

Executive summary: The impact of urban features on health outcomes

Authors: Nina Adhikari, Juliana Londono Alvarez

Problem statement: This project aimed to identify the urban features that most influence the amount of free time people spend on physical activity. It contributes to understanding how urban planning and design can impact a community's physical and mental health.

Stakeholders: The findings would serve designers involved in urban planning, as well as local governments, health and environmental agencies, residents and local businesses. We want to create a cost-benefit analysis tool to help city planners identify design interventions with the largest potential impact on health outcomes in their particular neighborhood.

Key performance indicators: The project prioritizes explainability by assessing feature importance; we use R-squared and adjusted R-squared scores to measure performance of our models.

Data: We collected data specifically for the District of Columbia using three different datasets:

- From [500 Cities: Local Data for Better Health](#), we obtained the health outcome variable: the percentage of adults aged ≥ 18 years that did not participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise, during the past month, outside their regular job (No Leisure-Time Physical Activity, LPA).
- From [Smart Location Database](#), we obtained the urban features: more than 90 different indicators associated with the built environment and location efficiency. Indicators include density of development, diversity of land use, street network design, and accessibility to destinations as well as various demographic and employment statistics.
- From [2017 Census data](#), we obtained median household income in the past 12 months (in 2022 inflation-adjusted dollars).

Methods:

For the data cleaning, the first step involved ensuring temporal consistency of the data. We selected datasets from the years 2017 and 2018 to maintain coherence in our analysis. We also removed incomplete census tracts (totaling 3 out of 450) from the dataset. Finally, we had to carefully manage the data to ensure coherence and comparability across different geographic units.

To begin exploring the most influential features, we first identified highly correlated features and removed those from our dataset to mitigate multicollinearity issues. Then, we employed iterative feature selection techniques, including lasso regression and manual selection, to reduce the feature space to the top 25 most influential variables. With these, we did a linear regression analysis, which revealed high correlations with wage-related data and potential nonlinear relationships, prompting further investigation into the role of wage-related features in predicting LPA. To address this, we ran a nonlinear regression using only wage-related features to fit LPA. We then used the urban features not directly related to wage or wealth to do a linear regression on the residuals of the wage-only nonlinear regression. The coefficients of that regression were the result of our analysis, as detailed below.

Results and implications: The study found that income and other **wage-related data** were the best predictors of leisure time physical activity. However, after accounting for these relationships, some urban markers (destination accessibility by auto & transit, frequency of transit service, and residential density) exhibited some correlation with LPA. While the correlations were weak, they suggest that those urban features may impact leisure physical activity levels. However, it's important to note that the study only establishes correlation, not causation.

Future directions: We aim to expand the analysis and model to encompass other cities and include some more advanced learning models like clustering to develop a neighborhood classification that can help predict LPA levels by neighborhood. It would also be interesting to integrate additional datasets, such as race demographics, park access, and healthcare resources, to make our analysis more robust. Lastly, we plan to conduct cost-benefit analyses per Census Block Group (CBG) to facilitate informed decision-making in urban planning and policy formulation.