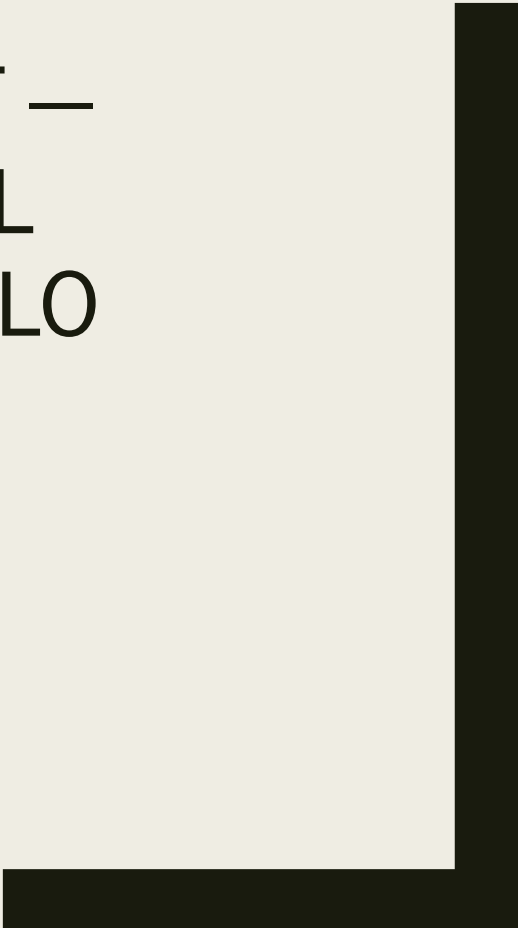
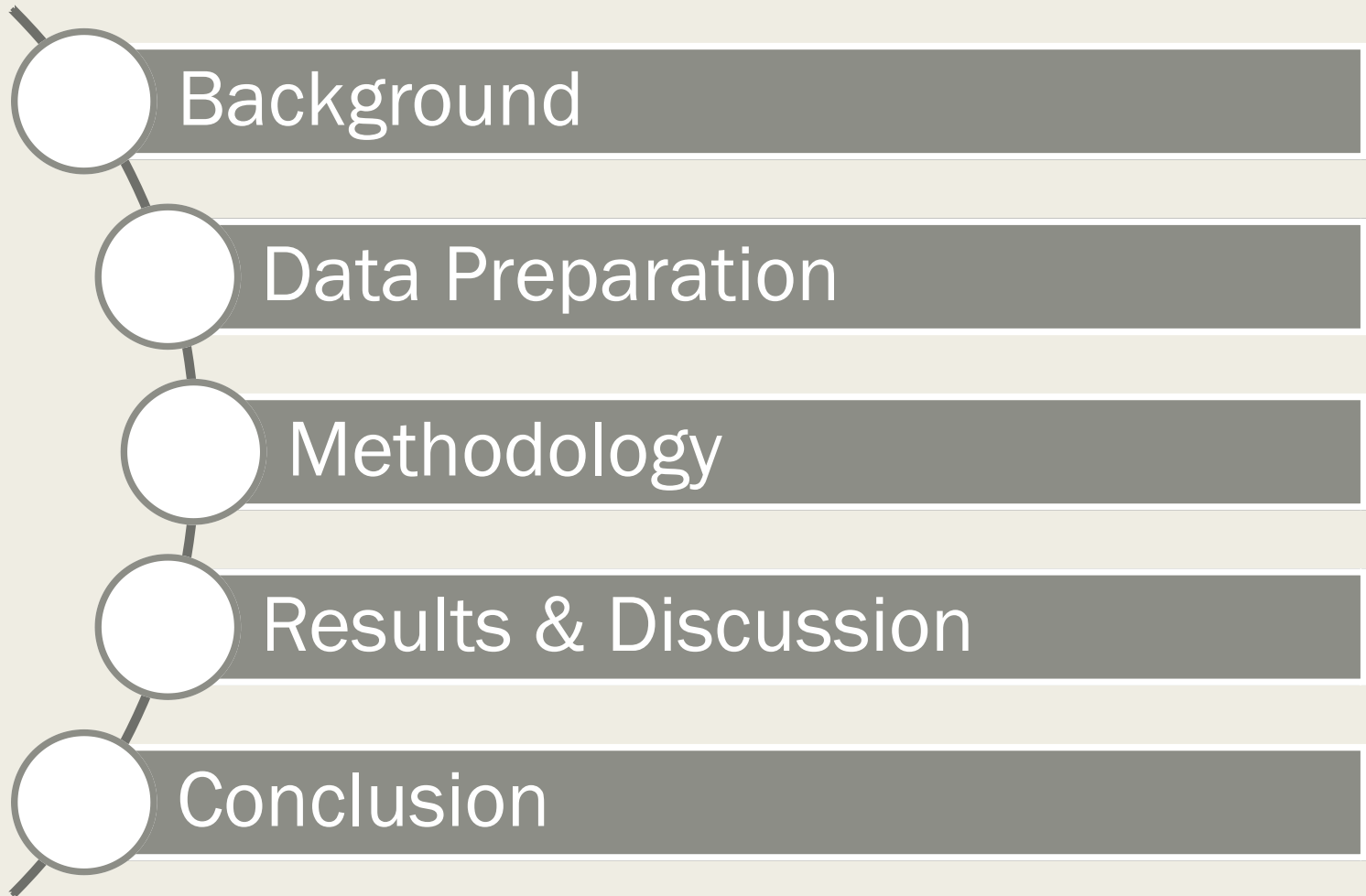


# CAPSTONE PROJECT — A STUDY ON SMALL BUSINESS IN BUFFALO

06/24/2020



# Outline



# Background

**Buffalo** is the second largest city in the U.S. state of New York and the largest city in Western New York.

**Business problem:** you live in Buffalo and plan to do some investment or start your own business.

**Target audience:** entrepreneurs or business owners who want to start a new business or invest in the existing businesses in Buffalo



[https://en.wikipedia.org/wiki/Buffalo,\\_New\\_York](https://en.wikipedia.org/wiki/Buffalo,_New_York)

## **Objective:**

- which district is likely to have more potential customers
- how many different types of venues there are in this district
- what the most popular venues in this area are

# Data Preparation

- The population data in each district of Buffalo:

<https://www.zip-codes.com/city/ny-buffalo.asp>

	Zip code	Type	County	Population	Area code
0	14201	Standard	Erie	11549	716
1	14202	Standard	Erie	3911	716
2	14203	Standard	Erie	1618	716
3	14204	Standard	Erie	8691	716
4	14205	P.O. Box	Erie	0	716

- The zip code latitude and longitude data of Buffalo:

<https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/?q=buffalo>

	Zip code	City	State	Latitude	Longitude	Timezone	Daylight savings time	geopoint
0	14201	Buffalo	NY	42.896407	-78.885150	-5	1	42.896407,-78.88515
1	14202	Buffalo	NY	42.886357	-78.877900	-5	1	42.886357,-78.8779
2	14203	Buffalo	NY	42.880107	-78.869900	-5	1	42.880107,-78.8699
3	14204	Buffalo	NY	42.884008	-78.861520	-5	1	42.884008,-78.86152
4	14205	Buffalo	NY	42.768413	-78.887094	-5	1	42.768413,-78.887094

- The information on venues in Buffalo:  
obtained by utilizing Foursquare API

# Methodology

Select the districts suitable for small businesses in Buffalo

Explore the venues in selected districts (Foursquare API)

What venues are popular in each district (one hot encoding)

K-Means Clustering

# Results & Discussion

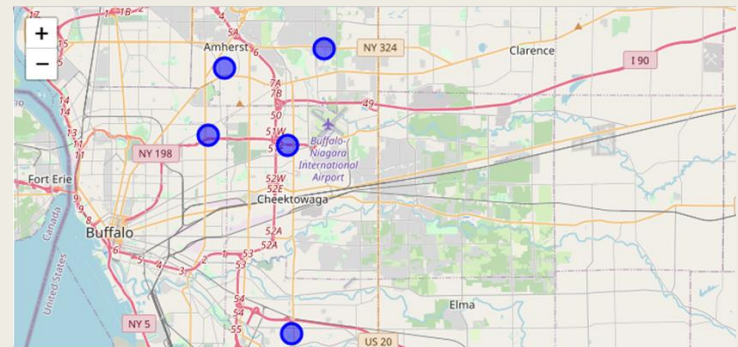
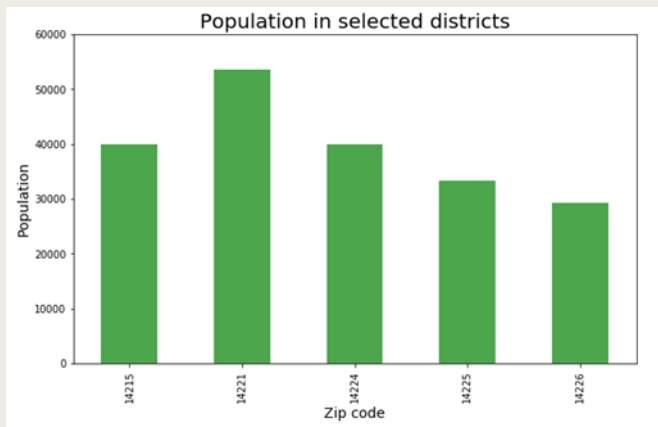
## Select the districts suitable for small businesses in Buffalo

- Merged the population data and geopoint data to get the population data in each district of Buffalo
- Find a place where there are more potential customers--The larger population, the better

population larger than 25,000

```
# Ignore cells with population less than 25000
df=df.drop(df[df['Population']<25000].index)
df.reset_index(drop=True, inplace=True)
df
```

	Zip code	Type	County	Population	Area code	City	State	Latitude	Longitude	Timezone	Daylight savings time	geopoint
0	14215	Standard	Erie	39999	716	Buffalo	NY	42.934757	-78.81180	-5	1	42.934757,-78.8118
1	14221	Standard	Erie	53555	716	Buffalo	NY	42.977456	-78.73356	-5	1	42.977456,-78.73356
2	14224	Standard	Erie	39889	716	Buffalo	NY	42.836858	-78.75557	-5	1	42.836858,-78.75557
3	14225	Standard	Erie	33385	716	Buffalo	NY	42.929891	-78.75813	-5	1	42.929891,-78.75813
4	14226	Standard	Erie	29267	716	Buffalo	NY	42.968057	-78.80047	-5	1	42.968057,-78.80047



# Results & Discussion

## Explore the venues in selected districts

- By utilizing the Foursquare API, a function to look up the venues was defined
- The venues in each district were obtained (includes the venue name, latitude, longitude and category data)

buffalo_venues					
	Zip code	Venue	Venue Latitude	Venue Longitude	Venue Category
0	14215	Hair World	42.925755	-78.813374	Cosmetics Shop
1	14215	Bailey Fish & Seafood	42.945048	-78.813887	Seafood Restaurant
2	14215	99 Fast Food Restaurant	42.947685	-78.813821	Vietnamese Restaurant
3	14215	Louie's Texas Red Hots	42.931152	-78.813312	Hot Dog Joint
4	14215	Dollar General	42.938669	-78.813275	Discount Store
...	...	...	...	...	...
323	14226	Walgreens	42.957649	-78.819353	Pharmacy
324	14226	Main & Kenmore	42.957353	-78.819001	Intersection
325	14226	Sweeney's Garage	42.965832	-78.823875	Auto Garage
326	14226	Starbucks	42.979629	-78.782035	Coffee Shop
327	14226	Basil Resale Sheridan	42.977818	-78.779820	Used Auto Dealership

328 rows × 5 columns

- Total number of unique categories in these selected areas is 130.

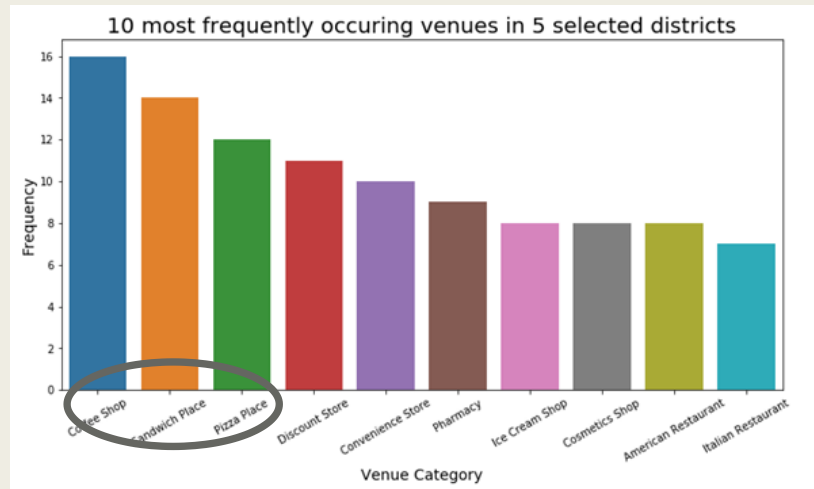
```
# Count how many unique categories in selected areas
print('There are {} unique categories in selected areas in Buffalo.'.format(len(buffalo_venues['Venue Category'].unique())))
```

There are 130 unique categories in selected areas in Buffalo.

# Results & Discussion

## Explore the venues in selected districts

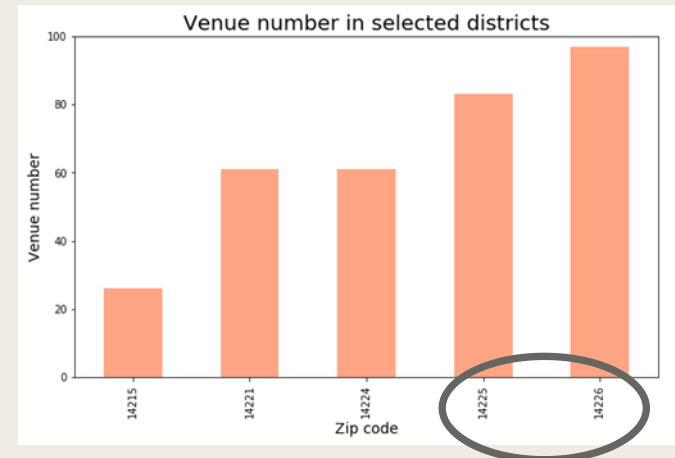
- The number of each unique category was counted and assorted (5 districts were treated as a whole)



The top three venue categories:

- coffee shops
- sandwich places
- pizza places

Although 14225 and 14226 districts were ranked as the fourth and fifth in population, the venue number in these two districts were higher than the other three districts





# Results & Discussion

## One hot coding

Transform categorical data into numerical data

```
buffalo_onehot = pd.get_dummies(buffalo_venues[['Venue Category']], prefix="", prefix_sep="")

# add the column back to dataframe
buffalo_onehot['Zip code'] = buffalo_venues['Zip code']

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

buffalo_venues_grouped = buffalo_onehot.groupby('Zip code').mean().reset_index()
buffalo_venues_grouped
```

	Zip code	ATM	Accessories Store	Airport Service	American Restaurant	Arcade	Arts & Crafts Store	Auto Garage	Auto Workshop	BBQ Joint	...	Video Store	Vietnamese Restaurant	Warehouse Store	Waterfall	
0	14215	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.038462	0.000000	0.000000	0
1	14221	0.000000	0.000000	0.000000	0.032787	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.016393	0.000000	0.000000	0.016393	0
2	14224	0.000000	0.000000	0.000000	0.016393	0.000000	0.000000	0.000000	0.016393	0.000000	...	0.000000	0.000000	0.000000	0.000000	0
3	14225	0.000000	0.012048	0.012048	0.036145	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.012048	0.000000	0.012048	0.000000	0
4	14226	0.010309	0.000000	0.000000	0.020619	0.010309	0.020619	0.020619	0.000000	0.010309	...	0.041237	0.000000	0.000000	0.000000	0

5 rows x 131 columns

- District 14221: coffee shops, Italian restaurants and ice cream shops are more popular
- District 14225: clothing stores, sandwich places and cosmetics shops are more popular
- District 14226: Chinese restaurants and pizza places are more popular

Zip code: 14215

	venue	freq
0	Discount Store	0.19
1	Convenience Store	0.12
2	Coffee Shop	0.12
3	Seafood Restaurant	0.08
4	Cosmetics Shop	0.04

Zip code: 14221

	venue	freq
0	Coffee Shop	0.08
1	Italian Restaurant	0.05
2	Ice Cream Shop	0.05
3	Greek Restaurant	0.03
4	Park	0.03

Zip code: 14224

	venue	freq
0	Pizza Place	0.07
1	Ice Cream Shop	0.05
2	Discount Store	0.05
3	Sandwich Place	0.05
4	Pharmacy	0.03

Zip code: 14225

	venue	freq
0	Clothing Store	0.06
1	Cosmetics Shop	0.05
2	Sandwich Place	0.05
3	Lingerie Store	0.04
4	American Restaurant	0.04

Zip code: 14226

	venue	freq
0	Chinese Restaurant	0.05
1	Pizza Place	0.05
2	Convenience Store	0.04
3	Coffee Shop	0.04
4	Pharmacy	0.04

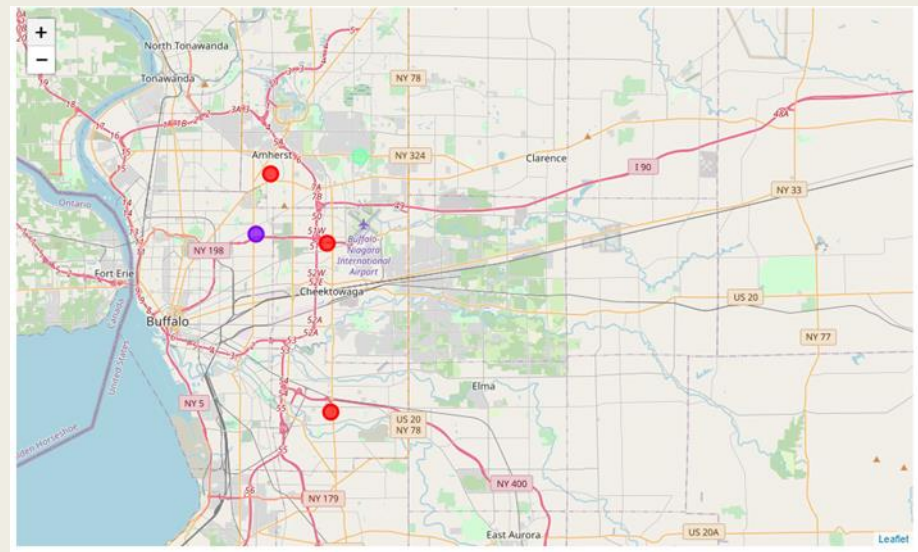
# Results & Discussion

## K Means Clustering

	Zip code	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	14215	Discount Store	Convenience Store	Coffee Shop	Seafood Restaurant	Mediterranean Restaurant	Sandwich Place	Cosmetics Shop	Pizza Place	Pharmacy	Paper / Office Supplies Store
1	14221	Coffee Shop	Italian Restaurant	Ice Cream Shop	Thai Restaurant	Golf Course	Greek Restaurant	Gym	Park	Café	Sandwich Place
2	14224	Pizza Place	Ice Cream Shop	Sandwich Place	Discount Store	Cosmetics Shop	Gas Station	Sports Bar	Clothing Store	Coffee Shop	Bar
3	14225	Clothing Store	Sandwich Place	Cosmetics Shop	Hotel	Lingerie Store	American Restaurant	Italian Restaurant	Optical Shop	Ice Cream Shop	Pizza Place
4	14226	Chinese Restaurant	Pizza Place	Pharmacy	Sandwich Place	Video Store	Convenience Store	Coffee Shop	Discount Store	Rental Car Location	Greek Restaurant

Three clusters

- Cluster 0: 14224, 14225 and 14226
- Cluster 1: 14215
- Cluster 2: 14221



# Conclusion

- Overall, coffee shops, sandwich places and pizza places are the three most popular venues in five major districts (14215, 14221, 14224, 14225 and 14226) in Buffalo
- District 14221 is more likely to have more potential customers, and coffee shops, Italian restaurants and ice cream shops are most popular in this area
- District 14225 and 14226 has more venues survived. In district 14225, clothing stores, sandwich places and cosmetics shops are more popular; in district 14226, Chinese restaurants and pizza places are more popular