THE BATTLE OF NEIGHBORHOODS

- WHERE TO START A SUSHI RESTAURANT IN CALGARY, ALBERTA ?

BY DAVID JIANZHE LI

AGENDA

- Introduction
- Data and EDA
- K-means Clustering
- Recommendations
- Future consideratons

INTRO

- Calgary is a city in the western Canadian province of Alberta, Canada
- 80 km east of the Canadian Rockies.
- Biggest city in Alberta and the second largest in western Canada after only Vancouver

BUSINESS PROBLEM:

Where to start a Sushi restaurant in Calgary, Canada?

Data sources

- Neighborhood and area -Wiki
- Neighborhood demographics – great-news.ca
- Geo data from Geocoder

	Community	Median Household Income	Population	Area	PopulationDensity	Latitude	Longitude
0	Abbeydale	55345.0	6071	1.7	3571.176471	51.05976	-113.92546
1	Acadia	46089.0	10969	3.9	2812.564103	50.97227	-114.05882
2	Albert Park / Radisson Heights	38019.0	6529	2.5	2611.600000	51.04200	-113.99683
3	Altadore	53786.0	9518	2.9	3282.068966	51.01601	-114.10558
4	Applewood Park	65724.0	6864	1.6	4290.000000	51.04544	-113.92513
5	Arbour Lake	70590.0	10987	4.4	2497.045455	51.13364	-114.20307
6	Aspen Woods	133939.0	7496	3.8	1972.631579	51.04519	-114.21160
7	Auburn Bay	84350.0	11127	4.5	2472.666667	50.88976	-113.96397
8	Banff Trail	49996.0	4204	1.5	2802.666667	51.07472	-114.11297
9	Bankview	32474.0	5416	0.7	7737.142857	51.03412	-114.10044

Venue data from Foursquare API - food Section ONLY

- https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&***section*=**
  {}&limit={}'.format(
- CLIENT_ID,
- CLIENT_SECRET,
- VERSION,
- lat,
- lng,
- radius,
- ***SECTION***,
- LIMIT)

	Community	Community Latitude	Community Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abbeydale	51.05976	-113.92546	Atlas Pizza and Sports Bar	51.052481	-113.941859	Pizza Place
1	Abbeydale	51.05976	-113.92546	A&W	51.068291	-113.933571	Fast Food Restaurant
2	Abbeydale	51.05976	-113.92546	Subway	51.059239	-113.934423	Sandwich Place
3	Abbeydale	51.05976	-113.92546	Subway	51.069623	-113.932907	Sandwich Place
4	Abbeydale	51.05976	-113.92546	Subway	51.052786	-113.942449	Sandwich Place

Feature Extraction

	Community	Median Household Income	PopulationDensity	venueCount	Chinese Restaurant	Japanese Restaurant	Sushi Restaurant	Dim Sum Restaurant
0	Abbeydale	55345.0	3571.176471	12.0	0.083333	0.0	0.00	0.00
1	Acadia	46089.0	2812.564103	50.0	0.020000	0.0	0.04	0.02
2	Albert Park / Radisson Heights	38019.0	2611.600000	27.0	0.000000	0.0	0.00	0.00
3	Altadore	53786.0	3282.068966	20.0	0.050000	0.0	0.15	0.00
4	Applewood Park	65724.0	4290.000000	8.0	0.000000	0.0	0.00	0.00

Data normalization – max/min scaler

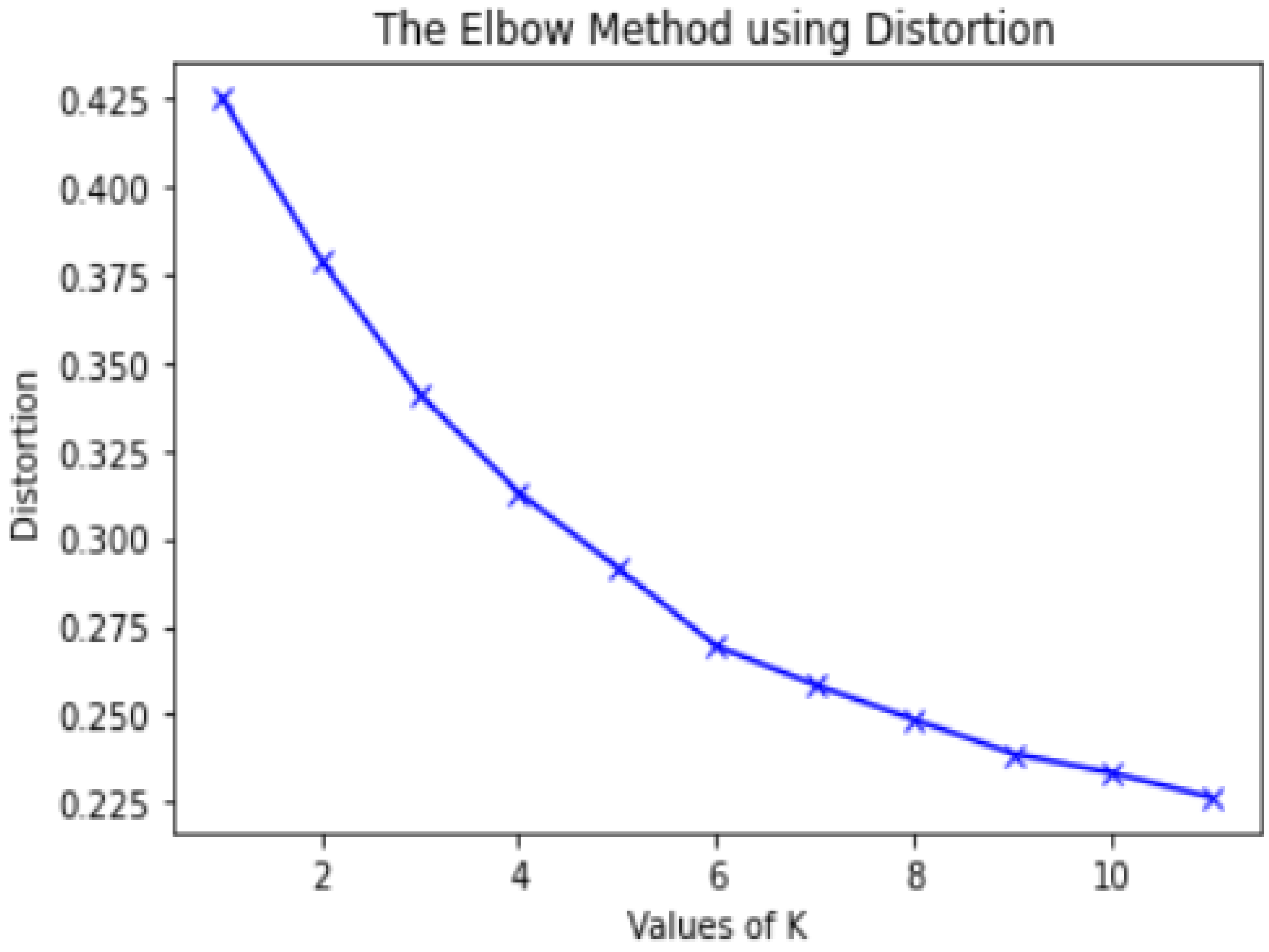
$$\text{featureDF\_scaled} = \frac{\text{featureDF} - \text{featureDF.min()}}{\text{featureDF.max()} - \text{featureDF.min()}}$$

	Median Household Income	PopulationDensity	venueCount	Chinese Restaurant	Japanese Restaurant	Sushi Restaurant	Dim Sum Restaurant
0	0.161939	0.319253	0.12	0.208333	0.0	0.00	0.00
1	0.101824	0.250325	0.50	0.050000	0.0	0.16	0.36
2	0.049412	0.232065	0.27	0.000000	0.0	0.00	0.00
3	0.151813	0.292985	0.20	0.125000	0.0	0.60	0.00
4	0.229347	0.384566	0.08	0.000000	0.0	0.00	0.00

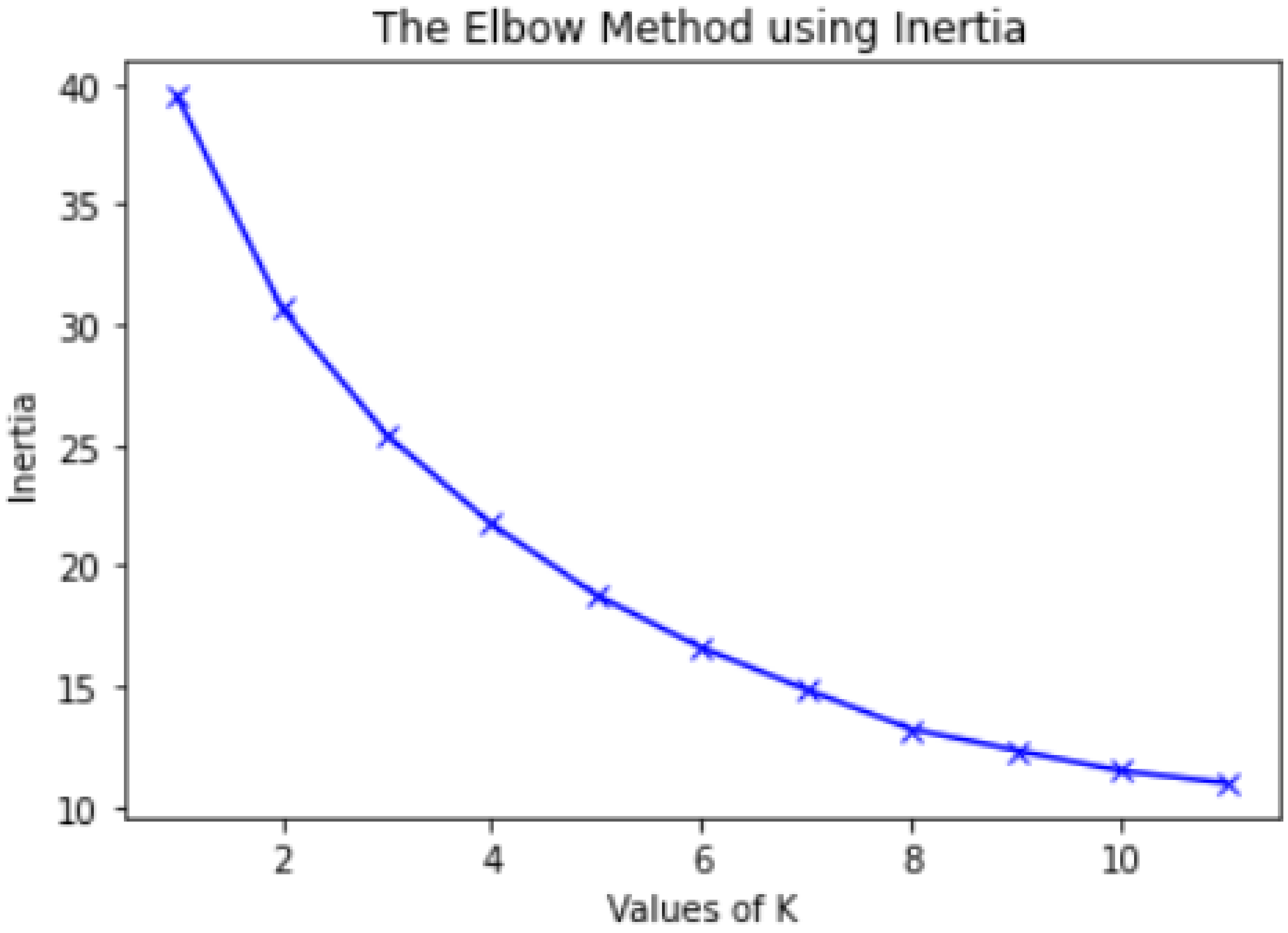
Find k with Elbow method – K-means clustering

- k: Optimal number of clusters
- Distortion: average of the squared distances from the cluster centers of the respective clusters.
- Inertia: sum of squared distances of samples to their closest cluster center.

Elbow method using Distortion(k = 5)



Elbow method using Inertia(k = 5)

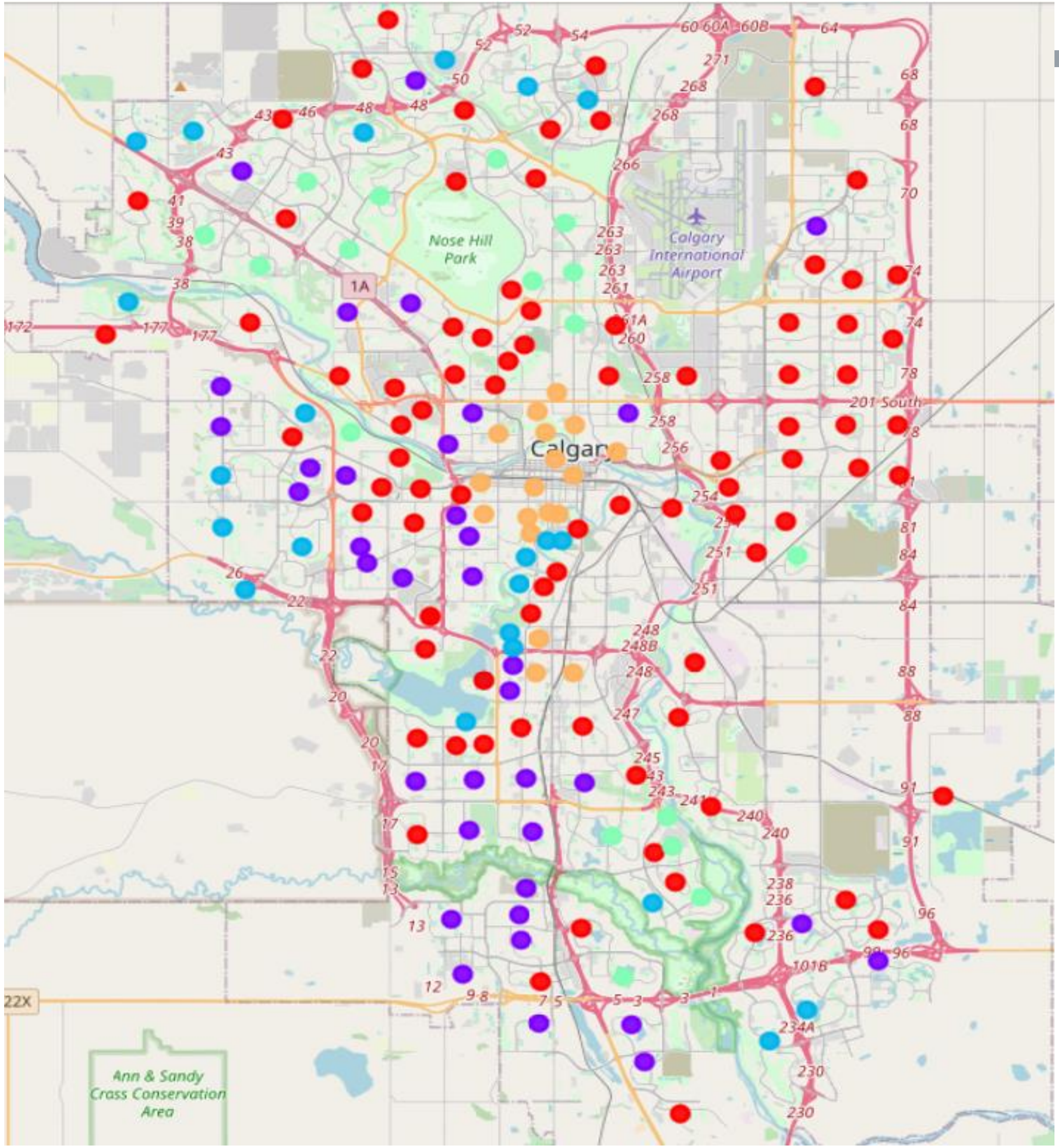


K-means clustering (k = 5)

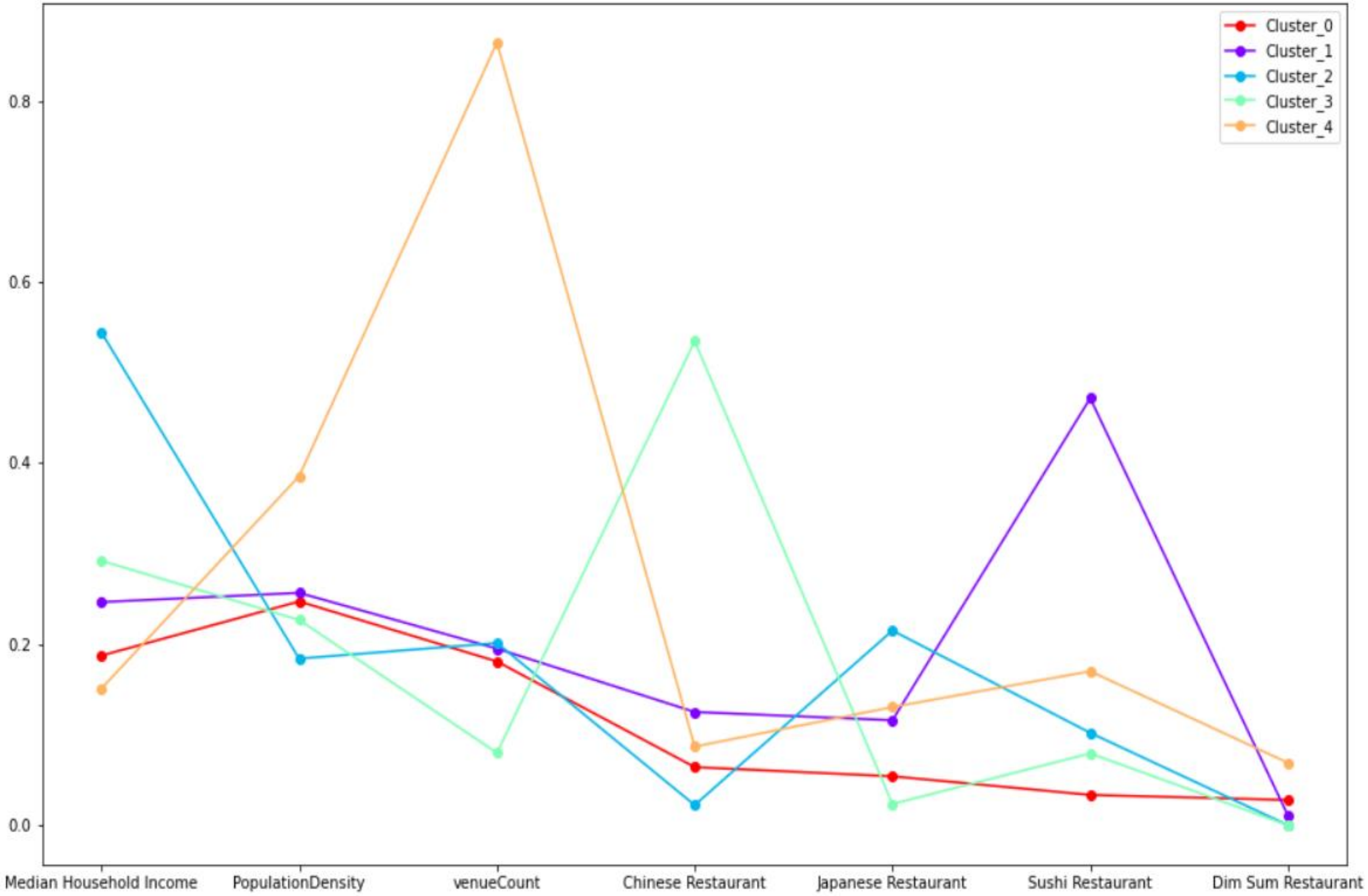
```
# set number of clusters
kclusters = 5

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

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
array([0, 0, 0, 1, 0, 1, 2, 2, 0, 4])
```

Clustering on the map



Clustering line graph



Clustering analysis

- Cluster 0: Lower family income, medium population density, fair overall food venue coverage and low East Asian restaurant coverage.
- Cluster 1: Medium family income, medium population density, medium overall food venue coverage with fair Chinese/Japanese restaurant coverage, but highest Sushi restaurant coverage.
- Cluster 2: Highest family income, lowest population density and fair overall food venue coverage, highest Japan restaurant coverage, but low Sushi restaurant coverage.
- Cluster 3: High family income, lower density and lowest overall food venue coverage, highest Chinese restaurant coverage, but low Sushi restaurant coverage
- Cluster 4: Lowest family income, highest population density and overall food venue coverage, high East Asian restaurant coverage.

Recommendations

- Cluster 0: Not recommended even if a low East Asian restaurant coverage, due to its low family income, fair overall food venue coverage.
- Cluster 1: Not recommended due to its highest Sushi restaurant coverage (fierce competition), medium population density, and medium family income.
- Cluster 2: Highly recommended for a high-end Sushi restaurant. Highest family income, low Sushi restaurant coverage. Some competition from Japanese restaurants.
- Cluster 3: Highly recommended for a high-end Sushi restaurant. High family income, low Sushi restaurant coverage. Some competition from Chinese restaurants.
- Cluster 4: Highly recommended for low-end fast-food like Sushi restaurant. Lowest family income, highest population density and overall food venue coverage. Downtown or hub areas.

Future considerations – ways to improve

- Improved Foursquare data
- More neighborhood data, such as:

  median rent, median age, proximity to C-train line, and etc.
- Try other clustering algorithm, Hierarchical clustering, and compare performance

THANK YOU