Applied Data Science Capstone Project
The Battle of Neighborhoods

Detailed research
of London neighborhoods
for the Art Dealer Company “NNN”

Author: Irina Chernova

01

Problem Background&Description

Is it easy to be an Art Dealer in London?

The very ambitious art dealer* company "NNN" is going to open office in London.
They going to study boroughs and neighborhoods of the city in order to find out which is the best for their office.

Main requirements :

- Art centers and commercial galleries nearby
- Moderate competition
- Café & restaurants, and other venues
- To be in touch with artistic life of the city

Result :

- List of recommendations

** An art dealer is a person or company that buys and sells works of art, so location of art centers and galleries, artistic workspaces etc is very important if we want to achieve the goal, because such facilities provide arts space, visual art gallery space, museum facilities where new artworks exhibit and promote.*

02

IBM Coursera

The Data we need

	borough_name	cnt	latitude	longitude
0	Camden	34	51.528	-0.137
1	City and County of the City of London	5	51.516	-0.092
2	City of Westminster	119	51.512	-0.145
3	Hackney	27	51.534	-0.075
4	Hammersmith and Fulham	6	51.490	-0.212
5	Islington	17	51.531	-0.104
6	Kensington and Chelsea	26	51.499	-0.185
7	Lambeth	15	51.483	-0.118
8	Southwark	20	51.491	-0.080
9	Tower Hamlets	26	51.525	-0.061
10	Wandsworth	9	51.458	-0.174

- **Open data provided by the Government of UK**
- **Common geographical data**
- **Foursquare location data and venues information**

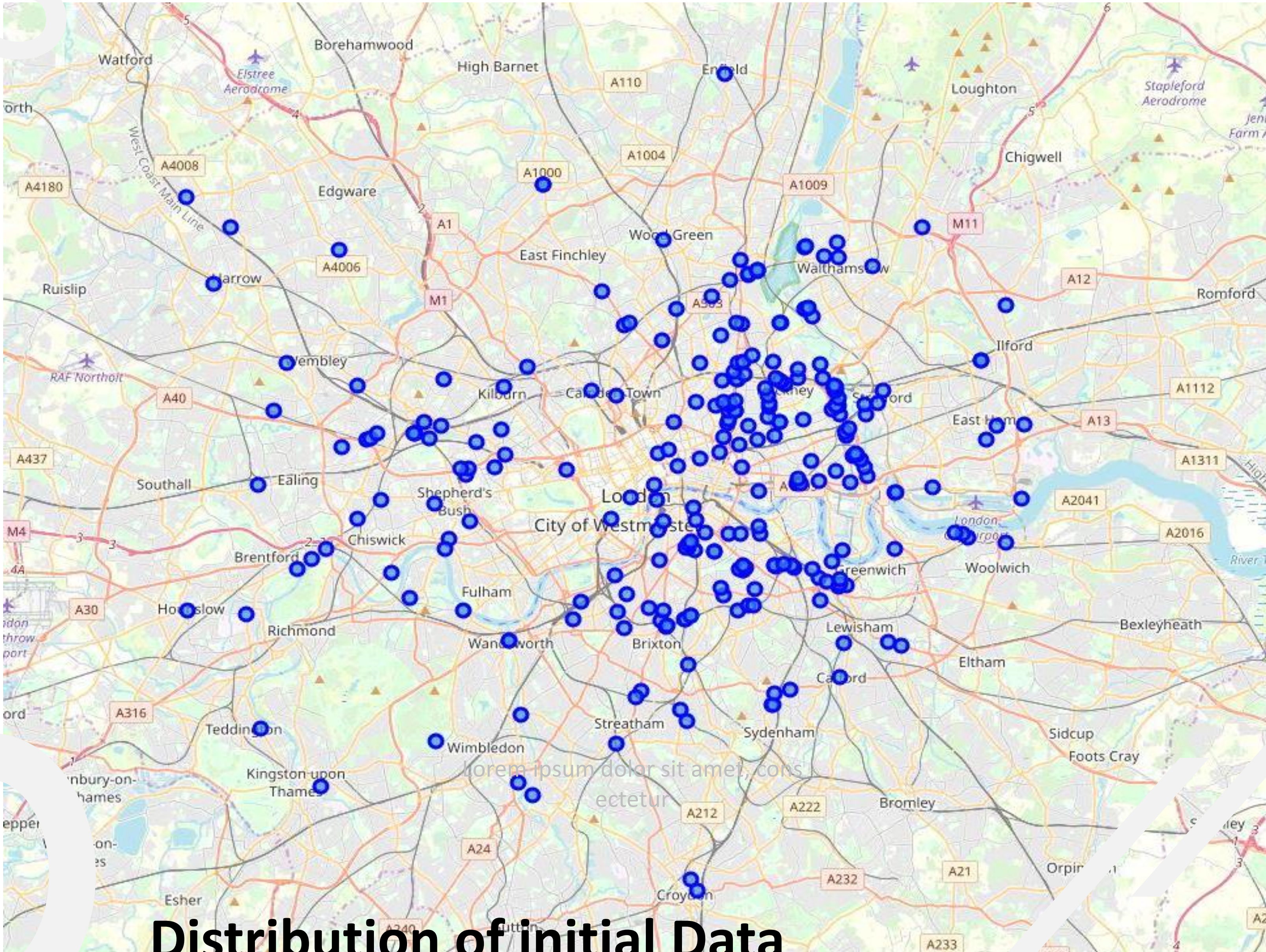
The final result set is a combination of data from all the sources described above.
It includes 11 boroughs which should be explored using analytical approach

All the data were loaded into pandas dataframe and processed according the given requirements

03

Data visualization using folium

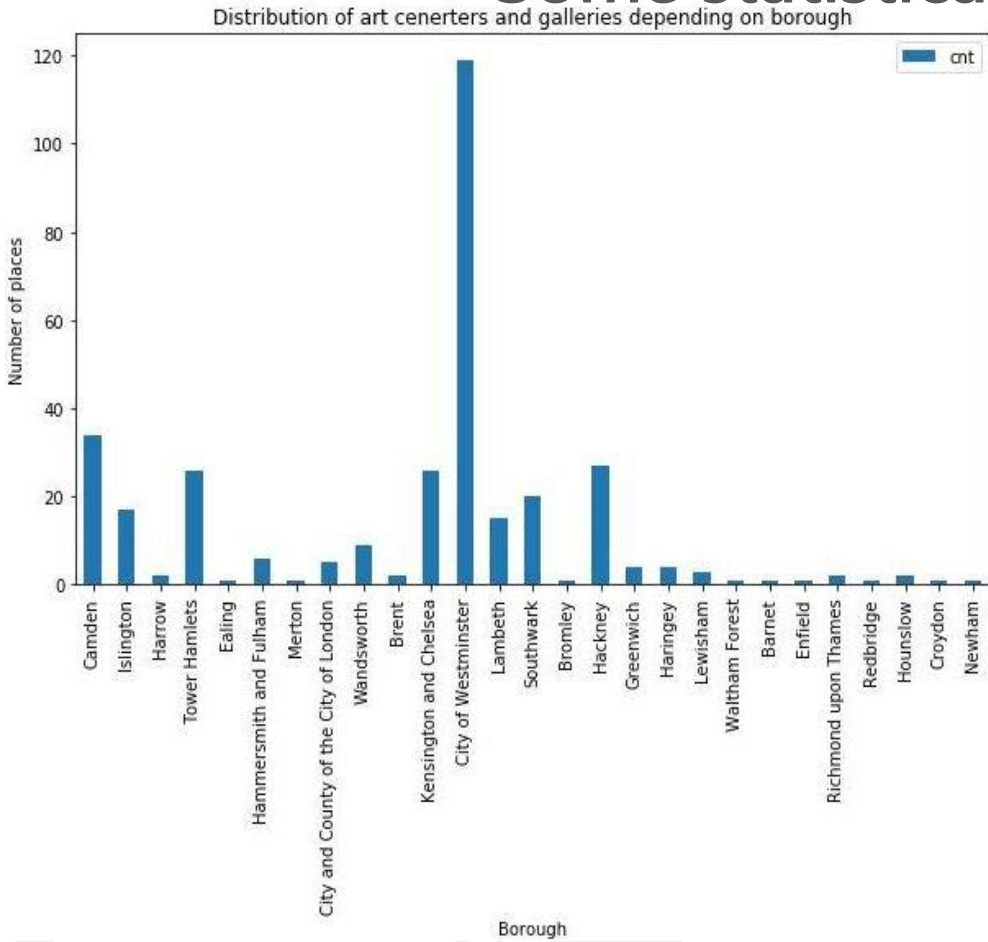
IBM Coursera



lorem ipsum dolor sit amet, cons ectetur

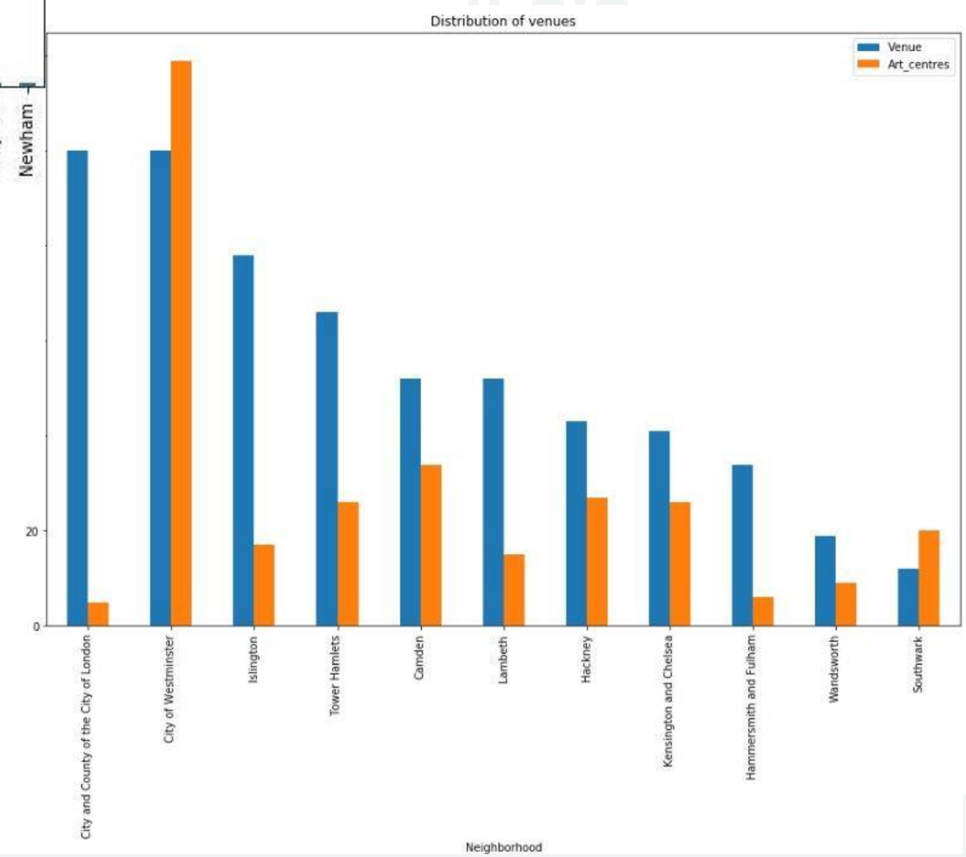
Distribution of initial Data

Some statistical calculations



Comparison of data from resulting dataset with the data, provided by Foursquare API

Here depicted the number of art centers and galleries compared with the total number of different venues provided by Foursquare API depending on borough



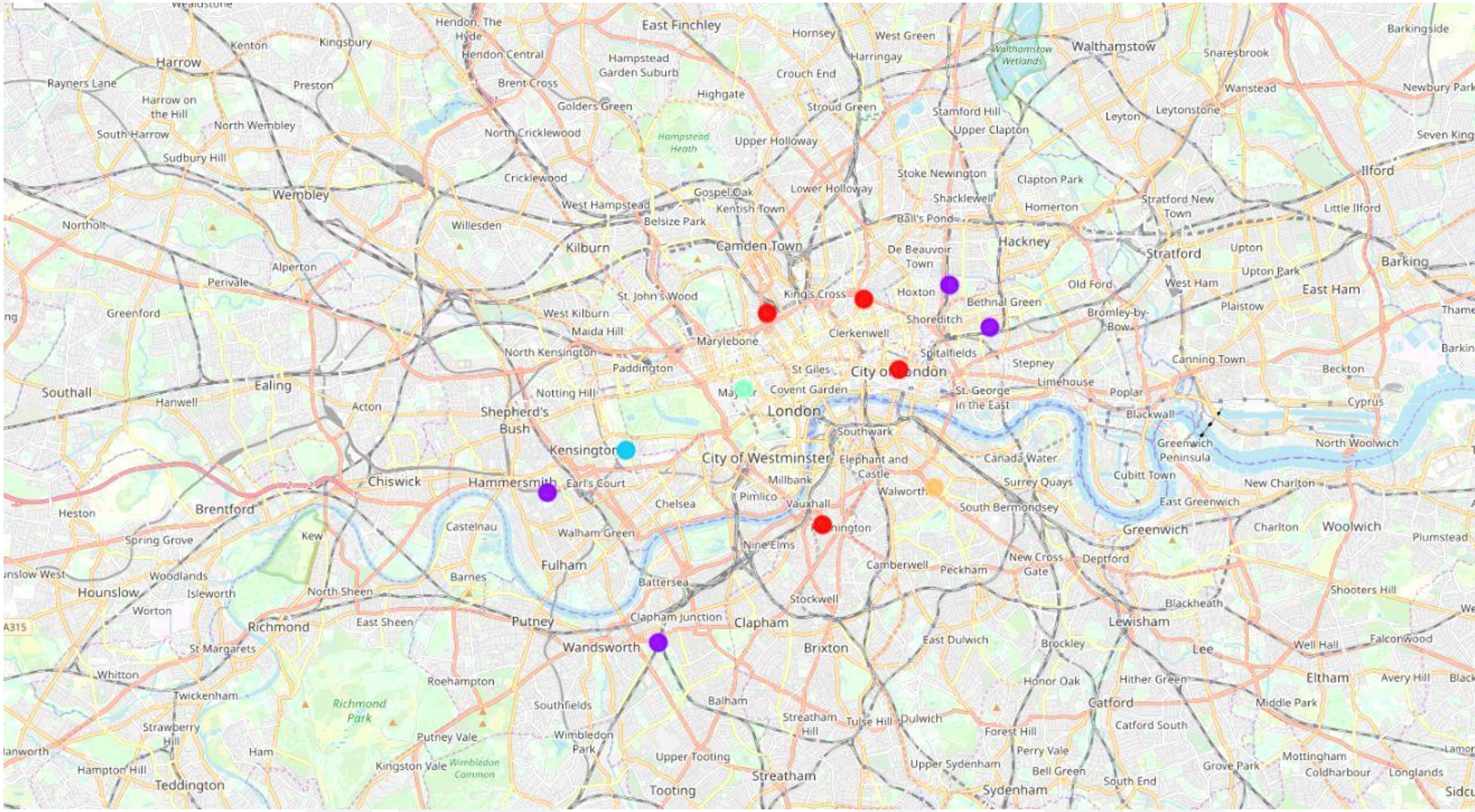
As we can see distribution of art centers and commercial galleries is quite irregular. But customer hardly be interested in boroughs where number of art centers less than 4, so such boroughs were exclude from the resulting dataset

05

IBM Coursera

Clusters

Cluster map built by K-means clustering algorithm



Methodology:

K-means clustering is one of the simplest and very effective unsupervised machine learning algorithms.

A cluster refers to a collection of data points aggregated together because of certain similarities . Target number was chosen as k = 5

Results

Five clusters provided by K- means algorithm

	borough_name	cnt	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	Camden	34	Coffee Shop	Gym / Fitness Center	Theater	Beer Bar	Indian Restaurant	Plaza	Pub	Café	Science Museum	Sushi Restaurant
1	City and County of the City of London	5	Coffee Shop	Italian Restaurant	Seafood Restaurant	Art Gallery	Sushi Restaurant	Restaurant	Vietnamese Restaurant	Steakhouse	History Museum	Roof Deck
5	Islington	17	Pub	Coffee Shop	Café	Sandwich Place	Gym / Fitness Center	Theater	Sushi Restaurant	French Restaurant	Nightclub	Supermarket
7	Lambeth	15	Café	Pub	Coffee Shop	Gay Bar	Nightclub	Indian Restaurant	Hotel	Italian Restaurant	Korean Restaurant	Cricket Ground

```
london_merged.loc[london_merged['Cluster Labels'] == 1, london_merged.columns[[0] +[1]+ list(range(5, london_merged.shape[1]))]]
```

	borough_name	cnt	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	Hackney	27	Vietnamese Restaurant	Café	Pub	Grocery Store	Bakery	Yoga Studio	Soccer Field	Middle Eastern Restaurant	Jewish Restaurant	Convenience Store
4	Hammersmith and Fulham	6	Pub	Grocery Store	Coffee Shop	Hotel	Thai Restaurant	Italian Restaurant	Café	Performing Arts Venue	Cocktail Bar	Pizza Place
9	Tower Hamlets	26	Café	Coffee Shop	Pub	Grocery Store	Park	Fast Food Restaurant	Pizza Place	Bakery	Turkish Restaurant	Church
10	Wandsworth	9	Café	Pub	Coffee Shop	Indian Restaurant	Bar	Pizza Place	English Restaurant	Beer Store	Chinese Restaurant	Thai Restaurant

```
london_merged.loc[london_merged['Cluster Labels'] == 2, london_merged.columns[[0]+[1] + list(range(5, london_merged.shape[1]))]]
```

	borough_name	cnt	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
6	Kensington and Chelsea	26	Hotel	Indian Restaurant	Chinese Restaurant	Restaurant	Italian Restaurant	French Restaurant	Clothing Store	Pub	Dance Studio	Coffee Shop

```
london_merged.loc[london_merged['Cluster Labels'] == 3, london_merged.columns[[0]+[1] + list(range(5, london_merged.shape[1]))]]
```

	borough_name	cnt	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
2	City of Westminster	119	Art Gallery	Clothing Store	Boutique	Indian Restaurant	French Restaurant	Lounge	Italian Restaurant	Cosmetics Shop	Hotel	Tailor Shop

```
london_merged.loc[london_merged['Cluster Labels'] == 4, london_merged.columns[[0]+[1] + list(range(5, london_merged.shape[1]))]]
```

	borough_name	cnt	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
8	Southwark	20	Pub	Public Art	Bar	Dessert Shop	Convenience Store	Pizza Place	Coffee Shop	Bus Stop	Supermarket	Hotel

07

IBM Coursera

Results discussion

- ✘ Cluster #3: Even though the City of Westminster satisfies our conditions in the best way I wouldn't recommend this borough.

Comment: Due to the large number of art centers&galleries it's highly competitive, and doing business here could be difficult for the young company. Also, apparently to rent the office here would be extremely expensive.

- ✓ Cluster #0: Is more suitable for our "NNN" company. I would strongly recommend these boroughs as the most suitable place for the office.

Comment: All boroughs of the cluster have highly developed social infrastructure and a lot of venues for the benefit of office employees. As well it provides a lot of possibilities in getting new customers and to be in touch with the artistic life.

- ✓ Clusters #2 and #4 appear to be promising and could be appropriate to rent an office.

Conclusion

/ Even though amount of data was limited and some aspects were not taken into account, we got a relevant list of recommendations/

It’s amazing how easy and quickly you can get the comprehensive analysis of economic and social infrastructure using only open data. This approach is quite flexible. There are a lot of ways for development and refinement of models and methods. Than more accurate and detailed data you have, than more relevant and precise result you would obtain. The results of such analysis are highly important and helpful in making sensible decision concerning business issues.

There’s always room for improvement