

The Battle of Neighborhoods

Brendan Hargrove

Introduction & Problem Statement

Many people are forced to relocate for their careers and/or choose to move to a new city for better opportunities. However, most still have a home neighborhood they are fond of, or a favorite neighborhood from their hometown. This notebook will attempt to characterize the venues of a chosen neighborhood from one city and compare it to all neighborhoods of a different city, returning the neighborhoods with similar businesses, restaurants, etc. In this way, a user could implement this notebook to match a known neighborhood in one city with similar neighborhoods in a different city, thereby creating a “short list” of potential neighborhoods to investigate for housing, demographics, etc. However, the housing and demographic data are not within the intended scope of this notebook.

Beyond the personal use of this notebook, finding a neighborhood in a new city that is most similar to a favorite neighborhood back home, this workflow has business applications as well.

Say a company is looking to expand into a new city. By characterizing the area around an already high-performing location, and comparing that characterization against all neighborhoods in the new city, the business might be able to predict which new locations would be most likely to perform well. Similarly, by characterizing very low-performing locations and utilizing the same workflow, it might be possible to predict which neighborhoods should be avoided.

Description of Data

The data required for this project includes: neighborhood names for Houston, TX and Seattle, WA, location information (latitude-longitude) for each neighborhood/city center, and a list of venues and venue categories. All of these data were available through public websites and/or online tools/repositories and required no previous cleaning, organizing, or processing.

Neighborhood names for Houston were obtained from https://en.wikipedia.org/wiki/List_of_Houston_neighborhoods.

Neighborhood names for Seattle were obtained from https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Seattle.

Location data for each neighborhood was gathered by iteratively feeding each neighborhood name into GeoPy Geocoder (<https://geopy.readthedocs.io/en/stable/>).

Once neighborhood locations were defined, each lat-long point was queried in Foursquare (<https://developer.foursquare.com/>) for venue information.

In addition, the following python libraries were used in data import, processing, analysis, and display: Pandas, Numpy, urllib, BeautifulSoup, GeoPy, json, Matplotlib, SKLearn, Folium.