Applied Data Science Capstone

Final assingment: The Battle of Neighborhoods

Name: Fernando Paolo Cortez Silva

June 2019

Índex

I. Introduction

Peru, is a country located in western South America. It is bordered in the north by Ecuador and Colombia, in the east by Brazil, in the southeast by Bolivia, in the south by Chile, and in the west by the Pacific Ocean.

Tourism in Peru is one of its principal sources of income, and has a significantly impact on PIB. In first place it is directed towards archaeological monuments (like the ones in Cusco), ecotourism in the Peruvian Amazon, cultural tourism in colonial cities, gastronomic tourism and many more. According to a Peruvian government study, the satisfaction rate for tourists after visiting Peru is 94%. Tourism is the most rapidly growing industry in Peru, growing annually at a rate of 25%.

Although tourism is mainly directed for cities outside Lima (the capital of Peru), an important part of tourist has to pass or stay in Lima in order to arrive or leave the country. One popular district for tourist to stay in Lima is Miraflores.

Miraflores is probably the most popular district in Lima from a tourist's perspective, featuring gorgeous coastal views, quality shopping, and world-class food.

II. Business Problem

When traveling for the first time to places around the world one problem tourists has to face is food. In the case of traveling to Peru many tourists will have to stay in Lima, probably Miraflores and it will be very useful if they know what type of restaurants they will find.

So, for this project I am comparing options of food of a common city like Los Angeles in the USA against Miraflores in Lima, Peru.

The main objective is to find out what type of food they will find in each city and how many options they will have.

III. Data description

In order to solve the problem of this case we need information about the restaurants around Miraflores and Los Angeles. More precisely, the number of existing restaurants in each location and their type

Following data sources will be needed to extract the required information:

- Centers districs will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding.
- Number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API.

IV. Methodology section

1. Objective

In this project we will direct our efforts on comparing with some basic statistics and visualizations food options between two cities: Miraflores and Los Angeles. Is important to note that the radius of search was 2 km around the given latitude and longitude.

2. Data collection

The first step was data collection, which consisted 1) in using a geolocator in order to obtain the latitude and longitude of a given address. 2) search for places related with food on Foursquare given some parameters and download the information on a JSON file.

This is the code I used to get the latitude and longitude of the address Miraflores Lima:

```
[ ]: address = 'Miraflores Lima'

geolocator = Nominatim(user_agent="foursquare_agent")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
```

Then we search for food on Foursquare

```
[4]: search_query = 'food'
     radius = 2000
     print(search_query + ' .... OK!')

     food .... OK!

[5]: lient_id={}&client_secret={}&ll={},{}&v={}&query={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, latitude, longitude, VERSION, search_query, radius, LIMIT)
```

Then we retrieve the data from the JSON file

```
[6]: results = requests.get(url).json()

[7]: # assign relevant part of JSON to venues
     venues = results['response']['venues']
```

3. Data transformation

The second step is data transformation, which consisted in converting the JSON file into a dataframe in order to use pandas library on it. For our analysis we only need two columns: name of the restaurant and its type. We finally group our dataframet by type of restaurant and add a counter (the number of restaurant types).

This is the code for transforming the JSON file into a dataframe:

```
# tranform venues into a dataframe
dataframe = json_normalize(venues)
dataframe.head()
```

[7]:

	categories	hasPerk	id	location.address	location.cc	location.city	location.country	location.crossStreet	location.distance	location.formattedAddres
0	[{'id': '4bf58dd8d48988d120951735', 'name': 'F...	False	4de67b75185067eb63f6b323	Benavides	PE	Miraflores	Perú	Alcanfores	411	[Benavides (Alcanfores), Miraflores, Perú
1	[{'id': '4bf58dd8d48988d120951735', 'name': 'F...	False	4fdbd90de4b0f18e76ff7b5a	Malecón de la Reserva 610	PE	Miraflores	Perú	NaN	1295	[Malecón de la Reserva 610, Miraflores, Perú
2	[{'id': '4bf58dd8d48988d1c4941735', 'name': 'R...	False	5ac5d5d4d033605386cb04b0	Calle Schell 385	PE	Miraflores	Perú	NaN	290	[Calle Schell 385, Miraflores, 15074, Perú
3	[{'id': '4bf58dd8d48988d16e941735', 'name': 'F...	False	5b0e15fa4acb19002c75b0ad	Jiron San Diego 261, Surquillo	PE	Lima	Perú	NaN	390	[Jiron San Diego 261, Surquillo, Lima, 15048, ...
4	[{'id': '50aa9e744b90af0d42d5de0e', 'name': 'H...	False	53a1fc20498eed61f3b47a4c	Ernesto Diez Canseco 316	PE	Miraflores	Perú	NaN	402	[Ernesto Diez Canseco 316, Miraflores, Perú

We only keep information related with location

[8]:
```
# keep only columns that include venue name, and anything that is associated with location
filtered_columns = ['name', 'categories'] + [col for col in dataframe.columns if col.startswith('location.')] + ['id']
dataframe_miraflores = dataframe.loc[:, filtered_columns]

# function that extracts the category of the venue
def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']

    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']

# filter the category for each row
dataframe_miraflores['categories'] = dataframe_miraflores.apply(get_category_type, axis=1)

# clean column names by keeping only last term
dataframe_miraflores.columns = [column.split('.')[-1] for column in dataframe_miraflores.columns]

dataframe_miraflores.head()
```

[8]:

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng	neighborhood	postalCode	state	
0	Food Court Vivanda	Food Court	Benavides	PE	Miraflores	Perú	Alcanfores	411	[Benavides (Alcanfores), Miraflores, Perú]	[{'label': 'display', 'lat': -12.1244250583423...	-12.124425	-77.028215	NaN	NaN	Lima	4c
1	Food Court - Larcomar	Food Court	Malecón de la Reserva 610	PE	Miraflores	Perú	NaN	1295	[Malecón de la Reserva 610, Miraflores, Perú]	[{'label': 'display', 'lat': -12.1325626847085...	-12.132563	-77.029608	NaN	NaN	Lima	4
2	Panhela - Healthy Food	Restaurant	Calle Schell 385	PE	Miraflores	Perú	NaN	290	[Calle Schell 385, Miraflores, 15074, Perú]	[{'label': 'display', 'lat': -12.123137, 'lng'...	-12.123137	-77.027980	NaN	15074	Municipalidad Metropolitana de Lima	5a

We only select two columns for the analysis and the change its names:

```
df_LA=dataframe_LA[['name','categories']]
df_M=dataframe_miraflores[['name','categories']]
```

Finally, we group by type of restaurant and add a counter.

```
df_M.rename(columns={'name':'Restaurant name','categories':'Type of restaurant'},inplace=True)
df_M_G=df_M.groupby(['Type of restaurant']).count()
```

4. Visualizations

The third step is data visualization, we used three types of it, the first one was displaying all the restaurants over a map in order to see the density. The second one was a bar chart in order to see which types of restaurants were the most and least common. Finally, the third one is a pie chart which let us know the proportion of the 5 main restaurants in each city.

Here is the code I used for displaying the places over the map:

```
venues_map = folium.Map(location=[latitude, longitude], zoom_start=13) # generate map centred around Miraflores

# add a red circle marker to represent the center of Miraflores
folium.features.CircleMarker(
    [latitude, longitude],
    radius=10,
    color='red',
    popup='Miraflores',
    fill = True,
    fill_color = 'red',
    fill_opacity = 0.6
).add_to(venues_map)

# add the Italian restaurants as blue circle markers
for lat, lng, label in zip(dataframe_miraflores.lat, dataframe_miraflores.lng, dataframe_miraflores.categories):
    folium.features.CircleMarker(
        [lat, lng],
        radius=5,
        color='blue',
        popup=label,
        fill = True,
        fill_color='blue',
        fill_opacity=0.6
    ).add_to(venues_map)

# display map
venues_map
```

Here is the code I used for creating a horizontal bar chart:

Bar chart for Los Angeles

```
df_M_G.plot(kind='barh', figsize=(10, 6))

plt.xlabel('Restaurant name') # add to x-label to the plot
plt.ylabel('Quantity of restaurants') # add y-label to the plot
plt.title('Types of restaurants in Miraflores') # add title to the plot

plt.show()
```

Here is the code I used for creating a pie chart of the 5 main types of restaurants:

Pie chart for Miraflores

```
# autopct create %, start angle represent starting point
dfM['Quantity of restaurants'].plot(kind='pie',
                            figsize=(5, 6),
                            autopct='%1.1f%%', # add in percentages
                            startangle=90,     # start angle 90°
                            shadow=True,       # add shadow
                            )

plt.title('Restaurants in Miraflores')
plt.axis('equal') # Sets the pie chart to look like a circle.

plt.show()
```

V. Results section

In this section I will show the information retrieved from Foursquare.

1. List of results

The first one is a simple list of the names of food stores that were found. On the following picture we can see the food stores in Miraflores. We found 26 options where eat in Lima.

```
dataframe_miraflores.name
```

```
0                  Food Court Vivanda
1              Food Court - Larcomar
2             Panhela - Healthy Food
3              Chepulino's Fast Food
4     TIKA - Peruvian Food Survenirs
5               Fitness Protein Food
6              Spice Food and Drinks
7              Food Truck Codornices
8                       Food Rockers
9                           Fit+Food
10                      Protein Food
11             Food court Open Plaza
12               Food Truck Aramburu
13                 Peru Natural Food
14                         Tarboüsh
15              fast food open plaza
16                    Begonias food
17             Big Bro - food trucks
18             Cordeone Typical Food
19                       Afgan food
20                      Boulevard 99
21                   Sopas Josefina
22                              KFC
23                La Industria Foods
24    Mondelez Peru (ex Kraft Foods)
25                          DeliFood
Name: name, dtype: object
```

On the other hand, these are the options found in Los Angeles. A total of 50 options were found in Los Angeles

```
dataframe_LA.name
```

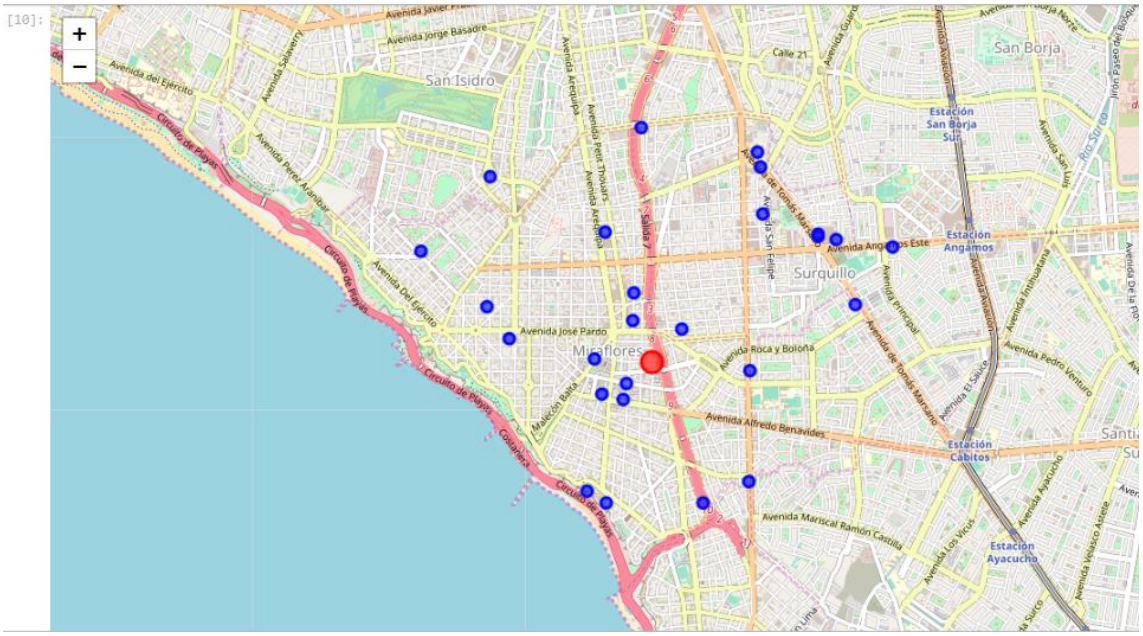
```
0                    Kabab and More Middle Eastern Food
1                                Lexus @ LA Food & Wine
2                                Artwalk Food Truck Lot
3                                   Broadway Food Court
4                                Holy Grill Food Truck
5                                 AB Chinese Fast Food
6                                  Heritage Food Truck
7                                      Pop's Food Mart
8                              Chunky Chiller Food Truck
9                                      Junk Food Clothing
10                  Begian's Catering and Food Service
11                Begins Cafe Catering and Food Service
12                                             S.L. food
13                               Cuchifritos Food Truck
14                             Just Food For Dogs (DTLA)
15                               Doña Estela Food Truck
16                                    Olga's Food Truck
17                             Corporation Food Hall
18                                          Famex Food
19                                La Times Food Bowl
20                Milk and Eggs - Farm & Food Delivery
21                                          Food Court
22                                   WILD Living Food
23                                Tem Pura Food Truck
24                           Slammin' Sliders Food Truck
25                                   Beyond Food Mart
26                                 Food, Fizz & Film
27                             Tokyo Doggies Food Truck

28                     Superior Snacks and Food Service
29                                       AAA Safe Food
30                               Canton Food Co.Warehouse
31                                     Just Food For Dogs
32                                  Blue Nova Food Truck
33                                      thai street food
34                          Ralphs Food 4 Less Auditorium
35    Street Food Cinema @ Los Angeles State Histori...
36                                 Disgusting Food Museum
37                                          Food Court
38                       Food Court Plaza - 818 Wilshire
39                                   L T Food Service
40                                        A Food Coma
41                                  Daniel's Food Truck
42                                            Food 630
43                                  Whole Foods Market
44                                    $1 Chineese Food
45                                             KE FOOD
46                           Game On! Gourmet Food Truck
47                           Prince of Venice Food Truck
48                                alameda food truck lot
49                              International Food Court
Name: name, dtype: object
```

Another observation is that Los Angeles has almost twice of food options than Miraflores.

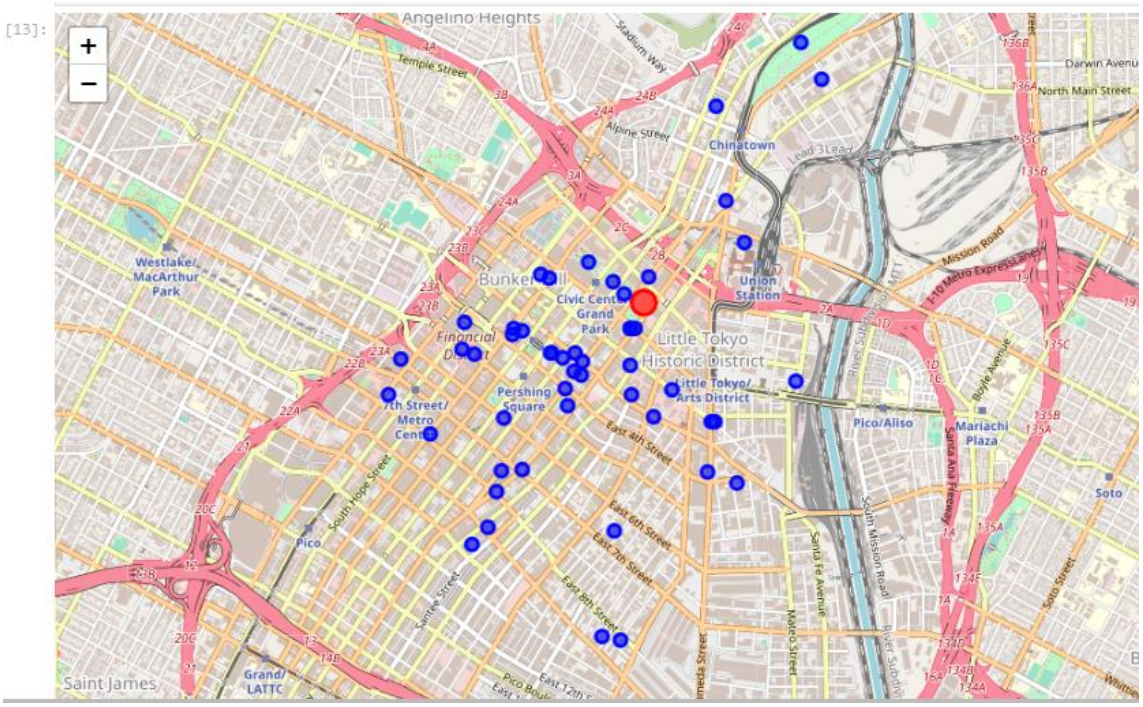
2. Maps

On the following images we can see the restaurants listed before on their respective map.

Miraflores map of food stores:



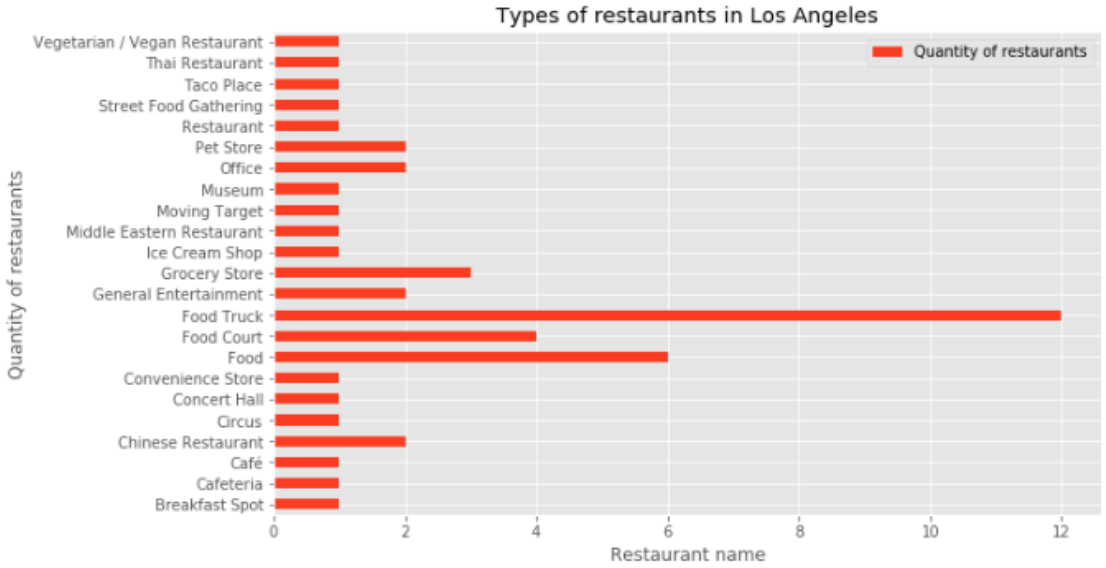
Los Angeles map of food stores



From the previous maps we can note that Los Angeles is denser than Miraflores in terms of restaurants over the same area (2 km in radius).

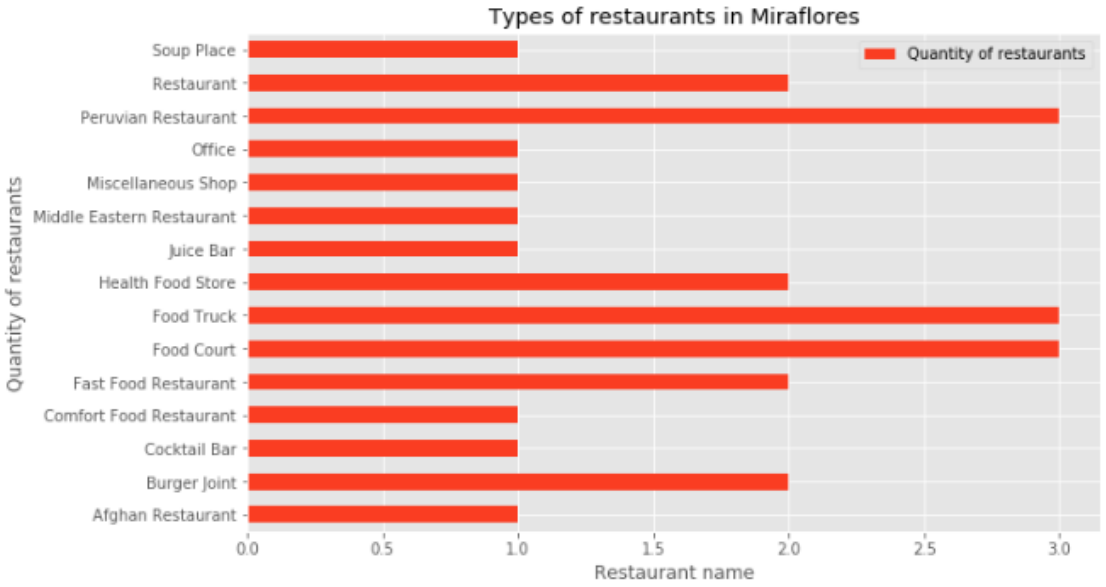
3. Bar chart

Then I used a bar chat to show the frequency of each type of restaurant. The first one is for Los Angeles which has 23 different types of food stores.



From it we can se it has a lot of food trucks (12) and food courts (5). We can note that one of its types does not show enough information (type: Food).

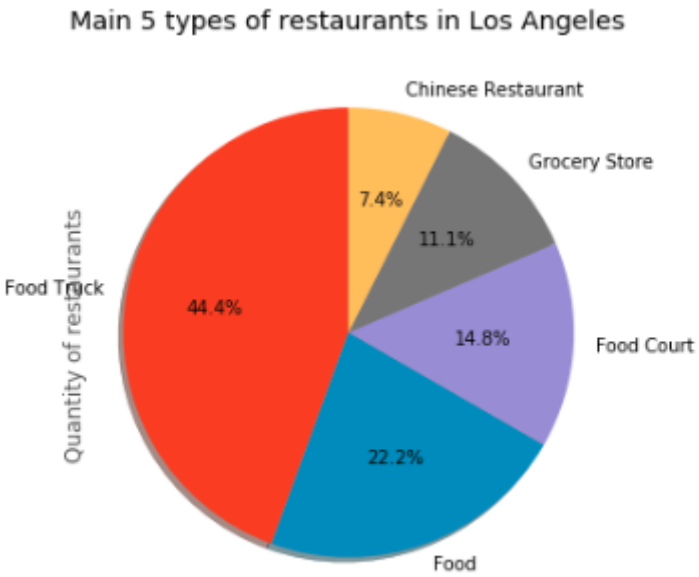
The second bar chart is from Miraflores, which has 15 different types of food stores.

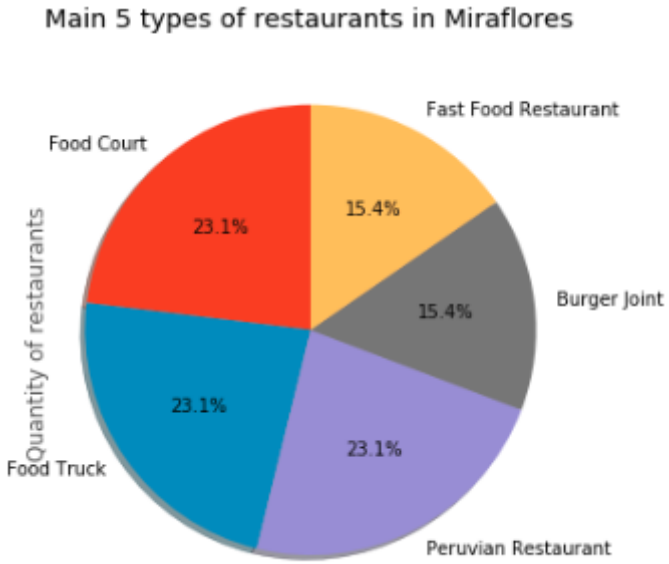


From it we note that Miraflores has a 3 main types of restaurants: Peruvian Restaurant, Food Truck and Food Court, also you can find a Health Food Store.

4. Pie chart

For our pie charts I used the 5 main type of food restaurants in order to see the proportion of each one.





From it we can note that Food Trucks has the largest proportion in each city, but then both cities are pretty different in terms of food options.

Here we can see the number of the 5 main options for both cities:

Los Angeles

```
[28]: dfLA
```

[28]:

Type of restaurant	Quantity of restaurants
Food Truck	12
Food	6
Food Court	4
Grocery Store	3
Chinese Restaurant	2

Miraflores

```
[29]: dfM
```

[29]:

Type of restaurant	Quantity of restaurants
Food Court	3
Food Truck	3
Peruvian Restaurant	3
Burger Joint	2
Fast Food Restaurant	2

VI. Discussion section

In this section I will make emphasis on the main points noted in the results section in order to make some recommendations to people who wants to travel to Perú.

- ✓ Los Angeles is denser than Miraflores in terms of restaurants, plus it has more restaurants in total.
- ✓ In both cities you can find restaurants like: Food Trucks, Food Courts and Middle Eastern Restaurant.
- ✓ Some interesting types of restaurants to eat you can only find in Miraflores are: Soup place, Afghan Restaurant, Health food store and of course Peruvian Restaurants.

- ✓ Some interesting types of restaurants to eat you can only find in Los Angeles are: Thai Restaurant, Chinese Restaurants, Taco Place and Street Food Gathering.
- ✓ Food Truck is the most popular food store in both cities.

VII. Conclusion

From our analysis I conclude that:

- ✓ Both cities have different options of food in terms of the types of restaurants but at the same time share their principal (Food trucks). If you are not a big fan of that kind of food you can try also Food courts and Eastern Restaurants in Miraflores.
- ✓ Miraflores has almost half of restaurants in total compared to Los Angeles and they not as close as they are in L.A. So maybe you will have to take a taxi o another mean of transport to get to your favorite place.
- ✓ Miraflores has a diverse portfolio of options you cannot find in L.A. like: Afghan Restaurant, Health food store and Peruvian Restaurants.