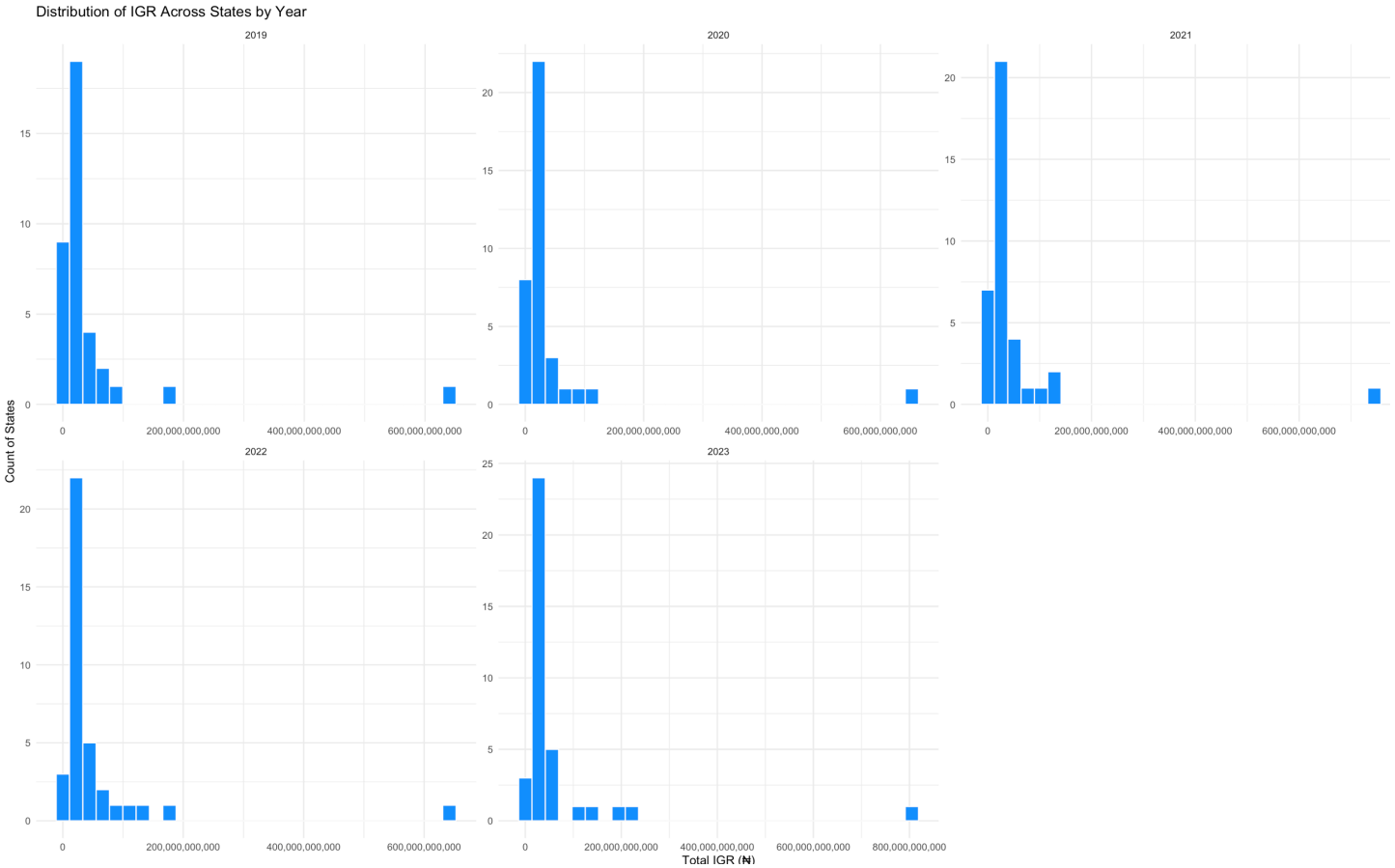

Figure 3: Histograms showing yearly skewed distribution of state-level IGRs.

Bar charts and histograms consistently revealed extreme financial disparity. Lagos State generated the highest IGR by a wide margin, accounting for a disproportionately large share of national totals each year. In contrast, most other states recorded modest revenues, typically below ₦100 billion annually. This right-skewed distribution reflects both structural economic imbalances and inconsistencies in tax mobilisation capacity.

3.3 Correlation Patterns

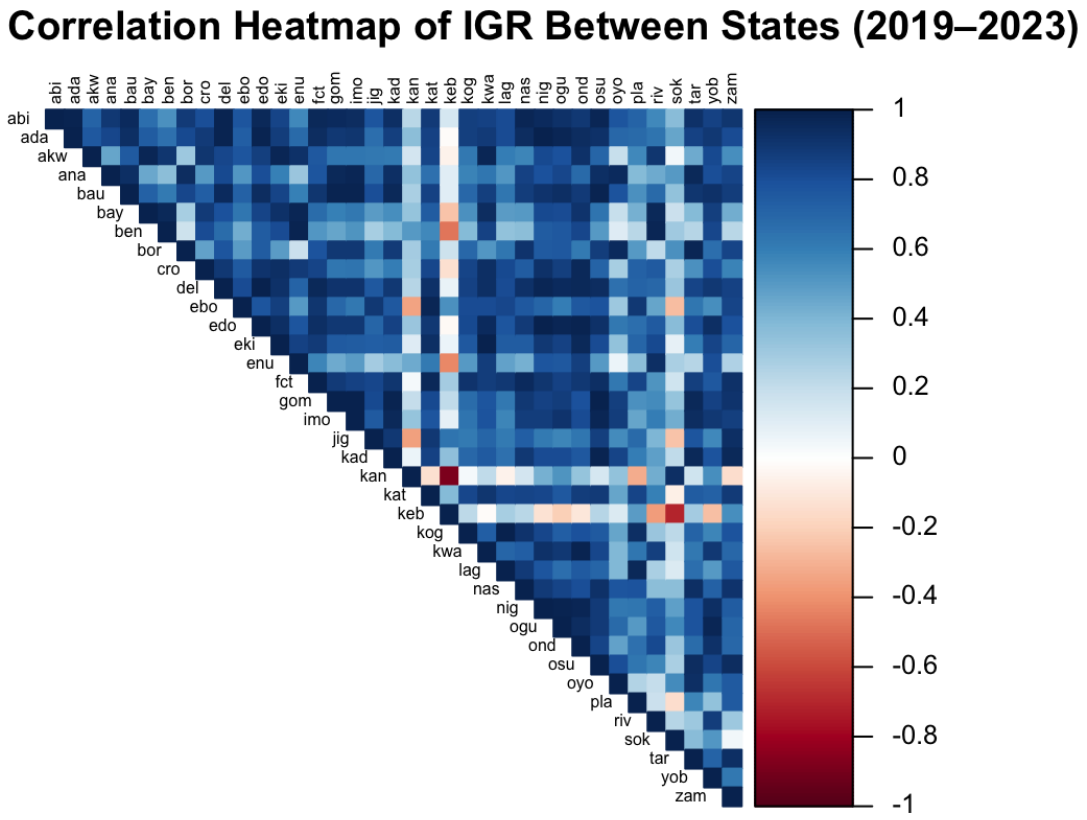

Figure 4: Correlation heatmap of state-level IGR patterns from 2019 – 2023.

The correlation heatmap indicated that some states' IGR movements are closely linked, potentially reflecting shared economic drivers or tax policy alignment. Conversely, weaker or inverse correlations may reflect regional divergence, external shocks or state-level inadequacies.

3.4 Clustering of States

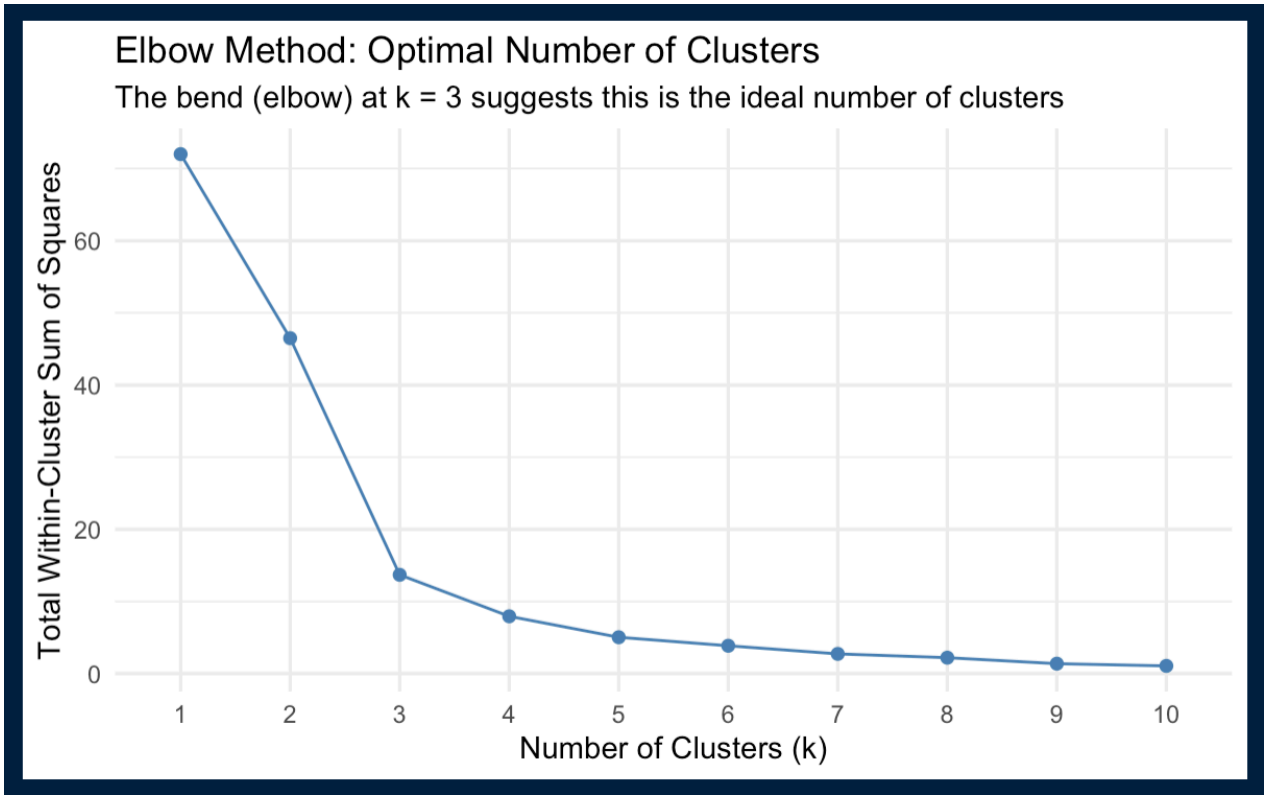

Figure 5: Elbow method plot identifying optimal number of clusters (k = 3).

Figure 6: Cluster scatter plot of scaled average IGR vs. growth (labelled by state).

K-means clustering effectively segmented states into three groups:

- Cluster 1: Lagos (outlier with high IGR and slow growth)
- Cluster 2: Moderately performing but high-growth states (e.g., FCT, Ebonyi, Katsina)
- Cluster 3: Low IGR, low or negative growth states (majority)

This grouping allows policymakers to tailor financial strategies by state performance.

3.5 Statistical Comparison

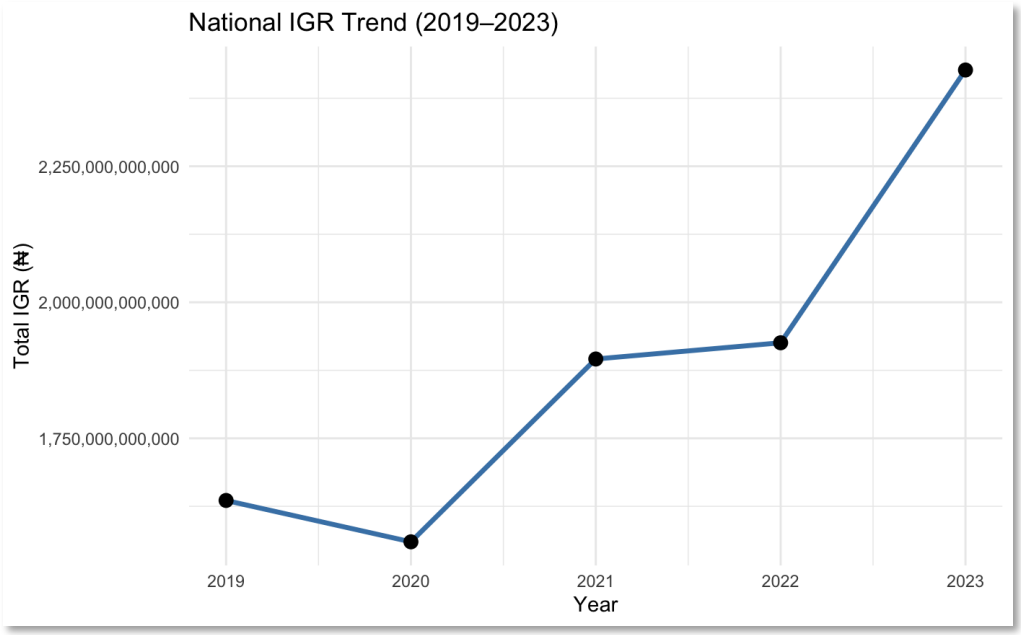

Figure 7: Boxplot comparing average IGR by cluster group.

Source	**Df**	**Sum of Squares**	**Mean Square**	**F Value**	**p-value**
Cluster	2	4.403×10^{23}	2.202×10^{23}	197.5	< 0.001
Residuals	34	3.789×10^{22}	1.115×10^{21}	-	-

Table 1: ANOVA test results comparing IGR between clusters.

Comparison	Mean Difference (₦)	Lower Bound	Upper Bound	Adjusted p-value
2 - 1	-677,109,108,513	-763,341,157,630	-590,877,059,396	< 0.0001
3 - 1	-671,024,625,315	-754,332,699,404	-587,716,551,225	< 0.0001
3 - 2	+6,084,483,198	-25,403,009,122	+37,571,975,519	0.8842

Table 2: Tukey HSD test results comparing IGR between clusters.

The ANOVA test confirmed significant differences in average IGR across the clusters ($F(2,34) = 197.5$, $p < 0.001$). While Tukey's post-hoc test revealed that Lagos (Cluster 1) significantly outperformed both Cluster 2 and Cluster 3 in mean IGR. However, no significant difference was found between Clusters 2 and 3, suggesting similar overall revenue levels despite differing growth trends.

3.6 Forecasting National IGR

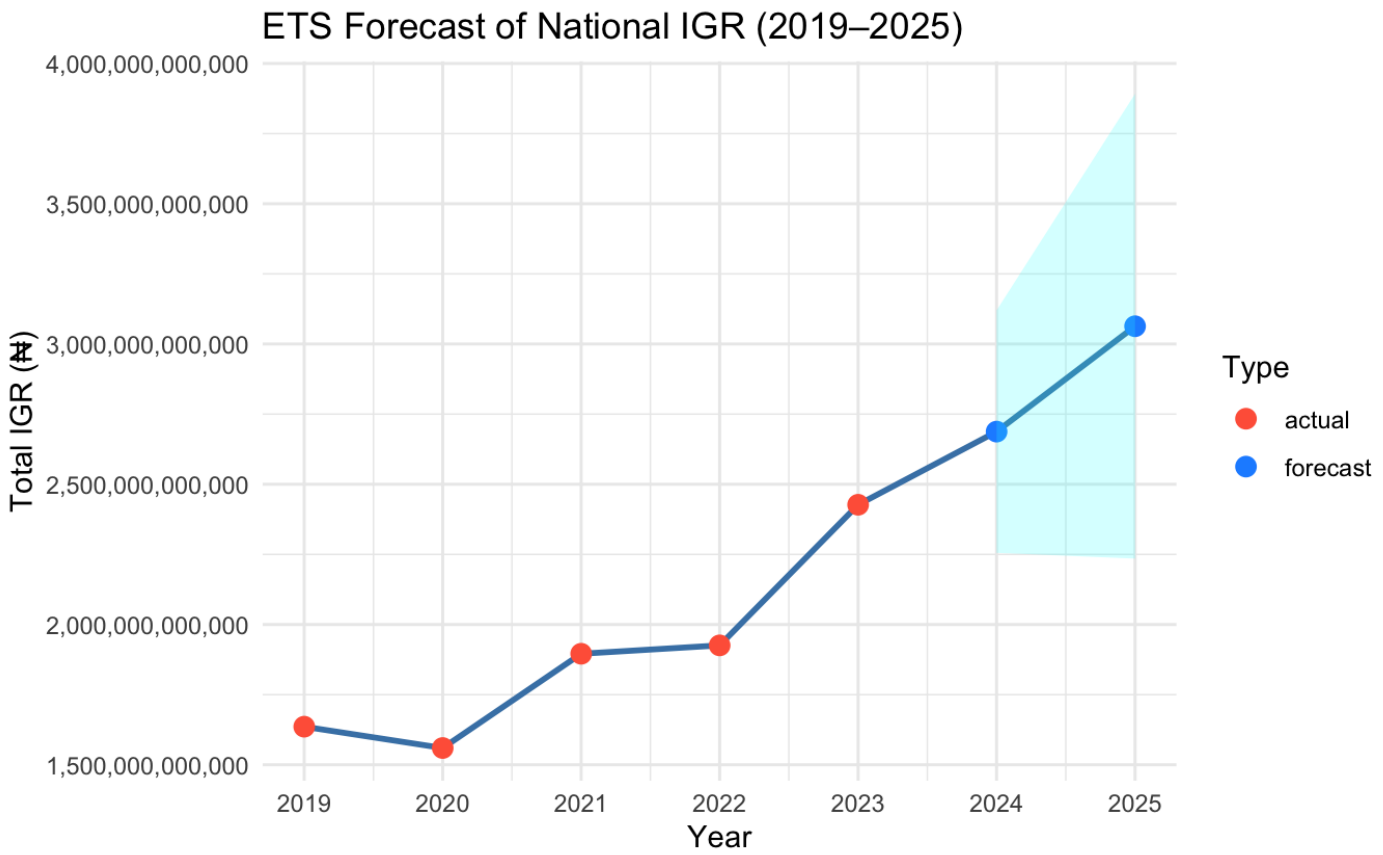

Figure 8: ETS model forecast of national IGR for 2024 and 2025 with 95% confidence interval.

The ETS forecast projected continued revenue growth into 2024 and 2025, with the total IGR expected to reach ₦3.1 trillion by 2025. The forecast plot demonstrated a smooth trajectory, with narrow confidence bands suggesting high model certainty. This trend indicates strong revenue momentum, excluding any economic disruption.

4. ETHICAL CONSIDERATIONS

Several ethical considerations were taken into account during this analysis. Firstly, the use of publicly available government data ensures transparency and compliance with open data standards. Measures were taken to ensure and preserve the accuracy of reported values, particularly during the data cleaning process, where state names and year alignments were standardised.

Secondly, the visualisation of disparities between states was handled cautiously to avoid misrepresentation. Lagos, due to its outlier status, was clearly labelled and discussed separately to ensure fair representation of other states. Where appropriate, scaling techniques were applied to maintain comparability across plots.

Thirdly, ethical data interpretation was prioritised by ensuring that statistical results were not overstated. For example, while clustering showed clear groupings, the underlying assumptions and limitations were acknowledged, such as the similarity in average IGR between Clusters 2 and 3 despite different growth patterns.

Finally, forecasting was carried out using a validated and context-aware model (ETS). It was visualised with confidence intervals to communicate the inherent uncertainty in prediction. No computations were made beyond a reasonable 2-year forecast window. All modelling and visual outputs were produced with reproducibility and transparency in mind.

5. CONCLUSION

This portfolio extends the findings of Assignment 2 by integrating descriptive, statistical and forecasting methods to deliver actionable insights for stakeholders engaged in public finance and financial planning. The analysis not only reveals key performance indicators but also provides a structured perspective through which state-level revenue mobilisation can be assessed and improved.

From a strategic viewpoint, the persistent dominance of Lagos and the emerging strength of certain mid-tier states indicate clear opportunities for both benchmarking and targeted investment. Meanwhile, the significant group of underperforming states signals a need for deep systemic reforms.

The national IGR forecast offers a forward-looking tool to support budget planning and resource allocation across ministries, departments and agencies. With tailored recommendations grounded in validated models and evidence-based clustering, this report equips stakeholders with practical guidance for enhancing financial sustainability across Nigeria's subnational units.

RECOMMENDATIONS

- Prioritise support for Cluster 2 states showing strong growth.
- Provide foundational reform and capacity-building in underperforming Cluster 3 states.
- Benchmark Lagos for best practices while addressing its growth slowdown.
- Encourage nationwide adoption of forecasting tools for revenue planning.
- Promote the adoption of forecasting tools and financial modelling for data-driven governance.

This report reaffirms the value of integrating data science techniques in public sector analysis and offers a reproducible framework for ongoing financial evaluation.

Word Count;

The report contains approximately 1402 words, figures and tables.