Final project report

Introduction

The room price for Singapore is based on Central area and non-central area. I want to use foursquare data to do a unsupervised clustering base on the neighbourhood, and find out the relationship with room price.

Method and data

I scrape the data from Singapore URA website for 2018 average room price in Singapore dollor.

My hypothesis is that housing price(HDB) is correlated with location clustering.

Data prep:
I get data from Singapore URA data. For simplicity, I only get 2018 4-bedroom price. Some of the area in Singapore don't have HDB. I ignore the NULL values here.
For clustering part, I used Foursquare API to get Singapore district data. For each district, get information about its neighbourhood. Follow the method we learnt in the course, I did a similar Kmeans clustering.

Modeling:
The model I used is Kmeans with cluster = 3. After few try and error, I decided the cluster 3 make more sense to pick.

evaluation:
By compare the average price of the 3 clusters, we find that location indeed has a impact on the price.