Battle of Neighborhoods in Los Angeles

Applied Data Science Capstone Project by IBM via Coursera

Aimaiti Rehemanjiang

Outlines

- The Business Problem
- Data Loading and Preparation
- Methodology
- Data Analysis
- Result and Discussion
- Conclusion

The Business Problem

Let us imagine that one of the largest coffee shops in Europe MeinCoffee is planning to expand its business to North Ameria. They are planning to open their very first five Coffee Shops in Los Angeles, California, USA. Since there are lots of coffee shops in Los Angeles, the stakeholders must have some reliable information about the best or optimal locations / neighbourhoods. Here optimal neibourhood might be neighborhood or neiborhoods with less or no number of existing coffee shops in certain near area e.g. 3 km range. The data science team must be able to delierver a fast report about the necessray location information, then the stakeholder can make safer decision that in which neighbourhoods are more suitable to open their new coffee shops.

Data Loading and Preparation

- Using data from online source

In [2]:

```
# https://usc.data.socrata.com/dataset/Los-Angeles-Neighborhood-Map/r8qd-yxsr
neigh_path = r'C:/pythonwork/kaggle/data/us_accidents/la_neighborhoods.csv'
# read the dataset into the pandas DataFrame
neigh_df=pd.read_csv(neigh_path)
# Let's look at the dataset
neigh_df.head()
```

Out[2]:

geom	kind	external_i	name	display_na	sqmi	type	name_1	slug_1	latitude	longitude	location
YGON 20541 897...	L.A. County Neighborhood (Current)	acton	Acton	Acton L.A. County Neighborhood (Current)	39.339109	unincorporated-area	NaN	NaN	-118.169810	34.497355	POINT(34.497355239240846 -118.16981019229348)
YGON 00012 109...	L.A. County Neighborhood (Current)	adams-normandie	Adams-Normandie	Adams-Normandie L.A. County Neighborhood (Curr...	0.805350	segment-of-a-city	NaN	NaN	-118.300208	34.031461	POINT(34.031461499124156 -118.30020800000011)
YGON 00009 029...	L.A. County Neighborhood (Current)	agoura-hills	Agoura Hills	Agoura Hills L.A. County Neighborhood (Current)	8.146760	standalone-city	NaN	NaN	-118.759885	34.146736	POINT(34.146736499122795 -118.75988450000015)
YGON	L.A. County			Agua Dulce							

Data Loading and Preparation(cont.)

- Using data via Foursquare API

In [11]:
```
# type your answer here
latitude =  dwtw_df.longitude.values[0]
longitude = dwtw_df.latitude.values[0]
LIMIT = 100
search_query = 'Italian'
radius = 500
url = 'https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&ll={},{}&v={}&radius={}&limit={}'.format(C
url
```

Out[11]:
```
'https://api.foursquare.com/v2/venues/explore?client_id=4KA2G0XZSTB0RYQRCEKQQB1SFRLNQKIEY4NJJNLKTXMDTOYS&clie
```

Now query the above prepared URL by utilizing the *requests* library.

In [12]:
```
results = requests.get(url).json()
results
```

Out[12]:

```
{'meta': {'code': 200, 'requestId': '5e524db8fb34b5001bc91441'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': '$-$$$$', 'key': 'price'}]},
  'headerLocation': 'Fashion District',
  'headerFullLocation': 'Fashion District, Los Angeles',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 19,
```

Data Loading and Preparation(cont.)

- Analyse friendly prepared dataset

In [25]:
```
LA_venues.rename(columns={"Unnamed: 0":"ID"},inplace=True)
LA_venues.set_index("ID", inplace=True)
print(LA_venues.shape)
LA_venues.head()
```

(3020, 7)

Out[25]:

ID	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Acton	34.497355	-118.169810	Epik Engineering	34.498718	-118.168046	Construction & Landscaping
1	Acton	34.497355	-118.169810	Alma Gardening Co.	34.494762	-118.172550	Construction & Landscaping
2	Adams-Normandie	34.031461	-118.300208	Orange Door Sushi	34.032485	-118.299368	Sushi Restaurant
3	Adams-Normandie	34.031461	-118.300208	Shell	34.033095	-118.300025	Gas Station
4	Adams-Normandie	34.031461	-118.300208	Sushi Delight	34.032445	-118.299525	Sushi Restaurant

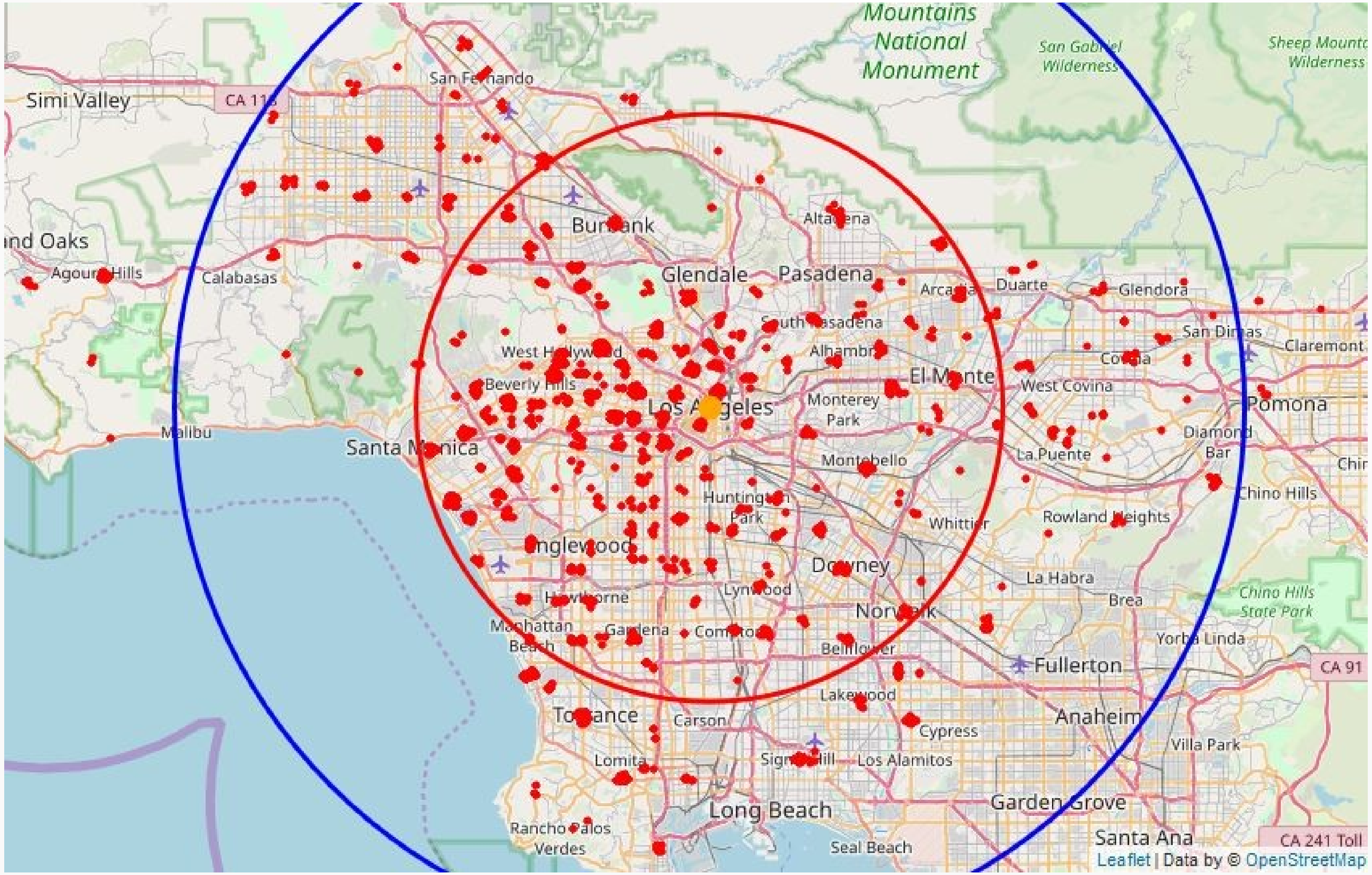
In [26]:
```
LA_venues.tail()
```

Out[26]:

ID	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
3015	Woodland Hills	34.159409	-118.615217	e-Motion Training	34.158111	-118.615593	Gym / Fitness Center
3016	Woodland Hills	34.159409	-118.615217	Deats Stewart Deats Design	34.159030	-118.612447	Wine Bar
3017	Woodland Hills	34.159409	-118.615217	Woodland Hills Carpet Restoration	34.158725	-118.618393	Carpet Store
3018	Woodland Hills	34.159409	-118.615217	Trunk Peter Productions	34.162271	-118.613903	Concert Hall
3019	Woodland Hills	34.159409	-118.615217	Le Zig Zag Club, Paris	34.162803	-118.615821	Asian Restaurant

Data Loading and Preparation(cont.)

- Overview of Los Angeles venues on map



Red circle shows the 23 km distance range

Blue circle shows the 42 km distance range

Methodology

- Foucus my analysis on locations within the range of 42 km from the center of Los Angeles
- Find out all venues of coffee shops from the prepared dataset in previous step.
- Filter out those venues such that there are at least one coffee shop in 2 km range (by assuming that most residents have their own cars and feel the 2 km is not too far in such a large city like Los Angeles)
- According to previous result, find out those neighborhoods and venues without any coffee shops within 2 km distance.
- Classify these venues into reasonable number of clusters and show corresponding cluster centers.
- Show how far these locations, namely these cluster centers

Data Analysis

- Coffee Shops

In [31]:
```
LA_venues_coffee = LA_venues[LA_venues['Venue Category']=='Coffee Shop']
# To avoid the 'Quota Ecxceeded' Error of Foursquare API
LA_venues_coffee.set_index('Neighborhood')
print(LA_venues_coffee.shape)
LA_venues_coffee.head(10)
```

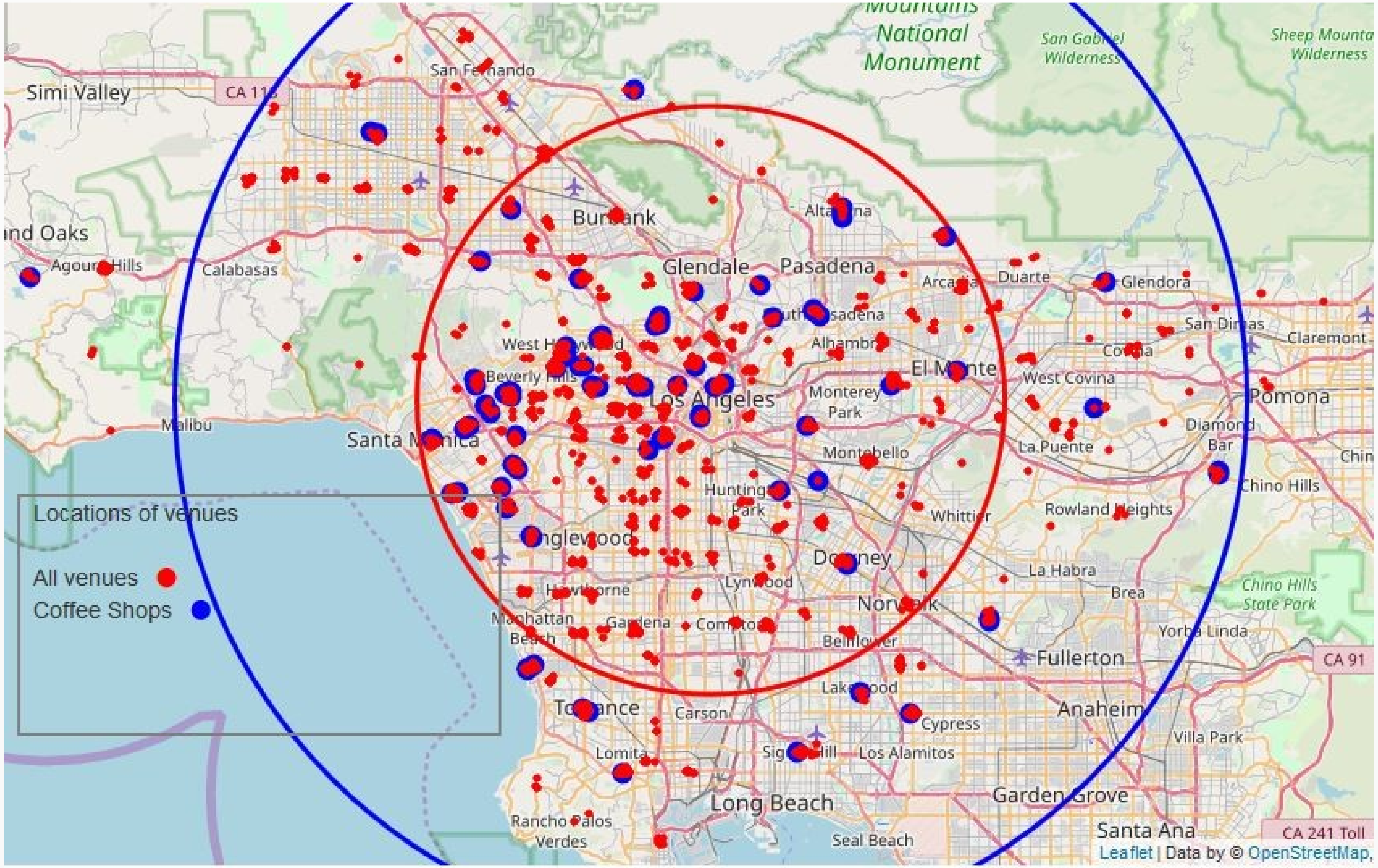
(89, 7)

Out[31]:

ID	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
159	Atwater Village	34.131066	-118.262373	Starbucks	34.129278	-118.258659	Coffee Shop
190	Azusa	34.137470	-117.912469	Starbucks	34.135670	-117.907500	Coffee Shop
336	Beverly Grove	34.076633	-118.376102	Starbucks	34.074911	-118.375322	Coffee Shop
396	Koreatown	34.064510	-118.304958	Bia Coffee	34.063580	-118.308221	Coffee Shop
418	Koreatown	34.064510	-118.304958	Starbucks	34.061339	-118.306407	Coffee Shop
434	Koreatown	34.064510	-118.304958	Starbucks	34.061796	-118.300898	Coffee Shop
523	Century City	34.055326	-118.415083	Starbucks	34.058445	-118.416640	Coffee Shop
532	Century City	34.055326	-118.415083	The Coffee Bean & Tea Leaf	34.058248	-118.413612	Coffee Shop
543	Century City	34.055326	-118.415083	The Coffee Bean & Tea Leaf	34.057721	-118.418984	Coffee Shop
555	Century City	34.055326	-118.415083	The Coffee Bean & Tea Leaf	34.058206	-118.414625	Coffee Shop

Data Analysis(cont.)

- Coffee Shops on map



Data Analysis(cont.)

- Distance between two geolocations

$$d = 2r \arcsin\left(\sqrt{\operatorname{hav}(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2)\operatorname{hav}(\lambda_2 - \lambda_1)}\right)$$
$$= 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right)$$

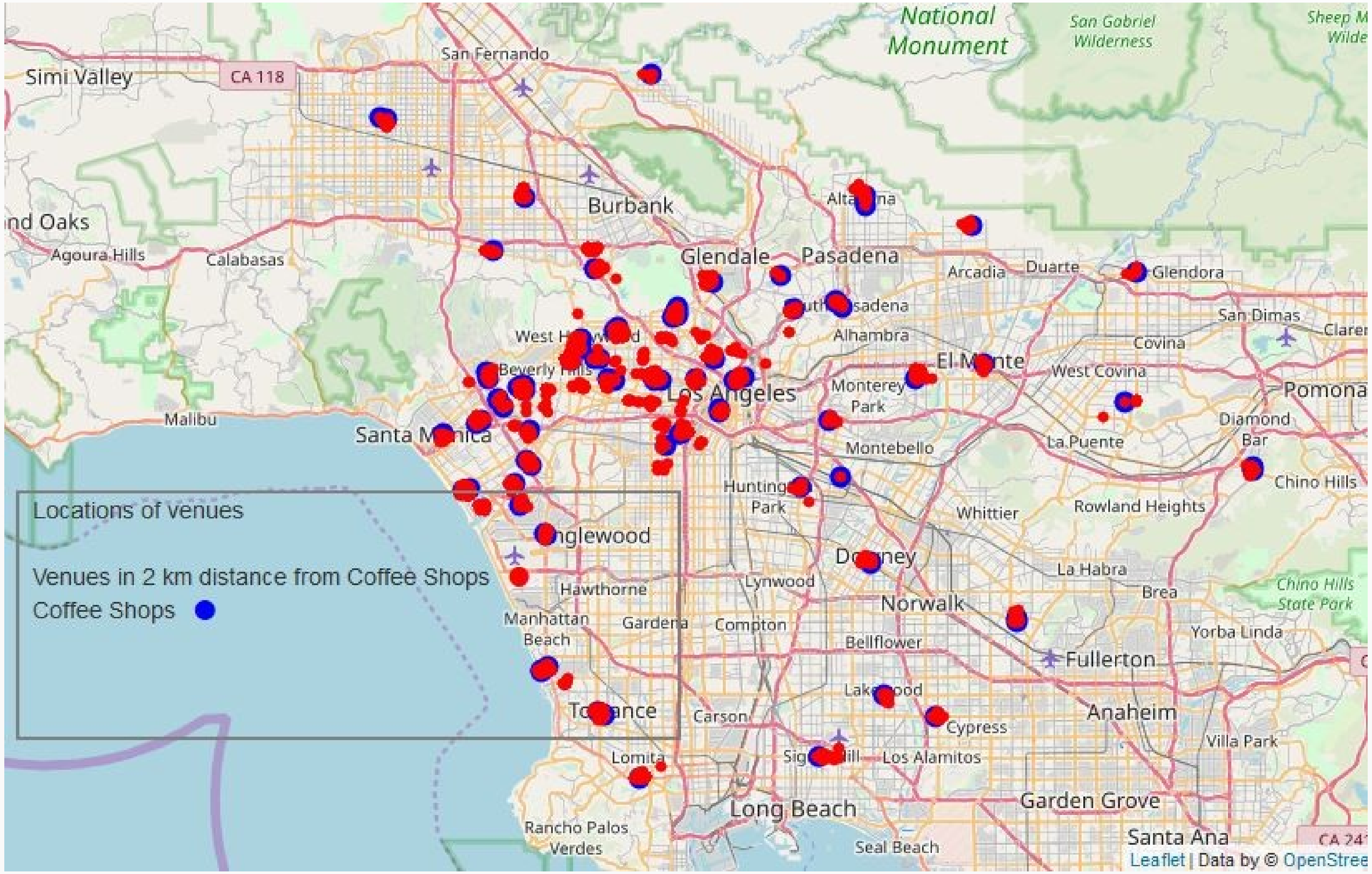
where

- φ_1, φ_2: latitude of point 1 and latitude of point 2 (in radians),
- λ_1, λ_2: longitude of point 1 and longitude of point 2 (in radians).

In [39]:
```
from math import radians, sin, cos, asin, sqrt
def haversine(lon1, lat1, lon2, lat2):
    lon1, lat1, lon2, lat2 = map(radians, [lon1, lat1, lon2, lat2])
    dlon = lon2 - lon1
    dlat = lat2 - lat1
    a = sin(dlat / 2) ** 2 + cos(lat1) * cos(lat2) * sin(dlon / 2) ** 2
    return 2 * 6371 * asin(sqrt(a))
```

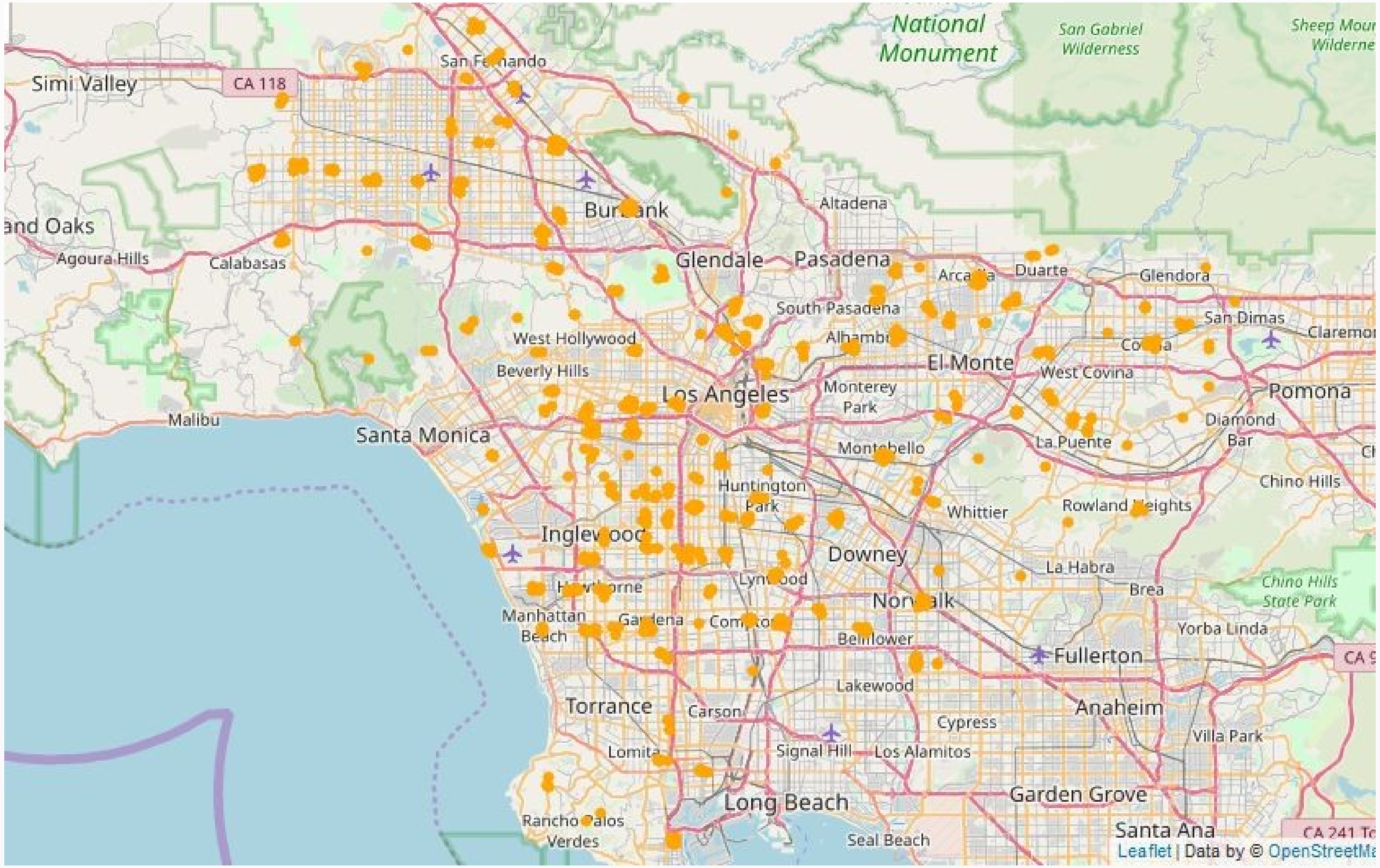
Data Analysis(cont.)

- Venues nearby Coffee Shops in 2 km range



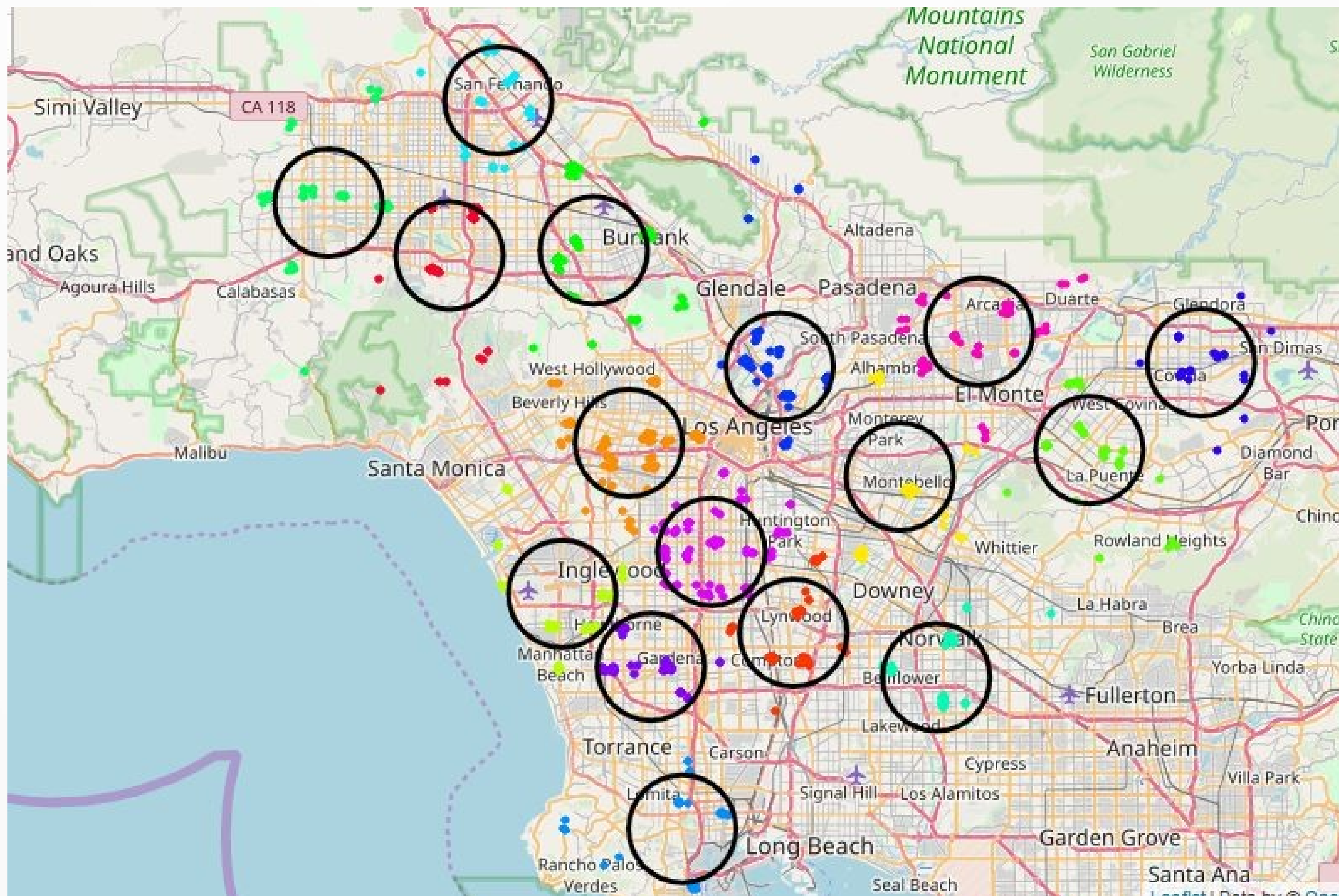
Data Analysis(cont.)

- Venues **without** Coffee Shops in 2 km range



Result and Discussion

- **Clustering** of Venues **without** Coffee Shops in 2 km range



Result and Discussion

- Population density of Los Angeles