THE RESTAURANT SCENE IN SINGAPORE

1. INTRODUCTION

1.1.Background

Singapore is a popular business and tourist center in southeast Asia. It also serves as one of the busiest transit points for people traveling across Asia. Singapore boasts of rich diversity in its population not only across various age groups but also across different ethnic groups. Singapore is well known for its hawker food centers and its vibrant restaurant scene which is constantly evolving. Several dining restaurants and chefs have been recognized at an international level (https://www.stb.gov.sg/content/stb/en/industries/dining-and-retail.html). The diverse culture along with the dining experience also helps Singapore in the tourism industry which is one of the major contributors to the country's economy.

Singapore is the second densest country (by population) in the world (https://www.singstat.gov.sg/find-data/search-by-theme/population/population-and-population-structure/latest-data). For sustainable living and to aid urban planning, the Singapore government divided the land into 55 planning areas (PA) across 5 regions. The census data for each of these PA is provided by the Statistics Department of Singapore.

1.2. Problem

This project is aimed at entrepreneurs to help them understand the restaurant scene in Singapore.

a) What are the different types of restaurants across different planning areas?
b) Where should an entrepreneur open an Indian restaurant based on the ethnic population?

2. DATA

2.1.Data Sources

The data used in this project is collected across various publicly available data platforms.

1. The geographical details of the various Planning Areas (PA) are obtained from https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore. The table columns used are 'Planning Area', 'Region' and 'Area (km^2)'.
2. The latest population data for different PA is collected from https://www.singstat.gov.sg/find-data/search-by-theme/population/population-and-population-structure/latest-data (updated 24/9/2020). The total population is obtained from the 'Residents by age group and type of dwelling' dataset

3. The resident population by planning area and ethnic group is available only for 2015 (https://data.gov.sg/dataset/resident-population-by-planning-area-subzone-ethnic-group-and-sex-2015?resource_id=d683afc9-d1e6-45a1-8d51-073710b7daca) . The fraction of each ethnic group in each PA is assumed to be the same for 2020.
4. The geolocations of the various PA are obtained using Nominatim which uses OpenStreetMap data to find locations based on Name and Address (https://nominatim.org/).
5. The details of the types of restaurants and other venue categories for different PA are collected using the FourSquare API with a search radius personalized on the PA area.

The datasets after cleaning will be used for

2.2.Data cleaning and Wrangling

2.2.1. 2020 Population dataset

The latest population dataset is in the form of a csv file. Using Pandas, the file was read and stored as a dataframe.

	PA	SZ	AG	Sex	TOD	Pop	Time
0	Ang Mo Kio	Ang Mo Kio Town Centre	0_to_4	Males	HDB 1- and 2-Room Flats	0	2011
1	Ang Mo Kio	Ang Mo Kio Town Centre	0_to_4	Males	HDB 3-Room Flats	10	2011
2	Ang Mo Kio	Ang Mo Kio Town Centre	0_to_4	Males	HDB 4-Room Flats	30	2011
3	Ang Mo Kio	Ang Mo Kio Town Centre	0_to_4	Males	HDB 5-Room and Executive Flats	50	2011
4	Ang Mo Kio	Ang Mo Kio Town Centre	0_to_4	Males	HUDC Flats (excluding those privatised)	0	2011

For the current problem, we require only the columns ‘PA’, ‘Pop’ for ‘Time’=2020. The data is then grouped by PA and the corresponding total population was calculated for each PA. Some of the planning areas like ‘Central Water Catchment’ do not have any population. Such rows were removed from the final dataset. This resulted in the number of Planning areas with population reducing from 55 to 42.

2.2.2. Planning Area/Regions and Area

The names of the different planning areas (PA) and the respective area in km^2 were read from https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore using Pandas.

	Name (English)	Malay	Chinese	Pinyin	Tamil	Region	Area (km2)	Population[7]	Density (/km2)
0	Ang Mo Kio	NaN	宏茂桥	Hóng mào qiáo	ஆங் மோ கியோ	North-East	13.94	163950	13400
1	Bedok	*	勿洛	Wù luò	பிடோக்	East	21.69	279380	13000
2	Bishan	NaN	碧山	Bì shān	பீஷான்	Central	7.62	88010	12000
3	Boon Lay	NaN	文礼	Wén lǐ	பூன் லே	West	8.23	30	3.6
4	Bukit Batok	*	武吉巴督	Wǔjí bā dū	புக்கிட் பாத்தோக்	West	11.13	153740	14000

The non-English columns were removed and only the Name, Region and Area columns were used. The Name column was renamed as 'Planning Area' and is used as the index of the resulting dataframe.

Planning Area	Region	Area (km2)
Ang Mo Kio	North-East	13.94
Bedok	East	21.69
Bishan	Central	7.62
Boon Lay	West	8.23
Bukit Batok	West	11.13

The above dataframe was merged with the population dataset from section 2.2.1. for common planning areas.

PA	Pop	Region	Area (km2)
Ang Mo Kio	162670	North-East	13.94
Bedok	277720	East	21.69
Bishan	87560	Central	7.62
Bukit Batok	158510	West	11.13
Bukit Merah	151700	Central	14.34

2.2.3. Ethnic group dataset

The resident population for different PA by ethnic groups is available as a csv file for the year 2015.

	year	level_1	level_2	level_3	value
0	2015	Total	Total	Total	3902690
1	2015	Total	Total	Ang Mo Kio- Total	174770
2	2015	Total	Total	Bedok- Total	289750
3	2015	Total	Total	Bishan- Total	90700
4	2015	Total	Total	Boon Lay- Total	30

The 'level_2' column provides the total population and the population of males and females for each ethnic group. Only the rows containing 'level_2'= 'Total' are used. The 'level_1' contains the ethnic groups classified as 'Total', 'Chinese', 'Malay', 'Indians' and 'Others'. The columns 'Year' and 'level_2' are dropped.

level_3	level_1	value
Ang Mo Kio	Total	174770
Bedok	Total	289750
Bishan	Total	90700
Boon Lay	Total	30
Bukit Batok	Total	139270

The idea is to obtain the fraction of each ethnic group for different PA. This is obtained by first separating the dataframe by ethnic groups.

level_3	typet	popt	typec	popc	typem	popm	typei	popi	typeo	popo
Ang Mo Kio	Total	174770	Chinese	143290	Malays	13060	Indians	14150	Others	4270
Bedok	Total	289750	Chinese	208880	Malays	43980	Indians	25110	Others	11780
Bishan	Total	90700	Chinese	77220	Malays	3760	Indians	7000	Others	2720
Boon Lay	Total	30	0	0	0	0	0	0	0	0
Bukit Batok	Total	139270	Chinese	102080	Malays	19570	Indians	13730	Others	3900

Some of the PA in 2015 showed 0 population for different ethnic groups. This could be possibly due to error in the census collection. Such rows are removed and ethnic fractions are calculated. The total number of PA with valid data is now 41. The ethnic fraction dataframe is now merged with the population data obtained in section 2.2.2. Two columns are added showing the ethnic group with the largest and 2nd largest fraction for each PA. The index is reset and the column PA is renamed as Planning Area.

	Planning Area	Pop	Region	Area (km2)	Chinese	Malay	Indians	Others	Max	2nd
0	Ang Mo Kio	162670	North-East	13.94	0.819878	0.074727	0.080964	0.024432	Chinese	Indians
1	Bedok	277720	East	21.69	0.720897	0.151786	0.086661	0.040656	Chinese	Malay
2	Bishan	87560	Central	7.62	0.851378	0.041455	0.077178	0.029989	Chinese	Indians
3	Bukit Batok	158510	West	11.13	0.732965	0.140518	0.098585	0.028003	Chinese	Malay
4	Bukit Merah	151700	Central	14.34	0.786768	0.085986	0.097023	0.030223	Chinese	Indians

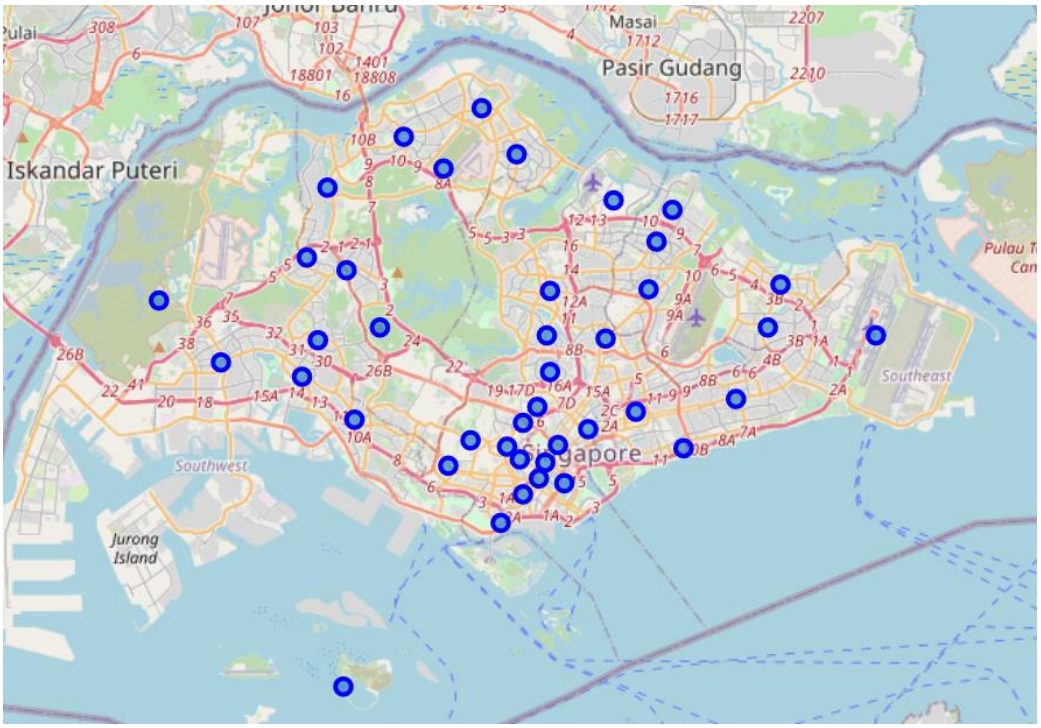
The main assumption in this study is that ethnic fractions for each PA is the same for 2020 as it was in 2015 regardless of the change in population.

2.2.4. Geolocation of each PA

The coordinates of Singapore are obtained using Nominatim which uses OpenStreetMap data. The resulting latitude and longitude (1.357107, 103.8194992) is used as the starting

location to obtain the coordinates of each planning area (PA). The latitudes and longitudes are then stored as separate columns in the dataframe and is plotted using the Folium library.

	Planning Area	Pop	Region	Area (km2)	Chinese	Malay	Indians	Others	Max	2nd	Latitude	Longitude
0	Ang Mo Kio	162670	North-East	13.94	0.819878	0.074727	0.080964	0.024432	Chinese	Indians	1.370080	103.849523
1	Bedok	277720	East	21.69	0.720897	0.151786	0.086661	0.040656	Chinese	Malay	1.323976	103.930216
2	Bishan	87560	Central	7.62	0.851378	0.041455	0.077178	0.029989	Chinese	Indians	1.350986	103.848255
3	Bukit Batok	158510	West	11.13	0.732965	0.140518	0.098585	0.028003	Chinese	Malay	1.349057	103.749591
4	Bukit Merah	151700	Central	14.34	0.786768	0.085986	0.097023	0.030223	Chinese	Indians	1.270439	103.828318



2.2.5. Nearby Venues using FourSquare API

The FourSquare API is used to obtain the nearby venues for each PA based on a search radius. The search radius in 'm' for each PA is calculated using the area of the PA as follows

$$Radius = 1000\sqrt{\frac{Area}{\pi}}$$

Based on the above radius, the FourSquare API reports the nearby venues along with the venue latitude and longitude and venue category.

	Planning Area	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ang Mo Kio	1.37008	103.849523	Bishan - Ang Mo Kio Park	1.362219	103.846250	Park
1	Ang Mo Kio	1.37008	103.849523	Aramsa ~ The Garden Spa	1.362292	103.847602	Spa
2	Ang Mo Kio	1.37008	103.849523	Old Chang Kee	1.369094	103.848389	Snack Place
3	Ang Mo Kio	1.37008	103.849523	FairPrice Xtra	1.369279	103.848886	Supermarket
4	Ang Mo Kio	1.37008	103.849523	Bangkok Street Mookata	1.365688	103.853186	BBQ Joint

Only restaurants and related categories like bar/pub/food courts are used for further analysis. The analysis is performed for each PA and is then studied along with the population dataset obtained in section 2.2.3.