

RECOMMENDATION OF OFFICE OR RENTAL SPACES FOR ENTREPRENEURS IN SINGAPORE

1. INTRODUCTION

Singapore is a very attractive place for entrepreneurs to set up a business. The low tax rates and the ease of doing business in Singapore are important factors that attracts entrepreneurs from around the world to this country. There are a lot of websites providing insights on how to fund your business or how to get grant on setting up a business.

Plus, there should be a proper plan to combat the staffing challenges. An important consideration to make is the number of employees you need to have on board in your first year. You need to get some solid figures to have a proper estimate of the relevant expenses so that you can plan around them.

1.1 Background

Singapore is becoming the preferred destination for first-time entrepreneurs. The country has signed a huge network of Agreements pertaining to Double Taxation Agreements, which is one of the key reasons for this. The Singaporean Government has successfully created a positive ambiance for businessmen so that their businesses can thrive. As so many merits are being offered by the Lion Capital, more entrepreneurs are interested to begin their businesses here.

Another important factor is getting a space to start up the business. Some new businesses prefer working from their homes to save money as well as other resources. The arrangement works fine if you are a sole proprietor.

However, if you want to see your business growing, you must shift to a separate office space soon. At the same time, you need to remember that rents for commercial spaces are high in Singapore.

It is imperative to find the right site to operate your business from for its survival. While setting up an office can be an issue in any other nations, it is not the same case in Singapore. In fact, the government of Singapore has made investments in several huge business parks all through the country, which can function as the ideal stage for the debut of your business.

1.2 Problem

You need to consider some factors before selecting the location for starting up your business. The main factor to be considered is rental needed for the office or retail space. If you are starting your entrepreneur career, setting aside a huge amount of rental is not worth it. But the location wise the space must be good enough. Another factors you have to consider is easily accessible location in terms of transportation facilities and also the food or beverages outlets nearby.

This project aims to recommend the suitable locations for office or retail spaces for entrepreneurs in terms of rental, nearby bus or train stations and nearby food courts in Singapore.

1.3 Interest

Since Singapore Government is providing grants for starting up your own business, it is one of the best locations to build up your business. It is also one of the business hubs in Asia. So for the entrepreneurs or any other local or international companies who wants to set up their business can look for appropriate location using this recommendation engine.

2. DATA PREPARATION

2.1 Data Sources

The main dataset is downloaded from Urban Redevelopment Authority (URA) website. The dataset contains the median rental of office and retail space in quarter 3 of 2019. The dataset contains the Street name and the associated rent in Singapore dollars (\$) per square meter (psm) per month. These rental values are categorized separately for Retail and Office space. Not all the street locations can be used for both office and retail businesses. Some of the street locations can be used for both office and retail spaces.

2.2 Data Cleaning

There are total 152 street locations in the dataset. Some of the values in Office and Retails spaces are NaN because those locations may not be used for office or retail spaces. So, these NaN values are replaced with value 0 and consider this value when segmenting the locations based on the criteria.

The Foursquare API is used to collect that data based on the locations. Foursquare is a technology company that built a massive dataset of accurate location data. Foursquare were very smart about building their dataset. They crowd-sourced their data and had people use their app to build their dataset and add venues and complete any missing information they had in their dataset.

Currently its location data is the most comprehensive out there, and quite accurate that it powers location data for many popular services. Their API and location data are currently being used by over 100,000 developers. Foursquare allows you to make requests through its web platform in order to get familiarized with the API, without complex steps such as user authentication.

This Foursquare API is used to generate the Venue Categories near each street location using the Coordinates and get the relevant venues such as Bus Station, Train Station and Food Court.

3. METHODOLOGY

3.1 Exploratory Data Analysis

The code of project has started with exploratory data analysis. Initially all the important libraries are imported such as **pandas** to perform the data analysis in the dataset, **NumPy** to perform the numerical analysis in Python.

Some other libraries such as **json** to handle JSON files and **requests** to handle requests are imported. **Matplotlib** which is used for plotting the data is also used for the project.

3.1.1 Geopy Library

Geopy is a Python 2 and 3 clients for several popular geocoding web services. Geopy makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources.

Datasets are rarely complete and often require pre-processing. Imagine some datasets have only an address column without latitude and longitude columns to represent your data geographically. In that case, you need to convert your data into a geographic format. The process of converting addresses to geographic information — Latitude and Longitude — to map their locations is called Geocoding.

Geocoding is the computational process of transforming a physical address description to a location on the Earth's surface (spatial representation in numerical coordinates). Geopy includes

geocoder classes for the OpenStreetMap Nominatim, Google Geocoding API (V3), and many other geocoding services. Geocoder classes are in `geopy.geocoders`. Nominatim is a Geocoding service, which is built on top of OpenStreetMap data.

3.1.2 Folium Library

`folium` builds on the data wrangling strengths of the Python ecosystem and the mapping strengths of the `Leaflet.js` library. Manipulate your data in Python, then visualize it in a Leaflet map via `folium`. `folium` makes it easy to visualize data that's been manipulated in Python on an interactive leaflet map. It enables both the binding of data to a map for choropleth visualizations as well as passing rich vector/raster/HTML visualizations as markers on the map.

The library has several built-in tilesets from OpenStreetMap, Mapbox, and Stamen, and supports custom tilesets with Mapbox or Cloudmade API keys. `folium` supports both Image, Video, GeoJSON and TopoJSON overlays.

3.2 Coding Environment

Watson Studio democratizes machine learning and deep learning to accelerate infusion of AI in your business to drive innovation. Watson Studio provides a suite of tools and a collaborative environment for data scientists, developers and domain experts.

Watson Studio provides you with the environment and tools to solve your business problems by collaboratively working with data. You can choose the tools you need to analyze and visualize data, to cleanse and shape data, to ingest streaming data, or to create and train machine learning models.

3.3 Data Preparation

The original dataset consists of 152 rows and 3 columns. This dataset is imported into python notebook. A sample of dataset is shown below.

	Street	Retail	Office
0	AIRPORT BOULEVARD	326.27	NaN
1	ALBERT STREET	89.32	52.02
2	ALEXANDRA ROAD	109.37	68.89
3	ANG MO KIO AVENUE 3	268.37	NaN
4	ANSON ROAD	82.41	65.91

As part of data cleaning, all the NaN values are replaced by value 0.

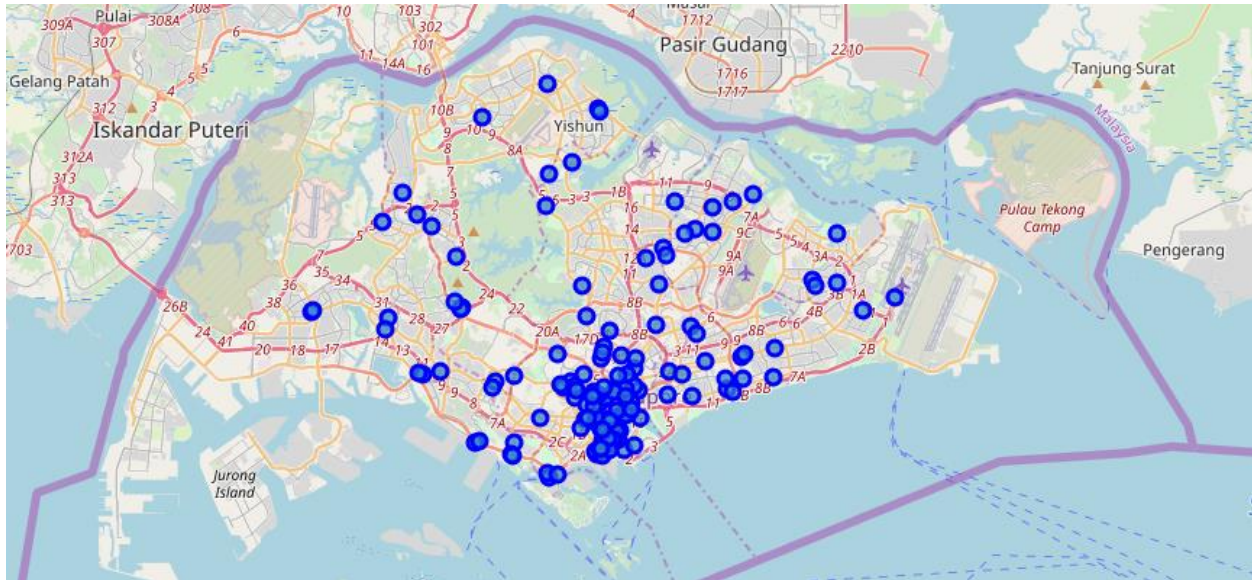
The first column in the data frame is Street which represents the location of office or retail space. Next, we have added a new column for representing the country. This column is filled with country name 'Singapore'. This is for confirming all the street names are in Singapore. Otherwise, the API returns the locations in other countries. The street column and Country column are combined to get the address of the location. This address is used to determine the coordinates of each location. For finding the coordinates Nominatim is called. Now the dataset become the following.

	Street	Retail	Office	Country	address	Coordinates
0	AIRPORT BOULEVARD	326.27	0.00	Singapore	AIRPORT BOULEVARD ,Singapore	(Airport Boulevard, Changi Business Park, Sout...
1	ALBERT STREET	89.32	52.02	Singapore	ALBERT STREET ,Singapore	(Albert Street, Clarke Quay, Selegie, Singapor...
2	ALEXANDRA ROAD	109.37	68.89	Singapore	ALEXANDRA ROAD ,Singapore	(Alexandra Road, Alexandra, Sentosa Cove, Sout...
3	ANG MO KIO AVENUE 3	268.37	0.00	Singapore	ANG MO KIO AVENUE 3 ,Singapore	(Ang Mo Kio Avenue 3, Serangoon North, Singapo...
4	ANSON ROAD	82.41	65.91	Singapore	ANSON ROAD ,Singapore	(Anson Road, Chinatown, Shenton Way, Singapore...

Next need to find the latitude and longitude values of the coordinates and the dataset looks like the following.

	Street	Retail	Office	Country	address	Coordinates	Latitude	Longitude
0	AIRPORT BOULEVARD	326.27	0.00	Singapore	AIRPORT BOULEVARD ,Singapore	(Airport Boulevard, Changi Business Park, Sout...	1.347311	103.983517
1	ALBERT STREET	89.32	52.02	Singapore	ALBERT STREET ,Singapore	(Albert Street, Clarke Quay, Selegie, Singapor...	1.303881	103.851074
2	ALEXANDRA ROAD	109.37	68.89	Singapore	ALEXANDRA ROAD ,Singapore	(Alexandra Road, Alexandra, Sentosa Cove, Sout...	1.278088	103.802975
3	ANG MO KIO AVENUE 3	268.37	0.00	Singapore	ANG MO KIO AVENUE 3 ,Singapore	(Ang Mo Kio Avenue 3, Serangoon North, Singapo...	1.370656	103.873373
4	ANSON ROAD	82.41	65.91	Singapore	ANSON ROAD ,Singapore	(Anson Road, Chinatown, Shenton Way, Singapore...	1.275815	103.846491

After finding the latitude and longitude values of street locations, we need to find out the latitude and longitude values of country to plot the country map as follows.



3.4 Foursquare API

Foursquare is a social location service that allows users to explore the world around them. Users can download the Foursquare application to their iPhone, Blackberry, or Android phone and sign up for free, then connect their Foursquare accounts to their other social media accounts.

After users download the free application and connect on Facebook or Twitter, they can connect with their friends who are also active on Foursquare. Whenever they or their friends check in to a place (that is, they're at a location and they tell others that they are there), the message is broadcast to their friends via Twitter or Facebook.

When a user checks in enough times, that user becomes the mayor of a location, which may or may not give the user access to special offers, depending on the business running a location. For example, a coffee shop might extend a free drink to anyone who becomes a mayor. Users can also earn badges as they explore and check in to locations, leave tips for other users (for instance, "The margaritas here are great!"), and may even can create locations that haven't appeared yet on the service.

3.4.1 Foursquare API

The Foursquare API allows application developers to interact with the Foursquare platform. The API itself is a RESTful set of addresses to which you can send requests, so there's really nothing to download onto your server.

There are both GET and POST methods you can use, which means you aren't just limited to reading from feeds, you can also do useful stuff like checking in and creating locations using the API. As for rate limits, your application is limited to 200 requests per hour per method, so you probably want to implement result caching to "play nice."

For the most part, you want to use basic or OAuth authentication to take full advantage of the various methods and services. The example in this article (/venues) doesn't require any authentication to get started, but it does have additional features available to those who do use authentication.

Next, we had created an URL and a get request function for returning nearby venues based on the location in the dataset.

	Street	Street Latitude	Street Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	AIRPORT BOULEVARD	1.347311	103.983517	World of Sports 3-Top T3	1.347745	103.984680	Sporting Goods Shop
1	AIRPORT BOULEVARD	1.347311	103.983517	Airport Boulevard	1.345959	103.983836	Road
2	AIRPORT BOULEVARD	1.347311	103.983517	Dnata Inflight Catering (CIAS)	1.347647	103.985100	Airport Terminal
3	AIRPORT BOULEVARD	1.347311	103.983517	Hugo Boss	1.347255	103.985423	Men's Store
4	AIRPORT BOULEVARD	1.347311	103.983517	Pizza Hut T3	1.349500	103.983649	Pizza Place

3.5 Data Analysis

Next, using the data frame found out the unique venue categories in the dataset. The venue categories for the location recommendation engine is Bus Station, Train Station, and Food Court.

So these venue categories are separated out from the data frame and created new one.

	Street	Street Latitude	Street Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
42	ALBERT STREET	1.303881	103.851074	Albert Centre Market & Food Centre	1.301024	103.854477	Food Court
80	ALEXANDRA ROAD	1.278088	103.802975	Bus Stop 18019	1.281130	103.802311	Bus Station
82	ALEXANDRA ROAD	1.278088	103.802975	Bus Stop 14369 (Gilman Hts Condo	1.281373	103.803256	Bus Station
94	ANG MO KIO AVENUE 3	1.370656	103.873373	Ghee Hong Eating House	1.369535	103.871813	Food Court
96	ANG MO KIO AVENUE 3	1.370656	103.873373	Cafe 107	1.369715	103.869961	Food Court

3.5.1 Categorical Features

One hot encoding is used for the categorical variables. That is, Encode categorical integer features as a one-hot numeric array.

The input to this transformer should be an array-like of integers or strings, denoting the values taken on by categorical (discrete) features. The features are encoded using a one-hot (aka ‘one-of-K’ or ‘dummy’) encoding scheme. This creates a binary column for each category and returns a sparse matrix or dense array.

By default, the encoder derives the categories based on the unique values in each feature. Alternatively, you can also specify the categories manually. The OneHotEncoder previously assumed that the input features take on values in the range $[0, \max(\text{values}))$. This behaviour is deprecated.

This encoding is needed for feeding categorical data to many scikit-learn estimators, notably linear models and SVMs with the standard kernels. One hot encoding is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction.

3.6 Data Processing

For data processing, K-Means clustering is used. K-means clustering is one of the simplest and popular unsupervised machine learning algorithms.

Typically, unsupervised algorithms make inferences from datasets using only input vectors without referring to known, or labelled, outcomes. A cluster refers to a collection of data points aggregated together because of certain similarities.

You’ll define a target number k , which refers to the number of centroids you need in the dataset. A centroid is the imaginary or real location representing the center of the cluster.

Every data point is allocated to each of the clusters through reducing the in-cluster sum of squares.

In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

The ‘means’ in the K-means refers to averaging of the data; that is, finding the centroid.

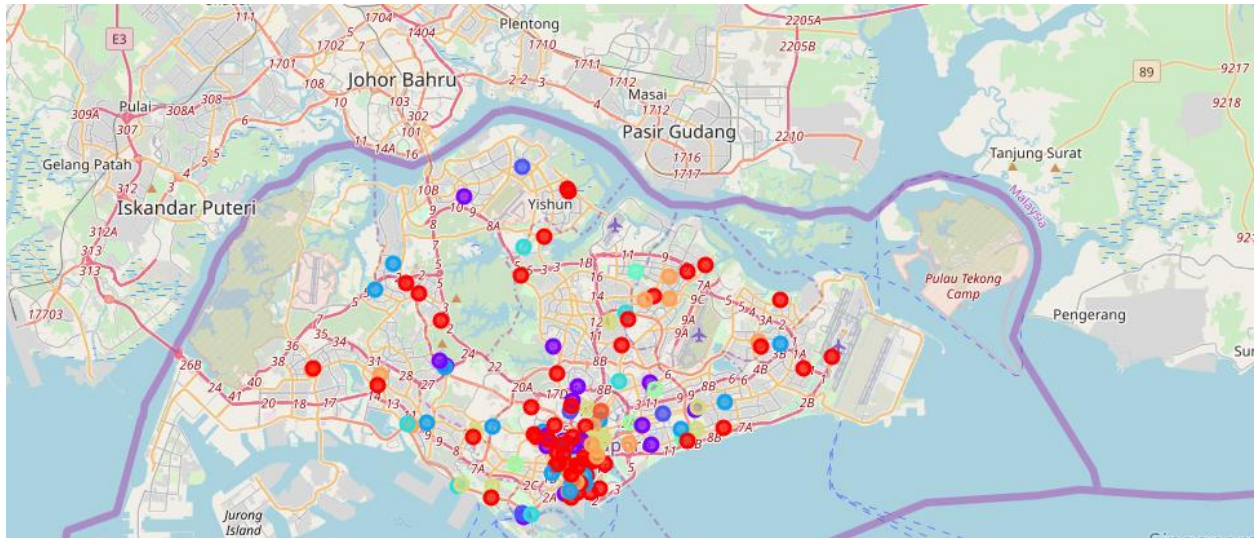
To process the learning data, the K-means algorithm in data mining starts with a first group of randomly selected centroids, which are used as the beginning points for every cluster, and then performs iterative (repetitive) calculations to optimize the positions of the centroids.

4. RESULTS

When the cluster labels are assigned, the final data frame is as follows.

	Street	Retail	Office	Country	address	Coordinates	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	AIRPORT BOULEVARD	326.27	0.00	Singapore	AIRPORT BOULEVARD ,Singapore	(Airport Boulevard, Changi Business Park, Sout...	1.347311	103.983517	0.0	NaN	NaN	NaN
1	ALBERT STREET	89.32	52.02	Singapore	ALBERT STREET ,Singapore	(Albert Street, Clarke Quay, Selegie, Singapor...	1.303881	103.851074	5.0	Food Court	Train Station	Bus Station
2	ALEXANDRA ROAD	109.37	68.89	Singapore	ALEXANDRA ROAD ,Singapore	(Alexandra Road, Alexandra, Sentosa Cove, Sout...	1.278088	103.802975	7.0	Bus Station	Train Station	Food Court
3	ANG MO KIO AVENUE 3	268.37	0.00	Singapore	ANG MO KIO AVENUE 3 ,Singapore	(Ang Mo Kio Avenue 3, Serangoon North, Singapo...	1.370656	103.873373	4.0	Food Court	Bus Station	Train Station
4	ANSON ROAD	82.41	65.91	Singapore	ANSON ROAD ,Singapore	(Anson Road, Chinatown, Shenton Way, Singapore...	1.275815	103.846491	2.0	Food Court	Train Station	Bus Station

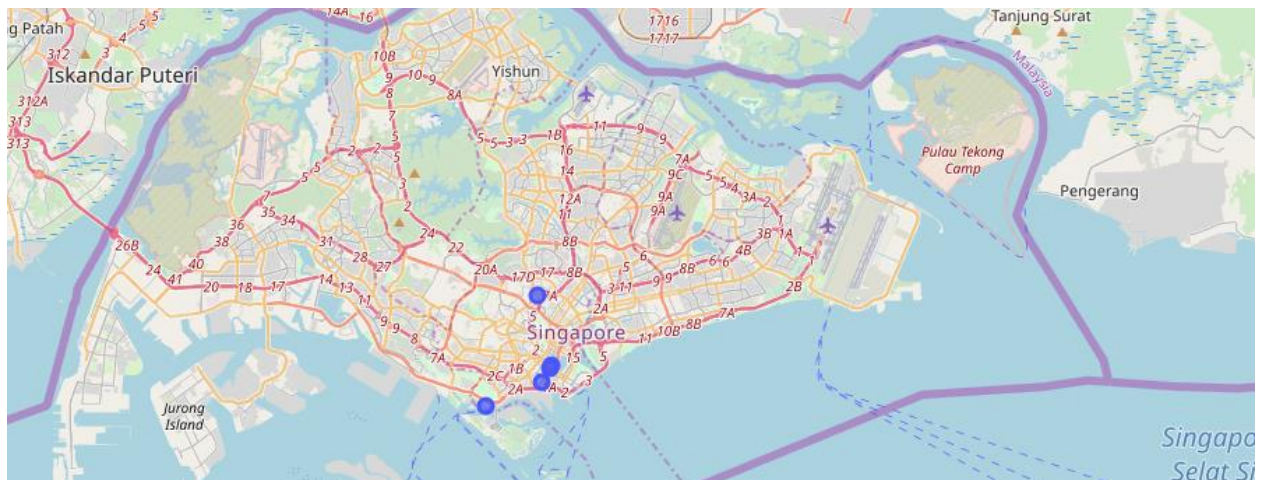
The map showing all the clusters is as follows.



Use Case 1: Locations of the retail spaces having rent less than \$100 (psm) per month and also near to bus or train stations and food courts for K=1.



Use Case 2: Locations of the office spaces having rent less than \$100 (psm) per month and also near to bus or train stations and food courts for $K = 2$.



5. DISCUSSION

During the model development, I have gone through some curdles mainly in the data analysis part. Since there will be same names for the streets in different countries, I recommend using the unique address to determine the coordinates. Also, during clustering I considered the rentals as well for the clustering the locations. In future, with more street locations or business park area, this recommendation engine can be trained more accurately.

6. CONCLUSION

In this study, I analyzed the relationship between the locations and rental values of office or retail spaces in these locations. I identified different venue categories have an important role in determining the rental values for each location. I built a clustering models and plotted a map to find out how the locations are spread out. These models can be very useful in helping entrepreneurs to find an appropriate location for their office space or retail space to set up their business.

.