**INEQUALITY IN SINGAPORE**

**IBM Data Science Capstone Project**

**Sean Woon**

**1. Introduction**

Inequality is a rising problem in Singapore, a city built on extracting wealth from the global elite and channelling it through the government's developmental state to further investments, creating a positive wealth cycle. Inequality is inevitable as some of the population cannot keep up with the rate of change in the global economy, need to and are unable to upskill, and/or lack initial starting capital/suffer indebtedness. The income gap between the bottom and top percentiles have been widening since 2009, to an average income difference of around SGD13,000 in 2019. (Department of Statistics Singapore 2019).

There is a lack of academic and governmental research into the effects of inequality on the physical spaces of Singapore. Research mainly concentrates on the lived experiences of those that are in lower-income groups, such as the publication of Teo You Yenn’s *This is What Inequality Looks Like* (2018). However, anecdotal observations suggest that parts of Singapore catering to high income individuals (e.g. the Central Business District or the luxury property districts such as Tanglin) are getting significantly different in terms of public facilities and the types of private businesses conducted in the area. In line with worldwide urban trends such as gentrification, there seems to be a shift to what is concerned ‘upper-class’ restaurants and cafes rather than the ubiquitous open-air coffeeshop in Singapore, serving Nanyang coffee for less than a dollar. Furthermore, there are areas with concentrations of top-scoring schools, such as in the Bukit Timah region.

Hence, this study will investigate whether there is any correlation between income and the physical spaces of the city, both as it organically develops (through commercial activity), and as it is constructed by the government (through state infrastructure of schools, parks, and transport).

**2. Methodology**

This study would first attempt to use K-means clustering on a dataset with only data about public infrastructure. Clustering would then be applied to Foursquare location data, which reflects commercial activity and tastes and preferences of individuals. Both clusters would then be analysed.

**3. Data**

Public infrastructure data will come from the Singapore government's Onemap API (for planning areas), the Land Transport Authority Datamall API (for number of bus stops in Singapore), and the data available on GovTech's data.gov.sg (for NParks .

Private business data will come from the Foursquare API.

**4. Results**

**A close up of a newspaper

Description automatically generated­**

*Fig 1: Clusters from Commercial Location Data*

**A close up of text on a white background

Description automatically generated**

*Fig 2: Clusters from Government Data on Public Infrastructure*

It appears that the second clustering with government data in fact more *poorly* reflects the distribution of wealth. While Tanglin, Outram, and Marine Parade are clustered together, there is more variation in other clusters, such as the Yishun being in the same cluster as Serangoon and Sembawang, even though it has a much higher proportion of its residents earning below the median individual wage in Singapore. While the commercial clustering better fits the income distribution, it is still not fully reflective, as in cluster ‘3’, Jurong West and Jurong East feature alongside Bukit Timah and Marine Parade. Yet, the inclusion of Jurong might be due to the heavy investment into the shopping districts in the area, with the recent addition in the last 10 years of several megamalls, including Westgate and JEM. This in turn reflects active management of land use in Singapore by the government, who reserves he rights on the sale of land according to plans made by the Singapore Land Authority. Hence, the initial thesis that income inequality would contribute significantly to differences in the shared spaces of the community are unfounded, supporting instead a narrative of active management of inequality.

**5. Conclusions**

While this study could be made better, in adding more dimensions to the government dataset — such as including other transport nodes, and perhaps government-built housing — this has shown that K-means clustering can indeed help to objectively determine if there is correlation between two variables. This study has thus shown that there is minimal correlation, and is a relief to those concerned by rising inequality.