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

Coursera Applied Data Science Certification – Capstone Project [12/19/2020]

Filippo Gronchi

The goal of this project is to identify the best neighborhood where to open a new Crossfit Box in the city of Denver, Colorado. Crossfit is a pretty new sport involving movements from different disciplines like gymnastics, running, rowing and weightlifting executed either individually or in group classes with motivational music background.

Requirements
- the facility 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 i.e. between18 and 65

DENVER NEIGHBORHOODS

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

Technical background and Datasets

The data analysis has been completed using *Python 3.8.5* on a dedicated *Jupyter Notebook*. This file along all the images and attached documentations have been placed under revision using *GitHub* in the repository named 'Capstone_Project' *(https://github.com/f1li/Coursera_Capstone)*.
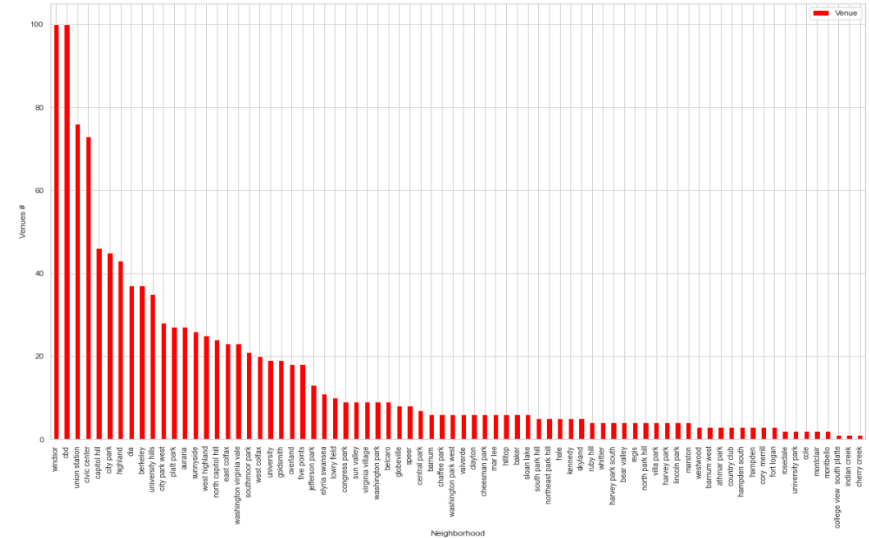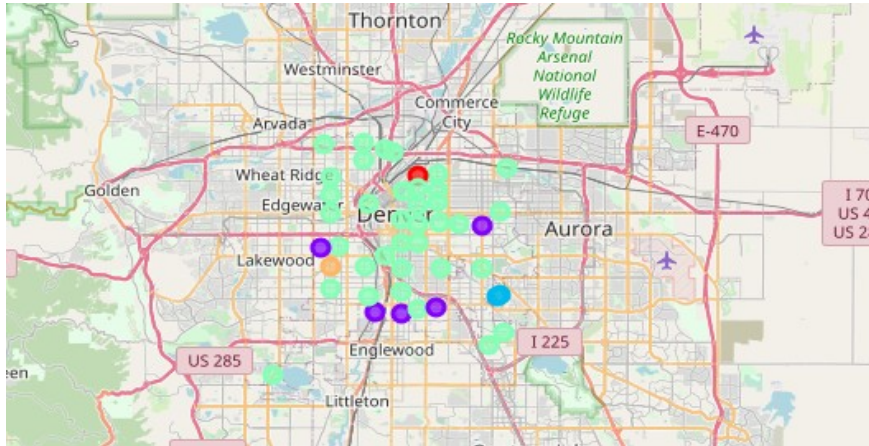
official source **www.denvergov.org**
- Neighborhood list – to have immediately available in a tabularar way the complete Denver neighborhood list and the correct naming convention.

- 2015-2020 Denver Crimes file list – to find out the top 10 Neighborhoods for Crimes and therefore exclude them from the final selection.

- 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).

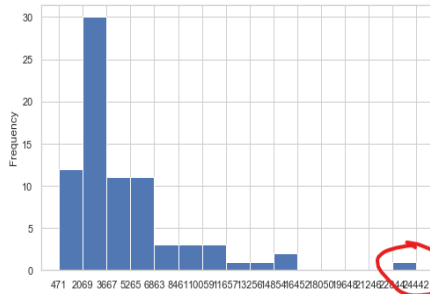# Venues analysis: Foursquare and Machine Learning

Using Foursquare API calls I built a full database of Denver Venues grouped by Neighborhood, excluded all the one having already at least one sport related venue. Using Machine Learning K-Means algorythm 5 different clusters have been identified.
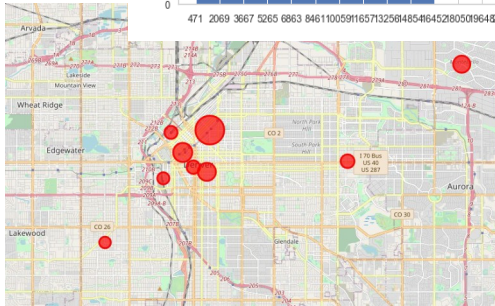

Purple and Green clusters are the one to be considered in our final selection (the other clusters are too small to be significant).
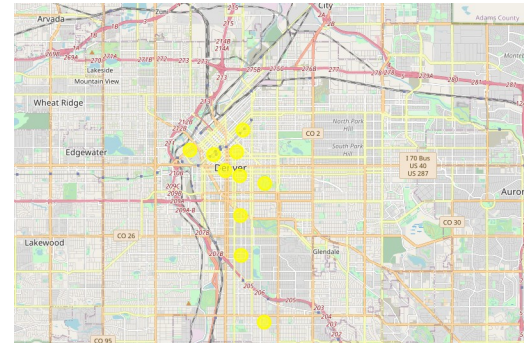
Crime and Demographic datasets

The top 10 neighborhoods for 2015-2020 crime events are derived after data cleaning and semplification (car accidents entries are discarded). The Crime distribution chart highlights the outlier neighboorhood (Five Points).

Datasets considers 2010 Census data already aggregated by Neighborhood. Top 10 neighborhhoods for citizens 18-65 percentage are the one to be considered for this project.

- five points
- cbd
- capitol hill
- montbello
- east colfax
- civic center
- union station
- lincoln park
- westwood
- gateway green valley ranch

- auraria
- cbd
- capitol hill
- civic center
- north capitol hill
- university
- speer
- five points
- cheesman park
- washington park west

Conclusion

The neighbordhoods that meet all the criteria established as prerequisites for this project are:

auraria
cheesman park
speer
university
washington park west

Future Improvements

- check the availability of the foreseen space for the sport facility in those areas in case of new facility or the presence of already suitable buildings to be purchased or rented.
- analisys on the land/rent price for each neighborhood.