Housing Markets and Air Quality Across Countries

Osman Şah

DSA210 – Introduction to Data Science
Fall 2025–2026

1 Introduction

Environmental conditions play an important role in shaping quality of life. Air quality, in particular, affects health outcomes and daily living conditions. This project studies whether differences in air quality across countries are reflected in housing market indicators such as rent levels, affordability, and population growth.

2 Research Objective

The main objective is to test whether housing market characteristics differ systematically between clean-air and polluted countries, and to identify which housing indicators are most strongly associated with air quality.

3 Data Description

Two country-level datasets are used:

- A global housing market dataset (2015–2024)
- A world air quality dataset (2024)

After standardizing country names, the datasets are merged into a single table containing housing indicators and air quality categories.

4 Exploratory Observations

Initial exploration shows that air quality categories are relatively balanced. Scatter plots and group summaries suggest that countries with cleaner air often differ in rent levels, population growth, and affordability measures.

5 Statistical Evidence

To formally test whether clean-air and polluted countries differ, I apply Welch's two-sample t-test.

5.1 Hypotheses

- H_0: There is no difference between clean-air and polluted countries.
- H_1: There is a difference between the two groups.

5.2 Test Result

The test yields:

$$t \approx -23.16, \quad p \approx 9.28 \times 10^{-118}$$

Since the p-value is far below 0.05, I reject the null hypothesis and conclude that the difference is statistically significant.

6 Predictive Modeling

I train several classification models to predict air quality category using housing indicators:

- Logistic Regression
- Decision Tree
- Random Forest
- XGBoost

Model accuracy ranges from approximately 52% to 53%, with tree-based models performing slightly better.

7 Feature Importance

Feature importance analysis reveals that:

- Population Growth (%) is the strongest predictor
- Rent Index and Affordability Ratio also contribute substantially

Variables such as GDP Growth, Inflation Rate, and Mortgage Rate have very low importance and can be removed without reducing performance.

8 Discussion

The results suggest that housing market dynamics reflect environmental conditions at the country level. In particular, demographic pressure and rental markets appear closely linked to air quality differences.

9 Limitations

This analysis is observational and does not establish causality. Additionally, country-level aggregation may hide regional variation.

10 Conclusion

Overall, the findings support the idea that air quality differences across countries are associated with housing market characteristics. Population growth emerges as the most important factor connecting the two.

11 Ethics and AI Use

Only public, aggregated datasets are used. No personal data is involved. AI tools were used for guidance and structuring, while analysis and conclusions were produced by the student.