ACCSCIENCE PUBLISHING

International Journal of Population Studies

RESEARCH ARTICLE

Beyond the census: Satellite and tax data for ranking South African municipalities

Lodewalt Venter[1]*, Clive Coetzee[2], and Waldo Krugell[1]

[1]Department of Economics, School of Economics, Faculty of Economics and Management Sciences, North-West University, Potchefstroom, North West, South Africa

[2]Department of Economics (Mil), Faculty of Military Science, Stellenbosch University, Saldanha, Western Cape, South Africa

Abstract

This study comprehensively analyzes South Africa's municipal rankings by employing unconventional data sources – specifically, satellite-derived and administrative tax data – to supplement traditional census counts. In light of recent controversies surrounding the accuracy of the 2022 Census data, we assessed the viability of alternative data sources in ensuring equitable and accurate municipal funding. We utilized data from South Africa's 2022 Census, Oak Ridge National Laboratory's LandScan 2022, Google Earth Engine's nightlight database, the South African Revenue Service – National Treasury (SARS-NT) spatial tax panel, and the Regional Explorer database to calculate rank-size distributions and Spearman rank correlations across 213 municipalities. Results indicate exceptionally high concordance between census data and satellite-derived population estimates (Spearman's $\rho = 0.98$), underscoring the robustness of satellite data in capturing demographic distributions. In addition, formal economic activity, as represented by SARS-NT establishment data, exhibits significant concentration, with metropolitan municipalities accounting for 67% of formal firms. Our findings also reveal that the population distribution among municipalities aligns closely with Zipf's rank-size rule, with modest deviations indicating slight primacy in larger metropolitan areas. This analysis advocates for integrating satellite-derived growth factors into fiscal allocation models and establishing real-time economic monitoring using satellite night lights. These measures promise to enhance intergovernmental transfer responsiveness, equity, and accuracy, directly reinforcing sustainable urban development and governance that supports Sustainable Development Goal 11. The study further recommends annual recalibration of municipal population data using satellite-based adjustments to address funding disparities and mitigate migration-induced inequities in fiscal distribution.

Keywords: Rank-size distribution; Urban economic analysis; Satellite data; Administrative tax data; South African municipalities; Population estimates

***Corresponding author:**
Lodewalt Venter
(Lodewalt.Venter@nwu.ac.za)

Citation: Venter, L., Coetzee, C., & Krugell, W. (2026). Beyond the census: Satellite and tax data for ranking South African municipalities. *International Journal of Population Studies*. 12(3):025110039. https://doi.org/10.36922/IJPS025110039

Received: March 10, 2025

1st revised: May 16, 2025

2nd revised: May 25, 2025

3rd revised: June 9, 2025

Accepted: July 1, 2025

Published online: July 22, 2025

1. Introduction

On October 10, 2023, Statistics South Africa (Stats SA) announced that the country's population had reached 62 million, and in the months that followed, a consortium of independent demographers argued that the 2022 census headline population count was overstated by roughly one million people, approximately 1.6% of the total (Kahn, 2024). This represents a mismeasurement large enough to redirect between R 1.5 billion and R 5 billion in equitable-share transfers across spheres of government (author's estimate based on the National Treasury's 2025 Budget Review). This clash between headline figures and fiscal reality underscores the importance of accurate and timely population data, motivating our exploration of satellite-derived and administrative alternatives.

By the same token, National Treasury (NT) acknowledges the funding risk: Both the 2024 Medium-Term Budget Policy Statement (MTBPS, Chapter 4) and the 2025 Budget Review (Chapter 6) refer to scenario work by the department's intergovernmental relations (IGR) division, showing that a ±1 million population adjustment would shift ≈ R 4.2 billion in formula-driven transfers – about R 1.6 billion through the local-government equitable share and a further R 2.6 billion through the population-sensitive components of the provincial equitable share (National Treasury, 2024; National Treasury, 2025). This controversy highlights how infrequent, single-shot enumerations can lead to sizable funding errors, underscoring the need for supplementary, higher-frequency population proxies – such as satellite-derived grids and administrative tax registers – to benchmark and, where necessary, correct official counts.

Economic growth or stagnation is documented in national accounts but occurs in specific places. South Africa's economic growth has been slow, resulting in a spatial pattern of unemployment and poverty. Many local governments are not financially viable, and as a result, service delivery has suffered. These challenges have finally drawn the attention of policymakers, and since mid-2024, Operation Vulindlela has set to focus on interventions at the local government level (Joffe, 2024). However, sound policymaking requires data, and the recent dispute about the accuracy of the 2022 Census data (Kahn, 2024) highlights the importance of understanding the location of people and economic activity. Comprehending the spatial characteristics and dynamics of a nation's population necessitates a thorough understanding of its spatial location, distribution, density, and size (Kleynhans & Coetzee, 2025).

Achieving sustainable development goal 11, *that is*, sustainable cities and communities, requires timely population counts and equitable, transparent data governance to reduce spatial inequalities in service delivery. Manioudis & Meramveliotakis (2022) argued that classical political-economy stages reveal that property regimes and class structures shape cities' abilities to pursue sustainable growth and resilience. In South Africa, where informal settlements often lie beyond the census's reach, combining satellite-derived estimates with administrative tax records creates a "living" data ecosystem that can underpin more just intergovernmental transfers, guide infrastructure investment, and monitor progress toward urban sustainability targets.

Fittingly, Von Fintel (2023) has recently surveyed the data landscape, specifically for South Africa's metropolitan economies, examining a range of sub-national data sources, from census to nightlight data. He compared metropolitan indicators in terms of population statistics, employment and earnings statistics, and output measures, and concluded that no "gold-standard" source of subnational data currently exists. Household survey data are calibrated to population estimates derived from infrequent census estimates. Various sources provide different estimates of employment, but the dynamics are similar. Data on remuneration and local-level gross domestic product (GDP) are subject to various biases. The South African Revenue Service – National Treasury (SARS-NT) spatial panel presents median earnings that follow similar trajectories to nightlight luminosity data, and the data appear to capture formal sector metropolitan employment reliably compared to other sources.

According to Von Fintel (2023), there is a pressing need for reliable and accessible data on local and metropolitan municipalities in South Africa, where cities are experiencing rapid growth. Consequently, researchers have worked hard over the years to extensively harness "conventional" data while simultaneously prospecting for sources of "unconventional" data to incorporate into the analysis of South Africa's urban development (Von Fintel, 2023). Coetzee (2012) and Bezuidenhout *et al.* (2021) have explored satellite and spatial tax data, respectively.

This study surveys the landscape of satellite-generated and administrative tax data, which have developed apace in the past decade. The paper presents a method for comparing different data sources by estimating the rank-size rule for South African municipalities. Much earlier research by Naudé & Krugell (2003) found that South African cities might be undersized. Recent work by Kleynhans & Coetzee (2023) concluded that a combination of the Pareto and the log-normal probability distribution is the optimal type of probability distribution to describe the size distributions of municipalities in South Africa. This paper uses the latest Stats SA Census data (2022), S&P Global's Regional Explorer (ReX) data (2022), the SARS-NT spatial tax panel (2022), and satellite data from Google Earth Engine (2022) and Oak Ridge National Laboratory's (ORNL's) LandScan (2023) to estimate rank-size distributions for South Africa.

We also estimate a Spearman rank correlation coefficient for the rankings produced from the datasets.

Our work, therefore, extends the strand of South African urban research that interrogates rank-size regularities (Naudé & Krugell, 2003; Kleynhans & Coetzee, 2025) by testing whether satellite-derived and administrative-tax indicators reproduce the same hierarchical ordering that census figures imply. Our work also advocates the following conceptual pillars for sustainable urban data governance: (i) Urban Growth and Equity: Population data distributions (*e.g.*, rank-size coefficients) directly inform fiscal allocations and service delivery priorities (the q > 1 'primacy' effect may signal over-concentration of resources in metros at the expense of smaller towns); (ii) Data Governance Models: Franke and Gailhofer (2021) identified three regimes of urban data rights – proprietary, shared/open, and civic – each with distinct implications for transparency, accountability, and participatory planning; and (iii) Spatial Inequality and Resilience: Sustainable development requires integrating data sources to capture the dynamics of informal settlements, thereby preventing service gaps that exacerbate poverty traps and undermine economic resilience.

In summary, this paper makes three key contributions to the literature on urban population estimation in South Africa. First, we conducted a comprehensive cross-comparison of five data sources – Stats SA Census 2022, ORNL LandScan 2022, Google Earth Engine, SARS-NT tax records, and the ReX database – across all 213 municipalities in South Africa. Second, we demonstrated how rank-size (q-coefficient) and rank-correlation (Spearman's ρ) metrics can quantify the supplementarity of unconventional data for decennial census counts – revealing, *inter alia*, a rank correlation of 0.98 between the Census and LandScan data, as well as comparable q-coefficients of 1.07 and 1.05 for the two population variables, respectively. Third, we translated these findings into actionable policy guidance, illustrating how the NT and provincial planners can incorporate rank-based adjustments to mitigate against inter-censal allocation volatility and address spatial inequality constraints.

The paper is structured as follows: Section 1.1 presents a brief review of the literature on city-size distribution and relevant evidence from South Africa. Section 2 describes the data sources and explains the analysis methods. Section 3 presents the results, Section 4 provides the discussion, and Section 5 concludes the paper.

1.1. Literature review

The location and concentration of firms and workers across South Africa's geographic landscape play a crucial role in the country's economic activity, growth, and prosperity (Bezuidenhout *et al.*, 2021). Metropolitan municipalities, or simply stated, major cities, generate positive "agglomeration effects" thanks to firms and households existing in proximity, thus catalyzing the process of "economic collisions," *that is*, interactions. Higher rates of economic growth improve efficiency, encourage entrepreneurship, and boost returns on investment, all of which produce economic products, such as knowledge, goods, and services (Turok, 2021).

The agglomeration of people in towns and cities ultimately leads to the concentration of economic activity in urban areas, which are generally more conducive to economic growth and development due to localization and urbanization economies (Naudé & Krugell, 2003). South Africa's cities can significantly enhance the country's economic growth prospects by enabling economies of scale and fostering positive externalities – key benefits of agglomeration (Turok, 2021).

Naudé & Krugell (2003) found that South Africa's cities were primarily offering urbanization economies due to their less-than-optimal size, which meant that, at the time, greater long-term economic growth could likely have been achieved with city-growth policies, *that is*, "urbanization stimulation." Naudé & Krugell (2003) further noted that each country has an optimal city-size set due to the onset of diminishing returns to localization and urbanization economies. Their estimated rank-size distribution explained the sizes of local and metropolitan municipalities; however, based on the existing evidence at the time, South Africa's cities were, in fact, too small.

This article partially builds on the work done by Kleynhans and Coetzee (2022a), who conducted a systematic empirical analysis based on population data of the rank-size distribution of South Africa's local and metropolitan municipalities from 1996 to 2016. They assert that understanding such rank-size distribution dynamics may benefit policymakers in efficiently and equitably delivering public goods and services and aid in "smart" development. This is especially important considering that the rural-urban arrangement in South Africa has undergone a dramatic reconfiguration due to rapid urbanization over the past three decades (Kleynhans & Coetzee, 2022a).

The empirical evidence on the rank-size distributions of municipalities reveals a complex picture. While some studies support the applicability of Zipf's Law or the rank-size rule (Veneri, 2016), others challenge its universality. González-Val (2012) argued that the relationship has been observed in many countries, but its validity depends on factors, such as municipality definition, sample size, and estimation method. City-size distributions often follow a log-normal distribution in the main body, with a Pareto distribution in the upper tail (Wang & Sun, 2024). The mean Pareto exponent across countries was found to be 1.136, slightly higher than the value of 1 implied by the rank-size rule (Rosen & Resnick, 1980).

Factors, such as socioeconomic development, urbanisation levels, and geographic characteristics influence municipality size distributions (Rosen & Resnick, 1980; Wang & Sun, 2024). In addition, the definition of cities and sample sizes can significantly affect the observed patterns (Rosen & Resnick, 1980). Henderson (1995) conducted a meta-analysis of 515 estimates from 29 studies and found that the Zipf coefficient for city-size distributions is significantly larger than 1.0, implying that municipalities are more right-tailed distributed – *that is*, populations (over) concentrate in metros – than suggested by Zipf's Law. Reggiani & Nijkamp (2012) argued that the rule's applicability extends beyond urban populations to socioeconomic spatial networks. They contended that the rank-size rule can be derived from an entropy maximization approach and is thus rooted in welfare theory. Different functional forms, such as exponential, log-normal, and Tanner functions, can be used to model the rank-size distribution and are theoretically justified. The rank-size coefficient can be interpreted as an indicator of the underlying network connectivity structure of the spatial system. The evidence suggests that no single distribution adequately describes the sizes of municipalities across all countries and time periods.

While the rank-size rule offers a parsimonious description of urban hierarchies, its theoretical underpinnings extend far beyond mere mathematical regularity. Manioudis & Meramveliotakis (2022) revisited classical political economy approaches to demonstrate how state capacity and class power influence whether data innovations lead to equitable urban governance. Simultaneously, Franke and Gailhofer (2021) highlighted that without robust civic data rights models, open satellite and tax data may reinforce existing inequalities if access and stewardship remain centralized. By synthesizing these perspectives, our paper situates rank-size and rank-correlation metrics within a broader debate on data justice, ensuring that empirical findings directly address the sustainable development objectives of inclusion, transparency, and resilience.

Kleynhans and Coetzee (2022a) determined that a log-normal probability (rank-size) distribution applied to all municipalities analyzed for each of the census years under their investigation (1996, 2001, and 2011). Their findings were robust and supported by previous research, with local and metropolitan rank-size distribution patterns and hierarchies remaining consistent over time, thereby providing sufficient evidence to accept log-normality for South Africa's local and metropolitan municipalities. Kleynhans and Coetzee (2022a) also found that the number of South African municipalities with populations exceeding one million remained constant from 1996 to 2016. However, the number of municipalities with populations between 100,000 and 500,000 increased significantly, rising from 94 to 116 over the same period.

Zipf's Law and agglomeration economies provide well-established lenses for understanding urban hierarchies (Rosen & Resnick, 1980; Naudé & Krugell, 2003), but our focus here is not on testing those theories *per se*. Instead, we deploy these tried-and-tested rank-size and rank-correlation techniques to bridge conventional census data and emerging satellite and tax data streams, assessing their concordance and policy implications.

2. Data and methods

2.1. The 2022 census: Stats SA

Stats SA is the government agency responsible for collecting and disseminating official statistics. It also facilitates collaboration among various data producers, both within and outside the government. As the official custodian of South African government data, Stats SA ensures that processed (and often underlying) data are accessible to policymakers, the public and academic researchers (Bezuidenhout *et al*., 2021).

Since the democratization of South Africa, Stats SA has conducted four censuses: 1996, 2001, 2011, and 2022. During this period, the Municipal Demarcation Board has made some consequential readjustments to the boundaries of local and metropolitan municipalities, resulting in some being partially or wholly incorporated into others, and new ones being demarcated. This has resulted in a significant reduction from 278 local municipalities in 1996 to 213 in 2016 (Kleynhans & Coetzee, 2022a).

The 2022 South African Census results were released on October 10, 2023. The results were presented to President Cyril Ramaphosa by Statistician-General Risenga Maluleke (Stats SA, 2023). Stats SA kindly provided us with the latest raw population data collected for each of the 213 local and metropolitan municipalities in South Africa during the 2022 Census, which are compared and analyzed in this article.

2.2. Google earth engine

The accessibility of supercomputers and high-performance computing systems is becoming increasingly commonplace, with large-scale cloud computing now universally available as a commodity. At the same time, petabyte-scale remote sensing (satellite) data archives have been made freely accessible from various United States government agencies and the European Space Agency. Furthermore, multiple tools have been developed for large-scale geospatial data processing (Gorelick *et al*., 2017).

Google Earth Engine is a cloud-based platform providing access to high-performance computing resources for processing large geospatial datasets. Unlike most supercomputing centers, Earth Engine is also designed to help researchers easily share their results with peers in the scientific community, government officials, policymakers, and the general public (Gorelick *et al*., 2017).

Earth Engine features a multi-petabyte data catalogue paired with a high-performance computation service. It holds many publicly available geospatial datasets, including satellite and aerial imagery in various wavelengths, environmental variables, weather forecasts, and socioeconomic data. These data are pre-processed into an accessible format, streamlining data management (Gorelick *et al*., 2017). Earth Engine is employed across various disciplines, encompassing topics, such as urban mapping (Gorelick *et al*., 2017).

Regarding the night-time lights dataset we built and employed, stable radiance was sourced from the VIIRS Day/Night Band (DNB) monthly cloud-free composites curated by the Earth Observation Group and hosted by NOAA NCEI (Google Earth Engine collection). The native grid is 15 arc-s (grid m). Whether accessed directly from EOG/NOAA servers or through GEE, this underlying 500 m raster is identical; only downstream user processing alters pixel values. We downloaded all 12 composites for January–December 2022, masked stray light and fire artifacts, and computed the annual median. To align with the LandScan 2022 population raster (1 km), we aggregated the VIIRS stack to 1 km through bilinear interpolation in GEE, and then calculated zonal means for each of the 2021 Municipal Demarcation Board polygons. Landsat-8 OLI surface-reflectance scenes (30 m) for 2022 were inspected only to verify land/water masking and flag anomalous low-light artifacts; Landsat data were not used in the statistical analysis.

2.3. LandScan global 2022: ORNL

Utilizing an innovative approach that integrates Geographic Information Science, remote sensing technology, and machine learning algorithms, ORNL's LandScan serves as the community standard for global population distribution. With a resolution of 30 arc-s (approximately 1 km), LandScan provides the highest-resolution global population distribution data available, which represents an "ambient population" (averaged over 24 h). The LandScan algorithm employs spatial data, high-resolution imagery exploitation, and a multi-variable dasymetric modeling approach to disaggregate census counts within an administrative boundary. LandScan population data are spatially explicit, unlike tabular census data (ORNL, 2023).

As no single population distribution model can encapsulate the variations in spatial data availability, quality, scale, accuracy, and differences in cultural settlement practices, LandScan population distribution models are customized to align with each country's and region's data conditions and geographical characteristics. By modeling an ambient population, LandScan Global captures the full scope of human activity throughout the day and night rather than merely the residential location (ORNL, 2023). Calka and Bielecka (2019) found good correspondence between the LandScan population distribution data and the census-derived population distribution data they used. According to ORNL (2023), the LandScan Global 2022 population data were published on July 1, 2023 (3 months before the 2022 South African census results were released).

2.4. Administrative tax data: SARS-NT

Administrative data are being increasingly utilized for research and policy analysis. South Africa now has an administrative tax database that links formal firms with their workers and other related information about the firms. Collaboration between the SARS, the NT, and the United Nations University World Institute for Development Economics Research (UNU-WIDER) illustrates how tax datasets can serve as a valuable resource for research and policymaking. The City Specialized Economic Data project, conducted through the cities support programme (CSP), recently introduced a spatial dimension to this dataset (Bezuidenhout *et al*., 2021).

As noted by Bezuidenhout *et al.* (2021), the data encompass the formal sector, including businesses and individuals who submit tax returns or have taxes paid on their behalf by their companies. It consolidates tax datasets from three primary areas: (1) Company Income Tax data (CIT/ITR14); (2) personal income tax: IRP5/IT3 (payroll) or ITR12 (assessed personal) tax; and (3) customs data. These three data sources have been linked, allowing firms to be matched with their employees and their engagement in international markets to be observed. Importantly, firms and workers are monitored over time, enabling the dynamics of economic relationships to be studied (Bezuidenhout *et al*., 2021).

The CSP is situated within the IGR division of the NT. The CSP directly supports metropolitan areas and collaborates across divisions in the NT, engaging with other national departments and broader stakeholders to improve the enabling policy environment and fiscal frameworks, thereby facilitating more efficient operations for cities (NT-CSP, 2012).

The CSP has worked to incorporate spatial components into the SARS-NT dataset. This project has utilized the postal codes of firm locations and addresses from IRP5s to identify the areas where individuals reside. These have been aggregated into equal-area hexagons using Uber's H3 index. This enables a highly disaggregated view of how firms and employment have evolved spatially over the past decade, encompassing aspects such as sector composition, the number of firms and full-time equivalent workers, firm turnover, export participation, wages, and others (Bezuidenhout *et al*., 2021). NT has been collaborating with the Human Sciences Research Council to regularly update this spatial tax panel and enhance the capacity within the NT and the metros to develop and utilize it (Bezuidenhout *et al*., 2021).

While the SARS-NT Spatial Tax Panel affords unparalleled coverage of South Africa's *registered* corporate economy, it is silent on the many enterprises and workers that operate outside the tax net. Recent technical commentary (Nell, 2024) and empirical analyses that juxtapose the panel with labor-force survey evidence (Turok & Visagie, 2025) emphasize that informal firms, sole proprietors, and self-employed workers are missing by design. Accordingly, it is essential to note that all SARS-NT-based findings reported here should be interpreted as reflecting formal-sector activity only and as lower-bound indicators in municipalities where informality is prevalent.

Because tax registration is a pre-requisite for inclusion in the SARS-NT panel, the dataset systematically omits informal micro-enterprises and own-account workers. In municipalities – particularly rural towns and peri-urban settlements – where informality dominates, establishment counts will therefore understate the true scale of economic activity. The SARS-NT establishment's variable should thus be regarded as a *lower-bound proxy*, with the magnitude of downward bias roughly proportional to the prevalence of informality (Nell, 2024; Turok & Visagie, 2025).

2.5. Rank correlation and rank-size distribution

We leveraged classic rank-size distributions (*e.g.*, Zipf's Law; Rosen & Resnick, 1980) and Spearman correlations, not to test theory, but as instruments to evaluate how well alternative data (LandScan, SARS-NT) track census-based ranks. This method-driven stance enables us to focus on the challenges and opportunities of harmonization. All methods assume consistent municipal boundaries; therefore, we harmonize the 2022 Census geographies with LandScan and SARS-NT panels, following Kleynhans and Coetzee (2022a).

2.5.1. Ranks-size rule

The rank-size rule is a pattern observed in the distribution of municipality sizes within a region or country. According to this rule, a municipality's population is inversely proportional to its rank in the municipal hierarchy. In other words, the second-largest city is about half the size of the largest city; the third-largest city is about one-third the size of the largest city, and so on.

Mathematically, it can be expressed as $P_r = \frac{P_1}{r}$, where P_r is the population of the city or town ranked r; P_1 is the population of the largest city; r is the rank of the city/town. This rule implies that plotting municipality populations against municipality rank on a log-log scale should show a roughly linear relationship. The rank-size rule helps to understand urban hierarchies and the distribution of urban resources and services.

Brakman *et al.* (1999) noted that Equation I, referred to as Zipf's Law, is formulated with *Rj*, the rank of local/metropolitan municipality *j* according to population rank; *Mj*, the size of local/metropolitan municipality *j* according to population count; and *Co* as the constant.

$$RjMj = Co; j = 1,\ldots,\mathrm{r} \quad \text{(I)}$$

Zipf's Law can be generalized to Equation II, the rank-size distribution, with *q* as a (positive) constant. The local or metropolitan municipality with the largest population is assigned *R* (rank) = 1.

$$R_j^q M_j = Co \quad \text{(II)}$$

Empirically calculating rank-size distributions requires transforming Equation II into a log-linear form, as shown in Equation III.

$$\log(Mj) = \log(Co) - q\log(Rj) \quad \text{(III)}$$

2.5.2. Rank correlation

We employed Spearman's rank correlation (ρ) rather than Pearson correlation because our primary interest lies in ordinal concordance of municipal population rankings across diverse sources, irrespective of scale differences (Giesen & Südekum, 2011). Spearman's ρ is robust to non-normality and heteroscedasticity, ensuring a meaningful comparison even when one dataset (*e.g.*, satellite estimates) systematically overestimates or underestimates absolute counts. In parallel, rank-size estimation assumes that municipality sizes follow a Pareto (or Pareto–log-normal) process (Kleynhans & Coetzee, 2022a).

Gauthier (2001) stated that Spearman's rank correlation coefficient is calculated using Equation IV, where *di* is the rank differential of each data pair (*xi, yi*), and *n* is the number of data pairs.

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^{\,2}}{(n^3 - n)} \quad \text{(IV)}$$

Agreement across population proxies was measured using Spearman's rank correlation coefficient (ρ) rather than Pearson's correlation coefficient (r). Spearman's rho was calculated on ordinal ranks, invariant to monotonic transformations and robust to heavy-tailed, right-skewed distributions that characterize municipality sizes (González-Val 2012). Pearson's moment-based statistic would be inappropriate in this context. Spearman also accommodates the numerous tied ranks in population-count, nightlight, and firm-count data without imposing arbitrary jitter. Correlation coefficients with a magnitude between 0.9 and 1.0 indicate variables that are considered very highly correlated. Correlation coefficients with a magnitude between 0.7 and 0.9 indicate variables that can be regarded as highly correlated.

A viable alternative to Spearman's rank-correlation coefficient (ρ) is Kendall's τ (Kendall, 1938), which is likewise calculated on ordinal ranks but uses concordant–discordant pair counts. Simulation studies have shown that τ and ρ deliver virtually identical inferences for $n \approx 200$ (González Val, 2012). Given this equivalence, and to avoid duplicating tables, we report only ρ in this paper.

2.6. Data integrity checks

Each dataset provides complete coverage of South Africa's 213 local and metropolitan municipalities, resulting in no missing values. We retained all values because the full population rather than a sample, was observed; extreme ranks reflect true municipal heterogeneity rather than sampling noise.

2.7. Limitations of satellite-derived estimates

While Google Earth Engine and LandScan deliver high-resolution ambient population grids, several biases warrant caution. Cloud cover and sensor malfunctions can obscure built-up areas, and urban morphology (*e.g.*, high-rise vs. sprawl) affects dasymetric allocation in the LandScan algorithm, potentially introducing systematic offsets. Calka and Bielecka (2019) found that LandScan underestimates populations by up to 5% in densely forested or informal-settlement zones. We therefore interpret satellite-derived metrics as complementary rather than substitutive, and emphasize rank-based comparisons over absolute counts.

3. Results

From Table 1, it is evident that the ranked populations of municipalities derived from the three population datasets included show nearly perfect correlations. Population data from the ReX database were incorporated into the rank correlation analysis as a control or sanity check. ReX is compiled by S&P Global Market Intelligence using proprietary algorithms that combine census counts, quarterly labor-force surveys, and municipal administrative registers. Therefore, the dataset is methodologically independent of the three public sources compared in this study. If the 2022 Census, 2022 LandScan, and VIIRS luminosity were spuriously correlated due to sharing a common bias, we would expect that correlation to weaken once ReX is added. Instead, Table 1 indicates $\rho > 0.90$ across population pairings, including those involving ReX. This outcome rules out a systemic rank-reversal bias and increases confidence that the high agreement among the public sources genuinely reflects spatial ordering rather than artifacts of a single enumeration method.

The coefficients shown in Table 1 indicate a very high rank correlation between places when using Stats SA population totals and LandScan data. The almost perfect rank correlation between the 2022 Census and the 2022 LandScan ($\rho = 0.98$) confirms that high-resolution ambient population grids can be used to benchmark decennial censuses reliably. Similar findings have been reported in Poland by Calka and Bielecka (2019) and in early LandScan validation work by Dobson *et al.* (2000), reinforcing confidence in satellite proxies for intercensal monitoring. This discovery directly addresses the data

Table 1. Rank correlation analysis

Spearman's rank correlation	Rho (ρ)
Rank correlation: ReX population total (2022) & Stats SA population total (2022)	0.99
Rank correlation: ReX population total (2022) & LandScan population total (2022)	0.99
Rank correlation: Stats SA population total (2022) & LandScan population total (2022)	0.98
Rank correlation: SARS-NT establishments total (2022) & ReX GVA-R total (2022)	0.82
Rank correlation: ReX GVA-R total (2022) & Google Earth Engine (2022)	0.76

Note: The ReX rows are included as a commercially curated control series. Because ReX blends census, survey, and administrative inputs, its rank order serves as an external benchmark to test whether correlations among purely public sources (Census, LandScan, VIIRS, SARS-NT) might be inflated by shared enumeration protocols. Abbreviations: GVA: Gross value added; ReX: Regional Explorer; SARS-NT: South African Revenue Service – National Treasury; Stats SA: Statistics South Africa.

gap issue highlighted by Von Fintel (2023) and supports the argument for rank-based fiscal adjustments advanced later in the paper. It is also notable that there is a large magnitude of the rank correlation between municipalities with more formal firms and their total gross value added. The coefficient of 0.76 indicates a somewhat weaker correlation when ranking municipalities using the ReX GVA-R measure and GEE's nightlight data. Nevertheless, the nightlight data may still have some potential for estimating present gross value added figures at a sub-national level.

Table 2 shows the number of larger municipalities compared to smaller ones, with each 'group' contributing nearly 50% of the total analyzed dataset. Part of this nearly equal share analysis indicates that, according to all three Population Total datasets used (Stats SA, ReX, LandScan), the 16 largest South African municipalities have approximately the same combined number of residents as the next 197 municipalities. These findings explain why policymakers tend to concentrate more on larger municipalities and also serve as a secondary validation of the potential usefulness of the LandScan database, given that all three population datasets produced the same municipal population distribution.

According to the ReX database, the total gross value added, as shown in Table 2, generated by the five largest South African municipalities (based on ReX GVA-R Total rankings) is roughly equal to the gross value added produced by the next 208 municipalities. Unsurprisingly, metropolitan municipalities receive more funding (based on population totals) and more attention from policymakers and academic researchers than smaller local municipalities. The ReX database also indicates that the 11 largest municipalities (based on ReX Unemployment Total rankings) have roughly the same number of unemployed people as the combined total of the next 202 municipalities. Finally, having access to the tax data of the entire population of formal firms in South Africa allows for more accurate conclusions than ever before, such that the four largest South African municipalities (based on SARS-NT Establishments Total rankings) have roughly the same number of formal firms as the next 209 municipalities combined.

Table 2. Near-equal share analysis

Near-equal shares	Approximately 50/50		Ratio of municipalities
Near-equal share: Stats SA population total (2022)	30648431	31090961	16:197 (metropolitan+local: local)
Near-equal share: ReX population total (2022)	30514085	30597751	16:197 (metropolitan+local: local)
Near-equal share: LandScan population total (2022)	29306032	31072023	16:197 (metropolitan+local: local)
Near-equal share: ReX GVA-R total (2022)	2989244969	2994214811	5:208 (metropolitan+local: local)
Near-equal share: SARS-NT establishments total (2022)	223738	220290	4:209 (metropolitan+local: local)
Near-equal share: ReX unemployment total (2022)	3831460	3975438	11:202 (metropolitan+local: local)

Abbreviations: GVA: Gross value added; ReX: Regional explorer; SARS-NT: South African Revenue Service – National Treasury; Stats SA: Statistics South Africa.

Because the Local Government and Provincial Equitable Share formulae scale transfers by population (and, indirectly, by local gross value added), these skews translate directly into proportionally larger fiscal envelopes for metros. The implication is two-fold. First, without a redistributive adjustment, transfer growth will track urban agglomeration, thereby accelerating fiscal divergence between metropolitan areas and small towns, as documented by Todes & Turok (2018). Second, the opportunity cost of mismeasuring the population is magnified: Shifting just 1% of the national headcount from small municipalities to metros would redirect approximately R 700 million in 2024/25 Local Government Equitable Share allocations. Policymakers may therefore wish to introduce (i) a cap on the marginal population weight beyond the first 500,000 residents, or (ii) a parallel needs-based component that cushions low-capacity municipalities against population-driven revenue erosion.

By demonstrating that the "near-equal share" threshold is confined to an unusually small subset of municipalities, our analysis reinforces the need to supplement infrequent census counts with higher-frequency proxies, so that transfer formulas respond to real migration trends rather than statistical artifacts.

Rank-size distribution analysis yields q-coefficients, which are reported in Table 3. If $(0 < q < 1)$, the curve's slope is flatter, resulting in a more even distribution of municipality sizes than what Zipf's Law predicts. If $q > 1$, large cities exceed those predictions, leading to a wider dispersion of municipality sizes (Brakman *et al*., 1999). Earlier South African analysis estimated the rank-size distribution for the 123 largest cities and towns in South Africa and found a q-coefficient (absolute value) of 0.75 (Naudé & Krugell, 2003).

Table 3. Rank-size distribution analysis

Rank-size distributions	q-coefficient (absolute value)	*p*-value
Rank-size distribution: ReX population total (2022)	1.06	<0.001
Rank-size distribution: Stats SA population total (2022)	1.07	<0.001
Rank-size distribution: LandScan population total (2022)	1.05	<0.001
Rank-size distribution: SARS-NT establishments total (2022)	1.44	<0.001
Rank-size distribution: ReX GVA-R total (2022)	1.23	<0.001
Rank-size distribution: ReX unemployment total (2022)	1.23	<0.001
Rank-size distribution (Biggest 123): Stats SA population total (2022)	0.84	<0.001

Abbreviations: GVA: Gross value added; ReX: Regional Explorer; SARS-NT: South African Revenue Service – National Treasury; Stats SA: Statistics South Africa.

Comparing the rank-size distributions from Table 3 across all three population databases consulted (ReX, Stats SA, and LandScan), it is evident that the q-coefficients vary only slightly – this serves as another method for assessing the substitutability of infrequent census data with the more present LandScan data. The rank-size distribution method not only considers the ranking of a municipality but also its size (population), which should dispel the idea that satellite data can only be useful in determining a city's or town's rank.

The population-only rank–size regressions, from Table 3, yield a composite estimate of q ≈ 1.06. This value sits just above Zipf's theoretical value of 1, signaling a modestly steeper urban hierarchy but well below the 1.20-plus range typical of strongly primate systems. Internationally, a q-coefficient in the 1.05 – 1.10 band is common among large, mature economies with diversified urban systems such as the United States (Giesen & Südekum, 2011), France (Rosen & Resnick, 1980), and the OECD functional area (Veneri, 2016). South Africa's q therefore points to incipient, not extreme, primacy: Johannesburg-Pretoria and Cape Town dominate the upper tail, yet secondary metros (*e.g.*, eThekwini, Nelson Mandela Bay) retain substantial scale.

A coefficient slightly above unity suggests that additional agglomeration could still be *welfare-enhancing*, provided congestion, housing, and service bottlenecks are managed (Capello, 2004). Rather than capping metro growth outright, planners might prioritize intra-metro infrastructure, faster commuter rail between primary and secondary cities, and targeted incentives that spread high-productivity activity to the next urban tier. Continuous monitoring with higher-frequency proxies (satellite grids, administrative tax panels) remains essential; should q drift upward toward the 1.15 – 1.20 range, more assertive spatial-redistribution tools would be justified.

The concentration of people in larger municipalities should be expected, given the movement of individuals from smaller municipalities who face unemployment and are driven by the prospect of finding work in the city. This is evidenced by the large q-coefficient calculated for SARS-NT Establishments of 1.44, as shown in Table 3, which indicates a significant concentration of formal firms in larger municipalities, alongside a ReX Unemployment q-coefficient of 1.23.

Figure 1 presents a Pareto curve illustrating the cumulative share of the total population accounted for by municipalities, based on the 2022 Census data, ranked from largest to smallest. On the x-axis, municipalities are ordered by descending population rank (1 = largest), while the right-hand y-axis shows the cumulative percentage of the national population. The steep initial slope indicates that the top 10 municipalities together account for almost 43% of South Africa's population. When we fit a simple log–log regression of municipality population on rank (rank-size rule), we obtain a slope of −1.07 (q = 1.07), reflecting stronger primacy than the q = 1 implied by Zipf's Law.

Figure 2 depicts a Pareto curve of SARS-NT–registered establishment counts, showing the cumulative share of total formal firms on the right-hand y-axis against municipalities ranked by descending firm counts on the x-axis (rank 1 = highest count). The steep initial slope reveals that the top 10 municipalities together account for an incredible 70% of all formal establishments in the country. A log–log regression of establishment count on rank (rank-size rule) produces a slope of −1.44 (q = 1.44), indicating a far more pronounced concentration of economic activity than the q = 1 benchmark of Zipf's Law. This slope is far steeper than that observed for population in Figure 1, suggesting that formal-sector firms are concentrated in an even smaller set of leading municipalities, predominantly the Gauteng metros. This elevated q-coefficient highlights that South Africa's urban system is more top-heavy – an outcome of lasting spatial constraints as documented by Todes & Turok (2018) – underscoring the value of rank-based metrics for identifying municipalities that dominate demographic and economic activity.

As shown in Table 4, although more than two-thirds of all formal firms in South Africa are located in the eight metropolitan municipalities, according to the 2022

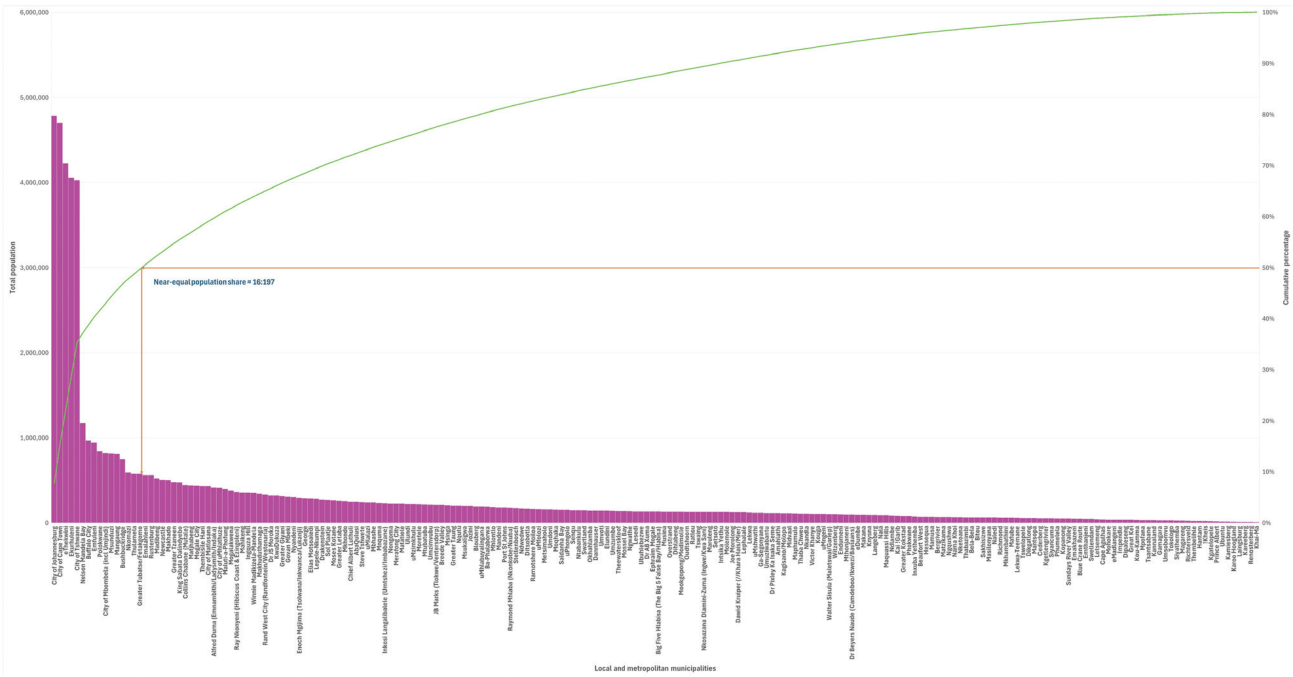

Figure 1. Pareto chart for population size analysis of local and metropolitan municipalities (2022)

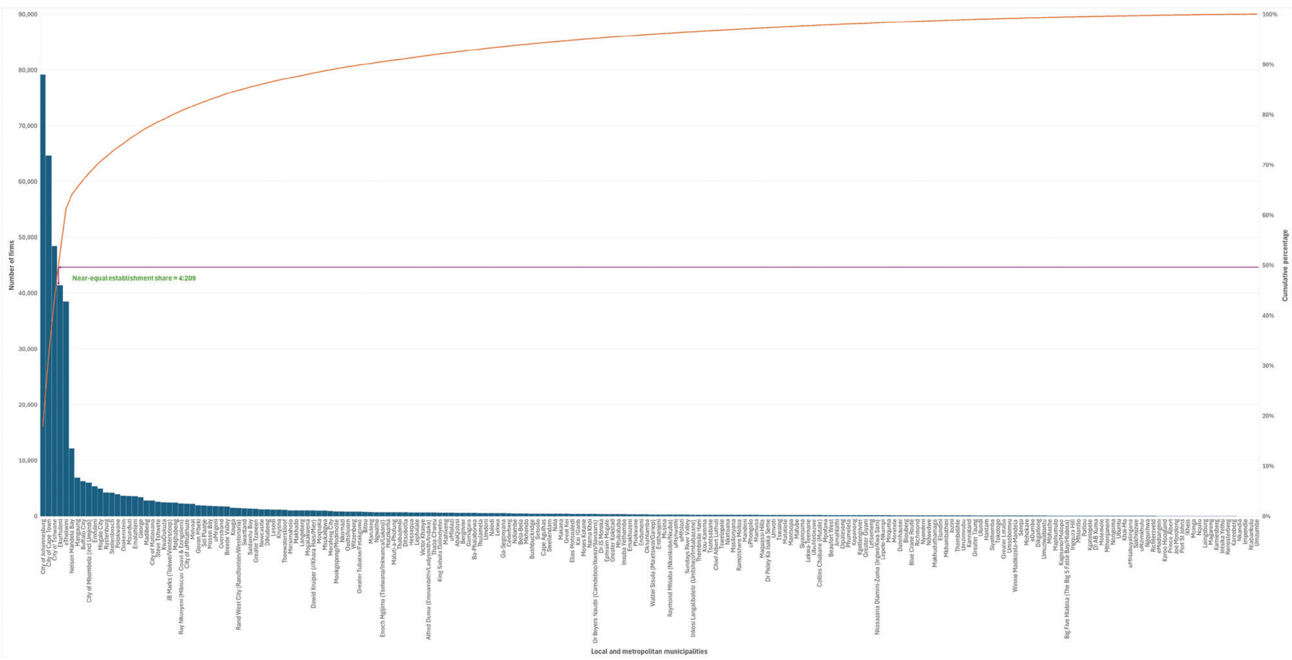

Figure 2. Pareto chart for South African revenue service – national treasury establishments count analysis (2022)

Census, only 40% of all South Africans reside in these municipalities. This economic phenomenon, studied strictly within the formal-firm environment, further reinforces the potential benefits of agglomeration, *that is*, positive externalities, as highlighted by Naudé & Krugell (2003) and Turok (2021).

4. Discussion

At a provincial or subnational level in South Africa and many developing countries, relevant and up-to-date economic data and information are scarce. Accessible data sources are often outdated, rendering them almost

Table 4. Metropolitan municipalities, SARS-NT establishments, and Stats SA population total and rank analysis

Metropolitan municipalities	SARS-NT establishments (2022)	SARS-NT rank	Total population based on Stats SA (2022)	Stats SA rank
City of Johannesburg	79189	1	4784240	1
City of Cape Town	64640	2	4698599	2
eThekwini	38499	5	4223177	3
Ekurhuleni	41410	4	4054357	4
City of Tshwane	48456	3	4025333	5
Nelson Mandela Bay	12121	6	1174572	6
Buffalo City	6272	8	967361	7
Mangaung	6898	7	808543	12
Sum	297485		24736182	
Share of total	**67%**		**40%**	

Abbreviations: SARS-NT: South African Revenue Service – National Treasury; Stats SA: Statistics South Africa.

useless, as they only describe a past landscape. The existing economic data and information are usually of little value for decision-making and as indicators of present trends and prospects in the subnational economy. This presents a significant challenge, as provincial and local governments often pursue an economic development agenda that necessitates relevant and present economic data and information (Coetzee & Kleynhans, 2021).

In many countries, including South Africa, obtaining high-quality, continuous, high-frequency economic data is challenging, especially at the sub-national level. High-frequency data refer to fine-scale time-series data, whereas annual data are low-frequency and often the standard in many places. This situation complicates effective policymaking, which relies on accurate, up-to-date data. Therefore, data-driven policymaking should integrate new sources, such as real-time remote sensing data and innovative processing techniques (Kleynhans & Coetzee, 2022b).

A significant gap exists between the demand and supply of subnational economic data, necessitating the need to seek reliable sources to bridge it. With scarce subnational data and increasing availability of satellite data, exploring the application of satellite data is necessary. Satellite data may be a reliable source of subnational economic information (Coetzee & Kleynhans, 2021).

Satellites have become a significant part of modern civilization. Future satellites will primarily focus on mapping the planet in minute detail through remote sensing. Remote sensing satellites are typically deployed to monitor resources vital for humanity (Ryan-Mosley *et al.*, 2019). The field of economics appears to be a major beneficiary of satellite technology, opening up applications that are not feasible using national statistics, particularly at a sub-national level (Coetzee & Kleynhans, 2021).

One would expect changes in the population of places between census surveys. Some cities and towns are net recipients of people, and others are net senders. This means that some places need to receive more funding from the equitable share grant, while others will receive less. Given the issues raised about the recent 2022 Census data, some local governments argue that the changes are not due to migration but rather constitute mismeasurement, and as a result, they will be at a disadvantage. Differentiating between these two scenarios is difficult without other reliable data.

Our hypothesis is that the absolute numbers are less significant if it is possible to benchmark the 2022 Census numbers against other sources. We ranked places by population from largest to smallest and compared the 2022 Census data to the population estimates from LandScan Global 2022. The rankings of the larger places are very persistent. If the equitable share grant is allocated according to these relative proportions, the larger places, which are likely to have been net recipients of people, will not lose much.

Our comparative analysis demonstrates that relative population rankings – whether derived from census counts, satellite imagery, or administrative tax records – constitute a robust basis for equitable intergovernmental transfers under SDG 11's mandate for sustainable cities. By harmonizing "conventional" and "unconventional" sources through rank-size and rank-correlation metrics, we reveal both the promise and pitfalls of alternative data streams for urban governance. Crucially, strong Spearman correlations ($\rho > 0.95$) affirm population rank stability, while pronounced Pareto exponents highlight metropolitan primacy pressures. Interpreted through the lens of data justice, these findings suggest that rank-based allocations can mitigate census undercounts in informal settlements without discarding absolute headcounts. Integrating dynamic data pipelines will be crucial for resilient urban systems that can adapt to shifting migration patterns and evolving settlement patterns.

5. Policy recommendations

5.1. Recommendation 1: Embed an annual "satellite growth factor" in the local-government equitable-share (LGES) formula

Budget-related matters, such as grant allocation, fall within the NT's remit, which can reevaluate and augment the present grant allocation formulae and approaches by incorporating an additional municipality "satellite growth factor" variable in their models. South Africa's LGES still relies on decennial census counts that grow stale long before the next enumeration. A practical solution is to recalibrate each municipality's census base annually with an objective growth factor derived from the LandScan global population raster. Treasury's Budget Office would compute the 4-year average LandScan growth rate for every municipality and apply it to the latest census total *before* the horizontal split of the LGES is calculated. Because LandScan is built from daytime satellite imagery, mobile phone data, and administrative registers, it captures real-time migration into fast-growing metropolitan areas and population losses in shrinking small towns – thereby aligning transfers with present service delivery needs. Stats SA would audit the raster for obvious artifacts (*e.g.*, mining voids misclassified as settlements), and the adjustment would be published in Annexure W1 of the Division of Revenue Bill each year, providing municipalities with a transparent, annually updated population weight without requiring them to wait for the next census.

Because the 2026 Budget is expected to be tabled toward the end of February 2026 (although the 2025 fiscal package famously slipped to May 2025), NT will still be working with the latest complete LandScan raster, *that is*, 2024, published in mid-2025. Using the data strictly to extract a growth signal, NT would take Municipality A's LandScan population counts of 98,000 (2021), 1,00,000 (2022), 1,03,000 (2023), and 1,06,000 (2024). The geometric mean of the 3 year-on-year growth factors is roughly 2.6%/annum. Starting from the 2022 Census anchor of 1,00,000 residents, and compounding this rate produces successive projections of about 1,02,600 (2023), 1,05,300 (2024), 1,08,000 (2025), and 1,10,800 (2026); the last figure would serve as the population weight in the 2026/27 Local-Government Equitable-Share calculation. A reality check is possible only at the provincial scale, because Stats SA's mid-year population estimates are published for South Africa as a whole and for the nine provinces, not for individual local or metropolitan municipalities. Consequently, any reconciliation between the LandScan-based roll-ups and official mid-year numbers informs the Provincial Government Equitable Share, whereas the LGES must rely on mechanically rolled-forward municipal projections until the next census.

5.2. Recommendation 2: Launch a public night-lights dashboard on the treasury CSP portal

To complement population updates, planners also need a high-frequency proxy for local economic buoyancy. A web-based dashboard built on monthly VIIRS night-time light composites can provide that signal. Hosted on the NT's CSP site, the dashboard would display: (i) Each municipality's median radiance level and percentile rank, (ii) month-on-month and year-on-year change maps, and (iii) downloadable CSV/GeoJSON feeds for analysts. The back-end script, running on Google Earth Engine, ingests the latest cloud-free VIIRS tile, applies a stray-light mask, aggregates radiance to a 1 km resolution, and calculates zonal means for the 2021 MDB boundaries. Results are refreshed automatically within 48 h of the NOAA release, warning finance officials early of the potential impact of power-supply disruptions, disaster aftermath, or unexpected growth surges. Operating costs are modest, and the open-data ethos aligns with the Protection of Personal Information Act (Act 4 of 2013) and the Open Government Partnership commitments that South Africa renewed in 2023.

The analysis is not without limitations. Satellite population models may misallocate dwellers in dense informal settlements or cloud-prone areas, while tax records systematically omit unregistered firms. Boundaries were held constant at the 2021 definitions; any redistricting would require reharmonization. Nevertheless, the rank-based approach is resilient to these imperfections because it focuses on ordinal, rather than cardinal, accuracy.

Looking ahead, we recommend three avenues for future research: (i) Informal-sector calibration: Combine labor-force or household surveys with machine-learning estimates of informal activity to refine administrative proxies; (ii) Temporal dynamics: Extend the framework into a panel to monitor how shocks (*e.g.*, COVID-19, load shedding) reshape municipal hierarchies in near real-time; and (iii) Cross-country replication: Apply the operational recipe to other census-constrained contexts in the Global South to test generalizability.

6. Conclusion

This paper set out to test whether readily available "unconventional" data – satellite-derived population grids and administrative tax records – can corroborate or even improve upon official census counts at the municipal level in South Africa. By harnessing five independent data sources and evaluating them using rank-size and rank-correlation techniques, we demonstrate that relative municipal rankings are remarkably stable across sources, even when absolute head counts differ. LandScan's ambient-population grids track Stats SA's 2022 Census almost one-for-one, while nightlight intensity and SARS-NT establishment counts capture complementary dimensions of local economic activity. The results confirm that metros dominate formal economic activity far more than they dominate population totals, underscoring the fiscal stakes of accurate big-city measurement.

From a policy perspective, the findings support three concrete actions. First, a simple "satellite growth factor" could be applied annually to update Equitable-Share allocations, mitigating the lag between decennial censuses and fast-moving urban growth. Second, integrating VIIRS night-light dashboards into the NT's CSP would provide planners with a near-real-time economic pulse. Third, administrative tax indicators should be interpreted as lower bounds in municipalities with high informality and should be paired with household survey evidence to minimize bias.

By embracing a *living* spatial-data ecosystem built on openly accessible satellite imagery stocks and routinely collected administrative records, South Africa can stabilize fiscal transfers, sharpen local development planning, and advance Sustainable Development Goal 11's call for sustainable and inclusive cities – all while reducing dependence on costly, infrequent censuses.

Harnessing the power of satellite data could go a long way in mitigating or smoothing out discrepancies in absolute population counts that are ostensibly reported. Government, policymakers, and researchers must continue to collaborate in mapping the "unconventional" data landscape and fully exploit these newly discovered resources, harnessing administrative tax and satellite data. Such "unconventional" data can enhance the accuracy of resource allocation, urban planning, and infrastructure development. In what other ways could these innovative data sources help officials grow their cities and towns more sustainably, efficiently, and equitably?

Acknowledgments

We would like to acknowledge the following organizations for their contributions to this research: Southern Africa – Toward Inclusive Economic Development, Spatial Economic Activity Data South Africa, Stats SA, The UNU-WIDER, and Economic Research Southern Africa. We would also like to thank the anonymous referees for their insights to help strengthen and elevate our work.

Funding

None.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Waldo Krugell
Data curation: Lodewalt Venter, Clive Coetzee
Formal analysis: Lodewalt Venter
Methodology: All authors
Investigation: Lodewalt Venter
Project administration: Waldo Krugell
Visualisation: Lodewalt Venter
Writing – original draft: Lodewalt Venter
Writing – review & editing: All authors

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Data can be accessed from: Statistics South Africa Census 2022 (https://superweb.statssa.gov.za/webapi/jsf/login.xhtml); S&P Global ReX (https://www.spglobal.com/market-intelligence/en/solutions/products/regional-explorer-economics-data-analytics); SARS-NT Spatial Tax Panel (https://spatialtaxdata.org.za/); Google Earth Engine (https://earthengine.google.com/); ORNL LandScan (https://landscan.ornl.gov/).

Further disclosure

The entire set of pre-reviewed findings has been presented at the Economic Research Southern Africa (ERSA) Cities Conference, "The Role of Cities as Drivers of Growth and Employment," in Muldersdrift, South Africa, on October 31, 2024.

References

Bezuidenhout, C., Malindi, K., & Rankin, N. (2021). A Report and Strategy for Spatialised Economic Activity Data. (Technical Report. City Support Programme. National Treasury: Pretoria).

Brakman, S., Garretsen, H., Van Marrewijk, C., & Van Den Berg, M. (1999). The return of Zipf: Towards a further understanding of the rank-size distribution. *Journal of Regional Science*, 39(1):183-213.

https://doi.org/10.1111/1467-9787.00129

Calka, B., & Bielecka, E. (2019). Reliability analysis of LandScan gridded population data. The case study of Poland. *ISPRS International Journal of Geo-Information*, 8(5):222.

https://doi.org/10.3390/ijgi8050222

Capello, R. (2004). Beyond optimal city size: Theory and evidence reconsidered. *Contributions to Economic Analysis*, 266:57-85.

https://doi.org/10.1016/S0573-8555(04)66003-2

Coetzee, C. (2012). Data resources "available" for local economic assessment in South Africa. *Skills at Work: Theory and Practice Journal*, 5(1):1-12.

https://doi.org/10.13140/2.1.3267.2007

Coetzee, C.E., & Kleynhans, E.P.J. (2021). Estimating economic activity of South African provinces using satellite technology: An econometric panel data application. *Remote Sensing Applications: Society and Environment*, 23(21):100564.

https://doi.org/10.1016/j.rsase.2021.100564

Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C., & Worley, B.A. (2000). LandScan: A global population database for estimating at-risk populations. *Photogrammetric Engineering and Remote Sensing*, 66(7):849-857.

https://doi.org/10.1201/9781482264678-24

Franke, J., & Gailhofer, P. (2021). Data governance and regulation for sustainable smart cities. *Frontiers in Sustainable Cities*, 3:763788.

https://doi.org/10.3389/frsc.2021.763788

Gauthier, T. D. (2001), Detecting trends using Spearman's rank correlation coefficient. *Environmental Forensics,* 2(4):359-362.

https://doi.org/10.1080/713848278

Giesen, K., & Südekum, J. (2011). Zipf's law for cities in the regions and the country. *Journal of Economic Geography*, 11(4):667-686.

https://doi.org/10.1093/jeg/lbq019

González-Val, R. (2012). Zipf's Law: Main issues in empirical work. *Région et Développement*, 36:147-164.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202(3):18-27.

https://doi.org/10.1016/j.rse.2017.06.031

Henderson, J.V. (1995). Innovation and agglomeration: Two parables suggested by city-size distributions: Comment. *Japan and the World Economy*, 7(4):391-393.

https://doi.org/10.1016/0922-1425(95)00022-4

Joffe, H. (2024). Operation Vulindlela Looks to Fix Local Government in Second Phase. Available from: https://www.businesslive.co.za/bd/economy/2024-06-26-operation-vulindlela-looks-to-fix-local-government-in-second-phase [Last accessed on 2024 Jul 30].

Kahn, T. (2024). Scientists Slam Stats SA's 2022 Census as "Work of Fiction". Available from: https://www.businesslive.co.za/bd/national/2024-07-11-scientists-slam-stats-sas-2022-census-as-work-of-fiction [Last accessed on 2024 Jul 30].

Kendall, M.G. (1938). A new measure of rank correlation. *Biometrika*, 30(1-2):81-93.

https://doi.org/10.2307/2332226

Kleynhans, E.P., & Coetzee, C.E. (2022a). The rank-size distribution of cities in South Africa. *GeoJournal*, 87(2): 4775-4790.

https://doi.org/10.1007/s10708-021-10527-x

Kleynhans, E.P., & Coetzee, C.E. (2022b). Temporal Disaggregation of gross domestic product data with night-time lights data for South Africa. *African Journal of Business and Economic Research*, 17(4):233-259.

https://doi.org/10.31920/1750-4562/2022/v17n4a11

Kleynhans, E.P.J., & Coetzee, C.E. (2025). Insights from a population grid of South Africa: An applied spatial satellite data analysis. *International Journal of Population Studies*, 11(2):30-42.

https://doi.org/10.36922/ijps.3297

Manioudis, M., & Meramveliotakis, G. (2022). Broad strokes towards a grand theory in the analysis of sustainable development: A return to the classical political economy. *New Political Economy*, 27(5):866-878.

https://doi.org/10.1080/13563467.2022.2038114

National Treasury. (2024). Medium-Term Budget Policy Statement 2024: Chapter 4 - Expenditure Priorities. Pretoria: Government Printer. Available from: https://www.treasury.gov.za/documents/mtbps/2024/mtbps/FullMTBPS.pdf [Last accessed on 2025 May 13].

National Treasury. (2025). Budget Review 2025: Chapter 6 - Division of Revenue and Spending by Provinces and Municipalities. Pretoria: Government Printer. Available from: https://www.treasury.gov.za/documents/national%20budget/2025/review/fullbr.pdf [Last accessed on 2025 May 13].

Naudé, W.A., & Krugell, W.F. (2003). Are South African cities too small? *Cities: The International Journal of Urban Policy and Planning*, 20(3):175-180.

https://doi.org/10.1016/S0264-2751(03)00011-8

Nell, A. (2024). Spatial Tax Panel - Overview [Earlier Version Presented to the City Economic Development Managers' Forum, 11 April 2023]. Government Technical Advisory Centre Public Economics Conference, Pretoria. Available from: https://www.gtac.gov.za/wp-content/uploads/2024/09/spatial-tax-panel-by-andrew-nell-from-sead-sa.pdf [Last accessed on 2025 May 24].

NT-CSP (National Treasury - Cities Support Programme). (2012). About the CSP. Available from: https://csp.treasury.gov.za [Last accessed on 2025 Feb 08].

ORNL (Oak Ridge National Laboratory). (2023). Landscan Global 2022. Tennessee: Oak Ridge National Laboratory.

https://doi.org/10.48690/1529167

Reggiani, A., & Nijkamp, P. (2012). Did zipf anticipate socio-economic spatial networks? University of Bologna: Quaderni-Working Paper DSE.

https://doi.org/10.2139/ssrn.2016979

Rosen, K.T., & Resnick, M.F. (1980). The size distribution of cities: An examination of the Pareto law and primacy. *Journal of Urban Economics*, 8(2):165-186.

https://doi.org/10.1016/0094-1190(80)90043-1

Ryan-Mosley, T., Winick, E., & Kakaes, K. (2019). The Number of Satellites Orbiting Earth Could Quintuple in the Next Decade. MIT Technology Review. Available from: https://www.technologyreview.com/2019/06/26/755/satellite-constellations-orbiting-earth-quintuple/ [Last accessed on 2025 Feb 09].

SAG (South African Government). (2024). South Africa Yearbook 2022/2023. Available from: https://www.gcis.gov.za/resources/south-africa-yearbook-202223 [Last accessed on 2025 May 14].

Stats SA (Statistics South Africa). (2023). MEDIA RELEASE: Census 2022 Population Count Results 10 October 2023. Available from: https://www.statssa.gov.za/?p=16716 [Last accessed on 2025 Feb 08].

Todes, A., & Turok, I. (2018). Spatial inequalities and policies in South Africa: Place-based or people-centred? *Progress in Planning*, 123:1-31.

https://doi.org/10.1016/j.progress.2017.03.001

Turok, I. (2021). Urbanization, Agglomeration, and Economic Development in South Africa. In: Oxford Handbook of the South African Economy. United Kingdom: Oxford University Press.

https://doi.org/10.1093/oxfordhb/9780192894199.013.28

Turok, I., & Visagie, J. (2025). Striking disparities in employment across South Africa: Evidence from a new spatial database. *Development Southern Africa*, 42(3):383-396.

https://doi.org/10.1080/0376835X.2025.2456252

Veneri, P. (2016). City size distribution across the OECD: Does the definition of cities matter? *Computers, Environment and Urban Systems*, 59:86-94.

https://doi.org/10.1016/j.compenvurbsys.2016.05.007

Von Fintel, D. (2023). Monitoring South Africa's Metropolitan Economies: A Survey of the Data Landscape. (Discussion Document 13. Economic Research Southern Africa). Available from: https://econrsa.org/download/27409/?tmstv=1684419483 [Last accessed on 2024 May 25].

Wang, Y., & Sun, B. (2024). The types of city size distributions and their evolution. *Cities*, 150(1):105045.

https://doi.org/10.1016/j.cities.2024.105045