

# Massachusetts ZIP Code Desirability Analysis

## Abstract

This project conducts a comprehensive analysis of 493 Massachusetts ZIP Code Tabulation Areas (ZCTAs) to evaluate community desirability using a weighted composite scoring model. The analysis incorporates demographic, economic, educational, and housing variables and applies min-max normalization, weighted aggregation, clustering, correlation analysis, and geospatial visualization. Results indicate strong disparities across ZIP codes, driven primarily by income, education, and employment factors. ZIP codes in the Boston metropolitan area outperform rural and inner-city regions. An interactive dashboard and choropleth map are developed to visualize these differences.

## 1. Introduction

Massachusetts exhibits substantial socio-economic variation across its ZIP-code regions. This project evaluates community desirability using a multi-factor approach that integrates five key indicators:

- Median household income
- Educational attainment
- Unemployment rate
- Housing rent burden
- Percentage of seniors

The study aims to produce a normalized, interpretable desirability index and perform regional, statistical, and cluster-based evaluations.

## 2. Data Inputs and Variables

The analysis dataset includes 493 ZIP codes and the following normalized variables:

Variable	Description
Median_income	Household income in past 12 months
Pct_bachelors	Percent of adults with Bachelor's degree
Unemployment	% unemployed labor force
Rent_burden	% renter households paying $\geq 35\%$ income
Pct_seniors	% aged 65+
Population	Total population

After cleaning and normalization, the dataset supports both statistical and spatial analysis.

## 3. Methodology

### 3.1 Normalization

Variables were normalized using min–max scaling.

Positive-impact factors (income, education): higher = better.

Negative-impact factors (unemployment, rent burden, seniors): lower = better.

### 3.2 Weighting Model

Dimension	Weight
Income	25%
Education	25%
Employment (inverse unemployment)	20%
Housing burden (inverse)	20%
Seniors (inverse)	10%

### 3.3 Composite Score

A ZIP code's overall desirability score ranges from **25.8 to 86.3**, with a **mean of 56.45**.

## 4. Statistical Analysis

### 4.1 Descriptive Statistics (Actual Values)

Metric	Value
Total ZIP Codes	<b>493</b>
Mean Overall Score	<b>56.45</b>
Median Score	<b>55.91</b>
Standard Deviation	<b>11.03</b>
Score Range	<b>25.77 – 86.27</b>

Income, education, and rent burden show high variance; employment is the least variable dimension.

## 4.2 Correlation Matrix (Actual Values)

Variable	Correlation with Overall Score
Median Income	<b>+0.862</b>
Education Level	<b>+0.833</b>
Unemployment (inverse)	<b>−0.533</b>
Rent Burden (inverse)	<b>−0.434</b>
Seniors (inverse)	<b>−0.103</b>

### Interpretation:

Community desirability is strongly predicted by income and education, moderately affected by employment and housing burden, and weakly influenced by age structure.

## 4.3 Score Distribution

- **Q1 (25th percentile):** 48.93
- **Q2 (Median):** 55.91
- **Q3 (75th percentile):** 63.33
- **IQR:** 14.40
- **Skewness:** +0.185 (slightly right-skewed)
- **Kurtosis:** −0.198 (light-tailed distribution)

Scores cluster around midrange values, with fewer extremely high or low performers.

# 5. Ranking of ZIP Codes

## 5.1 Top 5 Highest-Scoring ZIP Codes (Actual Values)

ZIP	Score	Median Income	Education %	Unemployment %
<b>01745</b>	86.27	180,833	80.0	0.0
<b>01770</b>	85.38	247,500	81.2	6.3
<b>02071</b>	85.16	207,878	75.0	4.3
<b>01740</b>	82.99	198,475	80.8	5.8
<b>02482</b>	82.80	240,069	83.5	4.0

These ZIP codes represent affluent, highly educated suburban communities.

## 5.2 Bottom 5 Lowest-Scoring ZIP Codes

ZIP	Score	Median Income	Education %	Unemployment %
01105	25.77	22,994	8.3	14.8
01022	29.60	56,307	28.6	11.1
02558	30.69	58,594	16.2	16.3
01107	31.03	31,969	14.4	13.4
01074	31.43	76,859	3.6	18.5

Common issues include high unemployment, extremely low education levels, and low median income.

## 6. Cluster Analysis (K-Means, Actual Results)

Four clusters emerged:

Cluster	ZIPs	Avg Score	Median Income	Avg Education	Avg Unemployment
High Performing	110	71.4	165,958	72.1%	3.6%
Moderate-High	209	56.9	107,674	42.7%	3.8%
Moderate-Low	58	50.3	93,327	52.1%	4.9%
Needs Improvement	116	44.6	70,582	30.9%	7.8%

### Insight:

The “Needs Improvement” cluster represents economically strained communities with high unemployment and low educational attainment.

## 7. Regional Analysis by ZIP Prefix

Region	Avg Score	Avg Income	Avg Education
01	55.22	106,475	43.8%
02	58.09	117,874	52.7%

Prefix “02” (Greater Boston and eastern MA) consistently outperforms prefix “01.”

## 8. Income Quintile Analysis

Income Group	Avg Score	Avg Education	Avg Unemployment	Avg Rent Burden
Lowest 20%	44.32	32.43%	7.02%	43.75%
Low 20%	50.94	38.52%	4.94%	39.00%
Middle 20%	55.15	45.70%	4.28%	41.24%
High 20%	61.27	53.31%	4.00%	35.61%
Highest 20%	70.54	67.93%	3.86%	37.50%

**Income is the strongest stratifier of community desirability.**

## 9. Geographic Analysis

Interactive choropleth maps clearly show:

- High-performing areas cluster around **Boston suburbs**.
- Lower-performing ZIP codes concentrate in **Springfield, New Bedford, and rural central/western MA**.
- Housing affordability and unemployment strongly influence geographic disparities.

## 10. Conclusion

This study demonstrates significant socio-economic inequality across Massachusetts ZIP codes. Income, educational attainment, and employment serve as the strongest predictors of community desirability. The composite scoring model, statistical insights, clustering structure, and map visualizations together provide a comprehensive tool for understanding regional disparities.