

A Cluster Analysis of Hong Kong Venues Data from Foursquare

IBM Applied Data Science Capstone Course via Coursera

Richard Williams, 21 January 2021.

Hong Kong

Visiting or intending to live in a large new city?



- Are there differences between the districts in Hong Kong? Important to understand what each has to offer.
- Will I be able to eat my favourite food?
- Will I find a good variety of cuisines?
- Until now, no analysis found to categorise Hong Kong's districts by type of venue.
- Is Foursquare an adequate guide to Hong Kong or should I refer to other sources?

Data Sources

Wikipedia

 <https://www.wikipedia.org>



F
SQ

- A list of Hong Kong's 19 districts from **Wikipedia**
- Latitude and longitude values of Hong Kong's 19 districts from **GeoCoder**
- Data about venues in Hong Kong from **Foursquare**

Methodology



- Code run in Jupyter notebook
- GeoCoder supplied latitude and longitude values for Hong Kong districts
- Foursquare provided venues for each district
- Scikit-learn's K-Means algorithm used to cluster districts

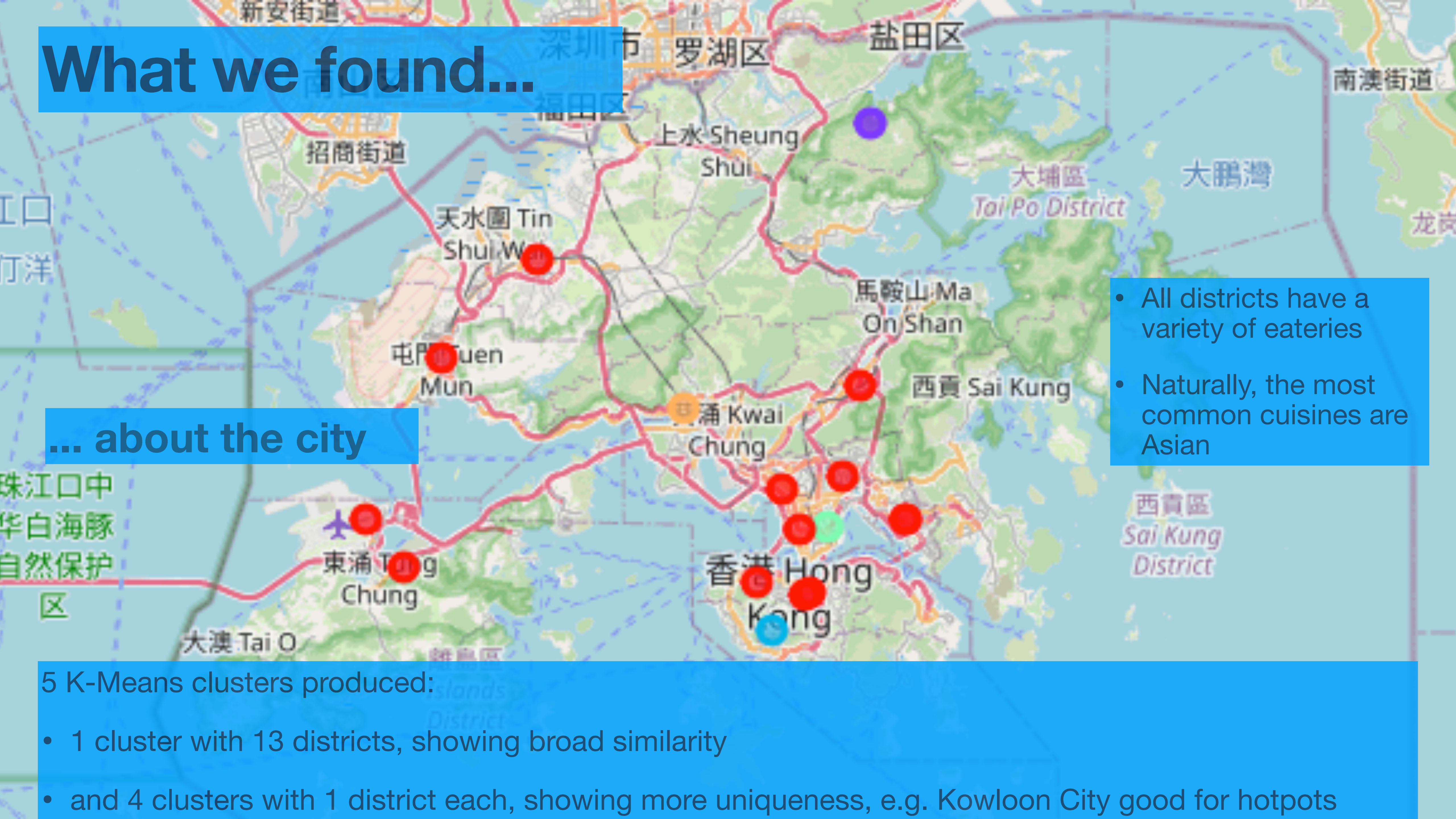
What we found...

... about the city

- All districts have a variety of eateries
- Naturally, the most common cuisines are Asian

5 K-Means clusters produced:

- 1 cluster with 13 districts, showing broad similarity
- and 4 clusters with 1 district each, showing more uniqueness, e.g. Kowloon City good for hotpots



What we found...

... about the data

- Comparatively small dataset for Hong Kong
- Venues weighted towards eateries
- Venue data on infrastructure needs to catch up



Questions?