Capstone Project – The Battle of Neighborhoods

Introduction:
A friend of mine is relocating to San Francisco, CA. He currently lives in Philadelphia, PA in an apartment and uses public transportation to work. He goes to gym daily and loves Thai cuisine. He frequently visits movie theaters and would like to live in an area which is similar to his life style.

The aim of this project is to identify similar locations between San Francisco and Philadelphia, classify these areas based on the accessibility, public transportation, type of restaurants etc.

Methodology:

1) We will identify the neighborhoods in Philadelphia , cluster them and mark the neighborhood where my friend currently lives.

2) We will identify the neighborhoods in San Francisco based on similar parameters such as
    a. Distance of public transportation
    b. Types of restaurants – Specifically Thai Cuisine
    c. Movie theaters
    d. Gym

3) Once we cluster and segment the neighborhoods we will make the recommendation

Data:

1) Neighborhood and zip code data will be scrapped from Wikipedia pages and other sources–
    https://en.wikipedia.org/wiki/List_of_Philadelphia_neighborhoods
    https://en.wikipedia.org/wiki/List_of_neighborhoods_in_San_Francisco
    https://spaceandcompany.com/philadelphia-zip-codes/
    https://www.zillow.com/browse/homes/pa/philadelphia-county/
    http://www.healthysf.org/bdi/outcomes/zipmap.htm

2) We will utilize google geocoder or similar method to get latitude and longitude from the neighborhood addresses.

3) We will use Foursquare API to pull nearby venues based on latitude and longitude

4) Once we have nearby venues we will use K-Means clustering (as in Labs) from the Scikit-learn library to cluster and identify/segment similar neighborhoods.