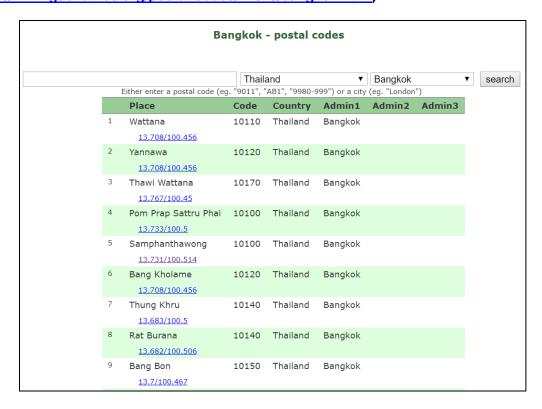# Data

We will use data from two main sources in this analysis

1. **Location data**

This will tell us about the name of each district area in Bangkok together with their Latitude and Longitude as Geographic coordinate that will be used to fetch the important places within each area from FourSquare API. In this study, the data will be scraped from geonames using the following site. **(https://www.geonames.org/postal-codes/TH/10/bangkok.html)**

	Place	Code	Country	Admin1	Admin2	Admin3
	Bangkok - postal codes					
1	Wattana 13.708/100.456	10110	Thailand	Bangkok		
2	Yannawa 13.708/100.456	10120	Thailand	Bangkok		
3	Thawi Wattana 13.767/100.45	10170	Thailand	Bangkok		
4	Pom Prap Sattru Phai 13.733/100.5	10100	Thailand	Bangkok		
5	Samphanthawong 13.731/100.514	10100	Thailand	Bangkok		
6	Bang Kholame 13.708/100.456	10120	Thailand	Bangkok		
7	Thung Khru 13.683/100.5	10140	Thailand	Bangkok		
8	Rat Burana 13.682/100.506	10140	Thailand	Bangkok		
9	Bang Bon 13.7/100.467	10150	Thailand	Bangkok		

From the webpage, you can clearly see that we have access to

- The name of the area (Place column)
- Postal Code (Code column)
- The Latitude and Longitude of each district in the form of Latitude/Longitude

2. **Popular places in each area**

Since we will cluster areas in Bangkok based on popular places in each area, we will rely on FourSquare API "Explore" request.

The request url will be in the form of https://api.foursquare.com/v2/venues/explore? following by user credentials and some parameters regarding the explore request. In this study, the parameters are as follows:

- ll: Latitude, Longitude

- radius: the area that we are going to observe. In this study, we will use 2,000 m radius to be consistent with the general size of each district
- limit: let's limit our search to top 80 places in each area