Capstone Project — The Battle of Neighborhoods Santiago

By Aníbal Fernández

Introduction

Santiago is the capital city and the largest of Chile, as well as one of the largest cities in America. It is located in the center of Chile, and it is the most densely populated region, the Santiago Metropolitan Region, which total population is 7 million, in which more than 6 million live in the city's continuous urban area. Given the current circumstances due to the coronavirus, Santiago is detained at a commercial and gastronomic level, therefore it is a good opportunity to analyze and determine where could be a good place to put a restaurant, due the economic rise that will be generated at the end of the coronavirus emergency, added to its fabulous characteristics, such as being a city with the highest density and being the capital of the country.

Data Requirements

Datasource : https://es.wikipedia.org/wiki/Anexo:Comunas_de_Chile_por_poblaci%C3%B3n

Description: We will Scrap Santiago communes* table from Wikipedia and transform the coordinates of the major communes using in google maps. Using foursquare APIs we will get all the venues in each communes and then determine which communes are more attractive to locate a restaurant based on the different services and affluence nearly the commune.

*The communes are the minor and basic administrative division of Chile. They correspond to what in other countries is known as a municipality.

Methodology

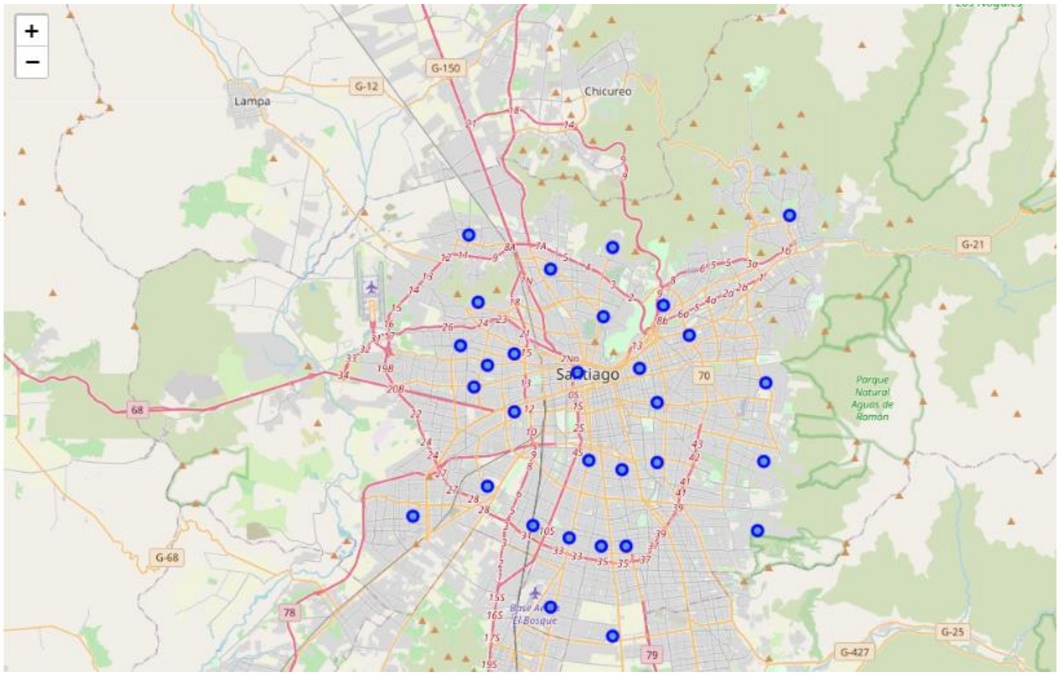
First, I used the table of the communes of Santiago in Wikipedia, and transform the latitud and longitud that is in degrees**,** minutes, seconds (DMS) coordinates to decimal degrees (dd) with Google maps and put the information in a Excel to import in a pandas data frame all the communes.

```
df = pd.read_excel('Santiago-comunas.xlsx')
df.head()
```

	Commune	Surface (km2)	Population	Density (hab./km2)	Latitud	Longitud	Latitude	Longitude
0	Santiago	23.2	404495	17435.1	-33°26'14"	-70°39'26"	-33.437222	-70.657222
1	Cerrillos	21.0	80832	3849.1	-33°30'0"	-70°43'0"	-33.500000	-70.716667
2	Cerro Navia	11.0	132622	12056.5	-33°25'19.2"	-70°44'6"	-33.422000	-70.735000
3	Conchalí	10.7	126955	11865.0	-33°22'48"	-70°40'30"	-33.380000	-70.675000
4	El Bosque	14.2	162505	11444.0	-33°34'1.2"	-70°40'30"	-33.567000	-70.675000

Then I used Python folium to create a map and visualize the

[5]:



Finaly we use foursquare API and get the top 100 venues that are near Providencia within a radius of 500 meter. In resume of the data 23 venues were returned by Foursquare.
Later we will concentrate only the restaurant and see the most frequently occurring venues

[12]:
```
print('{} venues were returned by Foursquare.'.format(nearby_venues.shape[0]))
```
```
23 venues were returned by Foursquare.
```

[13]:
```
print ('{} unique categories in Providencia'.format(nearby_venues['categories'].value_counts().shape[0]))
```
```
18 unique categories in Providencia
```

[14]:
```
print (nearby_venues['categories'].value_counts()[0:10])
```
```
Pizza Place                    4
Sandwich Place                 2
Peruvian Restaurant            2
Pool Hall                      1
Gym                            1
Southern / Soul Food Restaurant    1
South American Restaurant      1
Gastropub                      1
Coffee Shop                    1
Tattoo Parlor                  1
Name: categories, dtype: int64
```

Then we look the most popular type of restaurant and which communes have more restaurants.

[20]:
```
Santiago_5_Commune_Venues_Top10 = Santiago_Venues_only_restaurant['Venue Category'].value_counts()[0:10].to_frame(name='frequency')
Santiago_5_Commune_Venues_Top10=Santiago_5_Commune_Venues_Top10.reset_index()
#Santiago_5_Comm_Venues_Top10

Santiago_5_Commune_Venues_Top10.rename(index=str, columns={"index": "Venue_Category", "frequency": "Frequency"}, inplace=True)
Santiago_5_Commune_Venues_Top10
```

[20]:

	Venue_Category	Frequency
0	Restaurant	20
1	Chinese Restaurant	19
2	Sushi Restaurant	18
3	Peruvian Restaurant	13
4	Fast Food Restaurant	9
5	South American Restaurant	8
6	Japanese Restaurant	6
7	Italian Restaurant	6
8	Seafood Restaurant	4
9	Asian Restaurant	4

[24]:
```
Santiago_Venues_restaurant
```

[24]:
```
Commune
Cerrillos            2
Conchalí             1
Estación Central     3
La Cisterna          6
La Pintana           1
La Reina             3
Las Condes          11
Lo Barnechea         5
Lo Prado             5
Macul                3
Maipú                2
Peñalolén            2
Providencia          6
Pudahuel             2
Quilicura            7
Recoleta             2
San Miguel          10
Santiago            27
Vitacura            15
Ñuñoa                9
Name: Venue Category, dtype: int64
```

Then we analyze each commune to know about the top 5 venues of each one. We create a dataframe with pandas one hot encoding for the categories

[26]:
```
# one hot encoding
Santiago_onehot = pd.get_dummies(Santiago_Venues_only_restaurant[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
Santiago_onehot['Commune'] = Santiago_Venues_only_restaurant['Commune']

# move neighborhood column to the first column
fixed_columns = [Santiago_onehot.columns[-1]] + list(Santiago_onehot.columns[:-1])
Santiago_onehot = Santiago_onehot[fixed_columns]

Santiago_onehot.head()
```

[26]:

	Commune	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Chinese Restaurant	Falafel Restaurant	Fast Food Restaurant	French Restaurant	Greek Restaurant	Indian Restaurant	...	Latin American Restaurant	Mediterranean Restaurant	Peruvian Restaurant	Restaurant	Seafood Restaurant	South American Restaurant	Southern / Soul Food Restaurant	F
1	Santiago	0	0	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0	
2	Santiago	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	
3	Santiago	0	0	0	0	0	0	0	0	0	...	0	0	1	0	0	0	0	
4	Santiago	0	0	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0	
5	Santiago	0	0	0	0	0	0	0	0	0	...	0	0	1	0	0	0	0	

5 rows × 22 columns

[27]:
```
Santiago_onehot.shape
```

[27]: (122, 22)

Using pandas groupby on neighborhood column and calculate the mean of the frequency of occurrence of each venue category

```
[28]: Santiago_grouped = Santiago_onehot.groupby('Commune').mean().reset_index()
Santiago_grouped
```

[28]:

	Commune	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Chinese Restaurant	Falafel Restaurant	Fast Food Restaurant	French Restaurant	Greek Restaurant	Indian Restaurant	...	Latin American Restaurant	Mediterranean Restaurant	Peruvian Restaurant	Restaurant	Seafood Restaurant	South American Restaurant	Southern / Soul Food Restaurant
0	Cerrillos	0.000000	0.000000	0.000000	0.000000	0.000000	0.500000	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.500000	0.000000	0.000000	0.000000
1	Conchalí	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	Estación Central	0.000000	0.333333	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.333333	0.000000	0.000000	0.000000
3	La Cisterna	0.000000	0.000000	0.000000	0.333333	0.000000	0.166667	0.000000	0.000000	0.000000	...	0.0	0.000000	0.166667	0.000000	0.000000	0.000000	0.000000
4	La Pintana	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	La Reina	0.000000	0.000000	0.000000	0.333333	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Las Condes	0.000000	0.000000	0.000000	0.090909	0.000000	0.272727	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.454545	0.000000	0.000000	0.000000
7	Lo Barnechea	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.200000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.200000	0.000000	0.000000	0.000000
8	Lo Prado	0.000000	0.000000	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	Macul	0.000000	0.000000	0.000000	0.333333	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.666667	0.000000	0.000000	0.000000
10	Maipú	0.000000	0.000000	0.500000	0.000000	0.000000	0.500000	0.000000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Then output each neighborhood along with the top 5 most common venues:

```
[32]: num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Commune']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
Commune_venues_sorted = pd.DataFrame(columns=columns)
Commune_venues_sorted['Commune'] = Santiago_grouped['Commune']

for ind in np.arange(Santiago_grouped.shape[0]):
    Commune_venues_sorted.iloc[ind, 1:] = return_most_common_venues(Santiago_grouped.iloc[ind, :], num_top_venues)

Commune_venues_sorted.head(23)
```

[32]:

	Commune	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Cerrillos	Restaurant	Fast Food Restaurant	Thai Restaurant	Italian Restaurant	Argentinian Restaurant	Asian Restaurant	Chinese Restaurant	Falafel Restaurant	French Restaurant	Greek Restaurant
1	Conchalí	Fast Food Restaurant	Thai Restaurant	Italian Restaurant	Argentinian Restaurant	Asian Restaurant	Chinese Restaurant	Falafel Restaurant	French Restaurant	Greek Restaurant	Indian Restaurant
2	Estación Central	Japanese Restaurant	Argentinian Restaurant	Restaurant	Italian Restaurant	Asian Restaurant	Chinese Restaurant	Falafel Restaurant	Fast Food Restaurant	French Restaurant	Greek Restaurant
3	La Cisterna	Sushi Restaurant	Chinese Restaurant	Peruvian Restaurant	Fast Food Restaurant	Thai Restaurant	Indian Restaurant	Argentinian Restaurant	Asian Restaurant	Falafel Restaurant	French Restaurant
4	La Pintana	Sushi Restaurant	Thai Restaurant	Italian Restaurant	Argentinian Restaurant	Asian Restaurant	Chinese Restaurant	Falafel Restaurant	Fast Food Restaurant	French Restaurant	Greek Restaurant
5	La Reina	Italian Restaurant	Sushi Restaurant	Chinese Restaurant	Thai Restaurant	Argentinian Restaurant	Asian Restaurant	Falafel Restaurant	Fast Food Restaurant	French Restaurant	Greek Restaurant

Finally, we try to cluster the communes based on the venue categories and use K-Means clustering that they learn previously.

```
[33]: # set number of clusters
kclusters = 5

Santiago_grouped_clustering = Santiago_grouped.drop('Commune', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(Santiago_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
[33]: array([4, 0, 4, 1, 2, 1, 4, 1, 2, 4], dtype=int32)
```

```
[34]: # add clustering labels

Commune_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

Santiago_merged = Santiago

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
Santiago_merged = Santiago_merged.join(Commune_venues_sorted.set_index('Commune'), on='Commune')

Santiago_merged# check the last columns!
```

[34]:

	Commune	Surface (km2)	Population	Density (hab./km2)	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Santiago	23.2	404495	17435.1	-33.437222	-70.657222	1.0	Peruvian Restaurant	Japanese Restaurant	Sushi Restaurant	Asian Restaurant	Chinese Restaurant	South American Restaurant	Restaurant	Italian Restaurant	Falafel Restaurant	Fast Food Restaurant
1	Cerrillos	21.0	80832	3849.1	-33.500000	-70.716667	4.0	Restaurant	Fast Food Restaurant	Thai Restaurant	Italian Restaurant	Argentinian Restaurant	Asian Restaurant	Chinese Restaurant	Falafel Restaurant	French Restaurant	Greek Restaurant
2	Cerro Navia	11.0	132622	12056.5	-33.422000	-70.735000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	Conchalí	10.7	126955	11865.0	-33.380000	-70.675000	0.0	Fast Food Restaurant	Thai Restaurant	Italian Restaurant	Argentinian Restaurant	Asian Restaurant	Chinese Restaurant	Falafel Restaurant	French Restaurant	Greek Restaurant	Indian Restaurant
4	El Bosque	14.2	162505	11444.0	-33.567000	-70.675000	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	Estación Central	15.0	147041	9802.7	-33.459000	-70.699000	4.0	Japanese Restaurant	Argentinian Restaurant	Restaurant	Italian Restaurant	Asian Restaurant	Chinese Restaurant	Falafel Restaurant	Fast Food Restaurant	French Restaurant	Greek Restaurant

Results

We find a interesting insight and information based on the previous analysis.

The 3 communes that have more restaurants are Santiago city center (27), Vitacura (15), Las Condes (11).

The largest frequency of restaurant are restaurant (not classified as a specific cousine), Chinese restaurant and sushi restaurant.

We not have information about Cerro Navia, El bosque, Huechuraba, Independecia, La Florida, La Granja, Lo espejo, Pedro Aguirre Cerda, Quinta Normal, Renca, San Joaquin, San Ramon. Some of this are the communes of Santiago with the most poor income but that doesn´t mean that they are not suitable place to put a restaurant in term of population.

The commune of Santiago is in the center of the city, it is the most populate area, here are the patrimonial architecture buildings and government palace, so is the most connected area with more education establishments and private companies.

The clustering and the information result of the analysis are completely based from Foursquare data, for that reasons the result of the analysis have some bias, because it depends on other people who comment on the platform for us to use it.

Conclusion

Finaly we determine that the most tentative communes to put a restaurant are Santiago, Vitacura and Las condes. This information was taken by clustering the communes in Santiago City and is based on most common food restaurants. But some communes don't have any info in foursquare, but in any case this should be a reason to be discarded like La Florida that has a population of 366916 n and maybe it is a commune that can be attractive to explore or Huechuraba for instance that is a new business area in the city.