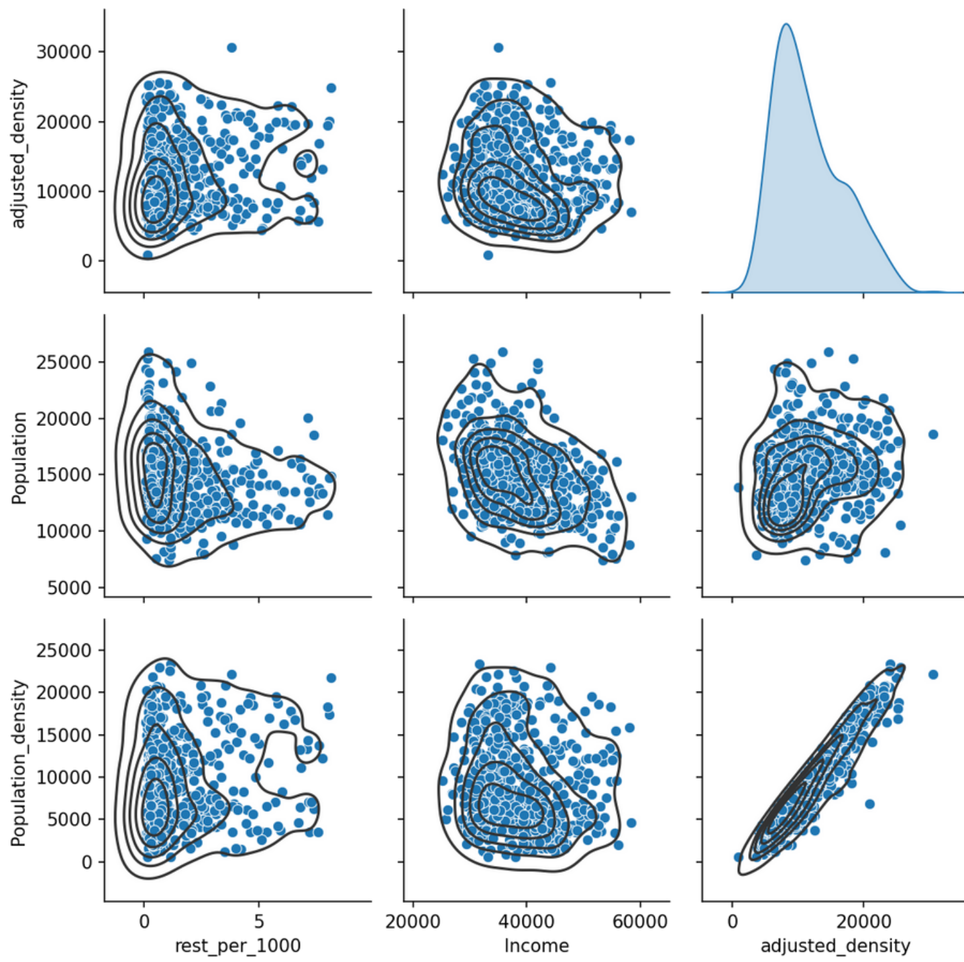
Applied Data Science Capstone Project

Opening restaurant in the Greater London area



Author: Ksenia Tabakova

A restaurant entrepreneur aims to open more restaurants in the Greater London area. The client already runs restaurants of American, Mexican and Italian cuisine.

The client wants to know answers for the following questions:

1. **Which Greater London areas are the best to open restaurant(s)?**
2. **Which cuisine(s) (American, Mexican or Italian) are the best option to open in the identified areas?**
3. **What price range for newly opened restaurant should be considered (low or high price tag)?**

Possible factors affecting the future of a new restaurant

Abundance of similair restaurants in the same area

It can be hard to compete for clients in the area where there are already many similar restaurants.

Population/housing density

Higher population density is expected to support more small businesses, including restaurants.

Income of citizens

Disposable income can define overall habits of citizens of a neighborhood as well as possible menu pricing.

Expected population growth in the area

The higher the population growth, the better outcome for a restaurant is expected

Note: There are other parameters that impact survival rate of a restaurant, e.g. statistics on sales of restaurants, distribution of population by age. I could not find any such data.

Data sources and description

LONDON DATASTORE

https://data.london.gov.uk/

Land Area and Population Density, Ward and Borough.

Variables: Population in 2021, Population density in 2021, Population in 2026.

Statistical GIS Boundary Files for London.

Variables: Geometry of wards in 2011 (polygons)

Household Income Estimates for Small Areas.

Variable: Median household income per ward, 2012/13

Access to Public Open Space and Nature by Ward

Variable: Percent of open space are in the ward, 2011

Foursquare data

Variables: food venues in wards

More on data

London Datastore provides open cohesive datasets with homogenized geographical coding system. All used datasets contained unique codes for each ward, making it easy to manipulate and merge dataframe.
Example of data:

```
income.head()
```

	Code	Income
0	E09000001	63620
1	E05000026	33920
2	E05000027	32470
3	E05000028	33000
4	E05000029	33920

- There were 624 wards in London at the last census in 2011.
- Data was easy to understand, manipulate and did not require much cleaning.
- Wards geometry was converted from easting and northing in meters to degrees of latitude and longitude.
- Adjusted population density was calculated using open spaces data.
- Expected population growth by the year 2026 was calculated.
- Restaurant number per 1000 people was calculated.

Foursquare API calls

Explore calls, 750 m radius, pagination when necessary (for wards with >100 venues), categoryIDs endpoint used to call only food venues.

Restaurants were divided by four types:

American

['Fried Chicken Joint', 'Diner', 'American Restaurant', 'Wings Joint', 'Bagel Shop', 'Burger Joint', Australian Restaurant', 'Southern / Soul Food Restaurant', 'Cajun / Creole Restaurant', 'BBQ Joint'].

Mexican

['Burrito Place', 'Mexican Restaurant', 'Taco Place'].

Italian

['Pizza Place', 'Italian Restaurant', 'Veneto Restaurant']

Other

Everything else

Note: Australian restaurants were attributed to American cuisine as they often serve similair types of dishes: burgers, steaks, lobsters.

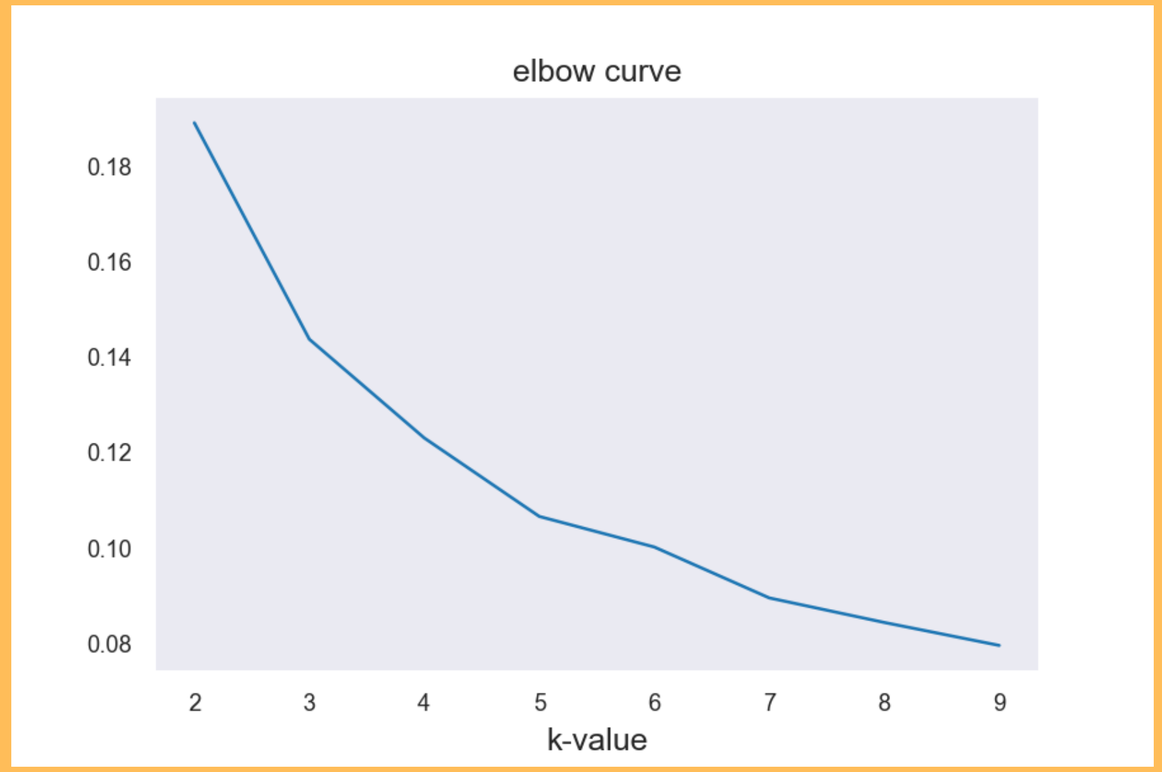
K-means clustering

Why?

Unlabeled data =>
unsupervised clustering

Which k?

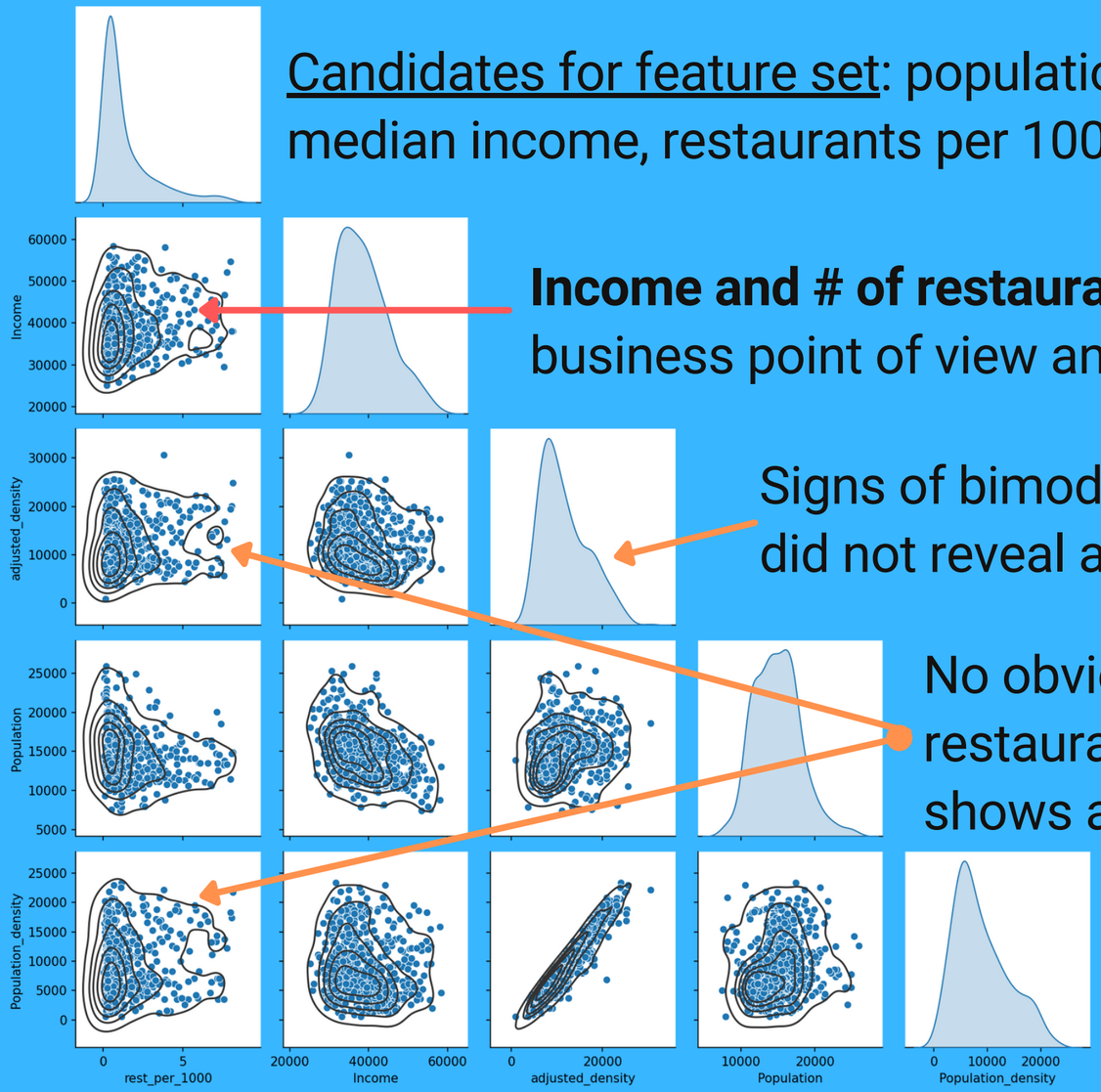
Elbow method (euclidian distance) - see next slide about featureset



Elbow curve is smooth - two possible k values: 3 and 5

Further analysis showed that having higher k did not yield better results or helped decision making. k=3 is selected for analysis.

Features selection



Candidates for feature set: population, population density, adjusted population density, median income, restaurants per 1000 people.

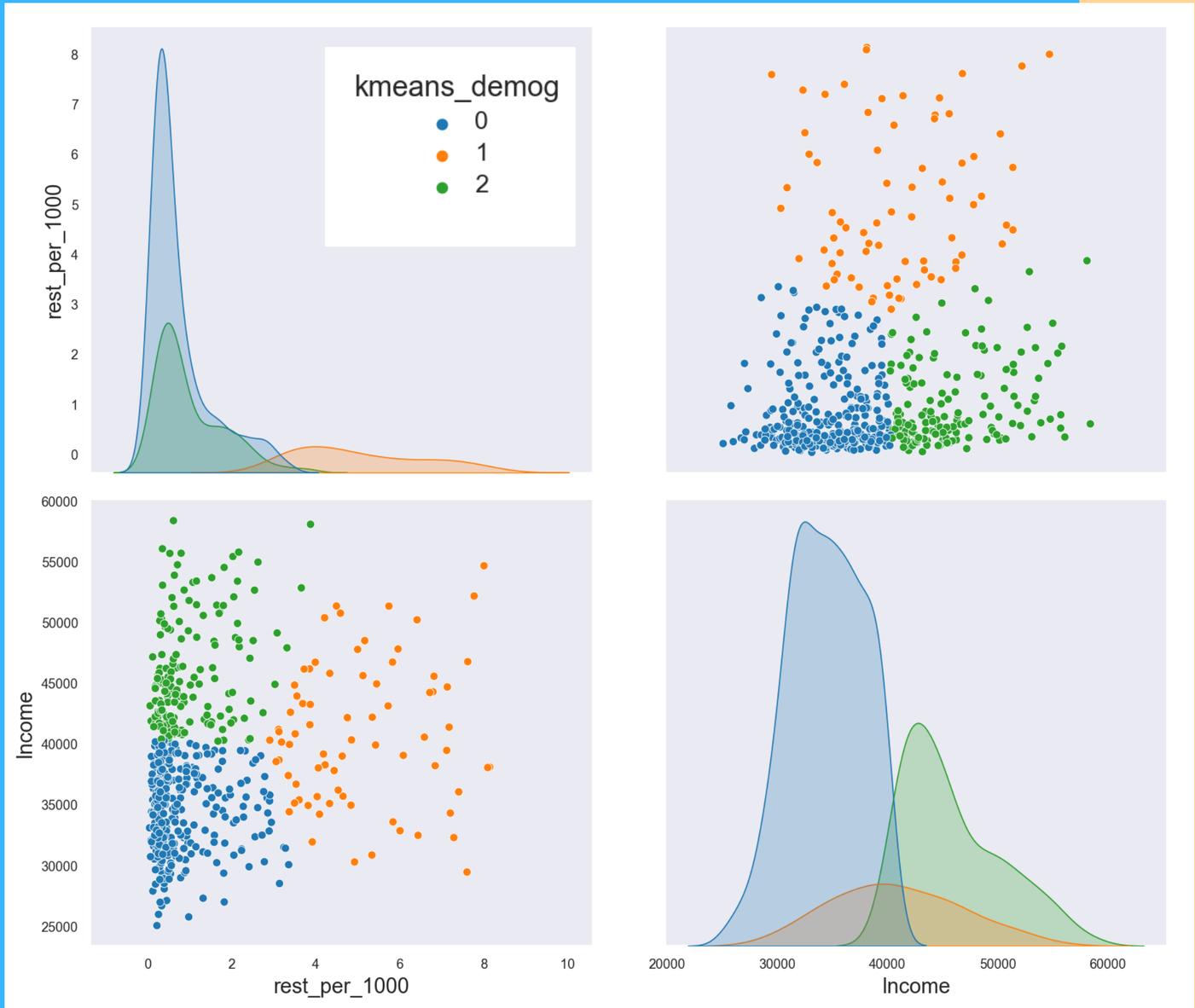
Income and # of restaurants per 1000 people are important metrics from business point of view and are selected for clustering.

Signs of bimodal distribution in population density. Deeper analysis did not reveal any hidden patterns.

No obvious connection between population density and restaurant number. Neither adjusted population density shows any connection.

Other variables will not bring extra information (was confirmed during analysis stage).

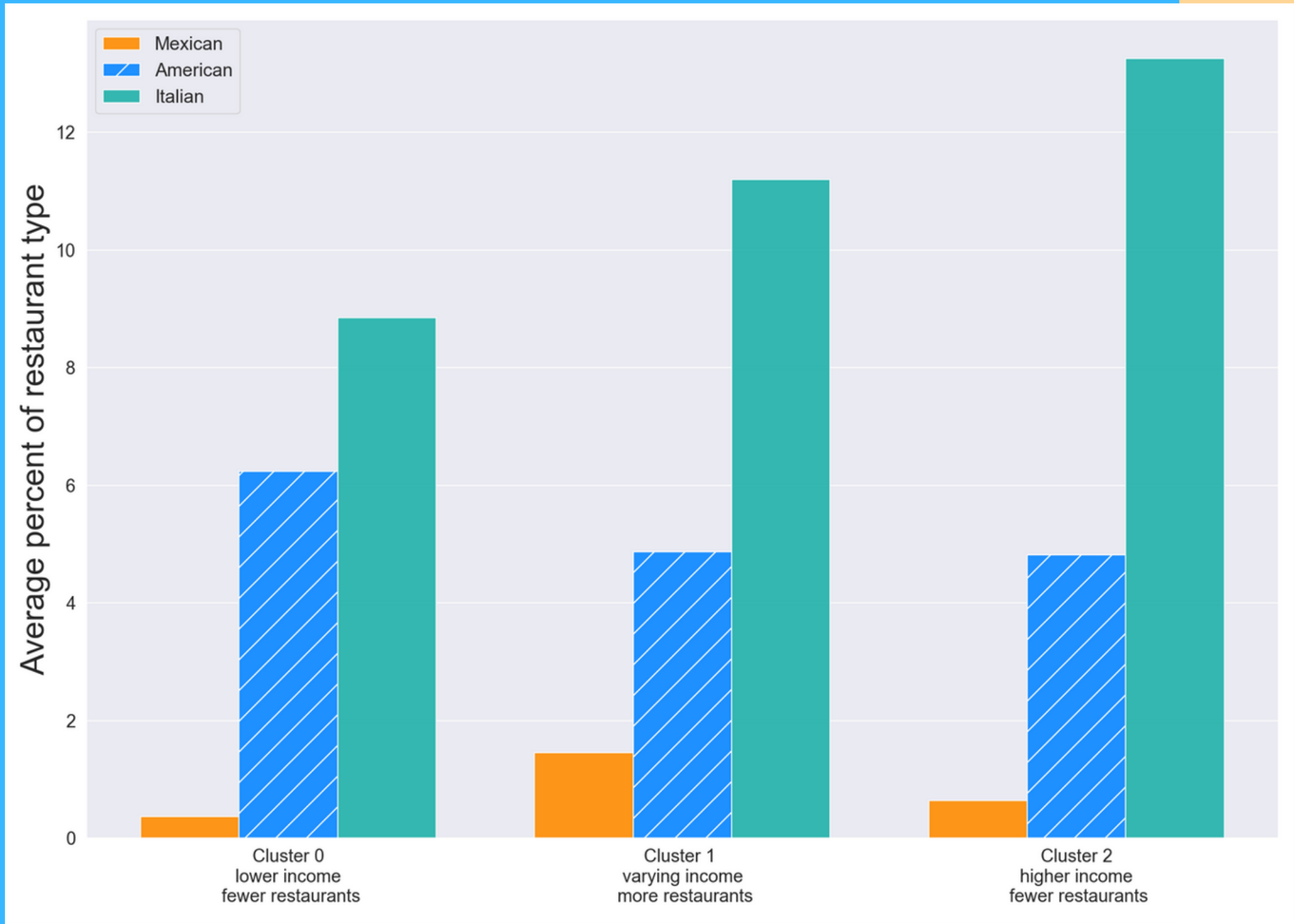
Clustering results



- **Cluster 0:** generally lower median income, relatively fewer restaurants per 1000 people. 58% of wards belong to this cluster.
- **Cluster 1:** varying income, more restaurants per 1000 people. These are wards with lively dining life. 13% of wards belong to this cluster.
- **Cluster 2:** generally higher income, fewer restaurants per 1000 people. 29% of wards make up this cluster.

Distribution of American, Mexican and Italian restaurants in clusters.

First recommendations for each ward clusters, including $ range:



- **Cluster 0:** (lower income, fewer restaurants per 1000) - Mexican; lower price tag.
- **Cluster 1:** (lively restaurant life regardless of income) - Mexican, American; both lower and higher price tag possible.
- **Cluster 2:** (high income, fewer restaurants per 1000) - Mexican, American, higher price tag possible.

Mexican is minority - low competition. American: 4-6% of food places, medium competition. Italian: 9-13% - high competition.

Expected population growth as a factor

Top 10 wards in each cluster:

	Ward_Name	Borough	Growth
0	Tokyngton	Brent	41.209948
1	Southall Broadway	Ealing	36.752350
2	Upper Edmonton	Enfield	34.864208
3	East Acton	Ealing	33.780928
4	River	Barking and Dagenham	31.392330
5	Royal Docks	Newham	30.947766
6	Woolwich Riverside	Greenwich	29.156801
7	Shepherd's Bush Green	Hammersmith and Fulham	27.902053
8	Livesey	Southwark	27.450753
9	South Hornchurch	Havering	27.143833

	Ward_Name	Borough	Growth
0	College Park and Old Oak	Hammersmith and Fulham	74.080560
1	Hammersmith Broadway	Hammersmith and Fulham	22.425598
2	Grove	Kingston upon Thames	14.137752
3	Wick	Hackney	11.865915
4	Thamesmead Moorings	Greenwich	11.692020
5	Tolworth and Hook Rise	Kingston upon Thames	11.020060
6	Abingdon	Kensington and Chelsea	10.816220
7	Petts Wood and Knoll	Bromley	8.571807
8	Hanworth Park	Hounslow	8.273296
9	St George's	Islington	8.026685

	Ward_Name	Borough	Growth
0	Peninsula	Greenwich	33.959538
1	Golders Green	Barnet	26.228589
2	Uxbridge North	Hillingdon	17.434613
3	Merton Park	Merton	16.453983
4	Bromley Town	Bromley	15.143942
5	Mill Hill	Barnet	13.643449
6	Fairfield	Wandsworth	12.936550
7	Wimbledon Park	Merton	10.707833
8	Coulsdon West	Croydon	10.362507
9	Bishop's	Lambeth	9.698699

Cluster 0:

- favorable expected population growth;
- #1 at 41%; #10 at 27% (still high) which is also high.
- #2 and #4 are from the same borough

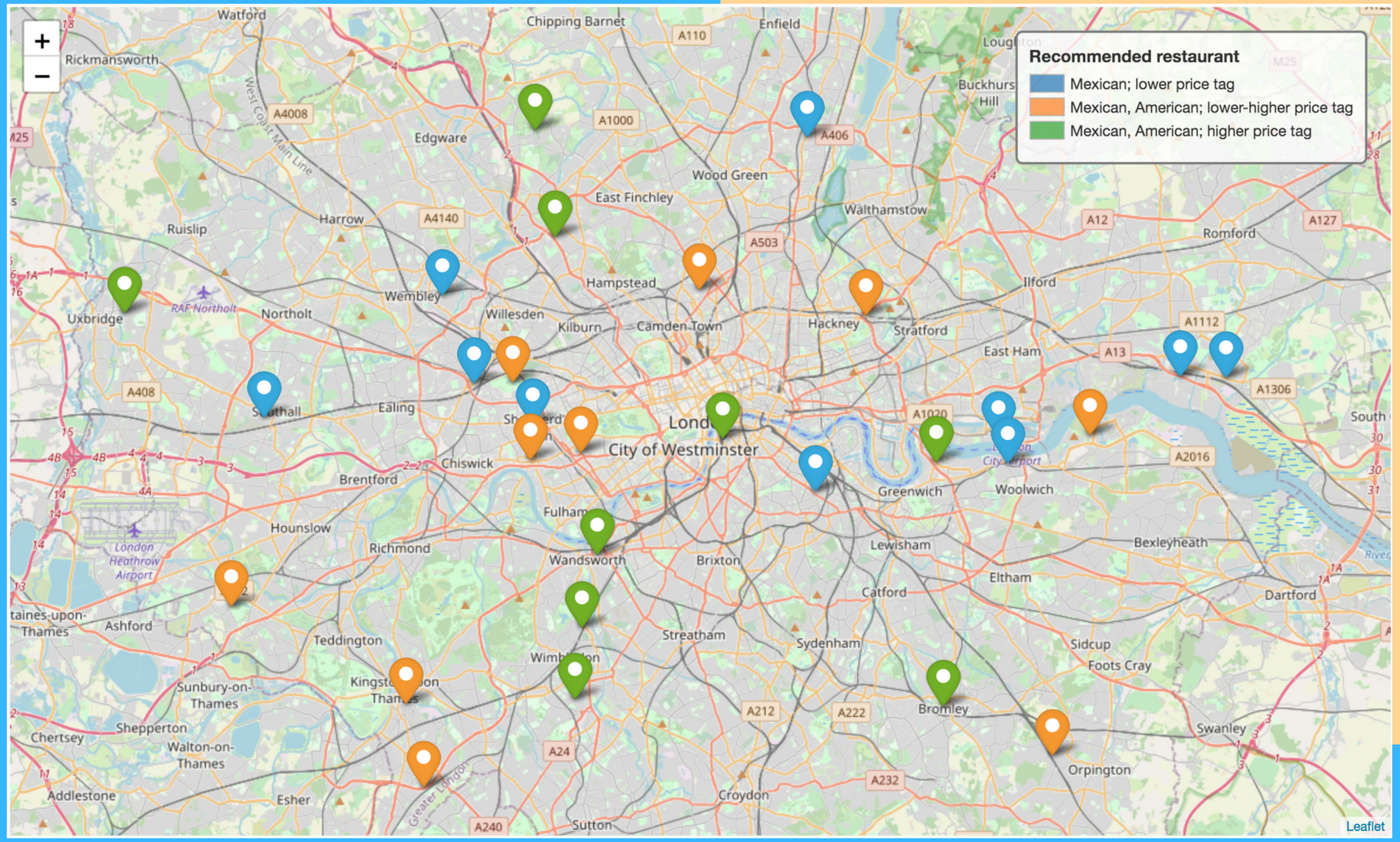
Cluster 1:

- expected growth range from 8 to 72%;
- #1 exceed second best wards by 50%.
- Top two wards are from the same borough - excellent candidates

Cluster 2:

- expected growth range from 10 to 34%
- #3 and #7 are from the same borough Merton, suggesting that it is a good candidate.

Top 3 wards to open new restaurant on map.



Conclusions

1. **Three types of wards were found based on the median income and # of restaurants per 1000 people.**

2. **Preferred cuisine was selected based on the current distribution of American, Mexican and Italian cuisine for each type of ward.**

3. **Preferred menu pricing strategy (low or high prices) were recommended.**

4. **Top 10 wards to open restaurant for each of the three ward types were selected based on the projected population growth for the next 5 years.**

References

1.**Land Area and Population Density, Ward and Borough.** https://data.london.gov.uk/dataset/land-area-and-population-density-ward-and-borough
Author: Greater London Authority.
Licence: UK Open Government Licence (OGL v3)
2. **Statistical GIS Boundary Files for London.** https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london
Author: Greater London Authority
Licence: UK Open Government Licence (OGL v2)
3.**Household Income Estimates for Small Areas.** https://data.london.gov.uk/dataset/household-income-estimates-small-areas
Author: Greater London Authority
Licence: UK Open Government Licence (OGL v2)
4. **Access to Public Open Space and Nature by Ward**https://data.london.gov.uk/dataset/access-public-open-space-and-nature-ward
Author: Greenspace Information for Greater London
Licence: UK Open Government Licence (OGL v2)