**SPATIAL DATA ANALYSIS (SDA)**

**SHOULD THE GOVERNMENT PROMOTE AND REGULATE THE ROLE OF SPATIAL DATA INFRASTRUCTURES IN THE DIGITAL TRANSFORMATION OF EDUCATION IN PAKISTAN FOR POLICY AND DECISION-MAKING?**

**AGBOOLA ISAAC OLUWATOMIWA**

**2ND SEPTEMBER 2024**

# **ABSTRACT**

The spatial information infrastructures are crucial to the digital transformation of the education sector. As educational institutions are increasing especially in Africa, Europe, and America the spatial data infrastructure can enhance educational planning, resource allocation, and performance assessment.

The study highlights the role, challenges, historical context, and application of spatial data infrastructure that could ensure the effective use of spatial data in resolving education access disparities and efficiently distributing educational facilities. The government is urged to foster collaboration between schools, government agencies, and technology providers to ensure the harnessing of spatial data for improving education outcomes.

The findings suggest a strategic partnership between the government and educational institutions to unlock the full potential of spatial data in Pakistan's education landscape.

**TABLE OF CONTENTS**

[ABSTRACT 2](#_Toc175796417)

[FIGURE 5](#_Toc175796418)

[TABLE 6](#_Toc175796419)

[CHAPTER ONE 7](#_Toc175796420)

[INTRODUCTION 7](#_Toc175796421)

[1.1 BACKGROUND 7](#_Toc175796422)

[1.2 HISTORICAL CONTEXT 7](#_Toc175796423)

[1.3 RESEARCH PROBLEM 8](#_Toc175796424)

[1.4 OBJECTIVES 9](#_Toc175796425)

[CHAPTER 2 10](#_Toc175796426)

[METHODOLOGY 10](#_Toc175796427)

[2.1 DATASET 10](#_Toc175796428)

[2.2 DATA COLLECTION 11](#_Toc175796429)

[2.3 SHAPE FILE 11](#_Toc175796430)

[2.4 DATA MAPPING 12](#_Toc175796431)

[2.5 ALGORITHM 13](#_Toc175796432)

[2.6 QUANTUM GEOGRAPHIC INFORMATION SYSTEM CODE RELIABILITY 13](#_Toc175796433)

[CHAPTER 3 18](#_Toc175796434)

[RESULT AND DISCUSSION 18](#_Toc175796435)

[3.6 PERFORMANCE 28](#_Toc175796436)

[CHAPTER FOUR 30](#_Toc175796437)

[CONCLUSION 30](#_Toc175796438)

[4.1 FUTURE WORK 31](#_Toc175796439)

[REFERENCES 32](#_Toc175796440)

# **FIGURE**

[Figure 1: The Map of Pakistan indicating the Seven Provinces 10](#_Toc175584607)

[Figure 2: The Architecture of QGIS 11](#_Toc175584608)

[Figure 3: The Shape File of Pakistan Map 12](#_Toc175584609)

[Figure 4: The Boundary Wall of the School 18](#_Toc175584610)

[Figure 5: The School Drinking Water 19](#_Toc175584611)

[Figure 6: The Education score of the school 20](#_Toc175584612)

[Figure 7: The Electricity of the School 21](#_Toc175584613)

[Figure 8: The Enrolment score of the school 22](#_Toc175584614)

[Figure 9: The Gender Parity of the School 23](#_Toc175584615)

[Figure 10: The Learning Score in School 24](#_Toc175584616)

[Figure 11: The Retention Score in School 25](#_Toc175584617)

[Figure 12: The Infrastructure Score in School 26](#_Toc175584618)

[Figure 13: The Toilet in School 27](#_Toc175584619)

# **TABLE**

[Table 1: The Data of the School Performance of the Seven Provinces 13](#_Toc175794880)

[Table 2: The Statistics Analysis of School Performance 28](#_Toc175794881)

[Table 3: The High School Performance 30](#_Toc175794882)

[Table 4: The Low School Performance 31](#_Toc175794883)

# **CHAPTER ONE**

# **INTRODUCTION**

# **1.1 BACKGROUND**

The digital transformation in the Pakistani education system is essential to improving the educational framework by enhancing accessibility, equality, and quality of education. Additionally, the Ministry of Education has a partnership with Google that aims to provide smart classrooms and digital tools to millions of students who are out of school by organising a digital education program aimed at educating over 130,000 students and training 43,000 teachers (Suleiman & Danmuchikwali, 2020).

According to (Zuberi, 2023), Spatial Data Infrastructure which encompasses policies, institutional mechanisms, technologies, data, and human resources, is designed to provide information about various locations to users or providers with relevant applications (Duhdra, 2023). In education, spatial data infrastructure could be instrumental in mapping and visually representing the locations of schools, resource allocation, school performance, and facilities nationwide (Hu & Li, 2017).

# **1.2 HISTORICAL CONTEXT**

In Pakistan, the national SDI process was launched in the early 2000s and is being led by the Survey of Pakistan with support from other government agencies. The Pakistan Spatial Data Infrastructure initiative will attempt to institute a harmonised system of management and distribution of geospatial data from various sectors, including infrastructure, natural resources, and disaster management. Also, the digital transformation started in the education sector by beginning initiatives like establishing Pakistan’s Virtual learning environment and integrating ICT curriculum in schools, the integration of these SDIs into sector-based applications has yet to be achieved (Rajabifard & Williamson, 2003).

Traditionally, the education sector of Pakistan has remained heavily reliant on administrative data and conventional planning methods; there has hardly been any use of spatial data and analysis for decision-making. Consequently, regional imbalances in educational provision and attainment levels have remained elusive to identify and rectify, and inefficiencies in resource and infrastructure allocation persist. For example, without spatial analysis, it has not been possible to target underserved areas to improve school enrollment and retention rates.

The Government of Pakistan has acknowledged digital transformation in education through the National Education Policy 2021. It highlights the importance of data-driven decision-making and the synthesis of information and communication technologies for improved educational planning, monitoring, and evaluation (Ali & Imran, 2021; Rajabifard & Williamson, 2003).

This historical backdrop underpins the need for a more strategic and coordinated approach toward developing SDI into the broader agenda of the digital transformation of the education sector in Pakistan, building on existing national SDI initiatives and bringing due attention to the particular challenges and opportunities facing this sector.

# **1.3 RESEARCH PROBLEM**

A comprehensive spatial data infrastructure can be created and deployed by enhancing educational planning, service delivery, and equity in Pakistan’s education system. Some specific studies within this context could be:

1. What are some of the obstacles or challenges that exist in adopting spatial data technologies in schools in Pakistan?
2. How may the government promote educational research and innovation through spatial data and analysis?
3. How can Spatial Data Infrastructure facilitate the development of digital skills and literacy among students and teachers in Pakistan?

# **1.4 OBJECTIVES**

* To create a plan for adopting and implementing spatial data infrastructure in Pakistan’s education sector.
* To assess the impact of spatial data infrastructure on education resource allocation and School Performance
* To explore the potential of spatial data infrastructure in improving access to quality education in Pakistan.

# **CHAPTER 2**

# **METHODOLOGY**

This research was done using the Quantum Geographic Information System platform, to develop a series of spatial analyses and visualisations to uncover patterns, trends, and disparities in school performance across the seven provinces in Pakistan. This includes creating graduated maps, centroid points to indicate the province, and Statistics Analysis (highest and lowest) performance.



Figure 1: The Map of Pakistan indicating the Seven Provinces

# **2.1 DATASET**

Spatial and non-spatial data are the major branches of the dataset that can be integrated and examined using Quantum Geographic Information System (QGIS). The spatial data provide the Shapefile obtained from Global Administrative Areas (GADM), and the non-spatial data offer the attribute information about the school performance obtained from the Pakistan open data.

QUANTUM GEOGRAPHIC INFORMATION SYSTEM

(QGIS)

COMMA SEPARATED VALUES (CSV)

SHAPE FILE (SHP)

Figure 2: The Architecture of QGIS

# **2.2 DATA COLLECTION**

The data for this analysis consolidated-educational-dataset also known as secondary data collected from Pakistan open data, having 51 columns and 581 rows. Data cleaning was performed to avoid unusual behaviour during analysis by identifying errors, changing the outlying values and completing the missing data. Once the data had been processed it was reduced to 15 columns and 581 rows and identified anomalies were removed from the set making it into shapefile format.

# **2.3 SHAPE FILE**

The administrative boundary shape files to the Geographic Information System database allowed spatial analysis at an appropriate level of detail to identify local patterns and trends in school performance. The Pakistan country map was picked from the GADM shape file with appropriate fields that match the data accordingly.

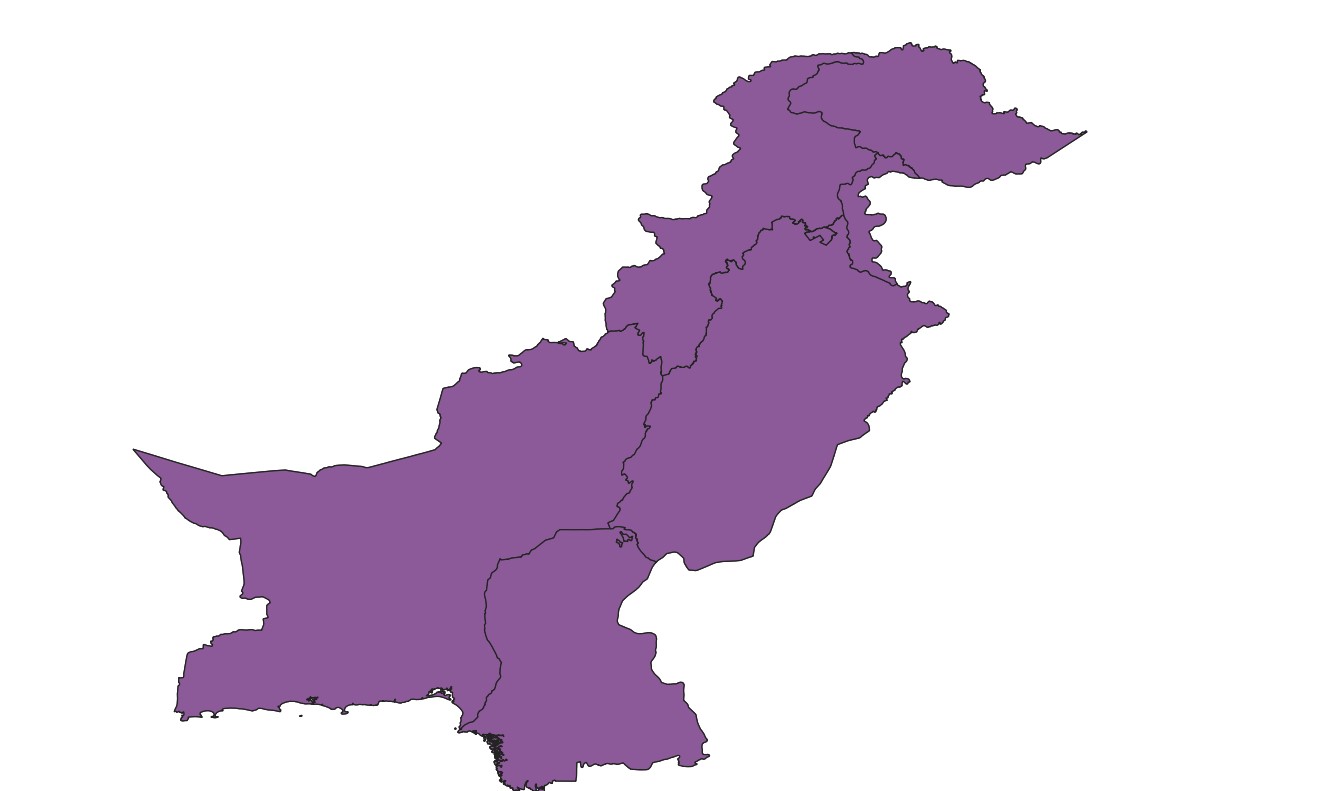


Figure 3: The Shape File of Pakistan Map

# **2.4 DATA MAPPING**

The collected data was represented in a comma-separated value (CSV) file, which was then transformed into a shapefile (.shp). The main attributes used for the examination were the properties that identify specific provinces within the dataset, which was compared to its equivalent in the spatial information of the shapefile.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Indicators | Sindh | Khyber Pakhtunkhwa | Balochistan | Azad Kashmir | Punjab | Gilgit Baltistan | Islamabad |
| Boundary wall | 52.01923 | 70.38414 | 37.899543 | 3.943662 | 98.08061 | 61.130742 | 96.858639 |
| Drinking water | 23.22115 | 38.78563 | 17.351598 | 4.225352 | 99.32822 | 39.575972 | 92.670157 |
| Education score | 52.15309 | 57.04867 | 43.627316 | 78.28199 | 69.00949 | 73.265443 | 85.740863 |
| Electricity | 20 | 25.27881 | 3.196347 | 0.84507 | 93.28215 | 35.335689 | 98.429319 |
| Enrolment score | 51.71426 | 55.64614 | 56.548519 | 68.18008 | 69.08697 | 57.993168 | 89.515444 |
| Gender parity score | 87.87311 | 74.14338 | 50.241065 | 98.07263 | 93.00099 | 91.0415 | 94.818199 |
| Learning score | 33.025 | 39.15283 | 46.719681 | 54.32341 | 54.95 | 49.808995 | 71.129807 |
| Retention score | 36 | 59.25234 | 21 | 92.55184 | 59 | 94.218109 | 87.5 |
| Infrastructure score | 35.73077 | 59.8513 | 19.31065 | 6.760563 | 94.84503 | 36.325088 | 86.596859 |
| Toilet | 54.61539 | 80.54523 | 15.525114 | 7.887324 | 97.88868 | 25.441696 | 95.287958 |

Table 1: The Data of the School Performance of the Seven Provinces

# **2.5 ALGORITHM**

STEP 1: Preparing, and preprocessing Geocoding and Spatial Data.

STEP 2: Join the CSV school performance with the Shapefile using spatial join operation.

STEP 3: Creating Interactive visualisation to communicate findings

STEP 4: Use Thematic Mapping and Data Visualization to visualise school performance on a map.

STEP 5: Using Statistical analysis for the performance.

# **2.6 QUANTUM GEOGRAPHIC INFORMATION SYSTEM CODE RELIABILITY**

STEP 1: Background

* Go to the project menu and select properties
* From the General on the left side click on the background colour
* Select any background colour of your choice
* Click apply, then the background will change in the project

STEP 2: Shapefile

* Open QGIS and click on the layer in the menu bar > Add layer > Add Vector Layer
* In the Add Vector layer where there is Vector dataset(s) select the shapefile (.shp) from the file explorer and click open
* Click add and close
* The shapefile will be added to the QGIS project as a separate layer, visible in the layer panel

STEP 3: Dissolve

* Select the shapefile layer in the layer panel
* Go to Vector > Geoprocessing Tools > Dissolve
* Choose the shapefile as the Input layer
* In the Dissolve field select the ADM1\_EN from the file explorer
* Click ok and Run to dissolve the shapefile
* Then right-click the dissolve on the layer panel > Export < Save features as
* In the save layer ensure that the format is on ESRI Shapefile, in the file name click the file explorer and save where you want your work to be.
* Save the dissolve to Pakistan\_sp click save and select ok.
* Then select the dissolve from the layer panel and remove the layer.

STEP 4: Centroids

* Select the Pakistan\_sp layer in the Layers panel
* Go to Vector > Geometry Tools > Centroids
* In the Centroids dialogue, select Pakistan\_sp as the input layer
* Click Run to create the centroid layer
* The new layer will contain point features representing the centroids of the input features
* You can change the colour by right-clicking the centroid in the layer panel and selecting properties > Symbology.
* From the drop-down list select Single Symbol, Project style select dot Red, apply and click ok
* Right Click on the Centroids in the layer panel and save it to Locations in the same format in Step 3

STEP 5: Command Separate Values

* Click the layer menu and Add Layer > Add Delimited Text Layer
* Create Layer from Delimited Text file dialogue, click the Browse button from the file name and navigate to the School Peformance.csv
* Select the School Peformance.csv file and click Open
* In the dialogue, configure the settings for your School Peformance.csv file, such as the regular expression delimiter from the file format should be a comma, in geometry Definition click on no geometry
* Click add, then the School Peformance.csv file is added to the layer in the QGIS project

STEP 6: Joins

* Identify the common field between the School Peformance.csv and Pakistan\_sp shapefile that you want to use to join which are ADM1\_EN.
* In the Layer Panel, right-click on the Pakistan\_sp shapefile layer and select properties > Joins
* In the Join where you will see + click on it
* In the Join layer select School Performance.csv, join field and target field select ADM1\_EN
* In Joined Fields select from the option on which one has to appear on the attribute table
* Click ok, apply and ok
* From the layer panel right click on Pakistan\_sp and click Open Attribute table to check if there are no null values and if correctly joined with School Peformance.csv

STEP 7: Thematic Mapping

* In the layer panel, right-click on the Pakistan\_sp layer and select the properties
* In the layer properties dialogue, navigate to the Symbology tab.
* In the Symbology dropdown, select Graduated
* In the Value dropdown, select the field from the School Peformance.csv data you want to use for the graduated mapping
* Choose an appropriate colour ramp to represent the values on the map
* Select Equal Count (Quantile) mode for the classification
* Adjust the classes to set the number of intervals to 7 to divide the data into
* Adjust the symbol size, colour, and other visual properties
* Click Apply
* Add labels, Select the single labels from the drop-down list.
* Adjust the legend, and configure other map element
* Use the Apply button in the layer properties dialogue to preview your graduated map
* Save it in the QGIS project

STEP 8: Print Layout

* Go to the project menu and select New Print layout
* In the Print Layout Window, click the Add Map button
* Click and drag on the layout canvas to create a map frame
* The map frame will display the map that was created in the project
* Double-click the map frame to open the item Properties panel
* In the main properties section adjust the scale, rotation, and other settings of the map
* Click the Add Legend button to add a legend to the layout
* Click and drag on the layout canvas to create a legend frame
* Double-click the legend frame to open the item Properties panel, where you can modify the legend
* Click the Add label button to add a title to the layout
* Click and drag on the layout canvas to create a label frame, then double-click it to edit the text.
* Add the north arrow, and scale bar using the corresponding toolbar
* Export it as an image document by going to the layout menu.

# **CHAPTER 3**

# **RESULT AND DISCUSSION**

The figure below contains different indicators on education quality and physical infrastructure within various provinces in Pakistan. Consequently, the study intends to examine spatial disparities and trends in school infrastructure as well as school performance.

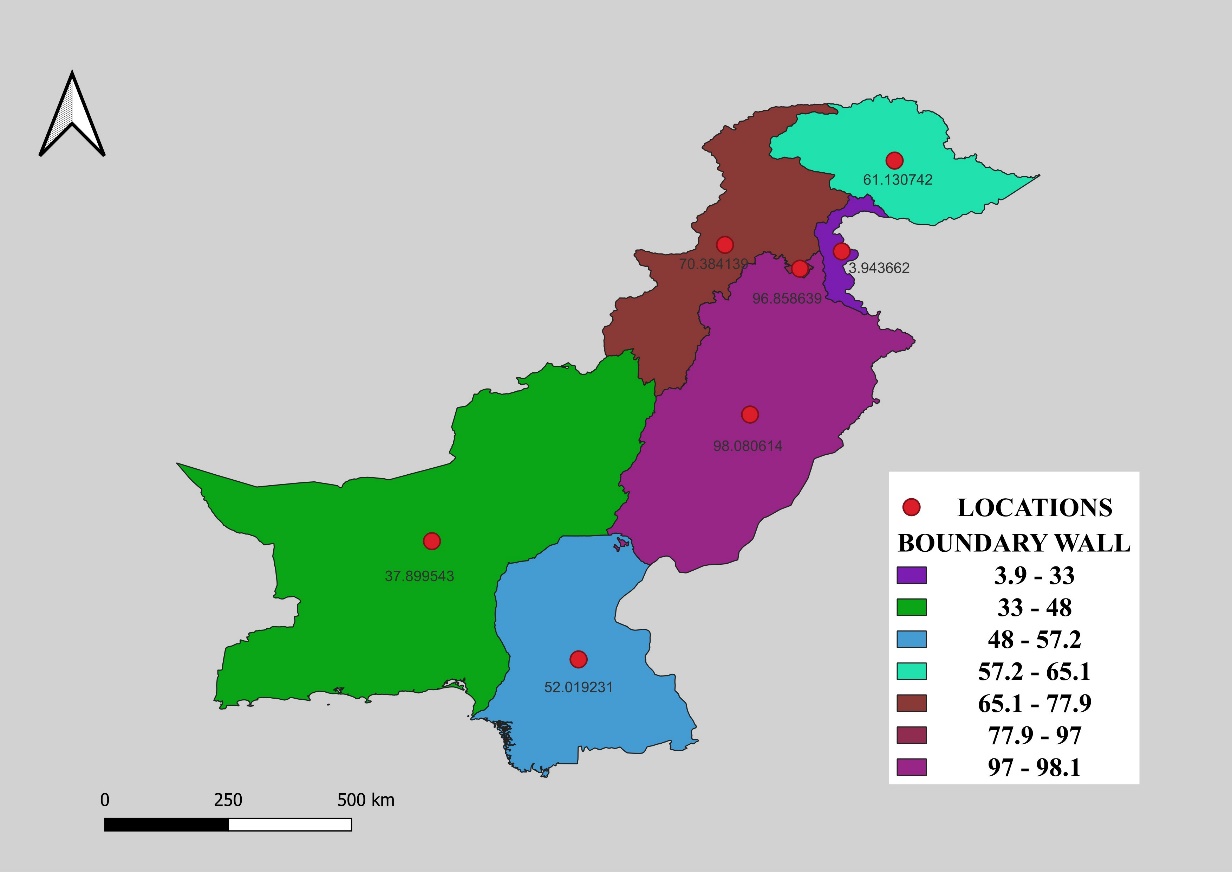


Figure 4: The Boundary Wall of the School

Figure 4 shows that Punjab (98.08%) has the highest percentage of schools with boundary walls followed by Islamabad (96.86%). That means provinces have better infrastructure for securing school premises, and Azad Kashmir (3.94%) and Balochistan (37.90%) have the lowest percentage suggesting a considerable infrastructure gap in this province they do not have a better infrastructure for securing school premises.

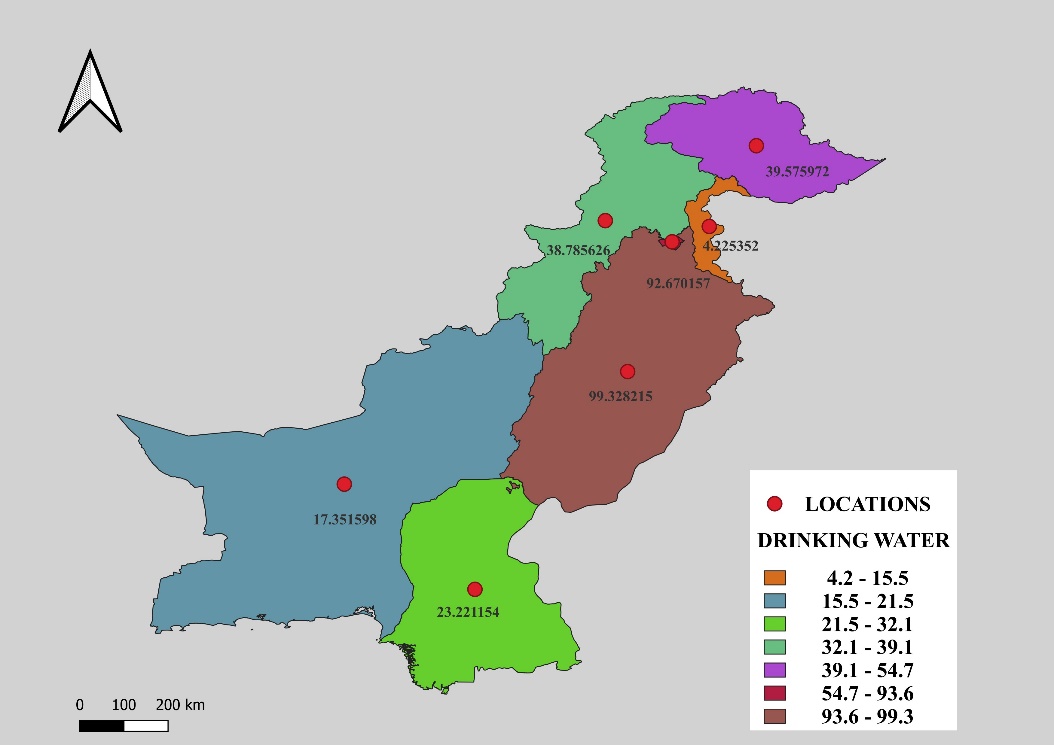


Figure 5: The School Drinking Water

Figure 5 shows that Punjab has the highest percentage of schools with access to drinking water at 99.32% followed by Islamabad (92.67%). Azad Kashmir (4.23%) and Balochistan (7.36%) have the lowest percentages indicating a significant gap in the distribution of this basic amenity in the region.

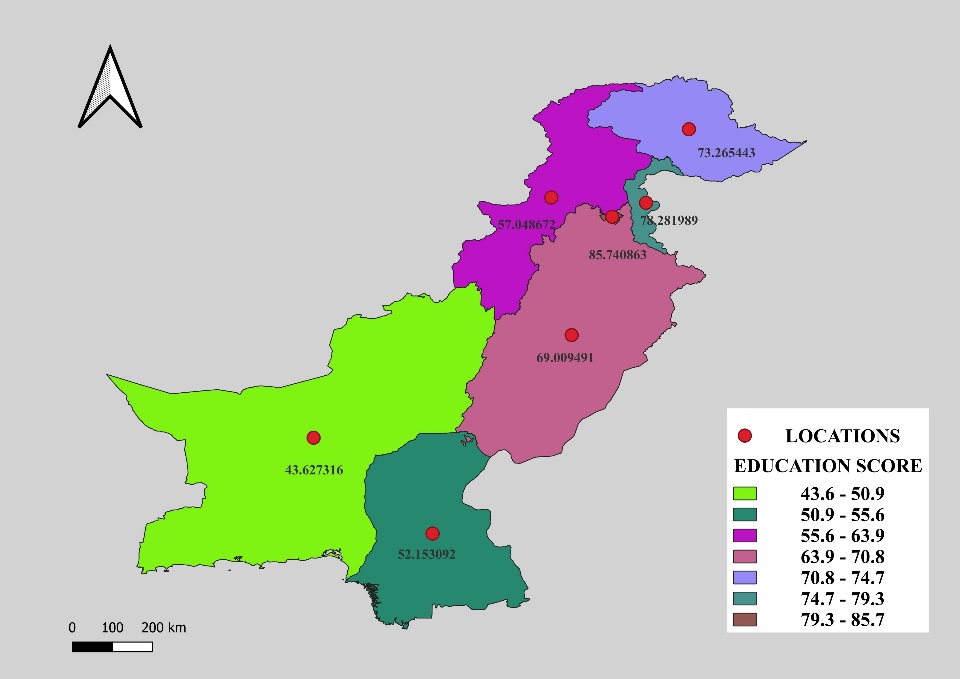


Figure 6: The Education score of the school

Figure 6 shows that Islamabad has the highest Education score of 85.74% followed by Azad Kashmir (78.28%). Balochistan has the lowest education score of 43.63% suggesting poorer overall educational outcomes than other parts of the province.

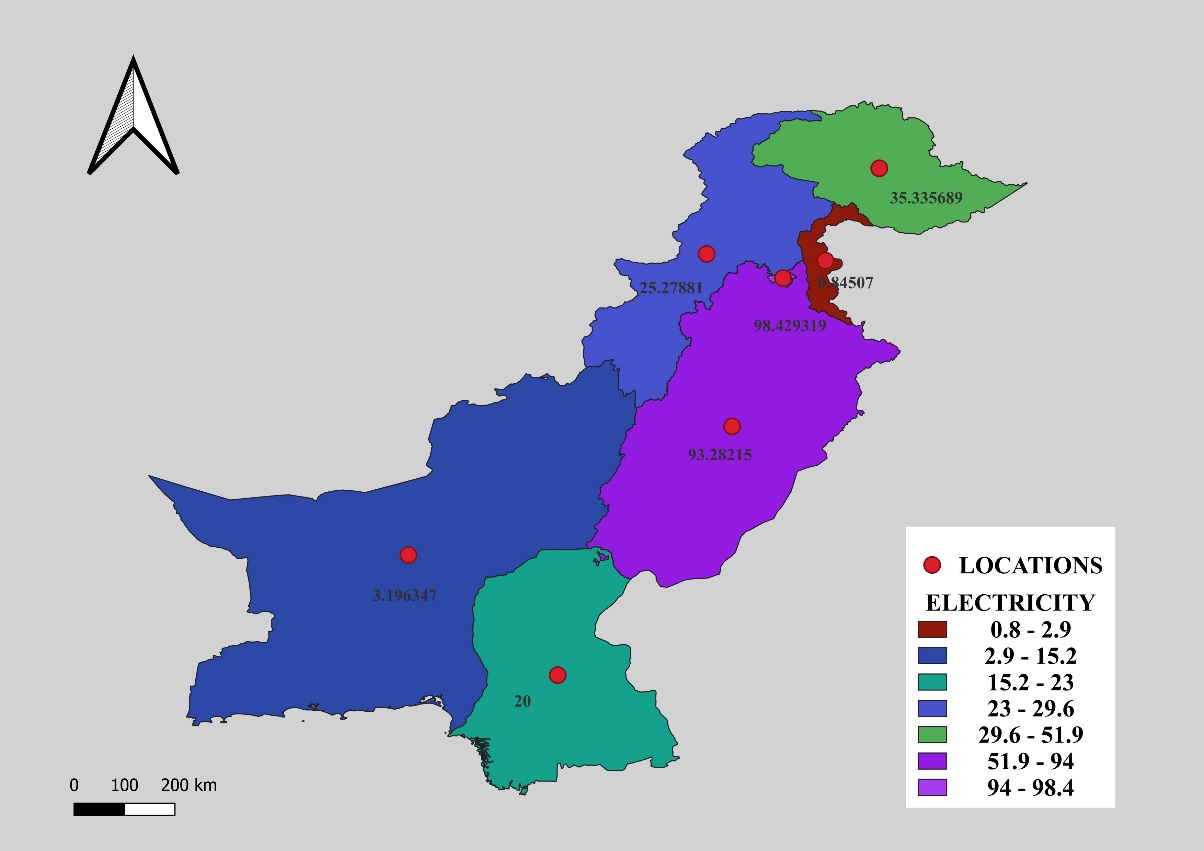


Figure 7: The Electricity of the School

Figure 7 shows that Islamabad has the highest percentage of Electricity at 98.43% followed by Punjab (93.28%). Azad Kashmir has the lowest electricity of 0.85% followed by Balochistan (3.20%). This disparity in electricity provision could impact the quality of education and the ability to use technology-based learning resources.

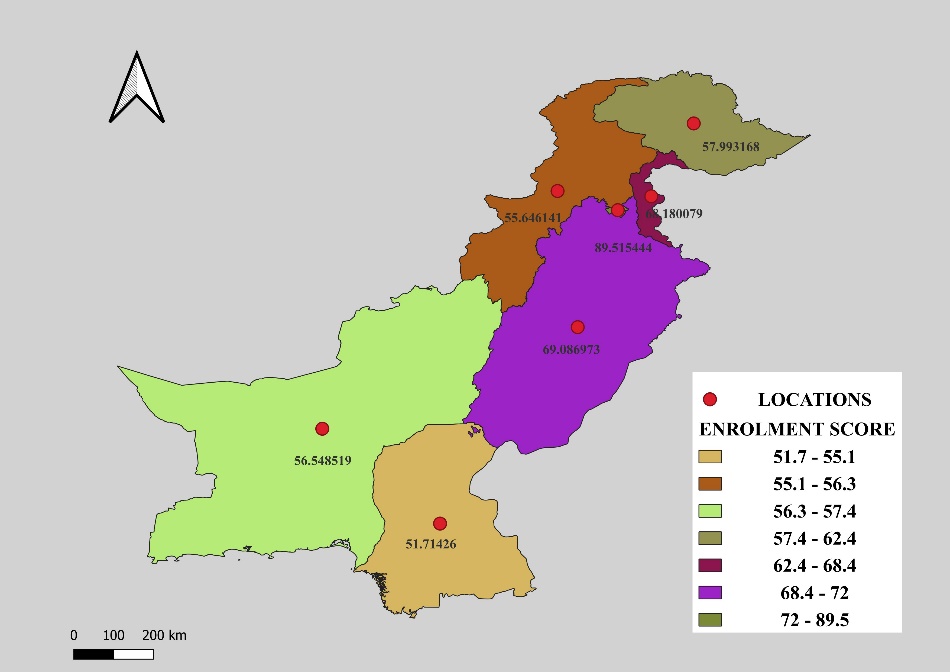


Figure 8: The Enrolment score of the school

Figure 8 shows that Islamabad has the highest Enrolment score of 89.52% followed by Punjab (69.09%). Sindh (51.71%) and Khyber Pakhtunkhwa (55.65%) have the lower enrolment score suggesting challenges in ensuring widespread access to education.

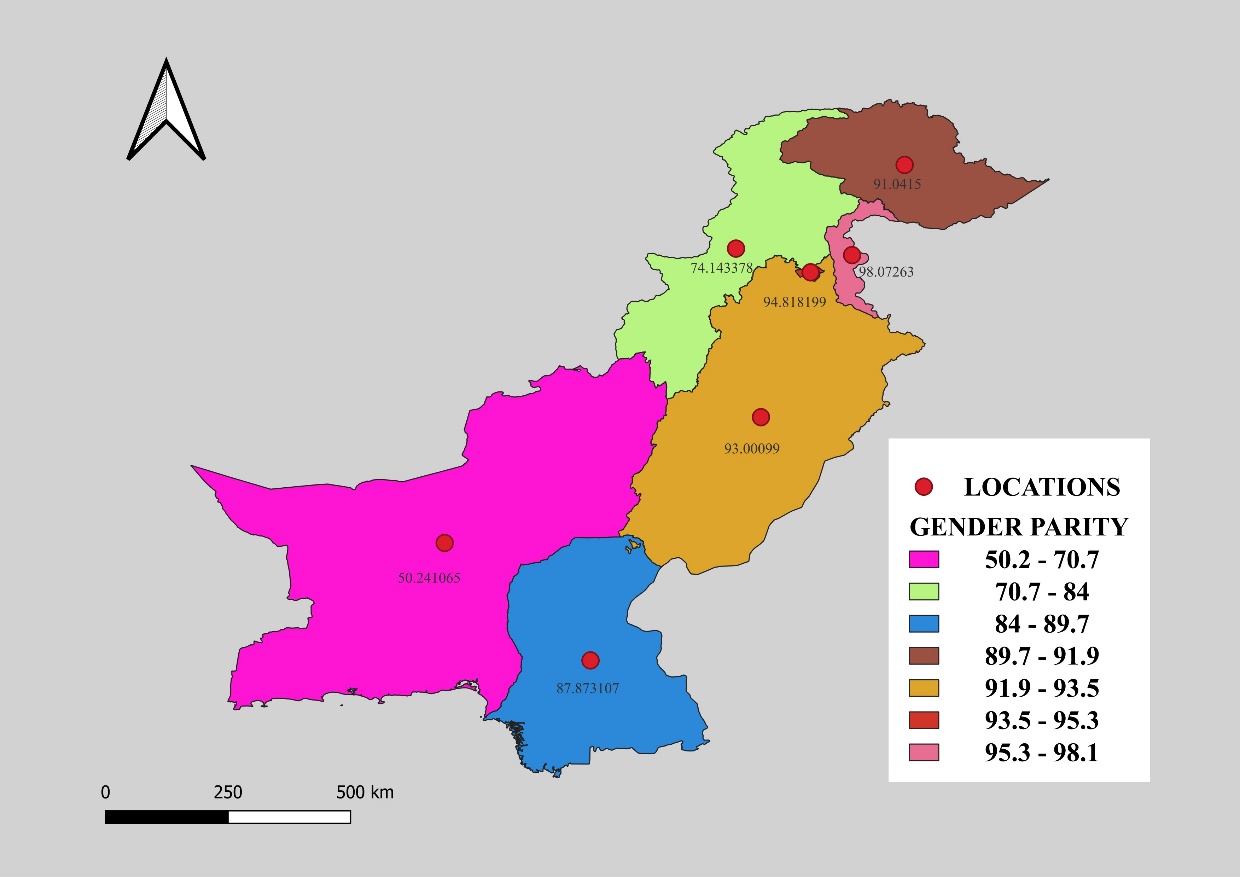


Figure 9: The Gender Parity of the School

Figure 9 shows that Azad Kashmir has the highest Gender Parity score of 98.07% followed by Islamabad (94.82%). This indicates relatively better gender equity in educational opportunities in these regions compared to others, such as Balochistan.

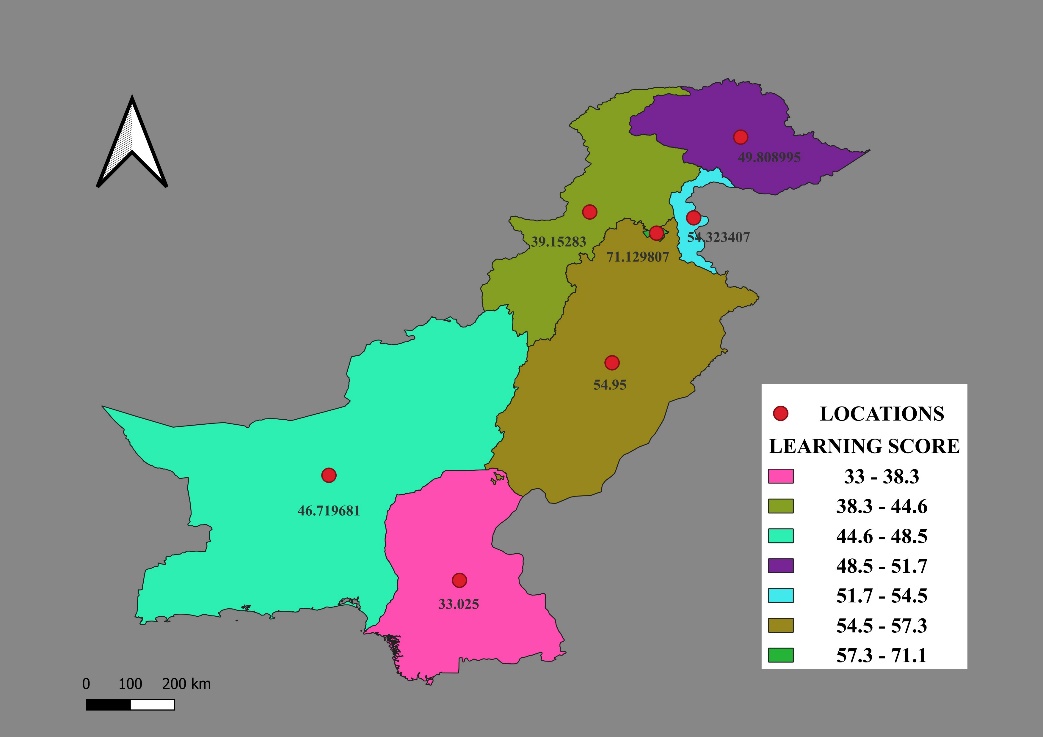


Figure 10: The Learning Score in School

Figure 10 shows that Islamabad has the highest Learning score of 71.13% followed by Punjab (54.95%). The lower learning scores in Sindh (33.03%) and Khyber Pakhtunkhwa (39.15%) suggest challenges in ensuring effective teaching and learning.

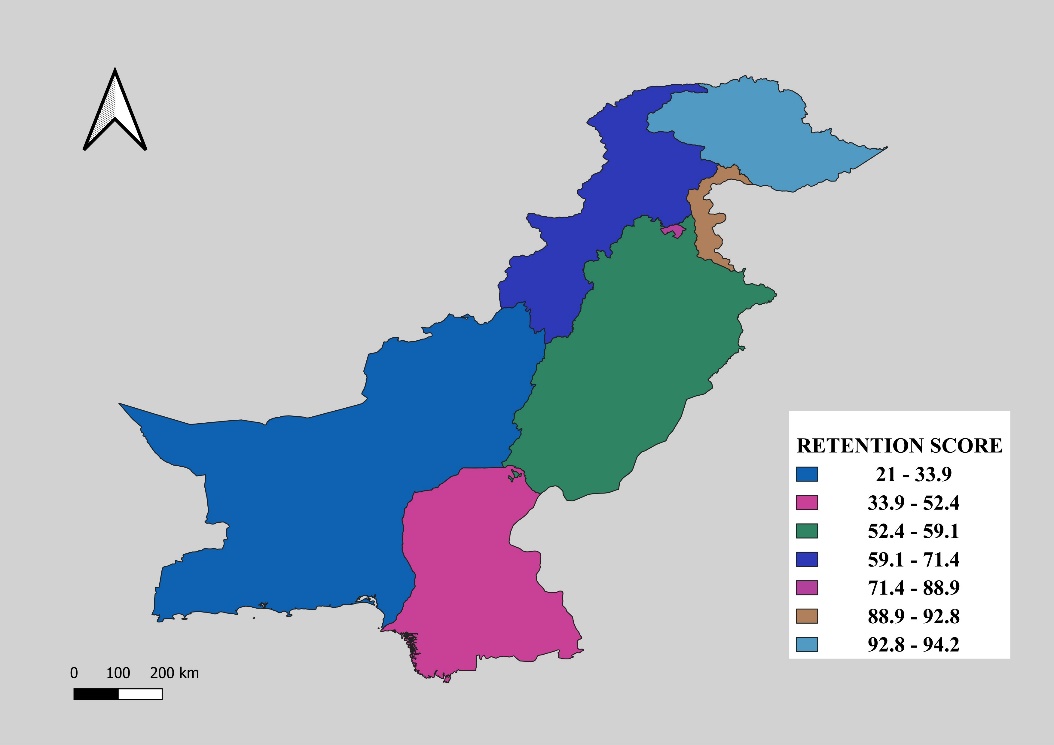


Figure 11: The Retention Score in School

Figure 11 shows that Gilgit Baltistan has the highest Retention score of 94.22% followed by Azad Kashmir (92.55%). Balochistan (21%) and Sindh (36%) have the lowest retention scores, highlighting the need for interventions to address dropout and retention issues in this region.

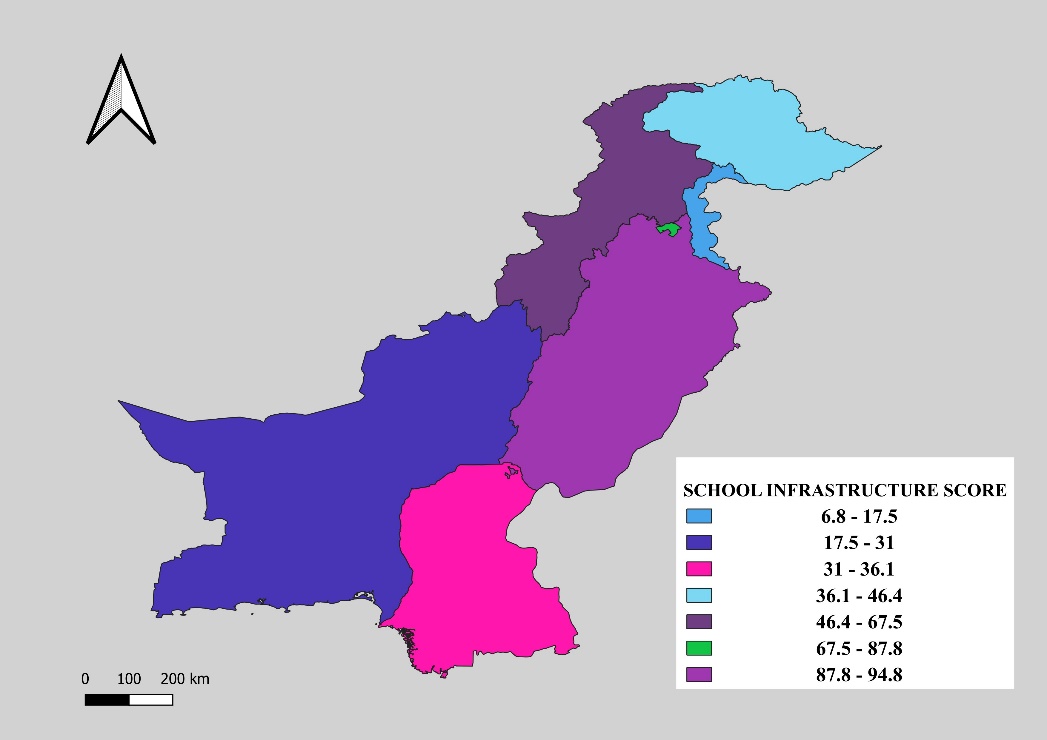


Figure 12: The Infrastructure Score in School

Figure 12 shows that Punjab have the highest Infrastructure score of 94.85% followed by Islamabad (86.60%) and Khyber Pakhtunkhwa (59.85%) while Azad Kashmir (6.76%) and Balochistan (19.31) have the lowest infrastructure score. This suggests that the overall spatial data infrastructure supporting school performance is relatively stronger in the former regions compared to the other.

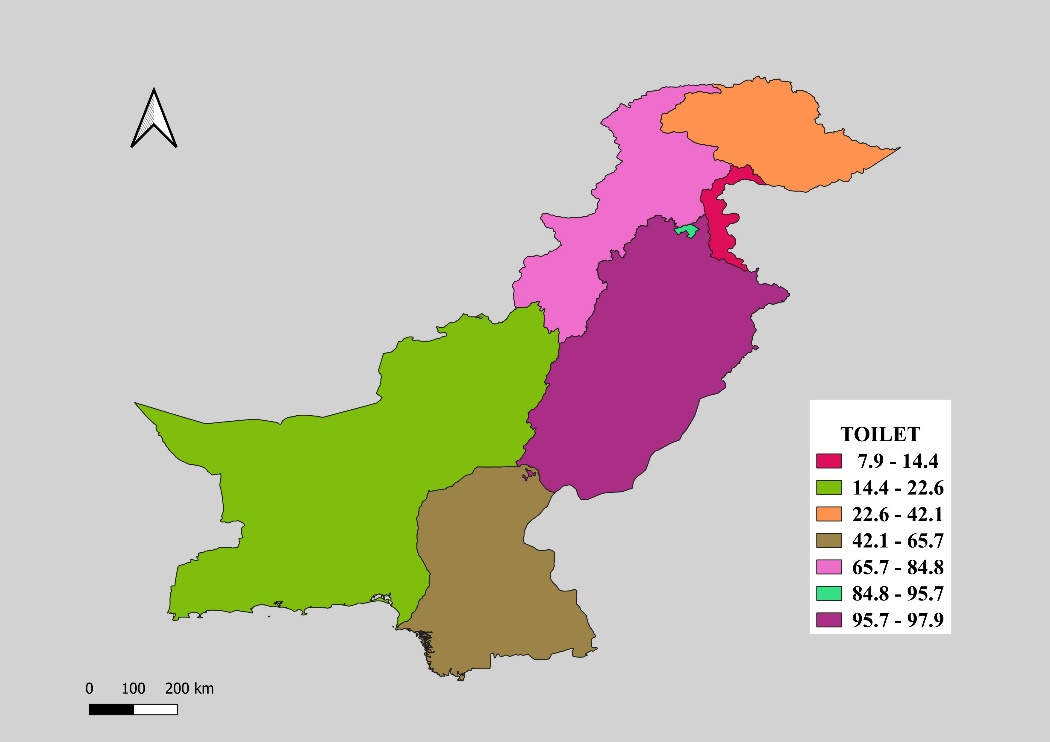


Figure 13: The Toilet in school

Figure 13 shows that Punjab has the highest Toilet of 97.89% followed by Islamabad (95.29%), while Azad Kashmir (7.89%) and Balochistan (15.53%) are the lowest. The availability of adequate sanitation facilities is crucial for student health and well-being, and the disparities observed across regions indicate a need for targeted improvements.

# **3.6 PERFORMANCE**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Indicators | Mean | Standard Deviation | Minimum | Maximum |
| Boundary wall | 60.045224 | 33.196945 | 3.943662 | 98.080614 |
| Drinking water | 45.022582 | 36.953480 | 4.225352 | 99.328215 |
| Education score | 65.589552 | 15.130617 | 43.627316 | 85.740863 |
| Learning score | 49.872817 | 12.289281 | 33.025000 | 71.129807 |
| Electricity | 39.481055 | 40.367426 | 0.845070 | 98.429319 |
| Enrolment score | 64.097798 | 12.952826 | 51.714260 | 89.515444 |
| Gender parity score | 84.170124 | 16.819086 | 50.241065 | 98.072630 |
| Retention score | 64.217470 | 28.754133 | 21.000000 | 94.218109 |
| Infrastructure score | 48.488608 | 33.250473 | 6.760563 | 94.845025 |
| Toilet | 53.884483 | 38.205842 | 7.887324 | 97.888676 |

Table 2: The Statistics Analysis of School Performance

The gender parity score was the highest, with a mean of 84.170124, with a standard deviation of 16.819086. This indicates that on average, the schools are performing well in terms of achieving gender equity, with the maximum score reaching 98.072630.

Electricity has the lowest mean of 39.481055, with a standard deviation of 40.367426. This suggests that the importance level of schools has limited or lacks access to reliable electricity, with the minimum score being as low as 0.845070. The shortfall in electricity supply could be a significant challenge in providing a conducive learning environment for students.

# **CHAPTER FOUR**

# **CONCLUSION**

This study analyses the influence of Spatial Data Infrastructures on digital transformation in education within Pakistan by looking at the present state of educational and infrastructural indicators across various provinces. It highlighted that disparities affect education and access, particularly in Azad Kashmir and Balochistan (Nguyen & Raju, 2014). The research showed that Islamabad and Punjab perform better than other areas, with good infrastructure, high enrolment rates, as well as conductive environments for learning as demonstrated in Table 3.

|  |  |  |
| --- | --- | --- |
| INDICATORS | PROVINCES | SCORE % |
| Boundary Wall | Islamabad | 98.08 |
| Drinking Water | Punjab | 99.33 |
| Education Score | Islamabad | 85.74 |
| Learning Score | Islamabad | 71.13 |
| Electricity | Islamabad | 98.43 |
| Enrolment Score | Islamabad | 89.52 |
| Gender Parity Score | Azad Kashmir | 98.07 |
| Retention score | Gilgit Baltistan | 94.22 |
| Infrastructure score | Punjab | 94.85 |
| Toilet | Punjab | 97.89 |

Table 3: The High School Performance

On the other hand, there were major deficiencies in terms of inadequate boundary walls, absence of drinking water and poor electricity supply in Azad Kashmir and Balochistan has low education and learning indices indicating systemic problems affecting students’ achievements as demonstrated in Table 4 below.

|  |  |  |
| --- | --- | --- |
| INDICATORS | PROVINCES | SCORE % |
| Boundary Wall | Azad Kashmir | 3.94 |
| Drinking Water | Azad Kashmir | 4.23 |
| Education Score | Balochistan | 43.63 |
| Learning Score | Sindh | 33.03 |
| Electricity | Azad Kashmir | 0.85 |
| Enrolment Score | Sindh | 51.71 |
| Gender Parity Score | Balochistan | 50.24 |
| Retention score | Balochistan | 21 |
| Infrastructure score | Azad Kashmir | 6.76 |
| Toilet | Azad Kashmir | 7.89 |

Table 4: The Low School Performance

Data analysis has brought out key insights that show glaring gaps in key services including electricity and water supply to most schools with an adverse impact on student health conditions thereby affecting their learning capabilities. Besides this retention rates together with gender parity figures reveal significant barriers which limit equal access to education.

# **4.1 FUTURE WORK**

With the integration of Spatial Data Infrastructures into Pakistan’s educational system, a multifaceted approach is required. Having a complete and cohesive framework that combines different types of academic and infrastructural data structures is essential.

Moreover, strategies should be designed to address the needs of poorly performing areas such as Azad Kashmir and Balochistan. This may involve identifying infrastructure gaps with facilitating access to essential services like clean drinking water and electricity which are key for effective learning environments.

# **REFERENCES**

[1] Ali, A. & Imran, M., 2021. The Evolution of National Spatial Data Infrastructure in Pakistan - Implementation Challenges and the Way Forward. *International Journal of Spatial Infrastructures Research,* June, Volume 15, pp. 110 - 142.

[2] Duhdra, R., 2023. Introduction to Higher Education 8625 unit 1 part-1. 26 Dec.

[3] Hu, Y. & Li, W., 2017. Spatial Data Infrastructures. *Geographic Information Science & Technology Body of Knowledge. April 2017,* 31 August.

[4] Pradheepan, R. & Neamat, E. G., 2019. Fraud Detection using Machine Learning and Deep Learning. *International Conference on Computational Intelligence and Knowledge Economy (ICCIKE),* 11-12 December.

[5] Rajabifard, A. & Williamson, P. I., 2003. Spatial data infrastructures: concept, SDI hierarchy and future directions. February.

[6] Sazp, K. M., 2013. Introduction to the Spatial Data Infrastructure. *1st National Workshop Bosnia and Herzegovina Mostar,* 16 May.

[7] Suleiman, M. M. & Danmuchikwali, G. B., 2020. Digital Education: Opportunities, Threats, and Challenges. 26 September.

[8] Williamson, I., Rajabifard, A. & Binns, A., 2006. The Role of Spatial Data Infrastructures in Establishing an Enabling Platform for Decision Making in Australia. 6 - 10 November.

[9] Zuberi, S., 2023. The Digital Transformation of Public Education in Pakistan. *UNICEF's Learning Passport Program empowers students,* 22 December.