APPLYING K-MEANS MACHINE LEARNING ALGORITHM TO CHOOSE THE BEST LOCATION FOR COFFEE SHOP START-UP IN HO CHI MINH CITY, VIETNAM

PHAN BAO DUNG

07TH JAN, 2021

Background

- People usually drinks coffee in the morning as culture in Ho Chi Minh city, Vietnam
- Coffee shop is the common place for customer meeting or hangout with friends
- The investment is not so high and not require the big team for start-up so the competitiveness is very high and agile.

Business Problem

- The objective of this article is to report the methodology to analyze and select the best locations in HO Chi Minh City, Vietnam for coffee shop start-up.
- By using the data science and machine learning techniques, this project aims to provide solutions to answer the business question: **if an investor is looking to open a new coffee shop, where would you recommend in Ho Chi Minh city, Vietnam?**

Data Sources

- List of district from Wikipedia page: https://en.wikipedia.org/wiki/Category:Districts_of_Ho_Chi_Minh_City
- Python Geocoder API to fetch the latitude and longitude coordinate of the district
- Foursquare API to explore the district data

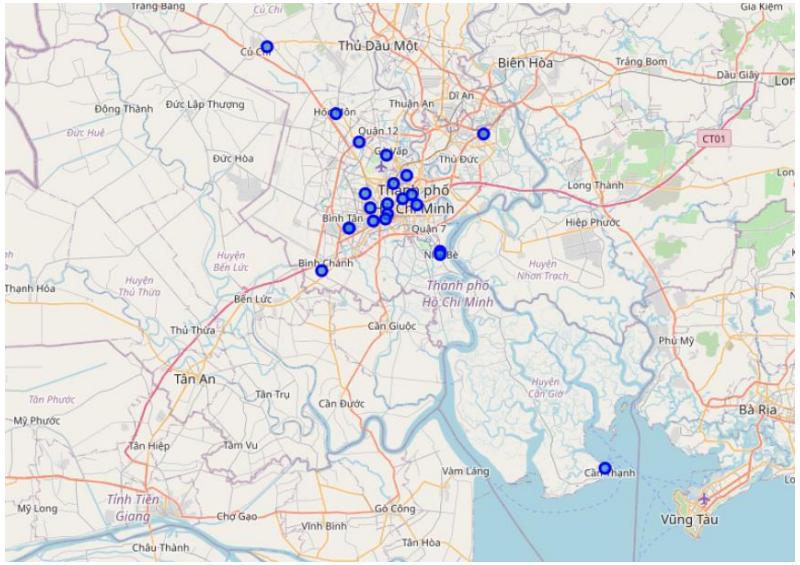
Data Cleaning

- Data downloaded are combined into one table. Remove the unnecessary data
- Merge the list of district into its coordinator

	Neighborhood	Latitude	Longitude
0	Bình Chánh District	10.67922	106.576540
1	Bình Tân District, Ho Chi Minh City	10.73684	106.614480
2	Bình Thạnh District	10.80608	106.692970
3	Cần Giờ District	10.41566	106.961300
4	Củ Chi District	10.97734	106.502230
5	District 1, Ho Chi Minh City	10.78096	106.699110
6	District 3, Ho Chi Minh City	10.77565	106.686720
7	District 4, Ho Chi Minh City	10.76670	106.706470
8	District 5, Ho Chi Minh City	10.75569	106.666370
9	District 6, Ho Chi Minh City	10.74597	106.647690
10	District 7, Ho Chi Minh City	10.70515	106.737480
11	District 8, Ho Chi Minh City	10.74771	106.663340
12	District 10, Ho Chi Minh City	10.76883	106.665990
13	District 11, Ho Chi Minh City	10.76316	106.643140

Exploratory Data Analysis

- Use Folium to show the location of each districts on the map
- Group the venues into categories and map them to the district name
- Create the data frame with location and quantity of coffee shop

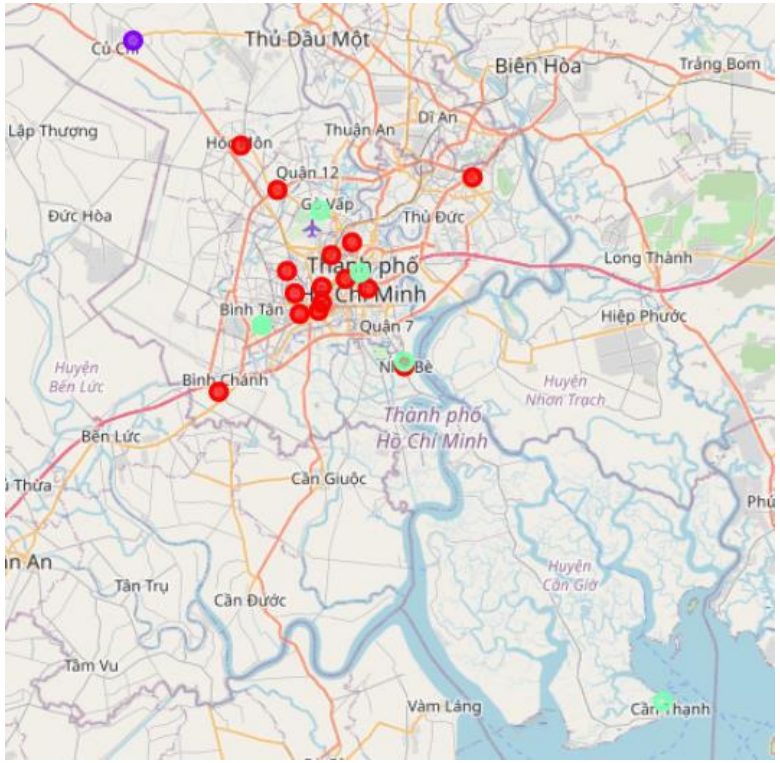



```
[22] #check out how many unique categories can be curated from all the returned values
     print('There are {} uniques categories.'.format(len(venues_df['VenueCategory'].unique())))

     There are 83 uniques categories.

[23] # displaying the first 50 Venue Category names
     venues_df['VenueCategory'].unique()[:50]

     array(['Shopping Mall', 'Multiplex', 'Bed & Breakfast', 'Dessert Shop',
            'Hotel', 'Coffee Shop', 'BBQ Joint', 'Pizza Place', 'Park',
            'Deli / Bodega', 'Sushi Restaurant', 'Vietnamese Restaurant',
            'Café', 'Whisky Bar', 'Noodle House', 'Indian Restaurant',
            'Vegetarian / Vegan Restaurant', 'Flower Shop', 'Beer Bar',
            'Asian Restaurant', 'Supermarket', 'Department Store',
            'Sandwich Place', 'Brewery', 'Hotel Bar', 'Tapas Restaurant',
            'Bar', 'Steakhouse', 'Massage Studio', 'Seafood Restaurant',
            'French Restaurant', 'German Restaurant', 'Italian Restaurant',
            'Spa', 'Burger Joint', 'Japanese Restaurant', 'Bookstore',
            'Nightclub', 'Hotpot Restaurant', 'Bistro', 'Clothing Store',
            'Ramen Restaurant', 'Middle Eastern Restaurant',
            'Korean Restaurant', 'Lounge', 'Golf Course', 'Mexican Restaurant',
            'Chinese Restaurant', 'Public Art', 'Health & Beauty Service'],
           dtype=object)
```

	Neighborhoods	Coffee Shop	Café
0	Bình Chánh District	4	7
1	Bình Thạnh District	2	7
2	Bình Tân District, Ho Chi Minh City	4	7
3	Cần Giờ District	2	8
4	Củ Chi District	0	9
5	District 1, Ho Chi Minh City	2	6
6	District 10, Ho Chi Minh City	3	6
7	District 11, Ho Chi Minh City	3	7
8	District 12, Ho Chi Minh City	3	7
9	District 3, Ho Chi Minh City	3	6
10	District 4, Ho Chi Minh City	2	6

Modeling

- Use K-Means algorithm to classify the data frame into 3 clusters



Conclusion

- Great number of coffee shop is located in center of Ho Chi Minh city (cluster 0 and 2). It is not good if we setup the new business here.
- At Cu Chi district, cluster 1, only coffee shops are operating. This is the right location for our customers to open the business.