



Cambodia Digital Desert Index

Bridging the Digital Divide by Uncovering Digital Deserts

Team Fortidata | ITU UMC Hackathon 2026

Mentor : José L. CERVERA-FERRI

Kimhong Set, Molika Taing, Kun Raksa, Phat Soklieng

08 September, 2026



01

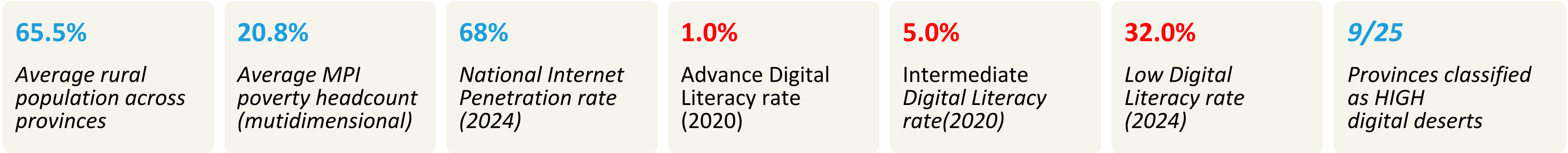
PROBLEM STATEMENT

01 PROBLEM STATEMENT

Problem statement — Geographical, economic & social context



KEY NATIONAL FIGURES



RESEARCH QUESTION

Where are **Cambodia’s digital deserts**, and what is **driving rural people out of meaningful connectivity**?

SCOPE AND UNIT OF ANALYSIS



OBJECTIVE

- 01
- Construct a validated province-level composite index measuring digital exclusion across **six dimensions**.
- 02
- Identify and rank Cambodia’s digital deserts—provinces where multi-dimensional exclusion is most severe.
- 03
- Diagnose **primary drivers of digital exclusion by pillar** and by cluster through statistical validation.
- 04
- Translate findings into evidence-based **policy recommendations** targeted at the most excluded provinces.

An aerial photograph of a city skyline, featuring a prominent curved skyscraper on the left and a dense urban landscape with various buildings and streets. A semi-transparent blue overlay covers the bottom half of the image, containing the text '02' and 'DATA SOURCES'.

02

DATA SOURCES



Data Sources as National and Provincial-Level

No	Indicator	Source Provider	Type	Page (if any)	Year	Source Link
1	Population Total	National Institute of Statistics	A	Page. 19	2024	Source
2	Rural Population (%)	National Institute of Statistics	A	Page.3	2024	Source
3	Internet Penetration Rate (%)	Internet Society Pulse	B		2024	Source
4	4G Coverage (%)	Ministry of Post and Telecommunications	A		2024	Source
5	Estimated 5G Coverage (%)	DataReportal	B		2026	Source
6	Fixed Broadband Subscriptions	Ministry of Post and Telecommunications	A		2025	Source
7	MPI Poverty Headcount	Oxford Poverty and Human Development	A	Page. 8	2021-2022	Source
8	MPI Index	Oxford Poverty and Human Development	A	Page. 8	2021-2022	Source
9	Basic Digital Literacy Rate (%)	UNESCO	B		2020	Source
10	Intermediate Digital Skills Rate	UNESCO	B		2020	Source
11	Female Mobile Phone Ownership (%)	National Institute of Statistics (Cambodia DHS)	A	Page. 336	2021-2022	Source
12	Male Mobile Phone Ownership (%)	National Institute of Statistics (Cambodia DHS)	A	Page. 337	2021-2022	Source
13	Computer Ownership (%)	Global Data Lab	A		2021	Source
14	Female Internet Usage Rate (%)	National Institute of Statistics (Cambodia DHS)	A	Page. 45	2021-2022	Source
15	Male Internet Usage Rate (%)		A	Page. 46	2021-2022	Source
16	Female Social Media Usage (%)	DataReportal	B		2026	Source
17	Male Social Media Usage (%)	DataReportal	B		2026	Source
18	Global Cybersecurity Index (GCI) Score	ITU	B		2024	Source
19	Online Safety (CSEA Rate)	UNICEF Cambodia	B		2022	Source

Type A: Direct provincial-level data

Type B: Estimated/proxy-based indicators

03

METHODOLOGY

03 CONCEPTUAL FRAMEWORK

Indicator Classification

A

10 Indicators

Type A where province-level observed values exist

Availability 4 Indicators

16.7%

internet_penetration_pct	B
4g_coverage_pct	A
estimate_5g_coverage_pct	B
fixed_broadband_subs	A

Affordability 3 Indicators

17.2%

rural_population_pct	A
mpi_poverty_headcount_pct	A
mpi_index	A

Skills 2 Indicators

15.7%

basic_digital_literacy_pct	B
intermediate_skills_pct	B

B

8 Indicators

Type B logit disaggregation uses provincial proxy variables to estimate provincial distributions anchored to national values.

Device 3 Indicators

16.7%

female_mobile_ownership	A
male_mobile_ownership	A
computer_ownership_pct	A

Usage 4 Indicators

16.7%

female_internet_pct	A
male_internet_pct	A
social_media_female_pct	B
social_media_male_pct	B

Safety 2 Indicators

17%

gci_score	B
csea_youth_pct	B

The rate under each pillar is derived from PCA weights.



SAE via Logit Disaggregation

Step 1: Formula Derivation ^a

For Type B indicators, a national-level anchor value Y_{nat} is available but no province level breakdown exists.

$$\text{logit}(Y_j) = \text{logit}(Y_{\text{nat}}) + \frac{1}{k} \sum_{p=1}^k s_p \left[\text{logit}(X_{pj}) - \text{logit}(\bar{X}_p) \right]$$

The raw estimate is then rescaled to recover the national anchor:

$$Y_j^{\text{final}} = \text{expit}(\text{logit}(Y_j)) \times \frac{Y_{\text{nat}}}{\bar{Y}_{\text{raw}}}$$

where $\text{expit}(x) = (1 + e^{-x})^{-1}$

Step 3: Normalisation ^a

$$\text{adj}(j) = \begin{cases} 1 - \hat{Y}(j) & \text{if indicator measures access} \\ \hat{Y}(j) & \text{if indicator measures deprivation} \end{cases}$$

$$\text{norm}(j) = \frac{\text{adj}(j) - \min_j[\text{adj}(j)]}{\max_j[\text{adj}(j)] - \min_j[\text{adj}(j)]}$$

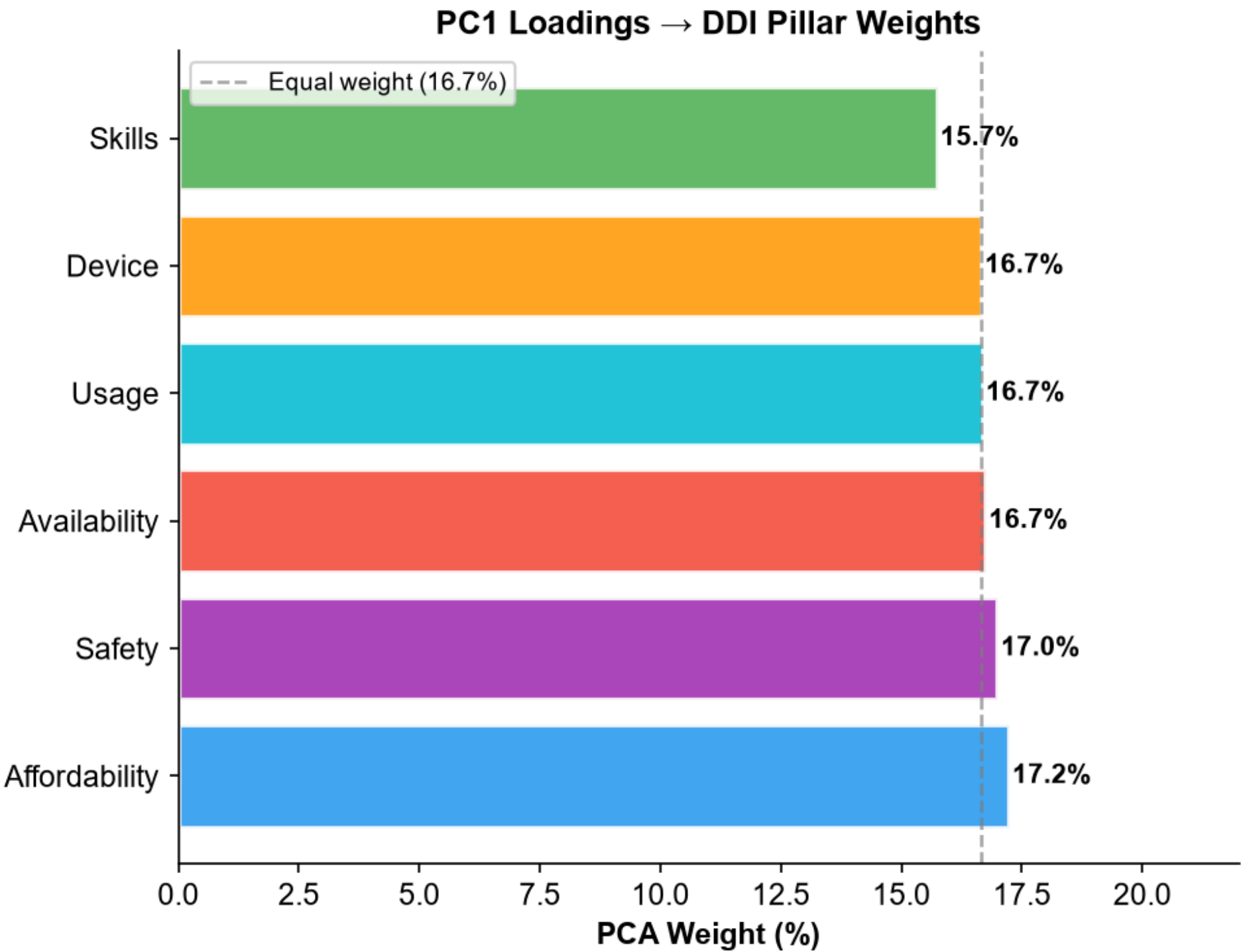
Step 2: Proxy Selection and Validation ^a

Evaluated using the Pearson r coefficient

- $|r| \geq 0.70$: **STRONG** — proxy retained
- $0.50 \leq |r| < 0.70$: **MODERATE** — proxy retained with documentation
- $|r| < 0.50$: **WEAK** — proxy removed iteratively until all remaining proxies are at minimum MODERATE

Indicator	R ²	Adjusted R ²
Internet Penetration	0.993	0.992
5G Coverage	0.745	0.694
Basic Digital Literacy	0.978	0.968
Intermediate Skills	0.969	0.962
Social Media (Female)	0.895	0.873
Social Media (Male)	0.833	0.799
GCI Score	0.897	0.876
CSEA Youth Exposure	0.986	0.983

Step 4: PCA-Based Pillar Weighting

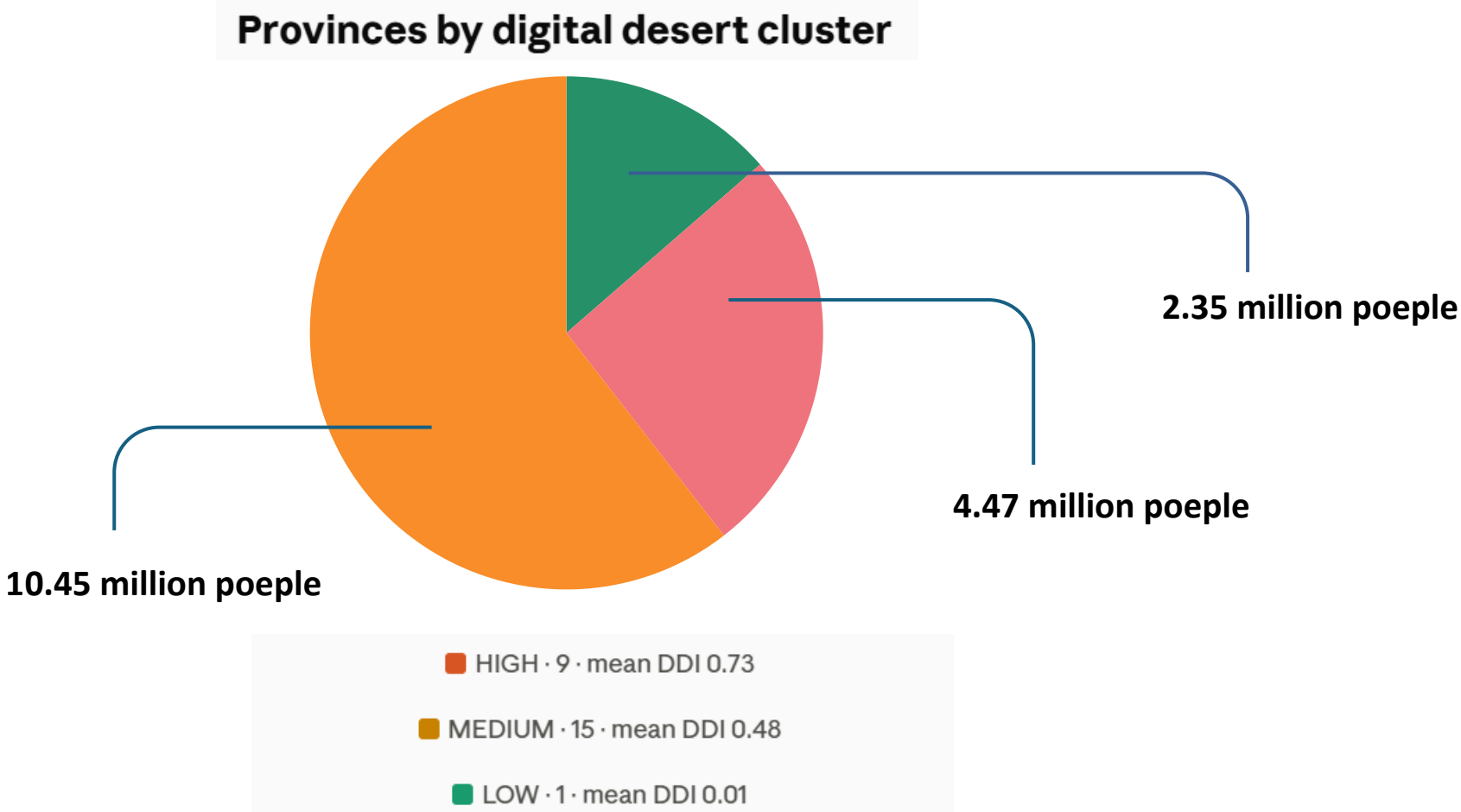


Step 5: DDI Score and Rank

The DDI score for province j is:

$$\text{DDI}(j) = \sum_{p=1}^6 w_p \cdot \text{pillar_norm}_p(j)$$

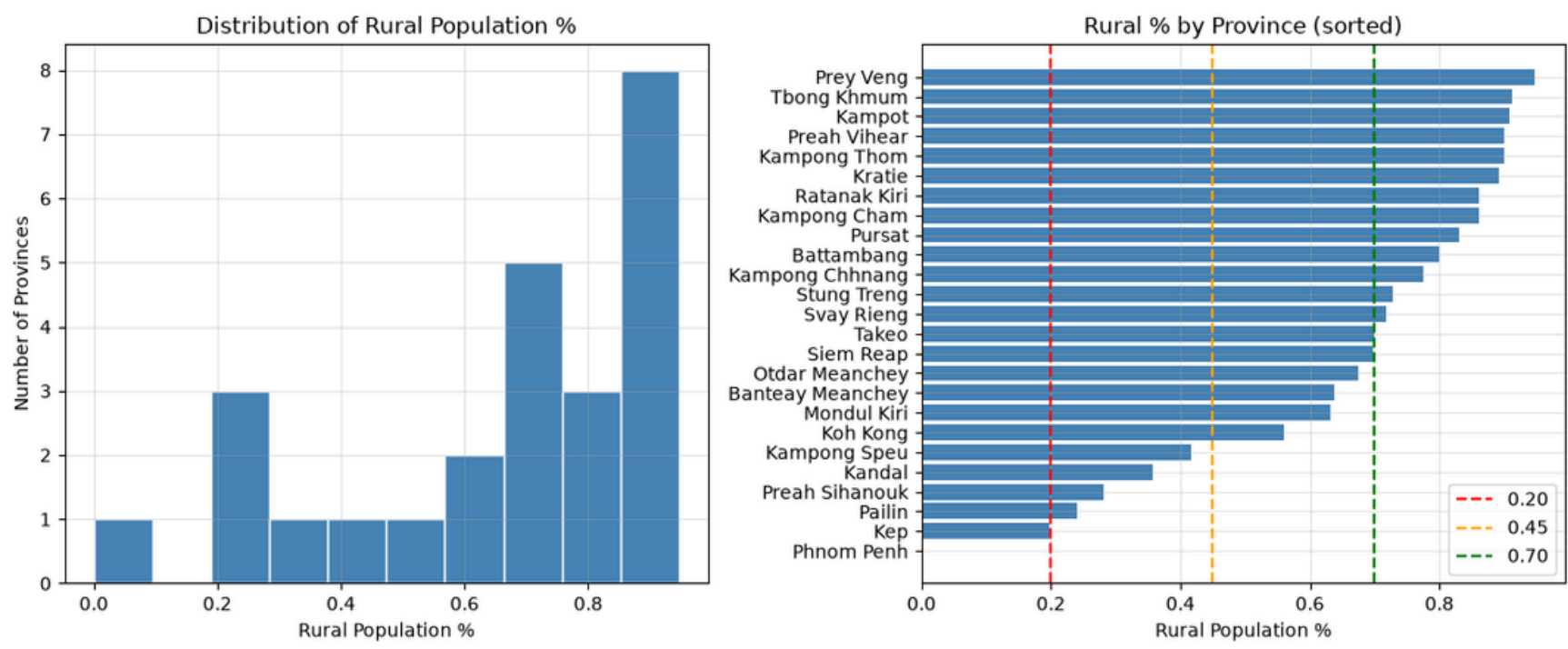
Step 6: Desert Cluster Classification



SAE via Logit Disaggregation (Cont)

Step 7: Zone Classification

Provinces are classified into four geographic zones using K-Means clustering (k = 4) applied to **rural_population_pct**



Quantile-based thresholds:
Q25 = 0.5590 → Urban below this
Q50 = 0.7186 → Peri-urban below this
Q75 = 0.8617 → Peri-rural below this
Above Q75 → Rural

K-means implied thresholds:
Urban min=0.0000 max=0.0000 n=1
Peri-urban min=0.1969 max=0.4161 n=5
Peri-rural min=0.5590 max=0.7276 n=8
Rural min=0.7742 max=0.9469 n=11

Step 8: OECD Variance Validation

Pillar	Indicators	All var > 0	Result
Availability	4	✓	PASS
Affordability	3	✓	PASS
Skills	2	✓	PASS
Device	3	✓	PASS
Usage	4	✓	PASS
Safety	2	✓	PASS
Total	18	✓	ALL PASS

All 18 normalised indicators are confirmed to have variance greater than zero

03 STATISTICAL VALIDATION



Test1: Independent Samples T-Test — HIGHvsMEDIUM/LOW

Hypothesis: high-cluster provinces have significantly higher ddi scores than medium+low-cluster provinces

- $H_0: \mu_{\text{HIGH}} = \mu_{\text{MED+LOW}}$
- $H_1: \mu_{\text{HIGH}} > \mu_{\text{MED+LOW}}$ (one-tailed)

A Mann–Whitney U test is also reported as a non-parametric alternative given $n = 25$.

Comparison	<i>t</i>	df	<i>p</i> -value	Cohen's <i>d</i>	Sig.
HIGH vs MEDIUM+LOW (DDI)	5.14	23	< 0.001	2.22	***
Mann–Whitney U	144.00	—	< 0.001	—	***

$n_{\text{HIGH}} = 9$, $n_{\text{MED+LOW}} = 16$. Note: OTHER cluster non-normal (Shapiro–Wilk $p = 0.002$), Mann–Whitney U is the primary result. Significance: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.



Test 2: Kruskal–Wallis Test — DDI Across Four Zones

Hypothesis: ddi scores differ significantly across the four geographic zones (Ur ban, Peri-urban, Peri-rural, Rural).

- H_0 : The DDI score distributions are identical across all four zones.
- H_1 : At least one zone has a different distribution.

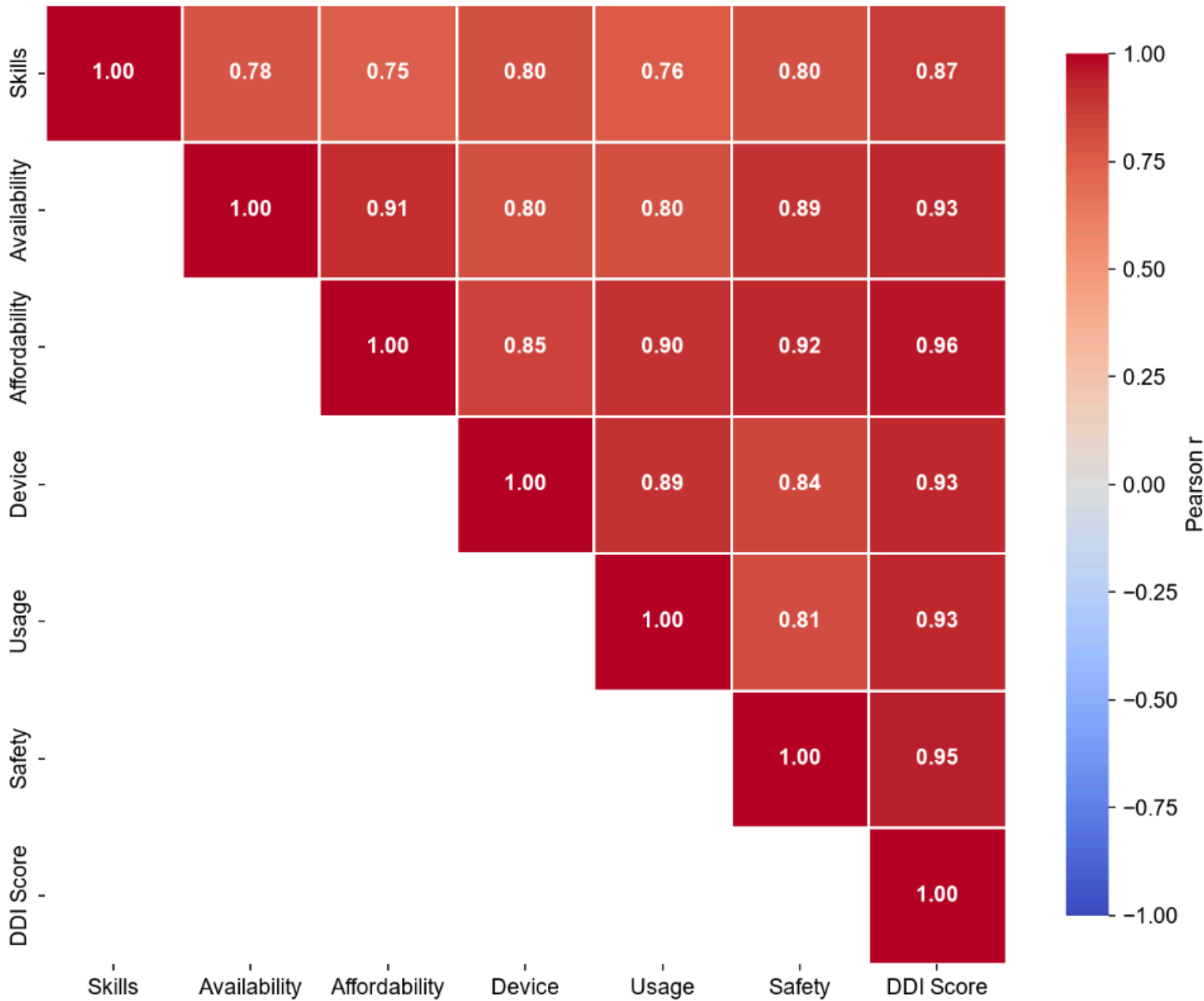
A Mann–Whitney U test is also reported as a non-parametric alternative given $n = 25$.

Zone	<i>n</i>	Mean DDI Score
Urban	1	0.012
Peri-urban	5	0.406
Peri-rural	8	0.571
Rural	11	0.654

Kruskal–Wallis: $H = 12.31$, $df = 3$, $p = 0.006^{**}$. DDI scores differ significantly across zones. DDI increases consistently from Urban to Rural.

Significance: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Figure 11: Pillar Correlation Matrix (including DDI Score)
All six pillars positively correlated with overall DDI



All six pillars are strongly and positively correlated with each other ($r = 0.75\text{--}0.92$) and with the overall DDI score ($r = 0.87\text{--}0.96$) — confirming each dimension captures a genuine, shared pattern of digital exclusion across provinces.

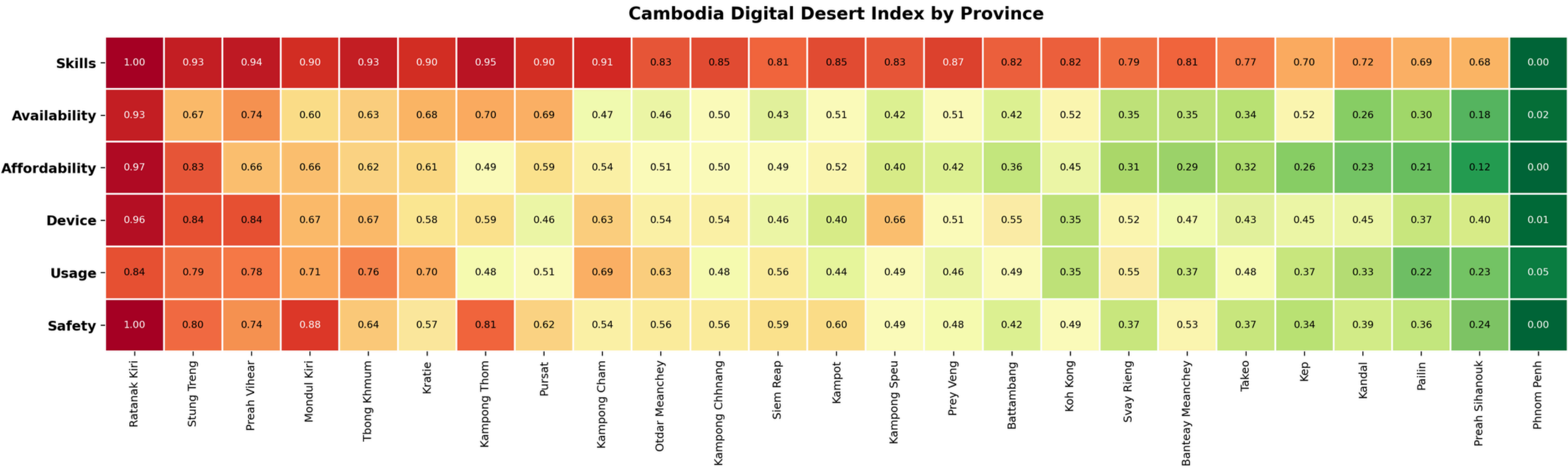
An aerial photograph of a city, likely in Southeast Asia, featuring a mix of modern high-rise buildings and older, lower-density structures. A large blue semi-transparent box is overlaid on the left side of the image, containing the number '04' in white. The sky is overcast with grey clouds.

04

RESULTS

WHAT/WHERE (RANKINGS, MAP, CLUSTER BREAKDOWN)

Pillar-Level finding and Top Five Most Excluded Provinces



Lowest % = perfectly connected, Highest % = completely excluded

**Availability**
Weigth: 16.7%

67.9%

4G & broadband weakest across rural provinces

**Affordability**
Weigth: 17.2%

66.4%

Poverty burden 8× higher in Ratanak Kiri

**Usage**
Weigth: 16.6%

69.6%

Female internet use trails male everywhere

**Device**
Weigth: 17.2%

69.4%

Computer ownership near-zero outside cities

**Skill**
Weigth: 16.6%

92.8%

Nationwide literacy crisis — the worst pillar

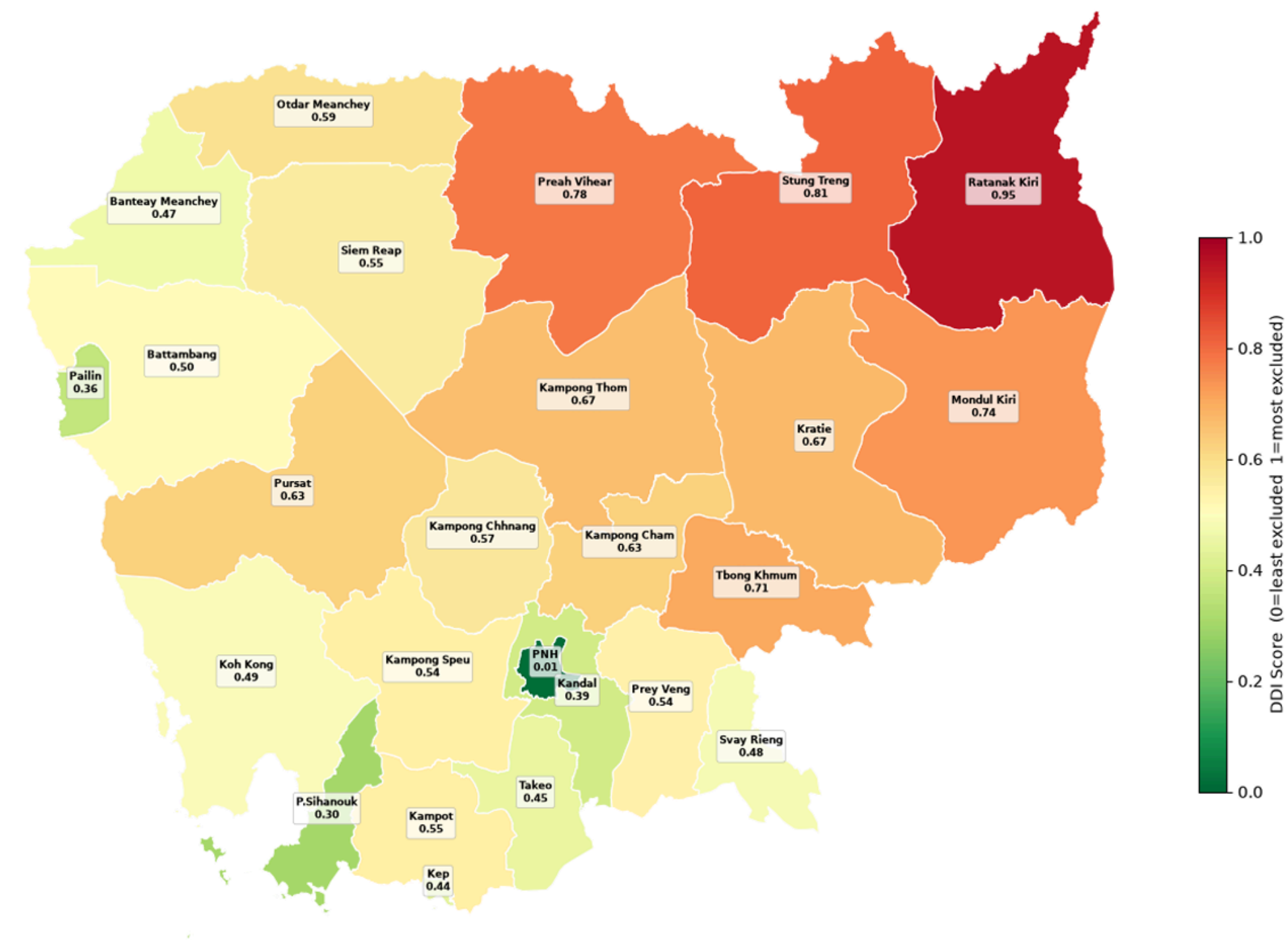
**Safety**
Weigth: 16.6%

73.2%

Youth exposure highest in the northeast

Overall DDI Distribution by Province

DDI Score in Cambodia



National DDI mean

0.552

std dev 0.187

High desert zones

9

36% of provinces

Medium desert zones

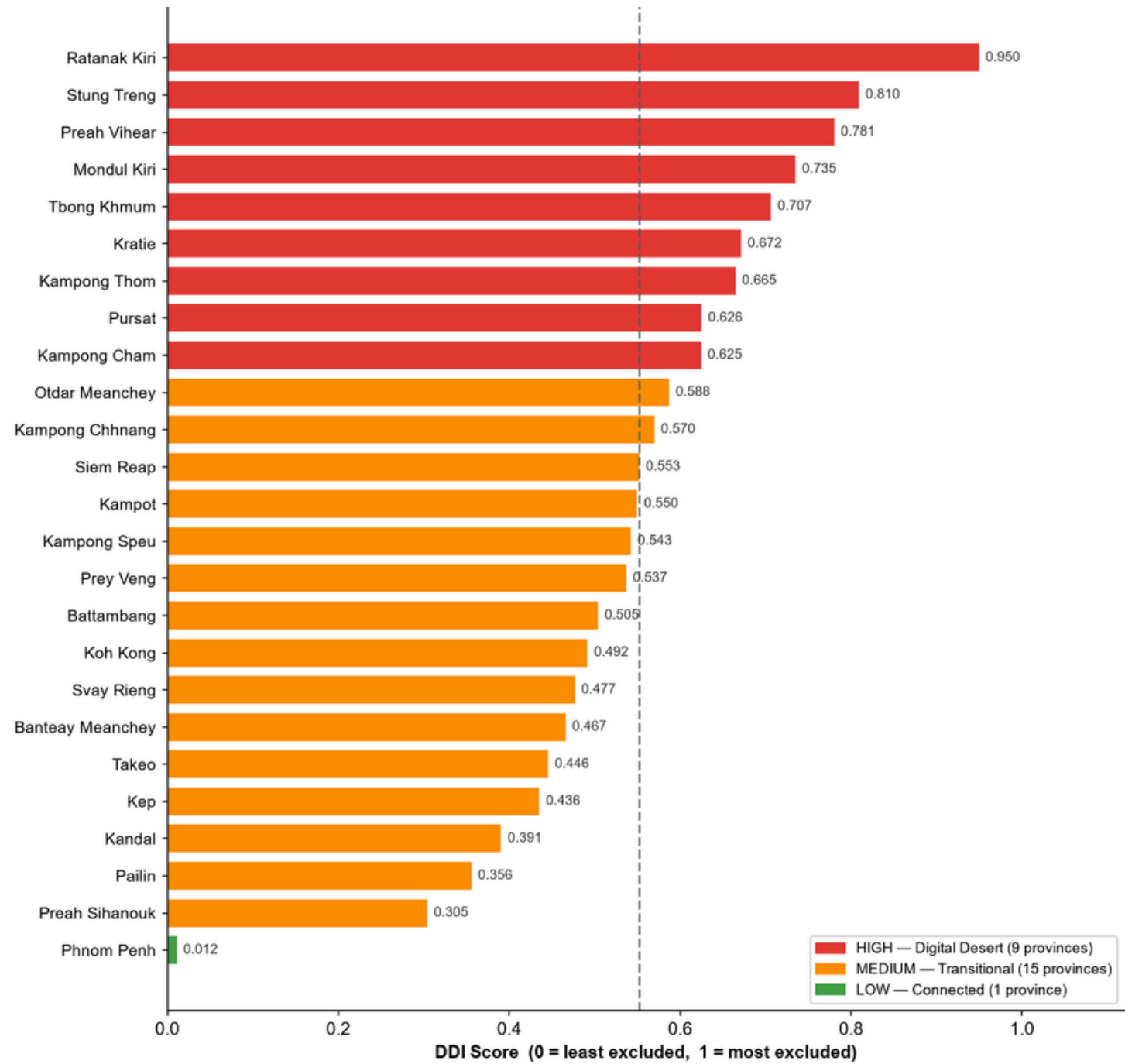
15

60% of provinces

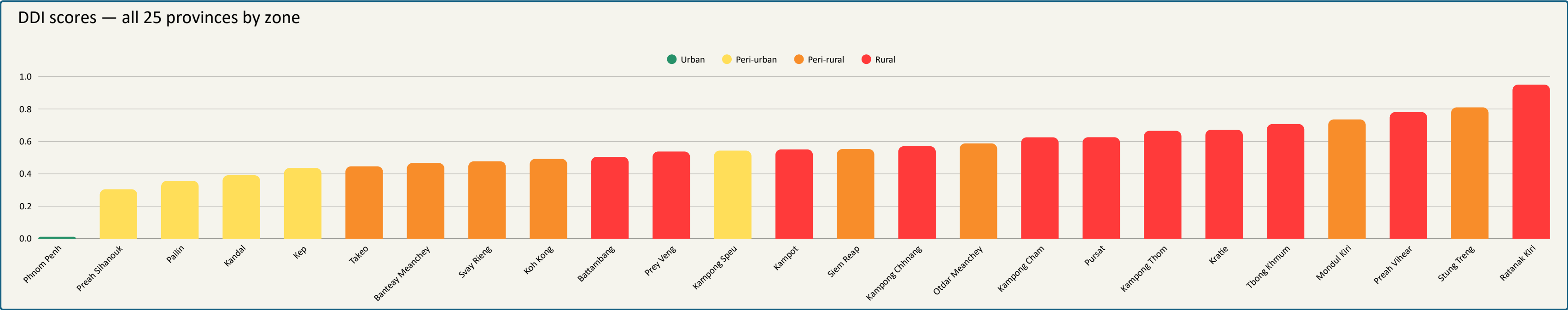
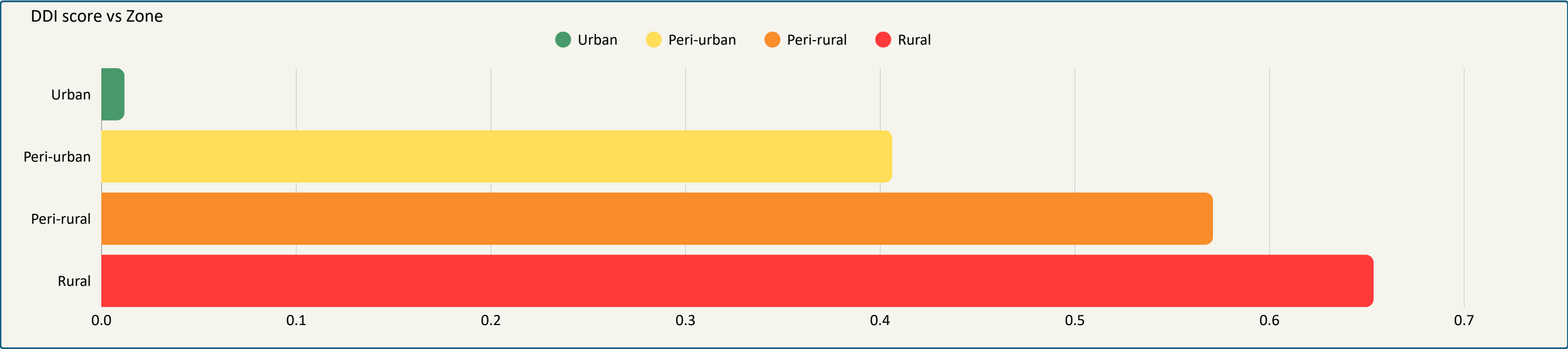
Low desert zones

1

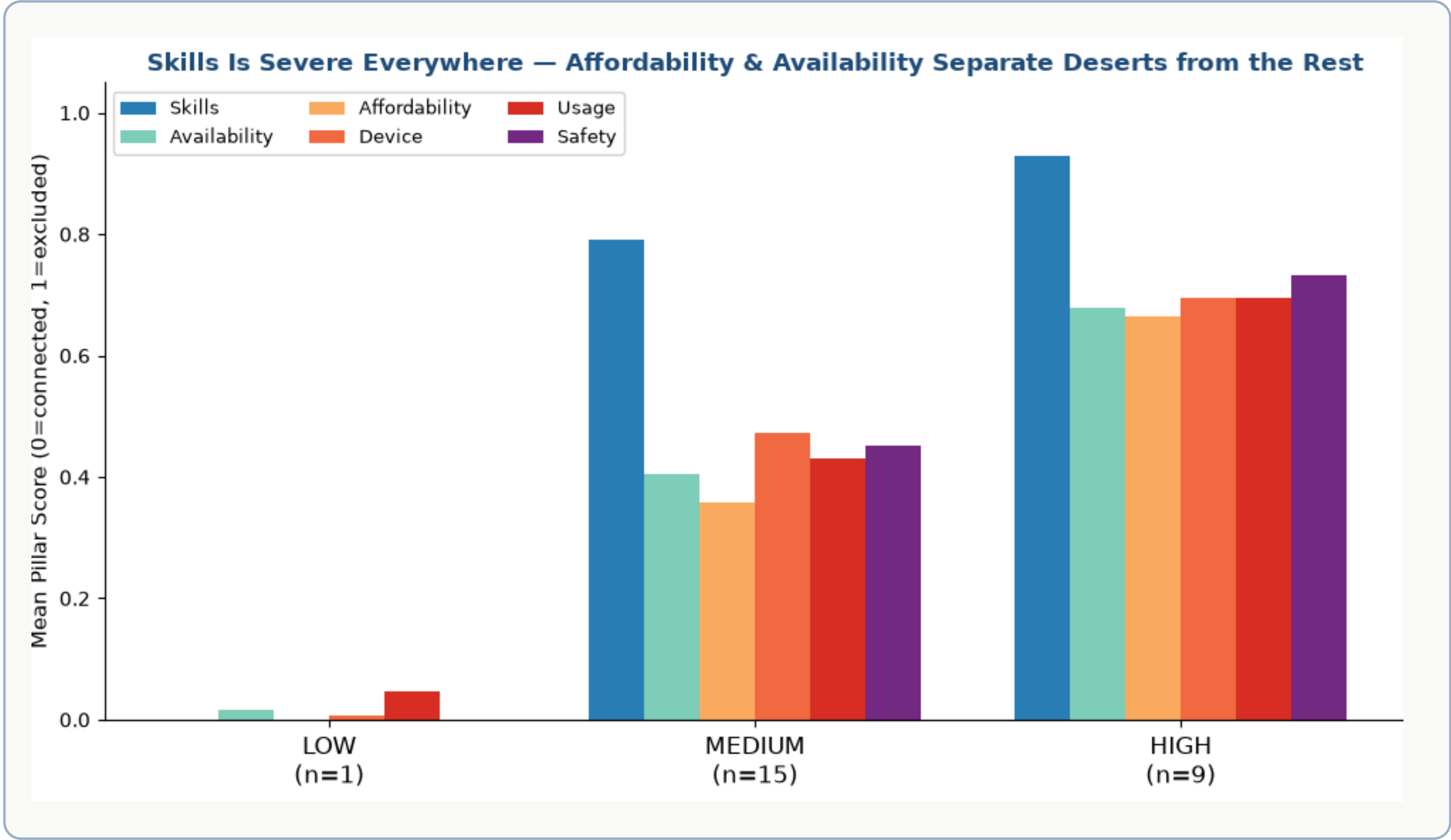
4% of provinces



Overall DDI Distribution



What Drives Digital Exclusion in Cambodia



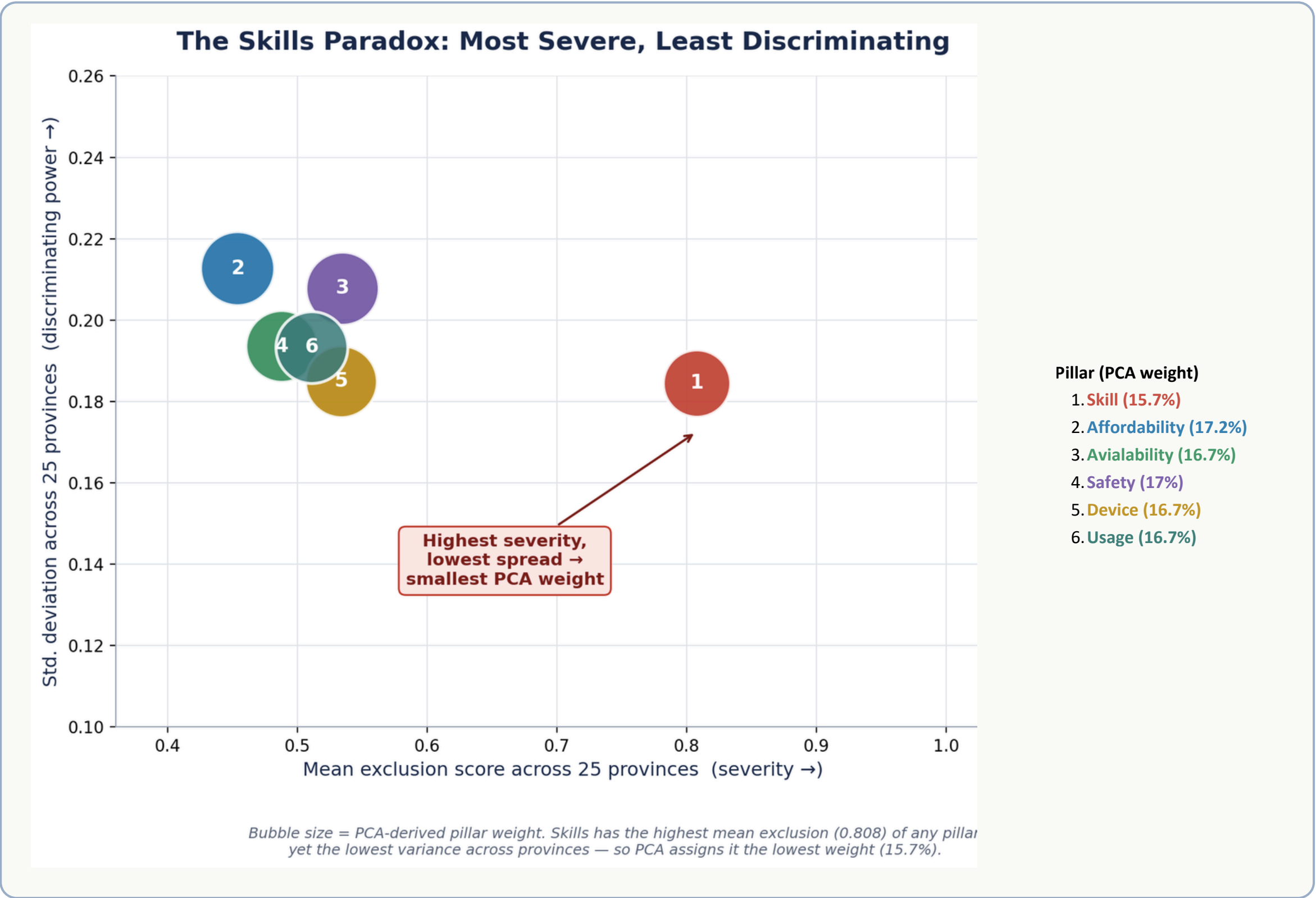
The Pattern Is Consistent — But Not Uniform

Skills deprivation is universal (mean = 0.808), even in MEDIUM/LOW clusters — it's the floor, not what separates provinces.

What actually drives the divide:

- Affordability: 0.358 → 0.664 (nearly 2×, MEDIUM → HIGH)
- Availability: 0.405 → 0.679 (nearly 2×, MEDIUM → HIGH)

▶ The Skill Paradox



Skills training can't just target the 9 worst provinces. It's a nationwide problem that needs a nationwide fix.

An aerial photograph of a city, likely Bangkok, featuring a mix of modern high-rise buildings and older, lower-density structures. A large, light blue semi-transparent rectangle is positioned on the left side of the image, containing the number '05' in white. The text 'POLICY RECOMMENDATIONS' is written in large, white, bold, sans-serif capital letters across the center of the image. The sky is overcast with grey clouds.

05

POLICY RECOMMENDATIONS

1

DIGITAL SKILLS PROGRAMMES

MoEYS and MLVT should embed digital literacy in school curricula and deploy 9 mobile skills units across HIGH-cluster provinces, reaching 1.23M youth and lifting literacy to 32% by 2030.

2

DEVICE LIFELINE

MEF and MoEYS should deliver part-subsidized device vouchers (50–70% of cost) through IDPoor, piloted in two provinces, to raise ownership from ~5% to 15% within three years.

3

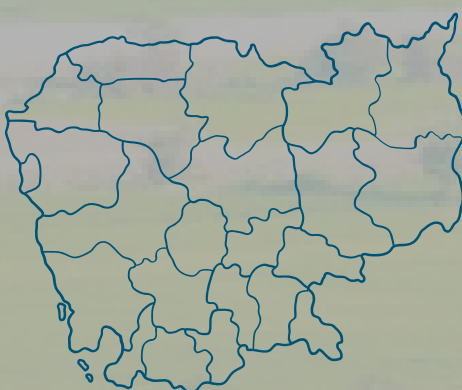
INFRASTRUCTURE MANDATE

TRC and MPTC should mandate 90%+ 4G coverage in all high-scoring DDI provinces by 2028, funded by a Universal Service Fund levy on operators — not the state budget.

4

ONLINE SAFETY

CamCERT (MPTC), and UNICEF, should launch a national online safety programme prioritizing Ratanak Kiri and Mondul Kiri



Cambodia's Digital Desert is a Skills and Poverty Crisis — Not an Infrastructure Crisis

Nine provinces are digital deserts, concentrated in the highland northeast. Rural people have signal — but lack the skills, devices, and affordability to use it meaningfully. The DDI provides a validated, reproducible evidence base for targeting Cambodia's Digital Economy and Society Policy Framework 2021–2035 at the province level.



“BRIDGING THE GAP STARTS WITH KNOWING EXACTLY WHERE IT IS.”