

IBM Data Science Capstone Project

The Battle of the Neighborhoods – Report

Introduction

Napa Valley is a wine country located in Northern California, 50 miles Northeast of San Francisco (SFO). It is known for world-class wine, as well as beautiful countryside and near perfect weather year-round. Napa valley is an appellation or American Viticultural Area (AVA) that is used to identify where the grapes for a wine are grown and there are many wine appellations in CA State.

Napa Valley is filled with iconic experiences, events and many things to do and it provides visitors with year-round reasons to visit. It has many wineries with plenty of grape varieties grown every year. Each winery is unique with tasting rooms, restaurants, wine shops, self-guided tours, and many more. Napa Valley consists of five distinctive towns or cities – Calistoga, St. Helena, Yountville, the city of Napa and American Canyon. Each town has its own unique identity that invites exploration.

The city of Napa (or Napa), falls under Napa county and is a heart of Napa Valley and a remarkable destination for its arts & culture. There are many things to do in Napa such as - Breweries & Distilleries, virtual tasting rooms, Wine Trains & Tours, Hot Air Balloons, Wine & Culinary Classes. The city of Napa is a good developed area, diverse, multicultural, business friendly environment and it provides lot of business opportunities. It has attracted many different players into the market and is a major center for wineries and restaurants in the USA.

As Napa is a good developed area, the cost of doing business is expensive and at the same time the market is highly competitive. So, any new business startups or expansion needs to be analyzed carefully. The insights derived from analysis will give good understanding of the business environment such as factors like competition, location, demographics, much more, which helps strategically in targeting the market. This will help in reducing the risk and results in good ROI.

Business Problem Description:

As Napa is diverse and multicultural city, there are many restaurants and many exciting places to eat breakfast, brunch, lunch and dinner prepared by world-class chefs. Also there are many local and international cuisines such as Italian, South Asian; European restaurants served food with many wine varieties made in Napa Valley and surroundings.

In order to open a new restaurant or to expand, it is clear and evident that to survive in such competitive market, strategic planning is crucial. Various factors need to be studied in order to decide on the Location in Napa such as:

- Target Market or Target Demographics
- Competitors and their Menus
- Untapped markets
- Accessibility and visibility
- Proximity to Suppliers - local farmers or wholesale markets
- Venues like Gyms, Entertainment zones, theme Parks
- Health Regulations and Zoning
- Segmentation of the Borough

& many more factors influences in opening new restaurant.

ABC catering Ltd, is a catering company offering their food services to companies' events and they decided to open it's first new theme based restaurant in Napa due to its uniqueness in culture and tourism. Even though ABC is well funded company, they need to choose the right location to start its first venture in Napa and if this is successful they can open their restaurants at other locations with the same theme or another within the Napa Vally. Getting success in the first move is very important, there by choosing the right location is crucial for the business.

Target Audience:

To recommend the right location to the management and board of directors of ABC catering Ltd. And as a Data Scientist my objective would be to locate which neighborhood of Napa city will be best choice to start a new theme based restaurant. The company also expects to understand the rationale of the recommendations made and it would be helpful to anyone who wants to start a new restaurant in Napa city. The success of the project will be based on a good recommendation of borough / neighborhood choice to ABC catering Ltd.

Data used for this assignment:

Napa city is the heart of Napa County. There are 9 cities within Napa County and they are: City of Napa, Calistoga, St. Helena, Yountville, American Canyon, Oakville, Rutherford, Angwin, Deek Park. For this assignment, restaurant location analysis is going to be done for the Napa city. Within the Napa city, there are many neighborhoods and this data is going to be sourced from the links mentioned below.

Napa county and its cities:

https://en.wikipedia.org/wiki/Category:Cities_in_Napa_County,_California

Napa city neighborhoods:

<https://www.cityofnapa.org/547/Neighborhoods>

And the suppliers data for the supply of ingredients can be sourced the local farmers market and can be obtained from the link mentioned below:

Napa local farmers' link:

<https://napafarmersmarket.org/>

The geocodes of the Napa neighborhoods can be obtained from the link below. The Napa neighborhoods data and the geocodes data merged together as a data frame and this entire data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood. And then Foursquare APIs' can be used to explore neighborhoods in the Napa city.

Geocodes of the Napa neighborhoods:

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

About Code

After importing required libraries, imported the raw data that has been downloaded from the link:

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

This data contains borough, city, neighborhood, geo codes information. Here is the image of the raw data below:

```
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

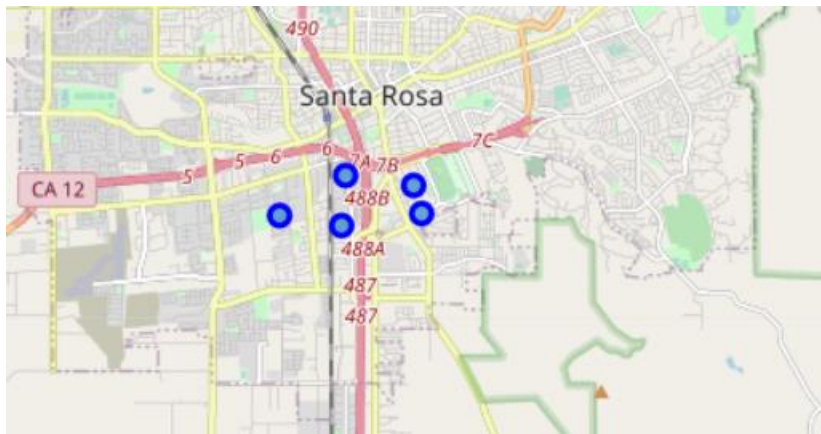
Next data was transformed in the required format and shown in the image below:

```
df.head()
```

Out[13]:

	Borough	Neighborhood	Latitude	Longitude
0	Santa Rosa	Aston Ave	38.424439	-122.705772
1	Santa Rosa	Olive Corby	38.429636	-122.718490
2	Santa Rosa	South Park	38.428237	-122.707064
3	Santa Rosa	Corby/Beachwood	38.422833	-122.719420
4	Santa Rosa	Sunset McMinn	38.424407	-122.729984

Next we used geopy library to get the latitude and longitude values of Santa Rosa city in Sonoma county. And created a Santa Rosa map with neighborhoods superimposed on top.



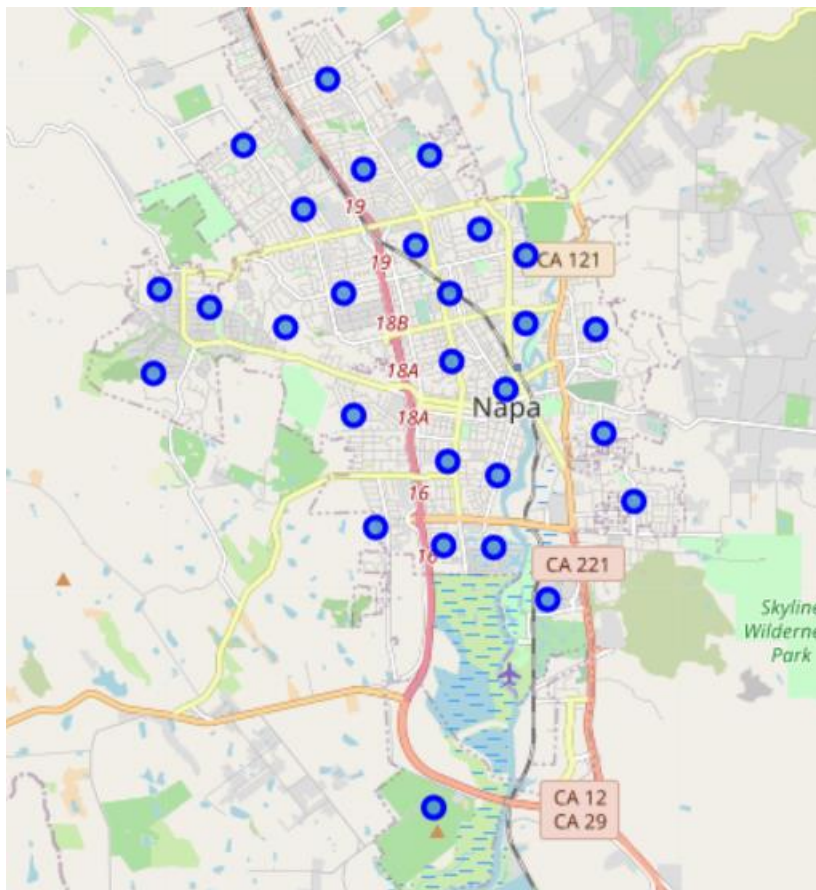
The above map is only for illustration purposes, It is simplified and considered the segmentation and clustering only the neighborhoods in Napa city. So let's slice the original data frame and create a new data frame of the Napa data.

```
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

The Napa map is created as shown below:



Next by using the Foursquare credentials, Napa neighborhoods are explored. It is shown in the image below. And Analyzed each venue category.

Exploring Neighborhoods in Napa

```
In [35]: napa_venues = getNearbyVenues(names=napa_data['Neighborhood'],
latitudes=napa_data['Latitude'],
longitudes=napa_data['Longitude']
)
```

Alta Heights
Beard
Bel Aire
Browns Valley Central
Browns Valley East
Browns Valley South
Browns Valley West
Central
Coventry Manor
Downtown
Fuller Park
Linda Vista
Mcpherson
Napa Abajo
Pueblo Park
Riverpark
Sheveland Ranch
Shurtleff
Soscol Gateway North
Soscol Gateway South
Springwood Estates
Stanly Ranch
Stonehouse
Terrace
Vineyard Estates
Vintage
Von Uhlit Ranch
Westwood

Analyzing Each Category as shown below:

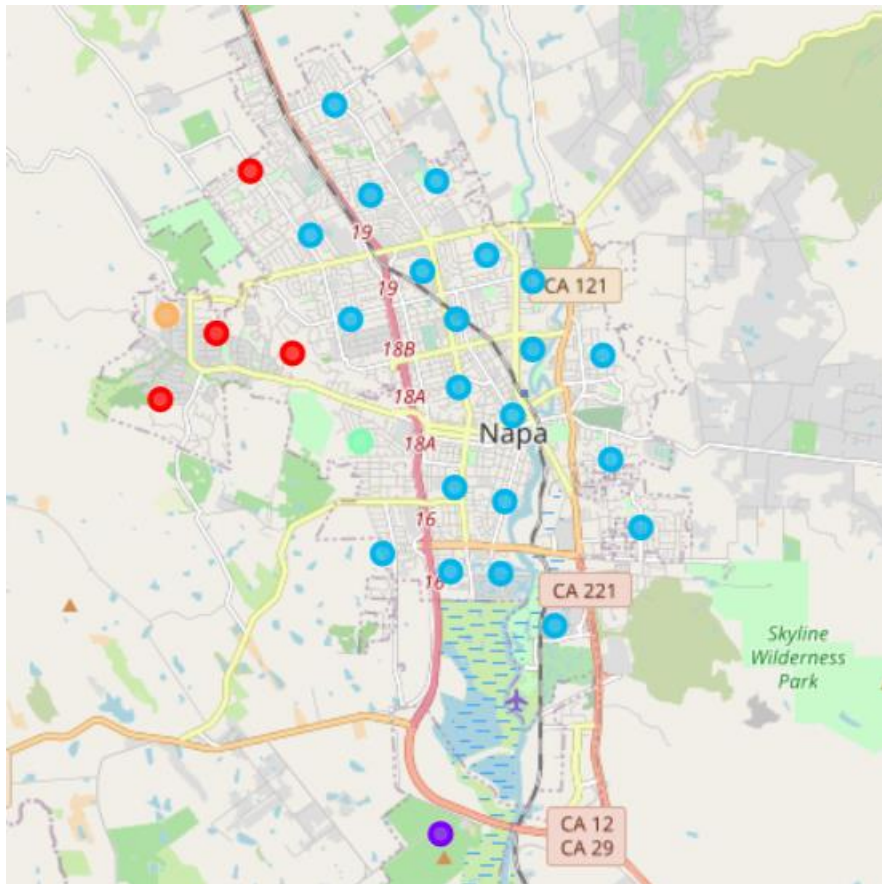
[illegible]

Later each neighborhood along with the top 5 most common venues are collected and put it in the data frame.

	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
4	Browns Valley East	Park	Construction & Landscaping	Grocery Store	Coffee Shop	Dessert Shop	Garden	Gastropub	Women's Store	Donut Shop	Fast Food Restaurant

Clustering Neighborhoods

Clustering neighborhoods are done and created visualization as shown in the picture below:



Examining Clusters: examining each cluster and determine the discriminating venue categories that distinguish each cluster.

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	Discount Store
4	Browns Valley East	Park	Construction & Landscaping	Grocery Store	Coffee Shop	Dessert Shop	Garden	Gastropub	Women's Store	Donut Shop	Fast Food Restaurant
5	Browns Valley South	Park	Lake	American Restaurant	Massage Studio	Scenic Lookout	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant
24	Vineyard Estates	Food Truck	Park	Bus Station	Baseball Stadium	Grocery Store	Mexican Restaurant	Hotel Pool	Trail	Italian Restaurant	Tour Provider

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 Shop

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

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

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, 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