Singapore Visitors Venue Recommendation

By Kanupuri Bharadwaja

Date: 14/04/2020

I. INTRODUCTION: Business Problem

Singapore is much more than the sum of its numerous attractions. It's constantly evolving, reinventing, and reimagining itself, with people who are passionate about creating new possibilities. It's where foodies, explorers, collectors, action seekers, culture shapers, and socialisers meet and create new experiences every day. Tourism is also a major industry and contributor to the Singaporean economy, attracting 18.5 million international tourists in 2018, more than 3 times of Singapore's total population. Singapore is the 5th most visited city in the world, and 2nd in the Asia-Pacific. However, travelers often find themselves confused with what to eat, where to stay, places to visit etc.. Even websites recommend usual tourist attractions and places to stay based on simple keywords like "Hotels", "Best Food","Travel" etc

The goal of this project is to collect best visited/reviewed places using Foursquare API and provide an accurate recommendation. We will be utilizing data retrieved from Singapore open data sources and FourSquare API venue recommendations.

The recommender system in this notebook will provide the following use case scenario:

If a person is planning to visit Singapore as a Tourist and looking for reasonable accommodation.

If the user wants to receive a venue recommendation where he can stay - places like Best Hotels or rent an apartment according to his budget with nearby places of interest or search category option.

With clear segmenting and cluster analysis I would like to present a comparison table of all possible town venues.

We will download the data from data.gov.sg and utilize the Median Rental prices

- Singapore Median Rental Prices by town.

- Popular Food venues in the vicinity. (Sample category selection)

Note: While this demo makes use of Food Venue Category, Other possible categories can also be used for the same implementation such as checking categories like:

- Outdoors and Recreation
- Nightlife
- Nearby Schools, etc.

As FourSquare API only allows 50 free venue queries, I would like to limit the scope of this search.

II. DATA ACQUISITION

We will use of the following data sources to make our recommendation:

Singapore Towns and median residential rental prices.

Data retrieved from Singapore open dataset from median rent by town and flattype from https://data.gov.sg website.

The data source contains median rental prices of Singapore HDB units from 2005 up to 4th quarter of 2019. I will retrieve rental the most recent recorded rental prices from this data source (Q4 2019) being the most relevant price available at this time. For this project, I will simplify the analysis by using the average rental prices of all available flat type.

Singapore Top Venue Recommendations from FourSquare API

(FourSquare website: www.foursquare.com)

I will be using the FourSquare API to explore neighborhoods in selected towns in Singapore. The Foursquare explore function will be used to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. The following information are retrieved on the first query:

- Venue ID
- Venue Name
- Coordinates : Latitude and Longitude
- Category Name

Another venue query will be performed to retrieve venue ratings for each location. Note that rating information is a paid service from FourSquare and we are limited to only 50 queries per

day. With this constraint, we limit the category analysis with only one type for this demo. I will try to retrieve as many ratings as possible for each retrieved venue ID.

III. Cleaning and Preparation

Singapore Towns List with median residential rental prices.

The source data contains median rental prices of Singapore HDB units from 2005 up to 4th quarter of 2019.

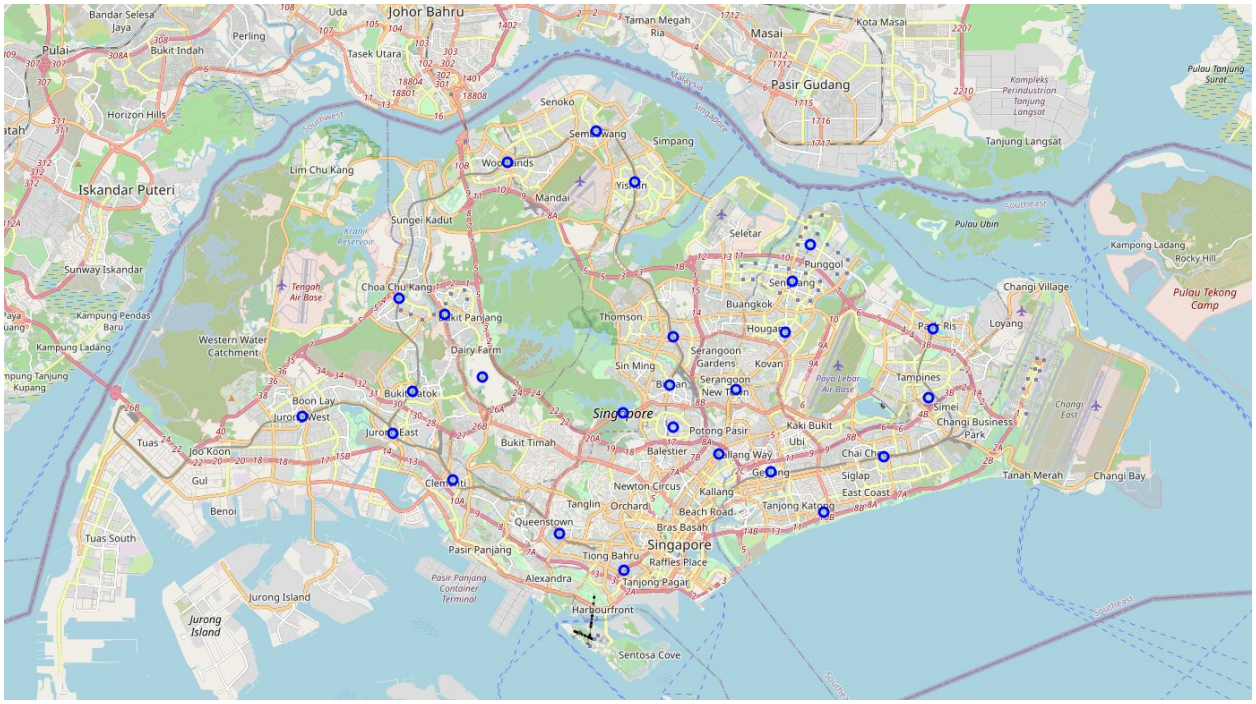
Data Cleanup and re-grouping. The retrieved table contains some un-wanted entries and needs some cleanup.

The following tasks will be performed:

- Drop/ignore cells with missing data.
- Use most current data records.
- Fix data types.

We will use the Geocode library to retrieve the coordinates (latitude and longitude of each town center. For this exercise, I just used the MRT stations as the center points of each evaluated town.
The town coordinates will be used in retrieval of Foursquare API location data.

Here is a folium map showing the locations of the data pulled.



IV. Performing Segmentation and Clustering Analysis in Singapore

Retrieving FourSquare Places of interest.

We will be using the Foursquare API - **explore** function to get the most common venue categories in each neighborhood and then group the neighborhoods into clusters.

We will use the *k*-means clustering algorithm for analysis.

Finally, the Folium library is used to visualize the recommended neighborhoods and their emerging clusters.

The function **getNearbyVenues** extracts the following information for the dataframe it generates:

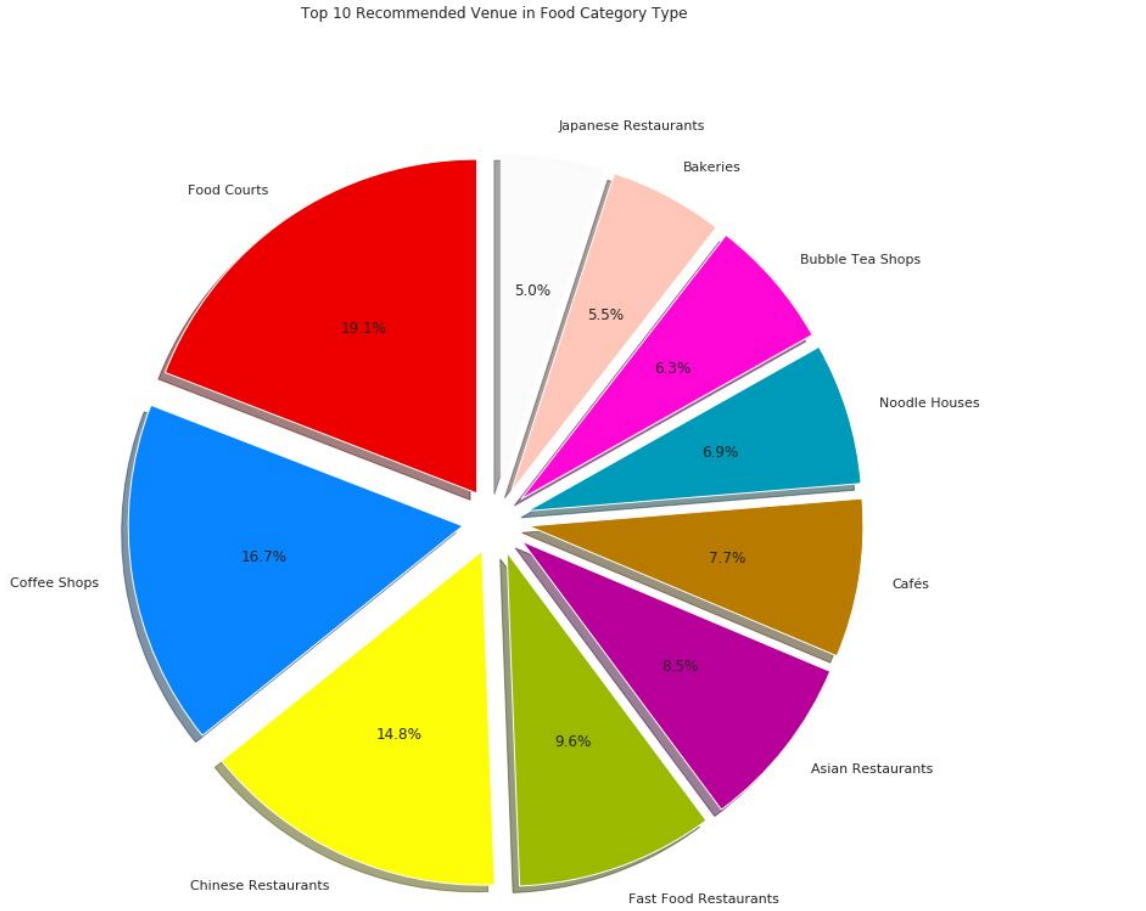
- Venue ID
- Venue Name
- Coordinates : Latitude and Longitude

- Category Name

The function **getVenuesByCategory** performs the following:

1. **category** based venue search to simulate user venue searches based on certain places of interest. This search extracts the following information:
    - Venue ID
    - Venue Name
    - Coordinates : Latitude and Longitude
    - Category Name
2. For each retrieved **venueID**, retrive the venues category rating.

Exploratory Data Analysis



Clustering Neighborhoods

Run *k*-means to cluster the Towns into 5 clusters.

Clustering Neighborhoods

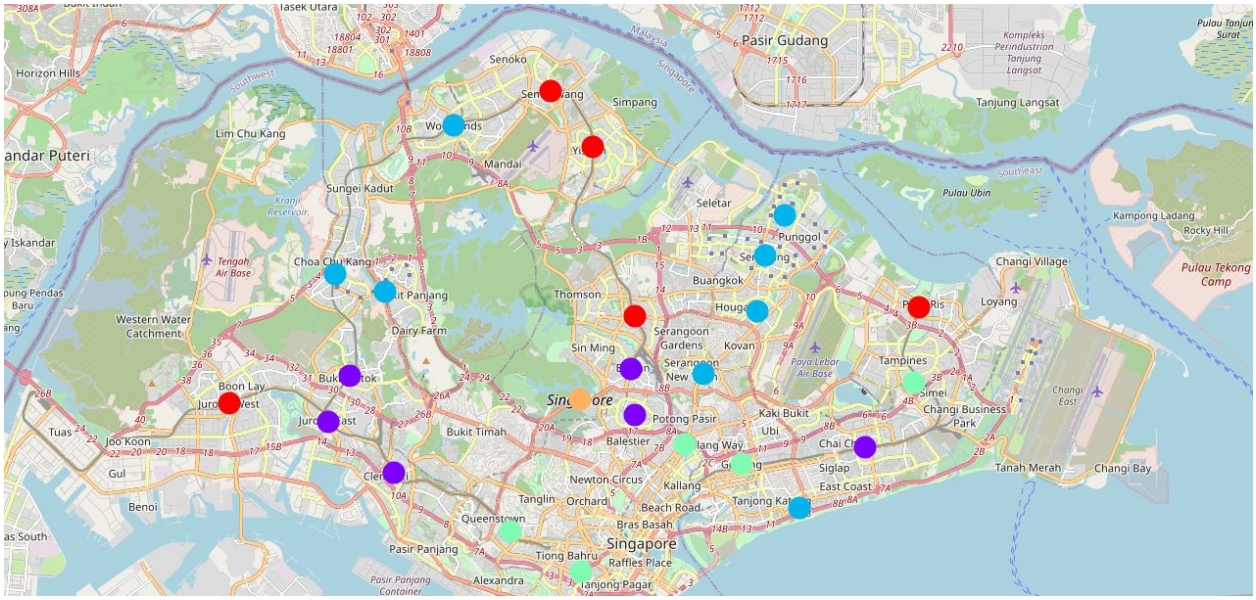
Run *k*-means to cluster the Towns into 5 clusters.

In [323]:
```
# set number of clusters
kclusters = 5
sg_grouped_clustering = sg_grouped.drop('Town', 1)
# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=1).fit(sg_grouped_clustering)

# check cluster labels generated for each row in the dataframe
print(kmeans.labels_[0:10])
print(len(kmeans.labels_))
```

```
[0 1 1 1 3 2 4 2 1 3]
25
```

Results



V. Conclusion

In this notebook, we've analyzed the best town venues and provided recommendations based on the Best Food venue category.

As singapore is a small country with a wide variety of interesting venues scattered around the town, the information extracted in this notebook present on the town areas, as this will be a good supplement to web based recommendations for visitors to find out nearby venues of interest and will be a useful aid in deciding a place to stay or where to go during their travel.

We used Foursquare API and collected a good amount of venue recommendations in Singapore Towns. Sourcing from the venue recommendations from FourSquare has its limitations, The list of venues is not an exhaustive list of all the available venues in the area but covers a good amount of important venues. Also, not all the venues found in the area were provided with store ratings. For this reason, the number of analyzed venues are only about 50% of all the available venues initially collected. The results therefore may significantly change, when more information is collected on those with missing data.

The generated clusters from our results shows that there are very good and interesting places located in areas where the median rents are cheaper. This kind of results may be very interesting for travelers who are travelling on a limited budget. Our results also yielded some interesting findings. For instance, The initial assumption among websites providing recommendations is that the Central Area that has the highest median rent also has better food venues. The results however shows that while Marine Parade, a cheaper location has better rated food courts. Results show that most popular food venues among Singaporeans, residents and visitors are **Food Courts, Coffee Shops, Chinese Restaurants and Fast Food Restaurants**.

Thank you.

References

https://data.gov.sg
https://www.foursquare.com