Bangkok Haunts Clustered

Nicholas Tantra

Introduction

Bangkok is one of the biggest cities in Asia - with over 10 Million inhabitants and over 1.5 Million square kilometers of land. The city is also home to the best restaurants and nightlife venues in the world and coupled with one of the more well developed transit systems in South East Asia - it is a truly thrumming metropolis.

So what now?

How would one pick between the 50 districts of Bangkok or over 170 Sub-districts with their own little quirky characteristics.

Problem

The data obtained from wikipedia provides the baseline information on the districts and subdistricts of Bangkok - as well as the coordinate information for the Districts. While the HERE API provides us with the coordinates of the Sub-districts of Bangkok.

While the data obtained using the Foursquare explore API can give a rough approximation of the types of venues present in each districts.

With both sets of data should be able to create a rudimentary clustering of districts and subdistricts according to their venue tendencies.

With the cluster data - visitors to Bangkok can use them to locate potential locations to visit or reside in.

___

Data Acquisition and Cleaning

We will use Wikipedia for Bangkok District and Subdistricts information

https://en.wikipedia.org/wiki/List_of_districts_of_Bangkok

https://en.wikipedia.org/wiki/Khwaeng

The coordinate data for the Districts are available through Wikipedia.

The Subdistricts coordinates will be acquired via HERE's API

(https://developer.here.com/develop/rest-apis)

The venues information will be acquired by using FourSquare's Explore API.

(https://developer.foursquare.com/)

The Data

There are

- 50 Districts
- 180 Sub-Districts (4 with no venues)
- At the District level
  - There are over 160 Venue Categories
  - And over 1000 Venues
- At the Sub-District level
  - There are over 250 Venue Categories
  - And over 3800 Venues

Clustering at the District level shows the three prominent clusters around the center of Bangkok and around the central cluster.
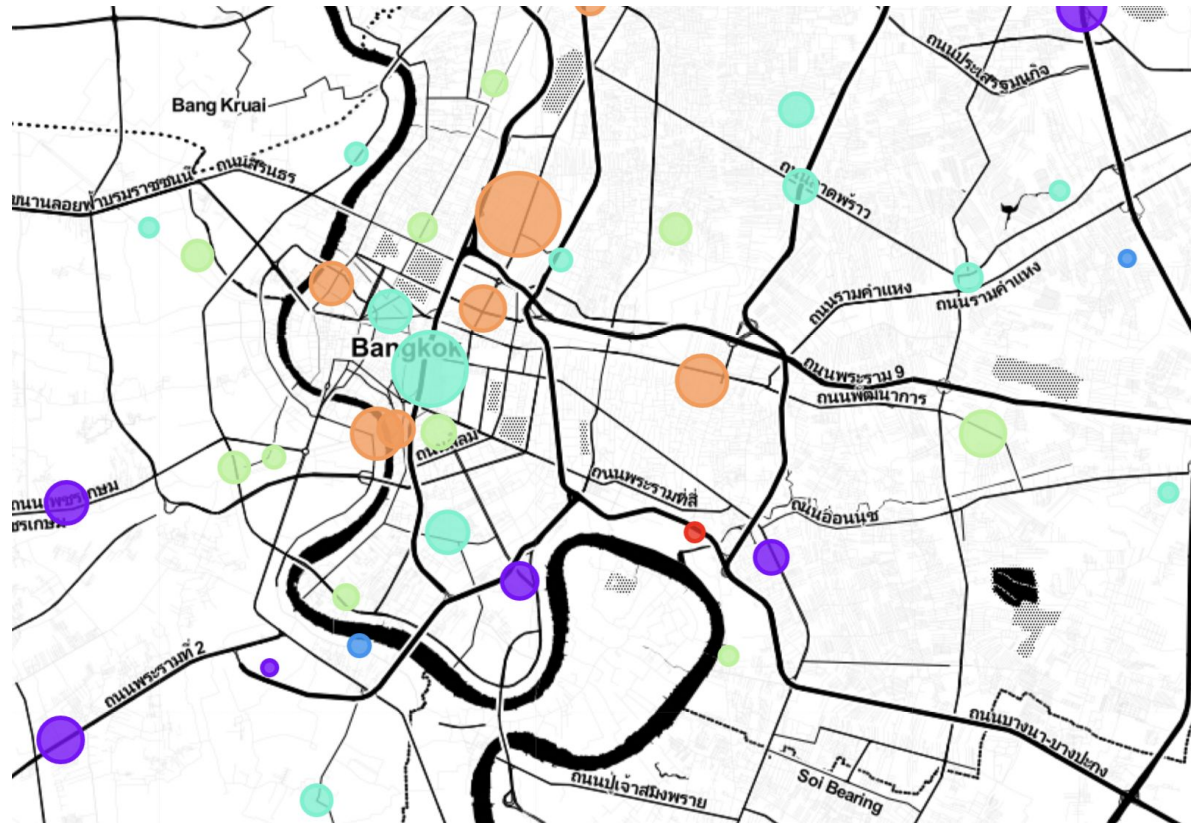
Clustering Bangkok Venues by District

Clustering at the Sub-district level also shows the three prominent clusters.

The clusters are a more clearly defined - with one to the northern side of Bangkok, one at the center and the last one at the south-east region of Bangkok.

The northern cluster features a heavy presence of Nightlife and Accomodation typed venue categories. Note that the northern cluster types are quite prominent around Bangkok in general due to the large numbers of accommodations, nightlife venues and general 'restaurant' category.

The central cluster on the other hand are formed due to the large concentrations of noodle houses and to some extent chinese restaurants.

While the southern clusters are formed due to the large numbers of cafes and western restaurants in the area.

Clustering Bangkok Venues by Sub-district

Clustering Bangkok Venues by Sub-district Using Reduced Categories

The removal of a large chunk of categories used - from over 20 categories to less than 7 categories. The resulting cluster seems to have a less defined shape and cluster.

Conclusions

By examining the clusters formed we can see that they have formed around certain areas of Bangkok in-line with what one would expect when viewing the tourist's guide to Bangkok.

By examining the three clustered visualisation - we can see that the Sub-district level are a more accurate offering a much more precise insight into the neighborhoods associated.

___

The northern cluster forms most of the Bangkok - as it clusters around Accomodations and Nightlife

The central cluster on the other hand are formed due to the large concentrations of noodle houses and to some extent chinese restaurants.

———

While the southern clusters are formed due to the large numbers of cafes and western restaurants in the area.

___

Reduced category clustering seems have lost a lot of details, this is due to the aggressive culling down of categories.

———

WHAT'S NEXT

Following the results of these findings, we can assume that the anomalies or irregularities in the venue clustering are caused my the irregularly shaped district and subdistrict boundaries.

Hence - it would be a great idea to move away from the District-Subdistrict clustering and perhaps focus on a interval-coordinate based system.