FinalProject

April 21, 2020

1 Applied Data Science Capstone - IBM Professional Certificate

1.0.1 Marcelo Porto

1.1 Introduction

The skills learned during this course give us tools to solve some interesting problem. If you are deciding to open a new business, it might be interesting to explore a city and find a neighborhood where your business is needed.
Let's say you want to open a coffee shop. It wouldn't be interesting to open in an area where there are 3 Starbucks' already. We can use the Foursquare API to find neighborhoods where there might be a need for your new coffee shop.

We can use clustering to find similar neighborhoods in the city that might support your plan: neighborhood X has a couple of coffee shops and it is similar to neighborhood Y. So it might be good to take our business to neighborhood Y if there isn't many coffee shops there already. Are neighborhoods X and Y similar enough if we remove coffee shops from the equation?

1.1.1 Target Audience

Prospective business owners could use this to find the best location for their business.
City officials or the appropriate city departments could use this project to identify areas to invest in and attract new business, enhancing the value of possible up-and-coming neighborhoods.

1.1.2 Data Required

Some cities do not have postal codes. Abu Dhabi, in the United Arab Emirates, is one of the cities. So first I will need to search for a list of neighborhoods in Abu Dhabi. Then, I will need to find coordinates for each one of these neighborhoods with the geopy service. After this is done, I will be able to search the city for different venues with the Foursquare API, using this data to cluster and compare neighborhoods.

For Abu Dhabi, Wikipedia provides a list of neighborhoods, which I can use to search for coordinates.

1.1.3 Structure

The Methodology section will explain how the process was done and methods applied. In Results, the findings will be presented along with the codes used. This reports ends with a brief discussion about the results and limitations of this exercise, and improvements that can be made.

1.2 Methodology

The Foursquare API will be used to extract information from business venues in the city. Geopy will be used to turn neighborhood names into latitude and longitude coordinates. The folium package will allows us to visualize these findings in beautiful interactive maps. And talking about beautiful, I will use the BeautifulSoup package to scrape neighborhood names for Abu Dhabi from Wikipedia.

Last but not least, I will apply K-means to cluster the neighborhoods based on what type of businesses are most common between them. After specifying how many clusters are to be created, this unsupervised machine learning algorithm chooses random points as cluster centers, and every other point is assigned to the closest center via the calculation of its Euclidean distances. The mean point of each cluster, or centroids, is calculated, and it becomes the new center for the cluster. This centroid minimizes the total squared distance of each point to the cluster center (Source: my thesis. Just trust me on this). There are limitations to this method, but this is out of the scope of this exercise.

All the coding here is done in Python, in a Jupyter Notebook, using the IBM Skill Network Labs platform, which is free and I **highly** recommend, it will make your life easier. As you can see, a Github repository is used to store the notebook. Feel free to clone it and use for your own learning!

(By the way, even though I am not using it here, I also suggest you to open an account on IBM Cloud. They have loads of free and cool resources for Data Science!)

1.3 Results

Note: Some outputs were cleared to make reading it easier.

Let's begin by loading and installing our tools!

[1]:
```
import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analsysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

!conda install -c conda-forge geopy --yes # uncomment this line if you haven't␣
↪completed the Foursquare API lab
from geopy.geocoders import Nominatim # convert an address into latitude and␣
↪longitude values

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas␣
↪dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors
```

```
# import k-means from clustering stage
from sklearn.cluster import KMeans

!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you␣
↪haven't completed the Foursquare API lab
import folium # map rendering library

print('Libraries imported.')
```

```
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home/jupyterlab/conda/envs/python

  added / updated specs:
    - geopy


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    geographiclib-1.50         |             py_0          34 KB  conda-forge
    geopy-1.21.0               |             py_0          58 KB  conda-forge
    ------------------------------------------------------------
                                           Total:          92 KB

The following NEW packages will be INSTALLED:

  geographiclib      conda-forge/noarch::geographiclib-1.50-py_0
  geopy              conda-forge/noarch::geopy-1.21.0-py_0



Downloading and Extracting Packages
geopy-1.21.0         | 58 KB     | ##################################### | 100%
geographiclib-1.50   | 34 KB     | ##################################### | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
Collecting package metadata (current_repodata.json): done
Solving environment: failed with initial frozen solve. Retrying with flexible
solve.
Collecting package metadata (repodata.json): done
```

```
Solving environment: done

## Package Plan ##

  environment location: /home/jupyterlab/conda/envs/python

  added / updated specs:
    - folium=0.5.0


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    altair-4.1.0               |             py_1         614 KB  conda-forge
    branca-0.4.0               |             py_0          26 KB  conda-forge
    brotlipy-0.7.0             |py36h8c4c3a4_1000         346 KB  conda-forge
    chardet-3.0.4              |py36h9f0ad1d_1006         188 KB  conda-forge
    cryptography-2.8           |   py36h45558ae_2         628 KB  conda-forge
    folium-0.5.0               |             py_0          45 KB  conda-forge
    pandas-1.0.3               |   py36h830a2c2_1        11.1 MB  conda-forge
    pysocks-1.7.1              |   py36h9f0ad1d_1          27 KB  conda-forge
    toolz-0.10.0               |             py_0          46 KB  conda-forge
    urllib3-1.25.9             |             py_0          92 KB  conda-forge
    vincent-0.4.4              |             py_1          28 KB  conda-forge
    ------------------------------------------------------------
                                           Total:        13.1 MB

The following NEW packages will be INSTALLED:

  altair             conda-forge/noarch::altair-4.1.0-py_1
  attrs              conda-forge/noarch::attrs-19.3.0-py_0
  branca             conda-forge/noarch::branca-0.4.0-py_0
  brotlipy           conda-forge/linux-64::brotlipy-0.7.0-py36h8c4c3a4_1000
  chardet            conda-forge/linux-64::chardet-3.0.4-py36h9f0ad1d_1006
  cryptography       conda-forge/linux-64::cryptography-2.8-py36h45558ae_2
  entrypoints        conda-forge/linux-64::entrypoints-0.3-py36h9f0ad1d_1001
  folium             conda-forge/noarch::folium-0.5.0-py_0
  idna               conda-forge/noarch::idna-2.9-py_1
  importlib-metadata conda-forge/linux-64::importlib-
metadata-1.6.0-py36h9f0ad1d_0
  importlib_metadata conda-forge/noarch::importlib_metadata-1.6.0-0
  jinja2             conda-forge/noarch::jinja2-2.11.2-pyh9f0ad1d_0
  jsonschema         conda-forge/linux-64::jsonschema-3.2.0-py36h9f0ad1d_1
  markupsafe         conda-forge/linux-64::markupsafe-1.1.1-py36h8c4c3a4_1
  pandas             conda-forge/linux-64::pandas-1.0.3-py36h830a2c2_1
  pyopenssl          conda-forge/noarch::pyopenssl-19.1.0-py_1
  pyrsistent         conda-forge/linux-64::pyrsistent-0.16.0-py36h8c4c3a4_0
```

```
  pysocks            conda-forge/linux-64::pysocks-1.7.1-py36h9f0ad1d_1
  pytz               conda-forge/noarch::pytz-2019.3-py_0
  requests           conda-forge/noarch::requests-2.23.0-pyh8c360ce_2
  toolz              conda-forge/noarch::toolz-0.10.0-py_0
  urllib3            conda-forge/noarch::urllib3-1.25.9-py_0
  vincent            conda-forge/noarch::vincent-0.4.4-py_1
  zipp               conda-forge/noarch::zipp-3.1.0-py_0



Downloading and Extracting Packages
pysocks-1.7.1        | 27 KB     | ##################################### | 100%
cryptography-2.8     | 628 KB    | ##################################### | 100%
toolz-0.10.0         | 46 KB     | ##################################### | 100%
chardet-3.0.4        | 188 KB    | ##################################### | 100%
folium-0.5.0         | 45 KB     | ##################################### | 100%
urllib3-1.25.9       | 92 KB     | ##################################### | 100%
brotlipy-0.7.0       | 346 KB    | ##################################### | 100%
pandas-1.0.3         | 11.1 MB   | ##################################### | 100%
altair-4.1.0         | 614 KB    | ##################################### | 100%
branca-0.4.0         | 26 KB     | ##################################### | 100%
vincent-0.4.4        | 28 KB     | ##################################### | 100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
Libraries imported.
```

First, let's test with some location, to see if there are entries in the Foursquare database for Abu Dhabi.

[ ]:

```
#Test with Abu Dhabi
CLIENT_ID = 'yourid' # your Foursquare ID
CLIENT_SECRET = 'yoursecret' # your Foursquare Secret
VERSION = '20200401' # Foursquare API version
neighborhood_latitude =24.4539
neighborhood_longitude =54.3773
# type your answer here
LIMIT = 100 # limit of number of venues returned by Foursquare API
radius = 500 # define radius
url = 'https://api.foursquare.com/v2/venues/explore?
 ↪&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    neighborhood_latitude,
    neighborhood_longitude,
    radius,
    LIMIT)
```

```
url # display URL

results = requests.get(url).json()
results
```

[3]:
```
# 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']
```

[4]:
```
venues = results['response']['groups'][0]['items']

nearby_venues = json_normalize(venues) # flatten JSON

# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat',␣
↪'venue.location.lng']
nearby_venues =nearby_venues.loc[:, filtered_columns]

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

# clean columns
nearby_venues.columns = [col.split(".")[-1] for col in nearby_venues.columns]

nearby_venues.head()
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/ipykernel_launcher.py:3: FutureWarning: pandas.io.json.json_normalize
is deprecated, use pandas.json_normalize instead
  This is separate from the ipykernel package so we can avoid doing imports
until
```

[4]:

	name	categories	lat	lng
0	Umm Al Emarat Park) (	Park	24.453300	54.381092
1	Home Bakery	Bakery	24.453337	54.381018
2	Café Arabia	Café	24.455760	54.379318
3	Mushrif palace park	Park	24.453375	54.374729
4	Murjan Asfar Hotel Apartment	Hotel	24.453511	54.377871

[5]:
```
nearby_venues.shape
```

[5]:
```
(5, 4)
```

Ok, apparently there are some. Let's plot these venues in a map of Abu Dhabi.

[6]:
```
# create map of AD using latitude and longitude values
latitude = 24.4539
longitude = 54.3773
#neighborhood_latitude =24.4539
#neighborhood_longitude =54.3773
map_AD = folium.Map(location=[latitude, longitude], zoom_start=10)

# add markers to map
for lat, lng, borough, neighborhood in zip(nearby_venues['lat'],␣
 ↪nearby_venues['lng'], nearby_venues['name'], nearby_venues['categories']):
    label = '{}, {}'.format(neighborhood, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_AD)

map_AD
```

[6]:
```
<folium.folium.Map at 0x7fb4cd72bdd8>
```

Cool! Now, those are venues. There are no postal codes in Abu Dhabi.
We can find a list of neighborhoods on Wikipedia.
Let's see if I can find coordinates for one of these neighborhoods, Al Karama.

[7]:
```
address = 'Al Karama, Abu Dhabi, United Arab Emirates'

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

```
25.244402800000003 55.30475541735386
```

Alright, it works. There aren't that many, so I could probably copy the list to a spreadsheet and load here as a csv file.

But that's no fun. Let's scrape the Wikipedia page to get the Neighborhood names, using the BeautifulSoup package.

[8]:
```
import requests
website_url = requests.get('https://en.wikipedia.org/wiki/Abu_Dhabi').text
!pip install BeautifulSoup4
from bs4 import BeautifulSoup
soup = BeautifulSoup(website_url,'html.parser')
```

```
Collecting BeautifulSoup4
  Downloading https://files.pythonhosted.org/packages/e8/b5/7bb03a696f2c9b
7af792a8f51b82974e51c268f15e925fc834876a4efa0b/beautifulsoup4-4.9.0-py3-none-
any.whl (109kB)
     |                 | 112kB 6.1MB/s eta 0:00:01
Collecting soupsieve>1.2 (from BeautifulSoup4)
  Downloading https://files.pythonhosted.org/packages/05/cf/ea245e52f55823f19992
447b008bcbb7f78efc5960d77f6c34b5b45b36dd/soupsieve-2.0-py2.py3-none-any.whl
Installing collected packages: soupsieve, BeautifulSoup4
Successfully installed BeautifulSoup4-4.9.0 soupsieve-2.0
```

The output is pretty long, so I suggest collapsing it for a better read of the notebook. But printing the whole thing is necessary to find what we want. Or you can check the source code of the page.

[ ]:
```
print(soup)
```

That was long, so I'm suppressing it for the purpose of your reading. By inspecting our "soup", we see that the list is under the class "div-col columns column-width".

[ ]:
```
#div-col columns column-width
hoods = soup.find(class_="div-col columns column-width")
print(hoods)
```

To get just the name of the neighborhood, we use this code:

[11]:
```
hoods.a.text
```

[11]:
```
'Al Aman'
```

I'll create a dataframe for the neighborhoods and iterate over our "soup" to get all the names.

[12]:
```
header=["Neighborhood","Lat","Lon"]

df = pd.DataFrame(columns=header)
hoods_rows = hoods.find_all('a')
for tr in hoods_rows:
    row = tr.text

    df = df.append({'Neighborhood': row}, ignore_index=True)
```

[13]:
```
df
```

[13]:
```
              Neighborhood  Lat  Lon
0                  Al Aman  NaN  NaN
1                Al Bateen  NaN  NaN
2               Al Dhafrah  NaN  NaN
3                 Al Falah  NaN  NaN
4                Al Karama  NaN  NaN
5              Al Khubeirah  NaN  NaN
6           Al Lulu Island  NaN  NaN
7                Al Madina  NaN  NaN
8         Al Maryah Island  NaN  NaN
9              Al Manaseer  NaN  NaN
10               Al Manhal  NaN  NaN
11               Al Maqtaa  NaN  NaN
12           Al Markaziyah  NaN  NaN
13                Al Meena  NaN  NaN
14               Al Moroor  NaN  NaN
15              Al Mushrif  NaN  NaN
16               Al Muzoon  NaN  NaN
17               Al Nahyan  NaN  NaN
18              Al Qubesat  NaN  NaN
19        Al Ras Al Akhdar  NaN  NaN
20                 Al Reef  NaN  NaN
21          Al Reem Island  NaN  NaN
22               Al Rehhan  NaN  NaN
23               Al Rowdah  NaN  NaN
24              Al Shamkha  NaN  NaN
25                Bani Yas  NaN  NaN
26                 Al Zaab  NaN  NaN
27              Al Zahiyah  NaN  NaN
28               Al Zahraa  NaN  NaN
29             Al Khalidyah  NaN  NaN
30         Bain Al Jisrain  NaN  NaN
31              Hideriyyat  NaN  NaN
32            Khalifa City  NaN  NaN
33          Marina Village  NaN  NaN
34             Masdar City  NaN  NaN
35  Mohammed Bin Zayed City  NaN  NaN
36         Saadiyat Island  NaN  NaN
37          Shakhbout City  NaN  NaN
38           Officers City  NaN  NaN
39            Qasr El Bahr  NaN  NaN
40          Qasr El Shatie  NaN  NaN
41              Yas Island  NaN  NaN
42                Al Samha  NaN  NaN
```

Great! Now we can put that into the geocoder to get our coordinates.

[14]:
```
address = '{}, Abu Dhabi, United Arab Emirates'
geolocator = Nominatim(user_agent="foursquare_agent")

for i in range(0, df.shape[0]):
    hood = df.iloc[i,0]
    location = geolocator.geocode(address.format(hood))
    if location is not None:
        df.iloc[i,1] = location.latitude
        df.iloc[i,2] = location.longitude
```

[15]:
```
df
```

[15]:

	Neighborhood	Lat	Lon
0	Al Aman	24.432	54.4266
1	Al Bateen	24.2151	55.6263
2	Al Dhafrah	24.4761	54.3694
3	Al Falah	24.4447	54.7282
4	Al Karama	25.2444	55.3048
5	Al Khubeirah	24.4652	54.3368
6	Al Lulu Island	24.4996	54.3457
7	Al Madina	24.3409	54.4907
8	Al Maryah Island	24.5021	54.3902
9	Al Manaseer	NaN	NaN
10	Al Manhal	24.4666	54.366
11	Al Maqtaa	24.4346	54.4544
12	Al Markaziyah	24.4933	54.3667
13	Al Meena	NaN	NaN
14	Al Moroor	NaN	NaN
15	Al Mushrif	24.4369	54.391
16	Al Muzoon	NaN	NaN
17	Al Nahyan	24.4684	54.3852
18	Al Qubesat	NaN	NaN
19	Al Ras Al Akhdar	NaN	NaN
20	Al Reef	24.4577	54.6737
21	Al Reem Island	24.4867	54.4105
22	Al Rehhan	NaN	NaN
23	Al Rowdah	24.4564	54.3596
24	Al Shamkha	24.3885	54.7184
25	Bani Yas	NaN	NaN
26	Al Zaab	NaN	NaN
27	Al Zahiyah	24.4933	54.3799
28	Al Zahraa	24.4392	54.4352
29	Al Khalidyah	NaN	NaN
30	Bain Al Jisrain	24.4092	54.4954
31	Hideriyyat	NaN	NaN

	Neighborhood	Lat	Lon
32	Khalifa City	24.4214	54.5774
33	Marina Village	NaN	NaN
34	Masdar City	24.4259	54.6193
35	Mohammed Bin Zayed City	24.3279	54.5624
36	Saadiyat Island	24.5303	54.4409
37	Shakhbout City	24.3622	54.631
38	Officers City	24.3824	54.535
39	Qasr El Bahr	24.4746	54.3966
40	Qasr El Shatie	NaN	NaN
41	Yas Island	24.4864	54.6091
42	Al Samha	NaN	NaN

We couldn't find all the neighborhoods, but that's ok, we still got plenty for this exercise. Let's remove the ones without coordinates.

[16]:
```
df=df.dropna()
```

[17]:
```
df.reset_index(inplace=True,drop=True)
df
```

[17]:

	Neighborhood	Lat	Lon
0	Al Aman	24.432	54.4266
1	Al Bateen	24.2151	55.6263
2	Al Dhafrah	24.4761	54.3694
3	Al Falah	24.4447	54.7282
4	Al Karama	25.2444	55.3048
5	Al Khubeirah	24.4652	54.3368
6	Al Lulu Island	24.4996	54.3457
7	Al Madina	24.3409	54.4907
8	Al Maryah Island	24.5021	54.3902
9	Al Manhal	24.4666	54.366
10	Al Maqtaa	24.4346	54.4544
11	Al Markaziyah	24.4933	54.3667
12	Al Mushrif	24.4369	54.391
13	Al Nahyan	24.4684	54.3852
14	Al Reef	24.4577	54.6737
15	Al Reem Island	24.4867	54.4105
16	Al Rowdah	24.4564	54.3596
17	Al Shamkha	24.3885	54.7184
18	Al Zahiyah	24.4933	54.3799
19	Al Zahraa	24.4392	54.4352
20	Bain Al Jisrain	24.4092	54.4954
21	Khalifa City	24.4214	54.5774
22	Masdar City	24.4259	54.6193
23	Mohammed Bin Zayed City	24.3279	54.5624
24	Saadiyat Island	24.5303	54.4409
25	Shakhbout City	24.3622	54.631

```
26            Officers City  24.3824   54.535
27             Qasr El Bahr  24.4746  54.3966
28               Yas Island  24.4864  54.6091
```

Now we can plot these in a map:

[18]:
```
# create map of Abu Dhabi using latitude and longitude values
latitude = 24.4539
longitude = 54.3773
map_AD = folium.Map(location=[latitude, longitude], zoom_start=10)

# add markers to map
for lat, lng, neighborhood in zip(df['Lat'], df['Lon'], df['Neighborhood']):
    label = '{}'.format(neighborhood)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_AD)

map_AD
```

[18]: <folium.folium.Map at 0x7fb4cce6aa58>

We use the Foursquare API to find venues in these neighborhoods.
Let's use that trusty function that was provided in the course.

[19]:
```
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?
↪&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)
```

```
        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item␣
↪in venue_list])
    nearby_venues.columns = ['Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

Running the function...

[20]:
```
AD_venues = getNearbyVenues(names=df['Neighborhood'],
                                   latitudes=df['Lat'],
                                   longitudes=df['Lon']
                                  )
```

```
Al Aman
Al Bateen
Al Dhafrah
Al Falah
Al Karama
Al Khubeirah
Al Lulu Island
Al Madina
Al Maryah Island
Al Manhal
Al Maqtaa
Al Markaziyah
Al Mushrif
Al Nahyan
Al Reef
```

```
Al Reem Island
Al Rowdah
Al Shamkha
Al Zahiyah
Al Zahraa
Bain Al Jisrain
Khalifa City
Masdar City
Mohammed Bin Zayed City
Saadiyat Island
Shakhbout City
Officers City
Qasr El Bahr
Yas Island
```

Now we have a list of venues with coordinates for each neighborhood.
Let's just look at the first 10 entries.

[36]:
```python
AD_venues.head(10)
```

[36]:
```
  Neighborhood  Neighborhood Latitude  Neighborhood Longitude  \
0      Al Aman              24.431990                54.42655
1      Al Aman              24.431990                54.42655
2      Al Aman              24.431990                54.42655
3      Al Aman              24.431990                54.42655
4      Al Aman              24.431990                54.42655
5      Al Aman              24.431990                54.42655
6   Al Dhafrah              24.476147                54.36936
7   Al Dhafrah              24.476147                54.36936
8   Al Dhafrah              24.476147                54.36936
9   Al Dhafrah              24.476147                54.36936

                        Venue  Venue Latitude  Venue Longitude  \
0                      Subway       24.435156        54.425187
1  Novotel Abu Dhabi Al Bustan       24.429431        54.429166
2                  McDonald's       24.434646        54.424393
3         Cafe Bonjour Bonsoir       24.429314        54.428250
4            Adagio Aparthotel       24.429396        54.428604
5                Coffee Planet       24.434725        54.423094
6         Jumeirah Etihad Tower       24.476051        54.367716
7              Starbucks (    )    24.477430       54.371626
8            Gudee Pizza & Café       24.477860        54.371012
9  Al Shater Hassan Restaurant       24.478750        54.369562

         Venue Category
0        Sandwich Place
1                 Hotel
2  Fast Food Restaurant
```

```
3                  Café
4                 Hotel
5           Coffee Shop
6                 Hotel
7           Coffee Shop
8           Pizza Place
9    Falafel Restaurant
```

How many unique type of businesses do we have in Abu Dhabi?

[37]:
```
print('There are {} unique businesses.'.format(len(AD_venues['Venue Category'].
 ↪unique())))
```

```
There are 125 unique businesses.
```

[58]:
```
AD_venues.groupby('Neighborhood').count()
```

[58]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	\
Al Aman	6	6	6	
Al Dhafrah	32	32	32	
Al Karama	38	38	38	
Al Khubeirah	17	17	17	
Al Lulu Island	1	1	1	
Al Madina	3	3	3	
Al Manhal	7	7	7	
Al Maqtaa	27	27	27	
Al Markaziyah	39	39	39	
Al Maryah Island	56	56	56	
Al Mushrif	4	4	4	
Al Nahyan	29	29	29	
Al Reef	12	12	12	
Al Reem Island	1	1	1	
Al Rowdah	8	8	8	
Al Shamkha	2	2	2	
Al Zahiyah	68	68	68	
Al Zahraa	4	4	4	
Bain Al Jisrain	27	27	27	
Khalifa City	4	4	4	
Masdar City	6	6	6	
Mohammed Bin Zayed City	1	1	1	
Officers City	1	1	1	
Saadiyat Island	1	1	1	
Shakhbout City	1	1	1	
Yas Island	100	100	100	

```
                         Venue Latitude  Venue Longitude  Venue Category
```

```
Neighborhood
Al Aman                              6        6        6
Al Dhafrah                          32       32       32
Al Karama                           38       38       38
Al Khubeirah                        17       17       17
Al Lulu Island                       1        1        1
Al Madina                            3        3        3
Al Manhal                            7        7        7
Al Maqtaa                           27       27       27
Al Markaziyah                       39       39       39
Al Maryah Island                    56       56       56
Al Mushrif                           4        4        4
Al Nahyan                           29       29       29
Al Reef                             12       12       12
Al Reem Island                       1        1        1
Al Rowdah                            8        8        8
Al Shamkha                           2        2        2
Al Zahiyah                          68       68       68
Al Zahraa                            4        4        4
Bain Al Jisrain                     27       27       27
Khalifa City                         4        4        4
Masdar City                          6        6        6
Mohammed Bin Zayed City              1        1        1
Officers City                        1        1        1
Saadiyat Island                      1        1        1
Shakhbout City                       1        1        1
Yas Island                         100      100      100
```

We see we have a problem with our data here. There are a few neighborhoods with very few entries. We could attempt to increase the radius of search. Or find another data source.
For this exercise we will limit our analysis to those areas with at least 5 businesses.

[70]:
```
temp = AD_venues.groupby('Neighborhood').count()
id = temp.index[temp.iloc[:,0] >= 5]
```

[74]:
```
#rpt[rpt['STK_ID'].isin(stk_list)]
AD_venues = AD_venues[AD_venues['Neighborhood'].isin(id)]
```

What type of venue is more frequent in each neighborhood?
How are we going to cluster these together? To start to answer these questions, we turn the venues into dummy variables.
We can get the dummies with one hot encoding.

[75]:
```
# one hot encoding
AD_onehot = pd.get_dummies(AD_venues[['Venue Category']], prefix="",␣
↪prefix_sep="")
```

```python
# add neighborhood column back to dataframe
AD_onehot['Neighborhood'] = AD_venues['Neighborhood']

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

AD_onehot.head()
```

[75]:

```
   Women's Store  Afghan Restaurant  African Restaurant  Airport Terminal  \
0              0                  0                   0                 0
1              0                  0                   0                 0
2              0                  0                   0                 0
3              0                  0                   0                 0
4              0                  0                   0                 0

   American Restaurant  Arcade  Asian Restaurant  BBQ Joint  Bakery  Beach  \
0                    0       0                 0          0       0      0
1                    0       0                 0          0       0      0
2                    0       0                 0          0       0      0
3                    0       0                 0          0       0      0
4                    0       0                 0          0       0      0

   Bed & Breakfast  Bistro  Bookstore  Boutique  Bowling Alley  \
0                0       0          0         0              0
1                0       0          0         0              0
2                0       0          0         0              0
3                0       0          0         0              0
4                0       0          0         0              0

   Breakfast Spot  Buffet  Burger Joint  Cafeteria  Café  Candy Store  \
0               0       0             0          0     0            0
1               0       0             0          0     0            0
2               0       0             0          0     0            0
3               0       0             0          0     1            0
4               0       0             0          0     0            0

   Chinese Restaurant  Clothing Store  Cocktail Bar  Coffee Shop  \
0                   0               0             0            0
1                   0               0             0            0
2                   0               0             0            0
3                   0               0             0            0
4                   0               0             0            0

   Convenience Store  Cosmetics Shop  Cupcake Shop  Department Store  \
0                  0               0             0                 0
1                  0               0             0                 0
```

```
2                  0                0             0                   0
3                  0                0             0                   0
4                  0                0             0                   0

   Dessert Shop  Donut Shop  Electronics Store  Ethiopian Restaurant  \
0             0           0                  0                     0
1             0           0                  0                     0
2             0           0                  0                     0
3             0           0                  0                     0
4             0           0                  0                     0

   Falafel Restaurant  Fast Food Restaurant  Filipino Restaurant  \
0                   0                     0                    0
1                   0                     0                    0
2                   0                     1                    0
3                   0                     0                    0
4                   0                     0                    0

   Fish & Chips Shop  Flower Shop  Food & Drink Shop  Food Court  \
0                  0            0                  0           0
1                  0            0                  0           0
2                  0            0                  0           0
3                  0            0                  0           0
4                  0            0                  0           0

   French Restaurant  Fried Chicken Joint  Furniture / Home Store  Gift Shop  \
0                  0                    0                       0          0
1                  0                    0                       0          0
2                  0                    0                       0          0
3                  0                    0                       0          0
4                  0                    0                       0          0

   Go Kart Track  Greek Restaurant  Grocery Store  Gym  Gym / Fitness Center  \
0              0                 0              0    0                     0
1              0                 0              0    0                     0
2              0                 0              0    0                     0
3              0                 0              0    0                     0
4              0                 0              0    0                     0

   Gym Pool  Gymnastics Gym  Harbor / Marina  Health & Beauty Service  \
0         0               0                0                        0
1         0               0                0                        0
2         0               0                0                        0
3         0               0                0                        0
4         0               0                0                        0

   Hookah Bar  Hostel  Hot Spring  Hotel  Hotel Bar  IT Services  \
```

```
0          0      0          0     0          0          0
1          0      0          0     1          0          0
2          0      0          0     0          0          0
3          0      0          0     0          0          0
4          0      0          0     1          0          0

   Ice Cream Shop  Indian Restaurant  Italian Restaurant  Japanese Restaurant  \
0               0                  0                   0                    0
1               0                  0                   0                    0
2               0                  0                   0                    0
3               0                  0                   0                    0
4               0                  0                   0                    0

   Jewelry Store  Juice Bar  Korean Restaurant  Lebanese Restaurant  \
0              0          0                  0                    0
1              0          0                  0                    0
2              0          0                  0                    0
3              0          0                  0                    0
4              0          0                  0                    0

   Lingerie Store  Lounge  Medical Center  Mediterranean Restaurant  \
0               0       0               0                         0
1               0       0               0                         0
2               0       0               0                         0
3               0       0               0                         0
4               0       0               0                         0

   Mexican Restaurant  Middle Eastern Restaurant  Miscellaneous Shop  \
0                   0                          0                   0
1                   0                          0                   0
2                   0                          0                   0
3                   0                          0                   0
4                   0                          0                   0

   Moroccan Restaurant  Movie Theater  Multiplex  Nail Salon Neighborhood  \
0                    0              0          0           0      Al Aman
1                    0              0          0           0      Al Aman
2                    0              0          0           0      Al Aman
3                    0              0          0           0      Al Aman
4                    0              0          0           0      Al Aman

   Nightclub  Optical Shop  Pakistani Restaurant  Park  Peruvian Restaurant  \
0          0             0                     0     0                    0
1          0             0                     0     0                    0
2          0             0                     0     0                    0
3          0             0                     0     0                    0
4          0             0                     0     0                    0
```

	Pet Store	Pharmacy	Pizza Place	Playground	Plaza	Pool	Pub	Racetrack	\
0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	

	Residential Building (Apartment / Condo)	Resort	Restaurant	\
0	0	0	0	
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	

	Sandwich Place	Seafood Restaurant	Shawarma Place	Shoe Store	\
0	1	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

	Shopping Mall	Snack Place	South Indian Restaurant	Spa	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

	Sporting Goods Shop	Sports Bar	Steakhouse	Supermarket	Sushi Restaurant	\
0	0	0	0	0	0	
1	0	0	0	0	0	
2	0	0	0	0	0	
3	0	0	0	0	0	
4	0	0	0	0	0	

	Tennis Court	Theater	Theme Park	Theme Park Ride / Attraction	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	

	Theme Restaurant	Toy / Game Store	Turkish Restaurant	\
0	0	0	0	
1	0	0	0	
2	0	0	0	

```
3                 0                 0                    0
4                 0                 0                    0

   Vegetarian / Vegan Restaurant  Wine Bar
0                              0         0
1                              0         0
2                              0         0
3                              0         0
4                              0         0
```

[76]:
```python
AD_grouped = AD_onehot.groupby('Neighborhood').mean().reset_index()
AD_grouped
```

[76]:
```
         Neighborhood  Women's Store  Afghan Restaurant  African Restaurant  \
0             Al Aman       0.000000           0.000000            0.000000
1          Al Dhafrah       0.000000           0.000000            0.000000
2           Al Karama       0.000000           0.000000            0.000000
3        Al Khubeirah       0.000000           0.000000            0.000000
4           Al Manhal       0.000000           0.000000            0.000000
5           Al Maqtaa       0.000000           0.000000            0.000000
6       Al Markaziyah       0.000000           0.000000            0.000000
7    Al Maryah Island       0.017857           0.017857            0.000000
8           Al Nahyan       0.000000           0.000000            0.000000
9             Al Reef       0.000000           0.000000            0.000000
10          Al Rowdah       0.000000           0.000000            0.000000
11         Al Zahiyah       0.000000           0.000000            0.014706
12   Bain Al Jisrain        0.000000           0.000000            0.000000
13        Masdar City       0.000000           0.000000            0.000000
14         Yas Island       0.000000           0.000000            0.010000

    Airport Terminal  American Restaurant  Arcade  Asian Restaurant  \
0           0.000000             0.000000    0.00          0.000000
1           0.000000             0.000000    0.00          0.000000
2           0.000000             0.000000    0.00          0.105263
3           0.000000             0.000000    0.00          0.000000
4           0.000000             0.000000    0.00          0.000000
5           0.037037             0.000000    0.00          0.000000
6           0.000000             0.025641    0.00          0.051282
7           0.000000             0.035714    0.00          0.000000
8           0.000000             0.034483    0.00          0.034483
9           0.000000             0.000000    0.00          0.000000
10          0.000000             0.000000    0.00          0.000000
11          0.000000             0.000000    0.00          0.029412
12          0.000000             0.000000    0.00          0.000000
13          0.000000             0.000000    0.00          0.000000
14          0.000000             0.010000    0.01          0.000000
```

	BBQ Joint	Bakery	Beach	Bed & Breakfast	Bistro	Bookstore	\
0	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	
1	0.000000	0.031250	0.03125	0.000000	0.000000	0.000000	
2	0.000000	0.078947	0.00000	0.000000	0.000000	0.000000	
3	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	
4	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	
5	0.000000	0.037037	0.00000	0.000000	0.000000	0.000000	
6	0.000000	0.025641	0.00000	0.025641	0.000000	0.000000	
7	0.000000	0.035714	0.00000	0.000000	0.017857	0.000000	
8	0.000000	0.068966	0.00000	0.000000	0.000000	0.000000	
9	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	
10	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	
11	0.014706	0.000000	0.00000	0.029412	0.000000	0.014706	
12	0.000000	0.000000	0.00000	0.000000	0.000000	0.037037	
13	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	
14	0.000000	0.010000	0.00000	0.000000	0.000000	0.010000	

	Boutique	Bowling Alley	Breakfast Spot	Buffet	Burger Joint	\
0	0.000000	0.000000	0.000000	0.000000	0.000000	
1	0.000000	0.000000	0.000000	0.000000	0.000000	
2	0.000000	0.000000	0.000000	0.000000	0.000000	
3	0.000000	0.000000	0.000000	0.000000	0.000000	
4	0.000000	0.000000	0.000000	0.000000	0.000000	
5	0.000000	0.000000	0.000000	0.000000	0.037037	
6	0.000000	0.025641	0.025641	0.000000	0.000000	
7	0.017857	0.000000	0.000000	0.017857	0.017857	
8	0.000000	0.000000	0.000000	0.000000	0.034483	
9	0.000000	0.000000	0.000000	0.000000	0.000000	
10	0.000000	0.000000	0.000000	0.000000	0.000000	
11	0.029412	0.000000	0.014706	0.000000	0.029412	
12	0.000000	0.000000	0.000000	0.000000	0.037037	
13	0.000000	0.000000	0.000000	0.000000	0.000000	
14	0.000000	0.000000	0.000000	0.000000	0.030000	

	Cafeteria	Café	Candy Store	Chinese Restaurant	Clothing Store	\
0	0.000000	0.166667	0.00	0.000000	0.000000	
1	0.000000	0.093750	0.00	0.000000	0.031250	
2	0.026316	0.026316	0.00	0.000000	0.000000	
3	0.000000	0.235294	0.00	0.000000	0.000000	
4	0.000000	0.428571	0.00	0.000000	0.000000	
5	0.000000	0.185185	0.00	0.000000	0.000000	
6	0.000000	0.076923	0.00	0.025641	0.000000	
7	0.000000	0.089286	0.00	0.035714	0.000000	
8	0.000000	0.172414	0.00	0.000000	0.000000	
9	0.000000	0.000000	0.00	0.000000	0.000000	
10	0.000000	0.625000	0.00	0.000000	0.000000	
11	0.000000	0.044118	0.00	0.029412	0.029412	

12	0.000000	0.037037	0.00	0.000000	0.000000
13	0.000000	0.166667	0.00	0.000000	0.000000
14	0.000000	0.110000	0.01	0.010000	0.040000

	Cocktail Bar	Coffee Shop	Convenience Store	Cosmetics Shop	\
0	0.000000	0.166667	0.000000	0.000000	
1	0.000000	0.031250	0.000000	0.000000	
2	0.000000	0.000000	0.026316	0.000000	
3	0.000000	0.235294	0.000000	0.000000	
4	0.000000	0.142857	0.000000	0.000000	
5	0.000000	0.074074	0.037037	0.000000	
6	0.000000	0.000000	0.000000	0.000000	
7	0.017857	0.071429	0.000000	0.017857	
8	0.000000	0.103448	0.034483	0.000000	
9	0.000000	0.083333	0.083333	0.000000	
10	0.000000	0.250000	0.000000	0.125000	
11	0.000000	0.044118	0.000000	0.014706	
12	0.037037	0.111111	0.000000	0.000000	
13	0.000000	0.000000	0.000000	0.000000	
14	0.000000	0.070000	0.000000	0.020000	

	Cupcake Shop	Department Store	Dessert Shop	Donut Shop	\
0	0.000000	0.000000	0.000000	0.000000	
1	0.000000	0.031250	0.000000	0.000000	
2	0.000000	0.000000	0.000000	0.000000	
3	0.000000	0.000000	0.000000	0.058824	
4	0.000000	0.000000	0.000000	0.000000	
5	0.000000	0.000000	0.037037	0.000000	
6	0.000000	0.025641	0.000000	0.025641	
7	0.000000	0.000000	0.017857	0.000000	
8	0.034483	0.000000	0.068966	0.000000	
9	0.000000	0.000000	0.000000	0.000000	
10	0.000000	0.000000	0.000000	0.000000	
11	0.000000	0.014706	0.044118	0.000000	
12	0.000000	0.000000	0.000000	0.000000	
13	0.000000	0.000000	0.000000	0.000000	
14	0.000000	0.030000	0.010000	0.000000	

	Electronics Store	Ethiopian Restaurant	Falafel Restaurant	\
0	0.000000	0.000000	0.000000	
1	0.000000	0.000000	0.031250	
2	0.000000	0.000000	0.000000	
3	0.000000	0.000000	0.000000	
4	0.000000	0.000000	0.000000	
5	0.000000	0.000000	0.000000	
6	0.000000	0.000000	0.000000	
7	0.000000	0.000000	0.000000	

8	0.000000	0.000000	0.034483
9	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000
11	0.029412	0.014706	0.000000
12	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000
14	0.010000	0.000000	0.000000

	Fast Food Restaurant	Filipino Restaurant	Fish & Chips Shop	Flower Shop	\
0	0.166667	0.000000	0.00	0.000000	
1	0.062500	0.031250	0.00	0.000000	
2	0.000000	0.000000	0.00	0.000000	
3	0.117647	0.000000	0.00	0.000000	
4	0.000000	0.000000	0.00	0.000000	
5	0.037037	0.000000	0.00	0.000000	
6	0.128205	0.025641	0.00	0.000000	
7	0.000000	0.000000	0.00	0.017857	
8	0.000000	0.000000	0.00	0.068966	
9	0.000000	0.000000	0.00	0.000000	
10	0.000000	0.000000	0.00	0.000000	
11	0.044118	0.000000	0.00	0.000000	
12	0.000000	0.000000	0.00	0.000000	
13	0.166667	0.000000	0.00	0.000000	
14	0.010000	0.000000	0.01	0.000000	

	Food & Drink Shop	Food Court	French Restaurant	Fried Chicken Joint	\
0	0.000000	0.00	0.000000	0.0000	
1	0.000000	0.00	0.000000	0.0625	
2	0.000000	0.00	0.000000	0.0000	
3	0.000000	0.00	0.000000	0.0000	
4	0.000000	0.00	0.000000	0.0000	
5	0.000000	0.00	0.037037	0.0000	
6	0.000000	0.00	0.000000	0.0000	
7	0.000000	0.00	0.017857	0.0000	
8	0.000000	0.00	0.034483	0.0000	
9	0.000000	0.00	0.000000	0.0000	
10	0.000000	0.00	0.000000	0.0000	
11	0.000000	0.00	0.014706	0.0000	
12	0.037037	0.00	0.037037	0.0000	
13	0.000000	0.00	0.000000	0.0000	
14	0.000000	0.01	0.010000	0.0000	

	Furniture / Home Store	Gift Shop	Go Kart Track	Greek Restaurant	\
0	0.0000	0.000000	0.00	0.000000	
1	0.0625	0.000000	0.00	0.000000	
2	0.0000	0.000000	0.00	0.000000	
3	0.0000	0.000000	0.00	0.000000	

4	0.0000	0.000000	0.00	0.000000
5	0.0000	0.000000	0.00	0.000000
6	0.0000	0.000000	0.00	0.000000
7	0.0000	0.000000	0.00	0.017857
8	0.0000	0.000000	0.00	0.000000
9	0.0000	0.000000	0.00	0.000000
10	0.0000	0.000000	0.00	0.000000
11	0.0000	0.014706	0.00	0.000000
12	0.0000	0.037037	0.00	0.000000
13	0.0000	0.000000	0.00	0.000000
14	0.0300	0.010000	0.01	0.010000

	Grocery Store	Gym	Gym / Fitness Center	Gym Pool	Gymnastics Gym \
0	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.031250	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.058824	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000
5	0.037037	0.037037	0.000000	0.037037	0.000000
6	0.000000	0.025641	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.034483	0.000000	0.000000
9	0.000000	0.166667	0.000000	0.000000	0.083333
10	0.000000	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.014706	0.014706	0.000000	0.000000
12	0.000000	0.037037	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000	0.000000

	Harbor / Marina	Health & Beauty Service	Hookah Bar	Hostel \
0	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.025641	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.014706	0.014706
12	0.037037	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.020000	0.000000

	Hot Spring	Hotel	Hotel Bar	IT Services	Ice Cream Shop \

0	0.000000	0.333333	0.000000	0.00	0.000000
1	0.000000	0.031250	0.000000	0.00	0.031250
2	0.000000	0.000000	0.000000	0.00	0.026316
3	0.000000	0.000000	0.000000	0.00	0.058824
4	0.142857	0.142857	0.000000	0.00	0.000000
5	0.000000	0.037037	0.000000	0.00	0.000000
6	0.000000	0.128205	0.025641	0.00	0.025641
7	0.000000	0.017857	0.000000	0.00	0.000000
8	0.000000	0.000000	0.000000	0.00	0.000000
9	0.000000	0.000000	0.000000	0.00	0.000000
10	0.000000	0.000000	0.000000	0.00	0.000000
11	0.000000	0.058824	0.000000	0.00	0.029412
12	0.000000	0.074074	0.000000	0.00	0.000000
13	0.000000	0.000000	0.000000	0.00	0.000000
14	0.000000	0.000000	0.000000	0.01	0.010000

	Indian Restaurant	Italian Restaurant	Japanese Restaurant	Jewelry Store	\
0	0.000000	0.000000	0.000000	0.000000	
1	0.062500	0.000000	0.000000	0.000000	
2	0.447368	0.000000	0.026316	0.000000	
3	0.000000	0.000000	0.000000	0.000000	
4	0.000000	0.000000	0.000000	0.000000	
5	0.037037	0.037037	0.000000	0.000000	
6	0.025641	0.051282	0.000000	0.000000	
7	0.017857	0.035714	0.017857	0.035714	
8	0.000000	0.034483	0.000000	0.000000	
9	0.000000	0.000000	0.000000	0.000000	
10	0.000000	0.000000	0.000000	0.000000	
11	0.073529	0.000000	0.000000	0.000000	
12	0.037037	0.074074	0.037037	0.000000	
13	0.000000	0.166667	0.000000	0.000000	
14	0.010000	0.040000	0.010000	0.000000	

	Juice Bar	Korean Restaurant	Lebanese Restaurant	Lingerie Store	\
0	0.000000	0.000000	0.000000	0.00	
1	0.000000	0.000000	0.000000	0.00	
2	0.000000	0.026316	0.000000	0.00	
3	0.058824	0.000000	0.000000	0.00	
4	0.000000	0.000000	0.000000	0.00	
5	0.000000	0.000000	0.000000	0.00	
6	0.000000	0.000000	0.000000	0.00	
7	0.000000	0.000000	0.000000	0.00	
8	0.000000	0.034483	0.000000	0.00	
9	0.000000	0.000000	0.000000	0.00	
10	0.000000	0.000000	0.000000	0.00	
11	0.014706	0.000000	0.000000	0.00	
12	0.000000	0.000000	0.037037	0.00	

13	0.000000	0.000000	0.000000	0.00
14	0.010000	0.000000	0.000000	0.01

	Lounge	Medical Center	Mediterranean Restaurant	Mexican Restaurant \
0	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.142857	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000
6	0.025641	0.000000	0.000000	0.000000
7	0.035714	0.000000	0.017857	0.035714
8	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000

	Middle Eastern Restaurant	Miscellaneous Shop	Moroccan Restaurant \
0	0.000000	0.000000	0.000000
1	0.156250	0.000000	0.000000
2	0.026316	0.000000	0.000000
3	0.058824	0.000000	0.000000
4	0.000000	0.000000	0.000000
5	0.074074	0.000000	0.000000
6	0.000000	0.000000	0.025641
7	0.035714	0.000000	0.000000
8	0.034483	0.000000	0.000000
9	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000
11	0.088235	0.014706	0.000000
12	0.037037	0.000000	0.000000
13	0.000000	0.000000	0.000000
14	0.020000	0.000000	0.000000

	Movie Theater	Multiplex	Nail Salon	Nightclub	Optical Shop \
0	0.000000	0.000000	0.000000	0.000000	0.00
1	0.062500	0.000000	0.000000	0.000000	0.00
2	0.000000	0.000000	0.000000	0.000000	0.00
3	0.000000	0.000000	0.000000	0.000000	0.00
4	0.000000	0.000000	0.000000	0.000000	0.00
5	0.000000	0.000000	0.000000	0.000000	0.00
6	0.025641	0.000000	0.000000	0.000000	0.00
7	0.000000	0.017857	0.000000	0.000000	0.00
8	0.000000	0.000000	0.034483	0.000000	0.00

9	0.000000	0.000000	0.000000	0.000000	0.00
10	0.000000	0.000000	0.000000	0.000000	0.00
11	0.014706	0.000000	0.000000	0.000000	0.00
12	0.000000	0.000000	0.000000	0.037037	0.00
13	0.000000	0.000000	0.000000	0.000000	0.00
14	0.030000	0.000000	0.000000	0.000000	0.01

	Pakistani Restaurant	Park	Peruvian Restaurant	Pet Store	Pharmacy	\
0	0.000000	0.000000	0.000000	0.000000	0.000000	
1	0.000000	0.000000	0.000000	0.000000	0.000000	
2	0.026316	0.026316	0.000000	0.000000	0.000000	
3	0.000000	0.000000	0.000000	0.000000	0.000000	
4	0.000000	0.000000	0.000000	0.000000	0.000000	
5	0.000000	0.000000	0.000000	0.000000	0.037037	
6	0.000000	0.025641	0.000000	0.000000	0.000000	
7	0.000000	0.000000	0.017857	0.000000	0.000000	
8	0.000000	0.000000	0.000000	0.000000	0.000000	
9	0.000000	0.000000	0.000000	0.083333	0.000000	
10	0.000000	0.000000	0.000000	0.000000	0.000000	
11	0.000000	0.000000	0.000000	0.000000	0.000000	
12	0.000000	0.000000	0.000000	0.000000	0.000000	
13	0.000000	0.000000	0.000000	0.000000	0.000000	
14	0.000000	0.000000	0.000000	0.000000	0.000000	

	Pizza Place	Playground	Plaza	Pool	Pub	Racetrack	\
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	
1	0.031250	0.000000	0.000000	0.000000	0.000000	0.00	
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	
5	0.037037	0.000000	0.000000	0.000000	0.000000	0.00	
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	
7	0.000000	0.000000	0.017857	0.017857	0.000000	0.00	
8	0.034483	0.000000	0.000000	0.000000	0.000000	0.00	
9	0.083333	0.000000	0.000000	0.250000	0.000000	0.00	
10	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	
11	0.029412	0.014706	0.000000	0.000000	0.014706	0.00	
12	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	
13	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	
14	0.010000	0.000000	0.000000	0.000000	0.000000	0.01	

	Residential Building (Apartment / Condo)	Resort	Restaurant	\
0	0.000000	0.000000	0.000000	
1	0.000000	0.000000	0.000000	
2	0.000000	0.000000	0.026316	
3	0.000000	0.000000	0.000000	
4	0.000000	0.000000	0.000000	

```
5                                  0.000000  0.037037    0.037037
6                                  0.000000  0.025641    0.025641
7                                  0.017857  0.017857    0.000000
8                                  0.000000  0.000000    0.000000
9                                  0.083333  0.000000    0.000000
10                                 0.000000  0.000000    0.000000
11                                 0.000000  0.000000    0.014706
12                                 0.000000  0.000000    0.037037
13                                 0.000000  0.000000    0.000000
14                                 0.000000  0.010000    0.000000

    Sandwich Place  Seafood Restaurant  Shawarma Place  Shoe Store  \
0         0.166667            0.000000        0.000000    0.000000
1         0.000000            0.031250        0.000000    0.000000
2         0.000000            0.000000        0.000000    0.000000
3         0.000000            0.000000        0.058824    0.000000
4         0.000000            0.000000        0.000000    0.000000
5         0.000000            0.000000        0.000000    0.000000
6         0.000000            0.025641        0.000000    0.000000
7         0.000000            0.000000        0.000000    0.017857
8         0.034483            0.034483        0.000000    0.000000
9         0.000000            0.000000        0.000000    0.000000
10        0.000000            0.000000        0.000000    0.000000
11        0.014706            0.000000        0.000000    0.000000
12        0.000000            0.000000        0.000000    0.000000
13        0.000000            0.000000        0.000000    0.000000
14        0.020000            0.000000        0.000000    0.000000

    Shopping Mall  Snack Place  South Indian Restaurant       Spa  \
0        0.000000      0.00000                 0.000000  0.000000
1        0.000000      0.03125                 0.000000  0.000000
2        0.000000      0.00000                 0.026316  0.000000
3        0.000000      0.00000                 0.000000  0.000000
4        0.000000      0.00000                 0.000000  0.000000
5        0.037037      0.00000                 0.000000  0.000000
6        0.000000      0.00000                 0.000000  0.000000
7        0.017857      0.00000                 0.000000  0.017857
8        0.000000      0.00000                 0.000000  0.000000
9        0.000000      0.00000                 0.000000  0.000000
10       0.000000      0.00000                 0.000000  0.000000
11       0.029412      0.00000                 0.000000  0.000000
12       0.037037      0.00000                 0.000000  0.074074
13       0.166667      0.00000                 0.000000  0.000000
14       0.010000      0.00000                 0.000000  0.000000

    Sporting Goods Shop  Sports Bar  Steakhouse  Supermarket  \
0                  0.00    0.000000    0.000000     0.000000
```

1	0.00	0.000000	0.000000	0.000000
2	0.00	0.000000	0.000000	0.026316
3	0.00	0.000000	0.000000	0.058824
4	0.00	0.000000	0.000000	0.000000
5	0.00	0.000000	0.000000	0.000000
6	0.00	0.025641	0.000000	0.000000
7	0.00	0.000000	0.035714	0.017857
8	0.00	0.000000	0.000000	0.000000
9	0.00	0.000000	0.000000	0.000000
10	0.00	0.000000	0.000000	0.000000
11	0.00	0.000000	0.000000	0.000000
12	0.00	0.000000	0.037037	0.000000
13	0.00	0.000000	0.000000	0.166667
14	0.04	0.000000	0.030000	0.010000

	Sushi Restaurant	Tennis Court	Theater	Theme Park \
0	0.000000	0.000000	0.00	0.00
1	0.000000	0.000000	0.00	0.00
2	0.000000	0.000000	0.00	0.00
3	0.000000	0.000000	0.00	0.00
4	0.000000	0.000000	0.00	0.00
5	0.000000	0.000000	0.00	0.00
6	0.000000	0.000000	0.00	0.00
7	0.035714	0.000000	0.00	0.00
8	0.000000	0.000000	0.00	0.00
9	0.000000	0.083333	0.00	0.00
10	0.000000	0.000000	0.00	0.00
11	0.000000	0.000000	0.00	0.00
12	0.000000	0.000000	0.00	0.00
13	0.166667	0.000000	0.00	0.00
14	0.000000	0.000000	0.01	0.03

	Theme Park Ride / Attraction	Theme Restaurant	Toy / Game Store \
0	0.00	0.000000	0.00
1	0.00	0.000000	0.00
2	0.00	0.026316	0.00
3	0.00	0.000000	0.00
4	0.00	0.000000	0.00
5	0.00	0.000000	0.00
6	0.00	0.000000	0.00
7	0.00	0.000000	0.00
8	0.00	0.000000	0.00
9	0.00	0.000000	0.00
10	0.00	0.000000	0.00
11	0.00	0.000000	0.00
12	0.00	0.000000	0.00
13	0.00	0.000000	0.00

```
14                            0.12          0.000000              0.03

    Turkish Restaurant  Vegetarian / Vegan Restaurant  Wine Bar
0              0.00000                       0.000000  0.000000
1              0.03125                       0.000000  0.000000
2              0.00000                       0.026316  0.000000
3              0.00000                       0.000000  0.000000
4              0.00000                       0.000000  0.000000
5              0.00000                       0.000000  0.000000
6              0.00000                       0.000000  0.000000
7              0.00000                       0.000000  0.017857
8              0.00000                       0.000000  0.000000
9              0.00000                       0.000000  0.000000
10             0.00000                       0.000000  0.000000
11             0.00000                       0.000000  0.014706
12             0.00000                       0.000000  0.000000
13             0.00000                       0.000000  0.000000
14             0.00000                       0.000000  0.000000
```

We now have the frequency for each neighborhood.
What are the top 5 venues for each neighborhood?

[77]:
```python
num_top_venues = 5

for hood in AD_grouped['Neighborhood']:
    print("----"+hood+"----")
    temp = AD_grouped[AD_grouped['Neighborhood'] == hood].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).
 ↪head(num_top_venues))
    print('\n')
```

```
----Al Aman----
                 venue  freq
0                Hotel  0.33
1  Fast Food Restaurant  0.17
2          Coffee Shop  0.17
3       Sandwich Place  0.17
4                 Café  0.17


----Al Dhafrah----
                      venue  freq
0  Middle Eastern Restaurant  0.16
1                      Café  0.09
```

```
2            Movie Theater  0.06
3      Fried Chicken Joint  0.06
4     Fast Food Restaurant  0.06


----Al Karama----
               venue  freq
0  Indian Restaurant  0.45
1   Asian Restaurant  0.11
2             Bakery  0.08
3        Supermarket  0.03
4               Park  0.03


----Al Khubeirah----
                  venue  freq
0           Coffee Shop  0.24
1                  Café  0.24
2  Fast Food Restaurant  0.12
3           Supermarket  0.06
4        Shawarma Place  0.06


----Al Manhal----
            venue  freq
0            Café  0.43
1           Hotel  0.14
2      Hot Spring  0.14
3  Medical Center  0.14
4     Coffee Shop  0.14


----Al Maqtaa----
                       venue  freq
0                       Café  0.19
1  Middle Eastern Restaurant  0.07
2                Coffee Shop  0.07
3               Burger Joint  0.04
4          French Restaurant  0.04


----Al Markaziyah----
                  venue  freq
0  Fast Food Restaurant  0.13
1                 Hotel  0.13
2                  Café  0.08
3      Asian Restaurant  0.05
4    Italian Restaurant  0.05
```

```
----Al Maryah Island----
               venue  freq
0               Café  0.09
1        Coffee Shop  0.07
2  Mexican Restaurant  0.04
3      Jewelry Store  0.04
4   Sushi Restaurant  0.04


----Al Nahyan----
         venue  freq
0         Café  0.17
1  Coffee Shop  0.10
2  Flower Shop  0.07
3       Bakery  0.07
4  Dessert Shop  0.07


----Al Reef----
                                    venue  freq
0                                    Pool  0.25
1                                     Gym  0.17
2                            Tennis Court  0.08
3                               Pet Store  0.08
4  Residential Building (Apartment / Condo)  0.08


----Al Rowdah----
           venue  freq
0           Café  0.62
1    Coffee Shop  0.25
2  Cosmetics Shop  0.12
3   Women's Store  0.00
4       Multiplex  0.00


----Al Zahiyah----
                      venue  freq
0  Middle Eastern Restaurant  0.09
1          Indian Restaurant  0.07
2                      Hotel  0.06
3               Dessert Shop  0.04
4       Fast Food Restaurant  0.04


----Bain Al Jisrain----
```

```
                venue  freq
0         Coffee Shop  0.11
1                 Spa  0.07
2  Italian Restaurant  0.07
3               Hotel  0.07
4 Lebanese Restaurant  0.04


----Masdar City----
                 venue  freq
0        Shopping Mall  0.17
1     Sushi Restaurant  0.17
2   Italian Restaurant  0.17
3 Fast Food Restaurant  0.17
4                 Café  0.17


----Yas Island----
                         venue  freq
0  Theme Park Ride / Attraction  0.12
1                          Café  0.11
2                   Coffee Shop  0.07
3            Italian Restaurant  0.04
4            Sporting Goods Shop  0.04
```

Let's put this information in a dataframe, with the top 5 most common venues for each neighborhood.

[78]:
```
#This function sorts the venues in descending order
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]
```

[79]:
```
num_top_venues = 5

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

# create columns according to number of top venues
columns = ['Neighborhood']
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))
```

```python
# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = AD_grouped['Neighborhood']

for ind in np.arange(AD_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] =␣
 ↪return_most_common_venues(AD_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

[79]:

```
    Neighborhood      1st Most Common Venue 2nd Most Common Venue  \
0        Al Aman                      Hotel           Coffee Shop
1    Al Dhafrah  Middle Eastern Restaurant                  Café
2      Al Karama          Indian Restaurant      Asian Restaurant
3   Al Khubeirah                       Café           Coffee Shop
4      Al Manhal                       Café            Hot Spring

  3rd Most Common Venue   4th Most Common Venue 5th Most Common Venue
0  Fast Food Restaurant          Sandwich Place                  Café
1   Fried Chicken Joint  Furniture / Home Store  Fast Food Restaurant
2                Bakery          Ice Cream Shop             Cafeteria
3  Fast Food Restaurant              Donut Shop  Gym / Fitness Center
4        Medical Center             Coffee Shop                 Hotel
```

Let's have a look at the whole thing…

[80]:

```python
neighborhoods_venues_sorted
```

[80]:

```
         Neighborhood         1st Most Common Venue 2nd Most Common Venue  \
0             Al Aman                         Hotel           Coffee Shop
1          Al Dhafrah     Middle Eastern Restaurant                  Café
2           Al Karama             Indian Restaurant      Asian Restaurant
3        Al Khubeirah                          Café           Coffee Shop
4           Al Manhal                          Café            Hot Spring
5           Al Maqtaa                          Café           Coffee Shop
6       Al Markaziyah          Fast Food Restaurant                 Hotel
7    Al Maryah Island                          Café           Coffee Shop
8           Al Nahyan                          Café           Coffee Shop
9             Al Reef                          Pool                   Gym
10          Al Rowdah                          Café           Coffee Shop
11         Al Zahiyah     Middle Eastern Restaurant     Indian Restaurant
12     Bain Al Jisrain                   Coffee Shop                   Spa
13        Masdar City             Sushi Restaurant    Italian Restaurant
14         Yas Island  Theme Park Ride / Attraction                  Café

        3rd Most Common Venue      4th Most Common Venue 5th Most Common Venue
```

```
0        Fast Food Restaurant             Sandwich Place                  Café
1         Fried Chicken Joint      Furniture / Home Store  Fast Food Restaurant
2                      Bakery             Ice Cream Shop             Cafeteria
3        Fast Food Restaurant                 Donut Shop  Gym / Fitness Center
4              Medical Center                Coffee Shop                 Hotel
5   Middle Eastern Restaurant                   Pharmacy         Shopping Mall
6                        Café         Italian Restaurant      Asian Restaurant
7            Sushi Restaurant  Middle Eastern Restaurant   American Restaurant
8                Dessert Shop                Flower Shop                Bakery
9                 Pizza Place          Convenience Store           Coffee Shop
10             Cosmetics Shop                   Wine Bar            Donut Shop
11                      Hotel                Coffee Shop  Fast Food Restaurant
12                      Hotel         Italian Restaurant   Lebanese Restaurant
13                       Café                Supermarket  Fast Food Restaurant
14                Coffee Shop             Clothing Store   Sporting Goods Shop
```

Now we can attempt to cluster these neighborhoods based on how similar its businesses are.

[81]:
```
# set number of clusters
kclusters = 5

AD_grouped_clustering = AD_grouped.drop('Neighborhood', 1)

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

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:5]
```

[81]:
```
array([0, 1, 3, 1, 2], dtype=int32)
```

[82]:
```
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster_Labels', kmeans.labels_)

AD_merged = df

# merge AD_grouped with AD_data to add latitude/longitude for each neighborhood
AD_merged = AD_merged.join(neighborhoods_venues_sorted.
↪set_index('Neighborhood'), on='Neighborhood')

AD_merged.head()
```

[82]:
```
  Neighborhood      Lat      Lon  Cluster_Labels      1st Most Common Venue  \
0      Al Aman   24.432  54.4266             0.0                      Hotel
1    Al Bateen  24.2151  55.6263             NaN                        NaN
2   Al Dhafrah  24.4761  54.3694             1.0  Middle Eastern Restaurant
3     Al Falah  24.4447  54.7282             NaN                        NaN
```

```
4    Al Karama  25.2444  55.3048            3.0          Indian Restaurant

  2nd Most Common Venue 3rd Most Common Venue   4th Most Common Venue  \
0           Coffee Shop  Fast Food Restaurant          Sandwich Place
1                   NaN                   NaN                     NaN
2                  Café    Fried Chicken Joint  Furniture / Home Store
3                   NaN                   NaN                     NaN
4      Asian Restaurant                Bakery           Ice Cream Shop

  5th Most Common Venue
0                  Café
1                   NaN
2  Fast Food Restaurant
3                   NaN
4             Cafeteria
```

[83]:
```
AD_merged
```

[83]:
```
               Neighborhood      Lat      Lon  Cluster_Labels  \
0                   Al Aman   24.432  54.4266             0.0
1                 Al Bateen  24.2151  55.6263             NaN
2                Al Dhafrah  24.4761  54.3694             1.0
3                  Al Falah  24.4447  54.7282             NaN
4                 Al Karama  25.2444  55.3048             3.0
5              Al Khubeirah  24.4652  54.3368             1.0
6            Al Lulu Island  24.4996  54.3457             NaN
7                 Al Madina  24.3409  54.4907             NaN
8          Al Maryah Island  24.5021  54.3902             1.0
9                 Al Manhal  24.4666   54.366             2.0
10                Al Maqtaa  24.4346  54.4544             1.0
11            Al Markaziyah  24.4933  54.3667             1.0
12               Al Mushrif  24.4369   54.391             NaN
13                Al Nahyan  24.4684  54.3852             1.0
14                  Al Reef  24.4577  54.6737             4.0
15           Al Reem Island  24.4867  54.4105             NaN
16                Al Rowdah  24.4564  54.3596             2.0
17               Al Shamkha  24.3885  54.7184             NaN
18               Al Zahiyah  24.4933  54.3799             1.0
19                Al Zahraa  24.4392  54.4352             NaN
20          Bain Al Jisrain  24.4092  54.4954             1.0
21             Khalifa City  24.4214  54.5774             NaN
22              Masdar City  24.4259  54.6193             1.0
23  Mohammed Bin Zayed City  24.3279  54.5624             NaN
24          Saadiyat Island  24.5303  54.4409             NaN
25           Shakhbout City  24.3622   54.631             NaN
26            Officers City  24.3824   54.535             NaN
27             Qasr El Bahr  24.4746  54.3966             NaN
```

```
28              Yas Island  24.4864  54.6091            1.0

          1st Most Common Venue 2nd Most Common Venue  \
0                         Hotel           Coffee Shop
1                           NaN                   NaN
2     Middle Eastern Restaurant                  Café
3                           NaN                   NaN
4             Indian Restaurant      Asian Restaurant
5                          Café           Coffee Shop
6                           NaN                   NaN
7                           NaN                   NaN
8                          Café           Coffee Shop
9                          Café            Hot Spring
10                         Café           Coffee Shop
11          Fast Food Restaurant                 Hotel
12                          NaN                   NaN
13                         Café           Coffee Shop
14                         Pool                   Gym
15                          NaN                   NaN
16                         Café           Coffee Shop
17                          NaN                   NaN
18    Middle Eastern Restaurant     Indian Restaurant
19                          NaN                   NaN
20                  Coffee Shop                   Spa
21                          NaN                   NaN
22             Sushi Restaurant    Italian Restaurant
23                          NaN                   NaN
24                          NaN                   NaN
25                          NaN                   NaN
26                          NaN                   NaN
27                          NaN                   NaN
28  Theme Park Ride / Attraction                  Café

       3rd Most Common Venue      4th Most Common Venue 5th Most Common Venue
0       Fast Food Restaurant             Sandwich Place                  Café
1                        NaN                        NaN                   NaN
2        Fried Chicken Joint     Furniture / Home Store  Fast Food Restaurant
3                        NaN                        NaN                   NaN
4                     Bakery             Ice Cream Shop             Cafeteria
5       Fast Food Restaurant                Donut Shop  Gym / Fitness Center
6                        NaN                        NaN                   NaN
7                        NaN                        NaN                   NaN
8           Sushi Restaurant  Middle Eastern Restaurant   American Restaurant
9             Medical Center               Coffee Shop                 Hotel
10  Middle Eastern Restaurant                  Pharmacy         Shopping Mall
11                      Café         Italian Restaurant      Asian Restaurant
12                       NaN                        NaN                   NaN
```

```
13              Dessert Shop           Flower Shop              Bakery
14               Pizza Place     Convenience Store          Coffee Shop
15                       NaN                   NaN                  NaN
16             Cosmetics Shop              Wine Bar          Donut Shop
17                       NaN                   NaN                  NaN
18                     Hotel           Coffee Shop  Fast Food Restaurant
19                       NaN                   NaN                  NaN
20                     Hotel   Italian Restaurant   Lebanese Restaurant
21                       NaN                   NaN                  NaN
22                      Café           Supermarket  Fast Food Restaurant
23                       NaN                   NaN                  NaN
24                       NaN                   NaN                  NaN
25                       NaN                   NaN                  NaN
26                       NaN                   NaN                  NaN
27                       NaN                   NaN                  NaN
28               Coffee Shop        Clothing Store  Sporting Goods Shop
```

Since we used our original dataframe some of the neighborhoods in which no venues were found on Foursquare reappeared here.
Let's drop these NA's. Also we will change the cluster values to integer.

[84]:
```python
AD_merged=AD_merged.dropna()
AD_merged['Cluster_Labels'] = AD_merged.Cluster_Labels.astype(int)
```

```
/home/jupyterlab/conda/envs/python/lib/python3.6/site-
packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

[85]:
```python
AD_merged
```

[85]:
```
         Neighborhood      Lat      Lon  Cluster_Labels  \
0             Al Aman   24.432  54.4266               0
2          Al Dhafrah  24.4761  54.3694               1
4           Al Karama  25.2444  55.3048               3
5        Al Khubeirah  24.4652  54.3368               1
8    Al Maryah Island  24.5021  54.3902               1
9           Al Manhal  24.4666   54.366               2
10          Al Maqtaa  24.4346  54.4544               1
11      Al Markaziyah  24.4933  54.3667               1
13          Al Nahyan  24.4684  54.3852               1
14            Al Reef  24.4577  54.6737               4
16          Al Rowdah  24.4564  54.3596               2
```

```
18        Al Zahiyah  24.4933  54.3799               1
20   Bain Al Jisrain  24.4092  54.4954               1
22       Masdar City  24.4259  54.6193               1
28        Yas Island  24.4864  54.6091               1

          1st Most Common Venue 2nd Most Common Venue  \
0                         Hotel           Coffee Shop
2     Middle Eastern Restaurant                  Café
4             Indian Restaurant      Asian Restaurant
5                          Café           Coffee Shop
8                          Café           Coffee Shop
9                          Café            Hot Spring
10                         Café           Coffee Shop
11         Fast Food Restaurant                 Hotel
13                         Café           Coffee Shop
14                         Pool                   Gym
16                         Café           Coffee Shop
18    Middle Eastern Restaurant     Indian Restaurant
20                  Coffee Shop                   Spa
22            Sushi Restaurant    Italian Restaurant
28  Theme Park Ride / Attraction                 Café

        3rd Most Common Venue      4th Most Common Venue 5th Most Common Venue
0        Fast Food Restaurant             Sandwich Place                  Café
2         Fried Chicken Joint     Furniture / Home Store  Fast Food Restaurant
4                      Bakery             Ice Cream Shop            Cafeteria
5        Fast Food Restaurant                Donut Shop  Gym / Fitness Center
8            Sushi Restaurant  Middle Eastern Restaurant   American Restaurant
9              Medical Center                Coffee Shop                 Hotel
10  Middle Eastern Restaurant                   Pharmacy         Shopping Mall
11                       Café         Italian Restaurant      Asian Restaurant
13               Dessert Shop                Flower Shop                Bakery
14                Pizza Place          Convenience Store           Coffee Shop
16             Cosmetics Shop                   Wine Bar            Donut Shop
18                      Hotel                Coffee Shop  Fast Food Restaurant
20                      Hotel         Italian Restaurant   Lebanese Restaurant
22                       Café                Supermarket  Fast Food Restaurant
28                Coffee Shop             Clothing Store    Sporting Goods Shop
```

Now we can map the neighborhoods differentiating the clusters with different colors.

[86]:
```
# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
```

```
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(AD_merged['Lat'], AD_merged['Lon'],␣
↪AD_merged['Neighborhood'], AD_merged['Cluster_Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

[86]:
```
<folium.folium.Map at 0x7fb4ca8455f8>
```

Let's inspect cluster 1, the most common one.

[87]:
```
AD_merged.loc[AD_merged['Cluster_Labels'] == 1, AD_merged.columns[[0] +␣
↪list(range(4, AD_merged.shape[1]))]]
```

[87]:
```
        Neighborhood        1st Most Common Venue 2nd Most Common Venue  \
2         Al Dhafrah     Middle Eastern Restaurant                  Café
5        Al Khubeirah                         Café           Coffee Shop
8    Al Maryah Island                         Café           Coffee Shop
10          Al Maqtaa                         Café           Coffee Shop
11      Al Markaziyah         Fast Food Restaurant                 Hotel
13          Al Nahyan                         Café           Coffee Shop
18         Al Zahiyah     Middle Eastern Restaurant     Indian Restaurant
20    Bain Al Jisrain                  Coffee Shop                   Spa
22        Masdar City            Sushi Restaurant    Italian Restaurant
28         Yas Island  Theme Park Ride / Attraction                 Café

        3rd Most Common Venue      4th Most Common Venue 5th Most Common Venue
2          Fried Chicken Joint     Furniture / Home Store  Fast Food Restaurant
5         Fast Food Restaurant                 Donut Shop  Gym / Fitness Center
8             Sushi Restaurant  Middle Eastern Restaurant   American Restaurant
10   Middle Eastern Restaurant                   Pharmacy         Shopping Mall
11                        Café         Italian Restaurant     Asian Restaurant
13                Dessert Shop                Flower Shop                Bakery
18                       Hotel                Coffee Shop  Fast Food Restaurant
20                       Hotel         Italian Restaurant   Lebanese Restaurant
```

```
22                         Café              Supermarket   Fast Food Restaurant
28                  Coffee Shop           Clothing Store    Sporting Goods Shop
```

We can already get some insights from this table. We see that our business idea, a coffee shop, is very present is most areas. But let's take a look at Al Zahiyah.

[88]:
```python
AD_merged.loc[AD_merged['Neighborhood'] == 'Al Zahiyah']
```

[88]:
```
   Neighborhood      Lat      Lon  Cluster_Labels      1st Most Common Venue  \
18   Al Zahiyah  24.4933  54.3799               1  Middle Eastern Restaurant

   2nd Most Common Venue 3rd Most Common Venue 4th Most Common Venue  \
18     Indian Restaurant                 Hotel           Coffee Shop

   5th Most Common Venue
18  Fast Food Restaurant
```

Ok, coffee shop appears only in 4th place. So perhaps there is an opportunity here. This could give a potential business owner a starting point to hers/his research.

I would like to remove coffee shops from the equation and recluster (is that a word?) the neighborhoods. It would be nice to see if maybe we get different neighborhoods that might be clustered together, which could suggest that a coffee shop would be a good investment in area X, because neighborhood Y has a couple, but it was clustered together with X when we removed coffee shops.

However, our current database does not have enough venues to attemp this.

Before we go, let's take a look at our other clusters.

[89]:
```python
AD_merged.loc[AD_merged['Cluster_Labels'] == 0, AD_merged.columns[[0] +␣
 ↪list(range(4, AD_merged.shape[1]))]]
```

[89]:
```
  Neighborhood 1st Most Common Venue 2nd Most Common Venue  \
0      Al Aman                 Hotel           Coffee Shop

  3rd Most Common Venue 4th Most Common Venue 5th Most Common Venue
0  Fast Food Restaurant        Sandwich Place                  Café
```

[90]:
```python
AD_merged.loc[AD_merged['Cluster_Labels'] == 2, AD_merged.columns[[0] +␣
 ↪list(range(4, AD_merged.shape[1]))]]
```

[90]:
```
   Neighborhood 1st Most Common Venue 2nd Most Common Venue  \
9     Al Manhal                  Café            Hot Spring
16    Al Rowdah                  Café           Coffee Shop

   3rd Most Common Venue 4th Most Common Venue 5th Most Common Venue
9         Medical Center           Coffee Shop                 Hotel
16        Cosmetics Shop              Wine Bar           Donut Shop
```

[91]:
```
AD_merged.loc[AD_merged['Cluster_Labels'] == 3, AD_merged.columns[[0] +␣
↪list(range(4, AD_merged.shape[1]))]]
```

[91]:
```
  Neighborhood 1st Most Common Venue 2nd Most Common Venue  \
4    Al Karama     Indian Restaurant      Asian Restaurant

  3rd Most Common Venue 4th Most Common Venue 5th Most Common Venue
4                Bakery        Ice Cream Shop             Cafeteria
```

[92]:
```
AD_merged.loc[AD_merged['Cluster_Labels'] == 4, AD_merged.columns[[0] +␣
↪list(range(4, AD_merged.shape[1]))]]
```

[92]:
```
   Neighborhood 1st Most Common Venue 2nd Most Common Venue  \
14      Al Reef                  Pool                   Gym

   3rd Most Common Venue 4th Most Common Venue 5th Most Common Venue
14           Pizza Place     Convenience Store           Coffee Shop
```

For this last cluster we see Al Reef, which is an up-and-coming residential area. Maybe it could use another coffee shop?

1.4 Discussion and Final Thoughts...

As you can see, there a lot of tools to work with here and make some interesting analysis. Of course, we could always use more data. We can find the locations of the other neighborhoods, even if we have to do it manually.

We showed that there aren't many venues listed on Foursquare for some areas. We could use other databases to see if we get better results. It would also be interesting to integrate this with information about if an area is a business area or a more residential one. As mentioned, we can also explore further the clustering method as well.

However, even with limitations, even if we do not get clear cut conclusions for this exercise, we do get some things from it: * We can at least get some insights and give a direction or focus for our research/work. * This was a great learning experience. Even though I had seen most of these methods during my Master's, I had done everything with R. A refreshment on the subject with Python was a great way to learn a new programming language. * Also, there was much more learned during the course that is not being used in this report. SQL, for example. Plus, as already mentioned in the Methodology, I was presented to a couple of really cool platforms to code and do analysis on, in IBM Skills Network Labs and IBM Cloud with Watson services. Have a look on those.

Thank you for reading this.

[ ]:
```

```