The Battle of the Neighborhoods

The Problem

- There are too many places in Lima, where a restaurant can be opened but too little information.

- This project aims to suggest the best District to open a Burger Joint based on the type of public around.

Data Sources & Cleaning

Data sources

We scrapped the Districts from Wikipedia. Also, we got the geolocalization of some category venues from the Foursquare API.

Data cleaning

The data scrapped from Wikipedia had some errors and some useless information within the cells.

Then, the data from Foursquare was pretty clear. It was not necessary to do any cleaning to the data.

Districts of Lima scrapped from Wikipedia

	Distritos	Ubigeo	area	densidad	CódigoPostal	poblacion
0	Ancón	150102	299.22	98.30	02	29413
1	Ate	150103	77.72	5399.70	03	419664
2	Barranco	150104	3.33	13790.40	04	45922
3	Breña	150105	3.22	29443.50	05	94808
4	Carabayllo	150106	346.88	544.20	06	188772
5	Chaclacayo	150107	39.50	1004.70	08	39685
6	Chorrillos	150108	38.94	6743.60	09	262595
7	Cieneguilla	150109	240.33	65.70	40	15789
8	Comas	150110	48.75	9533.20	07	464743
9	El Agustino	150111	12.54	13191.80	10	165425
10	Huaycán[4]	-	27.40	5839.40	-	159999
11	Independencia	150112	14.56	13551.40	28	197308
12	Jesús María	150113	4.57	12820.10	11	58587
13	La Molina	150114	65.75	1893.00	12	124464
14	La Victoria	150115	8.74	21764.10	13	190218
15	Lima	150101	21.88	13187.20	01	288535
16	Lince	150116	3.03	17202.30	14	52122

Then we added the Lat and Lon

```
In [17]: df
```

Out[17]:

	Distritos	Ubigeo	area	densidad	CódigoPostal	poblacion	LAT	LON
0	Ancón	150102	299.22	98.30	02	29413	-11.71	-77.1255
1	Ate	150103	77.72	5399.70	03	419664	-12.0267	-76.8896
2	Barranco	150104	3.33	13790.40	04	45922	-12.1437	-77.019
3	Breña	150105	3.22	29443.50	05	94808	-12.0585	-77.0507
4	Carabayllo	150106	346.88	544.20	06	188772	-11.7929	-76.9873
5	Chaclacayo	150107	39.50	1004.70	08	39685	-11.9948	-76.7683
6	Chorrillos	150108	38.94	6743.60	09	262595	-12.1876	-77.0078
7	Cieneguilla	150109	240.33	65.70	40	15789	-12.0948	-76.7567
8	Comas	150110	48.75	9533.20	07	464743	-11.93	-77.0535
9	El Agustino	150111	12.54	13191.80	10	165425	-12.048	-77.0007
10	Huaycán[4]	-	27.40	5839.40	-	159999	-12.0372	-77.0036
11	Independencia	150112	14.56	13551.40	28	197308	-11.9929	-77.0518
12	Jesús María	150113	4.57	12820.10	11	58587	-12.075	-77.0435
13	La Molina	150114	65.75	1893.00	12	124464	-12.082	-76.9282
14	La Victoria	150115	8.74	21764.10	13	190218	-12.0734	-77.0163
15	Lima	150101	21.88	13187.20	01	288535	-12.0464	-77.0428
16	Lince	150116	3.03	17202.30	14	52122	-12.0853	-77.0358

We got all the burger locations in Lima

In [23]:
```
# Use category id 4bf58dd8d48988d16c941735 to only get the burger joints
lima_venues_burger = getNearbyVenues(names=df['Distritos'], latitudes=df['LAT'], longitudes=df[
```

In [24]:
```
lima_venues_burger.head()
```

Out[24]:

	Localidad	Localidad Latitude	Localidad Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ate	-12.026700	-76.889584	Donde Walter Fast Food	-12.017339	-76.884570	Burger Joint
1	Ate	-12.026700	-76.889584	McDonald's	-12.022752	-76.881450	Burger Joint
2	Barranco	-12.143727	-77.019023	939 Burger House	-12.142231	-77.018110	Burger Joint
3	Barranco	-12.143727	-77.019023	brabazo	-12.143110	-77.018407	Burger Joint
4	Barranco	-12.143727	-77.019023	Coco loco	-12.141498	-77.020746	Burger Joint

Then, we got the locations of Schools, Universities, Offices and Nightlife places. In these places, we found a most of the potential customers.

In [33]:
```
lima_venues_office = getNearbyVenues(names=df['Distritos'], latitudes=df['LAT'], longitudes=df[
lima_venues_office.head()
```

Out[33]:

	Localidad	Localidad Latitude	Localidad Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ate	-12.0267	-76.889584	Colegio San Alfonso	-12.024123	-76.885875	High School
1	Ate	-12.0267	-76.889584	Laive	-12.027305	-76.893080	Factory
2	Ate	-12.0267	-76.889584	Capilla Colegio San Alfonso	-12.024633	-76.885929	Church
3	Ate	-12.0267	-76.889584	Industria Metalica Bullon sac	-12.026699	-76.892495	Factory
4	Ate	-12.0267	-76.889584	veterinaria car val	-12.027685	-76.893272	Veterinarian

In [36]:
```
lima_venues_nightlife = getNearbyVenues(names=df['Distritos'], latitudes=df['LAT'], longitudes=
lima_venues_nightlife.head()
```

Out[36]:

	Localidad	Localidad Latitude	Localidad Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ate	-12.026700	-76.889584	Abuela's Place	-12.017569	-76.891484	Bar
1	Barranco	-12.143727	-77.019023	Sargento Pimienta	-12.143782	-77.018755	Bar
2	Barranco	-12.143727	-77.019023	Red Cervecera Perú	-12.144339	-77.018878	Brewery
3	Barranco	-12.143727	-77.019023	Limbo Bar	-12.143327	-77.018503	Dive Bar
4	Barranco	-12.143727	-77.019023	Ex Teatro Salaverry	-12.146431	-77.018875	Nightclub

The data is merged in one table

Localidad	Ubigeo	area	densidad	CódigoPostal	poblacion	LAT	LON	Burger	High Schools	Universities	Offices	Ni
Ancón	150102	299.22	98.30	02	29413	-11.71	-77.1255	0.0	0.0	0.0	0.0	
Ate	150103	77.72	5399.70	03	419664	-12.0267	-76.8896	2.0	1.0	0.0	22.0	
Barranco	150104	3.33	13790.40	04	45922	-12.1437	-77.019	29.0	8.0	3.0	50.0	
Breña	150105	3.22	29443.50	05	94808	-12.0585	-77.0507	10.0	14.0	9.0	50.0	
Carabayllo	150106	346.88	544.20	06	188772	-11.7929	-76.9873	0.0	0.0	0.0	0.0	
Chaclacayo	150107	39.50	1004.70	08	39685	-11.9948	-76.7683	0.0	0.0	0.0	0.0	
Chorrillos	150108	38.94	6743.60	09	262595	-12.1876	-77.0078	4.0	4.0	5.0	48.0	
Cieneguilla	150109	240.33	65.70	40	15789	-12.0948	-76.7567	0.0	0.0	0.0	0.0	
Comas	150110	48.75	9533.20	07	464743	-11.93	-77.0535	2.0	1.0	1.0	43.0	
El Agustino	150111	12.54	13191.80	10	165425	-12.048	-77.0007	6.0	1.0	0.0	47.0	
Huaycán[4]	-	27.40	5839.40	-	159999	-12.0372	-77.0036	3.0	0.0	0.0	47.0	
Independencia	150112	14.56	13551.40	28	197308	-11.9929	-77.0518	5.0	2.0	6.0	49.0	
Jesús María	150113	4.57	12820.10	11	58587	-12.075	-77.0435	19.0	9.0	29.0	50.0	
La Molina	150114	65.75	1893.00	12	124464	-12.082	-76.9282	1.0	0.0	1.0	45.0	

A weight is given to each Category Venue

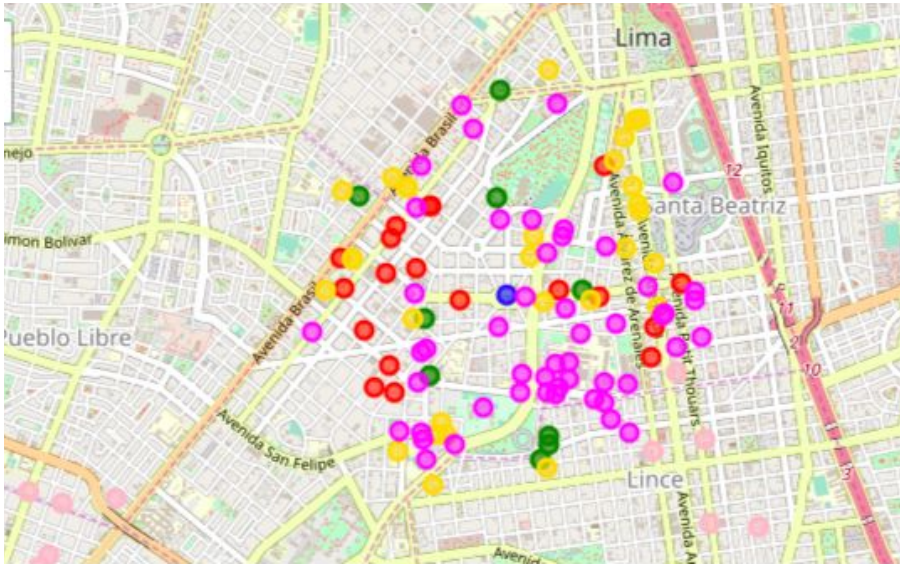
In [43]:
```
# negative weight, because Jeronimo wants to open a burger joint and thus wants to avoid concur
weight_burger = -2

# positive weight, because high school students are good customers
weight_schools = 1

# positive weight, because uni students are good customers
weight_uni = 1.5

# positive weight because employees are even better customers
weight_offices = 1.75

# positive weight because nighlifers are the best customers
weight_offices = 2
```

We found that Jesus María District is the best place to open a Burger Joint

In [46]:
```
df_weighted['Score'] = df['Burger'] * weight_burger + df['High Schools'] * weight_schools + df[
df_weighted = df_weighted.sort_values(by=['Score'], ascending=False)
df_weighted
```

Out[46]:

	Localidad	Score
12	Jesús María	114.5
15	Lima	112.5
3	Breña	107.5
6	Chorrillos	99.5
31	San Isidro	99.5
42	Surquillo	99.0
11	Independencia	99.0
21	Miraflores	96.5
17	Los Olivos	95.5
38	Santa María de Huachipa	93.5
14	La Victoria	93.5
30	San Borja	93.5



Recommendations

- Do some research to found if the informal burger streets sellers are considered in the Foursquare API.

- Research if it is possible to drill down in the geolocalization and found the best Zip Code with the new peruvian zip code format.