Name: ENCALLADO, Carl Francis T.	**Date:** December 12, 2025
Section: GF	**Machine Learning Cognate 1 Project**

Topic: Identifying Next-Wave IT-BPM Hubs in the Philippines Using GPS Connectivity and Employment Data

EXPLORATORY DATA ANALYSIS (EDA) OUTPUT

Analyzing the data set open-air consultants revealed a huge diversity in the digital infrastructure, population distribution, and labor characteristics in the different municipalities. The connectivity measures showed a strong positive skewness and a wide spread, which are typical indicators of the ICT infrastructure in developing countries with islands in the region (Akamatsu, N. 2022, Telecommunications Policy) Their download and upload speeds showed a wide spectrum, but the high throughput was consistently found in the major metropolitan areas and the emerging provincial urban centers, while the remote and island municipalities showed a very low performance and a high latency. The variation in the population levels was between the sparse rural settlements and the dense urban cores which had more than one million inhabitants, and the talent indicators were also similar in the densely populated regions but there were some secondary cities that showed an unexpectedly strong potential. The correlation analyses pointed out the very strong relationship between population density, network performance, and distance from Manila, thus showing the spatial dependence in the infrastructural development. The correlation between population density and connectivity is in line with the findings in ICT development literature (International Telecommunication Union, 2024 ICT Development Report). The IT-BPM hub scores had a positively skewed distribution with only a small number of municipalities reaching high values. The bar and map visualizations made it easier to see the potential sites' geographical distribution over the islands of Luzon, Visayas, and Mindanao, with a lot of the sites having different region-specific trade-offs in connectivity, accessibility, and talent availability.
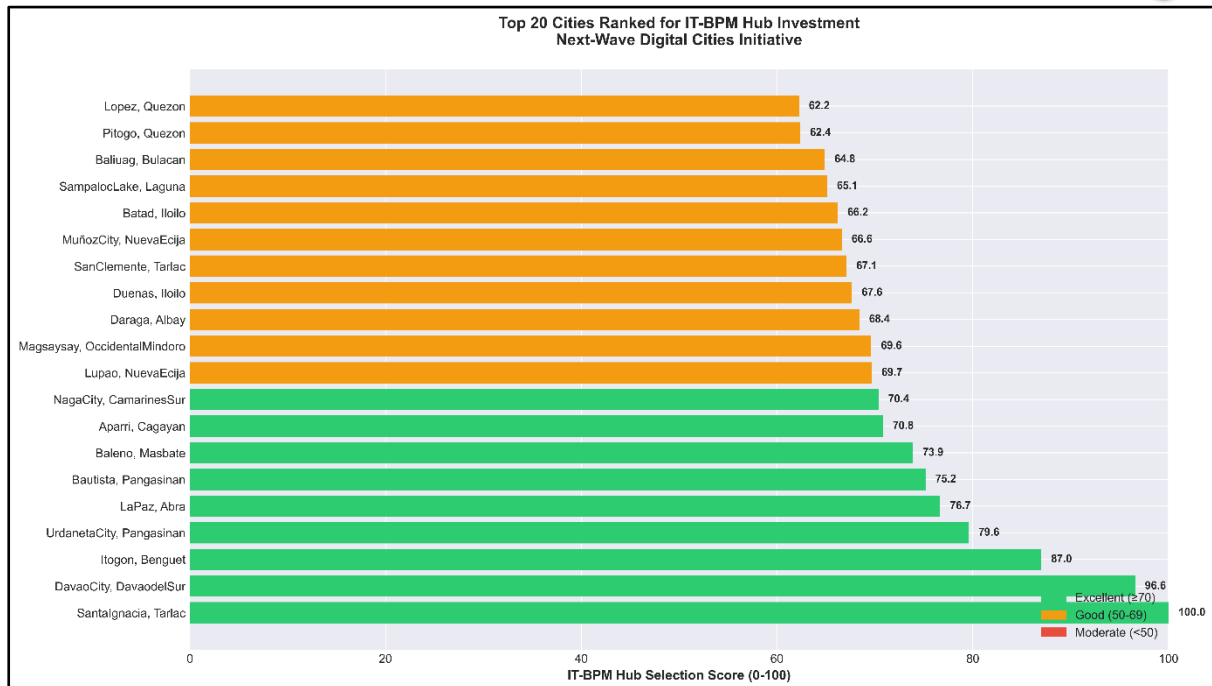
Figure 1. Top 20 Cities Ranked for IT-BPM Hub Investment

The figure ranks the top 20 places in the Philippines for IT-BPM investments as per the "Next-Wave Digital Cities Initiative" with Santa Ignacia, Tarlac getting the first rank with a score of 100.0 and Davao City with 96.6. The cities are categorized as "Excellent" (green) for the top ten and "Good" (orange) for the next ten, which indicates a strategic roadmap for the government to implement economic growth in rural areas rather than the already crowded metropolitan cities such as Manila. The analysis points out the different provincial centers, which are Itogon, Benguet to Daraga, Albay, and the theme of the local infrastructures' preparedness is reinforced by the data. Moreover, it also implicitly invites investors to consider these regions as their next destinations for the expansion, hence ultimately helping to attract foreign investments and creating high-value jobs that facilitate the closing of the digital divide across the territories.
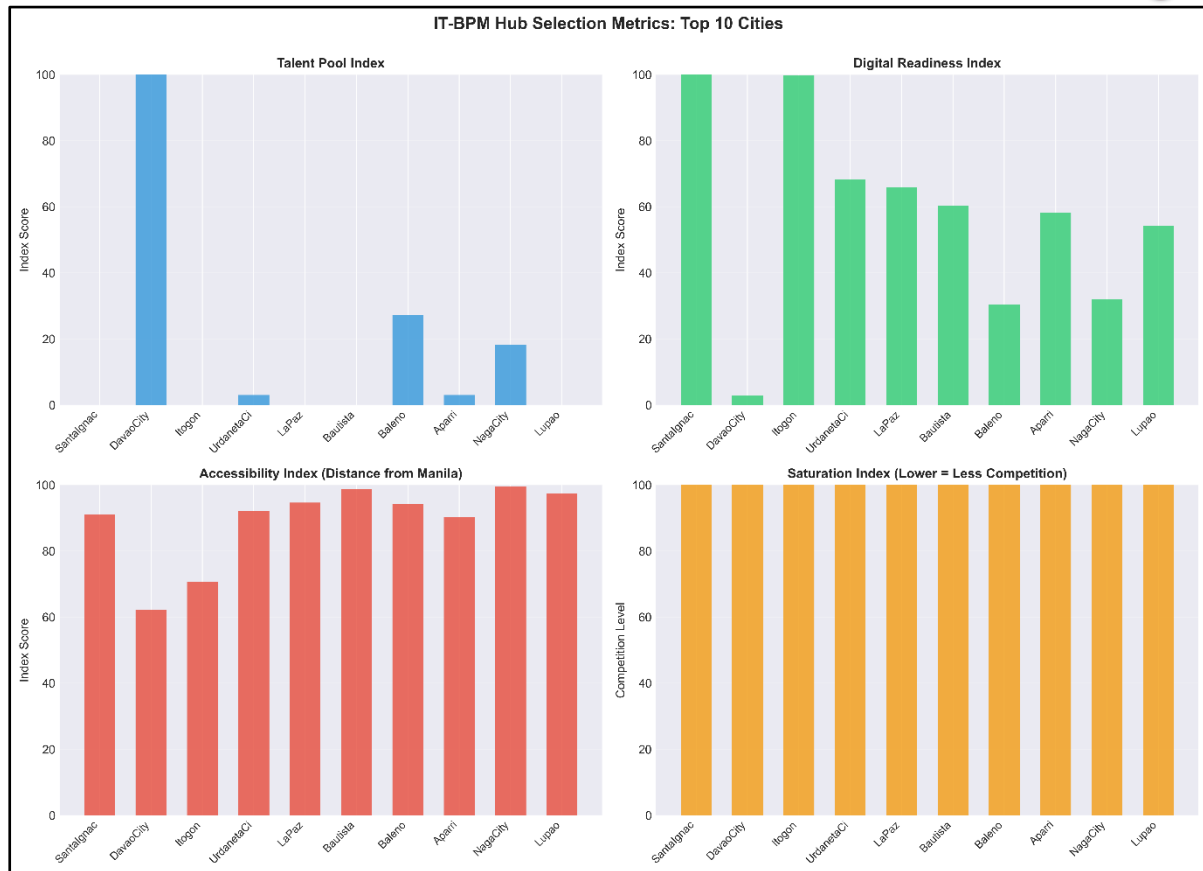
Figure 2. *IT-BPM Hub Selection Metrics: Top 10 Cities*

Analyzing the top 10 cities for IT-BPM businesses in terms of their human capital, infrastructure, and market openness reveals a situation in which these factors were leveled out and one common opportunity was opened up through the market being opened up. Davao City's position as the undisputed number one in the Talent Pool Index with a perfect score is incomparable with other cities' scores, but surprisingly its Digital Readiness score is very low compared with those smaller municipalities, i.e., Santa Ignacia and Itogon, achieving perfect scores. Although the Accessibility Index gives higher points to cities that are nearer to Manila leaving Davao with a lower score because of its location, the most important conclusion to be drawn is the Saturation Index where all cities scored 100 which is perfect. So it means that whether an investor goes for the huge labor force of Davao or the excellent digital infrastructure of Santa Ignacia, all ten places are offering practically unexploited, low-competition markets that are actually waiting for "first-mover" advantage.
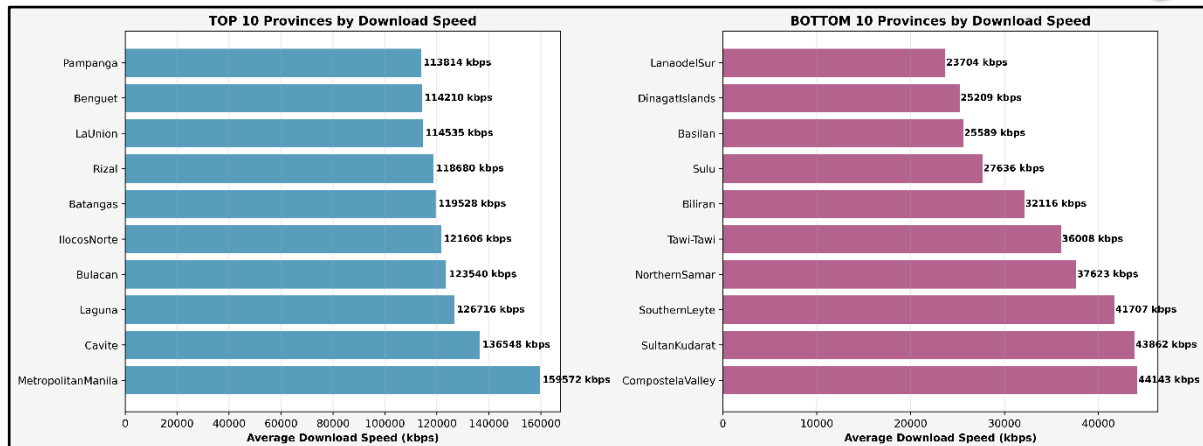
Figure 3. Top and Bottom 10 Provinces by Download Speed

The data exhibits a major "Digital Divide" where the high internet speeds of Metro Manila and CALABARZON (118-160 Mbps) are compared with, at the other end of the scale, the slowest provinces with about 23 Mbps, thus directly affecting the selection of the IT-BPM hub. Nevertheless, Davao City takes the lead in "Talent Pool" numbers while the smaller towns of Itogon and Santa Ignacia have scored "Digital Readiness" perfectly by, for example, taking advantage of the strong connectivity in the Benguet region. The importance of the location cannot be overstated: all the best cities feature a perfect "Saturation Index" which means there is still market potential waiting to be tapped; thus, the investors will have to make a rather difficult decision between the large and scalable workforce of the main urban areas or the fast, ready-to-use, "plug-and-play" infrastructure of these new provincial hubs.
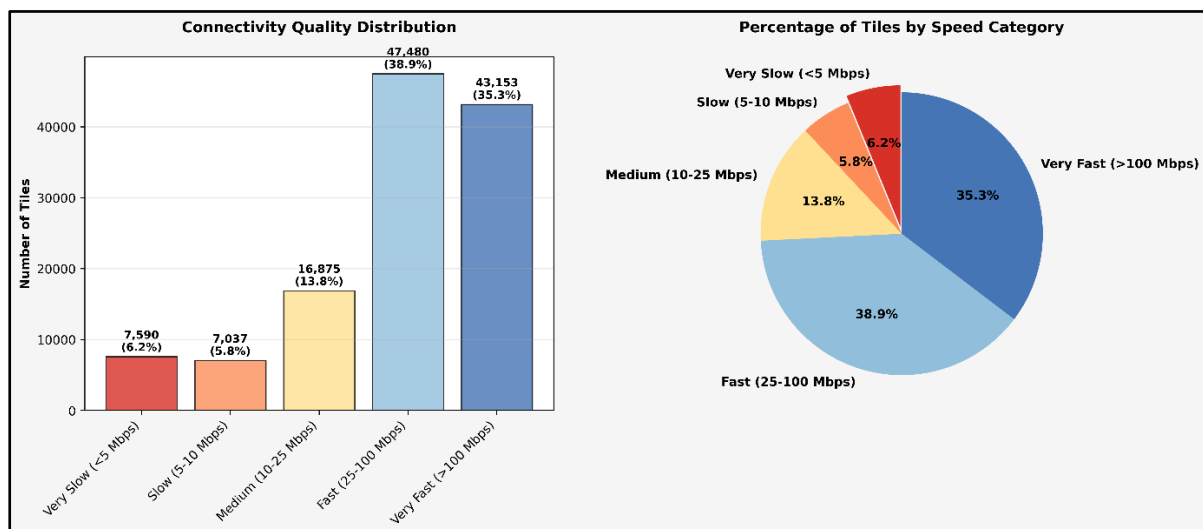

Figure 4. *Connectivity Quality Distribution and Percentage of Tiles by Speed Category*

The fourth visualization presents a bar and pie chart that give a summary of the different categories of connectivity. "Very Fast" speed (>100 Mbps) is the category under which 43,153 tiles (35.3%) are classified, on the other hand, the category of "Very Slow" (<5 Mbps) has 7,590 tiles (6.2%). The intermediate categories are being distributed in between them. More than 70% of tiles have speed greater than 25 Mbps, which demonstrates considerable improvement of the whole country in this aspect, however, the group of low performing tiles is still significant. In terms of geography, the fast categories are mainly identified through urban areas with high population density, while the slow categories are linked to rural or isolated areas. The pie chart indicates the percentage of each category, while the bar chart shows the differences in volume, thus together they depict high performance across the board but spatial imbalances still exist. The conclusions drawn from these observations suggest that there is a need for putting up new infrastructures in the neglected areas to improve the access of the residents there to the internet.
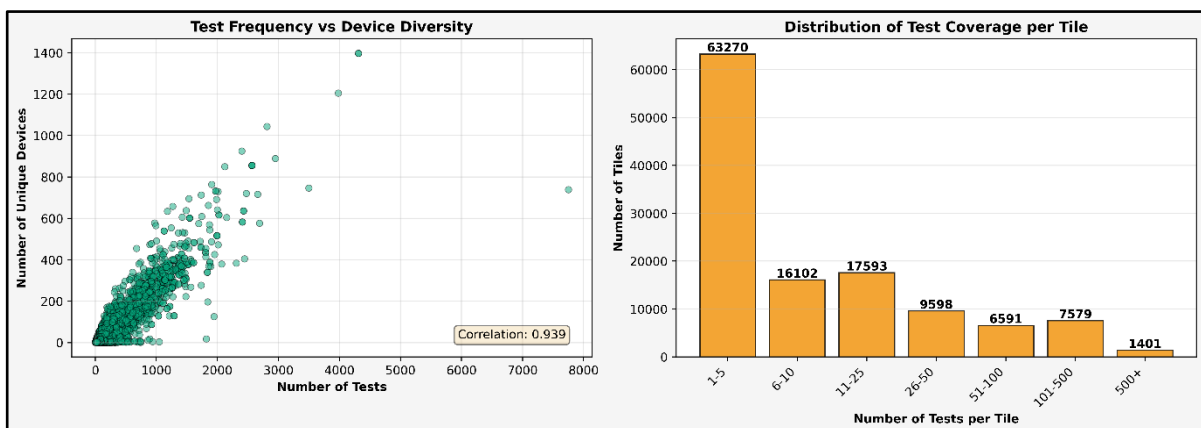

Figure 5. *Total Frequency vs Device Diversity and Distribution of Test Coverage per Tile*

Charts "Test Frequency vs. Device Diversity" and "Distribution of Test Coverage per Tile" play an important role in checking the quality of the internet speed data, which is the basis for the IT-BPM hub analysis, through a critical quality assurance process. The very strong positive correlation (0.939) assures that high-traffic places produce statistically strong data from a diverse collection of devices, thereby supporting the high-speed scores in the like of Metro Manila. Nonetheless, the massive spatial bias is revealed through the distribution histogram wherein a whopping number of geographic "tiles" (more than 63,000) are relying on very small sample sizes (1-5 tests)

indicating that urban connectivity data is very reliable whereas that for rural or "Bottom 10" provinces is based on fragile and sparse data hence physical verification on the ground is required rather than relying only on these crowd-sourced figures.
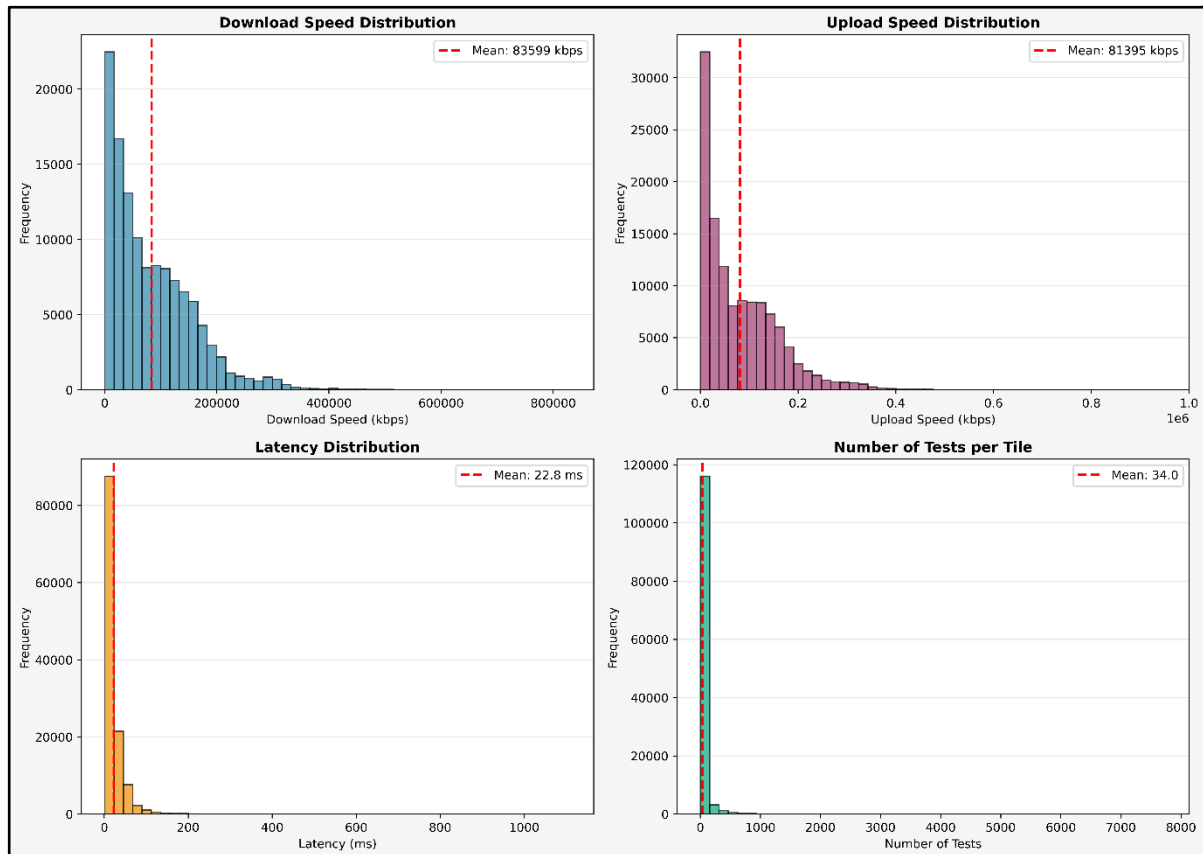

Figure 6. *Download and Upload Speed Distribution with Latency Distribution and Number of Tests per Tile*

The visualizations together show a clear "Digital Divide" in the Philippines that is a major factor in deciding the locations for IT-BPM investment. The Top 10 Provinces chart confirms the industrial superiority of Metro Manila and CALABARZON (with speeds of 118–160 Mbps) against the "Bottom 10" regions that have an average of around 23 Mbps. However, the Test Coverage histogram brings to light a significant data bias, as rural metrics are based on very small sample sizes (<5 tests) compared to the large data sets of the urban centers. This inequality in infrastructure is a direct factor in the Hub Selection Metrics, where every leading city has a perfect "Saturation Index" (which refers to the presence of markets that have not been tapped yet), but at the same time, it creates a difficult choice for the investors: they have to decide between the huge, scalable talent pool of Davao City or the excellent, "plug-and-play"

digital readiness of smaller municipalities like Santa Ignacia, which take advantage of the high-speed connectivity in their regions.
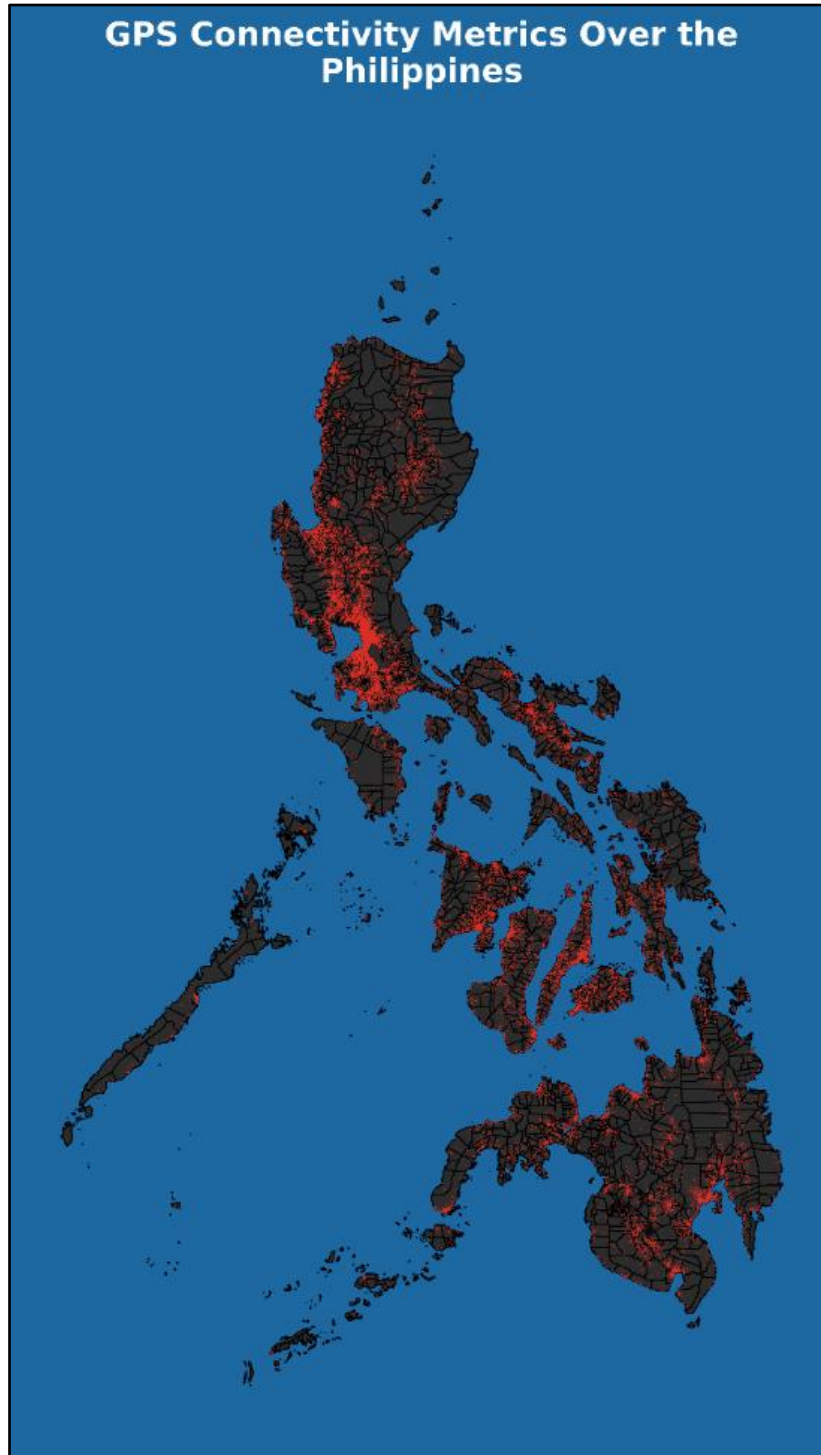

Figure 7. GPS Connectivity Metrics Over the Philippines

The seventh picture is a choropleth map with the label "GPS Connectivity Metrics Over the Philippines", showing with red shades the varying degrees of

performance. Central Luzon, Cebu, and Davao are the areas with strong red mixing, which then turn to dark red and even brown in the rural interior and less accessible islands. The highlighted areas characterize the differences in connectivity; urban centers enjoy significantly better access while remote areas are deprived of basic infrastructure. Statistical data exhibit a more or less direct relationship between red intensity and population/economic concentration, which is manifested as a clearly visible north-south gradient and the highlighting of urban agglomeration effects. Thus the map has confirmed that the spatial inequality and geographic isolation are the main causes of the digital divide.
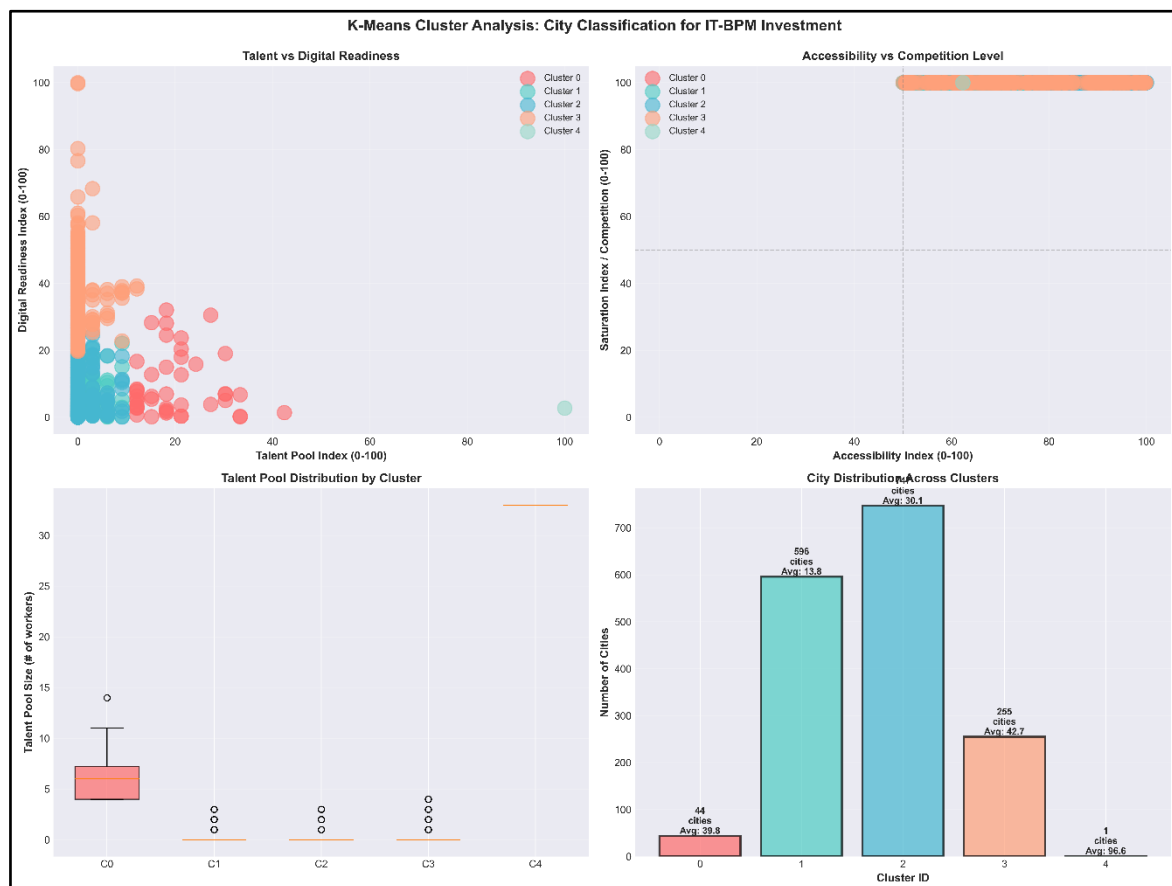

Figure 8. *K-Means Cluster Analysis: City Classification for IT-IBM Investment*

The data comes up with a very important "Digital Divide" where Davao City is positioned as a very special "Cluster 4" outlier, providing a very large talent pool (score: 100) even though the city does not have a good digital readiness, while smaller towns such as Santa Ignacia and Itogon count on their boosting digital readiness (score: 100) by taking advantage of strong regional speeds (e.g., Benguet's ~114 Mbps) that are far more than the "Bottom 10" provinces (~23 Mbps). This classification gives a clear

strategic trade-off to investors: the overall perfect Saturation Index, which is universal, confirms that all top cities are untapped "Blue Ocean" markets, however, the decision is between the scalable workforce of Davao or the excellent "plug-and-play" connectivity of provincial hubs, with the K-means analysis and test coverage histograms as indispensable checks to verify these locations against the low data reliability of rural areas.
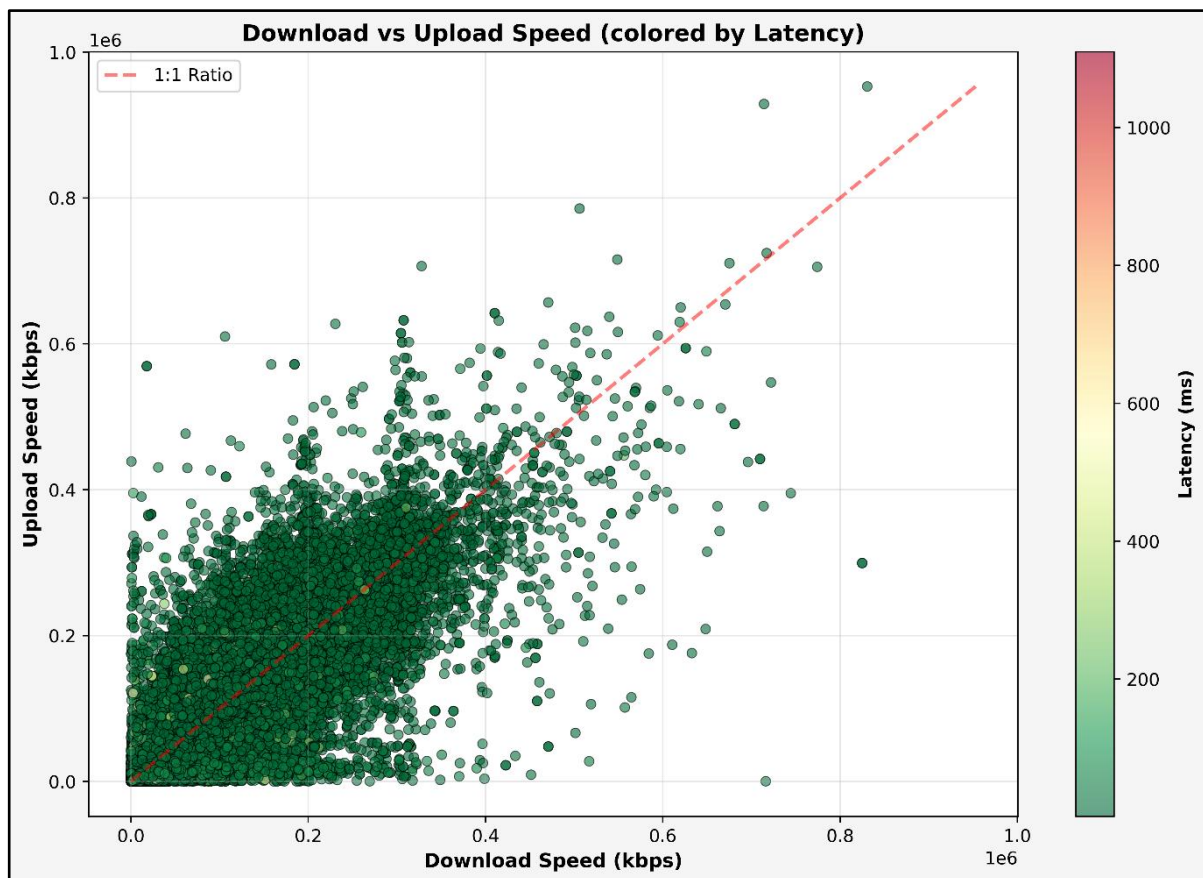

Figure 9. *Download vs Upload Speed (colored by Latency)*

The figure above clearly shows a significant "Digital Divide" with Davao City (considered as a single "Cluster 4" outlier) at the top in Talent Pool size, while very small hubs like Santa Ignacia, which are supported by fast internet in some provinces like Benguet where the speed is around 114-160 Mbps, are able to achieve perfect Digital Readiness. This difference gives rise to a straightforward strategic trade-off: the investors will have to decide whether to go for the large and easily scalable labor force of Davao or the much better "plug-and-play" infrastructure of the provincial municipalities. Meanwhile, the overall perfect Saturation Index indicates that the top locations are still undiscovered "Blue Ocean" markets; however, the Test Coverage

histogram gives a critical quality alert, disclosing that rural metrics are often based on weak data (1-5 tests per tile), therefore ground validation for the digital reliability of these new hubs is a must before the investment is done.

By looking at all these things, the data shows the existence of strong spatial inequalities, the connection between urbanization and connectivity, and the possibility of the second municipalities where talent and infrastructure are slowly merging. The proximity of Davao City to digital readiness, especially Davao City, and high-talent concentration points out the need for targeted interventions to facilitate even regional development.