CAPSTONE – IBM DATA SCIENCE

OPPORTUNITY FOR OPENNING PET COFFEE SHOP IN BINH TAN DISTRICT, HO CHI MINH CITY, VIETNAM

Khanh Trung Le

18-05-2018

A. Introduction

The growth rates of Vietnam in 2018 and 2019 are 7.08% and 7.02% respectively. This is a relatively high number in Southeast Asia in particular and the world in general.

In which, Ho Chi Minh City is considered as the economic leader of the whole country when the national budget contribution is at 25%.

At the same time, Ho Chi Minh City is also a place of modern and vibrant life thanks to the relatively high young population.

District 10 is one of the central districts of Ho Chi Minh City. There are Thanh Thai, February 3, To Hien Thanh and Su Van Hanh streets very lively with many types of entertainment, shopping, and entertainment.

My idea is to develop a coffee shop associated with pets like cats and dogs cafe in District 10. This will be a place where young people don't just come and chat with friends, but they can also have fun with these friendly animals. Science proves that people's moods will improve a lot after work pressure, life if they have positive experiences with cute pets.

B. Data Description

Google Cloud Platform API: I use service API called Geocoding API to find latitude and longitude; Reverse Geocoding API to find address based on latitude and longitude.

Foursquare API: I want to get nearby venues from circular places I made before. Then I filter venues that have categorie ID like Café, Cafeteria, Gaming Cafe, Pet Café,…

C. Methodology

First, I determine the longitude of Ho Chi Minh City. From there I draw the rhombus created by the small inner circle.

The main line consists of twenty circular areas of length 10000. I will access information about longitude, latitude, x y coordinates, distance. By using the Reverse Geocoding API, I found out the address information.

	latitude	longitude	distance_from_center	Address
0	10.768964	106.587490	37072.937380	63 Hẻm 55, Bình Trị Đông A, Bình Tân, Hồ Chí M...
1	10.768730	106.589336	36579.500363	3 Tây Lân, Bình Trị Đông A, Bình Tân, Hồ Chí M...
2	10.768496	106.591181	36107.021584	1087 QL1A, Bình Trị Đông A, Bình Tân, Hồ Chí M...
3	10.768262	106.593027	35656.334199	39-41, 16, Bình Trị Đông A, Bình Tân, Hồ Chí M...
4	10.768027	106.594872	35228.274573	953 HL2, Bình Trị Đông A, Bình Tân, Hồ Chí Min...

I explore venues near the addresses I have above. From those venues I get one hot category data.

Analyze each Neightborhood

```
# One hot encoding
firstneight_onehot = pd.get_dummies(firstneight_venues["Venue Category"], prefix="", prefix_sep="")

# Add neighborhood column to dataframe
firstneight_onehot["Neighborhood"] = firstneight_venues["Neighborhood"]
firstneight_onehot.head()
```

	Arts & Crafts Store	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Beer Garden	Bookstore	Breakfast Spot	Brewery	Bubble Tea Shop	Burger Joint	Café	Chinese Restaurant	Clothing Store	Coffee Shop	Comedy Club	Convenience Store	Department Store	Dessert Shop	Diner	Electronics Store	Fast Food Restaurant	Flower Shop	Food	Food & Drink Shop	Food Court	Food Truck	French Restaurant	Frozen Yogurt Shop
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

I created a dataframe about the common category of each address / neighborhood.

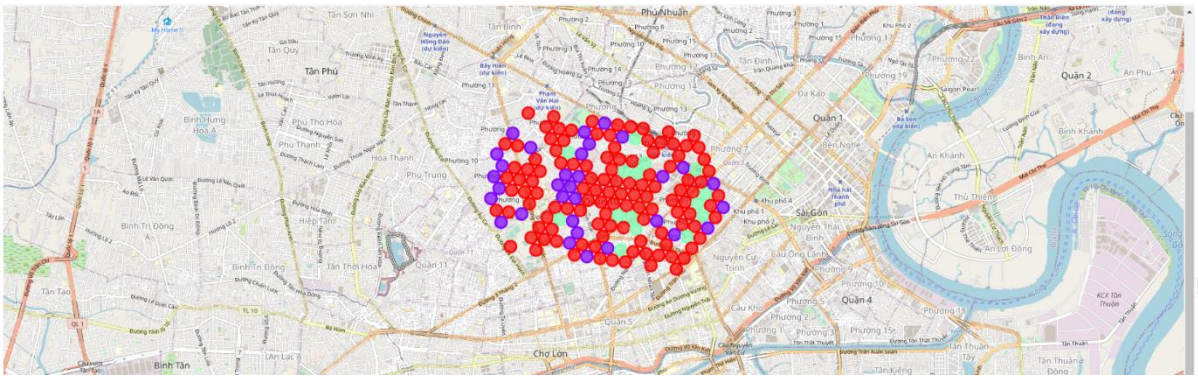
Create DataFrane after sorted by Most Common Category

```
neighborhoods_venues_sorted = pd.DataFrame(columns=cols)
neighborhoods_venues_sorted["Neighborhood"] = firstneight_grouped["Neighborhood"]
for i in np.arange(firstneight_grouped.shape[0]):
  neighborhoods_venues_sorted.iloc[i,1:] = return_most_common_categories(firstneight_grouped.iloc[i,:], num_top_venues)

neighborhoods_venues_sorted.head()
```

	Neighborhood	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category
0	10 Bắc Hải, Phường 15, Quận 10, Hồ Chí Minh, V...	Beer Garden	Vietnamese Restaurant	Food & Drink Shop	Dessert Shop	Diner
1	10/2 Đặng Lộ, Phường 7, Tân Bình, Hồ Chí Minh,...	Beer Garden	Vietnamese Restaurant	Food & Drink Shop	Dessert Shop	Diner
2	12 Hẻm 91 Hoà Hưng, Phường 12, Quận 10, Hồ Chí...	Coffee Shop	Vietnamese Restaurant	Food	Dessert Shop	Diner
3	125 Hoà Hưng, Phường 12, Quận 10, Hồ Chí Minh,...	Vegetarian / Vegan Restaurant	Coffee Shop	Food & Drink Shop	Diner	Electronics Store
4	131/12 Tô Hiến Thành, Phường 13, Quận 10, Hồ C...	Vietnamese Restaurant	Coffee Shop	Market	Furniture / Home Store	Frozen Yogurt Shop

Next I analyze Cluster with the number of cluster I chose 3. Here is the visual image.

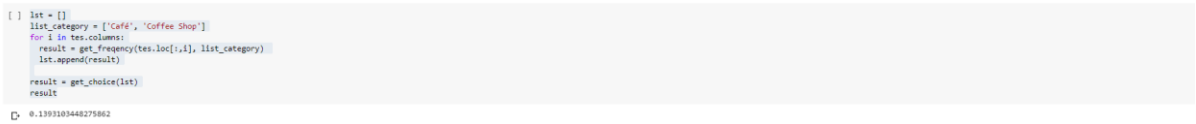


I made observations on each Cluster. I need to know that the

distribution of Café, Coffee Shop Category is in each Cluster. The goal is to find the cluster with the lowest density.

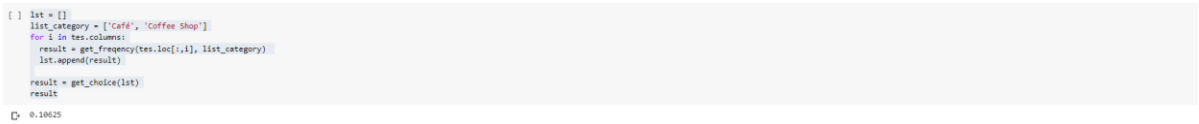
The analytical direction will be to find the frequency of Café, Coffee Shop Category in the columns from 1st common to 5th common and proceed to assign weights from high to low levels.

Cluster 1



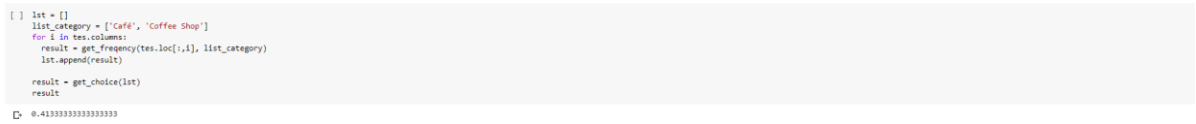
Result ~ 0.14

Cluster 2



Result ~ 0.1

Cluster 3



Result ~ 0.4

So I chose the second cluster.

Next I have a list of category IDs related to Café, Coffee Shop and Pet Cafe. I do a search for venues that are not pet cafe and venues are pet cafe.

The resulting json file returns and I proceed to filter the latitude longitude.

Not pet café venues

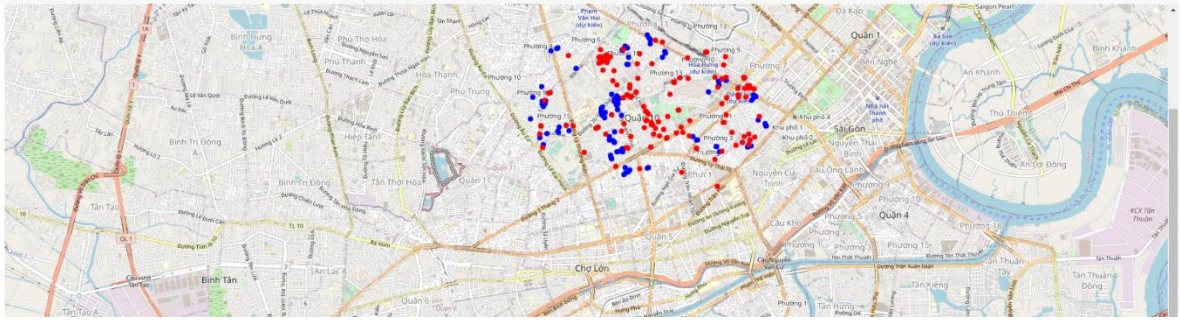
```
[ ] cafe_results
    [],
    [],
    [],
    [],
    [],
    [],
    [],
    [('4e50c45352b1e893ad6e46f0',
      'Cafe Football',
      [('Café', '4bf58dd8d48988d16d941735')],
      (10.7659061l, 106.668757),
      '234 Bà Hạt, Quận 10, Hồ Chí Minh, Việt Nam',
      61)],
    [('51d82db0498ebd04e0fd0fe5',
      'Milk shop (Hàng Phê)',
      [('Food & Drink Shop', '4bf58dd8d48988d1f9941735')],
      (10.765884399414062, 106.6685791015625),
      'Việt Nam',
      67)],
    [],
    [],
    [],
    [],
    [],
    [],
    [('50cc1b2fe4b0bd90982cf158',
      'Cafe Các',
      [('Café', '4bf58dd8d48988d16d941735')],
      (10.76493740081787l, 106.6651382446289),
      'Hẻm 583 Bà Hạt, Q10, Việt Nam',
      78),
     ('57cc0546498e5242eb75403f',
      'Nước mía PenTin',
      [('Café', '4bf58dd8d48988d16d941735')],
      (10.765351, 106.665855),
      'Thành phố Hồ Chí Minh, Việt Nam',
      13),
     ('51be509a498e40022f1c0a88',
      'Quan Kỳ Kỳ',
      [('Café', '4bf58dd8d48988d16d941735')],
      (10.765145, 106.665397),
      'Việt Nam',
```

Pet café venues

```
[ ] pet_cafe_results
      (10.76884479117253, 106.65320498559622),
      'Thiên Phước, Thành phố Hồ Chí Minh, Thành phố Hồ Chí Minh, Việt Nam',
      98)],
    [],
    [],
    [('51a025f3498eb546195cf16c',
      'H cafe',
      [('Café', '4bf58dd8d48988d16d941735')],
      (10.768129673372222, 106.65928168421578),
      '284/9 Lý Thường Kiệt, phường 14, quận 10, Hồ Chí Minh, Việt Nam',
      56)],
    [],
    [],
    [],
    [],
    [],
    [],
    [],
    [('4b923924f964a5201bed33e3',
      'Nhac Tre Coffee',
      [('Café', '4bf58dd8d48988d16d941735')],
      (10.765246260767658, 106.67364120483398),
      '50 Vinh Vien, D.10, Thành phố Hồ Chí Minh, Thành phố Hồ Chí Minh, Việt Nam',
      89)],
    [],
    [],
    [],
    [],
    [],
    [],
    [],
    [],
    [('5780f36e498e62d446084ff3',
      'The Coffee House',
      [('Café', '4bf58dd8d48988d16d941735')],
      (10.766124804318151, 106.66395287287132),
      '572 Ba tháng Hai, phường 14, quận 10, Việt Nam',
      70)],
    [],
    [('4f2ce24ee4b050b648972737',
```

Finally, I visualize that I found.

```
[ ] for i in pet_cafe_coor:
      lat = i[0]; lon = i[1]
      folium.CircleMarker([lat, lon], radius=3, fill=True, color='red', fill_opacity=1).add_to(map_cluster)
map_cluster
```



The blue dot is not pet cafe venues and vice versa the red dot.

D. Conclusion

The area created by February 3 road, Thanh Thai street, To Hien Thanh street and Ly Thuong Kiet street is located in ward 15 with the lowest density of pet cafe shop. It can be considered as a place for investment research.