The battle of neighborhoods: Need for coffee

Table of contents

1. Introduction

Problem description

Coffee is one of the most widely consumed beverages in the world. It is more than just a drink: it is a culture, an industry, an art, and a passion. Having a cup of coffee is an everyday ritual for a lot of people. Cafeterias and coffee shops have become favorite places for business meetings, romantic rendezvous, and to hang out with friends. Big cities, especially those districts that are populated with business or office facilities, historical or entertaining areas that attract a lot of tourists, usually, have numerous venues related to serving and selling coffee.

In my final report, I will explore coffee shops of San Francisco, CA, find out how many coffee shops are located in each neighborhood, build maps that represent the quantity and quality of them. Also, i will try to forecast where is the best location in San Francisco to open a new coffee shop. I will use data science powers to generate a few most promising neighborhoods as best possible locations that can be chosen by stakeholders.

2. Data

Data description

For the final assignment I will use Foursquare location data, data that we had been using in this course labs, and other datasets from open data sources.

1. **San Francisco neighborhoods data set with coordinates**

2. **Foursquare location data for a specific venue category (coffee)**

3. Methodology

In this assignment I will explore and analyse 5 most popular neighborhoods of San Francisco, CA from a touristic prospective:

Russian Hill

Marina District

Fisherman’s Wharf

Chinatown

Financial District. Our goal is to find out what are the most popular venues and how many of them are coffee shops. The **methodology** of the research is:

1. Upload files that contain San Francisco neighborhood data with their coordinates(.csv) and geojson file to build a map
2. Using Foursquare API, explore the downtown of San Francisco city, find the top 100 closest venues in 1000m radius
3. Using Foursquare API, find specific venues (coffee shops) for Lombard street
4. For each of five districts, find the top 100 closest venues in 1000m radius. Create data frames for all
   venues and separately for coffee shops and cafes.
5. For each of five districts, build graphics charts to visualise the data.
6. Combine all found coffee shops and cafes into one data set, then build a choropleth map

4. Analysis

For initial analysis let's upload San Francisco Neighborhoods data set and explore the centre of the city of San Francisco

[3]:

	LINK	the_geom	name
0	http://en.wikipedia.org/wiki/Sea_Cliff,_San_Fr...	MULTIPOLYGON (((-122.49345526799993 37.7835181...	Seacliff
1	NaN	MULTIPOLYGON (((-122.48715071499993 37.7837854...	Lake Street
2	http://www.nps.gov/prsf/index.htm	MULTIPOLYGON (((-122.47758017099994 37.8109931...	Presidio National Park
3	NaN	MULTIPOLYGON (((-122.47241052999993 37.7873465...	Presidio Terrace
4	http://www.sfgate.com/neighborhoods/sf/innerri...	MULTIPOLYGON (((-122.47262578999994 37.7863148...	Inner Richmond

Using Foursquare API let's extract the information about names, categories and coordinates of the venues for downtown of San Francisco city and put them into a Data Frame

:

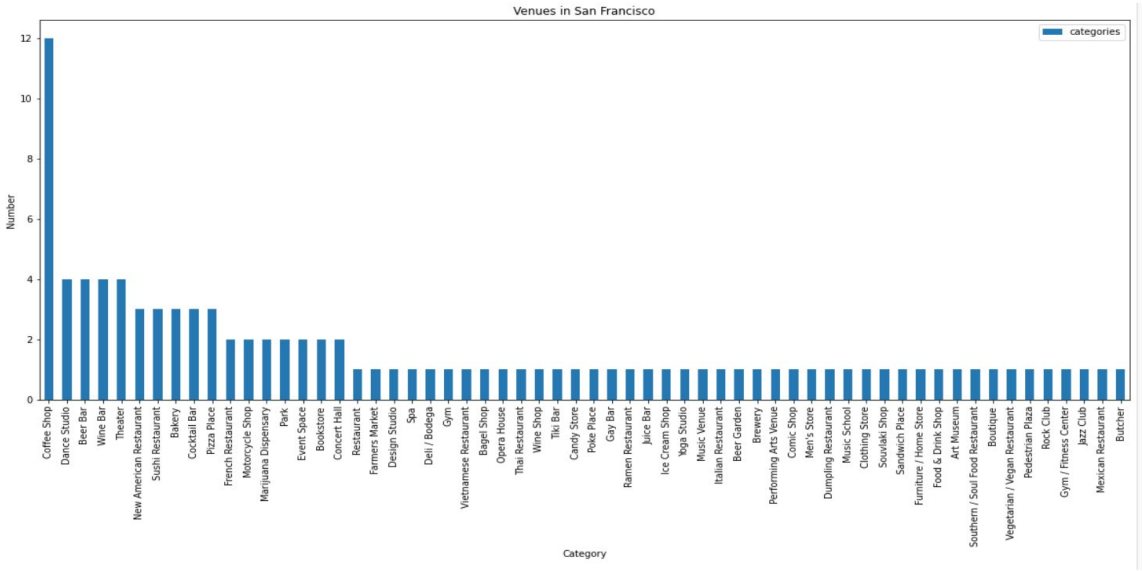
	name	categories	lat	lng
0	Blue Bottle Coffee	Coffee Shop	37.776286	-122.416867
1	Orpheum Theatre	Theater	37.779315	-122.414790
2	Fitness SF	Gym / Fitness Center	37.776161	-122.416413
3	Ananda Fuara	Vegetarian / Vegan Restaurant	37.777693	-122.416353
4	Coffee Cultures	Coffee Shop	37.776089	-122.414717

Since our objective is to explore venues that serve and sell coffee, let's group our data set by categories to see how many of them are Coffee shops or Cafe.

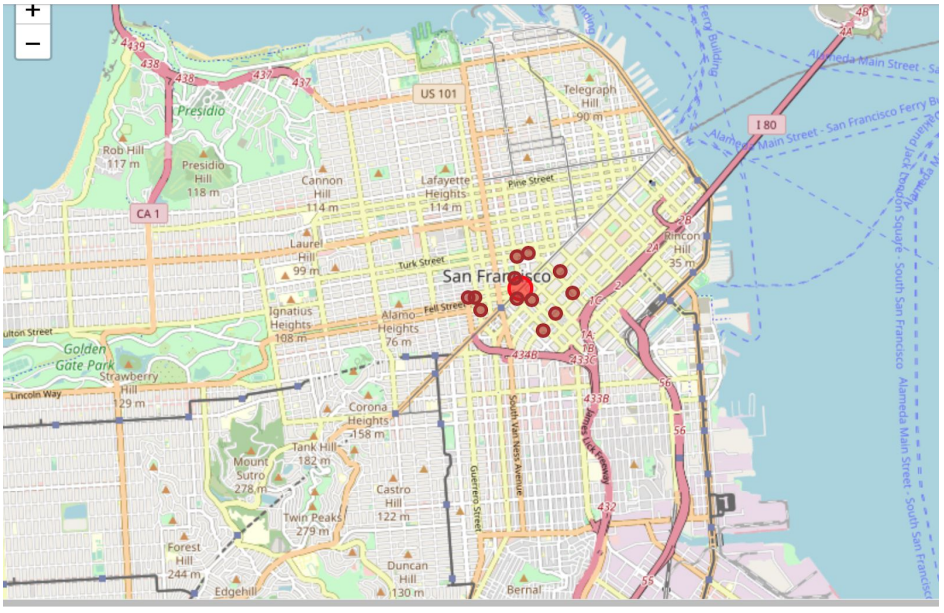
As we can see the largest number of venues of one specific category are Coffee Shops. Let's visualize the nearby_venues_counts data to see coffee shops number as well as other categories. In our data set we can spot such categories as **Café**. A café is a type of restaurant which typically serves coffee and tea, in addition to light refreshments such as baked goods or snacks. The term "café" comes from the French word meaning "coffee".

The difference between the number of coffee shops and other categories became even more significant!

Let's visualize our new data set



Let's visualize our coffee shops on a map



Now we know that coffee shops are the most common category of venues in the San Francisco Downtown. Lets use Foursquare service for a specific venue category (Coffee Shop, of course) in one of most touristic areas in San Francisco, Lombard Street (aka Crooked street). We will define a query to search for coffee that is within 500 metres.

search_query = 'Coffee'

radius = 500

print(search_query + ' .... OK!')

[34]:

	id	name	categories	referralId	hasPerk	location.address	location.lat	location.lng	loca
0	50a6999590e7783c56a9945c	Beacon Coffee & Pantry	[{'id': '4bf58dd8d48988d1e0931735', 'name': 'C...	v-1612156496	False	805 Columbus Ave	37.802293	-122.413210	[
1	4f9bff26e4b04f3d08fea156	Coffee Shop at Enoteca Musto	[{'id': '4bf58dd8d48988d16d941735', 'name': 'C...	v-1612156496	False	NaN	37.806255	-122.416235	[
2	519bfc0f498efb3ecbd8d442	Eva's Coffee	[{'id': '4bf58dd8d48988d1e0931735', 'name': 'C...	v-1612156496	False	1445 Lombard St	37.800892	-122.425348	[
3	52617140498ec4929863ea86	Saint Frank	[{'id': '4bf58dd8d48988d1e0931735', 'name': 'C...	v-1612156496	False	2340 Polk St	37.798461	-122.422249	[

[35]:
```
lombard_coffee.shape
```

[35]:
```
(4, 17)
```

Just four coffee shops! Not so much…

Let's explore five most touristic neighbourhoods of San Francisco according to: TripAdvisor, GoCity, SFTravel

https://www.tripadvisor.com/Attractions-g60713-Activities-c47-t34-San_Francisco_California.html

https://gocity.com/san-francisco/en-us/blog/san-francisco-neighborhoods and

https://www.sftravel.com/neighborhoods

1. Russian Hill

2. Marina

3. Chinatown

4. Financial District

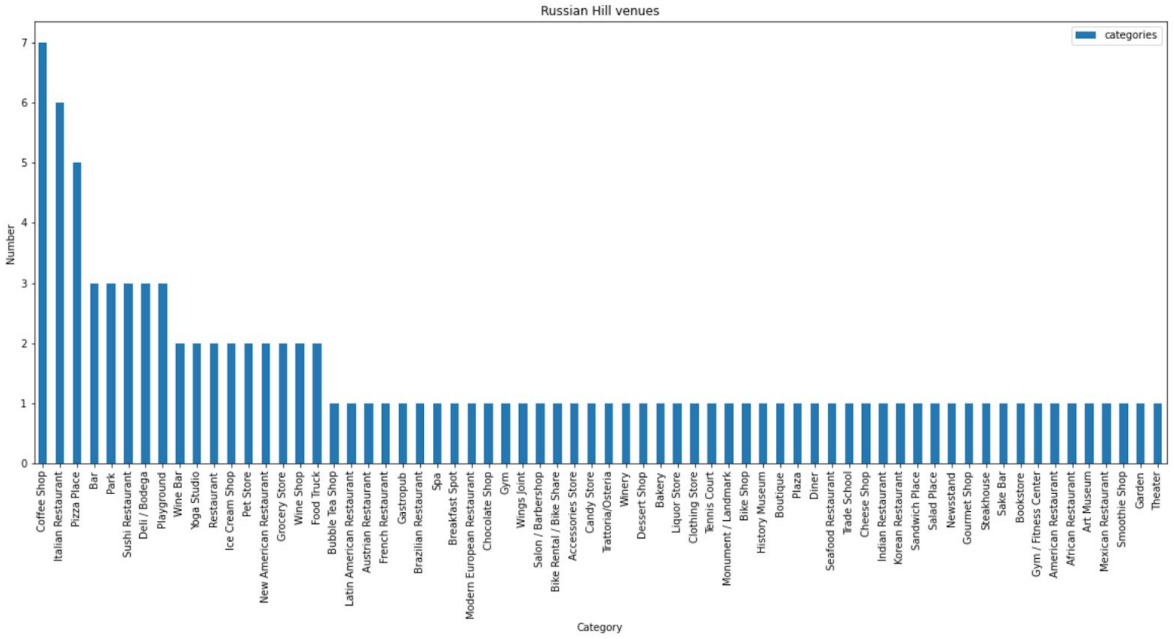
5. North Beach

4.1. Russian Hill

address = 'Russian Hill, San Francisco, CA'

```
geolocator = Nominatim(user_agent="foursquare_agent")

location = geolocator.geocode(address)

latitude = location.latitude

longitude = location.longitude

print('Russian Hill coordinate:', latitude, longitude)
```

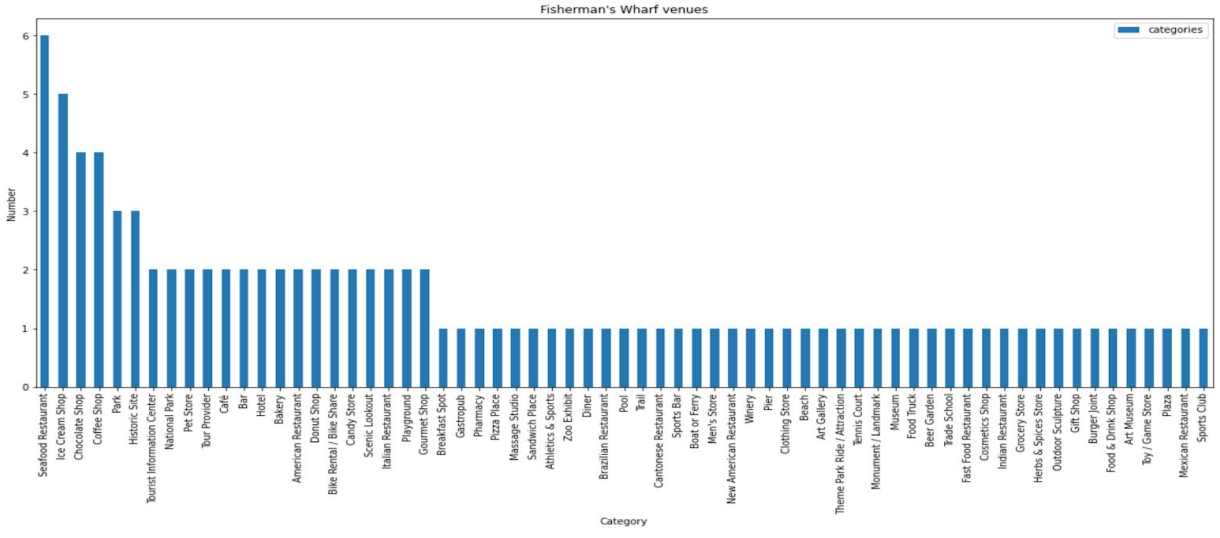
Let's visualize numbers of venues at Russian Hill



As we can see, Coffee Shops are preoccupied, but Italian Restaurants are at the second place.

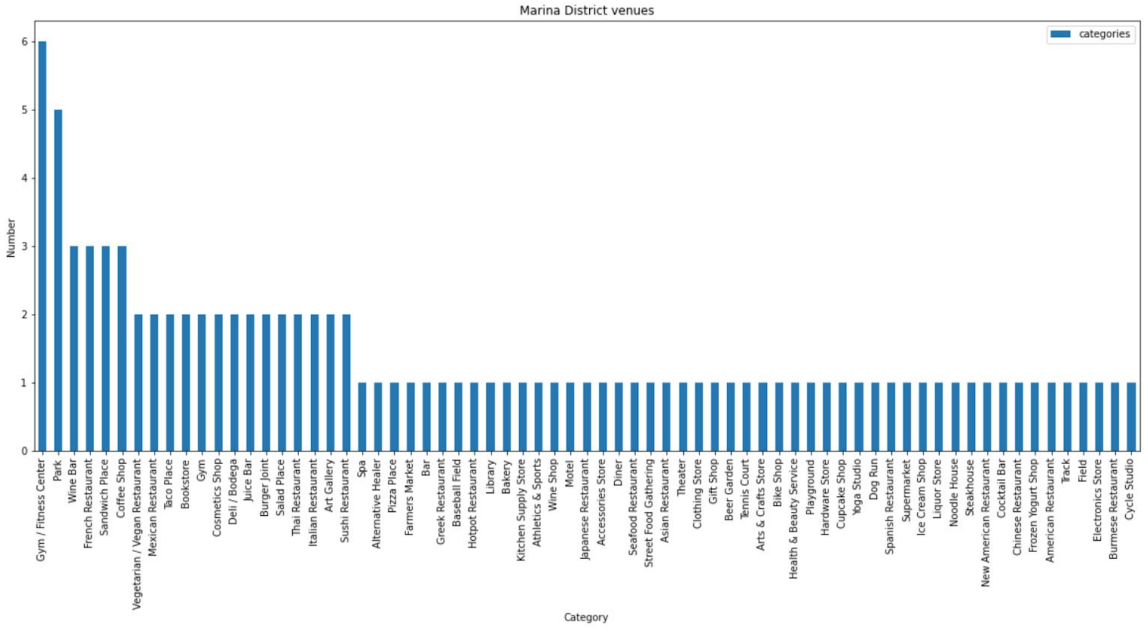
4.2. Fisherman's Wharf

Let's retrieve and visualize numbers of venues at Fisherman's Wharf



4.3. Marina District

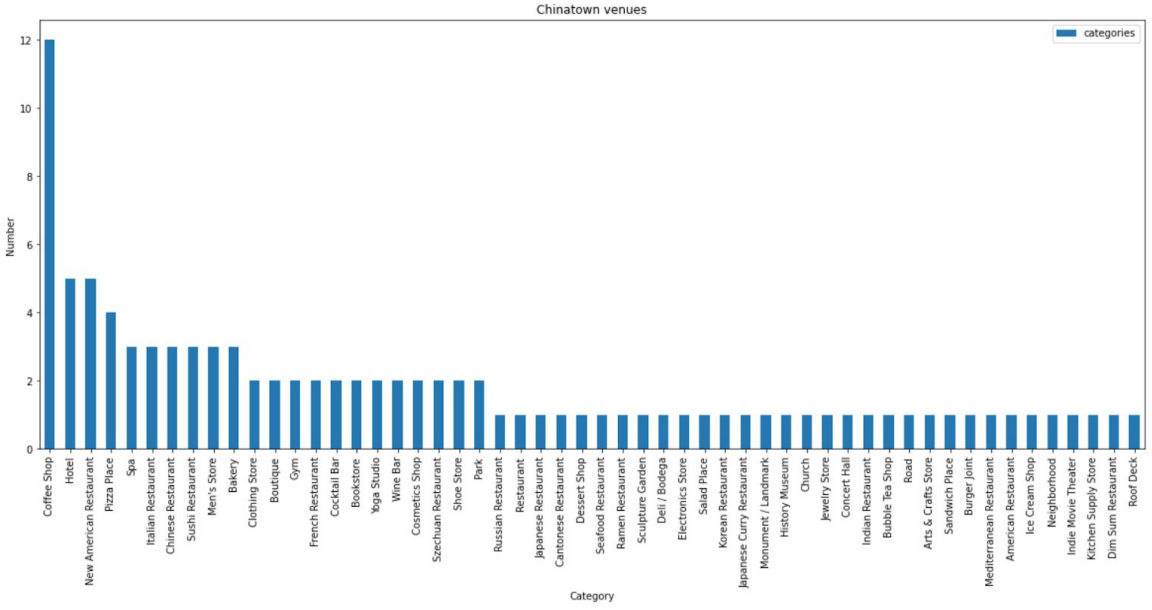
Let's retrieve and visualize numbers of venues at Marina District



We have a lot of Gim/Fitness Centers, but only three of Coffee Shops at Marina District

4.4. Chinatown

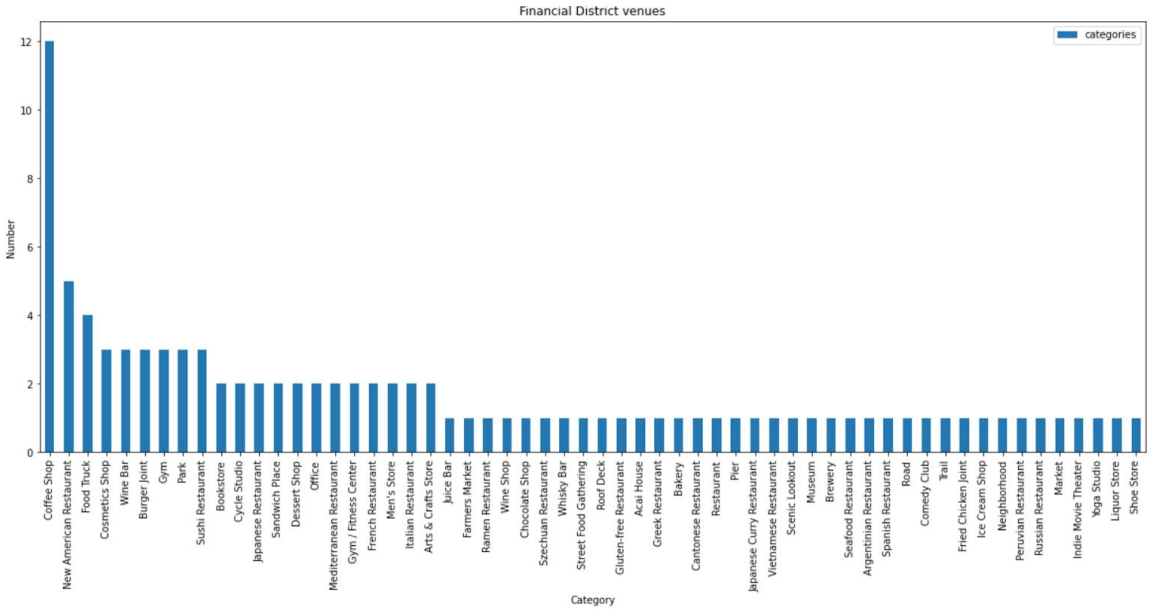
Let's retrieve and visualize numbers of venues at Chinatown



There are more coffee shops than Chinese restaurants in Chinatown!

4.5. Financial District

Finally, let's explore the Financial District of San Francisco



There are a lot of coffee shops at the Financial District in San Francisco. Let's reate a separate data frame for the coffee shops in this location

Place all found coffee shops in one data frame

:

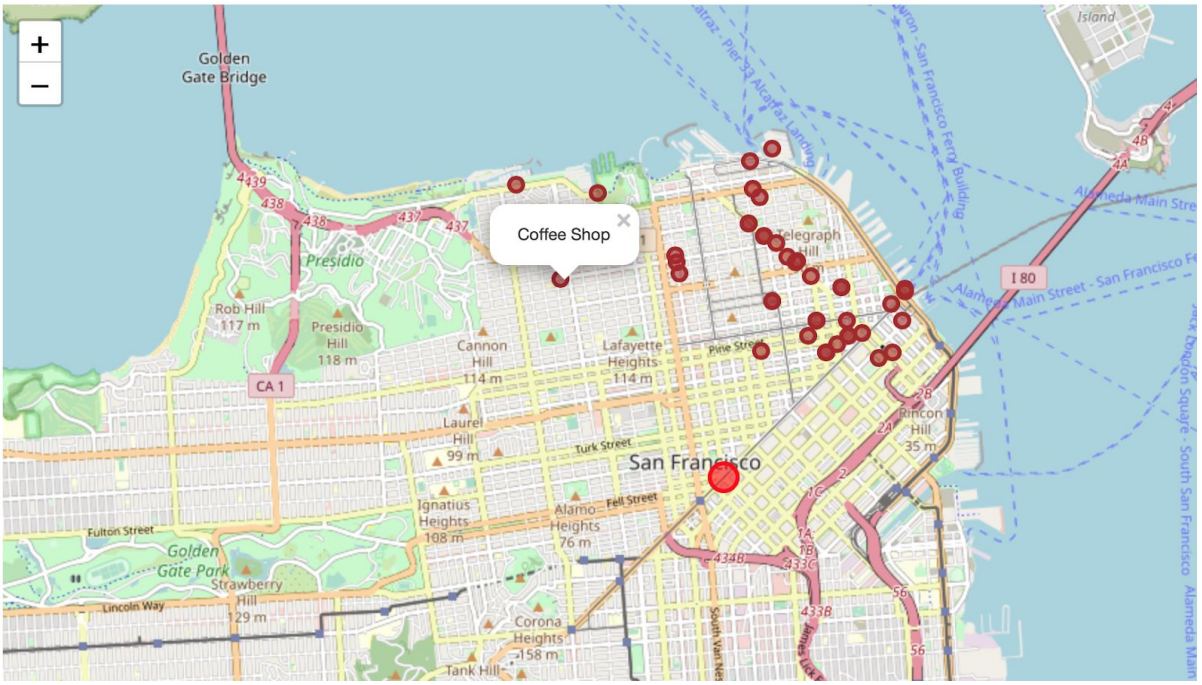
	index	name	categories	lat	lng	neighborhood
0	8	Philz Coffee	Coffee Shop	37.791651	-122.399157	Financial District
1	11	Blue Bottle Coffee	Coffee Shop	37.791320	-122.400983	Financial District
2	19	Philz Coffee	Coffee Shop	37.794422	-122.395636	Financial District
3	22	Cafe Me	Coffee Shop	37.796063	-122.401821	Financial District
4	32	Blue Bottle Coffee Kiosk	Coffee Shop	37.795691	-122.393896	Financial District

coffee_result.shape

(40, 5)

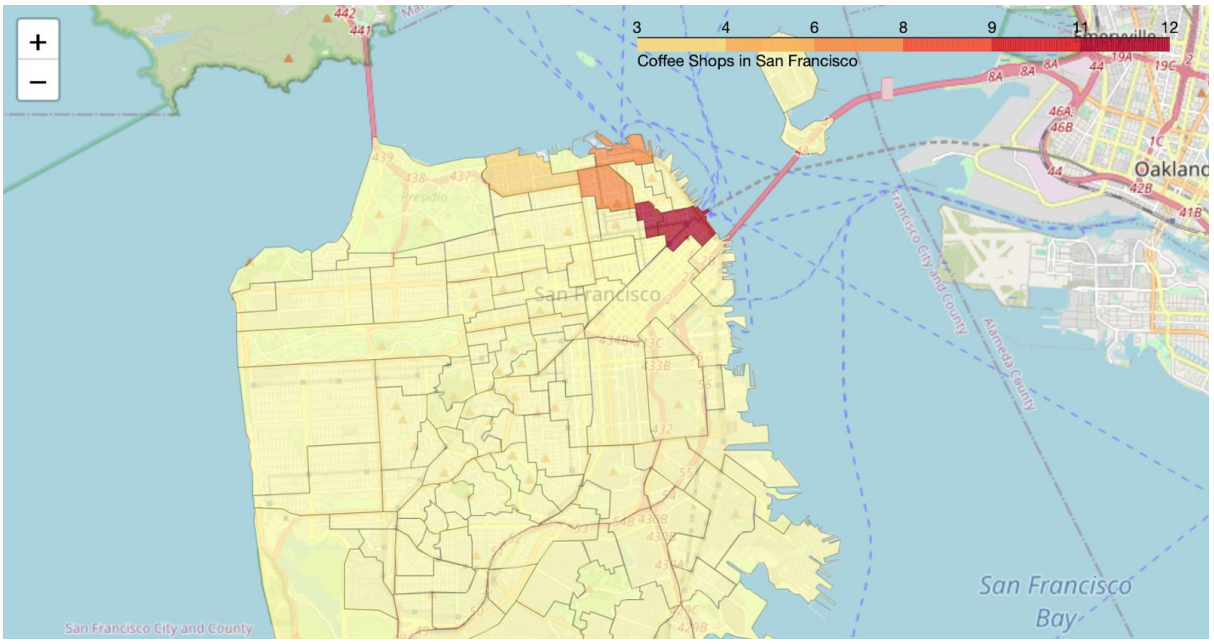
Let's place our coffee shops on a map

```
latitude = 37.7775
longitude = -122.416389

coffee_map = folium.Map(location=[latitude, longitude], zoom_start=13) # generate map centred arounc


folium.CircleMarker(
    [latitude, longitude],
    radius=10,
    color='red',
    popup='San Francisco',
    fill = True,
    fill_color = 'red',
    fill_opacity = 0.6
).add_to(coffee_map)


for lat, lng, label in zip(coffee_result.lat, coffee_result.lng, coffee_result.categories):
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        color='brown',
        popup=label,
        fill = True,
        fill_color='brown',
        fill_opacity=0.6
    ).add_to(coffee_map)

# display map
coffee_map
```



Let's build a Choropleth map

```
!wget --quiet https://cocl.us/sanfran_geojson -O san-francisco.geojson

gjson = r'https://raw.githubusercontent.com/Ok-sa/Coursera_Capstone/main/SF%20Find%20Neighborhoods.geojson'
latitude = 37.77
longitude = -122.42
sf_map = folium.Map(location=[latitude, longitude], zoom_start = 12)
sf_map.choropleth(
    geo_data = gjson,
    data = coffee_counts,
    columns = ['neighborhood', 'count'],
    key_on ='feature.properties.name',
    fill_color ='YlOrRd',
    fill_opacity = 0.7,
    line_opacity = 0.2,
    legend_name = 'Coffee Shops in San Francisco',
    reset = True)
sf_map
```



5. Results and Discussion

Our analysis shows that high demand and everyday coffee consumption, surges to open more coffee shops. Every good coffee shop eventually obtains its clients among local residents. Tourists make a huge impact on coffee consumption too. It is reasonable to open a coffee shop In places that are located nearby some attractions or places that are popular among tourists. We analyzed five most popular San Francisco neighbourhoods. Coffee shops and Cafes are the most popular categories of venues of four of them. We find out that Financial District and Chinatown have more coffee shops than Marina and Fisherman's Wharf. Marina District that has the lowest number of coffee shops would be a good place to open another one.

6. Conclusion

Purpose of this project was to identify San Francisco areas that are popular among tourists with low numbers of Coffee Shops in order to aid stakeholders in narrowing down the

search for the optimal location for a new Cafe. By placing on a map venues that belong to Coffee Shop or Cafe categories.

Final decision of optimal restauraIdeas for further rent location will be made by stakeholders based on specific characteristics of the location, taking into consideration additional factors that could make impact.

Ideas for further research in the area:

- tourism

- housing and employment

- criminal situation