



# Opening an English Restaurant in London Area - Analysis

Capstone Project

24Slides

# Problem Statement

When it comes to restaurants London is one of the most attractive and competitive place for restaurants. With high population (9 Mil) and large numbers of tourists (30 Mil per year), the capital of England host approximately 18,000 restaurants. Considering the high cost of opening a restaurant (500-1,000K\$ the need of having some data driven decision seems imperial. The location of the restaurant will impact the success rate along with the cost allocated. For this we will try to identify the best location to open an English restaurant by trying to segment the London areas and see if there are some similitudes between them. Once they are identified we can start identifying the best cohort and choosing the optimal location where a restaurant will work.

# DATA Collection



1. The data was selected Kaggle raw data containing info related each area and can be seen in the Table on the right
2. The data was enhanced with venues information collected via Foursquare API

#	Feature Name	#	Feature Name	#	Feature Name
1	New code	29	Employment rate (%) (2015)	57	Rented from Local Authority or Housing Association, (2014) %
2	Area name	30	Male employment rate (2015)	58	Rented from Private landlord, (2014) %
3	Inner/ Outer London	31	Female employment rate (2015)	59	% of area that is Greenspace, 2005
4	GLA Population Estimate 2016	32	Unemployment rate (2015)	60	Total carbon emissions (2013)
5	GLA Household Estimate 2016	33	Youth Unemployment (claimant) rate 18-24 (Dec-14)	61	Household Waste Recycling Rate, 2014/15
6	Inland Area (Hectares)	34	Proportion of 16-18 year olds who are NEET (%) (2014)	62	Number of cars, (2011 Census)
7	Population density (per hectare) 2016	35	Proportion of the working-age population who claim ou	63	Number of cars per household, (2011 Census)
8	Average Age, 2016	36	% working-age with a disability (2015)	64	% of adults who cycle at least once per month, 2013/14
9	Proportion of population aged 0-15, 2016	37	Proportion of working age people with no qualifications	65	Average Public Transport Accessibility score, 2014
10	Proportion of population of working-age, 2016	38	Proportion of working age with degree or equivalent an	66	Achievement of 5 or more A*- C grades at GCSE or equi
11	Proportion of population aged 65 and over, 2	39	Gross Annual Pay, (2015)	67	Rates of Children Looked After (2015)
12	Net internal migration (2014)	40	Gross Annual Pay - Male (2015)	68	% of pupils whose first language is not English (2015)
13	Net international migration (2014)	41	Gross Annual Pay - Female (2015)	69	% children living in out-of-work households (2014)
14	Net natural change (2014)	42	Modelled Household median income estimates 2012/13	70	Male life expectancy, (2012-14)
15	% of resident population born abroad (2014)	43	% adults that volunteered in past 12 months (2010/11 t	71	Female life expectancy, (2012-14)
16	Largest migrant population by country of birth	44	Number of jobs by workplace (2014)	72	Teenage conception rate (2014)
17	% of largest migrant population (2011)	45	% of employment that is in public sector (2014)	73	Life satisfaction score 2011-14 (out of 10)
18	Second largest migrant population by countr	46	Jobs Density, 2014	74	Worthwhileness score 2011-14 (out of 10)
19	% of second largest migrant population (201	47	Number of active businesses, 2014	75	Happiness score 2011-14 (out of 10)
20	Third largest migrant population by country o	48	Two-year business survival rates (started in 2012)	76	Anxiety score 2011-14 (out of 10)
21	% of third largest migrant population (2011)	49	Crime rates per thousand population 2014/15	77	Childhood Obesity Prevalance (%) 2014/15
22	% of population from BAME groups (2016)	50	Fires per thousand population (2014)	78	People aged 17+ with diabetes (%)
23	% people aged 3+ whose main language is n	51	Ambulance incidents per hundred population (2014)	79	Mortality rate from causes considered preventable 2012/13
24	Overseas nationals entering the UK (NINo), (2	52	Median House Price, 2014	80	Political control in council
25	New migrant (NINo) rates, (2014/15)	53	Average Band D Council Tax charge (£), 2015/16	81	Proportion of seats won by Conservatives in 2014 electio
26	Largest migrant population arrived during 20	54	New Homes (net) 2014/15 (provisional)	82	Proportion of seats won by Labour in 2014 election
27	Second largest migrant population arrived du	55	Homes Owned outright, (2014) %	83	Proportion of seats won by Lib Dems in 2014 election
28	Third largest migrant population arrived durin	56	Being bought with mortgage or loan, (2014) %	84	Turnout at 2014 local elections



# DATA Cleaning

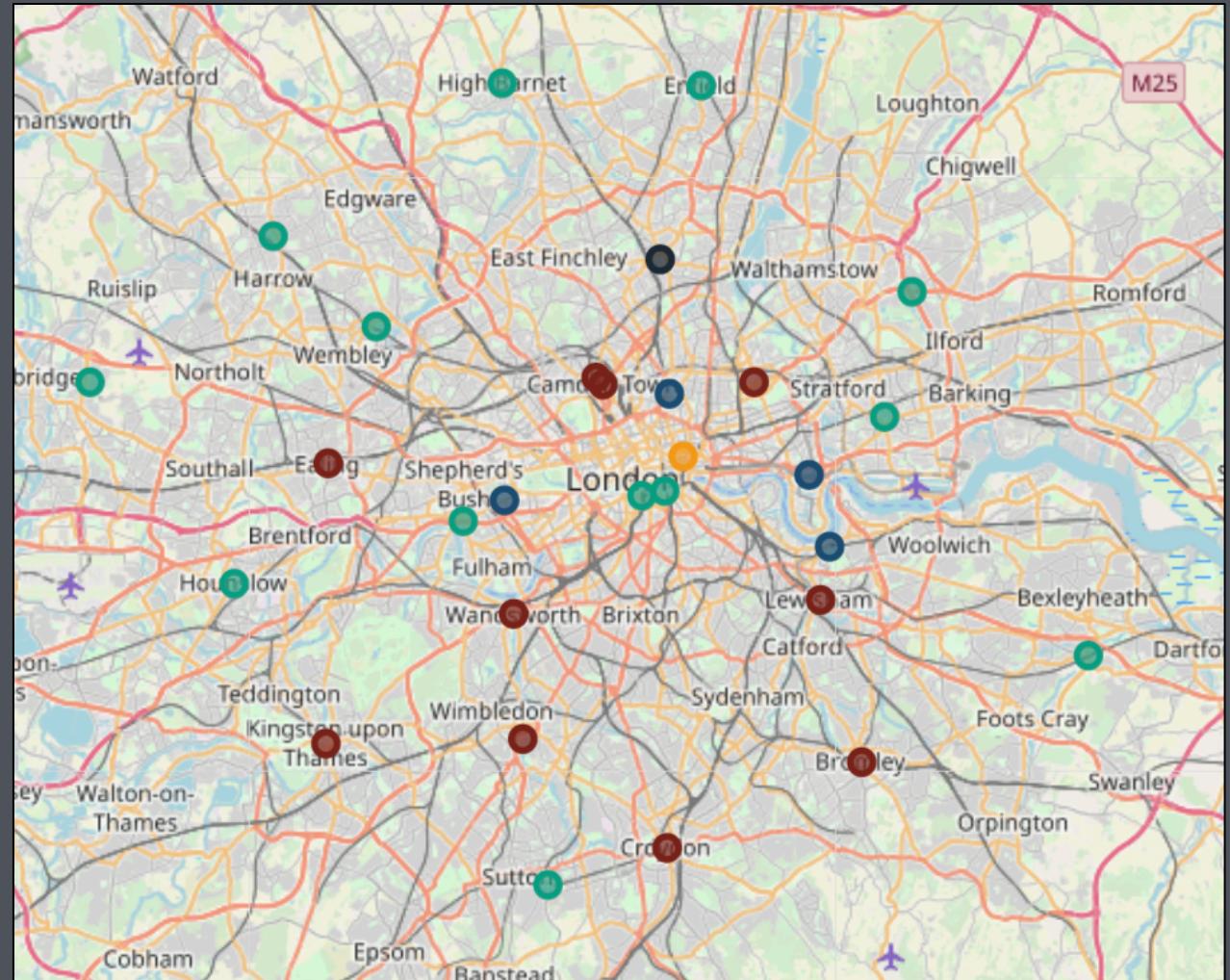
---

1. The data was clean by changing the formats of the features and eliminating the unnecessary symbols used like (commas or pound sign )
2. Some features – that are unnecessarily for restaurants analysis were dropped
3. Data related restaurant category, restaurant location were added to the data set.
4. Restaurants ranking collection was unsuccessful both via API and via web scrapping techniques
5. Feature selection was done in several ways
  1. Features intercorrelation. Features that had two higher correlation between them were eliminated as duplicated
  2. Features correlated with number of restaurants were grouped and used in a separate classification
  3. Trial and error, with multiple scenarios. The most relevant scenarios and results are showed in the next slide



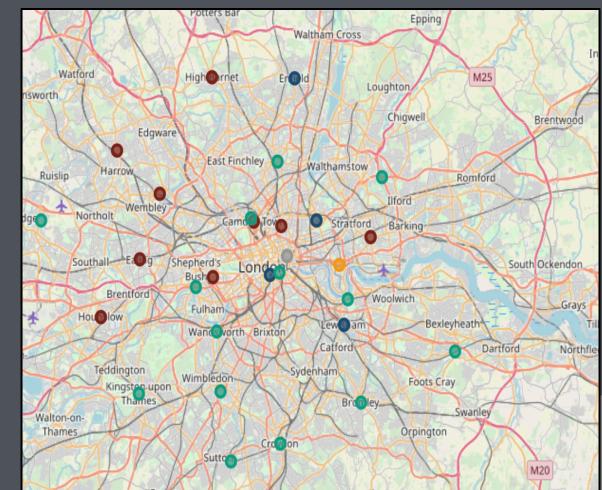
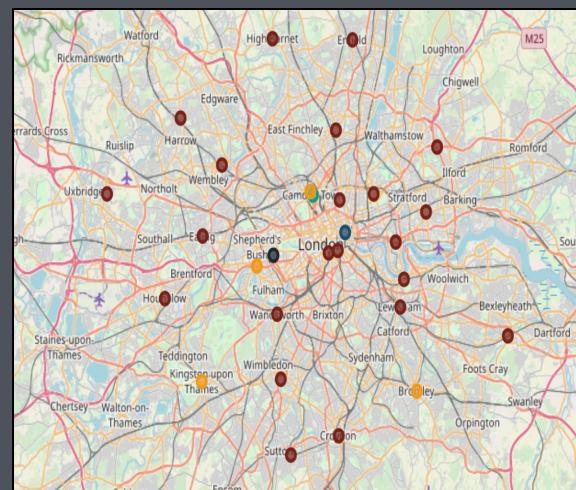
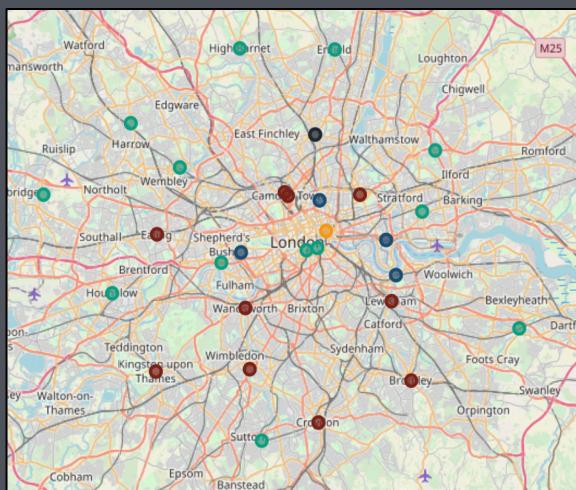
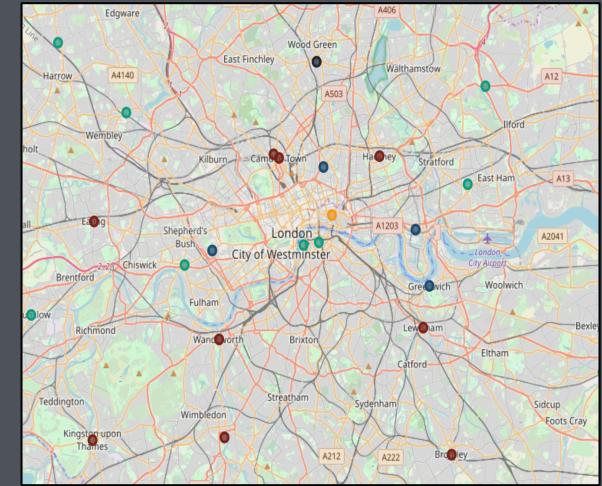
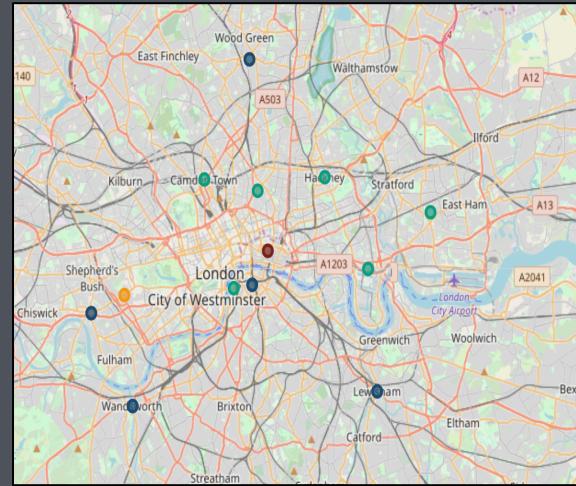
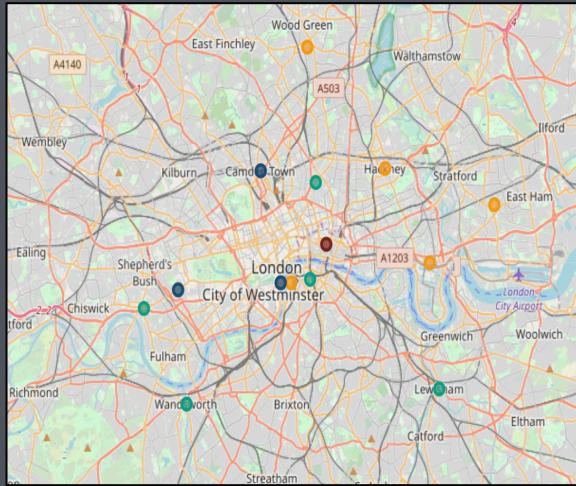
# Methodology

1. The ML algorithm selected was K-MEAN due to its suitability for our set of data
  2. The end goal was to define a set of clusters that can be used as support for future data analysis that will allow a recommendation for the best area of investment



# Results

1. Top 6 scenarios resulted based on different features selected are displayed on the right
2. The details of features selected are described in the report



# Results

1. Additional Results are the group of top 10 venues, restaurants category for each area, along with the most preeminent migrants category (Migrants were not used the final conclusions)
2. Top factors based segmented based on resulted clusters (Table on the right)

	0	1	2	3	4
<b>Population</b>	274803.00	282068.300	240129.2500	8548.00	289986.769231
<b>Population_Density</b>	92.70	68.330	122.7000	28.90	66.176923
<b>Working_Population</b>	90.70	87.690	90.1000	90.60	87.900000
<b>Youth_Population</b>	19.50	18.540	18.3250	27.20	20.446154
<b>Elderly_Population</b>	9.30	12.310	9.9000	9.40	12.100000
<b>Born_Abroad</b>	39.60	31.960	40.5500	0.00	38.607692
<b>Employment</b>	71.30	74.600	70.8250	64.60	73.100000
<b>Median_Income_H</b>	45860.00	52472.000	65347.5000	99390.00	46666.923077
<b>Job_Density</b>	0.48	0.793	1.1425	84.60	0.760769
<b>Active_Business</b>	11875.00	14807.500	14466.2500	19250.00	13952.692308
<b>Suvirval_Rate_2Years</b>	71.00	74.600	69.5000	63.00	72.000000
<b>Life_Satisfcation</b>	7.24	7.266	7.3000	6.59	7.330769
<b>Anxiety</b>	3.18	3.465	3.3650	5.57	3.150769
<b>Restaurants#</b>	10.00	14.700	18.7500	33.00	10.153846
<b>Bars#</b>	2.00	6.800	6.7500	7.00	4.307692
<b>Coffee#</b>	5.00	7.700	5.2500	12.00	4.230769
<b>Extra#</b>	15.00	26.300	31.7500	48.00	16.000000



# Results Clusters 0

Areas covered: Haringey

Properties:

- Low on restaurants

- Relative high population density but low job density

- Lower household income

Conclusion

- Not recommended to invest in restaurants in this area,  
seems more like an industrial area or a quiet  
residential area

- Does not seem the most attractive area for  
restaurant's landscape

	Rest_Tpye	Count
0	Indian Restaurant	1
1	Mediterranean Restaurant	1
2	Italian Restaurant	1
3	Polish Restaurant	1
4	African Restaurant	1
5	Middle Eastern Restaurant	1
6	Fast Food Restaurant	1
7	Bulgarian Restaurant	1
8	Falafel Restaurant	1
9	Turkish Restaurant	1



# Results Clusters 1

Areas covered:

Bromley, Camden, Croydon, Ealing, Hackney, Havering,  
Kingston upon Thames, Lewisham, Merton, Wandsworth

Properties:

- o Relative low density
- o Relative higher elderly population
- o High employment rate
- o Better than cluster 1 median earnings
- o Highest business success rate
- o Decent number of restaurants

Conclusion

- o Seems to be a stable attractive area
- o But does not seem to be an attractive/successful place for English restaurants

#	Rest Type	Count	#	Rest Type	Count
1	Burger Joint	15	16	French Restaurant	4
2	Italian Restaurant	13	17	Fish & Chips Shop	3
3	Restaurant	8	18	Ramen Restaurant	2
4	Asian Restaurant	8	19	American Restaurant	2
5	Vegetarian / Vegan Restaurant	7	20	Mexican Restaurant	2
6	Pizza Place	7	21	African Restaurant	2
7	Portuguese Restaurant	7	22	Malay Restaurant	2
8	Sandwich Place	7	23	Spanish Restaurant	2
9	Fast Food Restaurant	7	24	Middle Eastern Restaurant	2
10	Thai Restaurant	6	25	Falafel Restaurant	2
11	Vietnamese Restaurant	6	26	English Restaurant	2
12	Sushi Restaurant	6	27	New American Restaurant	1
13	Caribbean Restaurant	5	28	Japanese Restaurant	1
14	Turkish Restaurant	4	29	Latin American Restaurant	1
15	Indian Restaurant	4	30	Cajun / Creole Restaurant	1

# Results Clusters 2

Areas covered: '*Greenwich*', '*Islington*', '*Kensington and Chelsea*', '*Tower Hamlets*'

## Properties:

- Highest density populated of all districts
- Highest diversity
- Highest job density
- Lower elderly population suggesting a more dynamic and crowded area
- Lower survival rate suggesting a competitive landscape
- Considerable number of restaurants
- More interest shown for English restaurants
- Second highest income area

## Conclusion

- Seems to be an attractive area with good dynamics
- Seems to be the best place for a English restaurant

#	Rest Type	Count	#	Rest Type	Count
1	Burger Joint	11	16	Portuguese Restaurant	2
2	Sushi Restaurant	6	17	Sandwich Place	2
3	Pizza Place	5	18	Dumpling Restaurant	1
4	Japanese Restaurant	5	19	Ramen Restaurant	1
5	Italian Restaurant	5	20	Filipino Restaurant	1
6	English Restaurant	4	21	Persian Restaurant	1
7	Mediterranean Restaurant	4	22	Afghan Restaurant	1
8	French Restaurant	3	23	Vietnamese Restaurant	1
9	Restaurant	3	24	Kebab Restaurant	1
10	Mexican Restaurant	3	25	Austrian Restaurant	1
11	Steakhouse	3	26	Latin American Restaurant	1
12	Middle Eastern Restaurant	2	27	Thai Restaurant	1
13	Modern European Restaurant	2	28	Chinese Restaurant	1
14	Spanish Restaurant	2			
15	Indian Restaurant	2			



# Results Clusters 3

Areas covered: City of London

Properties:

Low density suggesting a relative expensive area with limited housing

Covers only the city center of London and seems to be the most peculiar cluster

Low employment rate

Lower elderly population suggesting a more dynamic and crowded area

Very high job density

Competitive business landscape with the lowest success rate of business

High bars and restaurants density suggesting a very promising area

High "Extra" attraction suggest a very tourist area

By far the highest median income area

Conclusion

Probably the most attractive area to invest with high potential but a risky one non the less.

Expected to have high competition, cost of operation and initial cost

English restaurants seems not be very prominent, but nevertheless a good restaurant in this area can be very successful

#	Rest_Type	Count	#	Rest_Type	Count
1	Steakhouse	3	10	Asian Restaurant	2
2	Restaurant	3	11	English Restaurant	1
3	French Restaurant	3	12	Mexican Restaurant	1
4	Italian Restaurant	3	13	Falafel Restaurant	1
5	Seafood Restaurant	3	14	Scandinavian Restaurant	1
6	Modern European Restaurant	2	15	Udon Restaurant	1
7	Sushi Restaurant	2	16	Latin American Restaurant	1
8	Indian Restaurant	2	17	New American Restaurant	1
9	Vietnamese Restaurant	2	18	Pizza Place	1

# Results Clusters 4

Areas covered: *Barnet, Bexley, Brent, Enfield, Hammersmith and Fulham, Harrow, Hillingdon, Hounslow, Lambeth, Newham, Redbridge, Southwark, Sutton*

## Properties

Median population density with a relative lower number of available restaurants

Relative low median income and lower job density than the rest of clusters

Fairly stable business continuity

## Conclusion

Can be an interesting opportunity for investment for a small restaurant with no high income expectations

Overall not the most attractive cluster to invest

#	Rest Type	Count	#	Rest Type	Count
1	Sandwich Place	16	18	Tapas Restaurant	2
2	Fast Food Restaurant	15	19	Sushi Restaurant	2
3	Italian Restaurant	11	20	Vegetarian / Vegan Restaurant	2
4	Indian Restaurant	11	21	Portuguese Restaurant	2
5	Chinese Restaurant	8	22	Japanese Restaurant	1
6	Restaurant	8	23	Romanian Restaurant	1
7	Pizza Place	7	24	Dim Sum Restaurant	1
8	Burger Joint	7	25	French Restaurant	1
9	Turkish Restaurant	6	26	Latin American Restaurant	1
10	English Restaurant	3	27	Afghan Restaurant	1
11	Fish & Chips Shop	3	28	Israeli Restaurant	1
12	Steakhouse	3	29	Greek Restaurant	1
13	Asian Restaurant	3	30	Vietnamese Restaurant	1
14	Korean Restaurant	3	31	Argentinian Restaurant	1
15	Ramen Restaurant	2	32	Modern European Restaurant	1
16	Eastern European Restaurant	2	33	Mexican Restaurant	1
17	Thai Restaurant	2	34	Spanish Restaurant	1



# Conclusions

---

1. There seems to be a lack of English restaurants resulting from the data collected using Foursquare, placing the English restaurants category outside of top ten categories after Italian, Burgers joins, Asian restaurants, Indian Restaurants and others
2. Surprisingly, the number of French, Portuguese or Vietnamese restaurants exceeds the English ones (this might be due to the limited data collection), nevertheless the 1400+ venues collected should be a significant sampler for making the above assumptions
3. Multiple analysis scenarios were performed where different features were selected, and the results are presented in the table xx above. Out of these scenarios one that seems to have the better cluster distribution was selected. Scenario 4 was selected
4. The analysis concluded above allow us to segregate the London restaurants landscape into 5 clusters based on restaurants types and numbers and the data was completed with other significant information like: population density, job density, household income, working population, business survival rate, life satisfaction, anxiety and coffee shops and bars numbers



# Conclusions

---

5. Thought the analysis the City of London itself was always placed in a separate individual cluster making us realize the uniqueness of this area
6. However the most promising scenario are Clusters 2 ('Greenwich', 'Islington', 'Kensington and Chelsea', 'Tower Hamlets') and Cluster 3 (City of London) showing high potential for new restaurants. Out of these two scenario the safest bet seems to be on Cluster 2 because it seems less riskier and challenging than Cluster 3. See detailed analysis and conclusions on the above chapters

# Future

---



1. Extending the data set that will require another level of account in Foursquare. The ability to add more data should results in better results
2. Add ranking for restaurants. The current analysis was performed without ranking due to limitations:
3. imposed by the Foursquare account (providing a limited number of rankings)
4. Inability to use web scraping techniques due to protection from Foursquare
5. Changing the account type should allow the collection of rating, but it comes with additional costs
6. The current analysis concluded that Cluster 2 is the best candidate, however the investigation can continue inside Cluster 2. The same type of analysis can be performed in order to detect which particular section of this Cluster is most suited for opening a restaurant. Adding cost of rent in the equation might also increase the output of the analysis



THANK YOU