

# The Battle of Neighborhoods - Singapore

## 1. Introduction

### 1.1. Background

Singapore is the most developed sovereign nation in Asia and is placed highly in key social indicators: education, healthcare, quality of life, personal safety, and housing, with a home-ownership rate of 91%. It is also a major financial and shipping hub, consistently ranked the most expensive city to live in since 2013, and has been identified as a tax haven. Singapore is also a popular tourist destination, with well-known landmarks such as the Merlion, Marina Bay Sands, Gardens by the Bay, the Jewel, the Orchard Road shopping belt, the resort island of Sentosa, and the Singapore Botanic Gardens, the only tropical garden in the world to be honored as a UNESCO World Heritage Site.

### 1.2. Problem

Singapore is a multicultural and vibrant city-state that is suitable for work, live, and play. Yet, the land constraint is also a problem faced by the country. Hence, buying or renting a place to live in Singapore is no trivial matter. This problem is compounded by the varying preferences and priorities of individuals.

### 1.3. Interest

People looking to buy or rent a place in Singapore will be interested in factors such as accessibility to amenities such as food places, supermarkets, parks, shopping centers, etc.

## 2. Data acquisition and cleaning

### 2.1. Data sources

Locations in Singapore may be classified based on postal codes which are administered by Singapore post. The postal district, postal sector which is associated with general locations may be accessed from [https://en.wikipedia.org/wiki/Postal\\_codes\\_in\\_Singapore](https://en.wikipedia.org/wiki/Postal_codes_in_Singapore). Additional information regarding the amenities near those general locations may be access through FourSquare API.

### 2.2. Data cleaning

Data scraped from the Wikipedia URL above was saved into a pandas data frame. Redundant data such as null data and duplicates were removed.

### 2.3. Feature selection

We will be looking at common venues in each district.

## 3. Modeling

We will apply the K-means clustering algorithm to cluster the districts. K-means clustering is an unsupervised clustering algorithm that is suitable to train the unlabelled data. K-means clustering will gradually learn how to cluster the unlabelled points into groups by analyzing the mean distance of the data points. In this case, we select a k value of 5, which represents the number of clusters in which

the data will be gathered. The algorithm will function to iterate through the clustering of data such that the error function is minimized.

### **3.1. Cluster examination**

Cluster 0 comprises of districts 1 to 18 and 20 to 24, which have mainly food places such as restaurants, coffee shops, bakery as their most common venue. Conversely, cluster 1 is likely to be grouped individually due to having zoo exhibit as one of its common venues. Notably, there is only one zoo, the Mandai zoo in Singapore which distinguishes district 19 from the other districts. Cluster 2 while having food places as its top 3 common venues, also has zoo exhibit and event space listed among its common venue. Cluster 3 is uniquely near to Singapore Changi Airport, the only airport in Singapore. Cluster 4 does not only have a zoo as one of its common venue but also a farm.

## **4. Conclusions**

For anyone deciding on a place to stay in Singapore, cluster 0 which consists of most districts in Singapore appears to be an obvious choice. This is assuming the most people would prefer to be close to amenities such as food places and shopping venues. On the other hand, people who enjoy wildlife and being close to nature may choose to stay in clusters 1 and 3. While people who enjoy a balance of access to zoo and food places may prefer cluster 2. Finally, people who need to travel via flight often may prefer to stay in cluster 3. In all, Singapore appears to be truly a food haven, based on this arbitrary analysis.