**Battle of Neighborhoods in Switzerland**

Yueming HU

# Introduction

## Background

I live in Thun, a medium sized *commune* in the Canton of Bern in Switzerland. *Commune* here in Switzerland is similar to the concept of neighborhood.

Recently, I got a new job in another canton, i.e. Canton of Geneva, which has 45 neighborhoods. I need to choose a neighborhood to live.

## Problem

What I’m satisfied most for my current neighborhood Thun is its number of food venues and the population size. Because I’m a foodie, so I hope there could be many different food-related venues, no matter it’s Café, restaurant, snack shop or even tearoom. And I also like medium sized neighborhood, neither too crowded nor too quiet as no one in the street in the evening.

Therefore, I want to find a similar neighborhood as Thun in the new destination Canton of Geneva.

## Target Audience

People who have to move to a new place and want to find a similar neighborhood as the one he or she currently lives in, with particular attention to food venues and population size.

# Data Acquisition and Cleaning

## Data sources

The websites <https://www.citypopulation.de/en/switzerland/geneve/> provides the names and population of each neighborhoods in Canton of Geneva, and the population of Thun is from Wikipedia (<https://en.wikipedia.org/wiki/Thun>).

Geopy: the geographical coordinates, i.e. latitude and longitude of all the neighborhood are obtained by using geopy.

Foursquare Api: the Food related information, including the name, category and number of food venues is obtained through Foursquare Api.

## Data cleaning

Pandas is to read the tables in the above-mentioned website. There are 2 tables, I use index to call the table contains population data. Since I only need the latest population data, I extracted the column of poplation datated in 2019-12-31 and the name of each municipality, then create a dataframe as below.

Table

Description automatically generated

The latitude and longitude of each neighborhood are obtained by using the geocode method of geopy.

Then the latitude and logitude data are used in the Foursquare requests to get relevant venue information. Since I want the information related to food, so I choose the category search to get the information. Accoding to [Foursquare](https://developer.foursquare.com/docs/build-with-foursquare/categories/), the category code of food is “4d4b7105d754a06374d81259”. The resulting json file contains many features, I extract the name and category of each food venue, as below.

Table

Description automatically generated

I use the same way to get all the names and categories of food venues in current neighborhood Thun.

## Feature selection

When doing analysis with k-means clustering, the features I will use are the population number and the food venue number. So I use the groupby method to get the number of venues in each neighborhood.

The population number is converted into population in thousand, so that the feature values would be comparable and can be used directly. The plot of neighborhood features is as below.

Chart, scatter chart

Description automatically generated