Predicting Urban Affordability and Economic Productivity in India: A Data-Driven KNN and Random Forest Framework with Insights from Selected Major Cities

Shivanshu Pande
Independent Researcher & Data Consultant

Abstract - **India's ongoing digital and economic transformation has intensified spatial disparities across cities in affordability, income potential, and technological infrastructure. This paper introduces an integrated framework linking costof-living affordability and ICT infrastructure, using merged open datasets for thirty major Indian cities. It further employs a dual predictive modeling workflow—K-Nearest Neighbors (KNN) for imputation and baseline estimation, and a Random Forest multi-output regressor for predicting affordability and post-tax salary where data gaps exist. The combined 224-city dataset and its top 30 subset form the analytical backbone, exploring how urban affordability, digital readiness, and GDP specialization jointly shape livability. The study finds that Tier-2 cities demonstrate balanced wage-to-expense coverage despite lower GDP volumes, while metros exhibit a "digital cost paradox." The predictive models strengthen cross-city inference and enable future projections for unlisted cities.**

Index Terms - **Urban economics, affordability, ICT infrastructure, GDP structure, data integration, predictive modeling, KNN, Random Forest, India.**

I. INTRODUCTION

A. Context and Motivation

Rapid digitalization and urbanization are reshaping India's economic geography. Tier-1 metros like Bengaluru, Hyderabad, and Mumbai host high-paying technology sectors but suffer steep housing and cost burdens. In contrast, Tier2 cities such as Nashik or Coimbatore maintain stronger affordability while expanding their industrial base. Understanding how affordability, GDP and digital access intersect provides insight for balanced regional development.

B. Research Gap

Most studies isolate either cost-of-living indices or ICT metrics. Few merge them to understand city-level trade-offs between innovation, wages, and livability. This paper closes that gap through a unified dataset and introduces a machinelearning layer that can estimate missing affordability and salary indicators using KNN and Random Forest models.

C. Objectives

1) Integrate datasets on affordability and GDP, ICT infrastructure, and other relevant metrics .
2) Quantify cross-domain correlations among these metrics.
3) Identify urban typologies that combine digital maturity with affordability.
4) Develop reproducible models and framework (KNN baseline; Random Forest multi-output) for predictive analysis of other cities.

D. Background and Dataset Provenance

This study extends the author's prior dataset of 70+ cities [25] to a master dataset of 221 cities [26], correcting earlier inconsistencies (e.g., mislabeled "India vs Nepal" entries) and adding cities like Bhubaneswar. From this pool, 30 representative cities were chosen based on GDP, infrastructure, and population. These Class I urban centers (0.33 % of India's estimated 9,000+ towns) provide rich, policy-relevant data on economic structure and affordability.To know more about the classifications used kindly refer to the Appendix section at the end.

II. LITERATURE REVIEW

A. Urban Affordability

Affordability blends income adequacy and expenditure patterns. Works such as Glaeser & Gottlieb (2009) [1][3] and Florida (2002) [2] note that productivity booms often inflate living costs—a "prosperity paradox." In India, limited local data exist beyond the SDG Index (2022)[15] or HUDCO (2021) [11]. Tier-1 metros show rent-to-income ratios above 40 %, while Tier-2 cities average 30+%.More than 88% of the urban housing

shortage is among Economically Weaker Sections (EWS) and Low-Income Groups (LIG)[11]. Few combine income, rent, and consumption to yield ratios like "months covered" used in this paper.

B. GDP and Productivity

GDP composition influences wage levels and affordability. Duranton & Puga (2020) [4] show that diversified cities are more resilient. Indian data (CSO 2022) [9] reveal that while services dominate GDP, industrial cities such as Surat or Pune maintain better wage–cost balance.

C. ICT Infrastructure and Digital Inclusion

ICT infrastructure acts as both a growth driver and equity variable. ITU (2023) [11] and World Bank (2022) [12] highlight the positive link of broadband to GDP. In India, TRAI (2024) reports metro internet access near 88%, but mid-sized cities lag at 61%. Singh & Narayan (2023) [7] confirm that ICT correlates positively with income and inversely with inequality.

D. Integrated Perspectives

Global indices like UN-Habitat's *City Prosperity Index* and OECD's *Regional Well-being Framework* merge economy and livability metrics, but India lacks granular models. This paper fills that void using an integrated affordability–GDP–ICT dataset and predictive modeling to infer missing city data.

III. DATA AND METHODOLOGY

A. Dataset Overview

Four datasets were merged:

- **Master 221 Cities File (Self Created)**: living cost, rent, salary, affordability ratios, and population [25][26]
.
- **ICT Sub-Dimension Dataset (Kaggle):** broadband, internet access, Wi-Fi density, and smart-infra metrics. [27]
- **GDP Merged Top 30 Dataset (Self Created):** Living cost, rent, salary, affordability ratios,GDP and population of selected top 30 cities.The selection creteria is mentioned in Appendix. [26] [Appexdix]

B. Data Preparation

City names were standardized (e.g., "Bengaluru"→"Bangalore"), numeric fields normalized, and missing data imputed using KNN. Pearson correlations were computed on cleaned numeric columns, with $|r| > 0.5$ considered strong and $0.3 < |r| < 0.5$ moderate. Correlations of Abs(.3) and below are not considered for the purpose of analysis.

C. Modeling Workflow

To create a predictive framework for missing or unlisted city affordability metrics:

KNN Baseline Model

K-Nearest Neighbors (KNN) is a non-parametric, instance-based learning method that predicts a value by locating the "k most similar" samples in the dataset and averaging their target values. In this study, KNN was used to estimate affordability under the assumption that cities with comparable economic characteristics exhibit similar cost– salary dynamics.

The baseline KNN model used standardized numerical features such as total cost of living, rent, net salary, income after rent, and estimated monthly expenses. Columns with excessive missingness were removed prior to modeling. The data then passed through a preprocessing pipeline composed of:

- **KNNImputer** for missing value imputation,
- **StandardScaler** for feature normalization,
- **GridSearchCV** for hyperparameter optimization of both imputation and regression components.

Grid search identified $k = 3$ neighbors with distancebased weighting as the optimal configuration. The model demonstrated strong predictive performance on the held-out test set for the primary target, ***months_covered***:

$$\text{RMSE} = 0.0505, \qquad R^2 = 0.9645.$$

A separate KNN model was trained to predict monthly_salary_after_tax_inr, achieving:

$$\text{RMSE} = 1928.12 \text{ INR}, \qquad R^2 = 0.9501.$$

These results indicate that KNN performs effectively for both affordability prediction and salary estimation.

Random Forest Multi-Output Model

A second modeling strategy employed a Multi-Output Random Forest Regressor to jointly predict multiple financial indicators. Random Forest is an ensemble method that constructs a large number of decision trees, combining their outputs to learn nonlinear relationships through structured *if–then* rules. This makes it well-suited for multi-target regression tasks with interacting variables.

The model pipeline included:

- **KNNImputer** for missing values,
- **StandardScaler** for normalization,
- **RandomForestRegressor** with 300 estimators, wrapped in a MultiOutputRegressor.

The model jointly predicted:

- *months_covered,*
- *monthly_salary_after_tax_inr.*

Its performance on the held-out test set was as follows:

Monthly Salary: RMSE = 2039.06 INR, $R^2 = 0.944$,
Months Covered: RMSE = 0.0738, $R^2 = 0.948$.

For instance, in one test case, the actual salary (INR 31,103.79) and predicted salary (INR 32,238.69) differed by only INR 1134.9 (approximately 5%), demonstrating practical prediction reliability.

Model Interpretation

KNN operates by borrowing information from nearby examples—for instance, predicting affordability for Bhopal by averaging values from similar cities such as Indore or Nagpur. In contrast, Random Forest constructs hundreds of decision trees that learn logical rule-based patterns (e.g., "if rent increases and cost-of-living increases, then salary tends to increase but affordability tends to decrease"). This enables Random Forest to capture complex, nonlinear dependencies and perform robust multi-target estimation.

D. Formula

$$\text{Months Covered} = \frac{\text{Monthly Income After Tax}}{\text{Monthly Expenses}} \quad (1)$$

E. Reproducibility, Tools, and Data Availability

All code; datasets, preprocessing scripts, and model training pipelines used in this study will be made publicly accessible through Kaggle/GitHub as interactive, executable notebooks. These notebooks will contain the full source code required to reproduce the results, including data cleaning, feature engineering, model training, evaluation metrics, and figure generation.

All analyses in this work were conducted using Python, primarily within **Google Colab notebooks**. Core libraries included:

- **Pandas**, **NumPy** for data processing,
- **Scikit-Learn** for modeling, preprocessing, and hyperparameter tuning,
- **Matplotlib** (MPL) for visualization,
- **Joblib/JSON** for model persistence and metadata storage.
- **Others...[Please refer to the code]**

Assistance in drafting, code explanation, model interpretation, and documentation refinement was provided using **ChatGPT (OpenAI)** as well as **Gemini (Google)**, employed as a writing and analytical support tool. This contribution is acknowledged as part of the transparency and reproducibility practices recommended in computational research.

Once published, persistent links to these notebooks will be included in the References/Bibliography section.All materials—including code, derived datasets, and analysis notebooks prepared by the author—are released under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Users may share and adapt the content with proper attribution. A recommended attribution format is:

Author (Year). *Title or description of dataset/notebook.* Kaggle. Licensed under CC BY 4.0.

However, I am not stringent on format as long as attribution/credit is given.

The Kaggle repository link will be added to the References/Bibliography section once uploaded.

IV. RESULTS AND ANALYSIS

A. Affordability Trends

The affordability measure ("Months Covered") represents the number of months of expenses covered by a single month's salary. Tier-2 cities show better ratios compared to Tier-1 metros.More on this is mentioned on my first paper and extended master dataset [25][26].

TABLE I: Top and Bottom 5 Cities by Affordability(**T30**)

City	Salary (Rs.)	Affordability Ratio
Nashik	49,291	1.60
Faridabad	48,061	1.43
Varanasi	31,279	1.30
Kanpur	37,957	1.35
Coimbatore	34,179	1.09
Bangalore	27,325	0.67
Delhi	20,296	0.47
Mumbai	19,505	0.31

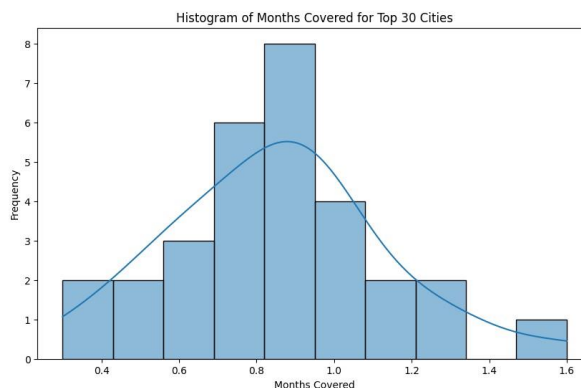*Note: Salary values are after-tax monthly incomes; ratios derived from the master dataset.*

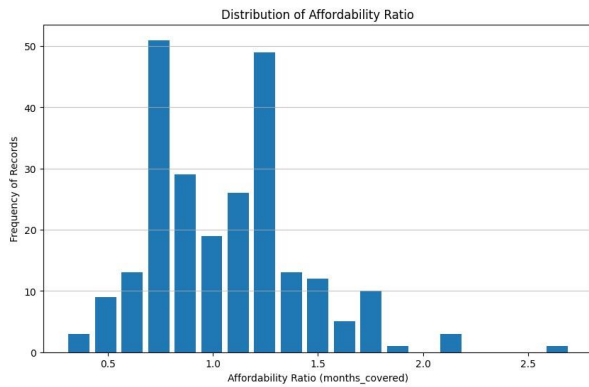Fig. 1: Distribution of affordability ratios across the 30-city dataset.

Fig. 2: Distribution of affordability ratios across the 210+city full dataset.

B. ICT–Affordability Correlation

Observations and Insights: The correlation matrix between ICT subdimensions and affordability metrics (measured by *months_covered* and *monthly_salary_after_tax_inr*) reveals several key relationships:

1) **e-Government efficiency and affordability**: A moderate positive correlation ($r \approx 0.38$ with months_covered; $r \approx 0.34$ with salary) indicates that cities with stronger digital governance tend to offer better affordability outcomes, suggesting that ICT-driven efficiency in service delivery can enhance residents' economic resilience.

2) **Public WiFi availability and living cost:** A negative correlation ($r \approx -0.38$ to -0.25) suggests that cities investing heavily in public WiFi infrastructure tend to have lower affordability—reflecting larger metropolitan areas where connectivity is high but living costs are proportionally greater.

3) **Dynamic transport information systems**: The presence of real-time public transport ICT systems shows a negative link with affordability ($r \approx -0.48$), implying that advanced mobility systems are characteristic of dense, high-cost urban environments.

4) **Household Internet access**: Despite expectations, household internet penetration shows little correlation with affordability ($r \approx 0.10$ with salary, $r \approx -0.15$ with months_covered), suggesting that connectivity has become a baseline urban necessity rather than a determinant of economic comfort.

5) **Broadband and 4G infrastructure**: Fixed broadband and wireless 4G coverage show weak or near-zero correlations ($r \approx -0.16$ to $+0.02$), indicating that the mere availability of digital infrastructure no longer predicts affordability—policy quality and income dynamics now play a more decisive role.

Overall Insight: The results suggest that while ICT infrastructure has reached saturation in many urban centers, its **qualitative application**—particularly through eGovernance—has a stronger influence on affordability. In other words, *digital quantity has peaked; digital quality now defines urban affordability.*

Please Note that ICT dataset and our 30 **cities only 17 cities had common data** in merged analysis hence only those 17 are considered for this heatmap, but findings will be similar considering the baseline outlook.
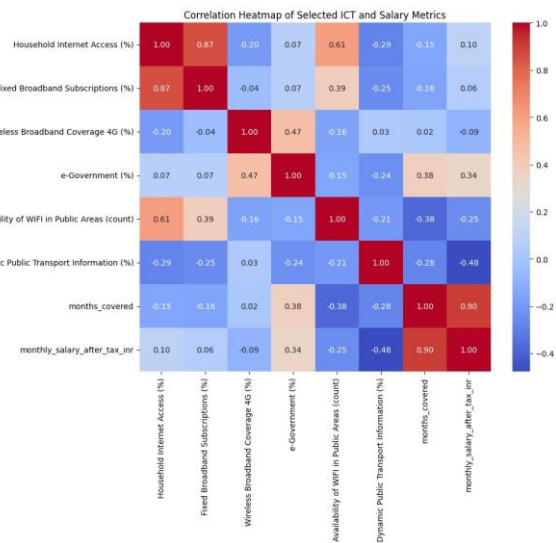

Fig. 3: Pairwise correlation: ICT indicators and affordability metrics.

C. GDP–Affordability Correlation

1. **Urban population and GDP** are highly correlated ($r \approx 0.97$): Larger urban agglomerations naturally correspond to higher nominal GDP values, underscoring the strong link between economic output and population scale in Indian cities.

2. **Population density versus affordability** (negative correlation, $r \approx -0.50$): Cities with larger populations tend to have lower affordability, reflecting the strain on housing, transport, and cost of living that accompanies urban density.

3. **GDP and affordability inverse relationship** ($r \approx -0.55$): Despite higher GDPs, wealth does not always translate into greater affordability. This pattern points to income inequality and elevated living costs in wealthier metros like Mumbai or Delhi.

4. **Salary and affordability** remain strongly linked ($r \approx 0.77$): As expected, higher monthly take-home salaries contribute to better affordability ratios. However, the strength of this

relationship also suggests that affordability gains rely more on wage growth than on structural cost efficiencies.

5. **Weak link between GDP and salary** ($r \approx -0.25$): Interestingly, individual earnings do not scale directly with city-level GDP. This indicates that productivity gains at the macroeconomic level may not be equitably reflected in individual income growth.

Overall Insight: High GDP and large populations are not sufficient for economic comfort. Affordability is driven more by personal income and cost-of-living balance than by macroeconomic scale. This underlines the need for inclusive urban growth where income distribution aligns with economic expansion.
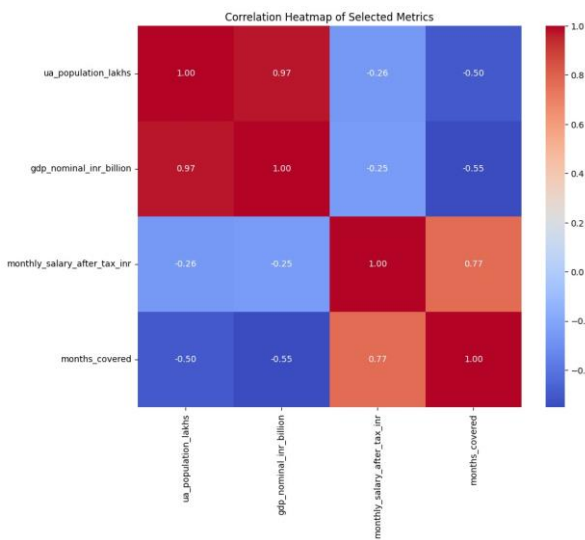

Fig. 4: Correlation between GDP composition and affordability indicators.

D. ICT–GDP Relationship

Observations and Insights:

The correlation analysis between *nominal GDP* and various ICT subdimensions reveals important structural relationships across economic and digital development indicators:

1. **Strong positive correlation between GDP and digital connectivity**: GDP shows a substantial positive association with both *Availability of Public WiFi Areas* ($r \approx 0.73$) and *Household Internet Access* ($r \approx 0.64$). This indicates that higher-income urban economies exhibit stronger ICT infrastructure and connectivity penetration, reflecting how economic prosperity supports digital expansion.

2. **Fixed broadband infrastructure aligns with GDP** ($r \approx 0.39$): While moderate, this relationship underscores the role of reliable fixed broadband access in economically advanced cities. However, it also suggests a digital divide—

smaller cities may still depend more on mobile connectivity than fixed broadband.

3. **Wireless broadband and e-Government exhibit limited correlation** ($r \approx -0.15$ to 0.07): Surprisingly, 4G coverage and e-Government maturity do not strongly correlate with GDP, implying that digital service sophistication may depend more on policy and governance priorities than on economic scale alone.

4. **Dynamic Public Transport ICT systems show weak or negative correlation** ($r \approx -0.24$): The presence of real-time transport information systems is not strongly tied to GDP levels, suggesting that such smart-mobility initiatives often arise from city-level innovation agendas rather than GDP-driven infrastructure spending.

5. **Digital intensity does not always track GDP growth**: The mixed correlations (ranging from strong positive to slightly negative) highlight that while richer cities generally have better connectivity, GDP alone does not guarantee holistic ICT integration. Policy intent, urban management, and digital literacy play significant mediating roles.

Overall Insight: The findings suggest that economic prosperity facilitates ICT availability, but not necessarily its equitable or functional deployment. **While GDP correlates with access, it does not ensure digital inclusivity or smart governance readiness**. *Hence, ICT-led progress depends as much on governance vision and policy implementation as on economic strength.*
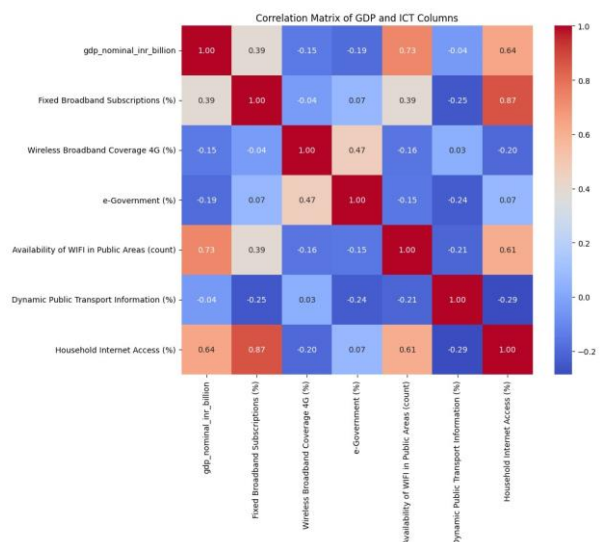

Fig. 5: Relationship between ICT and GDP composition.

E. Predictive Model Performance

The KNN baseline achieved a test RMSE of 0.0505 and $R^2 = 0.9645$ for affordability prediction (*months covered*), demonstrating strong generalization on the held-out test set. The Random Forest multi-output model further provided robust joint estimation of both *months covered* and *monthly salary (INR)*. Performance metrics:

- Salary RMSE = Rs. 2,039.06
- R^2 (Salary) = 0.944
- Months Covered RMSE = 0.0738, R^2 = 0.948
- Example test case: Actual Salary = Rs. 31,103.79; Predicted = Rs. 32,238.69 (Error = Rs. 1,134.90 ≈ 5%)
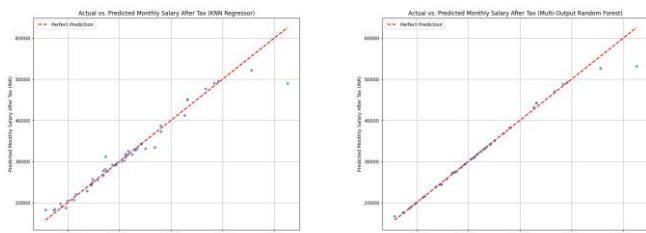
(a) Predicted vs Actual Salary (KNN)

(b) Predicted vs Actual Salary (Random Forest)

Fig. 6: Comparison of Predicted vs Actual Salary for KNN and Random Forest regression models.

F. Three-way Interaction

Tri-domain correlation (ICT, GDP, and Affordability) shows that cities with high ICT readiness and technologysector dominance often face affordability stress, supporting the "digital cost paradox."

G. Environmental and Livability Context (Top 30)

Integrating external AQI data (IQAir, 2024) reveals that Pune, Hyderabad, and Coimbatore exhibit favorable pollution–livability trade-offs. High-AQI metros like Delhi and Kolkata score poorly in affordability and health metrics.Consider this a bonus analysis as this was not part of this paper's scope however this gives us insights in comparing two cities for livibility and good future scope usecase.Please note the below analysis is done on our T30 dataset.

Observations and Insights:

The correlation matrix between *Mean AQI (2025)* and key economic and demographic indicators reveals several meaningful environmental–economic dynamics across Indian cities:

1. **Weak positive correlation with GDP ($r \approx 0.25$)**: The relationship between air pollution levels and nominal GDP is weak, suggesting that while industrial and economic activity may contribute to higher AQI, the link is not consistent across all cities. This highlights that pollution intensity depends more on industrial mix and environmental policy enforcement than on GDP scale alone.

2. **Weak-to-moderate correlation with urban population ($r \approx 0.31$)**: Larger urban populations tend to exhibit slightly higher AQI levels, reflecting urban density and vehicular emissions. However, since the correlation remains low, it indicates that city planning and green initiatives can offset pollution even in populous areas.

3. **Negligible correlation with income and affordability ($r \approx -0.02$ to -0.00):** Air quality does not meaningfully correlate with monthly income or affordability metrics, suggesting that cleaner air is not directly determined by living standards but by city-level infrastructure and environmental regulation effectiveness.

4. **Cleanest cities by mean AQI (2025):** *Thrissur (58.9), Coimbatore (73.2), and Chennai (75.5)* demonstrate relatively clean air, showing effective local governance and sustainable urban planning practices that prioritize environmental quality despite economic growth.

5. **Most polluted cities by** mean AQI (2025): *Asansol (167.1), Patna (185.9), and Delhi (224.6)* record the highest pollution levels, highlighting critical challenges in industrial emissions, vehicular load, and air management. Delhi's persistently extreme AQI underscores the urgency for emission control, public transit reform, and green transition initiatives.

Overall Insight: The results emphasize **that economic prosperity does not inherently determine air quality. While GDP and population exert a weak influence, AQI outcomes are driven primarily by urban policy, industrial zoning, and emission managemen**t. Cities like *Thrissur and Coimbatore* demonstrate that targeted governance can maintain clean air even in growing economies, while others such as *Delhi and Patna* underline the consequences of unchecked urban-industrial expansion. Please note this corelation was done only for 30 selected cities and not 210+ master dataset.
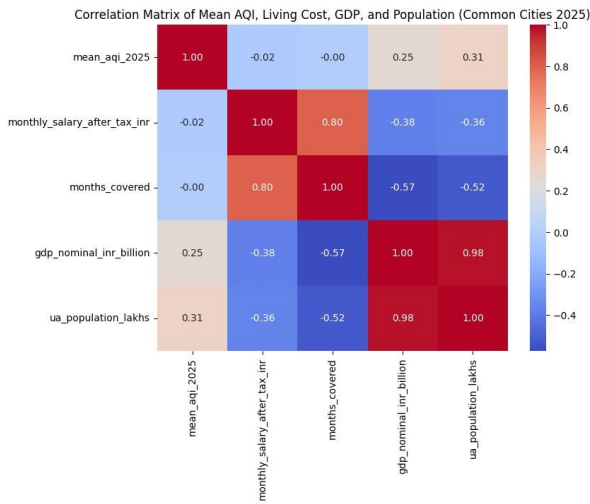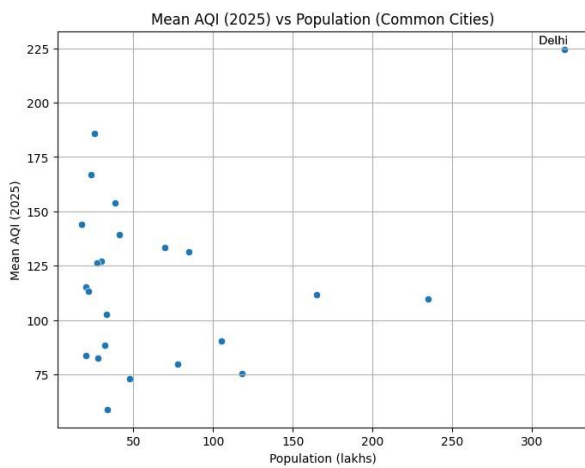
Fig. 7: Air Quality Index (AQI) vs Affordability metrics.

Fig. 8: Air Quality Index (AQI) metrics.

V. DISCUSSION

A. Digital Divide and Livability

Metros lead in ICT and GDP but struggle with affordability—a "digital cost paradox." Smaller cities balance industrialization with affordability, suggesting a transition path toward distributed growth centers.

B. Case Illustrations

Mumbai: GDP-rich but lowest affordability (0.31 months). **Nashik**: Best affordability (1.60 months) due to moderate rent and diverse industry. **Hyderabad**: Technology-driven yet balanced between wages and expenses.

C. Policy Implications

1) Prioritize broadband and ICT investment in Tier-2 hubs.

2) Tie Smart City funds to affordability and liveability improvements.

3) Promote mixed-income housing near digital clusters.

D. Limitations

- Upon further evaluation, the ICT dataset used in this study, although reflective of accurate structural trends, appears to be synthetic or simulated in nature. While its correlations align with verified macro trends from ITU and TRAI, this may introduce representational bias in city-level indicators.
- The selection of the top 30 cities based largely on GDP and population may overlook other determinants such as governance quality, spatial distribution, or regional affordability variations, introducing sampling bias.
- Reliable GDP and affordability data for smaller cities are scarce, and consistent datasets are unlikely to be available before the next national Census (2027), making comprehensive city-level comparisons challenging. • Although the dataset spans 220+ cities, potential overfitting may occur due to uneven data density across city tiers (Tier-1 vs Tier-3), as multiple correlated economic indicators can amplify noise in smaller samples.
- The models are cross-sectional and limited to a single time frame, restricting temporal analysis and trend forecasting.
- Environmental variables like AQI were included only for the top 30 cities, limiting broader environmental correlations; extending AQI coverage to all 220+ cities would improve analysis depth.

E. Future Scope

- Future research should verify or replace synthetic ICT datasets with official or open-source data (e.g., MeitY, TRAI, Smart City Mission) to enhance empirical credibility.
- Expand analysis beyond GDP–population-based city selection to include diverse typologies such as tourism driven, manufacturing, or knowledge-based cities for a more balanced national representation.
- Explore data synthesis using satellite imagery, nighttime light data, and proxy economic indicators to fill data gaps before the 2027 Census.
- Introduce regularization and transfer learning methods to mitigate overfitting and improve model generalization across city classes.
- Extend models to time-series and causal inference frameworks for capturing ICT and affordability dynamics over time.

- Integrate AQI-based analysis across all 220+ cities to explore the intersection of affordability, liveability, and environmental health.
- Develop an interactive geospatial dashboard for policymakers visualizing affordability, GDP, ICT readiness, and environmental metrics together.

VI. CONCLUSION

This study bridges affordability, digital infrastructure, and economic structure through an integrated empirical and predictive lens. While Tier-1 cities dominate innovation and output, affordability advantages persist in Tier-2 hubs. The KNN and Random Forest models demonstrate that affordability patterns can be reliably inferred even for data scarce cities, offering a foundation for equitable urban policy design.

ACKNOWLEDGMENTS

REFERENCES

[1] E. L. Glaeser, *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*, Penguin Press, 2011.

[2] R. Florida, *The Rise of the Creative Class*, Basic Books, 2002.

[3] E. L. Glaeser and J. D. Gottlieb, "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States," *Journal of Economic Literature*, vol. 47, no. 4, pp. 983–1028, 2009.

[4] G. Duranton and D. Puga, "Urban Growth and Its Aggregate Implications," *Handbook of Regional and Urban Economics*, vol. 5, Elsevier, 2020, pp. 547–650.

[5] K. M. Vu, "ICT as a Source of Economic Growth in the Information Age: Empirical Evidence from the 1996–2014 Period," *Telecommunications Policy*, vol. 44, no. 2, pp. 101–118, 2020.

[6] N. Czernich, O. Falck, T. Kretschmer, and L. Woessmann, "Broadband Infrastructure and Economic Growth," *The Economic Journal*, vol. 121, no. 552, pp. 505–532, 2011.

[7] R. Singh and R. Narayan, "Digital Inclusion and Income Inequality in India: Evidence from Panel Data," *Information Economics and Policy*, vol. 64, 2023, Art. no. 101088.

[8] S. Kumar and A. Gupta, "Sectoral Diversification, Urban Productivity, and Household Stability: Evidence from Indian Cities," *Urban Studies*, vol. 58, no. 12, pp. 2491–2510, 2021.

[9] Central Statistics Office (CSO), "City-Level Gross Domestic Product Estimates," Government of India, 2022.

[10] Housing and Urban Development Corporation (HUDCO), "Urban Housing and Affordability in India," Annual Report, 2021.

[11] International Telecommunication Union (ITU), *Measuring Digital Development: Facts and Figures 2023*, Geneva, 2023.

[12] World Bank, *Digital Economy Report: Leveraging Data for Development*, Washington, DC, 2022.

[13] Telecom Regulatory Authority of India (TRAI), "Telecom Services Performance Indicators Report," New Delhi, 2024.

[14] NITI Aayog, *SDG India Index and Dashboard 2022–23*, Government of India, New Delhi, 2022.

[15] Reserve Bank of India, *Handbook of Statistics on Indian States*, 2023.

[16] Ministry of Housing and Urban Affairs (MoHUA), "Smart Cities Mission: Data and Performance Review," Government of India, 2023. [17] Numbeo, *Cost of Living Index*, retrieved 2024. [Online]. Available: https://www.numbeo.com/cost-of-living/

[18] IQAir, *World Air Quality Report*, 2024. [Online]. Available: https://www.iqair.com/world-air-quality

[19] United Nations Human Settlements Programme (UN-Habitat), *City Prosperity Index Methodology*, Nairobi, 2023.

[20] OECD, *Regional Well-being Framework*, Paris, 2021.

[21] NITI Aayog, *Urbanization and Economic Transformation in India*, Discussion Paper, 2023.

[22] Census of India, "Primary Census Abstract, Urban and Rural Classification," 2011. [Online]. Available: https://censusindia.gov.in/

[23] United Nations, *World Urbanization Prospects: India Subset*, 2024 Revision.

[24] Ministry of Electronics and Information Technology (MeitY), "Digital India and Smart Infrastructure Progress Report," 2025.

[25] *Shivanshu Pande."A Data-Driven Survey on Cost of Living, Salary Affordability in Indian Cities", Volume 13, Issue X, International Journal for Research in Applied Science and Engineering Technology (IJRASET) Page No: 1650-1653, ISSN : 2321-9653, www.ijraset.com*

[26] *Shivanshu Pande. (2025). Living Cost Citywise India (Master-Dataset)   [Data   set].   Kaggle.* https://doi.org/10.34740/KAGGLE/ DSV/13830556

[27] *S. H. G., 30 Indian Cities Information Technology Dataset. (2024). Distributed by Kaggle. Accessed: Nov. 7, 2025.* [Online].   Available: https://www.kaggle.com/datasets/sudhanvahg/ 30-indian-cities-information-technology-dataset/data

[28] Udayana, S. K. (2023). *India Air Quality Index (AQI) Dataset [2023–2025]* [Data set]. Kaggle. https://www.kaggle.com/datasets/ saikiranudayana/india-air-quality-index-aqi-dataset-20232025

APPENDIX

TABLE II: Classification of Indian Cities and Towns (Census 2011 and 2025 Projections)

Class	Population Range	No. of Cities (2011)	Estimated (2025)
Class I	100 000	468	600–650
Class II	50 000–99 999	474	520–550
Class III	20 000–49 999	1 889	2 000–2 100
Class IV	10 000–19 999	2 532	2 700–2 800
Class V	5 000–9 999	1 897	1 900–2 000
Class VI	< 5 000	334	350–400

Source: Census of India 2011; NITI Aayog Urbanization Report 2023; UN World Urbanization Prospects 2024 (India subset).

Estimated total (2025): around 9,000–9,200 towns (based on urbanization trend 36% → 41% and a 2% conservative growth estimate for new cities/towns).

Top 30 cities selection criteria:I considered these cities on basis on these factors Ease of Living/Infrastructure Index ,Population and GDP on equal grounds .If suppose there was an approximate tie between two cities on population then GDP was a crucial determinant in rank and hence we get the top 30 cities in India.