The Battle of Neighborhoods

Exploring business opportunity around train stations' neighborhood in Singapore

1. Introduction

1.1 Background Information

Singapore railway system is made up of five Mass Rapid Transit (MRT) lines and two Light Rapid Transit (LRT) lines. The MRT & LRT 230km system has over three million daily ridership. There are a total of 187 Mass Rapid Transit (MRT)/Light Rapid Transit (LRT) train stations as of Jan 2019.

In general, the neighborhoods around MRT/LRT stations are highly populated and many shopping centers, shops, business offices are found near the MRT/LRT stations. The typical business or shops near the railway station include shopping malls, supermarkets, coffee shops, restaurants, dessert shops and etc. The kinds of business or shops clustered around the railway stations could vary significantly from station to station. As such, it will be interesting and beneficial for those who are keen in seeking business opportunity in the neighborhood area of railway stations to have a better idea of what are the most common business and/or shops that exist currently around these railway stations so as to make a better business decision.

1.2 Business Problem

To find out the existing common business or shops in the neighborhood around MRT/LRT stations and to explore business opportunity based on the analysis of the MRT/LRT trains' neighborhood data.

1.3 Target Audience

People who are interested in opening a shop for certain business of their interest at location near any of the MRT/LRT stations in Singapore

2. Data Acquisition and Cleaning

2.1 Location Data of MRT/LRT train stations

The location data (latitude & longitude coordinates) of MRT & LRT stations in Singapore is obtained from the following data source:
https://data.world/hxchua/train-stations-in-singapore

The following two methods can be used to retrieve the latitude & longitude data of train stations from the above website:

(1) Direct downloading of .csv file

```
!wget -q -O 'mrtlrt_sg.csv' https://data.world/hxchua/train-stations-in-singapore/workspace/file?filename=mrtsg.csv
```

(2) Calling pandas.read_csv() to read the location data from a provided URL into a dataframe:

```
import pandas as pd
url = 'https://query.data.world/s/heji2ztgpnwx5j5362iqgspyjioh2u'
df_sgp = pd.read_csv(url)
```

The csv data read in consists of 8 columns of data, namely 'OBJECTID', ;STN NAME', ;STN NO', 'X', 'Y', 'Latitude', 'Longitude' and 'COLOR', as shown in the following diagram:

```
url = 'https://query.data.world/s/heji2ztgpnwx5j5362iqgspyjioh2u'
df_sgp = pd.read_csv(url)
df_sgp
```

	OBJECTID	STN_NAME	STN_NO	X	Y	Latitude	Longitude	COLOR
0	12	ADMIRALTY MRT STATION	NS10	24402.1063	46918.1131	1.440585	103.800998	RED
1	16	ALJUNIED MRT STATION	EW9	33518.6049	33190.0020	1.316433	103.882893	GREEN
2	33	ANG MO KIO MRT STATION	NS16	29807.2655	39105.7720	1.369933	103.849553	RED
3	81	BAKAU LRT STATION	SE3	36026.0821	41113.8766	1.388093	103.905418	OTHERS
4	80	BANGKIT LRT STATION	BP9	21248.2460	40220.9693	1.380018	103.772667	OTHERS
...	...	...	...	...	...	...	...	...
182	175	WOODLANDS SOUTH MRT STATION	TE3	23607.8309	45444.7113	1.427260	103.793863	OTHERS
183	146	WOODLEIGH MRT STATION	NE11	32173.3186	35706.3794	1.339190	103.870808	PURPLE
184	6	YEW TEE MRT STATION	NS5	18438.9791	42158.0124	1.397535	103.747431	RED
185	41	YIO CHU KANG MRT STATION	NS15	29294.1283	40413.0820	1.381756	103.844944	RED
186	13	YISHUN MRT STATION	NS13	28187.6787	45686.0701	1.429443	103.835005	RED

187 rows × 8 columns

There are a total of 187 rows of data (with each row represents a train stations) retrieved. In our analysis, only 4 columns of data, i.e. 'STN NAME', 'STN NO', 'Latitude', 'Longitude', are needed. The unwanted columns of data will be discarded.

For some train stations, the same station name could be associated with more than one station number, for instance, the 'Bayfront MRT station' has two station numbers, i.e. 'DT16' and 'CE1', as these train stations comprise of platforms for more than one MRT lines and these platforms even though are linked but they could be at a distance from each other and hence each station no ('STN NO') is having its own latitude & longitude data. In our analysis, each station number is treated as a unique station and for the sake of station identification

completeness, the ‘STN NAME’ and ‘STN NO’ are combined to form the ‘Station ID’ (Station Identifier) and subsequent data analysis will be performed based on the ‘Station ID’ data.

The following diagram shows a new dataframe which contains only the required ‘Station ID’, ‘Latitude’ and ‘Longitude’ data.

```
df_stnloc = df_sgp[['Latitude', 'Longitude']].copy(deep=True)
df_stnloc['Station ID'] = df_sgp['STN_NAME'] + ' (' + df_sgp['STN_NO'] + ')'

# move Station_ID column to the first column
new_columns = [df_stnloc.columns[-1]] + list(df_stnloc.columns[:-1])
df_stnloc = df_stnloc[new_columns]
df_stnloc.head()
```

	Station ID	Latitude	Longitude
0	ADMIRALTY MRT STATION (NS10)	1.440585	103.800998
1	ALJUNIED MRT STATION (EW9)	1.316433	103.882893
2	ANG MO KIO MRT STATION (NS16)	1.369933	103.849553
3	BAKAU LRT STATION (SE3)	1.388093	103.905418
4	BANGKIT LRT STATION (BP9)	1.380018	103.772667

2.2 Latitude & Longitude Data of Singapore

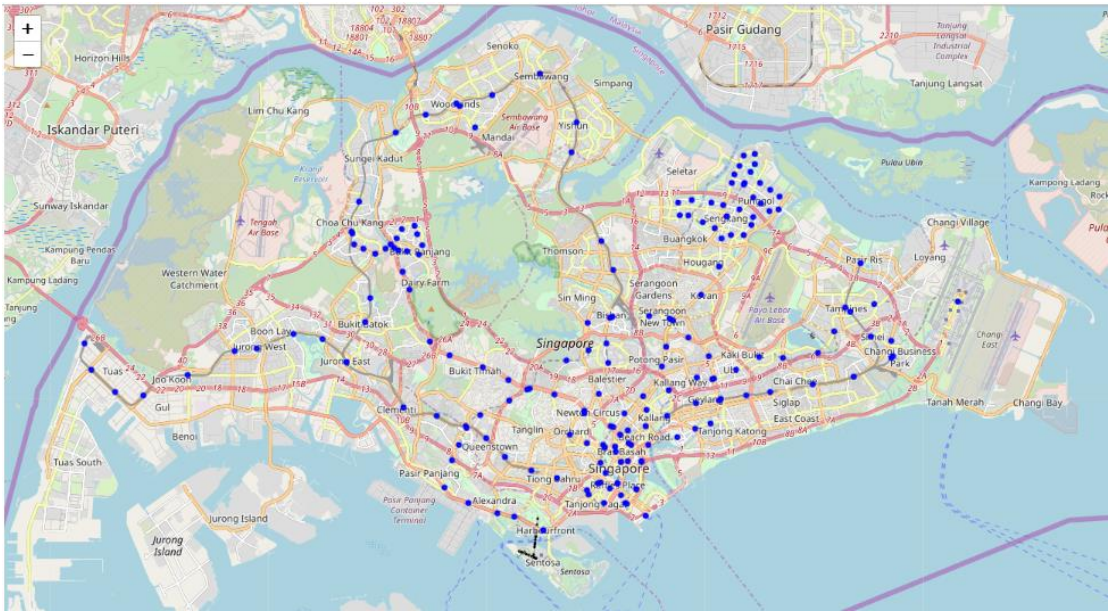
The Python’s geopy library is used to obtain the latitude and longitude values of Singapore.

```
address = 'Singapore'

geolocator = Nominatim(user_agent="sg_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude

print ('The latitude and longitude of Singapore is ({},{})'.format(latitude, longitude))
```

```
The latitude and longitude of Singapore is (1.340863,103.830391822121)
```

The latitude and longitude data of Singapore is needed for drawing a map of Singapore using Python’s folium library.

Together with the latitude & longitude data of train stations, a map of Singapore with markers of train station is generated which is shown in the diagram below:



2.3 Foursquare's Neighborhood Venues Data

Foursquare Places Data API is called to find out the venues (up to 50) located within 500m from each of the train station.

The Python code that calls the Foursquare Places Data API calls for retrieving the nearby venues of each train station is shown below:

```
LIMIT = 50    # limit of number of venues returned by Foursquare API
RADIUS = 500   # define radius

def getNearbyVenues(station_ids, latitudes, longitudes, radius=RADIUS):

    count = 1
    venues_list=[]
    for stn_id, lat, lng in zip(station_ids, latitudes, longitudes):
        print(count, stn_id)
        count = count + 1

        # create the API request URL
        url =
'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        venues_list.append([(
              stn_id,
```

```
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = [
                'Station ID',
                'Stn Latitude',
                'Stn Longitude',
                'Venue',
                'Venue Latitude',
                'Venue Longitude',
                'Venue Category']

    return(nearby_venues)


sgp_venues = getNearbyVenues(station_ids=df_stnloc['Station ID'],
                             latitudes=df_stnloc['Latitude'],
                             longitudes=df_stnloc['Longitude']
                             )
```

The venue, its latitude & longitude data as well as the venue category the venue belongs to are tabulated in a dataframe as shown in the diagram below. These data will then be further processed and analyzed so as to identify the most common venue categories for each of the train stations.

	Station ID	Stn Latitude	Stn Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ADMIRALTY MRT STATION (NS10)	1.440585	103.800998	Kampung Admiralty Hawker Centre	1.439939	103.800774	Food Court
1	ADMIRALTY MRT STATION (NS10)	1.440585	103.800998	Starbucks	1.439761	103.800659	Coffee Shop
2	ADMIRALTY MRT STATION (NS10)	1.440585	103.800998	NTUC Fairprice	1.439955	103.800761	Supermarket
3	ADMIRALTY MRT STATION (NS10)	1.440585	103.800998	Sawmudeen	1.439802	103.800750	Halal Restaurant
4	ADMIRALTY MRT STATION (NS10)	1.440585	103.800998	NTUC FairPrice	1.437707	103.797636	Supermarket
5	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Phoon Huat & Co	1.316521	103.881152	Kitchen Supply Store
6	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	The Skewer Bar	1.313674	103.883870	BBQ Joint
7	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Hong Qin Fish & Duck Porridge	1.315787	103.885663	Chinese Restaurant
8	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	No Signboard Seafood Restaurant	1.313155	103.882700	Seafood Restaurant
9	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Ian Ser Seng Herbs (Turtle) Restaurant 丰成山珍補品 ...	1.314068	103.879981	Chinese Restaurant
10	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	J.B. Ah Meng Restaurant	1.313735	103.886182	Chinese Restaurant
11	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	悦意坊 Yes Natural F & B Vegetarian Restaurant	1.315828	103.883807	Vegetarian / Vegan Restaurant
12	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Builders At Sims	1.317739	103.879848	Café
13	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Tai Dong Teochew Braised Duck Rice	1.317166	103.879990	Food Truck
14	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	The Lorong 24A shophouse series	1.312777	103.884045	Boarding House
15	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Koung's Wan Tan Mee 鉅氏雲吞面 (Koung's Wan Tan Mee)	1.314860	103.883855	Noodle House
16	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Shuang Shun Chicken Rice	1.312680	103.885536	Asian Restaurant
17	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	7-Eleven	1.312758	103.880738	Convenience Store
18	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Penang Seafood Restaurant	1.314833	103.882075	Seafood Restaurant
19	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Sims Vista Market & Food Centre	1.316978	103.879382	Food Court
20	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Aikido Shinju-kai (Singapore) HQ 心柔合木氣道场	1.315173	103.883155	Martial Arts Dojo
21	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Ci Hang Western & Chinese Vegetarian Fast Food	1.315744	103.883248	Vegetarian / Vegan Restaurant
22	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Hollywood Duck Rice + Duck Porridge	1.318095	103.879745	Chinese Restaurant
23	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Durians @ Lorong 21 Geylang	1.314700	103.879937	Farmers Market
24	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	The Ranch	1.316328	103.883760	Steakhouse
25	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Shi Wei Xian HongKong Tim Sum	1.319740	103.885760	Dim Sum Restaurant
26	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Tasvee Restaurant	1.313421	103.883765	Indian Restaurant
27	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	觀音堂 Kwan Inn Vegetarian Food	1.315932	103.886388	Vegetarian / Vegan Restaurant
28	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Swee Guan Hokkien Mee	1.313999	103.885706	Noodle House
29	ALJUNIED MRT STATION (EW11)	1.316433	103.882893	Sin Hin Restaurant	1.319785	103.885765	Asian Restaurant