E-report on a Demographical Behavior Dataset

Presented by:

Matilla, Gabriel Jeri

Richardson, Xoli Anastasia

July 2025

Introduction

This report explores a dataset of 942 patients (after data cleaning), aiming to summarize key health and demographic characteristics, as well as evaluate possible differences between rural and urban groups. The dataset includes information on BMI, lifestyle behaviors (smoking, drinking, physical activity), education, health literacy, and satisfaction with healthcare services. The goal is to identify general trends and assess whether region of residence significantly affects patient outcomes. All statistical analyses were performed using R.

Methods

Descriptive statistics were used to summarize patient characteristics, focusing on means, medians, standard deviations, and observed ranges for each variable. This provided insight into population health status and potential risk factors.

For inferential analysis, independent samples t-tests were used to compare key variables—Patient Satisfaction, Health Literacy Score, and Education Level—between patients living in rural (coded as 0) and urban (coded as 1) areas. Significance was evaluated at the 0.05 level.

Key Results and Figures

Descriptive Findings:

- Age: Mean age is 52.8 years (range 18–90), indicating the sample largely consists of middle-aged and elderly adults. Reflected in Figure 2.
- BMI: Average of 27.4, suggesting that many individuals fall into overweight or obese categories.
- Physical Activity: Mean of 8.3 hours/week, with a maximum of 16 — most patients are moderately active, though individual variation exists.
- Smoking Status: Mean of 1.04 (on a 0–2 scale); occasional smoking is most common.
- Drinking Status: Mean of 1.08; casual drinking is typical, but some heavy drinkers are present.
- Sex: Nearly balanced (Mean = 0.51), enabling unbiased gender-based comparisons.
- Region: Mean of 0.48 indicates a roughly equal rural–urban split.
- Education: Average level is 1.82 (mostly secondary education); formal schooling is common.

- Health Literacy Score: Mean of 3.29 (on a 1–5 scale), suggesting moderate health understanding among patients.
- Socioeconomic Class: Mean = 2.03 — most individuals are from the middle class.
- Patient Satisfaction Score: Average of 3.07 — indicates moderate satisfaction with healthcare services.
- Height and Weight: Mean height of 154 cm and weight of 64.4 kg, consistent with BMI results.

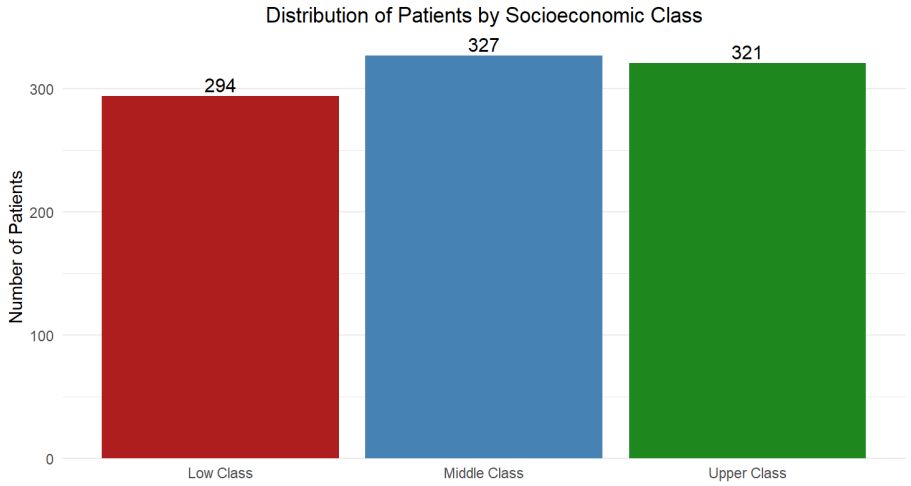

Figure 1. Distribution of patients by socioeconomic class

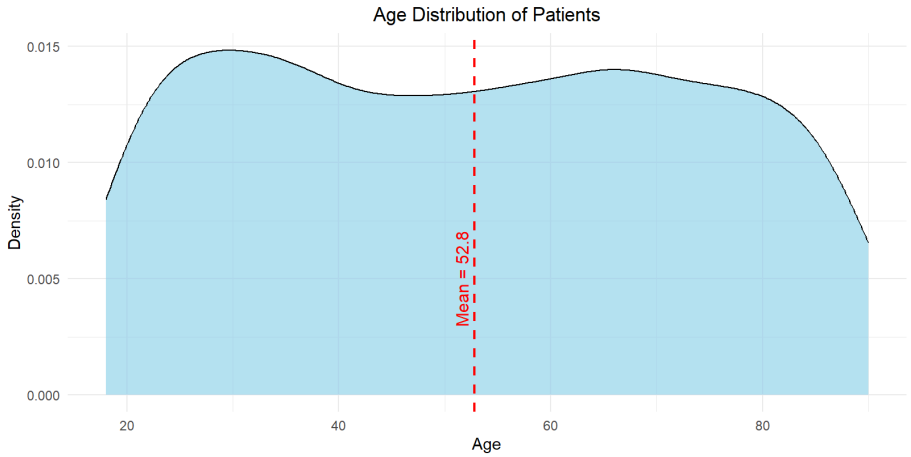

Figure 2. Distribution of age across patients

t-Test Results: Comparing Rural vs Urban Groups

Variable	Rural Mean	Urban Mean	t-value	p-value	95% CI of Difference	Interpretation
Patient Satisfaction	3.15	2.98	1.8185	0.069	[-0.013, 0.354]	Not significant (borderline)
Health Literacy Score	3.30	3.27	0.2878	0.774	[-0.154, 0.207]	No significant difference
Education Level	1.81	1.84	-0.4805	0.631	[-0.159, 0.096]	No significant difference

None of the comparisons between rural and urban groups were statistically significant. However, patient satisfaction showed a borderline result (p = 0.069), with rural patients reporting slightly higher satisfaction on average. These minimal differences can be seen in Figure 3, Figure 4, and Figure 5.

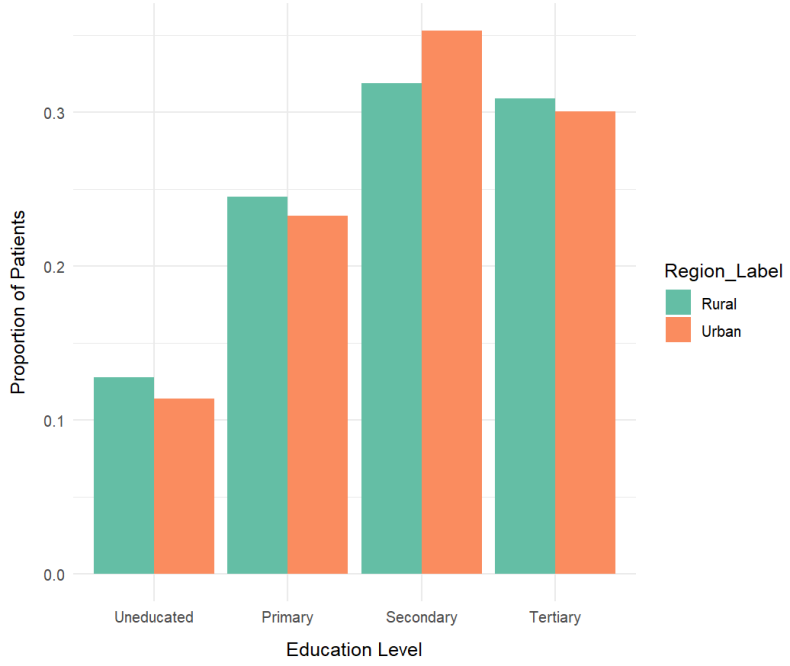

Figure 3. Distribution of patients by education level separated by region.

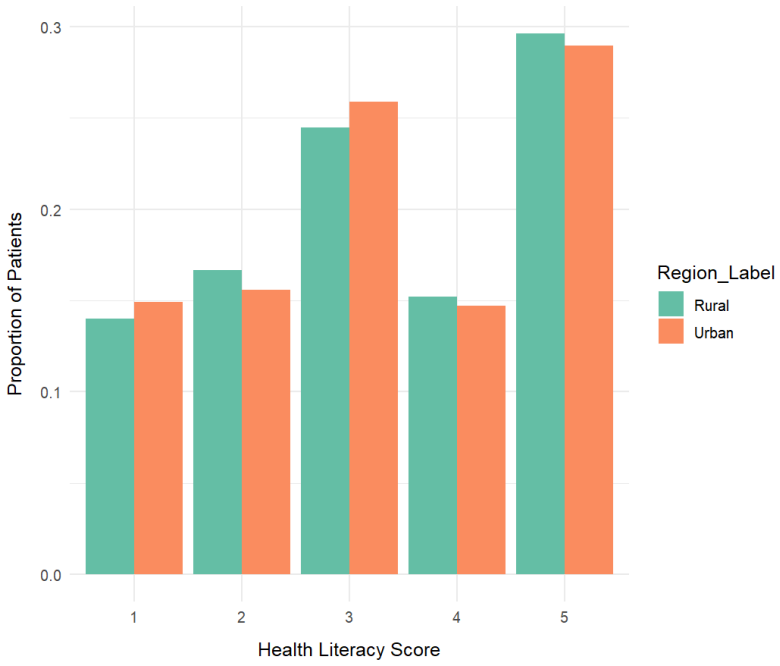

Figure 4. Distribution of patients by health literacy score separated by region

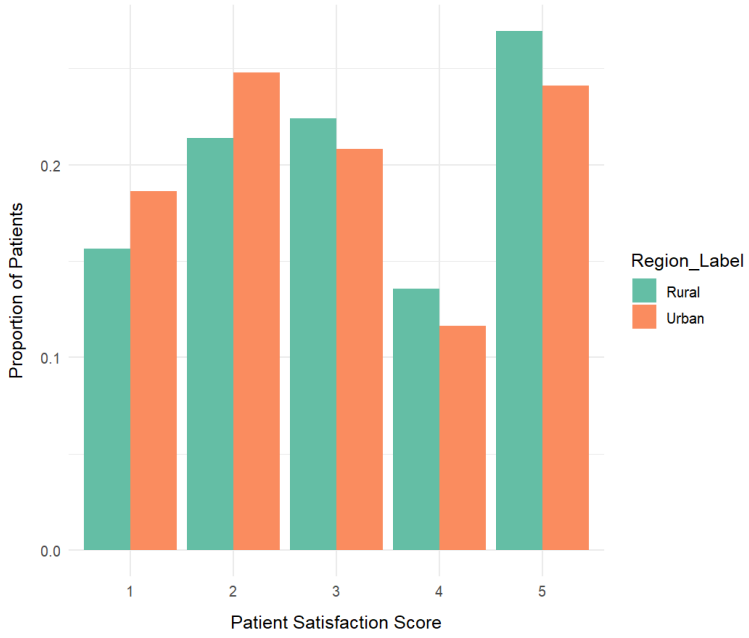

Figure 5. Distribution of patients by satisfaction score separated by region

Interpretation and Conclusion

The dataset reflects a population that is moderately health-literate, predominantly middle class, and balanced across sex and region. However, the high average BMI, occasional smoking, and low variation in education indicate potential areas for public health intervention.

Importantly, no major differences were found between rural and urban groups in terms of satisfaction, health literacy, or education. Still, the slight edge in satisfaction scores among rural patients might warrant further investigation, especially regarding access and expectations.

Overall, these findings help describe a relatively balanced but at-risk population and can support efforts in tailoring education, lifestyle interventions, and healthcare service improvements.