

How Income, Education Level, and Healthcare Access Influence Health Outcomes Among Underserved Communities in the United States

Abstract

Health outcomes are strongly influenced by social and economic factors. This study uses the 2024 Behavioral Risk Factor Surveillance System (BRFSS) dataset to explore how income, education level, and healthcare access relate to health outcomes among underserved communities in the United States. Exploratory data analysis was conducted using Python, including descriptive statistics and visualizations such as histograms, boxplots, and correlation matrices. Findings suggest that socioeconomic factors and healthcare access are associated with differences in self-reported health and health-related behaviors.

Introduction

Background

Health disparities remain a significant public health challenge in the United States. Individuals with lower income, lower educational attainment, and limited healthcare access often experience poorer health outcomes than those with greater resources.

Research Question

How do income, education level, and healthcare access influence health outcomes among underserved communities in the United States?

Dataset Overview

The study uses data from the BRFSS 2024 survey collected by the Centers for Disease Control and Prevention (CDC). The dataset contains 120,293 observations and 301 variables related to demographics, health conditions, healthcare access, and health behaviors.

[0]
✓ 7s

df = pd.read_csv('/content/brfss_survey_data_2024.csv')
df.head()

▼

	_STATE	FMONTH	IDATE	IMONTH	IDAY	IYEAR	DISPCODE	SEQNO	_PSU	CTELENM1	...	_LCSCSTN	_LCSPSTF	DRNKANY6
0	1.0	2.0	b'02282024'	b'02'	b'28'	b'2024'	1100.0	b'2024000001'	2.024000e+09	1.0	...	NaN	9.0	2.0 5
1	1.0	2.0	b'02212024'	b'02'	b'21'	b'2024'	1100.0	b'2024000002'	2.024000e+09	1.0	...	4.0	9.0	2.0 5
2	1.0	2.0	b'02212024'	b'02'	b'21'	b'2024'	1100.0	b'2024000003'	2.024000e+09	1.0	...	4.0	2.0	1.0 1.
3	1.0	2.0	b'02282024'	b'02'	b'28'	b'2024'	1100.0	b'2024000004'	2.024000e+09	1.0	...	NaN	9.0	2.0 5
4	1.0	2.0	b'02212024'	b'02'	b'21'	b'2024'	1100.0	b'2024000005'	2.024000e+09	1.0	...	3.0	9.0	2.0 5

5 rows x 301 columns

[7]
✓ 0s

df.shape

▼

(120293, 301)

[8]
✓ 0s

df.info()

▼

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 120293 entries, 0 to 120292
Columns: 301 entries, _STATE to _AIDTST4
dtypes: float64(296), object(5)
memory usage: 276.2+ MB

Data and Methodology

Data Cleaning

Missing values were identified and reviewed. Numeric variables were selected for analysis, and summary statistics were generated to understand the structure of the dataset.

Descriptive Statistics

[9]

✓ 1s

▼

df.describe()

	_STATE	FMONTH	DISPCODE	_PSU	CTELENM1	PVTRES1	COLGHOUS	STATERE1	CELPHON1	LADULT1	...
count	120293.000000	120293.000000	120293.000000	1.202930e+05	21054.0	21054.000000	4.0	21054.0	21054.0	21054.000000	...
mean	10.967105	6.335963	1120.459212	2.024004e+09	1.0	1.000190	1.0	1.0	2.0	1.001710	...
std	5.454508	3.242917	40.340493	3.151510e+03	0.0	0.013783	0.0	0.0	0.0	0.041316	...
min	1.000000	1.000000	1100.000000	2.024000e+09	1.0	1.000000	1.0	1.0	2.0	1.000000	...
25%	6.000000	4.000000	1100.000000	2.024002e+09	1.0	1.000000	1.0	1.0	2.0	1.000000	...
50%	12.000000	6.000000	1100.000000	2.024004e+09	1.0	1.000000	1.0	1.0	2.0	1.000000	...
75%	17.000000	9.000000	1100.000000	2.024006e+09	1.0	1.000000	1.0	1.0	2.0	1.000000	...
max	19.000000	12.000000	1200.000000	2.024013e+09	1.0	2.000000	1.0	1.0	2.0	2.000000	...

8 rows × 296 columns

Descriptive statistics were generated to summarize the distribution of key variables and identify potential outliers or unusual values.

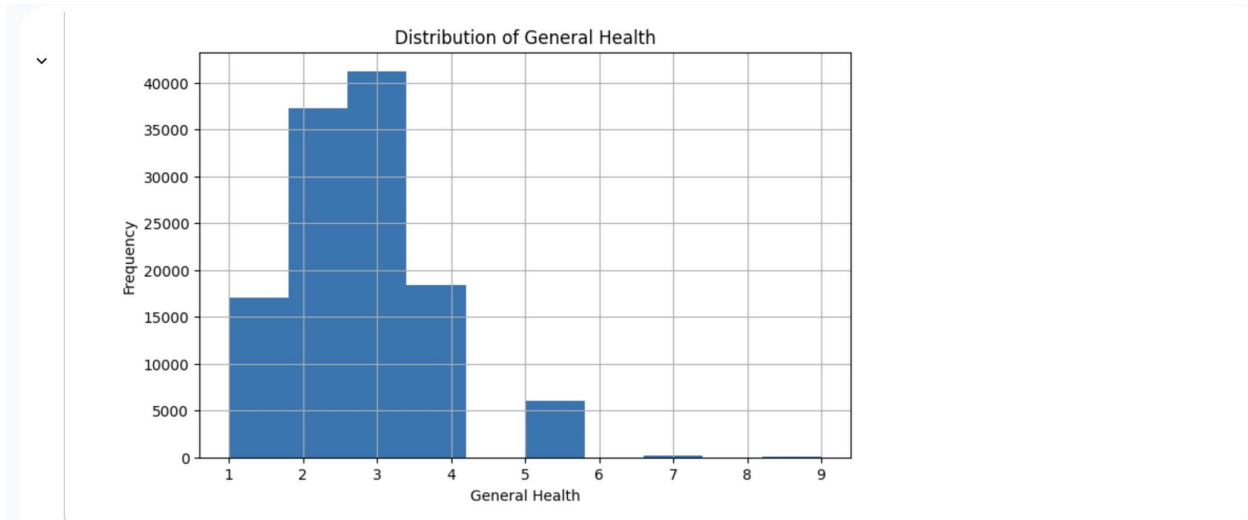
Exploratory Data Analysis

Several visualizations were created:

These visualizations helped identify distributions, outliers, and relationships among variables.

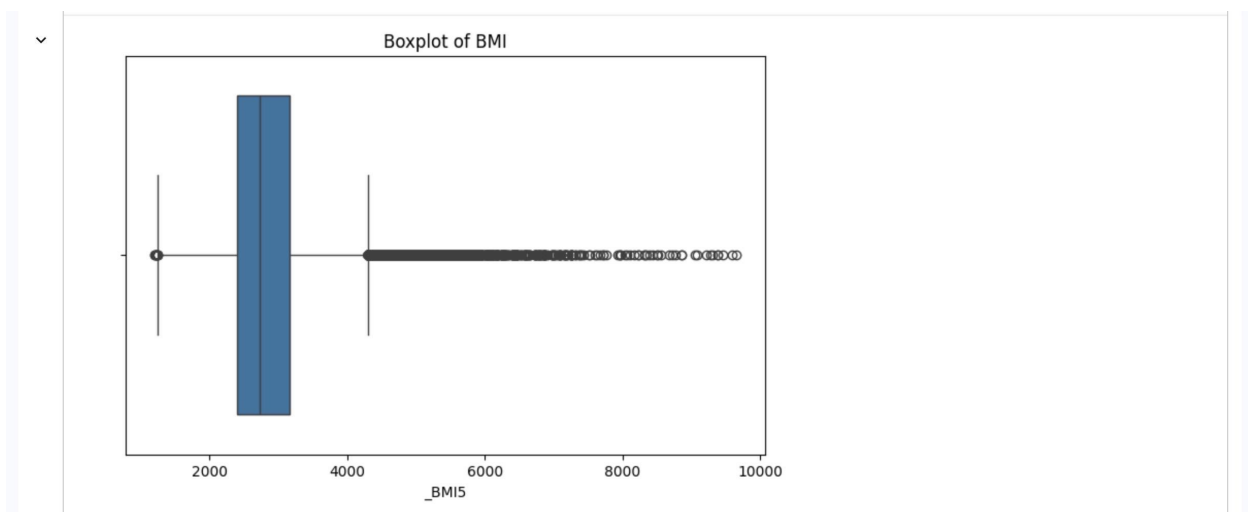
General Health Distribution

- Histogram of General Health



The histogram showed that most respondents reported fair to good health, while fewer individuals reported very poor health.

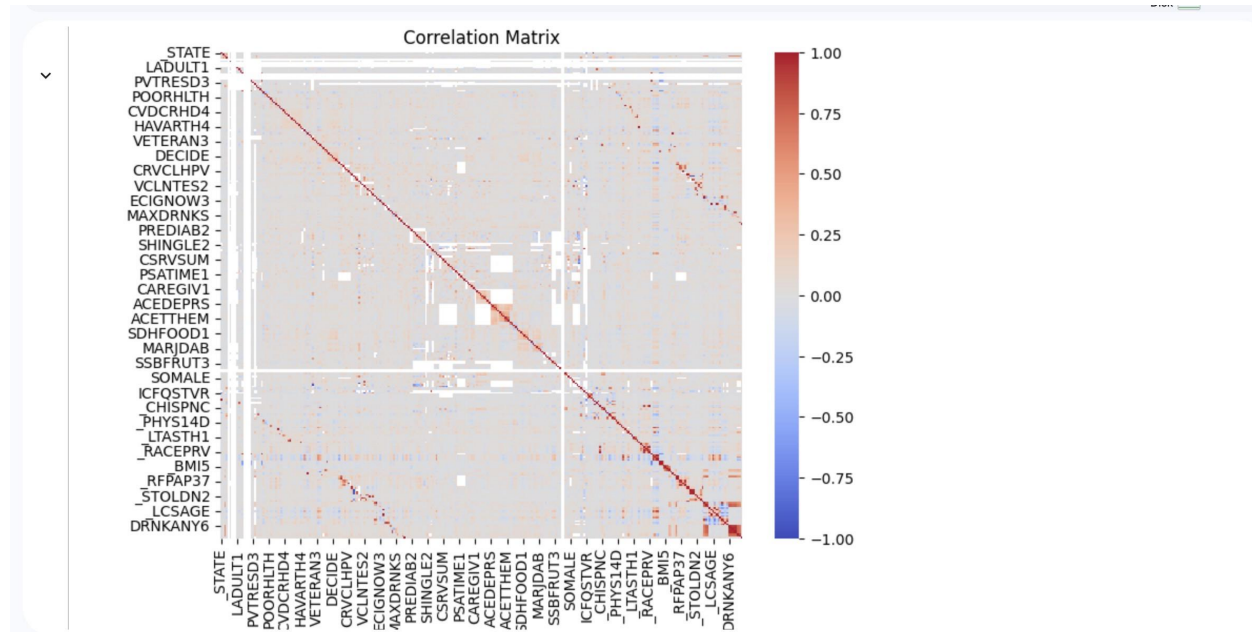
BMI Analysis



The BMI boxplot revealed several outliers and a right-skewed distribution, suggesting variation in body weight among respondents.

Correlation Analysis

The correlation matrix indicated relationships among several health and demographic variables. Most correlations were weak to moderate, suggesting that health outcomes are influenced by multiple interacting factors.



Healthcare Access and Socioeconomic Factors

Variables related to income, education, and healthcare access showed patterns associated with differences in self-reported health outcomes, supporting existing public health research on social determinants of health.

Conclusion

This analysis demonstrates that income, education level, and healthcare access play important roles in shaping health outcomes among underserved populations. Understanding these relationships can help policymakers and healthcare providers develop targeted interventions to reduce health disparities and improve community health.