Spearman’s Rank Correlation Test – Boston Housing Dataset

College: Cornerstone International Community College of Canada - CICCC

Student: Amir Lima Oliveira

Due Date: May 26rd, 2025

Dataset Information

Dataset Title: Boston House Prices – Advanced Regression Techniques

Source: Kaggle

Link: https://www.kaggle.com/datasets/fedesoriano/the-boston-houseprice-data

Description: The Boston house-price data of Harrison, D. and Rubinfeld, D.L. 'Hedonic prices and the demand for clean air', J. Environ. Economics & Management, vol.5, 81-102, 1978.

Objective

Analyze the relationship between socioeconomic factors and housing values using Spearman’s Rank Correlation Test.

Feature Pairs Selection for Analysis

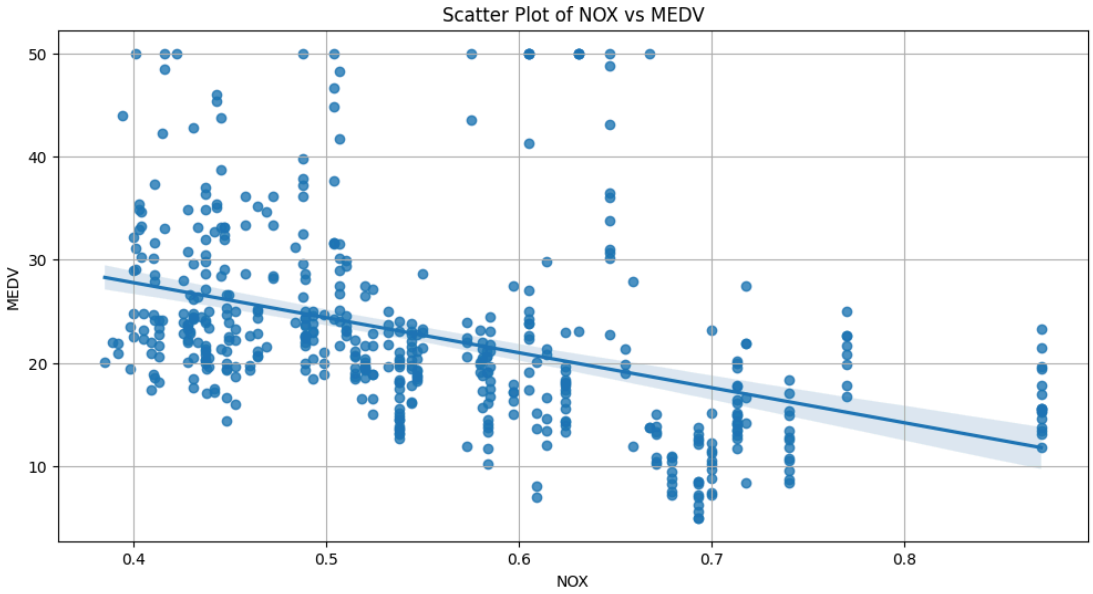
Variable Pairs to Analyze:

- 1. NOX & MEDV – Does pollution relate to a decrease in housing value?
- 2. NOX & INDUS – Is pollution associated with industrial land use?
- 3. NOX & LSTAT – Is pollution more concentrated in lower-status populations?
- 4. NOX & CRIM – Could pollution correlate with higher crime rates in degraded areas?

Spearman's Rank Correlation Results

NOX & MEDV

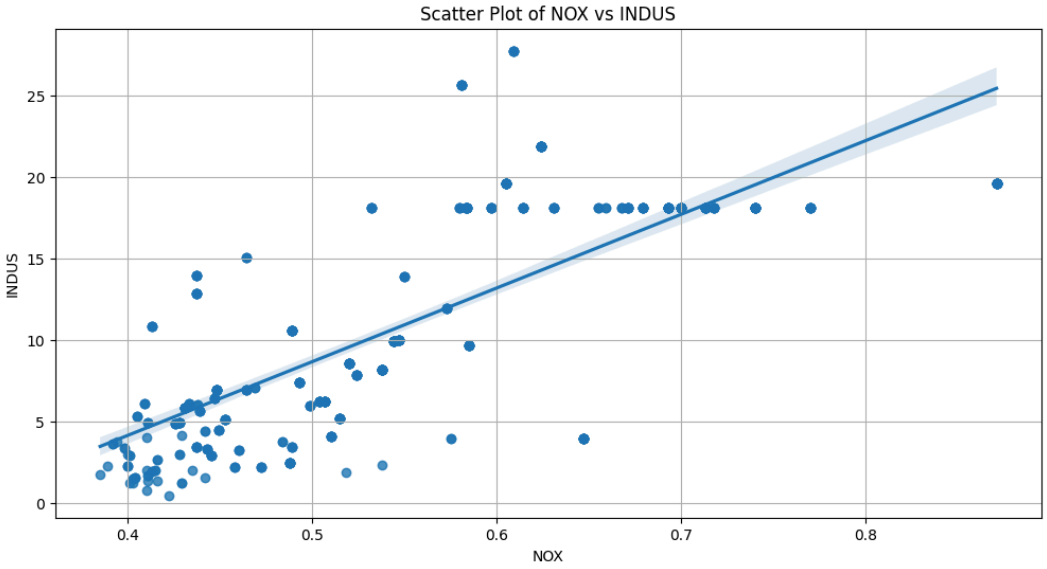
$\rho = -0.5626$, $p = 0.0$



Moderate negative correlation; higher pollution is associated with lower housing prices.

NOX & INDUS

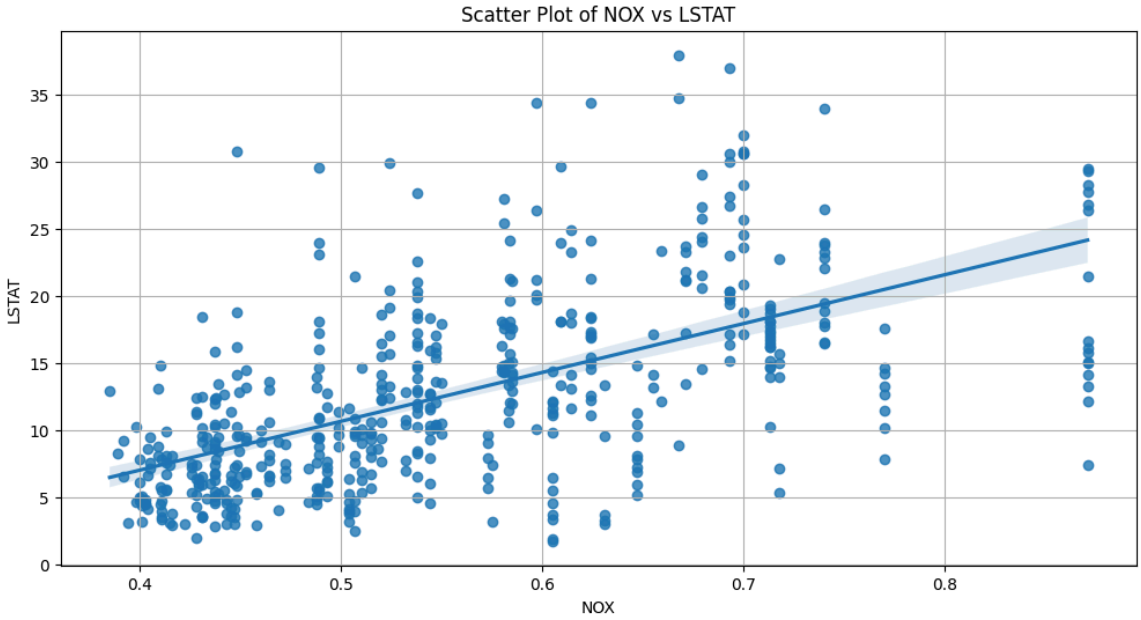
$\rho = 0.7912$, $p = 0.0$



Strong positive correlation; industrial land use relates to increased pollution.

NOX & LSTAT

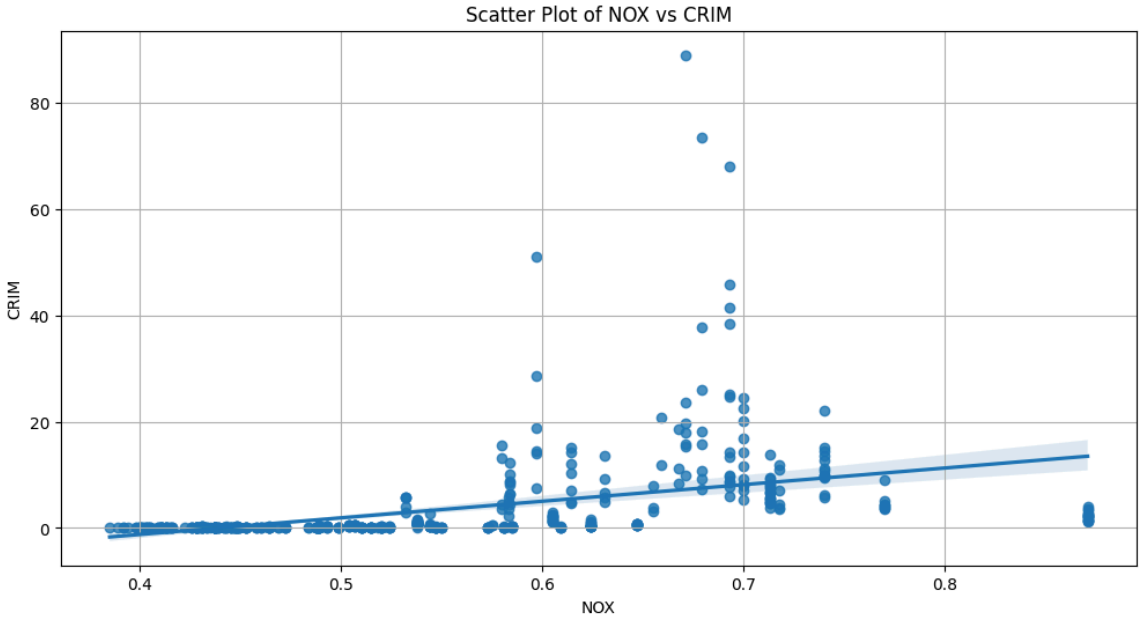
$\rho = 0.6368$, $p = 0.0$



Moderate positive correlation; pollution is more present in low-status areas.

NOX & CRIM

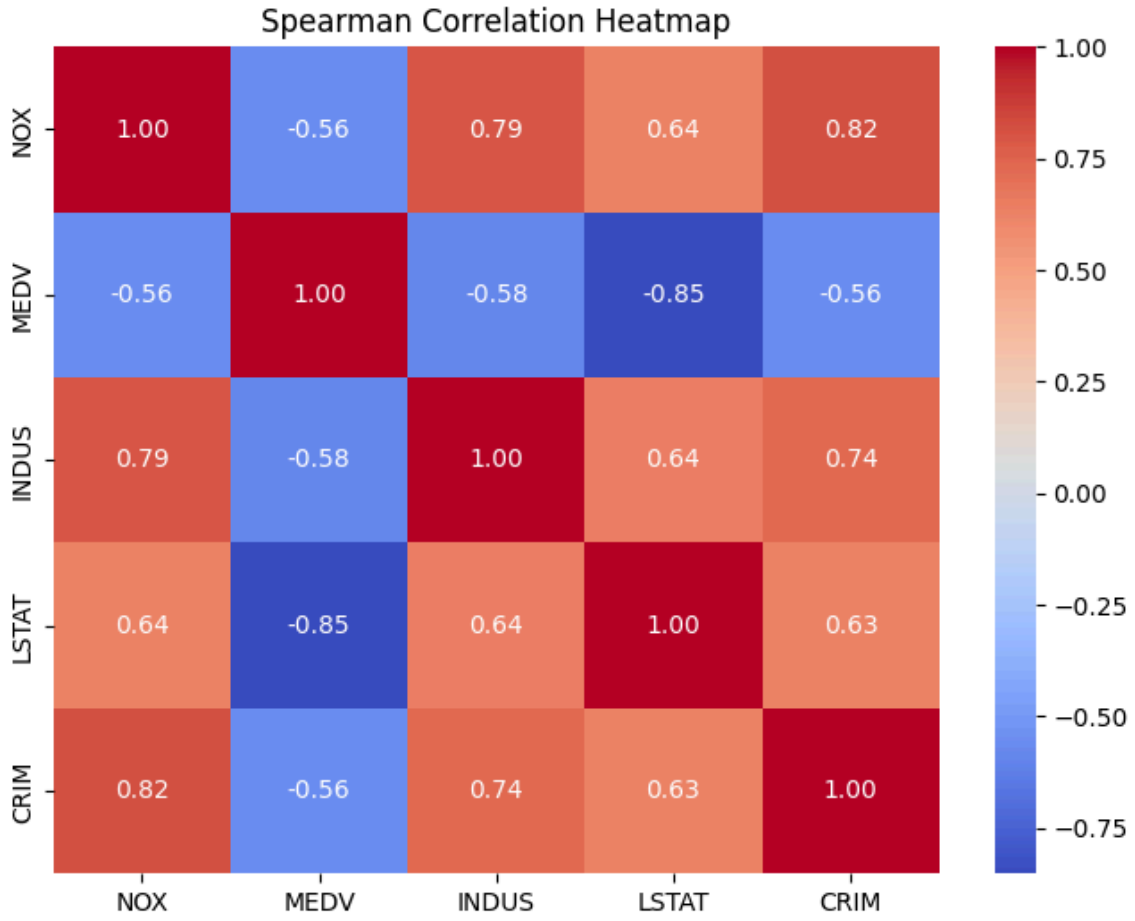
$\rho = 0.8215$, $p = 0.0$



Strong positive correlation; more pollution correlates with higher crime rates.

Spearman Correlation Heatmap

The heatmap visually summarizes the Spearman correlation coefficients between the selected variables.



Conclusions and Environmental Interpretation

This analysis reveals that pollution (NOX) is not just an environmental metric — it may also reflect socioeconomic disparities and signs of urban decay. The patterns show that areas with higher NOX levels tend to have lower housing values, higher industrial land use, more marginalized populations, and higher crime rates.

These findings highlight the importance of integrating environmental quality into housing policy, urban planning, and social equity initiatives. By framing this project through the lens of environmental data science, we emphasize how environmental conditions can be quantitatively linked to urban well-being.

Looking Ahead: Expanding the Analysis

Future studies could include:

- • Urban parks (count or total area in km^2)
- • Green space coverage (km^2)
- • Proximity to urban rivers (distance from housing zones)
- • City industry density or proximity
- • Cement or asphalt land cover rate (km^2)

Dataset Credits

This analysis used the Boston House Prices – Advanced Regression Techniques dataset

from Kaggle, shared by Federico Soriano.

- Link: https://www.kaggle.com/datasets/fedesoriano/the-boston-houseprice-data

Original citation:

Harrison, D., & Rubinfeld, D.L. (1978). Hedonic prices and the demand for clean air.

Journal of Environmental Economics and Management, 5(1), 81–102.