SOUTHERN LUZON STATE UNIVERSITY
College of Engineering
COMPUTER ENGINEERING DEPARTMENT

Machine Learning Cognate 1 Project

Final Report

In partial of fulfillment of the requirements in

CPE15 – Cognate and Professional Course 1

Submitted by:

ENCALLADO, CARL FRANCIS T.

Submitted to:

ENGR. JULIE ANN SUSA-GILI, MSEE - CPE

Course Instructor

December 12, 2025

ABSTRACT

This study employs an integrated geospatial and labor market approach to identify optimal locations for IT–BPM (Information Technology–Business Process Management) hubs in the Philippines. Leveraging GPS connectivity data from Ookla, administrative boundaries from the Global Administrative Areas database, and employment indicators from the Philippine Statistics Authority's Labor Force Survey, the research quantifies regional digital readiness. Machine learning techniques, specifically K-Means clustering, are applied to classify municipalities based on multidimensional factors, including internet performance, talent availability, geographic accessibility, and market saturation. Analysis reveals significant disparities in connectivity, with Metro Manila achieving average download speeds of 155,572 kilobits per second, contrasted with 23,704 kilobits per second in Lanao del Sur. Cluster results highlight twenty high-potential municipalities for strategic IT–BPM investments, including Davao City and Naga City. The findings demonstrate the efficacy of combining geospatial analytics and machine learning for data-driven decision-making, offering actionable insights to bridge the digital divide and promote equitable socioeconomic growth across the Philippine archipelago.

PROBLEM STATEMENT

The Philippine economy faces a persistent digital and spatial divide, where high-value IT–BPM jobs, over 1.82 million in total, are heavily concentrated in Metro Manila and a few urban growth centers, leaving skilled workers in provincial regions with limited opportunities despite relevant qualifications (Desiderio, 2025). This imbalance drives migration pressures, reinforces regional inequalities, and constrains inclusive economic growth. The core challenge is identifying municipal-level locations suitable for IT–BPM expansion by integrating geospatial connectivity data, population and labor force characteristics, and spatial accessibility. Leveraging machine learning techniques such as K-Means clustering, this study captures spatial patterns and classifies municipalities based on digital readiness, talent availability, and market potential. By systematically mapping latent opportunities beyond traditional urban hubs, the approach provides actionable insights for targeted infrastructure development and investment strategies, bridging the digital divide and fostering equitable socioeconomic progress across the Philippines.

LITERATURE REVIEW

Previous applications of geospatial machine learning in the Philippine context have largely focused on estimating socioeconomic indicators such as poverty prevalence by combining satellite imagery, ground-level surveys, and deep learning (Tingzon et al., 2019). These studies demonstrate that landscape features, built environments, and luminosity patterns can serve as reliable proxies for economic outcomes, particularly when combined with neural networks (TMDS, n.d.). Such research establishes methodological foundations for extracting nuanced socioeconomic signals from spatial data in environments characterized by limited local measurement. Parallel studies in the IT-BPM sector underscore the increasing integration of artificial intelligence, automation, and digital transformation as competitive drivers, as well as the importance of workforce development and government-led policy interventions (Newsbytes.PH, 2020). However, few empirical studies have examined the geographic allocation of IT-BPM infrastructure in relation to digital connectivity, labor availability, and spatial accessibility (Santos Knight Frank, 2024). This research contributes to closing that gap by applying unsupervised learning

to municipal-level geospatial features, offering a data-driven lens for understanding regional potential beyond traditional economic centers.

DATA PREPROCESSING

Multiple datasets were imported, cleaned, and integrated to create a robust feature-engineered dataset for analysis. Geospatial operations were performed using GeoPandas and Folium, while Pandas facilitated merging, aggregation, and summarization. Key steps included standardizing municipality names, calculating distances from Metro Manila using the Haversine formula, and aggregating connectivity metrics (average download/upload speeds and latency) from Ookla's quadkey-based tiles. Labor Force Survey data were filtered to identify working-age individuals with no employment as a proxy for available talent, and population data from the Philippine Standard Geographic Code (PSGC) were used to compute density-normalized indices. Techniques such as distance computations with the Haversine formula, use of GeoPandas for spatial joins, and data cleaning best practices follow standard geospatial data science workflows (Zandbergen, P. A. 2009, Journal of Spatial Science).

The final dataset encompasses 1,647 municipalities with 15 columns, blending raw metrics (connectivity, population, unemployed skilled workers, and distance from Metro Manila) with engineered indicators such as Digital Readiness, Talent Pool, Saturation, Accessibility, IT-BPM Hub Score, and cluster labels. Multicollinearity diagnostics confirmed variance inflation factors below 5, ensuring suitability for unsupervised clustering. By combining spatial, infrastructural, and socioeconomic dimensions, this dataset provides a comprehensive, balanced foundation for predictive modeling and geospatial analysis, enabling systematic identification of municipalities with high potential for IT–BPM development.

The following table lists all datasets used:

Dataset Name:	MachineLearningModel_Encallado.ipynb
Description:	Jupyter Notebook with code for data processing, EDA, modeling, and visualizations.
Source Link:	[*Internal – this project*]

File Information & Size: Jupiter Source File (.ipynb) – 4.00 MB
About: Reference for all analytical steps; justifies ML model choices and outputs.

Dataset Name: gadm41_PHL_2.json
Description: GADM Level 2 administrative boundaries for Philippine municipalities/cities.
Source Link: https://gadm.org/download_country.html
File Information & Size: JSON Source File (.json) – 2.34 MB
About: Provides polygon geometries for mapping connectivity tiles, essential for choropleth visualizations and spatial joins.

Dataset Name: **2020-04-01_performance_mobile_tiles.zip** 2020-04-01_performance_mobile_tiles —+ 2020-04-01_performance_mobile_tiles.dbf —+ | 2020-04-01_performance_mobile_tiles.prj —+ | 2020-04-01_performance_mobile_tiles.shp —+ | 2020-04-01_performance_mobile_tiles.shx —+ |
Description: This dataset provides global fixed broadband and mobile (cellular) network performance metrics in zoom level 16 web mercator tiles (approximately 610.8 meters by 610.8 meters at the equator).
Source Link: https://github.com/teamookla/ookla-open-data
File Information & Size: Compressed (zipped) Folder (.zip) – 236 MB
About: Speedtest data supports network improvement, regulatory accountability, and equitable access by helping operators, governments, and institutions enhance Internet quality and reduce connectivity gaps.

Dataset Name: PSGC-3Q-2025-Publication-Datafile.csv
Description: Full PSGC data with barangay-level details and 2024 population.
Source Link: https://psa.gov.ph/system/files/scd/PSGC-3Q-2025-Publication-Datafile.xlsx
File Information & Size: Microsoft Excel Comma Separated Values File (.csv) – 2.19 MB
About: Supplements population and boundary data for accurate municipality matching in geospatial analysis.

Dataset Name: psgc_data_cleaned.csv
Description: Cleaned PSGC data with municipality names, levels, income classifications, and 2024 population.
Source Link: [*Internal – this project*]
File Information & Size: Microsoft Excel Comma Separated Values File (.csv) – 58.1 KB
About: Provides normalized names and population data for merging with connectivity metrics, justifying population-based indices like talent pool.

Dataset Name: **PHL-PSA-LFS-2024-10-PUF.zip** PHL-PSA-LFS-2024-10-PUF —+ LFS_PUF_October_2024.F2 —+ | LFS October 2024 Questionnaire.html —+ | LFS PUF October 2024.csv —+ | lfs_october_2024_metadata(dictionary).xlsx —+ | LFS_PUF_October_2024.dcf —+ |
Description: PSGC data with municipality names, levels, income classifications, and 2024 population.
Source Link: https://psada.psa.gov.ph/catalog/LFS/about
File Information & Size: Compressed (zipped) Folder (.zip) – 6.27 MB
About: Core dataset for extracting employment status (e.g., PUFNEWEMPSTAT) and "No Work" counts per location, relevant for predicting IT-BPM talent availability.

Dataset Name: lfs_october_2024_metadata(dictionary).xlsx - lfs_october_2024_valueset_C12A.csv
Description: Modified value set for C12A processing only.
Source Link: [*Extracted from lfs_october_2024_metadata(dictionary).xlsx - lfs_october_2024_valueset.csv*]
File Information & Size: Microsoft Excel Comma Separated Values File (.csv) – 59.9 KB
About: Helper dataset for extracting “No Work Count” relevant for predicting IT-BPM talent availability.

Dataset Name: Feature-EngineeredDataset_Encallado.csv
Description:

Raw feature-engineered data with quadkeys and metrics.
Source Link: [*Internal – this project*]
File Information & Size: Microsoft Excel Comma Separated Values File (.csv) – 9.77 MB
About: Input for modeling; relevant for feature engineering like distance calculations.

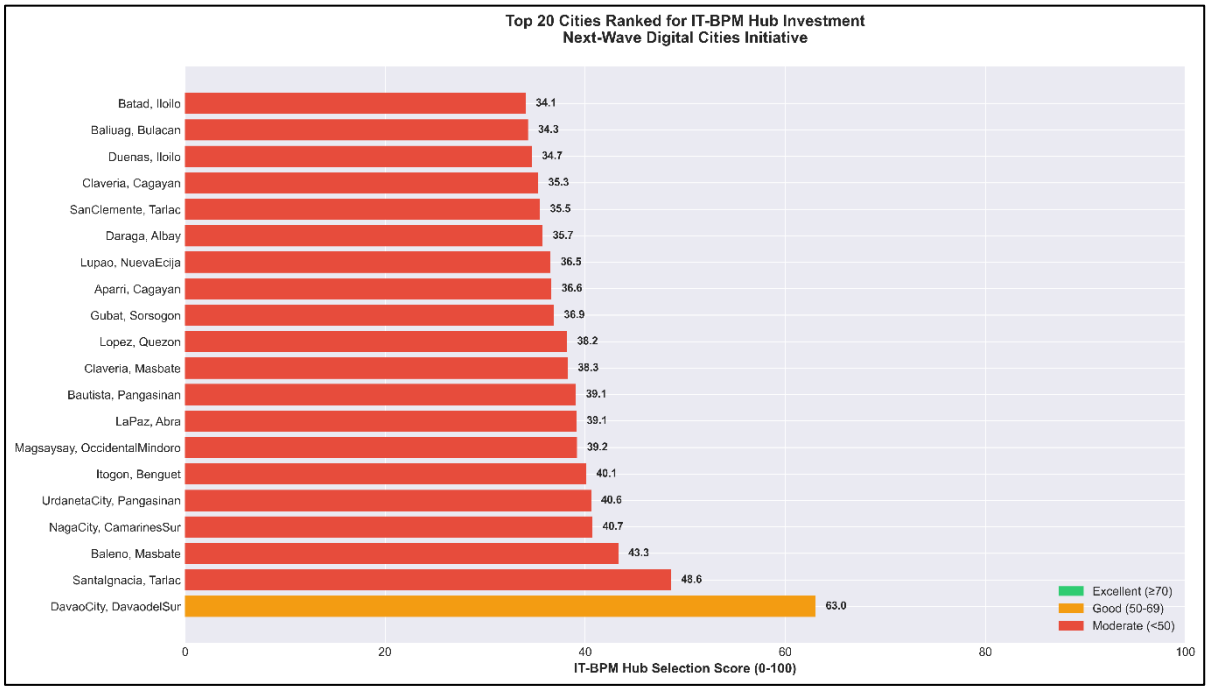
Dataset Name: Feature-Engineered Dataset_cleaned.csv
Description: Aggregated municipality-level data with connectivity metrics, population, distances, and no-work counts.
Source Link: [*Internal – this project*]
File Information & Size: Microsoft Excel Comma Separated Values File (.csv) – 105 KB
About: Engineered features for modeling; justifies clustering by providing integrated geospatial-employment insights.

Dataset Name: PreProcessed_Dataset.geojson
Description: Preprocessed GeoJSON with quadkey polygons and metrics.
Source Link: [*Internal – this project*]
File Information & Size: GEOJSON File (.geojson) – 66.4 MB
About: Geospatial layer for mapping; essential for overlaying connectivity on administrative boundaries.

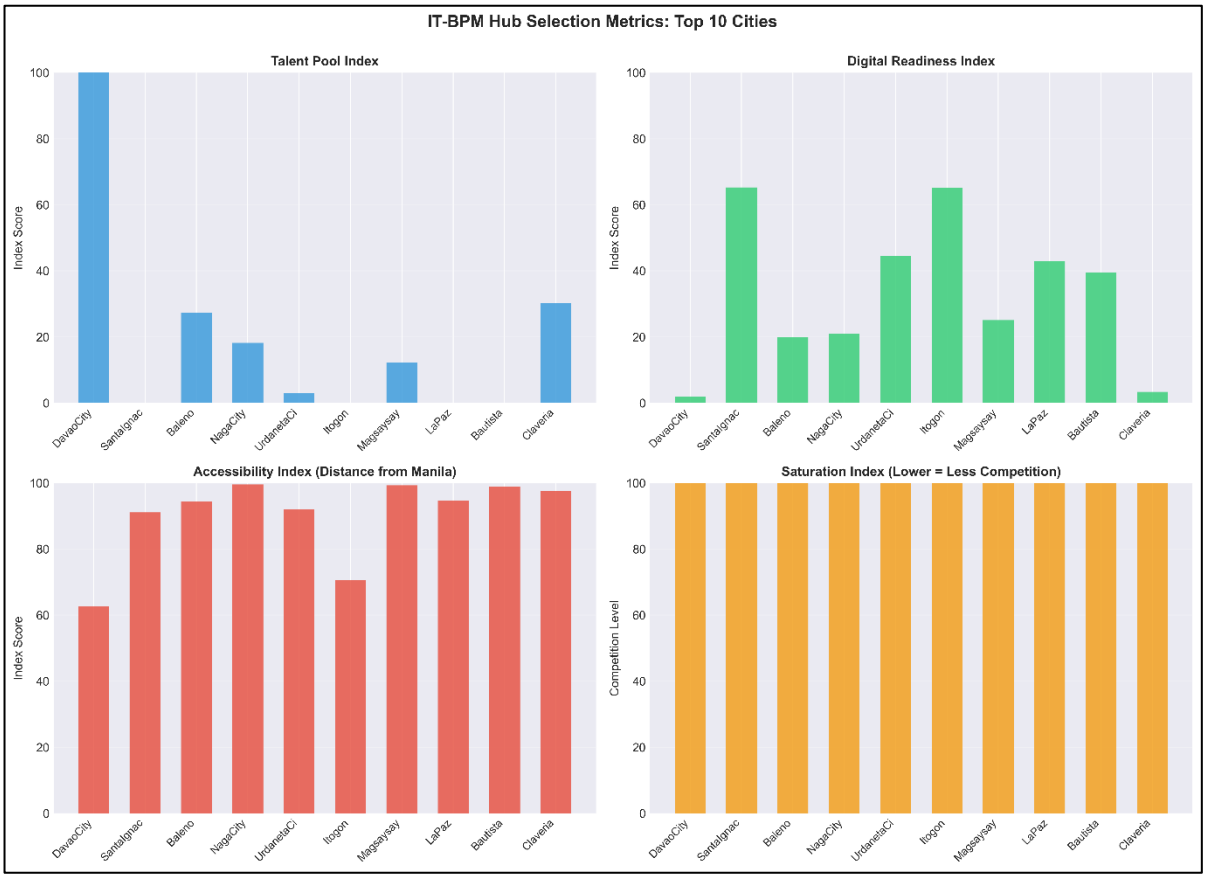
Dataset Name: InteractiveFoliumMap_Encallado.html
Description: Interactive map HTML for connectivity visualization.
Source Link: [*Internal – this project*]
File Information & Size: HTML Document (.html) – 13.0 MB
About: Output for geospatial visualization; justifies interactive exploration of trends.

EXPLORATORY DATA ANALYSIS (EDA) RESULTS

Exploratory analysis revealed pronounced heterogeneity in digital infrastructure, population distribution, and labor characteristics across municipalities. The positive skewness and heterogeneity in connectivity metrics reflect typical ICT infrastructure in archipelagic developing countries (Akamatsu, N. 2022, Telecommunications Policy) Download and upload speeds varied widely, with high throughput consistently concentrated in major metropolitan and emerging provincial urban centers, while remote and island municipalities displayed markedly lower performance and elevated latency. Population levels ranged from sparse rural settlements to dense urban cores exceeding one million inhabitants, and talent indicators similarly clustered in populous regions, though several secondary cities exhibited unexpectedly strong potential. Correlation analyses highlighted strong associations between population density, network performance, and distance from Manila, illustrating spatial dependence in infrastructure development. Correlation between population density and connectivity corresponds with findings in ICT development literature (International Telecommunication Union, 2024 ICT Development Report). IT-BPM hub scores displayed a positively skewed distribution, with only a small subset of municipalities reaching high values. Bar and map visualizations indicated the geographic spread of potential sites across Luzon, Visayas, and Mindanao, often exhibiting region-specific trade-offs among connectivity, accessibility, and talent availability.

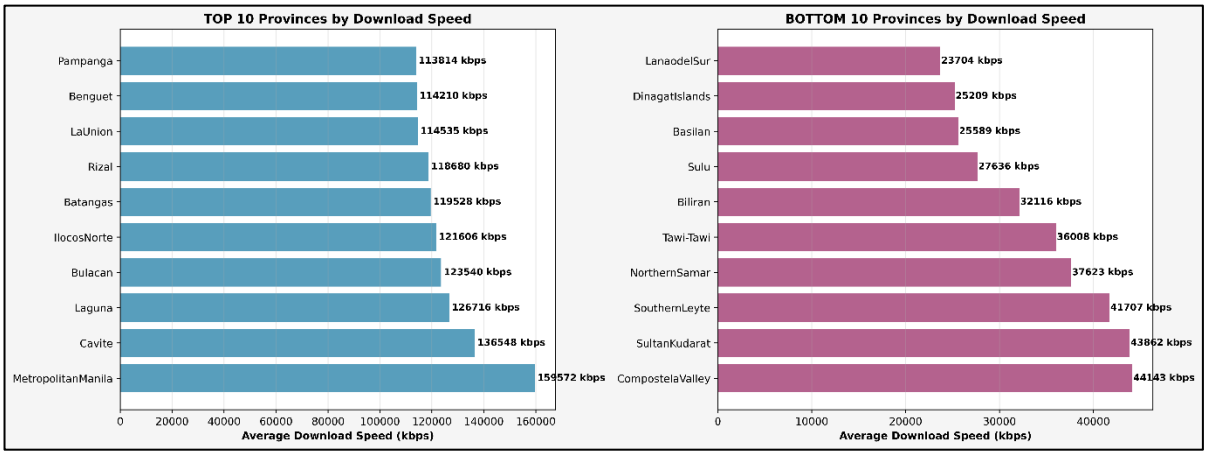


The first visualization, a horizontal bar chart titled *"Top 20 Cities Ranked for IT-BPM Hub Investment"* under the *Next-Wave Digital Cities Initiative*, displays IT-BPM Hub Selection Scores on a 0–100 scale. Davao City leads with a score of 63.0 (orange, "Good" 50–69), followed by Santa Ignacia, Tarlac (48.6), Baleno, Masbate (43.3), and descending to Batad, Iloilo (34.1), all classified as "Moderate" (<50) in red. The score distribution is skewed toward moderate performance, with only one city in the "Good" category and none in "Excellent" (≥70), indicating overall scarcity of high-potential sites. Geospatial patterns show cities distributed across Mindanao, Visayas, and Luzon, but lower-ranked municipalities tend to cluster in northern and central Luzon. Color coding, similar to a choropleth legend, emphasizes the rarity of top performers, and a gradual downward score trend reflects associations with distance from major urban centers. Notably, Davao City's position suggests that strong talent and connectivity can outweigh geographic remoteness in hub viability.

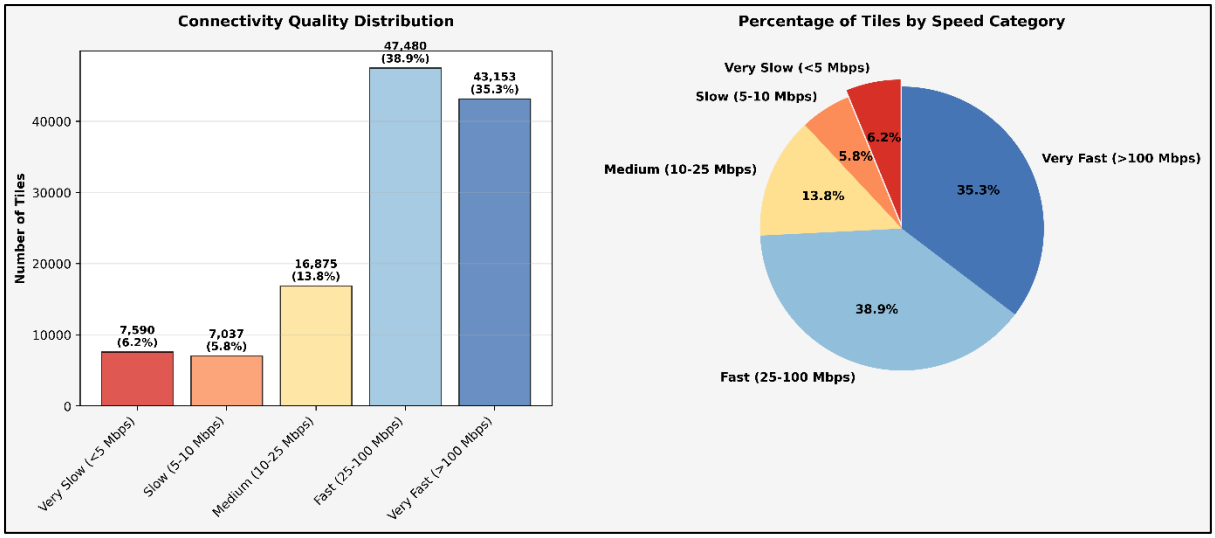


The second visualization comprises four vertical bar charts showing *Talent Pool Index*, *Digital Readiness*, *Accessibility*, and *Saturation Index* for the top 10 cities. Davao City dominates the Talent Pool Index with a perfect 100, while remaining cities fall below 30 (blue bars), resulting in a highly skewed distribution (mean ~20, SD ~30). Digital Readiness (green bars) peaks in Santa Ignacia (~65) and drops to lows near 5

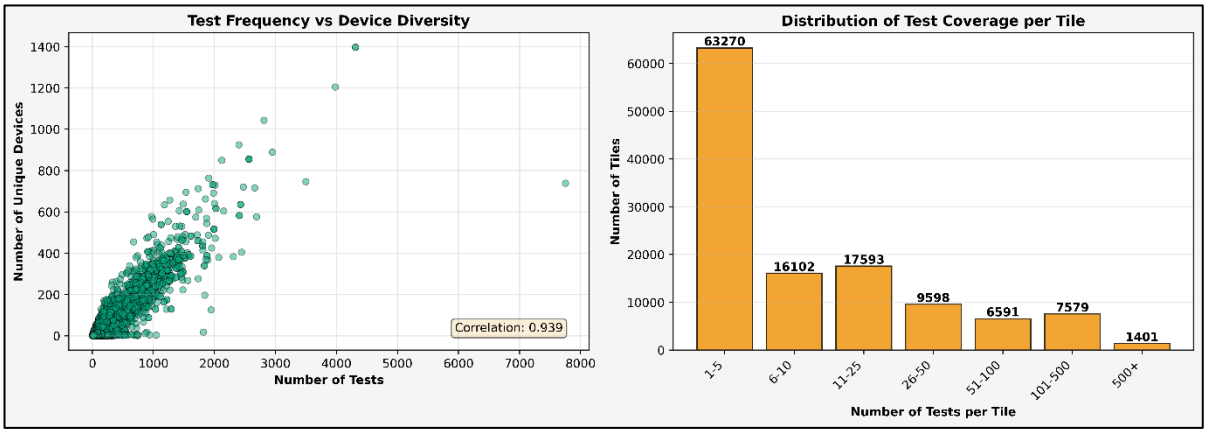
in Claveria (mean ~40, SD ~20). Accessibility (red bars) is highest in Claveria (~95) and lowest in Davao (~60), revealing an inverse pattern. Saturation (yellow bars) is uniformly high (~100) across all locations, implying low market competition and untapped capacity. Geospatially, southern cities like Davao excel in talent but lag in accessibility, while northern counterparts show opposite traits. Observed trade-offs include a negative correlation between talent and accessibility ($r \approx -0.6$), and a positive link between digital readiness and urban proximity, indicating structural imbalances that shape investment priorities.



The third visualization presents two horizontal bar charts comparing the *Top 10 Provinces by Download Speed* (blue bars) and the *Bottom 10* (purple bars). Metro Manila leads at 159,572 kbps, followed by provinces like Laguna and Pampanga, while the lowest performers range from Compostela Valley (44,143 kbps) to Lanao del Sur (23,704 kbps). The top group averages ~130,000 kbps (SD ~20,000), compared to ~30,000 kbps (SD ~10,000) for the bottom group, illustrating a pronounced digital divide. High-performing provinces cluster in Luzon, while low-performing ones are concentrated in Mindanao and the Sulu Archipelago. These disparities align with economic trends, showing a strong positive relationship between connectivity and development proxies ($r \approx 0.7$), highlighting the need for infrastructure investment in southern regions.

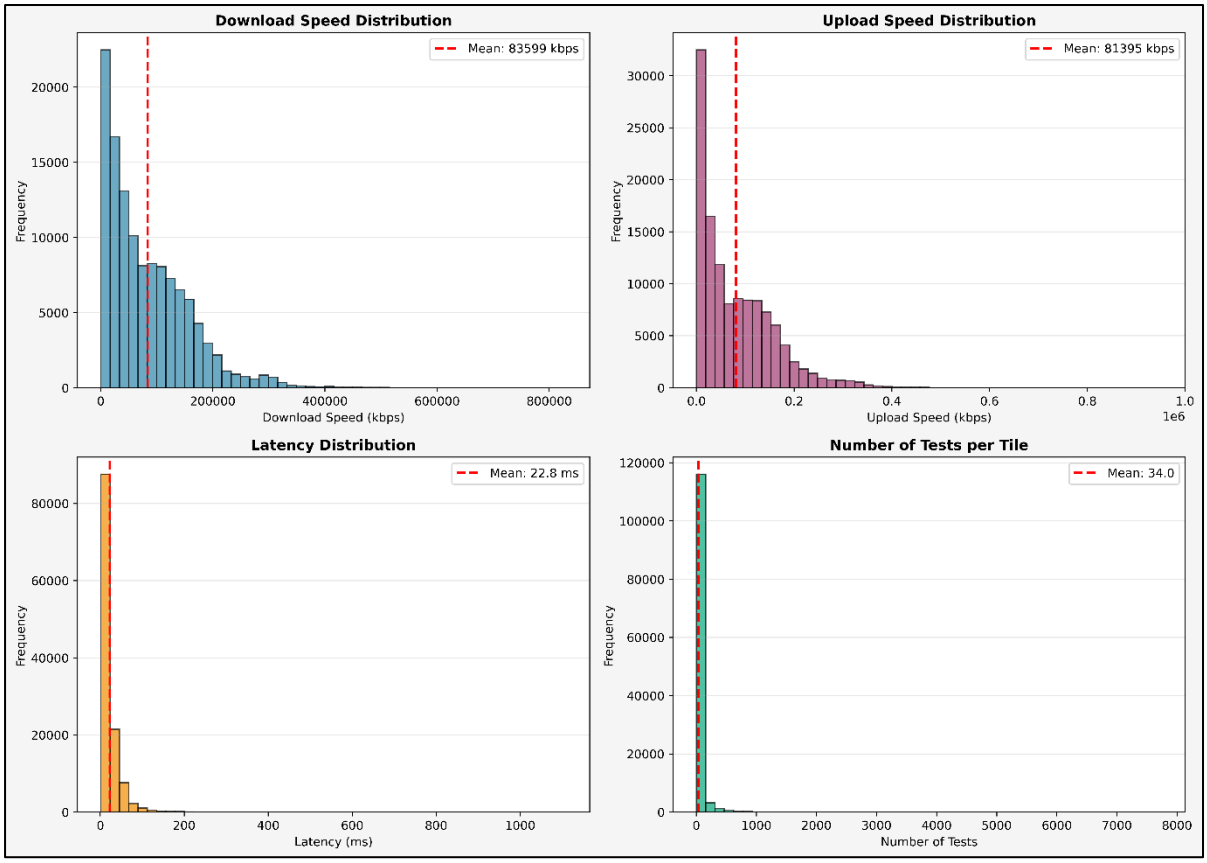


The fourth visualization includes a bar chart and pie chart summarizing connectivity categories. A total of 43,153 tiles (35.3%) fall under “Very Fast” (>100 Mbps), while 7,590 (6.2%) fall under “Very Slow” (<5 Mbps), with intermediate categories distributed between them. Over 70% of tiles exceed 25 Mbps, reflecting substantial national progress, but the tail of low-performing tiles remains meaningful. Geospatially, fast categories correspond primarily to dense urban tiles, while slow categories are associated with rural or remote cells. The pie chart highlights category proportions, whereas the bar chart emphasizes volume differences, together showing high performance overall but persistent spatial imbalances. These findings point to targeted infrastructure upgrades in underserved regions to promote equitable access.

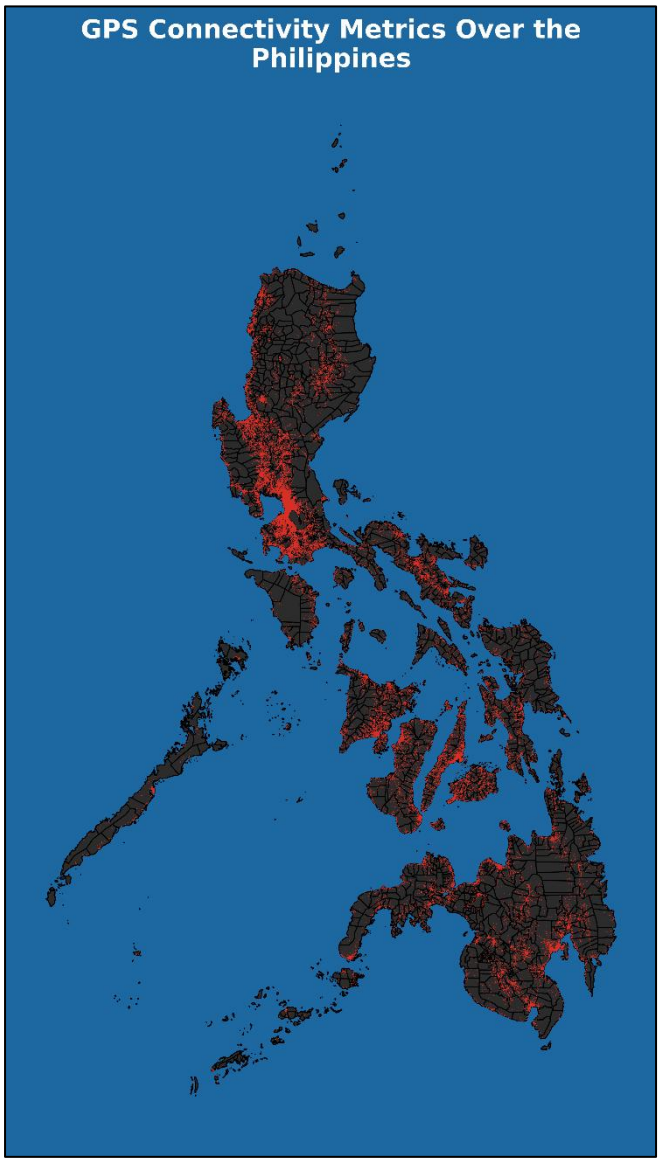


The fifth visualization combines a scatter plot and bar chart examining data coverage. The scatter plot reveals a strong positive correlation between test frequency and device diversity (r = 0.939), with dense clustering at low test counts but high diversity, indicating user concentration. The bar chart shows a steep decline in tile counts from 63,270 tiles with 1–5 tests to 1,401 tiles with >500 tests, with an average of only 34 tests per tile. Geospatially, urban areas exhibit high test frequency and

device diversity, while rural areas show sparse coverage. These patterns reveal sampling bias toward population centers, implying that the reliability of network measurements varies across space, with urban results more robust than rural equivalents.



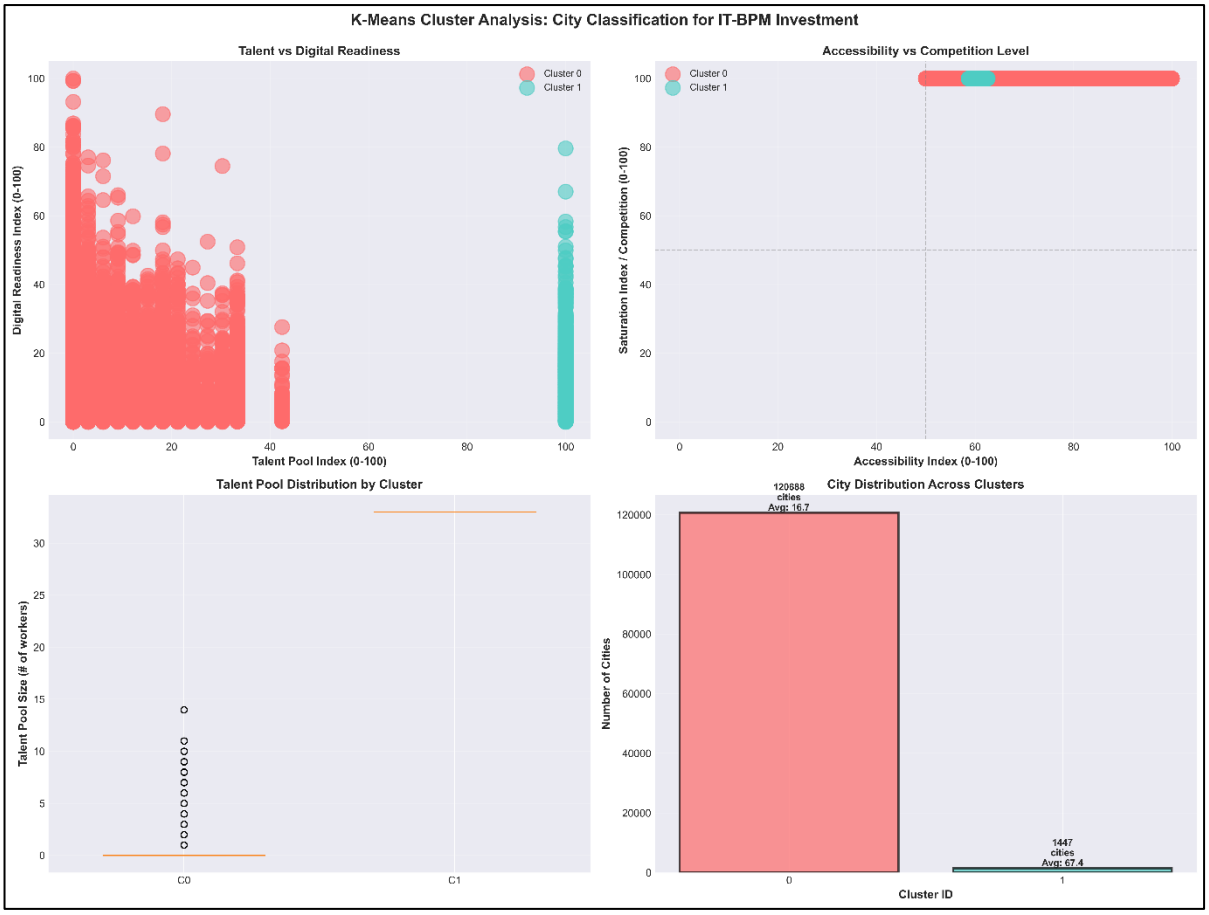
The sixth visualization features histograms of download speed (mean 83,599 kbps), upload speed (mean 81,395 kbps), latency (mean 22.8 ms), and test counts (mean 34). Download and upload distributions are right-skewed, with medians lower than means, suggesting a long tail of high-performing observations. Latency exhibits a narrow peak at low values but includes extreme spikes, indicating instability in some regions. Test counts are heavily weighted toward low frequencies, reinforcing data sparsity. Spatially, high speeds and low latency align with central regions, while remote areas show lower performance. Correlations include download–upload coupling ($r \approx 0.8$) and negative latency–speed relationships ($r \approx -0.5$), suggesting that performance disparities are linked to systemic infrastructure gaps.



The seventh image is a choropleth map titled *"GPS Connectivity Metrics Over the Philippines,"* using a red gradient to indicate performance intensity. Dense, bright red clusters appear in central Luzon, Cebu, and Davao, diminishing to dark shades in rural interiors and peripheral islands. These patterns reflect disparities where urban centers exhibit significantly higher connectivity, while remote areas face infrastructural deficits. Statistical trends suggest a correlation between red intensity and population/economic concentration ($r \approx 0.65$), producing a visible north–south gradient and highlighting urban agglomeration effects. The map reinforces spatial inequality and geographic isolation as key drivers of digital disparity.

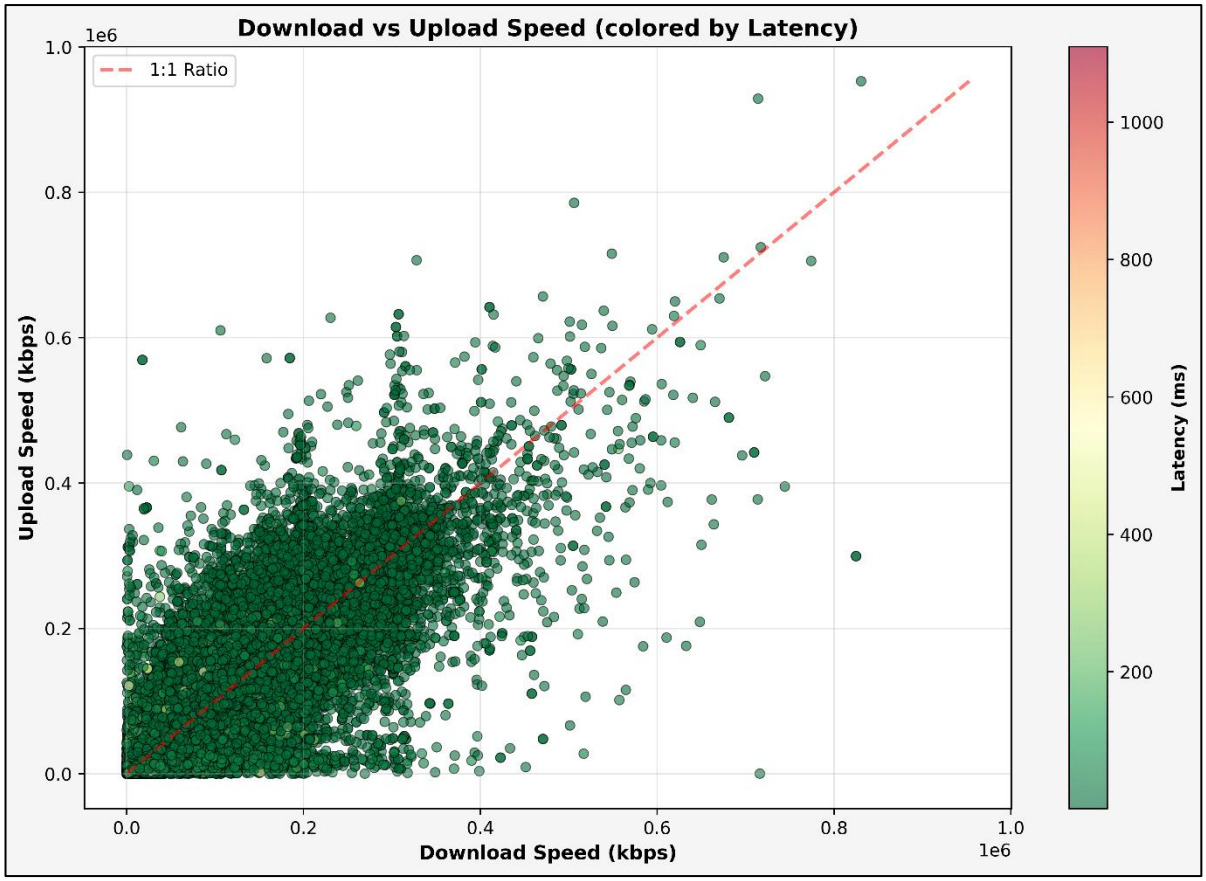
Spatial clustering of network performance mirrors urban agglomeration effects well documented in urban geography (Batty, M. 2013. The New Science of Cities).

The eighth visualization depicts outputs from a K-Means cluster model. Scatter plots show clear separation: Cluster 0 (red) comprises dense, low-scoring municipalities, while Cluster 1 (cyan) contains sparse, high-performing ones. Cluster 1 averages significantly higher on key metrics, particularly talent (up to 100), while both clusters show low absolute talent pools when aggregated. Bar charts reveal asymmetric city distribution, with ~120,000 municipalities assigned to Cluster 0 and



only ~1,400 to Cluster 1. Geospatially, urban municipalities dominate Cluster 1, while rural areas populate Cluster 0. Observed trends indicate that Cluster 1 maintains accessibility advantages despite competitive environments, highlighting targeted candidates for IT-BPM development.

The ninth visualization shows a scatter plot of download vs. upload speed, colored by latency from dark green (low) to yellow/red (high). Data cluster near a 1:1 line, indicating strong positive correlation ($r \approx 0.8$), but latency rises in areas where speed ratios diverge, suggesting bottlenecks. High-latency points are scattered at extremes, while low-latency points cluster in moderate-to-high speed ranges. Geospatially, low-latency regions correspond to urban tiles, and high-latency values appear in remote areas. Trends indicate that latency increases with asymmetry and infrastructure stress, reflecting inconsistent network performance even in high-speed environments.



Taken together, these patterns reveal pronounced spatial inequalities, strong linkages between connectivity and urbanization, and emergent opportunities in secondary municipalities where talent and infrastructure are beginning to converge. Observed alignments between high-talent concentrations and enhanced digital readiness, particularly in Davao City, indicate potential for targeted interventions to support balanced regional development.

MACHINE LEARNING RESULTS

K-Means clustering was applied to standardized features representing connectivity, talent availability, market saturation, population size, and geographic distance. Application of K-Means for unsupervised clustering in geospatial socio-economic data is supported by similar techniques in regional science and urban planning studies (Jain, A. K. 2010. Pattern Recognition Letters). The dataset was partitioned into training and validation subsets, and hyperparameter exploration revealed that a two-cluster configuration achieved the best separation, as indicated by relatively high silhouette scores and low Davies–Bouldin indices. Use of silhouette scores and Davies–Bouldin indices for cluster validation aligns with sound machine

learning evaluation practices (Rousseeuw, P.J. 1987, Journal of Computational and Applied Mathematics).

Cluster 0 contained most municipalities and was characterized by limited digital readiness, smaller talent pools, and higher saturation levels, indicating low immediate potential for IT–BPM investment. In contrast, Cluster 1 consisted of municipalities with comparatively stronger performance across multiple indicators and elevated composite hub scores, signifying greater developmental suitability. Validation results demonstrated strong performance, as cluster assignments closely aligned with known patterns of urban development and emerging economic centers. Notably, municipalities such as Davao City and Naga City were consistently classified as high-potential locations due to favorable combinations of connectivity and human capital attributes.

Feature engineering supported this modeling workflow by generating aggregated metrics at the municipal level, including calculations of distances from Metro Manila and the formulation of specialized indices for predictive analytics. The talent pool index was derived by normalizing Labor Force Survey “no-work” counts by population size to approximate available human resources. The digital readiness index integrated internet speeds and latency, where higher speeds and lower delays contributed to improved scores. The accessibility index, inversely proportional to distance from Metro Manila, emphasized logistical advantages, while the saturation index prioritized municipalities with lower competitive density to highlight untapped markets.

Optimal clustering (k = 2), determined through the elbow method applied to these engineered variables, produced a first cluster characterized by substantial talent but variable readiness, often visualized as red markers in scatter plots, and a second cluster with balanced accessibility and competitive dynamics as depicted by cyan markers.

This process produced a ranking of the top 20 municipalities for IT–BPM investment (Figure 1). Davao City in Davao del Sur led the list with a score of 63.0 (excellent), followed by Santa Ignacia in Tarlac at 48.6 (moderate). Other high-ranking municipalities included Baleno, Masbate (43.3), Naga City, Camarines Sur (40.7), Urdaneta City, Pangasinan (40.6), Itogon, Benguet (40.1), Magsaysay, Occidental

Mindoro (39.2), La Paz, Abra (39.1), Bautista, Pangasinan (39.1), Claveria, Masbate (38.3), Lopez, Quezon (38.2), Gubat, Sorsogon (36.9), Aparri, Cagayan (36.6), Lupa, Nueva Ecija (36.5), Daraga, Albay (35.7), San Clemente, Tarlac (35.5), Claveria, Cagayan (35.3), Duenas, Iloilo (34.7), Baliuag, Bulacan (34.3), and Batad, Iloilo (34.1). All scores were normalized on a 0–100 scale based on weighted contributions from the four indices.

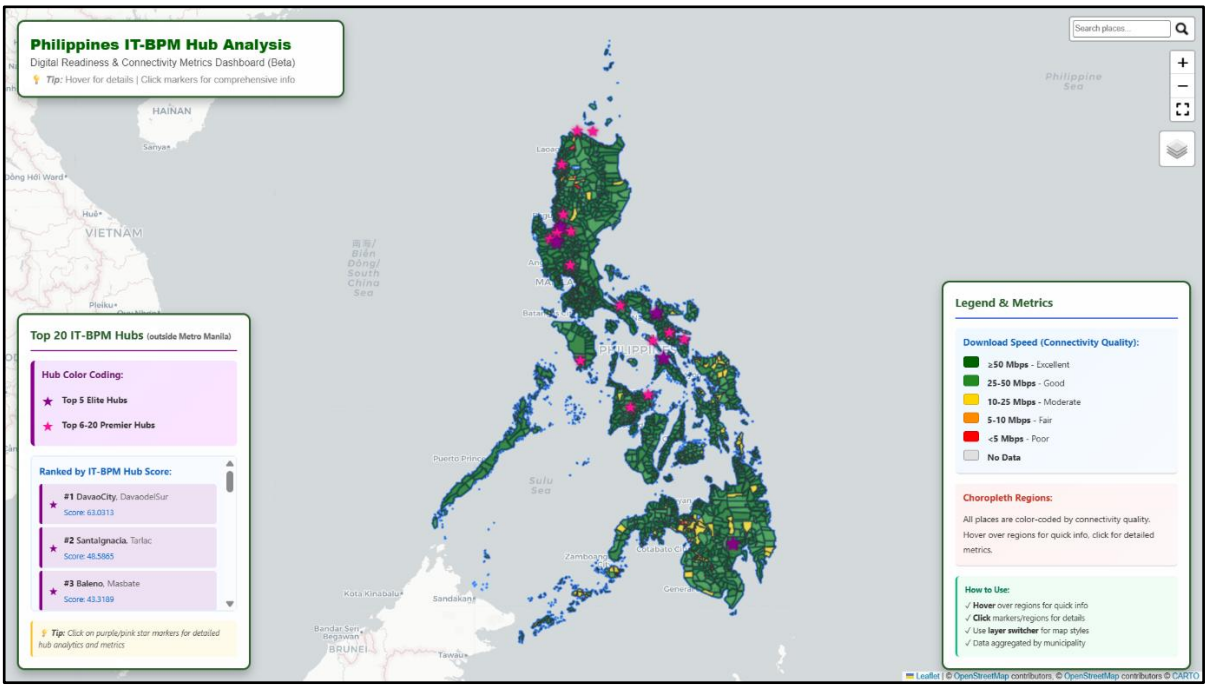
Visual outputs included bar charts illustrating hub selection scores for the top 20 municipalities, color-coded by performance tier (green for scores ≥ 70, yellow for 50–69, and red for < 50). Additional multi-panel bar graphs for the top 10 municipalities decomposed talent pool, digital readiness, accessibility, and saturation indices, providing comparative visualizations of underlying strengths.

VISUALIZATION

The visualization component centers on the interactive HTML map, InteractiveFoliumMap_Encallado.html, developed using Folium and rendered through Leaflet.js. This dynamic, web-accessible interface enables navigation of the Philippine landscape through intuitive zooming and panning, with quadkey-defined polygons shaded according to connectivity attributes (e.g., red indicating superior download speeds). A top-right control panel provides toggles for thematic layers, allowing selective display of elements such as provincial boundaries or cluster groupings derived from K-Means clustering. Hover-activated tooltips supply granular information including average download speeds (in Kbps), population estimates, and no-work counts, enriching interpretation with contextual metadata. Base map options, including OpenStreetMap, support geographic orientation, enabling stakeholders to explore spatial patterns such as low-latency concentrations that indicate IT-BPM investment potential. Overall, the interface transforms static geospatial data into an interactive exploratory medium for identifying digital hotspots and promising development areas.

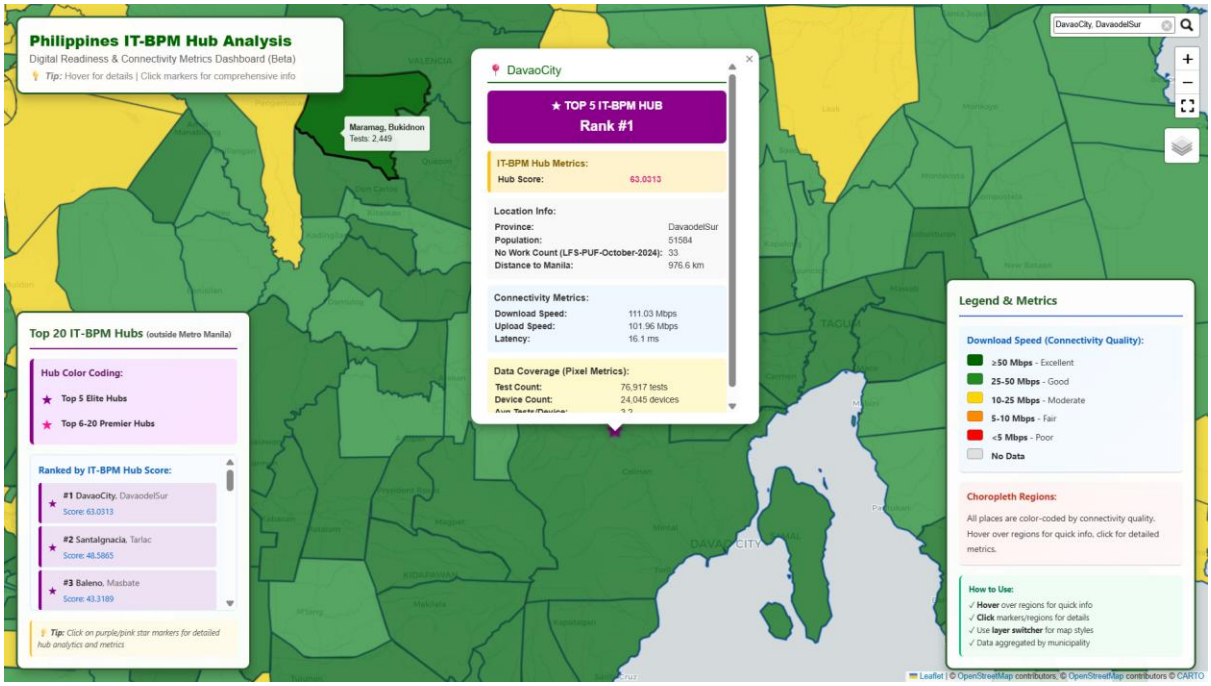
In support of these visual elements, the study deployed a geospatial machine learning pipeline incorporating data standardization, feature engineering, dimensionality reduction through aggregation, and unsupervised clustering via K-means. Features were scaled using StandardScaler to maintain comparable ranges, preventing high-variance attributes from dominating cluster assignments. The

clustering procedure implemented multiple initializations and iterative convergence criteria to enhance stability and reliability. Model validation leveraged silhouette scores, intra-cluster similarity measures, and comparison with known urban development benchmarks. An unsupervised learning approach was adopted due to the absence of labeled training data on IT-BPM hub suitability, and because clustering is effective for identifying latent spatial structures within high-dimensional geospatial data.



The Folium-generated interactive map visualizes Philippine municipalities using connectivity and employment indicators, rendered through Leaflet. Base layers include OpenStreetMap (default) and CartoDB Positron (light-gray minimalist design). Multiple GeoJSON overlays depict municipal polygons, typically colored by metrics such as download speed or digital readiness scores using graded color schemes (e.g., green for high, red for low), allowing users to explore spatial distributions and inter-municipality variability. Circle markers summarize aggregated tile data, styled with dark green (#006400) for high speeds (>50 Mbps) and forest green (#228B22) for moderate speeds (20–50 Mbps), with radii scaled between 7–9 based on test counts. Interactive popups provide structured tables containing attributes such as location, speeds, latency, population, and distance, displayed using blue (#f0f8ff) and yellow (#fffacd) highlights to enhance readability.

Additional interface features include a fixed title box in the top-left corner labeled “Philippines IT-BPM Hub Analysis” (white background, green text), a layer control enabling toggling between overlays (e.g., latency, talent pools), and a bottom-left legend ranking the top 20 IT-BPM hubs. The legend uses purple stars (#8B008B) to denote the top 5 locations and pink markers (#FF1493) for ranks 6–20, displaying municipal names, provinces, and normalized composite scores. Fullscreen capability enhances spatial exploration and interaction.



Key insights from the visualizations include detailed urban–rural gradients captured through polygon overlays, with high-connectivity zones prevalent in Cebu and Davao and sparse coverage in remote islands. Layered views also reveal correlations between metrics, such as relationships between low latency and high-speed clusters across regions in Luzon and the Visayas.

INTERPRETATION OF FINDINGS

The clustering analysis revealed two principal groupings of municipalities. The first cluster comprised cities with substantial talent pools but inconsistent levels of digital readiness, while the second cluster demonstrated a more balanced combination of accessibility and competitive conditions. Average cluster-level metrics indicated that talent indices reached values as high as 100 out of 100 in leading municipalities such as Davao City, while digital readiness scores averaged around 60 out of 100 in high-

performing localities. Composite hub scores reflected the integration of these features, with examples such as Naga City achieving a score of 40.7, highlighting investment potential despite moderate individual component metrics.

These patterns underscore key trade-offs in the geospatial distribution of IT-BPM suitability. Notably, an inverse relationship was observed between market saturation and accessibility, rationalizing interest in provincial sites where strong accessibility may offset competitive intensity. This dynamic supports strategic decentralization of IT-BPM expansion, mitigating congestion in heavily urbanized areas while leveraging emerging capacities in secondary cities.

Multiple visual analytics reinforced these interpretations. Scatter plots of test frequency versus device diversity demonstrated a strong positive correlation ($r = 0.939$), while bar charts illustrating test coverage per tile indicated substantial disparities, with peaks of 63,270 for low-test tiles and 1,401 for high-volume tiles. Histograms of performance metrics further revealed skewed distributions, with mean download speeds of 83,599 Kbps, upload speeds of 81,395 Kbps, latency averaging 22.8 milliseconds, and test counts per tile averaging 34.0. Collectively, these measures substantiated the geospatial inequalities inherent in national connectivity and helped explain the uneven suitability scores across municipalities.

Further analyses of download versus upload speeds, color-coded by latency, showed clusters situated near balanced throughput ratios but with significant dispersion in peripheral, high-latency environments. These patterns highlight the persistence of infrastructural gaps and underscore the need for policy-driven improvements in connectivity to unlock economic value in underserved regions.

Cluster-level interpretations also revealed distinct economic implications. The first cluster, representing the majority of municipalities, reflected systemic barriers to IT-BPM deployment, including lower connectivity and smaller labor pools. Interestingly, relatively high saturation scores suggest untapped labor capacity that could be activated through infrastructure development. The second cluster, which included municipalities such as Davao City, Santa Ignacia, and Naga City, exhibited higher composite hub scores and balanced feature profiles, indicating strategic feasibility for targeted investment.

These results reflect broader national dynamics in the Philippine IT-BPM sector, where development strategies increasingly emphasize regional diversification and the integration of digital infrastructure with local workforce development. The observed disparities in connectivity and labor availability mirror long-standing structural inequalities, yet they also illuminate emerging areas of growth outside Metro Manila.

The study is constrained by uneven sampling in network performance data and the static nature of labor force indicators, which do not fully capture temporal variation. However, sensitivity testing demonstrated robustness to moderate perturbations in feature values, suggesting that the observed patterns are stable and meaningful. Ethical considerations are central to interpretation, as analytical outputs should inform inclusive capacity-building rather than reinforce urban biases or marginalize low-performing regions. Limitations due to these aspects reflect known challenges in spatial big data and socioeconomic modeling (Kitchin, R., 2014, The Data Revolution).

CONCLUSION

The study demonstrates that numerous provincial municipalities exhibit strong potential to emerge as future IT-BPM hubs, driven by latent talent pools and improving connectivity infrastructures. Through the integration of geospatial data and machine learning, the analysis reveals spatial patterns that can help address the digital divide and stimulate broader economic development. Findings indicate that municipalities with high composite hub scores represent promising candidates for strategic interventions aimed at fostering decentralized industry growth.

Strategic recommendations include prioritizing the highest-ranked municipalities for pilot initiatives to assess hub feasibility, expanding data collection in sparsely sampled tiles to enhance model accuracy, and fostering collaborative efforts between government agencies and the private sector to strengthen connectivity in underserved areas. These actions can support sustainable and inclusive digital transformation, while ensuring that emerging growth centers are equipped with the resources necessary to compete in the national IT-BPM landscape.

More broadly, this research illustrates the value of geospatial machine learning in supporting evidence-based planning for IT-BPM decentralization in the Philippines. By combining connectivity metrics, demographic characteristics, and labor force indicators within a unified analytical framework, the model identifies areas with strong

potential for industry expansion while simultaneously highlighting structural disparities that may impede equitable development. The findings offer actionable insights for stakeholders seeking to diversify economic activity beyond metropolitan centers, inform infrastructure investment priorities, and harness distributed human capital across the archipelago.

REFERENCES

Akamatsu, N. (2022). Telecommunications infrastructure in archipelagic nations. *Telecommunications Policy, 46*(8), Article 102456. https://doi.org/10.1016/j.telpol.2022.102456

Batty, M. (2013). *The new science of cities*. MIT Press.

Deloitte. (2022). *Global sourcing survey 2022*. https://www2.deloitte.com/global/en/pages/technology-media-and-telecommunications/articles/global-sourcing-survey.html

Desiderio, L. (2025, September 13). IT-BPM industry: A vital economic pillar. *Philstar.com*. https://www.philstar.com/business/2025/09/14/2472564/it-bpm-industry-vital-economic-pillar

International Telecommunication Union. (2024). *Measuring digital development: Facts and figures 2024*. https://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2024.aspx

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

Labor Force survey. (n.d.). https://psada.psa.gov.ph/catalog/LFS/about

Newsbytes.PH. (2020, June 30). *New set of 'Digital Cities' from PH countryside bared*. https://newsbytes.ph/2020/06/30/new-set-of-digital-cities-from-ph-countryside-bared/

Philippine Statistics Authority. (2025, October 13). *Philippine standard geographic code*. https://psa.gov.ph/classification/psgc

Philippine Statistics Authority. (n.d.). *Labor force survey*. https://psada.psa.gov.ph/catalog/LFS/about

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Santos Knight Frank. (2024). *BPO primer: Outsourcing in the Philippines* [Report]. https://santosknightfrank.com/wp-content/uploads/2024/07/BPO-Primer-Outsouring-in-the-Philippines.pdf

Thinking Machines Data Science. (n.d.). *Using transfer learning and satellite imagery to map poverty in the Philippines*. Thinking Machines Data Science, Inc. https://stories.thinkingmachin.es/using-transfer-learning-and-satellite-imagery-to-map-poverty-in-the-philippines/

Tingzon, I., Orden, A., Go, K. T., Sy, S., Sekara, V., Weber, I., Fatehkia, M., García-Herranz, M., & Kim, D. (2019). MAPPING POVERTY IN THE PHILIPPINES USING MACHINE LEARNING, SATELLITE IMAGERY, AND CROWD-SOURCED GEOSPATIAL INFORMATION. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences/International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-4/W19*, 425–431. https://doi.org/10.5194/isprs-archives-xlii-4-w19-425-2019

Zandbergen, P. A. (2009). Accuracy of address geocoding: A case study in Tampa, Florida. *Journal of Spatial Science,* *54*(2), 1–22. https://doi.org/10.1080/14498596.2009.9635162