Prediction of the Walkability Index of US Neighborhoods
Level of Interest: 7/10
Cody Daniels

Task Description
The United States Environmental Protection Agency (EPA) occasionally publishes a Walkability Index score for neighborhoods across the country. According to the EPA, a walkable area means it is easy to walk around. Walkability is important because it encourages an active lifestyle, which leads to better holistic health outcomes. Walkability also means that there are less cars on the road, which is good for environmental health too—hence the EPA is concerned with monitoring the metric. Other benefits of walkability include reduced infrastructure costs, a sense of community, and safer streets.

The aim of this project is to see which demographic features correlate with highly walkable neighborhoods. We already know how the index is calculated, but what can we say about the people in those areas? For instance, are highly paid individuals more privileged to live in walkable areas? Does certain geography lend itself to walkability? Living in a walkable neighborhood should be accessible to everyone, and these conclusions can provide insight into how to make that happen.

Data Features
Below is an example of some of the features that might be useful in this task. I would withhold the principal components in the calculation of the index in order to focus on which other demographic factors correlate with high walkability.

Population	Land area	Water area	Housing units	% HU with no vehicle	Intersection density	Employers	% low wage workers	Target: Walkability
1202	73.5950285	0	460	0.163120567	115.9817473	66	0.240291262	14

The “Employers” feature could be broken down by industry, such as retail, office, entertainment, service, and industrial. Along with “% HU with no vehicle,” there is also data for one and two or more vehicles. There are plenty of urban planning metrics similar to “Intersection density” to take a look at.

Gathering the Data
The EPA published a CSV with 220,740 census areas along with basic metadata and their walkability scores. Each row in the EPA’s dataset has identifying information that can be used to join it with additional data from the United States Census Bureau. Both of these datasets are quite large, so it will take some effort to line them up correctly and select relevant features. The last Walkability Index scores were published in January 2021. The last substantial Census data release was in 2020, although some less detailed data is published annually.