

Capstone Project-Battle of the Neighborhoods-Saint Louis

Kalpana Joshi

Introduction

➤ Background :

- In this project we will be exploring the various Neighborhoods in the city of Saint Louis and the popular venues in each of these.

➤ Problem:

- This analysis is aimed at finding out the best suited locations to start up a new business like a restaurant or a pub.

➤ Interested Parties:

- The result of this project will benefit the entrepreneurs in Saint Louis to determine the location that is best for them to set up shop.

Data Acquisition

➤ Data Sources

- 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.
- Foursquare API to determine the venues at each of these locations.

➤ **Data Cleaning:** First 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)

➤ **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 with their corresponding Latitudes and Longitudes.

	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

Methodology: Saint Louis Map with the Zip Codes

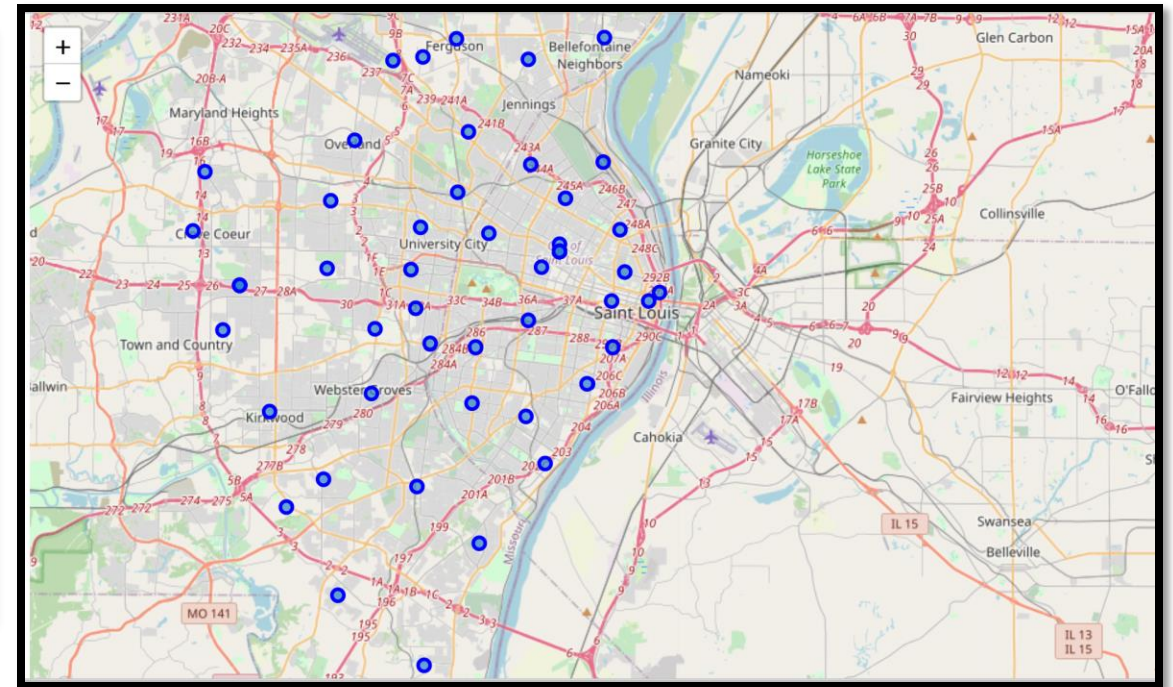
- We can visualize the dataframe with Zip codes and latitudes/longitudes on the map of Saint Louis using Folium library as shown below

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



Methodology: Foursquare API information

```
er
X CapstoneProject_Week4_niX
X Code git Submit Notebook ...
+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 can utilize the Foursquare API to get the venues around each of the Zip codes in Saint Louis and then arrange the information in a pandas dataframe

	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

Methodology: Dataframe with Venue categories for all Zip Codes

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 are a total of 630 venues across different Zip codes out of which 169 are unique values.

Methodology: Data Normalization for KMeans Clustering

```
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

- For further analysis, we normalize our data by performing ‘One-hot coding’ (by creating new columns for all 169 venues across all the Zip codes and assigning dummy values according to their presence in the location) . Now, we can analyse the Top venues at each of the Zip codes

Results: Cluster Labels

```
# 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

- **Machine Learning:** KMeans Clustering methodology is used to segregate the zip codes into clusters based on their venue categories as follows. We initialize our clusters to 5 for this analysis. Find the cluster labels and add them to the existing dataframe

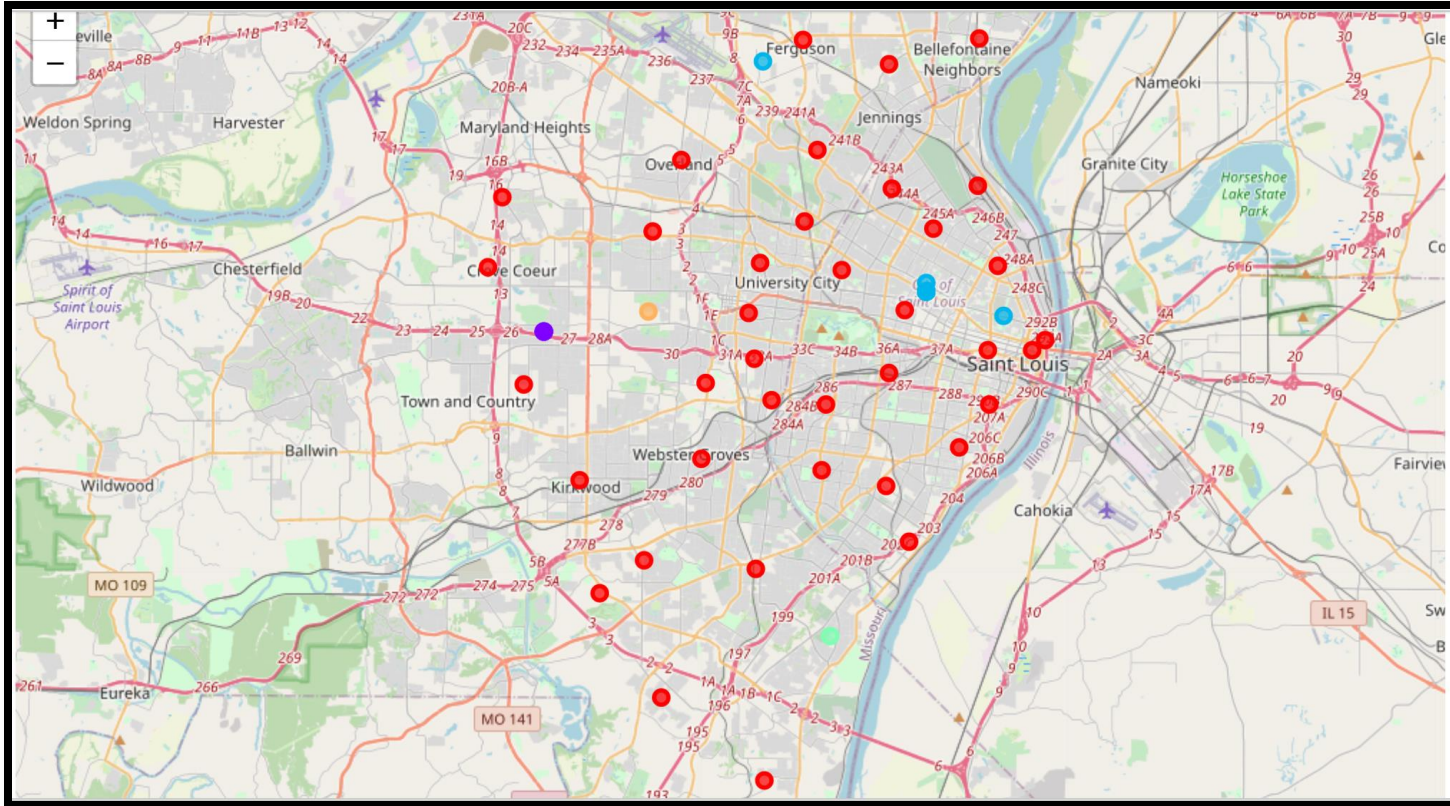
Results

```
# 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

- 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:

Results : Saint Louis Map with the Clusters



- We can clearly observe the 5 clusters in the map above and will delve into the details of each in the following slides

Discussion: Observations (Cluster 1)

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

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

Discussion: Observations (Cluster 2)

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)
```

- This cluster has a total of 4 Zip codes with similar venue categories throughout.
- The most popular venue categories are 'Fast Food Restaurants', 'Resort', 'Zoo', 'Event Service', 'Food'

Discussion: Observations (Cluster 3)

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

- 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.

Discussion: Observations (Cluster 4 & 5)

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', 'Theatre' and 'Zoo' as some of the top 5 venues

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.

Discussion: Recommendations-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

Recommendations: Businesses with Preferred Cluster locations

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)	

Business	Preferred Cluster
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 and Drink Shop	
Food (General)	

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 location using Foursquare API
- This gave us a high-level understanding of the business types and their popularity in these areas.
- 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