Scientific Research Publishing

Journal of Geographic Information System, 2026, 18(4), 196-211
https://www.scirp.org/journal/jgis
ISSN Online: 2151-1969
ISSN Print: 2151-1950

Automated Delineation of New Enumeration Areas by Using Administrative Data and Hierarchical Split and Merge Approach in ESRI ArcGIS Pro: A Case Study of Abu Dhabi Emirate

Rizwan Ali, Ali Abdulrasool Alrasheed, Marwa Mohammed Ali Alsuwaidi

Statistics Centre Abu Dhabi, Abu Dhabi, United Arab Emirates
Email: riali@scad.gov.ae

How to cite this paper: Ali, R., Alrasheed, A.A. and Alsuwaidi, M.M.A. (2026) Automated Delineation of New Enumeration Areas by Using Administrative Data and Hierarchical Split and Merge Approach in ESRI ArcGIS Pro: A Case Study of Abu Dhabi Emirate. *Journal of Geographic Information System*, **18**, 196-211.
https://doi.org/10.4236/jgis.2026.184011

Received: May 11, 2026
Accepted: August 18, 2026
Published: August 21, 2026

Abstract

As populations shift and urban landscapes evolve, enumeration area (EA) boundaries require regular updates to support accurate data collection, analysis, and dissemination, particularly for census operations. Existing approaches for creating or updating EAs lack automation and controls to prevent boundaries from intersecting building footprints. This often results in individual buildings being split across multiple EAs, leading to inaccuracies in household allocation and poor cartographic representation when overlaid with building and satellite data. To address these limitations, this study presents a fully automated tool developed in ArcGIS Pro using Python. The tool systematically subdivides lowest administrative units (community boundaries) into smaller geographic entities, termed primary units, by hierarchically intersecting road networks, legal parcel/plot boundaries, and building footprints as required. These primary units are then intelligently aggregated to generate EAs, ensuring that each building is entirely contained within a single EA. The resulting EAs are evaluated through both quantitative and visual validation. Household distributions are compared against legacy EAs, and spatial alignment is assessed through overlay with roads, parcels, building footprints, and satellite imagery. Applied to Abu Dhabi Emirate, the methodology generated 3298 EAs covering the emirate-wide household universe, representing a 57.6% increase in geographic granularity over the 2092 legacy EAs. The results demonstrate complete (100%) conformity with building boundaries, with 73.3% of populated EAs falling within the target household threshold of 75 - 200 and the maximum household count reduced from 87% above the 200-HH ceiling (legacy) to 63% above the ceiling (new). This approach significantly enhances pre-census mapping workflows in Abu Dhabi by reducing time and cost while improving spatial accuracy. The proposed automated tool is also applicable to other regions where EA boundaries are outdated or unavailable.

Keywords

Enumeration Area, Spatial Zoning, Census Geography, Split and Merge, GIS Automation, Building Footprint

1. Introduction

The United Nations recommends that each country conduct a national population census at least once every ten years to support evidence-based planning, governance, and resource allocation [1]. Census is a comprehensive process encompassing the collection, compilation, evaluation, analysis, and publication of demographic, economic, and social data pertaining to all individuals within a country at a specific point in time. A fundamental component of census operations is the subdivision of a country into small, manageable geographic units known as Enumeration Areas (EAs) [1] [2]. EAs represent the smallest operational units for data collection and serve as the foundation for statistical analysis, sampling frameworks, and dissemination of census outputs. Typically, an EA is designed to contain a workload that can be covered by a single enumerator within a defined time frame, ensuring complete coverage without omission or duplication [3].

To be operationally effective, EAs must satisfy several well-established criteria, including being mutually exclusive and exhaustive (no overlaps or gaps), geographically compact, aligned with administrative hierarchies, and easily identifiable on the ground using physical features such as roads or natural boundaries [3] [4]. Additionally, EAs should maintain a relatively balanced population distribution, be neither too large nor too small, and remain flexible enough to support diverse statistical and administrative applications. The quality and integrity of EA boundaries directly influence census efficiency, data accuracy, and the comparability of results across time and regions [5]. As highlighted in previous studies, poorly delineated EAs can lead to enumeration errors, inefficient field operations, and biased statistical outputs [6].

Traditionally, the creation and updating of EAs have relied on manual approaches, including field surveys and on-screen digitization using high-resolution satellite imagery. While these methods allow for human judgment in boundary definition, they are inherently time-consuming, labor-intensive, and costly [6] [7]. Large-scale census mapping exercises can take several years and require significant financial and human resources. Moreover, manual processes are prone to inconsistencies, subjectivity, and topological errors such as overlaps, gaps, and misalignment with real-world features [6]. A critical limitation observed in manual and semi-digital approaches is the inability to consistently enforce spatial constraints, particularly the preservation of building integrity, often resulting in individual buildings being split across multiple EAs [7].

With the advancement of geospatial technologies, there has been a global shift toward Geographic Information System (GIS)-based approaches for census mapping. GIS enables the integration of satellite imagery, administrative boundaries, infrastructure data, and population estimates to support more efficient and accurate EA delineation [4]. This transition from manual to digital mapping has significantly improved accuracy, consistency, and quality control, while reducing both time and cost requirements [4] [6]. Several national statistical organizations have adopted semi-automated GIS workflows, leveraging spatial modelling techniques to generate more coherent and balanced EAs.

Recent research has explored a variety of algorithmic approaches for automating EA delineation. For example, split-and-merge techniques and region-growing algorithms have been applied to optimize EA compactness and population balance [7]. Methods such as AZ Tool, quadtree decomposition, and max-p regionalization have demonstrated effectiveness in generating statistically homogeneous and spatially contiguous units [6] [7]. Other approaches, including Hierarchical Land Parcel Aggregation (HeLP) and SKATER, emphasize clustering based on geographic contiguity and attribute similarity. Additionally, polygon-splitting techniques using road networks and region-merging strategies have been used to create urban blocks and statistical zones [7].

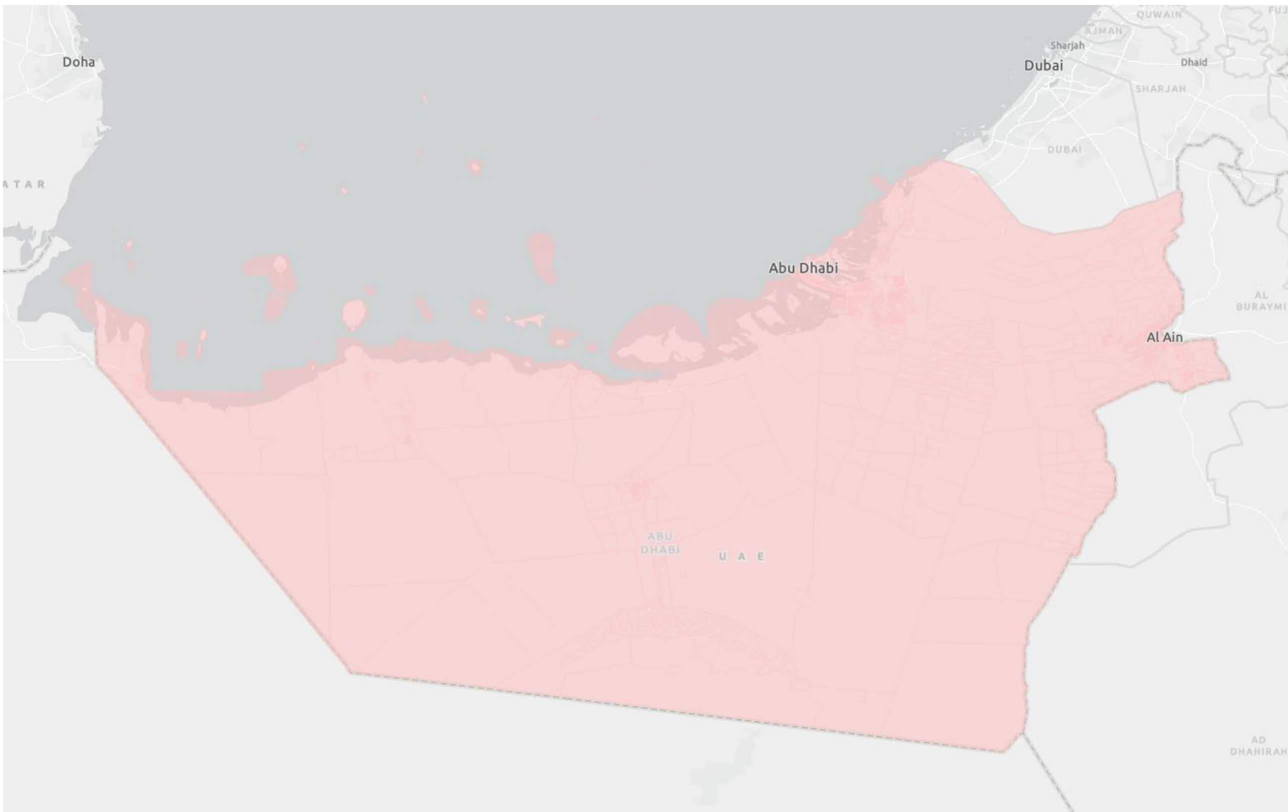
Despite these advancements, several limitations remain. Many existing approaches are implemented in standalone statistical or research-oriented environments, limiting their usability within operational workflows of national statistical offices [6] [7]. More importantly, most methods do not explicitly enforce constraints related to real-world features such as building footprints, leading to boundaries that may intersect structures and reduce their practical usability in field operations [7]. Furthermore, reliance on outdated or unavailable base units (e.g., legacy EAs) restricts applicability in rapidly evolving urban environments or regions where baseline data is incomplete.

In rapidly urbanizing regions such as Abu Dhabi, where high-resolution administrative and geospatial datasets are available, there is a critical need for fully automated, GIS-integrated solutions that can generate accurate, consistent, and operationally meaningful EAs. Such solutions must not only optimize population distribution and spatial compactness but also ensure strict adherence to physical features, particularly preventing EA boundaries from crossing building footprints.

This study addresses these gaps by proposing a fully automated EA delineation approach implemented within ArcGIS Pro using Python. The methodology leverages authoritative administrative data and applies a hierarchical split-and-merge strategy to systematically subdivide communities into smaller primary units based on road networks, parcel/plot boundaries, and building footprints. These units are subsequently aggregated to form EAs that strictly preserve building integrity while maintaining spatial coherence and balanced workloads. By embedding the entire workflow within a widely used GIS platform, the proposed approach provides a practical, scalable, and replicable solution for modern census mapping, significantly enhancing efficiency, accuracy, and operational usability with minimal human intervention.

2. Study Area

The Abu Dhabi Emirate, the largest of the seven emirates of the United Arab Emirates (UAE), covers approximately 87% of the country's landmass. It comprises three major administrative regions—Abu Dhabi (Central), Al Ain (Eastern), and Al Dhafra (Western)—each characterized by distinct geographic, demographic, and infrastructural attributes. Abu Dhabi city is a major metropolitan center with high-rise residential and commercial buildings, while Al Ain features a more traditional urban fabric surrounded by agricultural zones. Al Dhafra is characterized by sparsely populated desert and coastal settlements. This spatial diversity necessitates flexible geoprocessing tools capable of handling heterogeneous data (**Figure 1**).



Base Map Source: Esri, TomTom, Garmin, FAO, NOAA, USGS, OpenStreetMap contributors and GIS User Community.

Figure 1. Study Area—Abu Dhabi Emirate.

3. Methodology

The automated Enumeration Area (EA) delineation tool is implemented as a Python-based geoprocessing toolbox within ArcGIS Pro. The workflow is structured into three principal phases—Preprocessing, Processing, and Postprocessing, which collectively transform raw administrative and geospatial datasets into a validated EA layer (**Figure 2**).

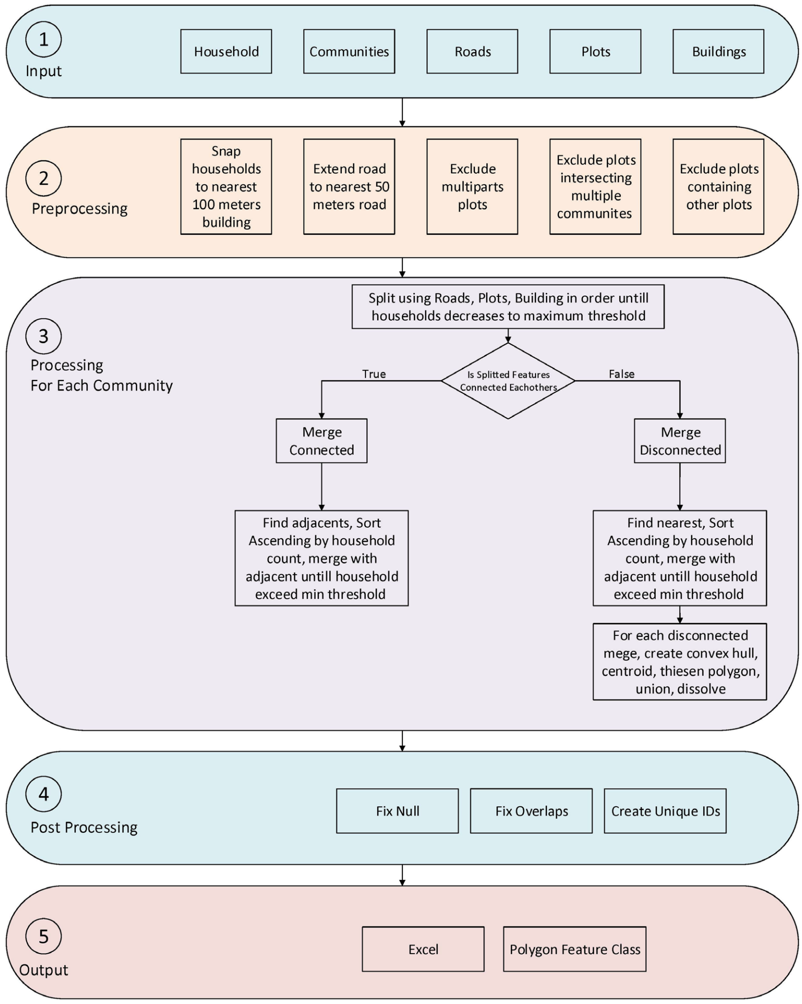


Figure 2. Flowchart showing the methodology stages.

3.1. Preprocessing

The preprocessing phase prepares five core datasets—households, buildings, plots (parcels), roads, and community boundaries—for downstream spatial operations (Table 1).

Table 1. Datasets summary.

Sr No	Dataset Name	Geometry	Source	Records	Spatial Coverage	Reference Date
1	Households	Point	Statistics Centre Abu Dhabi	Emirate HH universe	Abu Dhabi Emirate	Dec 2022
2	Legacy EAs	Polygon	Statistics Centre Abu Dhabi	2,092	Abu Dhabi Emirate	Dec 2022
3	Buildings	Polygon	*/353	382,102	Abu Dhabi Emirate	June 2025
4	Plots	Polygon	Department of Municipalities and Transport	379,698	Abu Dhabi Emirate	June 2025
5	Road Centerline	Line	*/101	75,491	Abu Dhabi Emirate	June 2025
6	Communities	Polygon	*/2	1,855	Abu Dhabi Emirate	June 2025

(* https://arcgis.sdi.abudhabi.ae/agspublish/rest/services/OpenData/ADSDI_OpenData/MapServer).

Household–Building Alignment: Household point locations are frequently misaligned with their associated building footprints in administrative datasets. To correct this, building centroids are computed and a nearest-neighbor (tolerance 100 meters) spatial association is performed, snapping each household point to its corresponding building centroid. The assignment is strictly one-to-one: each household is linked to the single nearest building within the 100-meter search radius, and multiple households may be associated with the same building if they are each independently nearest to that building. Household points for which no building falls within the 100-meter tolerance are retained at their original recorded location and attributed to the containing community, ensuring no households are lost from the dataset. This ensures that household counts are accurately attributed during subsequent spatial aggregation processes.

Road Network Topology Correction: Road datasets often contain disconnected segments due to digitization inconsistencies. These discontinuities are resolved through a topology correction procedure that extends road geometries (tolerance 50 meters) to form a fully connected planar network. Specifically, the terminal endpoint of each disconnected segment is extended linearly along the azimuth of its final bearing until it intersects or comes within 50 meters of the nearest endpoint or vertex of another segment, at which point the two segments are snapped and merged into a single connected arc. Road segments that remain disconnected after this extension step—because the nearest connectable segment lies beyond 50 meters—are retained as isolated features and do not participate in community partitioning; the affected community boundary is instead used directly as the splitting boundary for that area. This step is critical, as road networks serve as primary partitioning boundaries in the splitting stage.

Parcel Geometry Validation: Parcel geometries are assessed for multipart structures. Parcels with multiple geometry parts are identified and tested against community boundaries. Where parcels intersect administrative boundaries, they are clipped to ensure strict containment within a single community. This prevents cross-boundary inconsistencies during subdivision.

The preprocessing stage outputs three validated datasets, households, roads, and plots—which, together with buildings and community boundaries, form the complete input set for subsequent processing.

3.2. Hierarchical Splitting

The splitting stage applies a hierarchical subdivision strategy, conceptually aligned with split-based spatial partitioning approaches in EA design [6] [7]. A predefined household threshold is applied throughout the splitting and merging stages to control EA size: a minimum of 75 households and a maximum of 200 households per EA. This range reflects the operational requirements of Abu Dhabi's census framework, ensuring that each EA represents a manageable workload for a single enumerator while remaining statistically meaningful.

Community polygons are subdivided into progressively finer spatial units—

termed primary blocks—using three hierarchical levels.

Tier 1 Road-Based Splitting: Communities are first partitioned using the road network. Resulting polygons (roadblocks) are assigned unique identifiers and household counts. Blocks exceeding the predefined household threshold are passed to the next tier.

Tier 2 Parcel-Based Splitting: Oversized roadblocks are further subdivided using parcel boundaries. Household counts are recalculated, and only blocks exceeding the threshold proceed to the final tier.

Tier 3: Building-Based Splitting: Remaining oversized units are subdivided using building footprints, producing the smallest possible spatial units.

At this level, no further subdivision is possible. These units are retained regardless of threshold compliance. All resulting blocks are consolidated into a unified layer containing:

- UNIQUE_ID—unique feature identifier;
- COUNT—household count;
- GROUP_ID—community reference.

A critical constraint enforced throughout this stage is:

- No block boundary intersects a building footprint.

This ensures that each building is entirely contained within a single spatial unit, addressing a key limitation identified in existing methodologies.

3.3. Merge Connected Features

Following subdivision, primary blocks are aggregated into candidate EAs using a spatially constrained merging process, conceptually related to region-growing and aggregation methods [5] [6]. Blocks are processed in ascending order of household count to prioritize under-threshold units. Two blocks are defined as adjacent if they share a boundary segment of non-zero length (*i.e.*, edge-contiguity; point-only touches are excluded). For each under-threshold block, candidate neighbors are ranked in ascending order of their current household count, so that the smallest available neighbor is merged first, minimizing over-aggregation. The greedy merge continues iteratively until the combined household count reaches or exceeds the minimum threshold of 75; if no combination of available adjacent blocks within the same community can bring the total to 75 without exceeding 200, the block is retained as a sub-threshold single-EA unit consistent with the minimum-one-EA-per-community design rule.

For each block:

- If the household count exceeds the threshold → retained as-is;
- Otherwise → merged with adjacent blocks within the same community.

Neighbor selection follows a greedy strategy, prioritizing minimal aggregation required to meet the threshold. This ensures:

- spatial contiguity;
- balanced EA sizes;
- minimal over-aggregation.

The output consists of spatially contiguous EA candidates that satisfy threshold conditions wherever possible.

3.4. Merge Disconnected Features

Some blocks remain below the threshold due to spatial isolation (e.g., separated by roads or natural barriers). These are handled through a proximity-based merging strategy. Proximity is measured as centroid-to-centroid Euclidean distance; in cases where two blocks have equal centroid distances, the block with the lower UNIQUE_ID is selected as the merge partner.

For each under-threshold block:

- The nearest block within the same community is identified;
- Household counts are cumulatively aggregated until the threshold is met.

Spatial Reconciliation: Because merged units may be non-contiguous, geometric consistency is restored through:

- Convex hull generation;
- Voronoi (Thiessen) tessellation;
- Union and dissolve operations.

This process ensures gap-free coverage, non-overlapping geometries, and topological consistency. The output is a set of spatially valid EA candidates derived from previously disconnected units. The convex hull is used to define a bounding envelope around the non-contiguous constituent blocks of each merged unit, and Voronoi tessellation within that hull partitions the intermediate space in proportion to proximity to each constituent block, thereby producing a single connected polygon without arbitrary boundary decisions. Following union and dissolve, all final EAs produced through this disconnected merging pathway are single-part, topologically contiguous polygons; no multipart geometries persist in the output dataset. This reconstruction is acceptable for census operations because the Voronoi-defined boundary provides a deterministic, proximity-based area assignment that field enumerators can verify against physical landmarks, while each constituent block—and therefore the reconstructed EA—fully preserves building containment.

3.5. Postprocessing

The final stage performs quality assurance and prepares the EA dataset for operational use.

Attribute Validation: Household counts are recomputed to eliminate null or inconsistent values.

Overlap Resolution: Geometric overlaps are detected and resolved using boundary-based assignment rules, ensuring mutual exclusivity.

Boundary Alignment: EA boundaries are snapped to authoritative spatial features (roads, parcels, administrative boundaries) to improve spatial accuracy and cartographic consistency.

Unique ID Assignment: A standardized EA identifier is generated for each polygon, replacing intermediate attributes.

The resulting dataset represents a topologically valid, spatially consistent, and operationally optimized EA layer, characterized by:

- full building containment;
- balanced household distribution;
- alignment with real-world features;
- minimal manual intervention.

This methodology provides a scalable and reproducible solution for automated EA delineation, particularly suited to rapidly evolving urban environments such as Abu Dhabi. A key design principle embedded in the workflow is that every community boundary must contain at least one EA, regardless of its household count. Where an entire community falls below the minimum household threshold of 75—due to being uninhabited, sparsely settled, or predominantly industrial or desert land—it is retained as a single EA rather than being merged across community boundaries. This ensures complete spatial coverage of all administrative units, preserves the integrity of the community hierarchy, and avoids the creation of EAs that span multiple administrative boundaries, which would compromise both field navigation and the statistical coherence of census outputs.

3.6. Validation

The output EA dataset was subjected to both quantitative outcome checks and structured visual inspection. Three explicit validation criteria were applied: 1) Building boundary conformity rate—a spatial intersection test was run between all EA boundary segments and the building footprint polygons dataset; zero intersections were recorded, confirming 100% conformity (*i.e.*, no EA boundary crosses any building footprint across the emirate-wide household universe). 2) Topological completeness—the final EA layer was tested for gaps and overlaps against the union of all 1855 community boundaries; the result confirmed 0% gap area and 0% overlap area, demonstrating complete, non-redundant spatial coverage of the emirate. 3) Contiguity check—all 3298 final EAs were tested for single-part polygon geometry; EAs produced through the disconnected merging pathway were verified to be topologically contiguous single-part polygons following convex hull and Voronoi reconstruction, with no multipart geometries present in the output. Visual inspection was conducted by overlaying new EA boundaries on high-resolution satellite imagery and legacy EA polygons for a stratified sample of 50 communities across the three administrative regions (Abu Dhabi, Al Ain, and Al Dhafra), with particular attention to boundary alignment with roads, parcel edges, and building footprints.

4. Results

Table 2 summarizes the outputs of the hierarchical splitting stage. The results indicate that the majority of subdivisions occurred at the parcel level (64,750 plot blocks), highlighting the importance of parcel boundaries in refining spatial granularity beyond road-based segmentation. Building-level splitting was limited to 6313 blocks, confirming that finer subdivision was only required in dense urban areas where household thresholds could not be satisfied at higher levels.

Table 2. Summary of splitting process outputs.

Sr No	Output Name	Geometry	Records	Spatial Coverage
1	Community Blocks	Polygon	1885	Abu Dhabi Emirate
2	Roadblocks	Polygon	16,384	Community blocks exceeding threshold
3	Plot Blocks	Polygon	64,750	Roadblocks exceeding threshold
4	Building Blocks	Polygon	6313	Plot blocks exceeding threshold

Table 3 presents the results of the merging process. The connected merging stage effectively reduced 15,783 roadblocks to 1221 EA candidates, demonstrating strong spatial consolidation efficiency. Similarly, disconnected merging reduced fragmented plot and building blocks into a manageable number of spatially coherent units, with postprocessing further refining the output to eliminate geometric inconsistencies. The sum of intermediate outputs across all four pipeline stages totals 4161 EA candidates (1,392 + 1,221 + 1,332 + 216). The postprocessing stage then consolidated these into the final 3304 EAs through overlap resolution and boundary alignment operations, which merged 857 residual boundary artefacts into their neighboring units. It is also worth noting that **Table 2** reports 1885 community blocks rather than the 1855 input communities listed in **Table 1**; the additional 30 blocks arise from the multipart parcel clipping step in preprocessing, where communities intersecting parcel boundaries are split into sub-units before downstream processing.

Table 3. Summary of merging process outputs.

Sr No	Output Name	Input Records	Merging Process	Merged Output	Post-processing Output
1	Community Blocks below threshold	1392	Not Applicable	Not Applicable	Not Applicable
2	Roadblocks below threshold	15,783	Merge Connected	1221	Not Applicable
3	Plot Blocks below threshold	64,617	Merge Disconnected	1350	1332
4	Building Blocks	6313	Merge Disconnected	239	216

The workflow generated 3298 enumeration areas across Abu Dhabi Emirate, covering the emirate-wide household universe, achieving complete spatial coverage with no overlaps, no gaps. Of these, 2483 EAs (75.3%) contain at least one household. The remaining 815 EAs (24.7%) have zero households, and a further 632 EAs (19.2%) contain between 1 and 74 households—below the minimum threshold of 75. All 1447 sub-threshold EAs are the sole EA within their respective community boundary: as a core design principle, the tool guarantees at least one EA per administrative unit regardless of household count, thereby preventing any EA from spanning multiple community boundaries. As confirmed by **Table 3**, 1392 whole communities fell below the minimum threshold and were each retained as a single EA directly, representing 75.0% of the 1855 input communities. The methodology ensured 100% conformity with building footprints, with no instances of building splitting observed—addressing a key limitation of existing EA delineation approaches.

Household distribution analysis reveals a marked improvement in spatial coverage and granularity compared to the legacy EA dataset. Table 4 presents a comparative summary of key descriptive statistics for both old and new EAs. The new delineation produced 3298 EAs, representing a 57.6% increase from the 2092 legacy EAs, reflecting finer spatial resolution better suited to Abu Dhabi's evolving urban landscape. Among populated EAs (those containing at least one household), the new EAs have a mean of 79% of the threshold midpoint (137.5 HH) households and a median of 80% of the threshold midpoint, compared to a mean of 94% of the threshold midpoint and median of 89% of the threshold midpoint for the legacy dataset. The new EAs achieve a more distributed workload, with 73.3% of populated EAs falling within the target household threshold range of 75 - 200, compared to 92.9% for legacy EAs. Notably, both datasets share nearly identical 95th percentile values (195 for new EAs versus 193 for legacy EAs), confirming that the upper tail of the distribution is well-controlled in both cases. The new EAs show a substantial reduction in over-threshold units: only 1.3% of populated new EAs exceed 200 households, compared to 3.5% in the legacy dataset.

Table 4. Comparative descriptive statistics of legacy and new enumeration areas across Abu Dhabi Emirate.

Metric	Legacy EAs	New EAs
Total EAs	2092	3298
Populated EAs (HH > 0)	2091 (99.95%)	2483 (75.3%)
Uninhabited EAs (HH = 0)	1 (0.05%)	815 (24.7%)
Total Households Covered	Legacy HH universe	New HH universe
Household Statistics (Populated EAs)		
Mean Households per EA	94% of midpoint	79% of midpoint
Median Households per EA	89% of midpoint	80% of midpoint
Standard Deviation (HH)	37.8	60.5
Minimum Households	1	1
Maximum Households	87% above ceiling	63% above ceiling
Coefficient of Variation (%)	29.2%	55.5%
EAs within 75 - 200 HH Range	1943 (92.9%)	1819 (73.3%)
Spatial Characteristics		
Median Area (km^2)	0.17	0.39
Mean Area (km^2)	27.15	32.24
Median Boundary Length (km)	1.94	3.21
Mean Boundary Length (km)	7.02	10.05

The higher overall coefficient of variation (CV) for new populated EAs (55.5% versus 29.2% for legacy) reflects the spatial heterogeneity of Abu Dhabi, which encompasses dense urban centers alongside vast sparsely settled desert regions, rather than indicating poor balance within comparable settlement types. The new EAs also cover larger median geographic areas (0.39 km^2 versus 0.17 km^2) and longer median boundary perimeters (3.21 km versus 1.94 km), consistent with the broader spatial footprint of the emirate's expanded coverage zones. The household distribution across all populated EAs is further illustrated in **Figure 3**. The spatial extent and boundary configuration of new EA is presented in **Figure 4**, enabling direct visual comparison of geographic coverage and EA granularity across the emirate.

The aggregate household total differs slightly between the legacy dataset and the new EA dataset, a net increase of less than 0.1% of the legacy household universe. This difference is not a counting error but reflects the expanded spatial coverage of the new delineation: the legacy EA polygons did not cover the full extent of all 1855 community boundaries, leaving a small number of household points outside any legacy EA polygon. During the household–building preprocessing step, these previously uncovered household points—located within valid administrative community boundaries but outside legacy EA extents—were spatially associated with their nearest building centroid and incorporated into the new EA layer. The new EAs, by design, achieve complete coverage of all community boundaries, thereby capturing these previously unassigned households.

Regional analysis of the new EA dataset reveals significant variation in household distribution across Abu Dhabi's three administrative regions, reflecting the emirate's diverse geographic and demographic character. **Table 5** presents a breakdown of EA counts, household totals, and threshold compliance by region for both legacy and new datasets. The Abu Dhabi region, the most densely populated, accounts for 2053 new EAs (62.2% of total) covering 64.9% of the emirate household universe, with 77.0% of populated EAs meeting the 75 - 200 household target. Al Ain contributes 765 EAs covering 29% of the emirate household universe, with 75.5% threshold compliance among populated EAs. Al Dhafra, the vast desert region, accounts for 480 EAs but only 6.1% of the emirate household universe; its lower compliance rate of 44.0% among populated EAs directly reflects the structural sparsity of settlement in that region, consistent with the "minimum one EA per community" design rule that retains 232 uninhabited EAs as valid spatial coverage units. Notably, the new EA dataset expands the legacy coverage from 2092 to 3298 EAs across all three regions, with the greatest absolute growth in Abu Dhabi (+706 EAs) and the most proportional growth in Al Dhafra (+250.4%), reflecting more granular spatial delineation of previously coarse desert administrative boundaries.

Processing performance demonstrated substantial efficiency gains. The full dataset was processed in less than eight hours, compared to several weeks required for manual delineation. This represents a reduction of over 90% in manual effort, while maintaining consistency and reproducibility.

Table 5. Regional breakdown of legacy and new enumeration areas across Abu Dhabi Emirate's three administrative regions.

Region	Legacy EAs	New EAs	Legacy EAs HH	New EAs HH	Zero-HH New EAs	75 - 200 HH New EAs	Max HH New EAs
Abu Dhabi	1348	2053	64.9% of the emirate HH universe	64.9% of the emirate HH universe	532	1171 (77.0%)	325
Al Ain	607	765	29% of emirate HH universe	29% of emirate HH universe	51	539 (75.5%)	299
Al Dhafra	137	480	6.1% of emirate HH universe	6.1% of emirate HH universe	232	109 (44.0%)	200
Total	**2092**	**3298**	**Legacy HH universe**	**New HH universe**	**815**	**1819 (73.3%)**	**824**

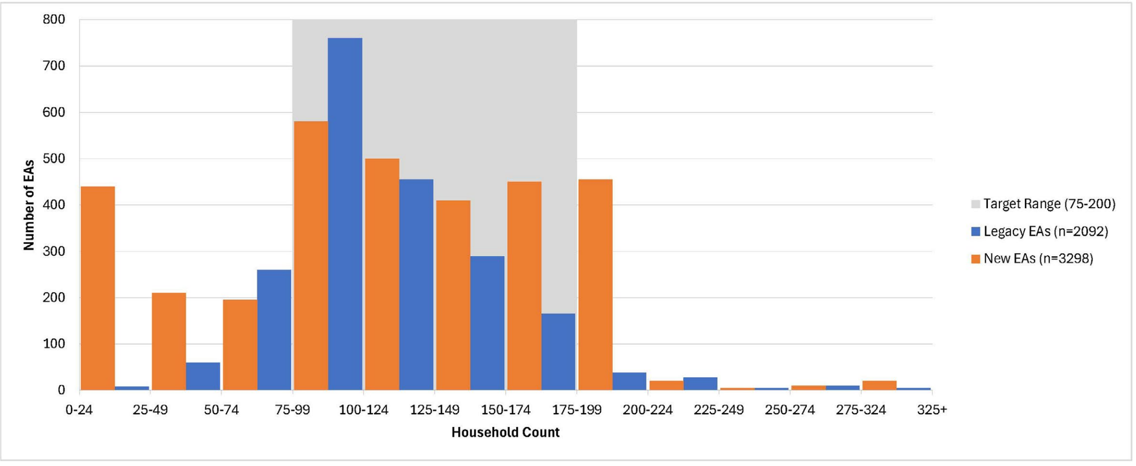
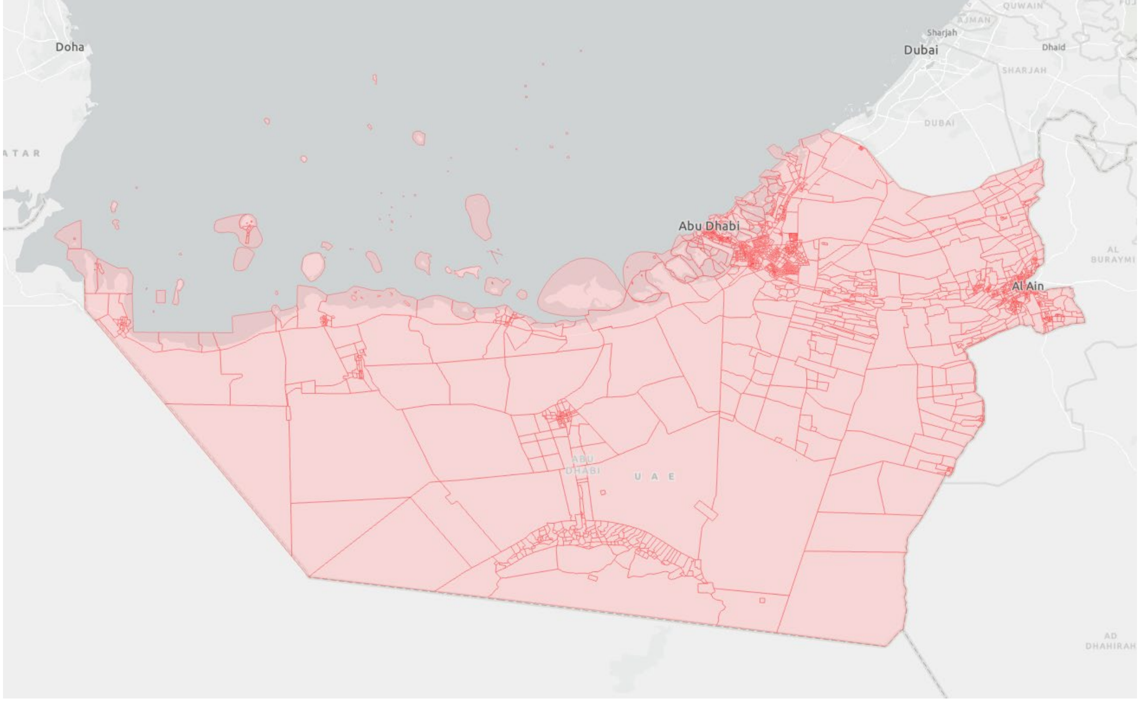


Figure 3. Household distribution of legacy and new enumeration areas (populated EAs only). The shaded band indicates the target threshold range of 75 - 200 households per EA.



Base Map Source: Esri, TomTom, Garmin, FAO, NOAA, USGS, OpenStreetMap contributors and GIS User Community.

Figure 4. Spatial distribution of new enumeration areas across Abu Dhabi Emirate.

5. Discussion

The results demonstrate that the proposed methodology significantly improves both the spatial accuracy and operational efficiency of EA delineation. Unlike traditional manual approaches and existing semi-automated techniques, the integration of hierarchical splitting and constraint-based merging enables the generation of spatial units that are both demographically balanced and geographically meaningful. Quantitatively, the new approach generated 3298 EAs covering the emirate-wide household universe, a 57.6% increase in spatial granularity over the 2092 legacy EAs, while reducing the maximum EA size from 87% above the threshold ceiling (legacy) to 63% above the ceiling (new) and achieving 100% building conformity with no boundary-building intersections.

A key contribution of this study lies in the explicit enforcement of building-level constraints. Previous approaches, including region-growing and clustering methods, often neglect the integrity of physical structures, resulting in boundaries that intersect buildings. In contrast, the proposed method guarantees full containment of each building within a single EA, improving both data accuracy and field usability.

The hierarchical splitting strategy proved effective in adapting to varying urban densities. Road-based segmentation provided efficient first-level partitioning, while parcel and building-level refinements ensured threshold compliance in dense areas. This multi-scale approach aligns with findings from previous studies but extends them by integrating administrative datasets within a unified GIS workflow. The predefined household threshold of 75 - 200 was met by 73.3% of populated new EAs, compared to 92.9% for legacy EAs. The lower in-range proportion for new EAs is a direct and expected consequence of the "minimum one EA per community" design rule: where an entire community has fewer than 75 households—whether uninhabited, desert, or sparsely settled—it is retained as a single EA rather than being merged across administrative boundaries. This constraint, which 1392 of the 1855 input communities (75.0%) triggered, is essential for preserving administrative hierarchy and field navigation integrity. Crucially, these sub-threshold EAs do not reflect algorithmic failure; they are the intended output of a spatial coverage guarantee. Importantly, only 1.3% of all new EAs exceed the upper threshold of 200 households, compared to 3.5% for legacy EAs, indicating strong control of over-sized units. Furthermore, the 95th percentile household count is nearly identical between datasets (195 new vs. 193 legacy), confirming comparable workload distribution at the upper range of the scale.

From an operational perspective, the reduction in manual effort and processing time represents a substantial improvement. The ability to generate high-quality EAs within hours enables more frequent updates, which is particularly important in rapidly urbanizing regions such as Abu Dhabi.

However, the methodology is dependent on the availability and quality of administrative data. In areas with incomplete or outdated datasets, additional pre-processing or manual validation may be required. Sparse desert regions present a structural challenge for threshold compliance: communities with very low or zero household density will always fall below the minimum of 75 households, but the design intentionally retains them as single EAs rather than merging across community lines. This trade-off between threshold compliance and administrative integrity is deliberate—the 1447 sub-threshold EAs in the output are not algorithmic residuals but a guaranteed spatial coverage outcome. Additionally, the current implementation requires ArcGIS Pro, which may limit adoptability in resource-constrained settings; future open-source adaptations would broaden applicability.

Future research should explore the integration of predictive models using remote sensing and mobility data to dynamically estimate population distribution, further enhancing automation and reducing dependency on static datasets.

6. Conclusions

This study presents a fully automated, GIS-integrated methodology for the delineation of enumeration areas, addressing critical limitations of existing approaches in accuracy, efficiency, and spatial consistency. By combining hierarchical splitting with constraint-based merging, the proposed method ensures balanced household distribution while preserving the integrity of real-world features, particularly building footprints. Applied to Abu Dhabi Emirate, the methodology generated 3298 EAs covering the emirate-wide household universe—a 57.6% increase in spatial granularity over the 2092 legacy EAs—with 100% conformity to building boundaries, a reduction in maximum EA size from 87% above the threshold ceiling (legacy) to 63% above the ceiling (new), and a processing time of under eight hours compared to several weeks for manual approaches.

The implementation within ArcGIS Pro enables seamless integration into existing workflows of national statistical offices, providing a practical and scalable solution for census mapping. The results from Abu Dhabi demonstrate that the methodology can significantly reduce processing time and manual intervention while improving the quality and usability of EA boundaries.

The primary contribution of this research lies in the development of a building-aware, fully automated delineation framework that bridges the gap between theoretical spatial clustering methods and operational GIS implementation. This approach is particularly relevant for rapidly evolving urban environments where frequent updates to census geography are required.

Future work will focus on enhancing the adaptability of the model through integration with real-time data sources and machine learning techniques, enabling dynamic and continuous EA updates. The methodology is readily transferable to other regions, offering a robust foundation for modernizing census mapping practices globally.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] United Nations (2017) Principles and Recommendations for Population and Housing Censuses, Rev. 3, Statistical Papers Series M No. 67/Rev.3. United Nations.

[2] Camara, M., Bengaly, S. and Coulibaly, O. (2025) A Semi-Automated Geographic Information System Approach for Pre-Census Mapping. *Journal of Geography, Environment and Earth Science International*, **29**, 12-26. https://doi.org/10.9734/jgeesi/2025/v29i11965

[3] United Nations Economic Commission for Europe (UNECE) (2018) Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses. United Nations.

[4] United Nations Statistics Division (2009) Handbook on Geospatial Infrastructure in Support of Census Activities, Studies in Methods, Series F, No. 103. United Nations.

[5] Cockings, S. and Martin, D. (2005) Zone Design for Environment and Health Studies Using Pre-Aggregated Data. *Social Science & Medicine*, **60**, 2729-2742.

[6] Qader, S., Lefebvre, V., Tatem, A., Pape, U., Himelein, K., Ninneman, A., *et al.* (2021) Semi-Automatic Mapping of Pre-Census Enumeration Areas and Population Sampling Frames. *Humanities and Social Sciences Communications*, **8**, Article No. 3. https://doi.org/10.1057/s41599-020-00670-0

[7] Qader, S., Lefebvre, V., Ninneman, A., Himelein, K., Pape, U.J., Bengtsson, L., Tatem, A. and Bird, T. (2019) A Novel Approach to the Automatic Designation of Predefined Census Enumeration Areas and Population Sampling Frames: A Case Study in Somalia. World Bank Policy Research Working Paper No. 8972.