

The Battle for Schools

1. Introduction

- Hong Kong is a city with extremely high population density and diverse culture. Different districts may possess very different features and environment. High gap of wealth is also a characteristic of this society, which can potentially bring along many social problems.
- Education has always been a hot topic and attracts a lot of attentions in HK society. Both students and parents face high competitive pressure. There are three types of primary schools, government funded, private and international schools. While the educational resources and cultural influences may differ in these three types of schools, the neighborhood around these schools may also have certain patterns that might be influential to the students.
- In an attempt to focus on education in the context of Hong Kong districts, the following questions/aims are raised for this project:
 - A. Are there any general features for venues nearby primary schools? Can these features help group schools into clusters?
 - B. If the schools can be clustered, does it correlate to its category (government, private or international?), or the district where it is located?
 - C. Is educational resource correlated with other features of the district, e.g. household income, rent, etc.?
 - D. From the above observations, generate implications for students, parents or educators

2. Data

- List for primary schools with classified types and location (including which district it belongs to)
The top nearby venues of each school will be extracted using Foursquare with the same searching parameter. The numbers of venues belonging to each category will be summarized and used as features for clustering using Kmeans. A label will be assigned to each school based on the clustering result and its relationship will be explored with the type of the school (government, private or international) and the district it belongs to.
- Statistics for each district including population, median monthly domestic household rent, median monthly domestic household income, median mortgage payment and loan repayment, median rent to income ratio, etc. For each district, the number of schools for each category and each cluster based on nearby venues will be summarized. The correlations between these data with the upper-listed statistics will be explored.