

To analyse clustering city of Penang (Pulau Pinang) state in Malaysia

Introduction:

The aim of this project is to demonstrate how Foursquare data can be used to explore and compare neighbourhoods or cities of choice and which problems can it solve.

Problem description:



Penang (a.k.a Pulau Pinang) is famous tourist state in Malaysia. I frequently visit Penang state during long holiday, one of our intention is to have good food. This let me to have a thought about food business in around the city of Penang. I decided to make a clustering analysis on most common venues around different city in Penang where most common venues attract visitors and that will be the source of business for food.

Target audience:

In the future, this approach could be used as a service, helping food business entrepreneur to understand which category of food business is preferred in each city of Penang.

Data description:

I will use the dataset from https://www.weatherdatasource.com/MY-Pulau_Pinang order to get the list of city in Penang. I then use the list of city data and retrieve latitude and longitude coordinates by using Geopy. With City, Latitude and Longitude data, 'Venue Category' can be retrieve using FourSquare API. With JSON data from FourSquare API, I will retrieve top 10 common venues, and use Kmeans to cluster those top common venue. City cluster will be presented via Folium map. The cluster city will show data of food business category is more common and blue ocean strategy can be applied to avoid stiff competition (or too many of same food category in the same city area).