HOLIDAYS IN SINGAPORE

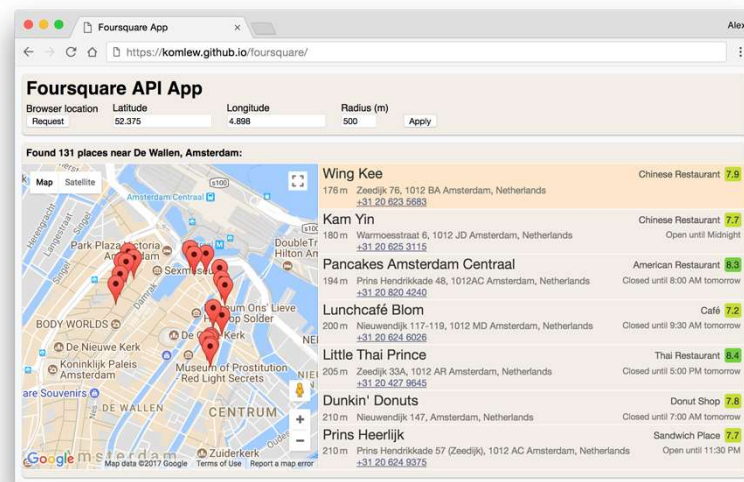BATTLE OF NEIGHBOURHOOD

1. Introduction

- Where to stay in Singapore as a family?

- How can we classify these accommodations into clusters and find
  more easily the right accommodation that corresponds to the
  budget and preferences of the travelers?

- For travelers looking for entire place accommodations

2. Data

Data required and source

2. Data

Data Understanding and preparation


```
#Fetch Singapore listing from AirBnB
df = pd.read_csv('http://data.insideairbnb.com/singapore/sg/singapore/2020-12-29/visualisations/listings.csv', low_memory = False)
df.head()
```

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number
0	49091	COZICOMFORT LONG TERM STAY ROOM 2	266763	Francesca	North Region	Woodlands	1.44255	103.79580	Private room	79	180	1
1	50646	Pleasant Room along Bukit Timah	227796	Sujatha	Central Region	Bukit Timah	1.33235	103.78521	Private room	80	90	18
2	56334	COZICOMFORT	266763	Francesca	North Region	Woodlands	1.44246	103.79667	Private room	66	6	20
3	71609	Ensuite Room (Room 1 & 2) near EXPO	367042	Belinda	East Region	Tampines	1.34541	103.95712	Private room	174	90	20
4	71896	B&B Room 1 near Airport & EXPO	367042	Belinda	East Region	Tampines	1.34567	103.95963	Private room	93	90	24

2. Data

Final dataframe used

```
#Fetch Singapore listing from AirBnB
df = pd.read_csv('http://data.insideairbnb.com/singapore/sg/singapore/2020-12-29/visualisations/listings.csv', low_memory = False)
df.head()
```

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number
0	49091	COZICOMFORT LONG TERM STAY ROOM 2	266763	Francesca	North Region	Woodlands	1.44255	103.79580	Private room	79	180	1
1	50646	Pleasant Room along Bukit Timah	227796	Sujatha	Central Region	Bukit Timah	1.33235	103.78521	Private room	80	90	18
2	56334	COZICOMFORT	266763	Francesca	North Region	Woodlands	1.44246	103.79667	Private room	66	6	20
3	71609	Ensuite Room (Room 1 & 2) near EXPO	367042	Belinda	East Region	Tampines	1.34541	103.95712	Private room	174	90	20
4	71896	B&B Room 1 near Airport & EXPO	367042	Belinda	East Region	Tampines	1.34567	103.95963	Private room	93	90	24

3. Methodology
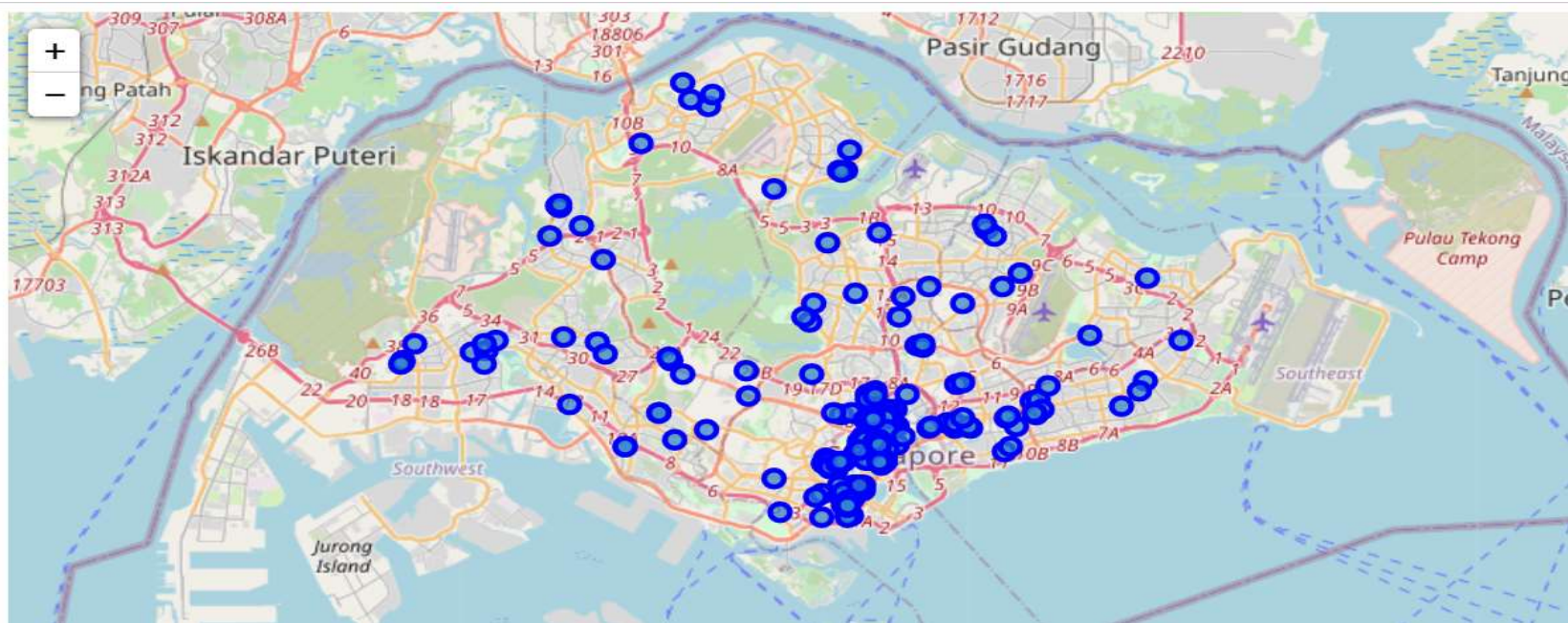
The methodology is made up of 4 sections listed below:

Part 1: Exploratory Data Analysis on Accomodation listing

Part 2: Get Venues around accommodation

Part 3: One Hot Encoding and K-means Clustering.

Part 4: Further analysis on each cluster¶

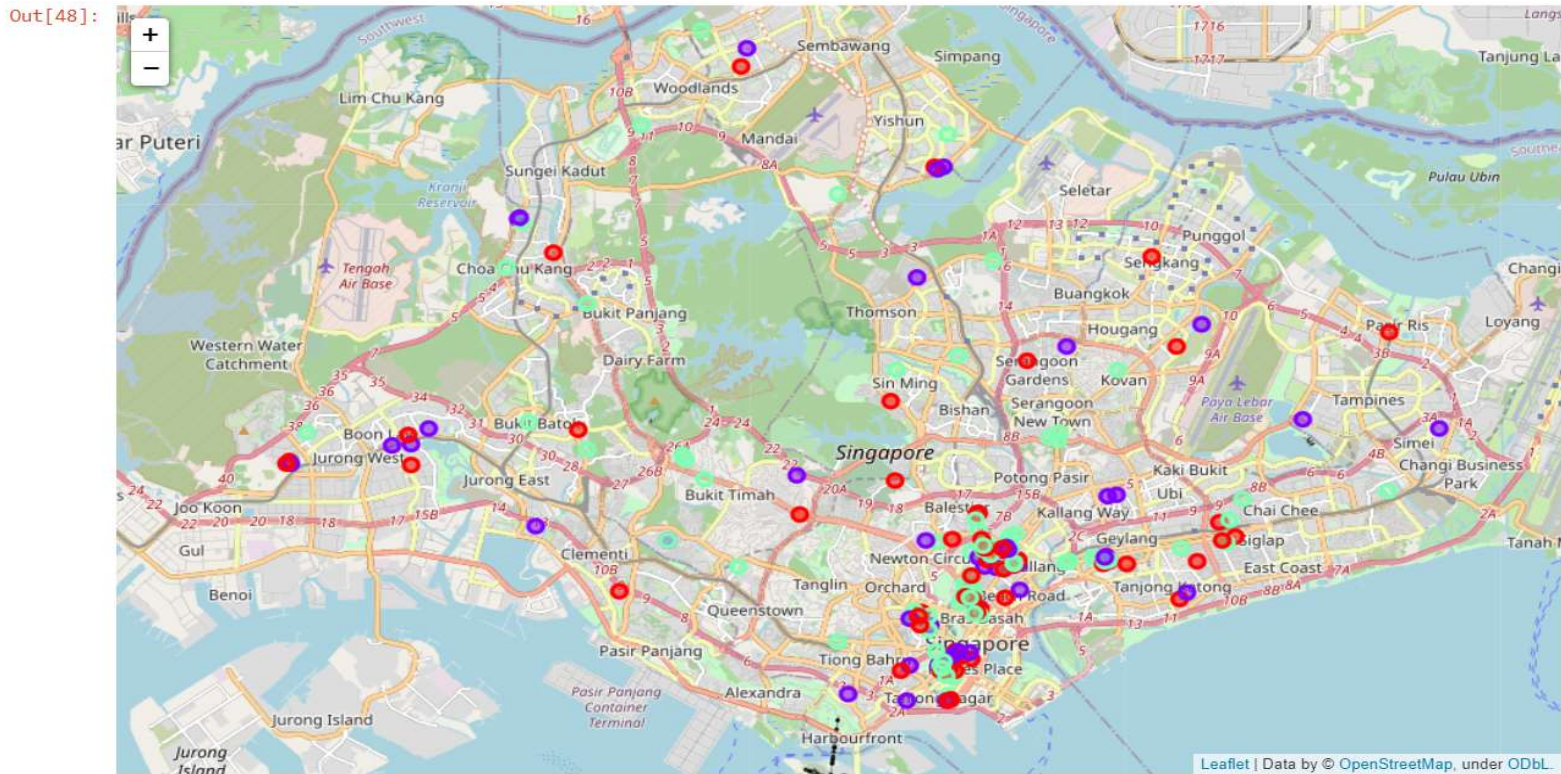# 3.1 Exploratory Data Analysis on Accommodation listing

3.2: Get Venues around accomodation

Final table used for k-means clustering in next step.

	id	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	custom_category
0	56334	1.44246	103.79667	Asia Ghani	1.437431	103.795388	Malay Restaurant	Food
1	56334	1.44246	103.79667	Kampung Admiralty Hawker Centre	1.439939	103.800774	Food Court	Food
2	56334	1.44246	103.79667	Starbucks	1.439761	103.800659	Coffee Shop	Food
3	56334	1.44246	103.79667	McDonald's	1.445931	103.798101	Fast Food Restaurant	Food
4	56334	1.44246	103.79667	KFC	1.437409	103.795428	Fast Food Restaurant	Food

3.3 One Hot Encoding and K-means Clustering

4.0 Discussion

- The cluster are not so well separated as maybe the radius of venue explore is set too high for the scale of Singapore at 700

- However, we can see that the prices are similar whether it is downtown or further away from city centre, all amenities are very close be it restaurant, transport, outdoor or culture.

5. Conclusion

- The clusters are overlapping, and this may be due to small size of the city state. Nevertheless, with this result, we can see that the different neighbor offers similar amenities (restaurants and malls), outdoor activities and convenience of transport. Hence, we conclude base on this finding that any cluster is suitable for families.

- To further improve the analysis, we could try with a smaller radius in the parameters used in the Foursquare API to see with it, the clusters formed maybe more distinct.