Where to open a brand new CrossFit box in Denver, CO

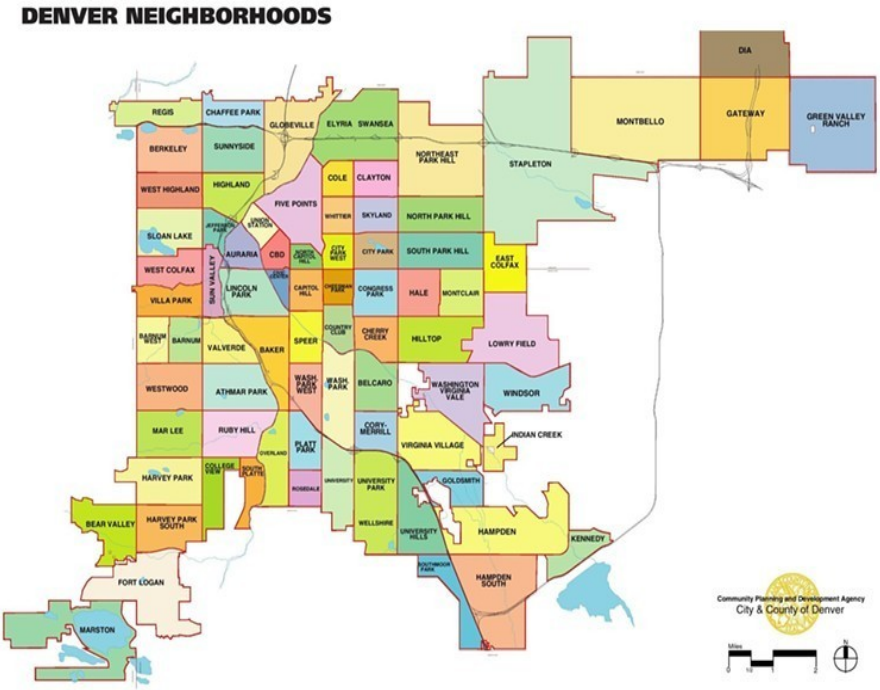
Goal and Stakeholder

The goal of this project is to identify the best suitable neighborhood where to open a new Crossfit Box in the city of Denver, Colorado. Crossfit is a rather new sport involving movements from different disciplines like gymnastics, running, rowing and weightlifting executed either individually or in group classes with motivational music background. This new activity will comprise a large area dedicated to sport (weights, machines and racks, hanging bars, rings, small track for running and floor mats for bodyweight gymnastic), changing rooms/showers plus a recreation and cafe area. Nice to have then a small outdoor areas for summer workouts.

Basic requirements to achieve are:

- the box should not be close to another similar activity
- the structure should preferably be located in a low crime area
- the selected neighborhood shall present a large young/middle age population, between18 and 65 (although Crossfit could be practiced also by younger and older people and the box owner is encourage to work to involve those people).

Possible stakeholder for this task would be a business owner or a fitness company who wants to to get into the CrossFit world in one of the most beautiful cities of US.



Project Data

The data to be collected for this analysis come from the several public datasets available on the web (official source www.denvergov.org).

In particular we will make use of:

- Neighborhood list – to have immediately available in a tabularar way the complete Denver neighborhood list and the correct naming convention. From this CSV file we will use the NBHD_NAME column as the baseline for the Denver Neighboorhood Dataframe.

- 2015-2020 Denver Crimes file list – to find out to the top10 Neighborhoods for Crimes and therefore exclude them from the final selection. This huge CSV needs a bit of elaboration and cleaning in order to remove the car accidents keeping only crimes and then grouping all the remaing entries by NEIGHBORHOOD_ID (i.e. Neighborhood name).

- 2010 Census Demographic Data – to identify the Neighborhoods with the majority of the population in the target age range. In this case data are already aggregated by NBHD_ID (i.e. Neighborhood name). The column PCT_LESS_18 and PCT_65_PLUS will be used to derive the people percentage in the range 18-65 for each NBHD_ID.

To quickly fill up the Neighboorhood dataframe with geographical data (latitude and longitude values) Python GEOPY package will be imported and called within a custom function.

For venue categorization analysis and neighborhood clustering, Foursquare data shall be used via avalable API.