

IBM Data Science

Capstone Project – Battle of
Neighborhoods

1. Importing all the required Libraries

```
import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analysis
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 # transform 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('Required Libraries imported.')
```

2. Downloading and reading the Neighborhood data

The raw dataset is downloaded and imported into python. Data is downloaded from the opendatasoft site:

<https://public.opendatasoft.com/explore/dataset/zillow-neighborhoods/export/?refine.county=Napa>

```
df_data_0.head()
```

Out[4]:

	State	County	City	RegionID	Name	Geo Point	Geo Shape
0	CA	Sonoma	Santa Rosa	761261	Aston Ave	38.4244388347, -122.705772408	NaN
1	CA	Sonoma	Santa Rosa	761260	Olive Corby	38.4296357683, -122.718490047	NaN
2	CA	Sonoma	Santa Rosa	761262	South Park	38.4282367292, -122.707063534	NaN
3	CA	Sonoma	Santa Rosa	761259	Corby/Beachwood	38.4228326645, -122.719420258	NaN
4	CA	Sonoma	Santa Rosa	761458	Sunset McMinn	38.4244068703, -122.729983643	NaN

3 . Data Wrangling

Data transformations is done to bring the data to required shape.
Considered only required columns.

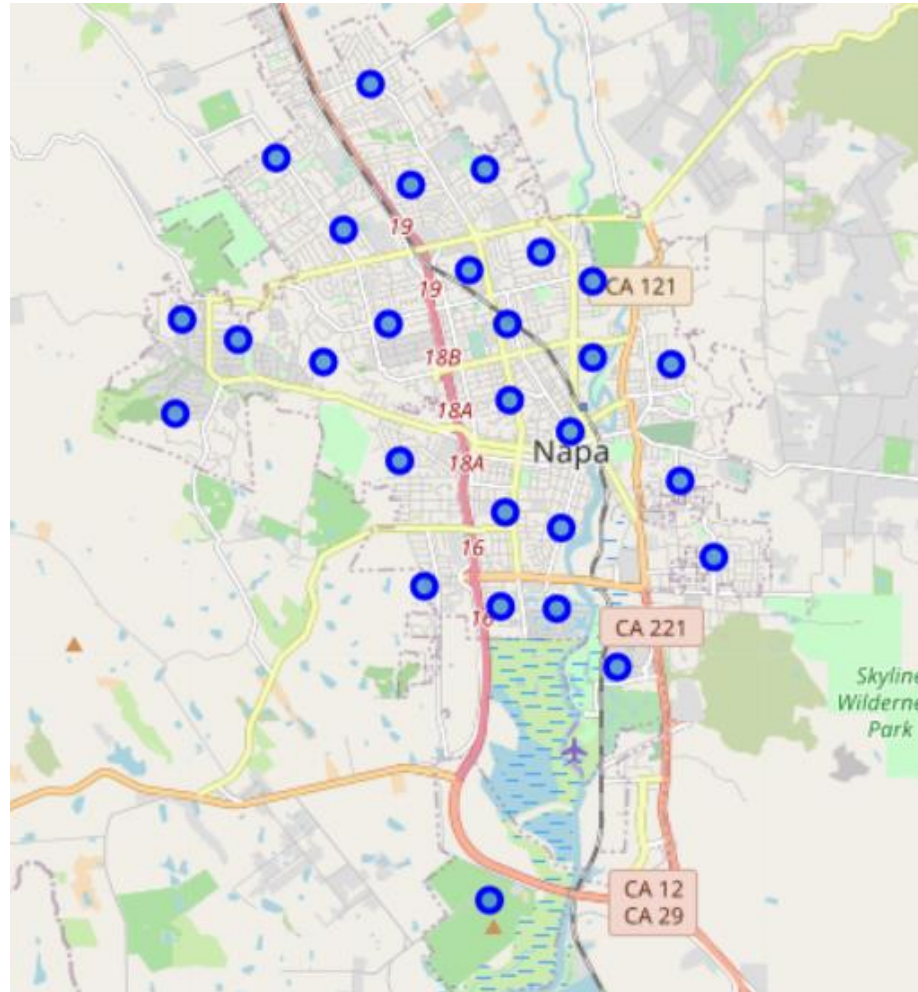
```
napa_data.head()
```

Out[21]:

	Borough	Neighborhood	Latitude	Longitude
0	Napa	Alta Heights	38.307360	-122.270429
1	Napa	Beard	38.318796	-122.301167
2	Napa	Bel Aire	38.329107	-122.310260
3	Napa	Browns Valley Central	38.310494	-122.336698
4	Napa	Browns Valley East	38.307756	-122.323485

4 . Mapping of Neighborhoods on the Map

Napa neighborhoods are mapped with the help of geopy library



5 . FourSqaure API – Explore Neighborhoods

With the help of Foursquare credentials, explored the neighborhoods and extracted venues within the radius of 1000 meters.

Out[36]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alta Heights	38.30736	-122.270429	Big D Burgers	38.301350	-122.276545	Burger Joint
1	Alta Heights	38.30736	-122.270429	Napa Tire Inc	38.310177	-122.276994	Automotive Shop
2	Alta Heights	38.30736	-122.270429	El Guadalajara	38.305510	-122.277364	Mexican Restaurant
3	Alta Heights	38.30736	-122.270429	RiverPointe Napa Valley Resort	38.311415	-122.279088	Resort
4	Alta Heights	38.30736	-122.270429	Compadres Rio Grille	38.309173	-122.279436	Mexican Restaurant

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Alta Heights	1	1	1	1	1	1
Beard	45	45	45	45	45	45
Bel Aire	3	3	3	3	3	3
Browns Valley Central	2	2	2	2	2	2
Browns Valley South	3	3	3	3	3	3
Browns Valley West	3	3	3	3	3	3
Central	4	4	4	4	4	4
Coventry Manor	1	1	1	1	1	1
Downtown	100	100	100	100	100	100

6. Analyze each Neighborhood

Each neighborhood in Napa is analyzed about venues and venue categories and selected 5 top venues.

	Neighborhood	ATM	Accessories Store	American Restaurant	Antique Shop	Art Gallery	Arts & Crafts Store	Asian Restaurant	Automotive Shop	BBQ Joint	Bagel Shop	Bakery	Bank
0	Alta Heights	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.0625	0.062500	0.000000	0.000000	0.000000	0.000000
1	Beard	0.012048	0.000000	0.024096	0.00	0.00	0.000000	0.0000	0.000000	0.012048	0.012048	0.012048	0.024096
2	Bel Aire	0.000000	0.000000	0.060000	0.00	0.00	0.000000	0.0000	0.000000	0.000000	0.000000	0.020000	0.000000
3	Browns Valley Central	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000
4	Browns Valley East	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000
5	Browns Valley South	0.000000	0.000000	0.166667	0.00	0.00	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Browns Valley West	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000
7	Central	0.010000	0.010000	0.060000	0.01	0.01	0.000000	0.0000	0.010000	0.030000	0.000000	0.000000	0.000000
8	Coventry Manor	0.000000	0.000000	0.000000	0.00	0.00	0.166667	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000
9	Downtown	0.000000	0.000000	0.080000	0.01	0.01	0.000000	0.0000	0.000000	0.010000	0.000000	0.020000	0.000000
10	Fuller Park	0.016393	0.016393	0.016393	0.00	0.00	0.016393	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000
11	Linda Vista	0.000000	0.000000	0.066667	0.00	0.00	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000

7. Analyze each Neighborhood

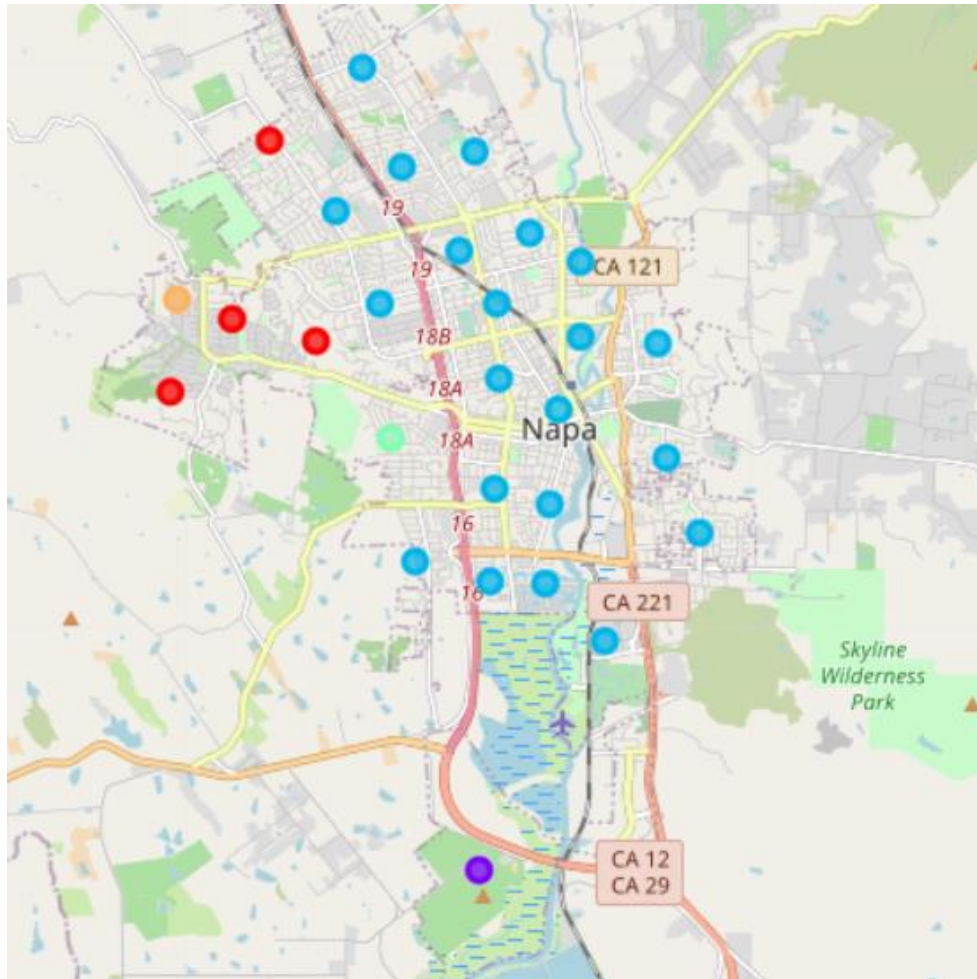
Each neighborhood in Napa is analyzed about venues and venue categories and selected common venues.

Neighborhood		ATM	Accessories Store	American Restaurant	Antique Shop	Art Gallery	Arts & Crafts Store	Asian Restaurant	Automotive Shop	BBQ Joint	Bagel Shop	Bakery	Bank
0	Alta Heights	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.0625	0.062500	0.000000	0.000000	0.000000	0.000000
1	Beard	0.012048	0.000000	0.024096	0.00	0.00	0.000000	0.0000	0.000000	0.012048	0.012048	0.012048	0.024096
2	Bel Aire	0.000000	0.000000	0.060000	0.00	0.00	0.000000	0.0000	0.000000	0.000000	0.000000	0.020000	0.000000
3	Browns Valley Central	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000	0.000000

Neighborhood		1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alta Heights	Mexican Restaurant	Restaurant	Trail	Child Care Service	Coffee Shop	Gas Station	Resort	Track	Burger Joint	Automotive Shop
1	Beard	Fast Food Restaurant	Coffee Shop	Mexican Restaurant	Grocery Store	Pizza Place	Steakhouse	Cosmetics Shop	Pet Store	Mobile Phone Shop	Liquor Store
2	Bel Aire	Hotel	American Restaurant	Fast Food Restaurant	Mexican Restaurant	Coffee Shop	Restaurant	Furniture / Home Store	Italian Restaurant	Mobile Phone Shop	Pharmacy
3	Browns Valley Central	Home Service	Grocery Store	Massage Studio	Gastropub	Coffee Shop	Park	Dessert Shop	Department Store	Diner	Discount Store

8. Clustering Neighborhoods

Running k-means to cluster the neighborhood into 5 clusters ($k = 5$).



9. Examining the Clusters

Cluster-1

```
napa_merged.loc[napa_merged['Cluster Labels'] == 0, napa_merged.columns[[1]+ list(range(5, napa_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Browns Valley Central	Home Service	Grocery Store	Massage Studio	Gastropub	Coffee Shop	Park	Dessert Shop	Department Store	Diner	Dis
4	Browns Valley East	Park	Construction & Landscaping	Grocery Store	Coffee Shop	Dessert Shop	Garden	Gastropub	Women's Store	Donut Shop	Fast Rest
5	Browns Valley South	Park	Lake	American Restaurant	Massage Studio	Scenic Lookout	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Farm	F Rest
24	Vineyard Estates	Food Truck	Park	Bus Station	Baseball Stadium	Grocery Store	Mexican Restaurant	Hotel Pool	Trail	Italian Restaurant	Tour Pr

Cluster-2

```
napa_merged.loc[napa_merged['Cluster Labels'] == 1, napa_merged.columns[[1]+ list(range(5, napa_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
21	Stanly Ranch	Cosmetics Shop	Farm	Food Truck	Flower Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Event Space	Donut

9. Examining the Clusters

Cluster- 3

```
napa_merged.loc[napa_merged['Cluster Labels'] == 2, napa_merged.columns[[1]+ list(range(5, napa_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alta Heights	Mexican Restaurant	Restaurant	Trail	Child Care Service	Coffee Shop	Gas Station	Resort	Track	Burger Joint	Automotive Shop
1	Beard	Fast Food Restaurant	Coffee Shop	Mexican Restaurant	Grocery Store	Pizza Place	Steakhouse	Cosmetics Shop	Pet Store	Mobile Phone Shop	Liquor Store
2	Bel Aire	Hotel	American Restaurant	Fast Food Restaurant	Mexican Restaurant	Coffee Shop	Restaurant	Furniture / Home Store	Italian Restaurant	Mobile Phone Shop	Pharmacy
7	Central	Hotel	Wine Bar	American Restaurant	Mexican Restaurant	Bed & Breakfast	Italian Restaurant	BBQ Joint	Pizza Place	Spa	Restaurant
8	Coventry Manor	Construction & Landscaping	Playground	Arts & Crafts Store	Café	Food Truck	Donut Shop	Women's Store	Falafel Restaurant	Flower Shop	Financial or Legal Service

Cluster- 4

```
napa_merged.loc[napa_merged['Cluster Labels'] == 3, napa_merged.columns[[1]+ list(range(5, napa_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
27	Westwood	Clothing Store	Shoe Store	Sporting Goods Shop	Accessories Store	American Restaurant	Coffee Shop	Mexican Restaurant	Burger Joint	Scenic Lookout	Food & Drink Shop

Cluster- 5

```
napa_merged.loc[napa_merged['Cluster Labels'] == 4, napa_merged.columns[[1]+ list(range(5, napa_merged.shape[1]))]]
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
6	Browns Valley West	Vineyard	Hotel	Seafood Restaurant	Wine Bar	Scenic Lookout	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Event Space

10. Conclusion

Based on the Clusters above, the company should start new themed restaurant in the Browns Valley West of cluster- 5 because it has farm and farmers market for the supply of ingredients and other raw materials. It has also Scenic Layout and Vineyard that attracts tourists & other visitors, businesses and is suitable to open a new themed restaurant.

Second option would be opening of new restaurant in Browns Valley South of Cluster-1 as this has also park, lake and farmers market. And last option would be in Stanly Ranch in cluster-2.