Battle of the Neighborhoods in Saint Louis

Kalpana Joshi

August 26th 2020

1. Introduction

1.a Background

Saint Louis is a quiet city in Missouri, right in the heart of the United States of America. Its widely known for its very popular baseball team, the Cardinals, which attracts a lot of tourists during the game season. The city has several natural places to visit during summer time like lakes, adventure trails and parks. There are many facilities provided to the public free of cost like the Saint Louis Zoo, Art Museum, History Museum and the Science Centre. Along with these free avenues there are several malls, restaurants, eateries which provide the residents a plethora of activities to keep themselves busy during weekends or holidays. In this project we will be exploring the various Neighborhoods in the city of Saint Louis and the popular venues in each of these.

1.b Problem

The project aims to explore the venues in the different parts of Saint Louis and determine the most popular ones in a certain area. This analysis is aimed at finding out the best suited locations to start up a new business like a restaurant or a clothing store.

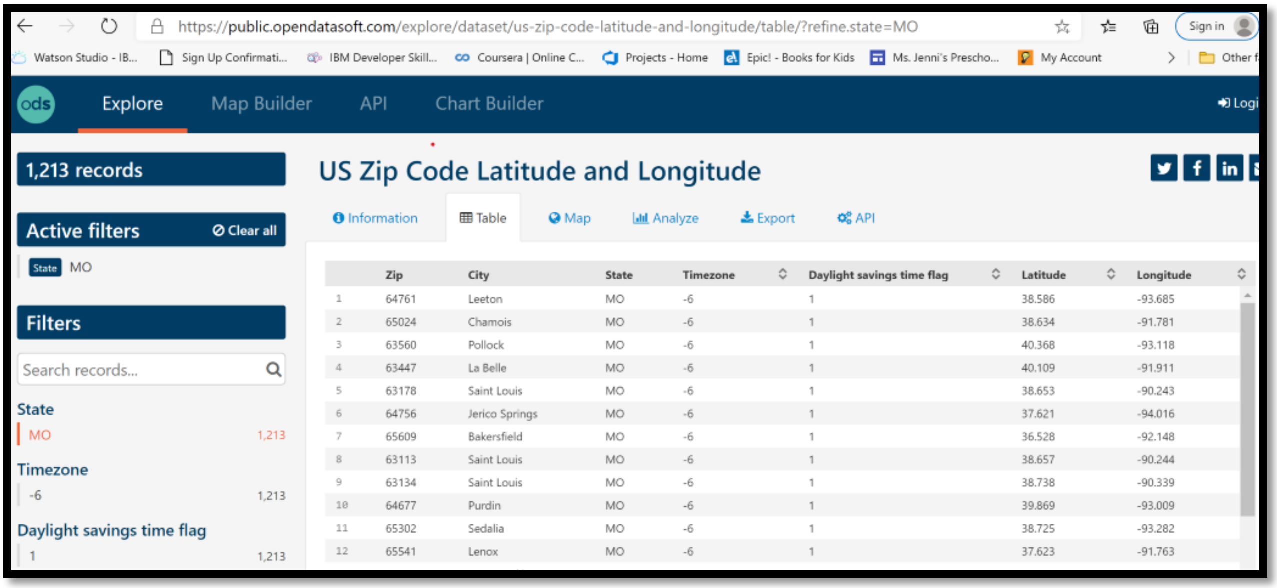
1.c Interested Parties

The result of this project will benefit many of the entrepreneurs in Saint Louis to determine the location that is best for them to set up shop. Since we will be assessing the popular venues in various locations, we will get an idea of the kind of customers and their purpose while visiting these locations. This will increase the chances of them visiting the stores of these entrepreneurs if they set up their business accordingly.

2. Data Acquisition

2.a Data Sources

The data for this project has been accessed from the website: https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/ , which has a detailed list of all the zip codes of the US cities along with other information such as City, State, Longitude, Latitude, Time Zone, Daylight savings time flag and Geopoint. This data can be filtered by State. Below is a snapshot of the data filtered for the State of Missouri where Saint Louis is located:



1,213 records

US Zip Code Latitude and Longitude

Active filters

State MO

Filters

State

MO 1,213

Timezone

-6 1,213

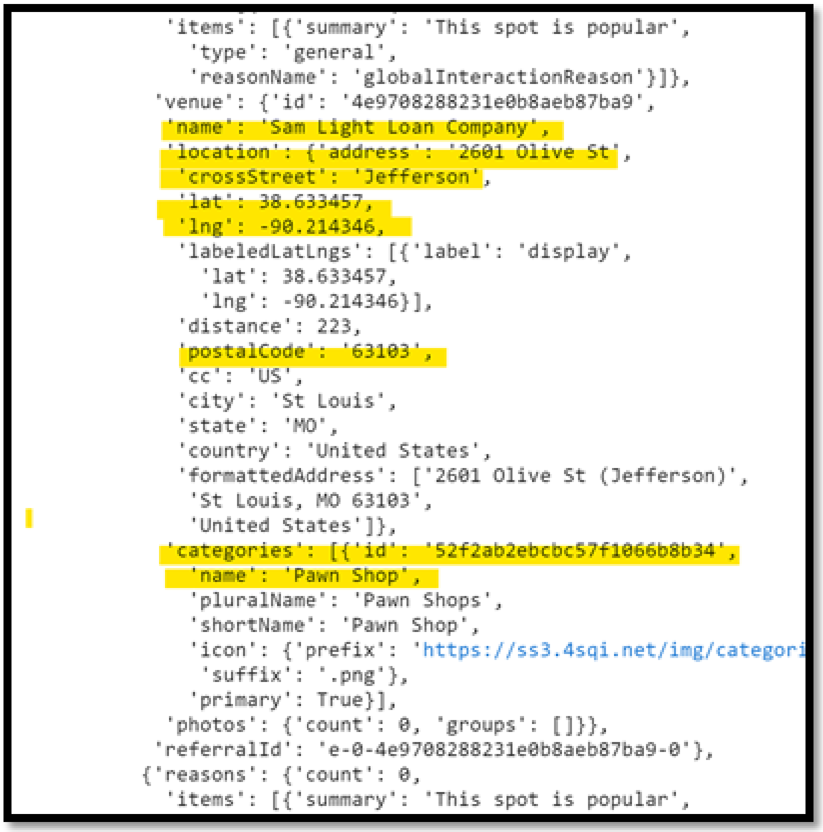
Daylight savings time flag

1 1,213

	Zip	City	State	Timezone	Daylight savings time flag	Latitude	Longitude
1	64761	Leeton	MO	-6	1	38.586	-93.685
2	65024	Chamois	MO	-6	1	38.634	-91.781
3	63560	Pollock	MO	-6	1	40.368	-93.118
4	63447	La Belle	MO	-6	1	40.109	-91.911
5	63178	Saint Louis	MO	-6	1	38.653	-90.243
6	64756	Jerico Springs	MO	-6	1	37.621	-94.016
7	65609	Bakersfield	MO	-6	1	36.528	-92.148
8	63113	Saint Louis	MO	-6	1	38.657	-90.244
9	63134	Saint Louis	MO	-6	1	38.738	-90.339
10	64677	Purdin	MO	-6	1	39.869	-93.009
11	65302	Sedalia	MO	-6	1	38.725	-93.282
12	65541	Lenox	MO	-6	1	37.623	-91.763

Above data can be exported in the csv format and then read into a pandas Dataframe.

We will also be utilizing the Foursquare API to determine the venues at each of these locations. Below is a screenshot of the information received from Foursquare for a specific location. The highlighted portions are the ones we will be focusing on during our analysis.



```
     'items': [{'summary': 'This spot is popular',
       'type': 'general',
       'reasonName': 'globalInteractionReason'}]},
    'venue': {'id': '4e9708288231e0b8aeb87ba9',
     'name': 'Sam Light Loan Company',
     'location': {'address': '2601 Olive St',
      'crossStreet': 'Jefferson',
      'lat': 38.633457,
      'lng': -90.214346,
      'labeledLatLngs': [{'label': 'display',
        'lat': 38.633457,
        'lng': -90.214346}],
      'distance': 223,
      'postalCode': '63103',
      'cc': 'US',
      'city': 'St Louis',
      'state': 'MO',
      'country': 'United States',
      'formattedAddress': ['2601 Olive St (Jefferson)',
       'St Louis, MO 63103',
       'United States']},
     'categories': [{'id': '52f2ab2ebcbc57f1066b8b34',
       'name': 'Pawn Shop',
       'pluralName': 'Pawn Shops',
       'shortName': 'Pawn Shop',
       'icon': {'prefix': 'https://ss3.4sqi.net/img/categori
        'suffix': '.png'},
       'primary': True}],
     'photos': {'count': 0, 'groups': []}},
    'referralId': 'e-0-4e9708288231e0b8aeb87ba9-0'},
   {'reasons': {'count': 0,
     'items': [{'summary': 'This spot is popular',
```

2.b Data Cleaning

The data extracted from the website below is in the csv format.

https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/

Fist we need to create a Dataframe with the data in the above file. This data is at the state level, so we need to filter it further by city (Saint Louis).

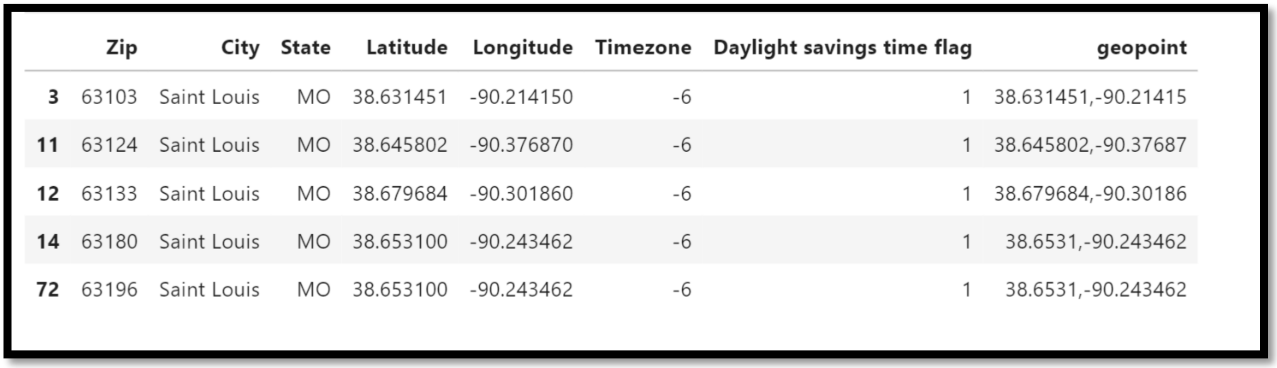
2.c Feature Selection:

We can further drop fields which will not be used during our analysis like ‘Time zone’, ‘Daylight Saving time flag’ and ‘Geopoint’. The final Dataframe had about 71 samples or information about 71 Postal Codes in the city of Saint Louis with their respective Latitudes and Longitudes.

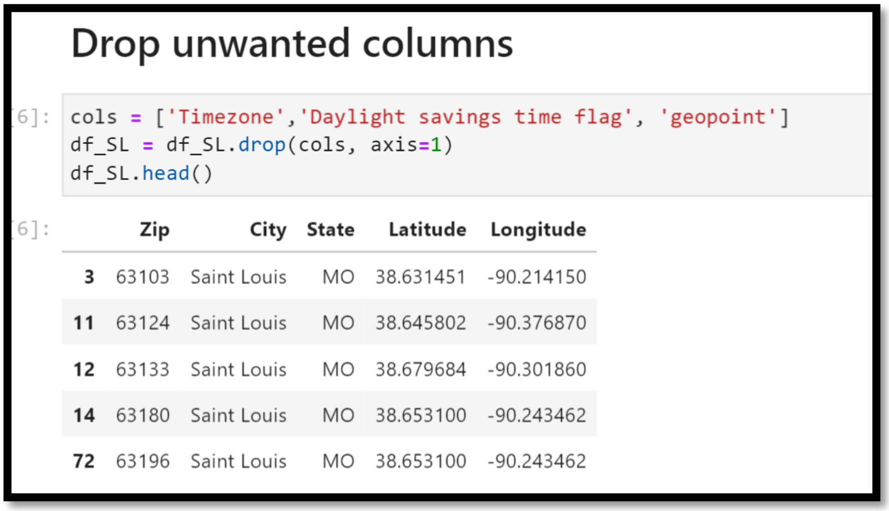
3. Methodology

3.a Exploratory data analysis:

The dataframe created from the csv file needs to be cleaned in order to move forward with our analysis. So, we first need to filter it by city of Saint Louis. For feature selection, certain columns or fields were dropped such as Time zone, Daylight saving time flag and geopoint. Both these steps are depicted in the screenshots below.



	Zip	City	State	Latitude	Longitude	Timezone	Daylight savings time flag	geopoint
3	63103	Saint Louis	MO	38.631451	-90.214150	-6	1	38.631451,-90.21415
11	63124	Saint Louis	MO	38.645802	-90.376870	-6	1	38.645802,-90.37687
12	63133	Saint Louis	MO	38.679684	-90.301860	-6	1	38.679684,-90.30186
14	63180	Saint Louis	MO	38.653100	-90.243462	-6	1	38.6531,-90.243462
72	63196	Saint Louis	MO	38.653100	-90.243462	-6	1	38.6531,-90.243462



Drop unwanted columns

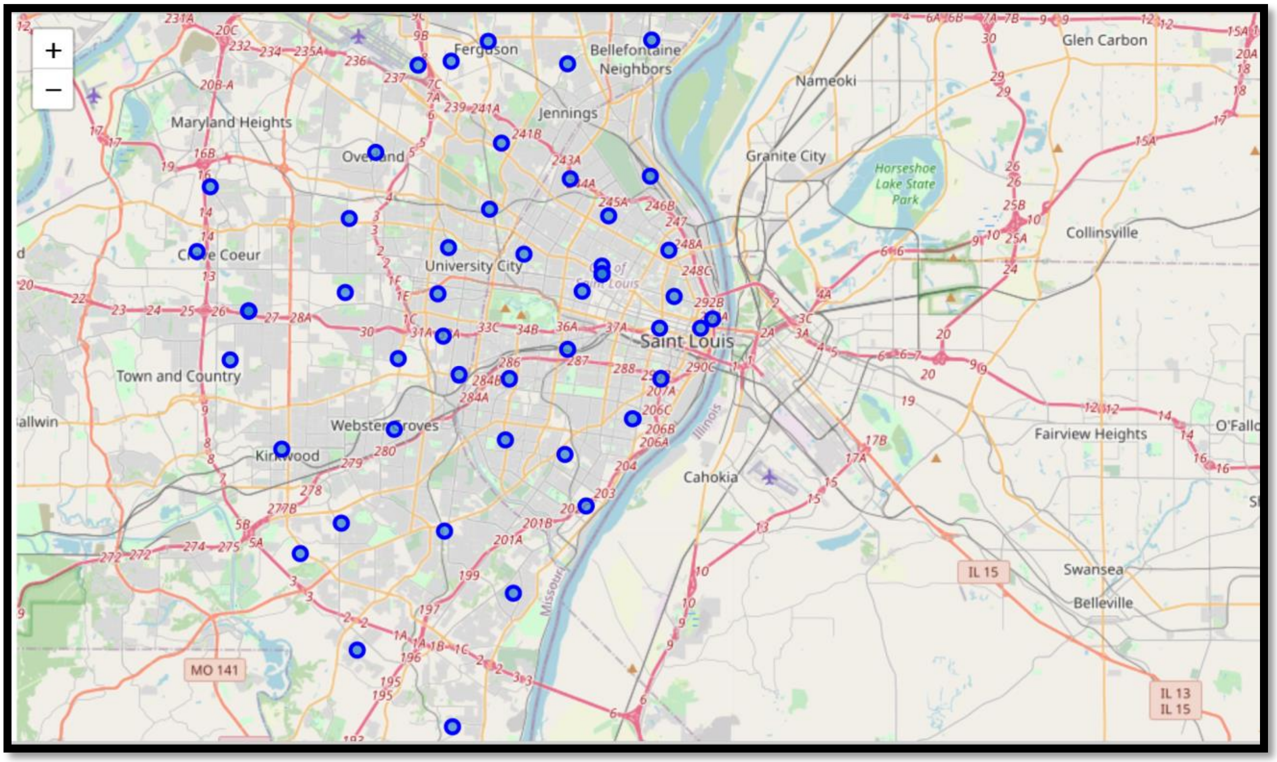
```
6]: cols = ['Timezone','Daylight savings time flag', 'geopoint']
    df_SL = df_SL.drop(cols, axis=1)
    df_SL.head()
```

6]:

	Zip	City	State	Latitude	Longitude
3	63103	Saint Louis	MO	38.631451	-90.214150
11	63124	Saint Louis	MO	38.645802	-90.376870
12	63133	Saint Louis	MO	38.679684	-90.301860
14	63180	Saint Louis	MO	38.653100	-90.243462
72	63196	Saint Louis	MO	38.653100	-90.243462

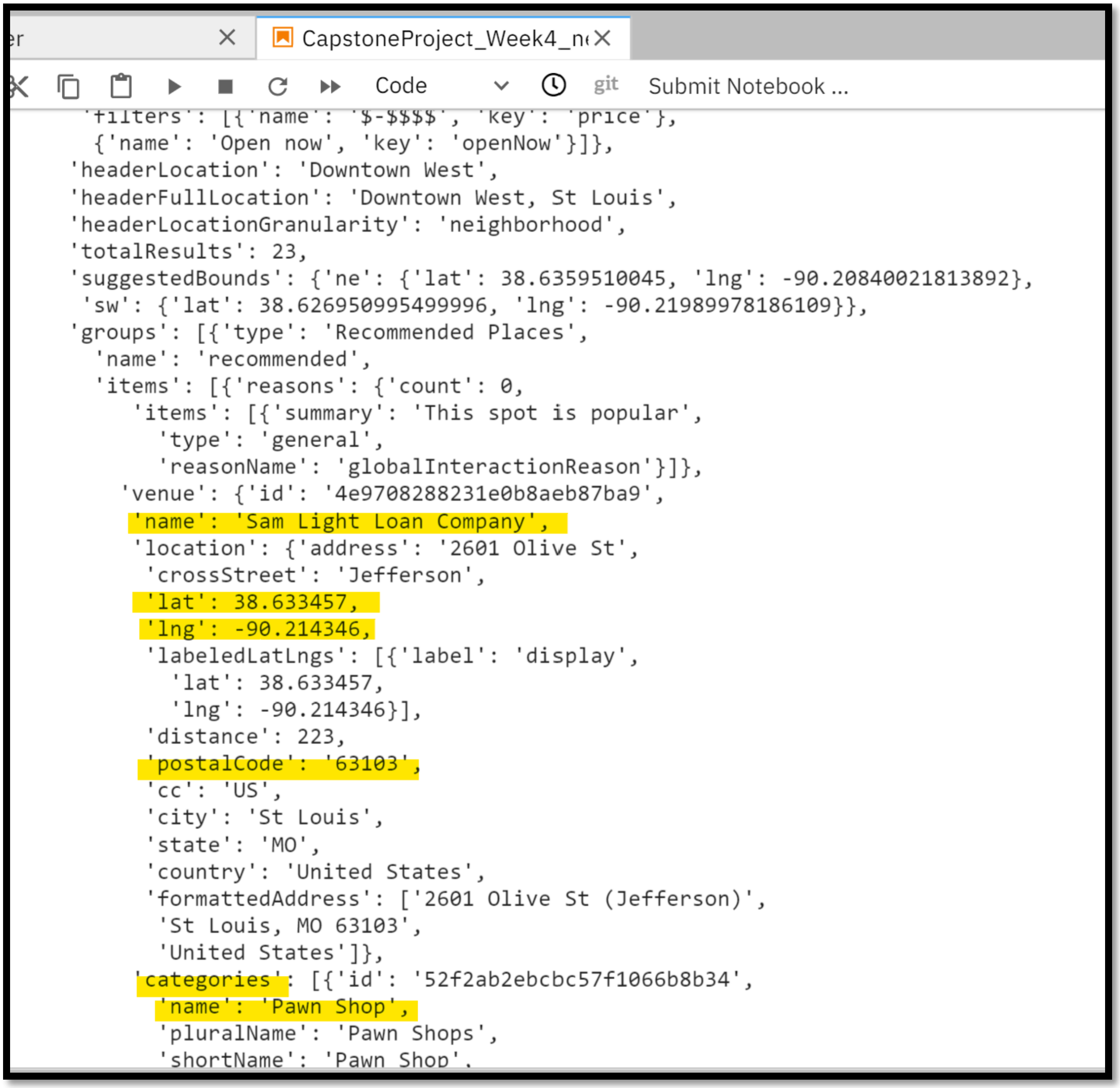
<u>Visualizing the City of Saint Louis on the Map with all the zip codes highlighted:</u>

We can visualize the city of Saint Louis by first finding the coordinates of the city (Latitude: 38.6268039, Longitude: -90.1994097) and then plotting all the Zip codes that are present in our database. Later we will be diving these Zip codes into clusters based on the types of venues.



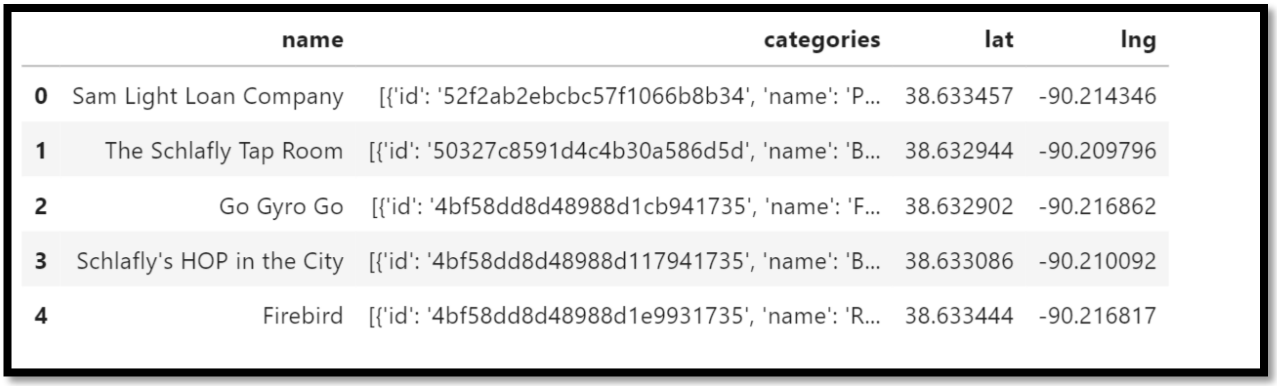
Make API calls to get all the possible venues in the surroundings of the first Zip:

Now, we can begin our exploration by analyzing the very first Zip code (63178 with coordinates: 38.6531, -90.243462) by making an API call to Foursquare. The result obtained in the screenshot shown below. The details which we need to focus on are Venue Name, Venue Category, Postal Code, Latitude and Longitude. We can see in the example below that the first venue in the list is ‘Sam Light Loan Company’ which is a ‘Pawn Shop’.



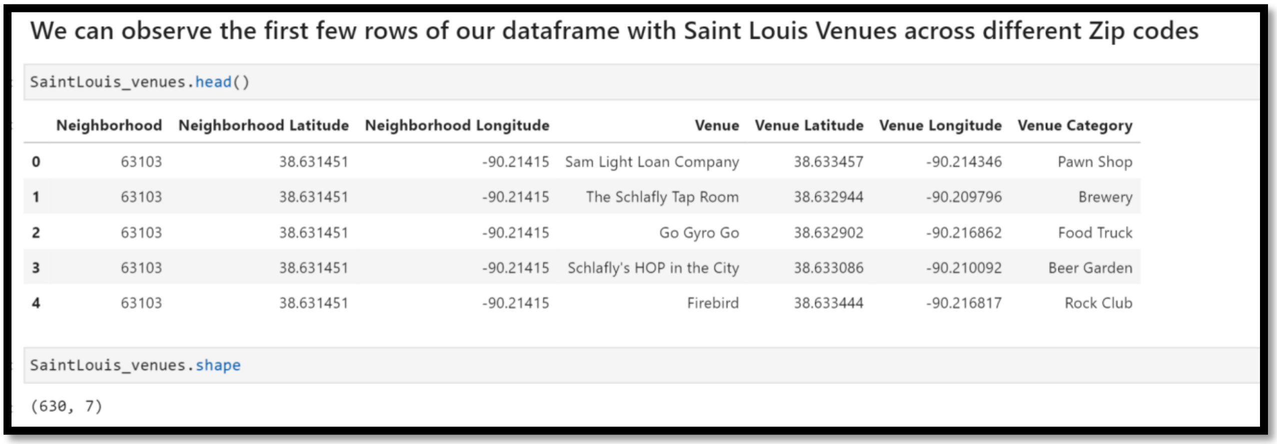
```
'filters': [{'name': '$-$$$$', 'key': 'price'},
  {'name': 'Open now', 'key': 'openNow'}]},
'headerLocation': 'Downtown West',
'headerFullLocation': 'Downtown West, St Louis',
'headerLocationGranularity': 'neighborhood',
'totalResults': 23,
'suggestedBounds': {'ne': {'lat': 38.6359510045, 'lng': -90.20840021813892},
 'sw': {'lat': 38.626950995499996, 'lng': -90.21989978186109}},
'groups': [{'type': 'Recommended Places',
  'name': 'recommended',
  'items': [{'reasons': {'count': 0,
     'items': [{'summary': 'This spot is popular',
       'type': 'general',
       'reasonName': 'globalInteractionReason'}]},
    'venue': {'id': '4e9708288231e0b8aeb87ba9',
     'name': 'Sam Light Loan Company',
     'location': {'address': '2601 Olive St',
      'crossStreet': 'Jefferson',
      'lat': 38.633457,
      'lng': -90.214346,
      'labeledLatLngs': [{'label': 'display',
        'lat': 38.633457,
        'lng': -90.214346}],
      'distance': 223,
      'postalCode': '63103',
      'cc': 'US',
      'city': 'St Louis',
      'state': 'MO',
      'country': 'United States',
      'formattedAddress': ['2601 Olive St (Jefferson)',
       'St Louis, MO 63103',
       'United States']},
     'categories': [{'id': '52f2ab2ebcbc57f1066b8b34',
       'name': 'Pawn Shop',
       'pluralName': 'Pawn Shops',
       'shortName': 'Pawn Shop',
```

We will now find out the details about all the other venues around the first Zip Code and then arrange them in a dataframe like below.



	name	categories	lat	lng
0	Sam Light Loan Company	[{'id': '52f2ab2ebcbc57f1066b8b34', 'name': 'P...	38.633457	-90.214346
1	The Schlafly Tap Room	[{'id': '50327c8591d4c4b30a586d5d', 'name': 'B...	38.632944	-90.209796
2	Go Gyro Go	[{'id': '4bf58dd8d48988d1cb941735', 'name': 'F...	38.632902	-90.216862
3	Schlafly's HOP in the City	[{'id': '4bf58dd8d48988d117941735', 'name': 'B...	38.633086	-90.210092
4	Firebird	[{'id': '4bf58dd8d48988d1e9931735', 'name': 'R...	38.633444	-90.216817

Here we can find the venue names and categories very clearly which will help us in determining the types of businesses in each area. Therefore, we will now create a dataframe with all the venue names, categories and coordinates for all the zip codes in the city of Saint Louis:



We can observe the first few rows of our dataframe with Saint Louis Venues across different Zip codes

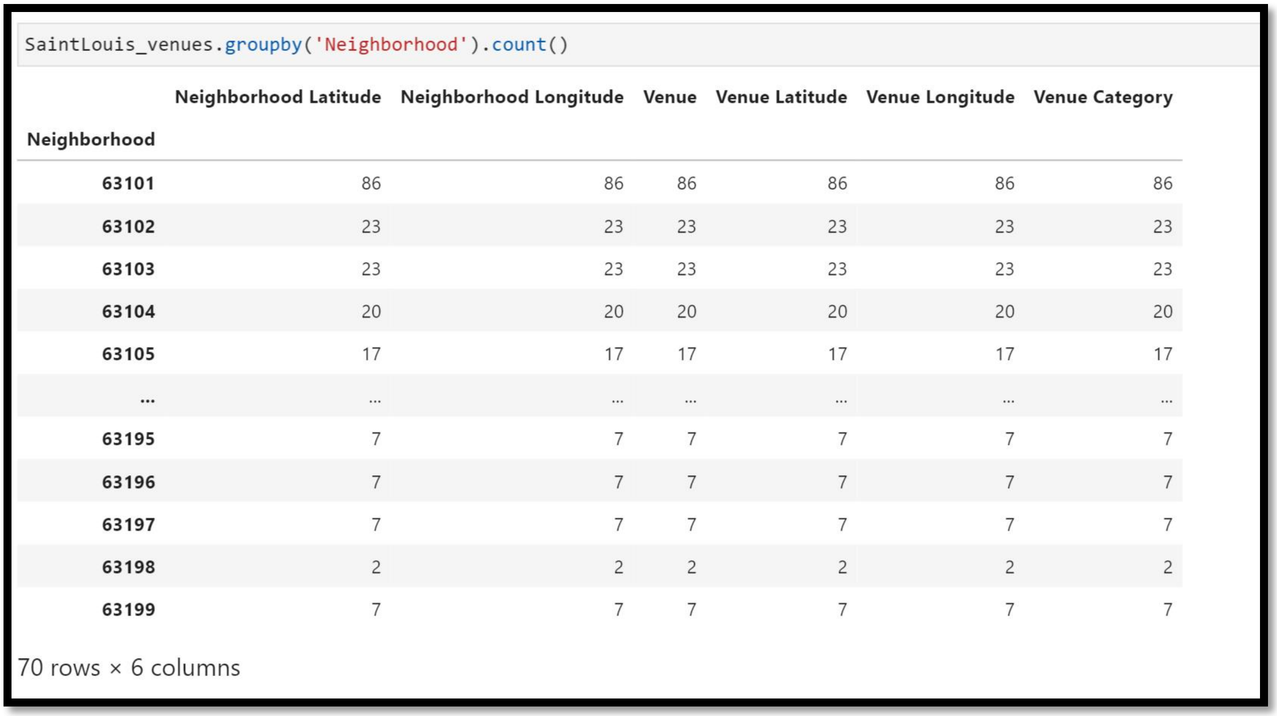
```
SaintLouis_venues.head()
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	63103	38.631451	-90.21415	Sam Light Loan Company	38.633457	-90.214346	Pawn Shop
1	63103	38.631451	-90.21415	The Schlafly Tap Room	38.632944	-90.209796	Brewery
2	63103	38.631451	-90.21415	Go Gyro Go	38.632902	-90.216862	Food Truck
3	63103	38.631451	-90.21415	Schlafly's HOP in the City	38.633086	-90.210092	Beer Garden
4	63103	38.631451	-90.21415	Firebird	38.633444	-90.216817	Rock Club

```
SaintLouis_venues.shape
```

```
(630, 7)
```

We can see that there is a total of 630 venues across different Zip codes out of which 169 are unique values. Following table shows the number of venues across each zip code:

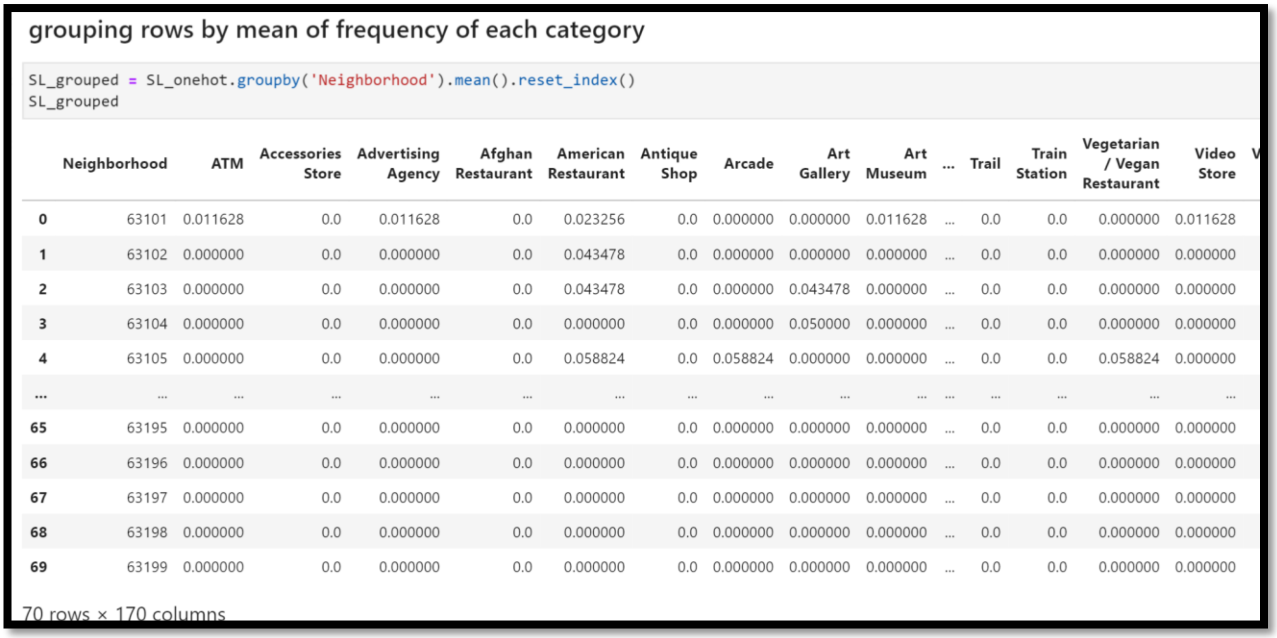


```
SaintLouis_venues.groupby('Neighborhood').count()
```

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
63101	86	86	86	86	86	86
63102	23	23	23	23	23	23
63103	23	23	23	23	23	23
63104	20	20	20	20	20	20
63105	17	17	17	17	17	17
...	...	...	...	...	...	...
63195	7	7	7	7	7	7
63196	7	7	7	7	7	7
63197	7	7	7	7	7	7
63198	2	2	2	2	2	2
63199	7	7	7	7	7	7

70 rows × 6 columns

For further analysis, we normalize our data by performing 'One-hot coding' (by creating new columns for all 164 venues across all the Zip codes and assigning dummy values according to their presence in the location)



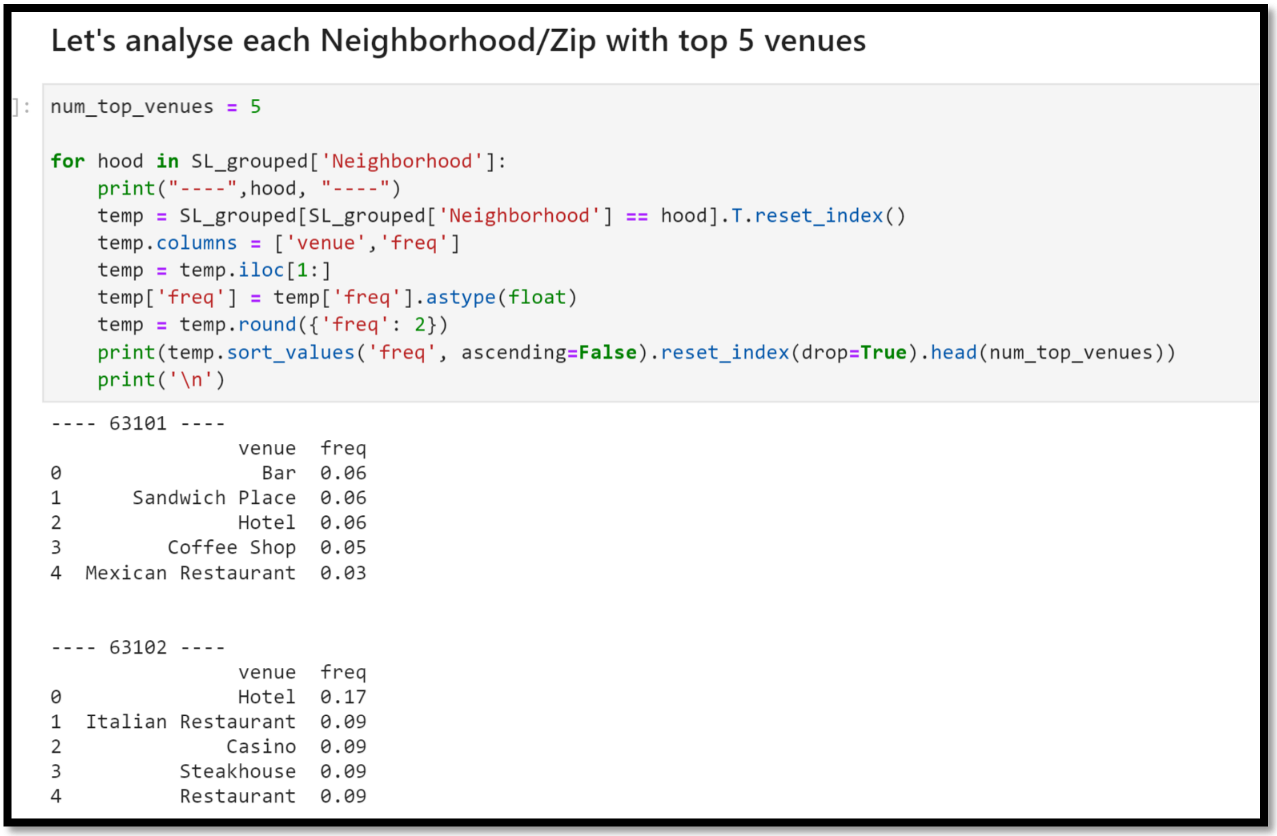
grouping rows by mean of frequency of each category

```
SL_grouped = SL_onehot.groupby('Neighborhood').mean().reset_index()
SL_grouped
```

	Neighborhood	ATM	Accessories Store	Advertising Agency	Afghan Restaurant	American Restaurant	Antique Shop	Arcade	Art Gallery	Art Museum	...	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Store	V
0	63101	0.011628	0.0	0.011628	0.0	0.023256	0.0	0.000000	0.000000	0.011628	...	0.0	0.0	0.000000	0.011628	
1	63102	0.000000	0.0	0.000000	0.0	0.043478	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000000	0.000000	
2	63103	0.000000	0.0	0.000000	0.0	0.043478	0.0	0.000000	0.043478	0.000000	...	0.0	0.0	0.000000	0.000000	
3	63104	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.000000	0.050000	0.000000	...	0.0	0.0	0.000000	0.000000	
4	63105	0.000000	0.0	0.000000	0.0	0.058824	0.0	0.058824	0.000000	0.000000	...	0.0	0.0	0.058824	0.000000	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
65	63195	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000000	0.000000	
66	63196	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000000	0.000000	
67	63197	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000000	0.000000	
68	63198	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000000	0.000000	
69	63199	0.000000	0.0	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000000	0.000000	

70 rows × 170 columns

Now we can analyze the top 5 venues for each of the locations:

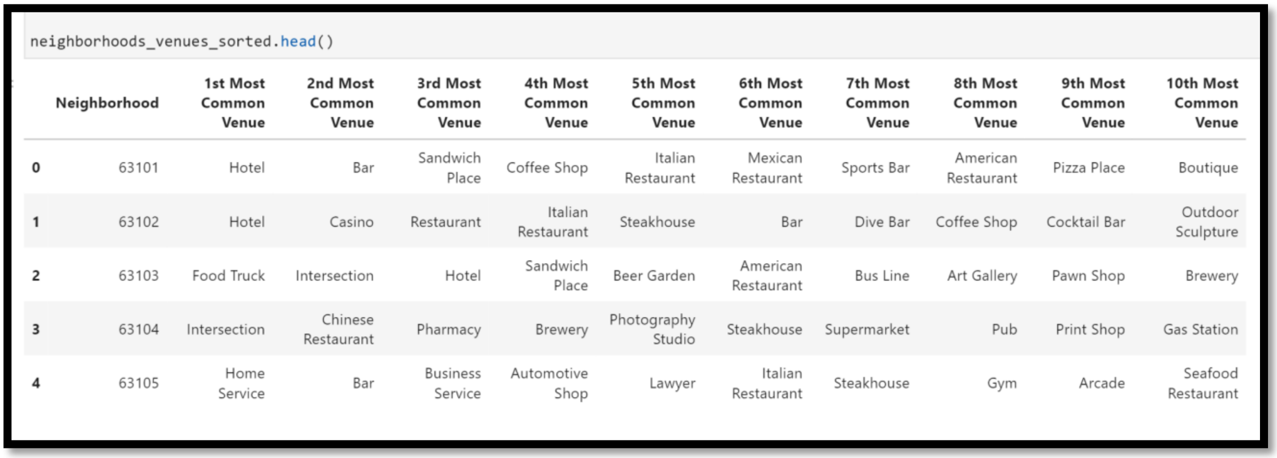


Let's analyse each Neighborhood/Zip with top 5 venues

```
num_top_venues = 5

for hood in SL_grouped['Neighborhood']:
    print("----",hood, "----")
    temp = SL_grouped[SL_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')
```

```
---- 63101 ----
                venue  freq
0                 Bar  0.06
1      Sandwich Place  0.06
2               Hotel  0.06
3         Coffee Shop  0.05
4  Mexican Restaurant  0.03


---- 63102 ----
                venue  freq
0               Hotel  0.17
1  Italian Restaurant  0.09
2              Casino  0.09
3          Steakhouse  0.09
4          Restaurant  0.09
```

Following is the dataframe with the Top 10 venues across each Zip code:



```
neighborhoods_venues_sorted.head()
```

	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	63101	Hotel	Bar	Sandwich Place	Coffee Shop	Italian Restaurant	Mexican Restaurant	Sports Bar	American Restaurant	Pizza Place	Boutique
1	63102	Hotel	Casino	Restaurant	Italian Restaurant	Steakhouse	Bar	Dive Bar	Coffee Shop	Cocktail Bar	Outdoor Sculpture
2	63103	Food Truck	Intersection	Hotel	Sandwich Place	Beer Garden	American Restaurant	Bus Line	Art Gallery	Pawn Shop	Brewery
3	63104	Intersection	Chinese Restaurant	Pharmacy	Brewery	Photography Studio	Steakhouse	Supermarket	Pub	Print Shop	Gas Station
4	63105	Home Service	Bar	Business Service	Automotive Shop	Lawyer	Italian Restaurant	Steakhouse	Gym	Arcade	Seafood Restaurant

3.b Machine Learning usage:

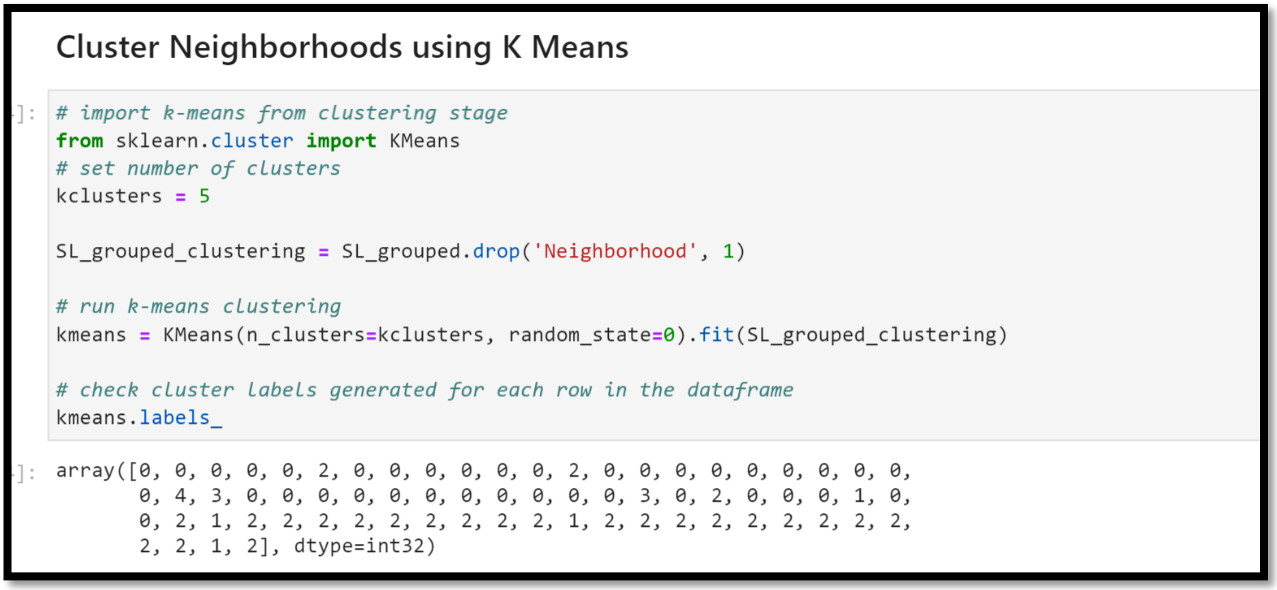
Now, we can use the KMeans Clustering methodology to segregate the zip codes into clusters based on their venue categories as follows. We initialize our clusters to 5 for this analysis:

Reason for using KMean Clustering:

KMeans Clustering is a machine learning methodology that can be used for both supervised and unsupervised learning. In our example, we do not have the clusters defined to begin with. Therefore, we need to train our model using the unsupervised method for which KMeans is the most appropriate. We begin with a set number of clusters (i.e. 5) and start creating our model based on the available data around the venue categories across each zip code. The model looks for the zip codes with similar patterns in the top 10 venue categories and groups them together.

4. Results

We use the KMeans clustering method define above to find out the labels for each of the Zip codes in the final dataframe with Top 10 venues

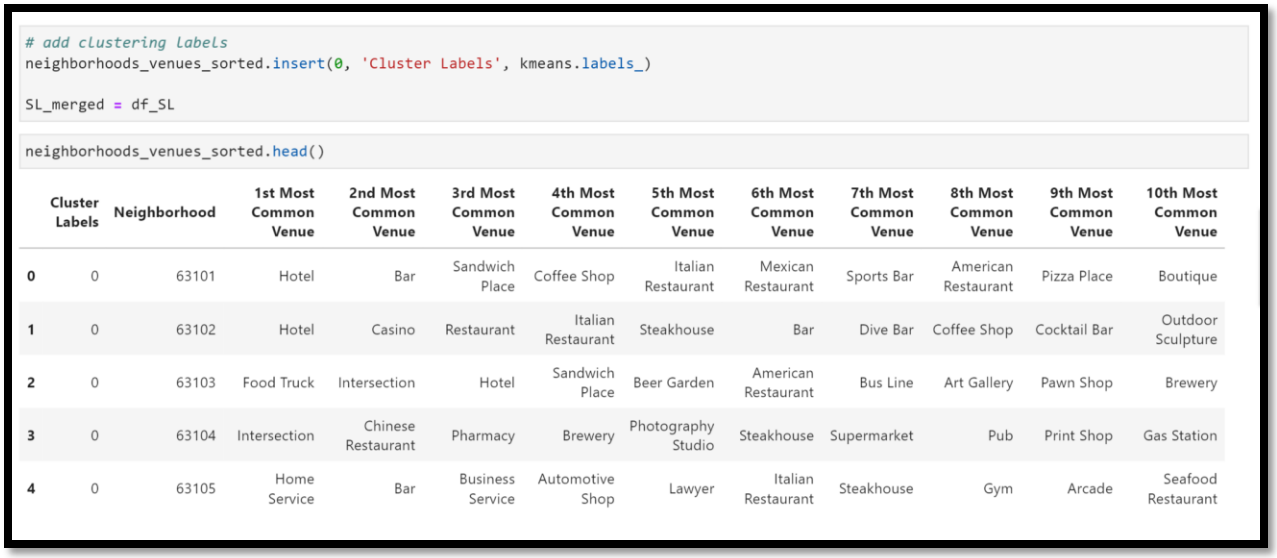


Cluster Neighborhoods using K Means

```
# import k-means from clustering stage
from sklearn.cluster import KMeans
# set number of clusters
kclusters = 5

SL_grouped_clustering = SL_grouped.drop('Neighborhood', 1)

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

# check cluster labels generated for each row in the dataframe
kmeans.labels_
```

```
array([0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 4, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0, 2, 0, 0, 0, 1, 0,
       0, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 1, 2], dtype=int32)
```

Assign these labels to the Neighborhoods with Top 10 venues:

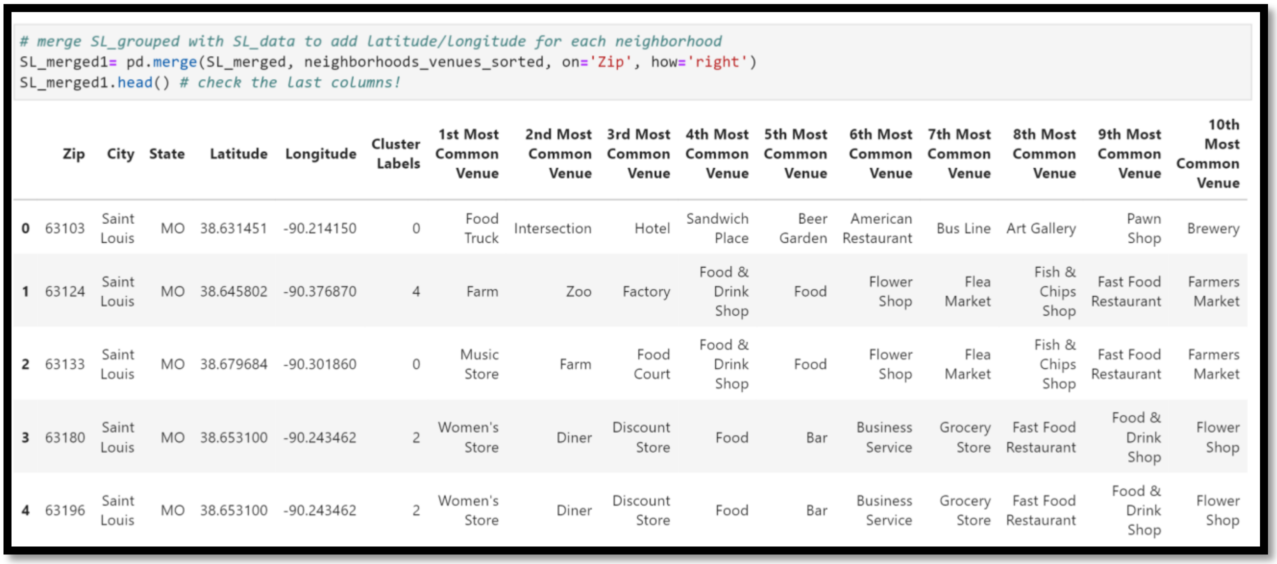


```
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

SL_merged = df_SL
```

```
neighborhoods_venues_sorted.head()
```

	Cluster Labels	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	0	63101	Hotel	Bar	Sandwich Place	Coffee Shop	Italian Restaurant	Mexican Restaurant	Sports Bar	American Restaurant	Pizza Place	Boutique
1	0	63102	Hotel	Casino	Restaurant	Italian Restaurant	Steakhouse	Bar	Dive Bar	Coffee Shop	Cocktail Bar	Outdoor Sculpture
2	0	63103	Food Truck	Intersection	Hotel	Sandwich Place	Beer Garden	American Restaurant	Bus Line	Art Gallery	Pawn Shop	Brewery
3	0	63104	Intersection	Chinese Restaurant	Pharmacy	Brewery	Photography Studio	Steakhouse	Supermarket	Pub	Print Shop	Gas Station
4	0	63105	Home Service	Bar	Business Service	Automotive Shop	Lawyer	Italian Restaurant	Steakhouse	Gym	Arcade	Seafood Restaurant

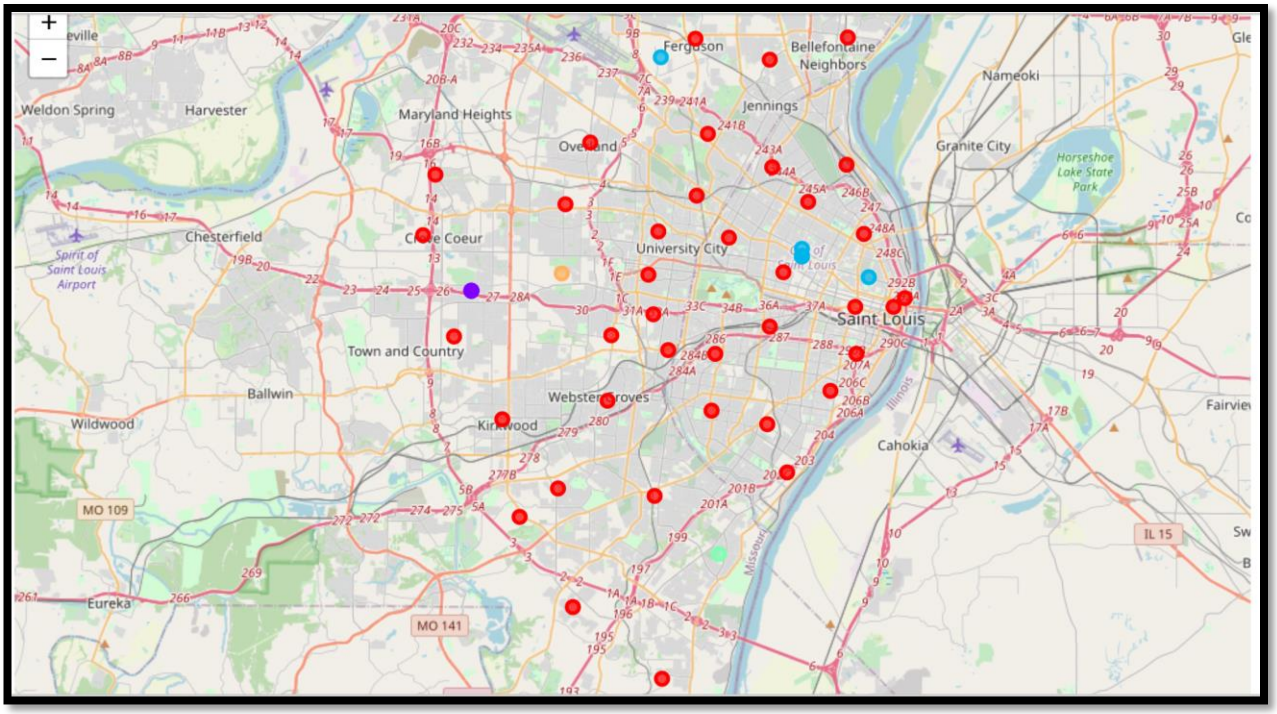
Now, merge this dataframe with the original dataframe of Saint Louis city to get a detailed picture with the Zip Code, City name, State, Longitude, Latitude as follows:



```
# merge SL_grouped with SL_data to add latitude/longitude for each neighborhood
SL_merged1= pd.merge(SL_merged, neighborhoods_venues_sorted, on='Zip', how='right')
SL_merged1.head() # check the last columns!
```

	Zip	City	State	Latitude	Longitude	Cluster Labels	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	63103	Saint Louis	MO	38.631451	-90.214150	0	Food Truck	Intersection	Hotel	Sandwich Place	Beer Garden	American Restaurant	Bus Line	Art Gallery	Pawn Shop	Brewery
1	63124	Saint Louis	MO	38.645802	-90.376870	4	Farm	Zoo	Factory	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish & Chips Shop	Fast Food Restaurant	Farmers Market
2	63133	Saint Louis	MO	38.679684	-90.301860	0	Music Store	Farm	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish & Chips Shop	Fast Food Restaurant	Farmers Market
3	63180	Saint Louis	MO	38.653100	-90.243462	2	Women's Store	Diner	Discount Store	Food	Bar	Business Service	Grocery Store	Fast Food Restaurant	Food & Drink Shop	Flower Shop
4	63196	Saint Louis	MO	38.653100	-90.243462	2	Women's Store	Diner	Discount Store	Food	Bar	Business Service	Grocery Store	Fast Food Restaurant	Food & Drink Shop	Flower Shop

We can now visualize these clusters on the map of Saint Louis as follows:

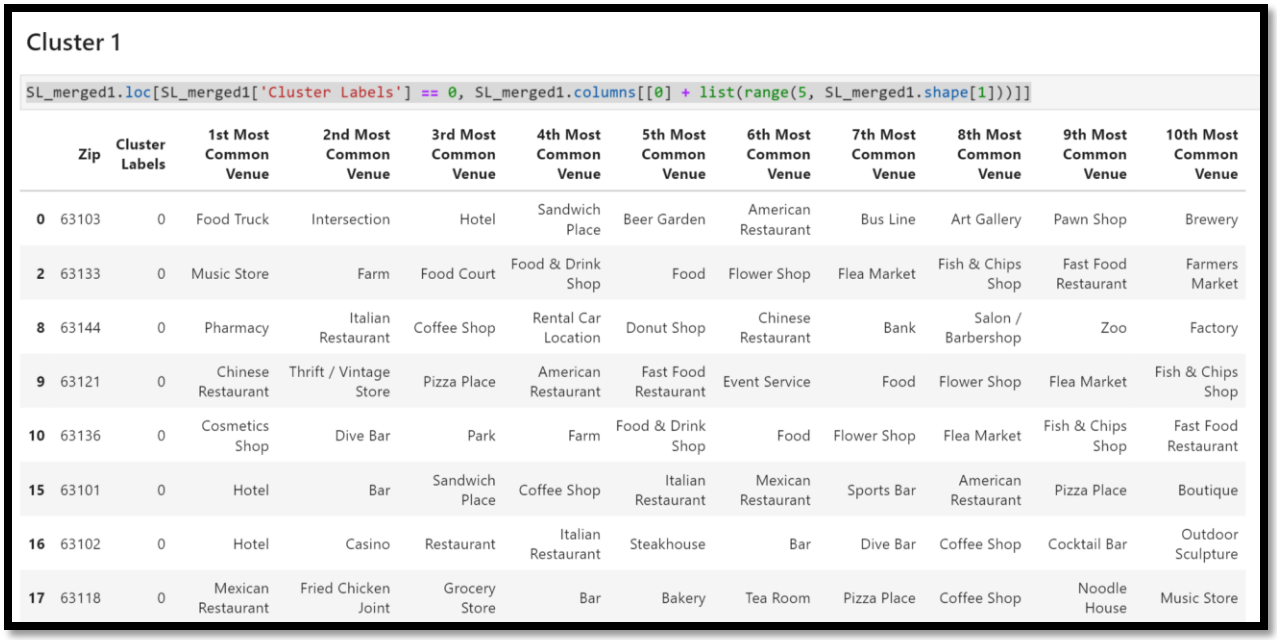


5. Discussion

a. Observations

Let's further look at each of the clusters:

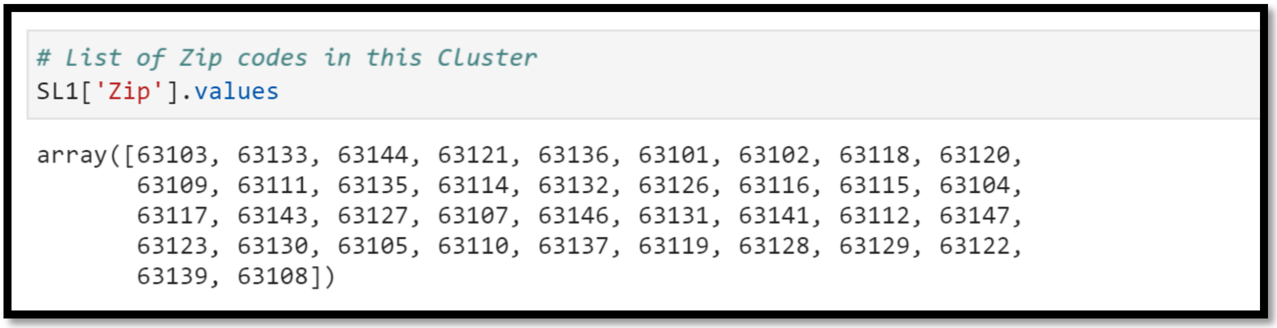
<u>1st Cluster:</u>



Cluster 1

```
SL_merged1.loc[SL_merged1['Cluster Labels'] == 0, SL_merged1.columns[[0] + list(range(5, SL_merged1.shape[1]))]]
```

	Zip	Cluster Labels	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	63103	0	Food Truck	Intersection	Hotel	Sandwich Place	Beer Garden	American Restaurant	Bus Line	Art Gallery	Pawn Shop	Brewery
2	63133	0	Music Store	Farm	Food Court	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish & Chips Shop	Fast Food Restaurant	Farmers Market
8	63144	0	Pharmacy	Italian Restaurant	Coffee Shop	Rental Car Location	Donut Shop	Chinese Restaurant	Bank	Salon / Barbershop	Zoo	Factory
9	63121	0	Chinese Restaurant	Thrift / Vintage Store	Pizza Place	American Restaurant	Fast Food Restaurant	Event Service	Food	Flower Shop	Flea Market	Fish & Chips Shop
10	63136	0	Cosmetics Shop	Dive Bar	Park	Farm	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish & Chips Shop	Fast Food Restaurant
15	63101	0	Hotel	Bar	Sandwich Place	Coffee Shop	Italian Restaurant	Mexican Restaurant	Sports Bar	American Restaurant	Pizza Place	Boutique
16	63102	0	Hotel	Casino	Restaurant	Italian Restaurant	Steakhouse	Bar	Dive Bar	Coffee Shop	Cocktail Bar	Outdoor Sculpture
17	63118	0	Mexican Restaurant	Fried Chicken Joint	Grocery Store	Bar	Bakery	Tea Room	Pizza Place	Coffee Shop	Noodle House	Music Store

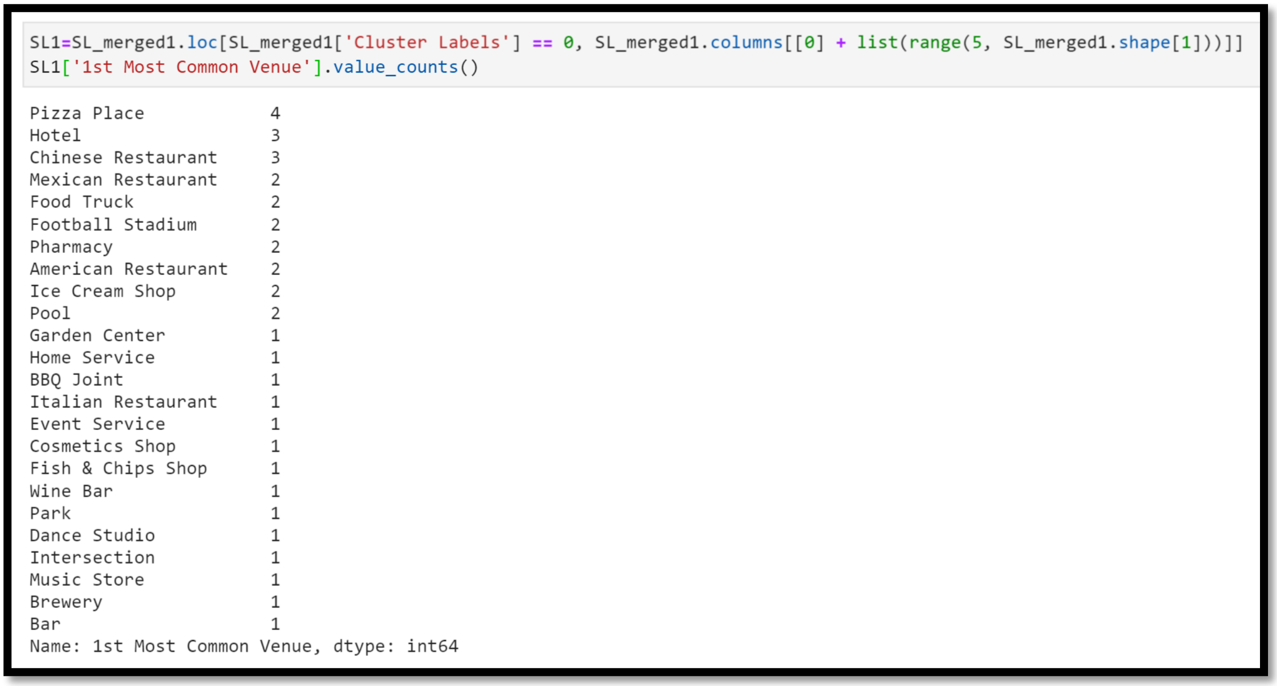


```
# List of Zip codes in this Cluster
SL1['Zip'].values
```

```
array([63103, 63133, 63144, 63121, 63136, 63101, 63102, 63118, 63120,
       63109, 63111, 63135, 63114, 63132, 63126, 63116, 63115, 63104,
       63117, 63143, 63127, 63107, 63146, 63131, 63141, 63112, 63147,
       63123, 63130, 63105, 63110, 63137, 63119, 63128, 63129, 63122,
       63139, 63108])
```

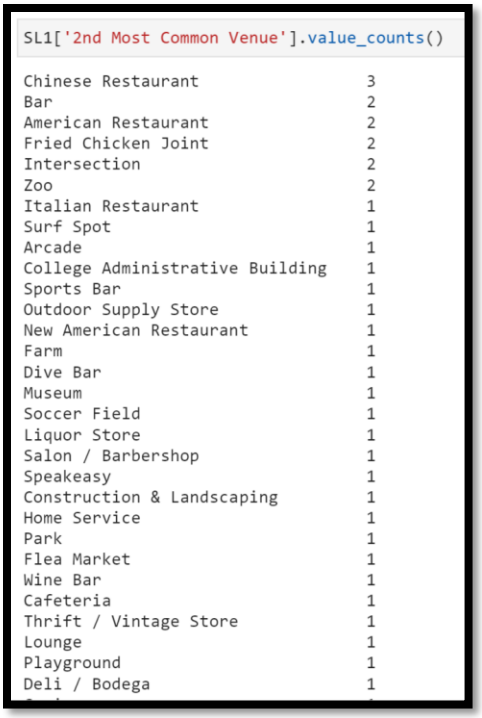
63103, 63133, 63144, 63121, 63136, 63101, 63102, 63118, 63120, 63109, 63111, 63135, 63114, 63132, 63126, 63116, 63115, 63104, 63117, 63143, 63127, 63107, 63146, 63131, 63141, 63112, 63147, 63123, 63130, 63105, 63110, 63137, 63119, 63128, 63129, 63122, 63139, 63108

In this cluster we have a total of 38 zip codes with different types of categories featuring in the 1st Most common Venue

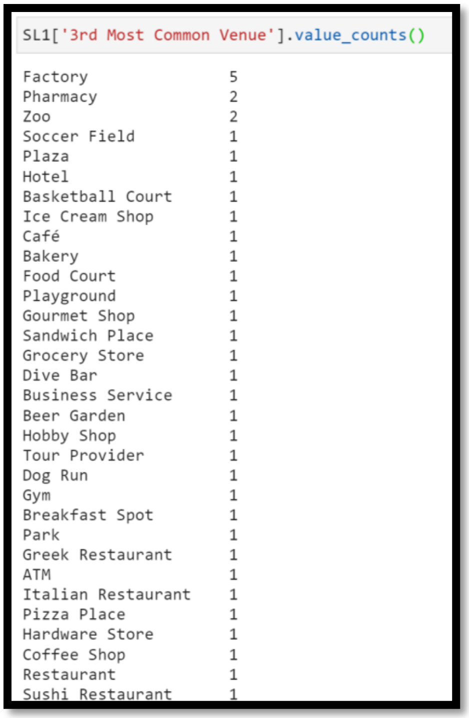


```
SL1=SL_merged1.loc[SL_merged1['Cluster Labels'] == 0, SL_merged1.columns[[0] + list(range(5, SL_merged1.shape[1]))]]
SL1['1st Most Common Venue'].value_counts()
```

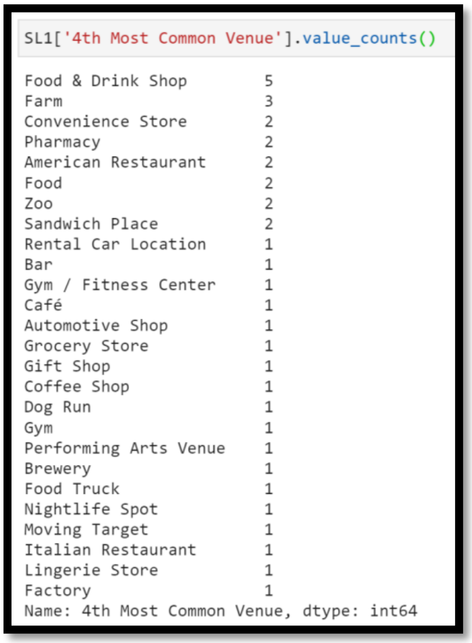
```
Pizza Place            4
Hotel                  3
Chinese Restaurant     3
Mexican Restaurant     2
Food Truck             2
Football Stadium       2
Pharmacy               2
American Restaurant    2
Ice Cream Shop         2
Pool                   2
Garden Center          1
Home Service           1
BBQ Joint              1
Italian Restaurant     1
Event Service          1
Cosmetics Shop         1
Fish & Chips Shop      1
Wine Bar               1
Park                   1
Dance Studio           1
Intersection           1
Music Store            1
Brewery                1
Bar                    1
Name: 1st Most Common Venue, dtype: int64
```



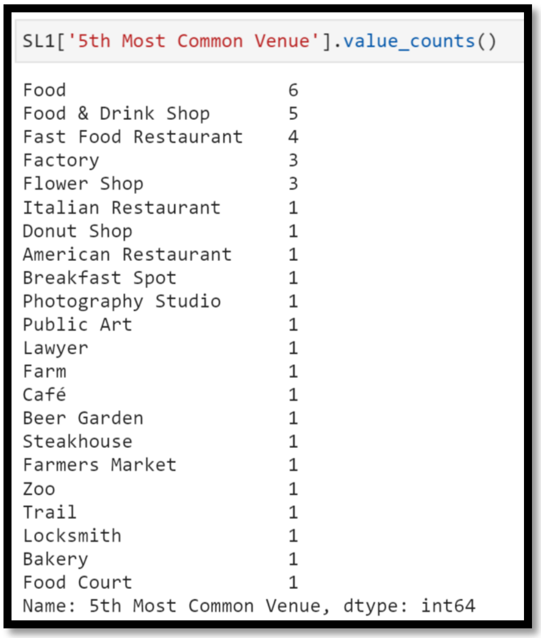
```
SL1['2nd Most Common Venue'].value_counts()

Chinese Restaurant                3
Bar                               2
American Restaurant               2
Fried Chicken Joint               2
Intersection                      2
Zoo                               2
Italian Restaurant                1
Surf Spot                         1
Arcade                            1
College Administrative Building   1
Sports Bar                        1
Outdoor Supply Store              1
New American Restaurant           1
Farm                              1
Dive Bar                          1
Museum                            1
Soccer Field                      1
Liquor Store                      1
Salon / Barbershop                1
Speakeasy                         1
Construction & Landscaping        1
Home Service                      1
Park                              1
Flea Market                       1
Wine Bar                          1
Cafeteria                         1
Thrift / Vintage Store            1
Lounge                            1
Playground                        1
Deli / Bodega                     1
```



```
SL1['3rd Most Common Venue'].value_counts()

Factory              5
Pharmacy             2
Zoo                  2
Soccer Field         1
Plaza                1
Hotel                1
Basketball Court     1
Ice Cream Shop       1
Café                 1
Bakery               1
Food Court           1
Playground           1
Gourmet Shop         1
Sandwich Place       1
Grocery Store        1
Dive Bar             1
Business Service     1
Beer Garden          1
Hobby Shop           1
Tour Provider        1
Dog Run              1
Gym                  1
Breakfast Spot       1
Park                 1
Greek Restaurant     1
ATM                  1
Italian Restaurant   1
Pizza Place          1
Hardware Store       1
Coffee Shop          1
Restaurant           1
Sushi Restaurant     1
```



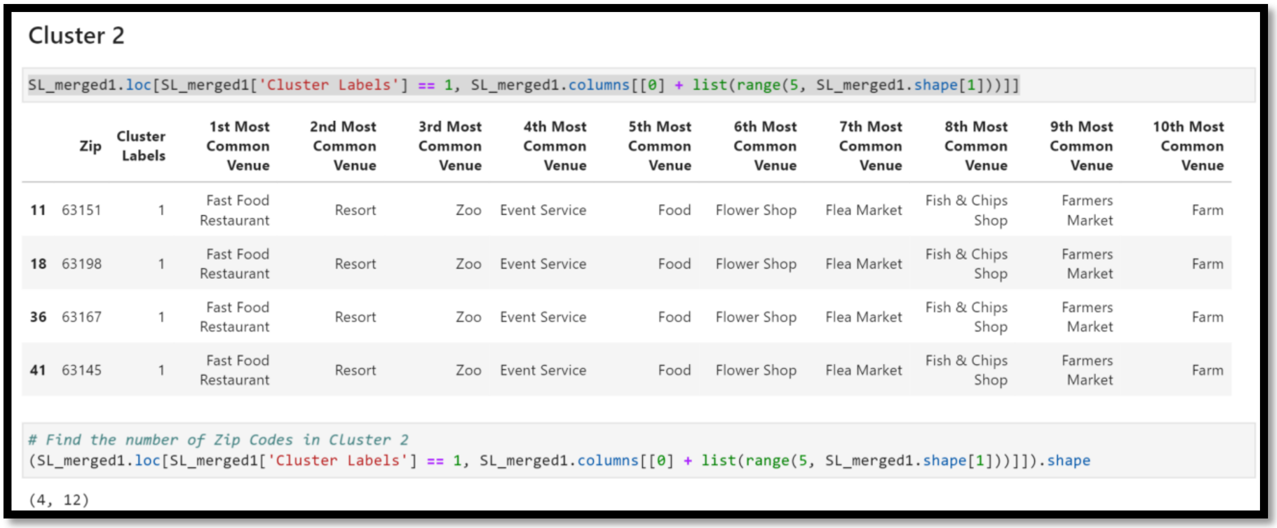
```
SL1['4th Most Common Venue'].value_counts()

Food & Drink Shop        5
Farm                     3
Convenience Store        2
Pharmacy                 2
American Restaurant      2
Food                     2
Zoo                      2
Sandwich Place           2
Rental Car Location      1
Bar                      1
Gym / Fitness Center     1
Café                     1
Automotive Shop          1
Grocery Store            1
Gift Shop                1
Coffee Shop              1
Dog Run                  1
Gym                      1
Performing Arts Venue    1
Brewery                  1
Food Truck               1
Nightlife Spot           1
Moving Target            1
Italian Restaurant       1
Lingerie Store           1
Factory                  1
Name: 4th Most Common Venue, dtype: int64
```



```
SL1['5th Most Common Venue'].value_counts()

Food                    6
Food & Drink Shop       5
Fast Food Restaurant    4
Factory                 3
Flower Shop             3
Italian Restaurant      1
Donut Shop              1
American Restaurant     1
Breakfast Spot          1
Photography Studio      1
Public Art              1
Lawyer                  1
Farm                    1
Café                    1
Beer Garden             1
Steakhouse              1
Farmers Market          1
Zoo                     1
Trail                   1
Locksmith               1
Bakery                  1
Food Court              1
Name: 5th Most Common Venue, dtype: int64
```

After doing further analysis, we can observe that the Top 5 most commonly venue categories for this cluster are: Pizza Place, Chinese restaurant, Factory, Food & Drink Shops, Food. There are other venues as well, but mostly this cluster of Zip codes is popular for Food joints.

2nd Cluster:



Cluster 2

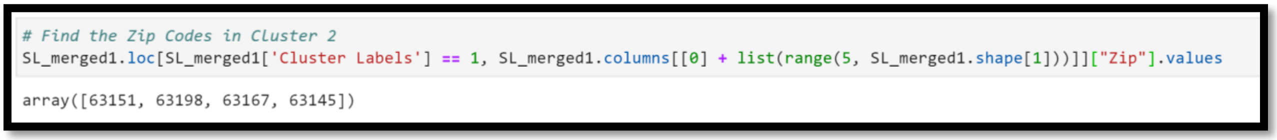
```
SL_merged1.loc[SL_merged1['Cluster Labels'] == 1, SL_merged1.columns[[0] + list(range(5, SL_merged1.shape[1]))]]
```

	Zip	Cluster Labels	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
11	63151	1	Fast Food Restaurant	Resort	Zoo	Event Service	Food	Flower Shop	Flea Market	Fish & Chips Shop	Farmers Market	Farm
18	63198	1	Fast Food Restaurant	Resort	Zoo	Event Service	Food	Flower Shop	Flea Market	Fish & Chips Shop	Farmers Market	Farm
36	63167	1	Fast Food Restaurant	Resort	Zoo	Event Service	Food	Flower Shop	Flea Market	Fish & Chips Shop	Farmers Market	Farm
41	63145	1	Fast Food Restaurant	Resort	Zoo	Event Service	Food	Flower Shop	Flea Market	Fish & Chips Shop	Farmers Market	Farm

```
# Find the number of Zip Codes in Cluster 2
(SL_merged1.loc[SL_merged1['Cluster Labels'] == 1, SL_merged1.columns[[0] + list(range(5, SL_merged1.shape[1]))]]).shape
```

```
(4, 12)
```

63151, 63198, 63167, 63145

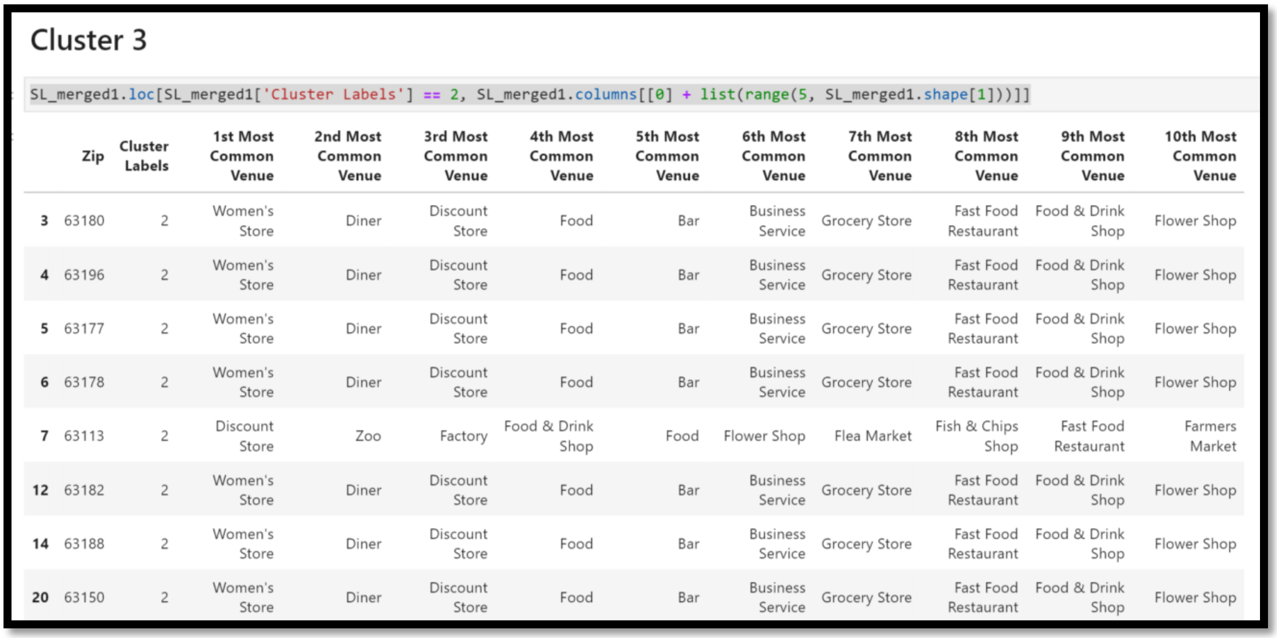


```
# Find the Zip Codes in Cluster 2
SL_merged1.loc[SL_merged1['Cluster Labels'] == 1, SL_merged1.columns[[0] + list(range(5, SL_merged1.shape[1]))]]["Zip"].values
```

```
array([63151, 63198, 63167, 63145])
```

This cluster has a total of 4 Zip codes with similar venue categories throughout.

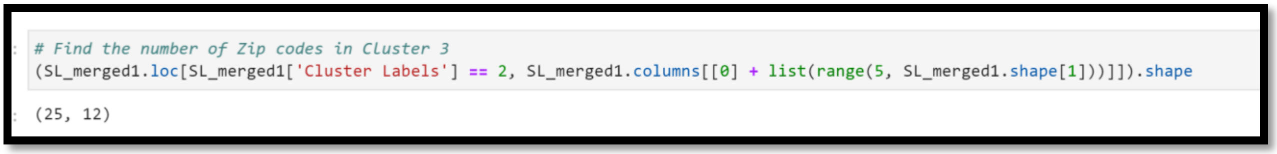
3rd Cluster:



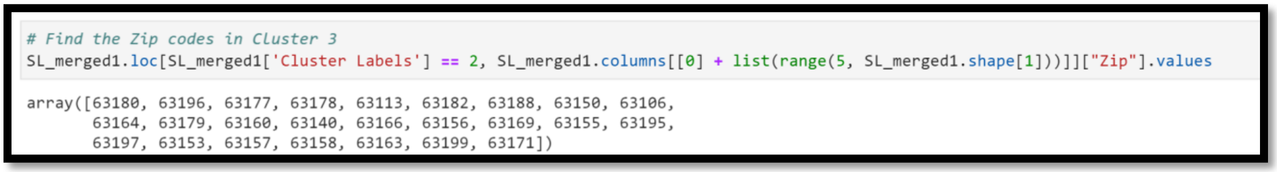
Cluster 3

```
SL_merged1.loc[SL_merged1['Cluster Labels'] == 2, SL_merged1.columns[[0] + list(range(5, SL_merged1.shape[1]))]]
```

	Zip	Cluster Labels	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	63180	2	Women's Store	Diner	Discount Store	Food	Bar	Business Service	Grocery Store	Fast Food Restaurant	Food & Drink Shop	Flower Shop
4	63196	2	Women's Store	Diner	Discount Store	Food	Bar	Business Service	Grocery Store	Fast Food Restaurant	Food & Drink Shop	Flower Shop
5	63177	2	Women's Store	Diner	Discount Store	Food	Bar	Business Service	Grocery Store	Fast Food Restaurant	Food & Drink Shop	Flower Shop
6	63178	2	Women's Store	Diner	Discount Store	Food	Bar	Business Service	Grocery Store	Fast Food Restaurant	Food & Drink Shop	Flower Shop
7	63113	2	Discount Store	Zoo	Factory	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish & Chips Shop	Fast Food Restaurant	Farmers Market
12	63182	2	Women's Store	Diner	Discount Store	Food	Bar	Business Service	Grocery Store	Fast Food Restaurant	Food & Drink Shop	Flower Shop
14	63188	2	Women's Store	Diner	Discount Store	Food	Bar	Business Service	Grocery Store	Fast Food Restaurant	Food & Drink Shop	Flower Shop
20	63150	2	Women's Store	Diner	Discount Store	Food	Bar	Business Service	Grocery Store	Fast Food Restaurant	Food & Drink Shop	Flower Shop

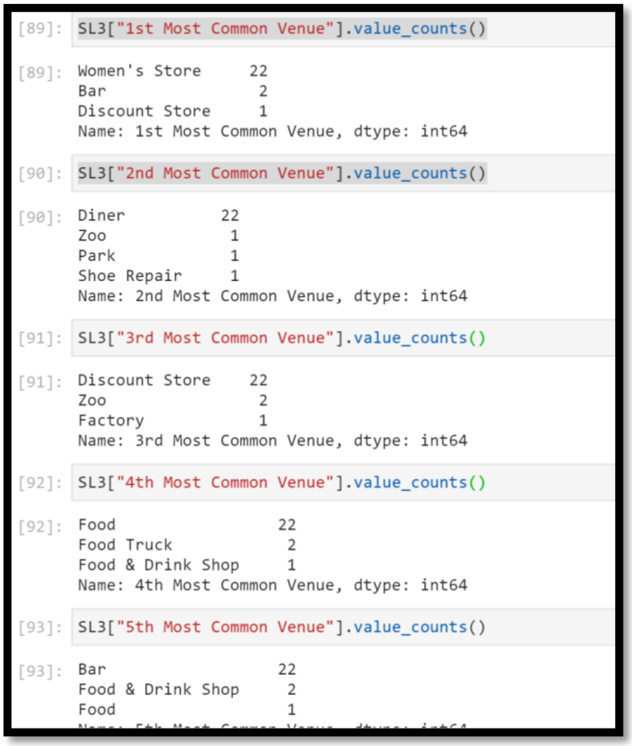


```
# Find the number of Zip codes in Cluster 3
(SL_merged1.loc[SL_merged1['Cluster Labels'] == 2, SL_merged1.columns[[0] + list(range(5, SL_merged1.shape[1]))]]).shape

(25, 12)
```



```
# Find the Zip codes in Cluster 3
SL_merged1.loc[SL_merged1['Cluster Labels'] == 2, SL_merged1.columns[[0] + list(range(5, SL_merged1.shape[1]))]]["Zip"].values

array([63180, 63196, 63177, 63178, 63113, 63182, 63188, 63150, 63106,
       63164, 63179, 63160, 63140, 63166, 63156, 63169, 63155, 63195,
       63197, 63153, 63157, 63158, 63163, 63199, 63171])
```

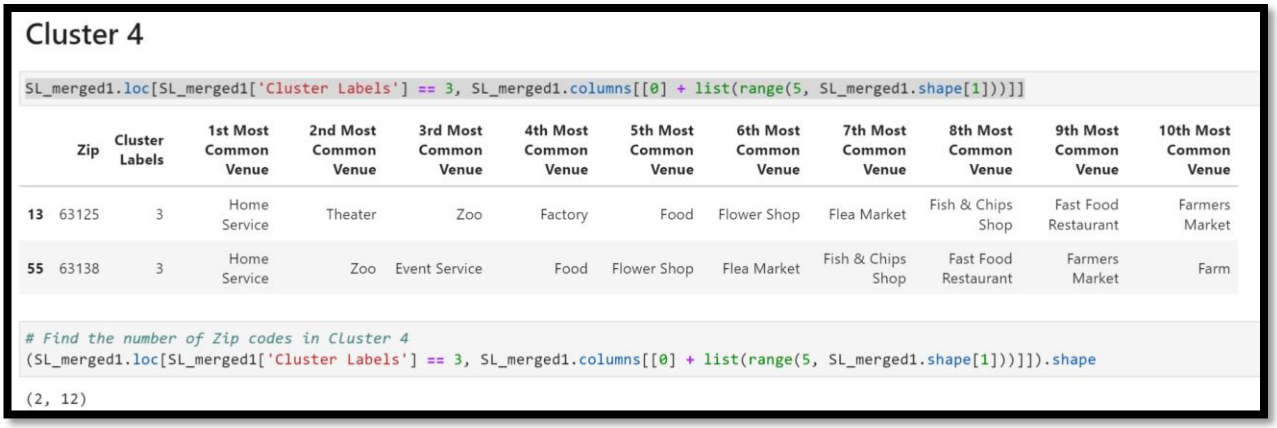
63180, 63196, 63177, 63178, 63113, 63182, 63188, 63150, 63106, 63164, 63179, 63160,63140, 63166, 63156, 63169, 63155, 63195, 63197, 63153, 63157, 63158, 63163, 63199, 63171



```
[89]: SL3["1st Most Common Venue"].value_counts()

[89]: Women's Store     22
      Bar                2
      Discount Store     1
      Name: 1st Most Common Venue, dtype: int64

[90]: SL3["2nd Most Common Venue"].value_counts()

[90]: Diner          22
      Zoo             1
      Park            1
      Shoe Repair     1
      Name: 2nd Most Common Venue, dtype: int64

[91]: SL3["3rd Most Common Venue"].value_counts()

[91]: Discount Store    22
      Zoo                2
      Factory            1
      Name: 3rd Most Common Venue, dtype: int64

[92]: SL3["4th Most Common Venue"].value_counts()

[92]: Food                22
      Food Truck           2
      Food & Drink Shop    1
      Name: 4th Most Common Venue, dtype: int64

[93]: SL3["5th Most Common Venue"].value_counts()

[93]: Bar                 22
      Food & Drink Shop    2
      Food                 1
      Name: 5th Most Common Venue, dtype: int64
```

We can see that in this cluster, there are a total of 25 Zip codes. There is a consistent pattern throughout the top 10 common venues with ‘Women -Store’ as the most common, followed by ‘Diner’, ‘Discount Store’, ‘Food’ and ‘Bar’ as the top 5 venues.

<u>4th Cluster:</u>



Cluster 4

```
SL_merged1.loc[SL_merged1['Cluster Labels'] == 3, SL_merged1.columns[[0] + list(range(5, SL_merged1.shape[1]))]]
```

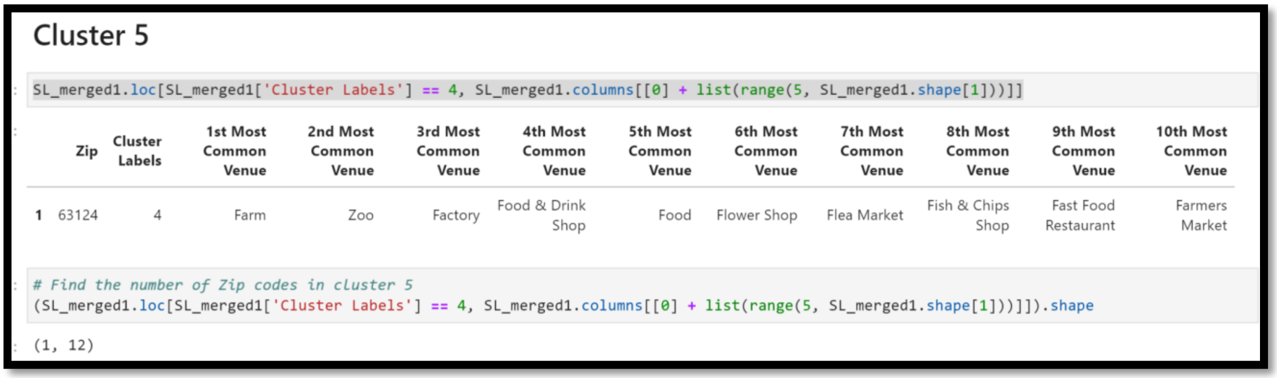
	Zip	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
13	63125	3	Home Service	Theater	Zoo	Factory	Food	Flower Shop	Flea Market	Fish & Chips Shop	Fast Food Restaurant	Farmers Market
55	63138	3	Home Service	Zoo	Event Service	Food	Flower Shop	Flea Market	Fish & Chips Shop	Fast Food Restaurant	Farmers Market	Farm

```
# Find the number of Zip codes in Cluster 4
(SL_merged1.loc[SL_merged1['Cluster Labels'] == 3, SL_merged1.columns[[0] + list(range(5, SL_merged1.shape[1]))]]).shape
```

```
(2, 12)
```

We have just 2 Zip codes in this cluster with 'Home Service' as most common followed by 'Theatre' and 'Zoo' as one of the top 5 venues

5th Cluster:



Cluster 5

```
SL_merged1.loc[SL_merged1['Cluster Labels'] == 4, SL_merged1.columns[[0] + list(range(5, SL_merged1.shape[1]))]]
```

	Zip	Cluster Labels	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
1	63124	4	Farm	Zoo	Factory	Food & Drink Shop	Food	Flower Shop	Flea Market	Fish & Chips Shop	Fast Food Restaurant	Farmers Market

```
# Find the number of Zip codes in cluster 5
(SL_merged1.loc[SL_merged1['Cluster Labels'] == 4, SL_merged1.columns[[0] + list(range(5, SL_merged1.shape[1]))]]).shape
```

```
(1, 12)
```

We have 1 Zip code in this cluster with 'Farm', 'Zoo' 'Factory', 'Food & Drink Shop' and 'Food' in the top 5 venues.

b. Recommendations

Since our problem statement was to identify which location will be best for the Entrepreneurs to open their businesses in the city of Saint Louis, we can have the certain recommendations based on our assessment of the individual clusters:

Cluster Number	Zip Codes
Cluster 1	63103, 63133, 63144, 63121, 63136, 63101, 63102, 63118, 63120, 63109, 63111, 63135, 63114, 63132, 63126, 63116, 63115, 63104, 63117, 63143, 63127, 63107, 63146, 63131,

	63141, 63112, 63147, 63123, 63130, 63105, 63110, 63137, 63119, 63128, 63129, 63122, 63139, 63108
Cluster 2	63151, 63198, 63167, 63145
Cluster 3	63180, 63196, 63177, 63178, 63113, 63182, 63188, 63150, 63106, 63164, 63179, 63160,63140, 63166, 63156, 63169, 63155, 63195, 63197, 63153, 63157, 63158, 63163, 63199, 63171
Cluster 4	63125, 63138
Cluster 5	63124

We can analyze the Most popular venue categories across different clusters to come up with the following set of Businesses that can be opened in these areas as there is already an existing demand for such businesses.

Business	**Preferred Clusters**
Pizza Place	Cluster1 (Most Popular Venues)
Chinese Restaurant	
Mexican Restaurant	
American Restaurant	
Fried Chicken Joint	
Food & Drink Shop	
Bar	Cluster 1 (Lesser in number but popular)
Cafe	
Ice cream shop	
Dance Studio	
Music Store	
Fast Food restaurant	Cluster 2
Resort	
Event Service	
Food (General)	
Womens’ Store	Cluster 3
Diner	
Discount Store	
Food (General)	
Bar	

Home Service	Cluster 4
Event Service	
Food (General)	
Flower Shop	
Theatre	
Factory	Cluster 5
Food (General)	

6. Conclusion

In this project, I tried to analyze the different areas in the city of Saint Louis with respect to the venue categories in each. This gave us a high-level understanding of the business types and their popularity in these areas. This information can be readily used to perform some competitor analysis by entrepreneurs to find out the locations which are best suitable to start their businesses. Dividing the locations into clusters gave us an insight into groups of locations which are similar and a Business would flourish by operating in these similar locations. The recommendations are provided above and can be used as a starting point to build up a business case based on existing information about these locations.