LONDON BOROUGHS CLUSTERING

IBM Data Science Professional Certificate

Capstone Project

Tomás Bertoglia

Munich, 25. Jun 2020

INTRO: TWO CHALLENGES ONE RESULT

- Our stakeholder requires a selection of boroughs to place his coffee schop chain
- Hig-end customer affluence and suitable commercial profile are the requirements
- The ultimate goals: guarantee brand positioning and business success

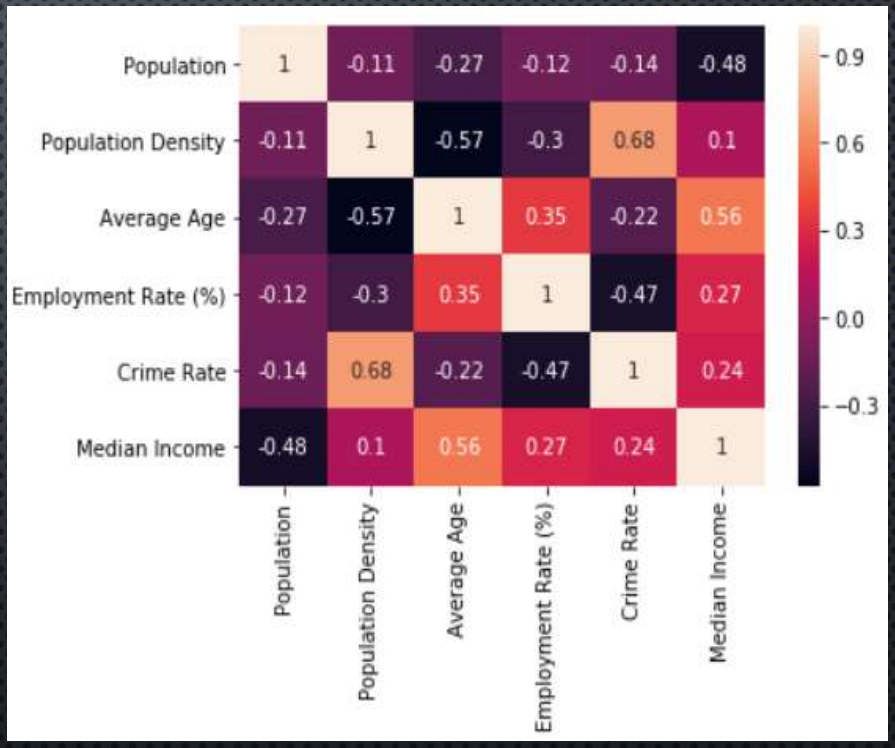
DATA & PREPARATION

- Boroughs and Coordinates Dataset crawled from wikipedia: https://en.wikipedia.org/wiki/List_of_London_boroughs
- ✓ *BeautifulSoup used for scrapping*

- Demographc Data downloaded from Greater London Authority (GLA) website: https://data.london.gov.uk/dataset/london-borough-profiles
- ✓ 6 out of 81 Features selected
- ✓ Features correlation tested trhough *heatmap*
- ✓ *StandardScaler* package used for normalization

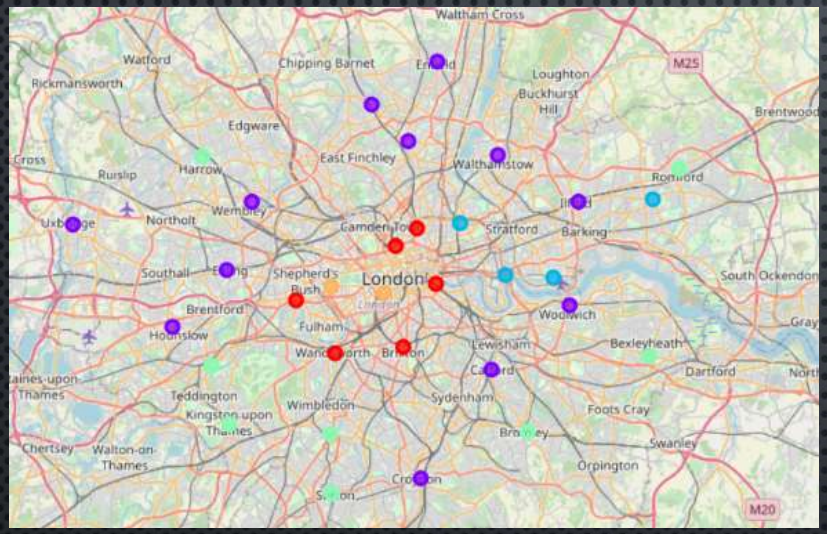
- Venues Data obained through the Foursquare API: https://api.foursquare.com/v2/venues/explore

DATA & PREPARATION



- DEMOGRAPHIC DATA FEATURES SELECTION: HIGH CORRELATION BETWEEN ‚HOUSE PRICE‘ AND 3 OTHER FEATURES SUGGEST HIS REMOVAL
- ✓ ‘MEDIAN HOUSE INCOME’ WOULD PROVIDE THE CLUSTERING INFO ASSOCIATED TO ‚HOUSE PRICE‘

DEMOGRAPHIC CLUSTERING

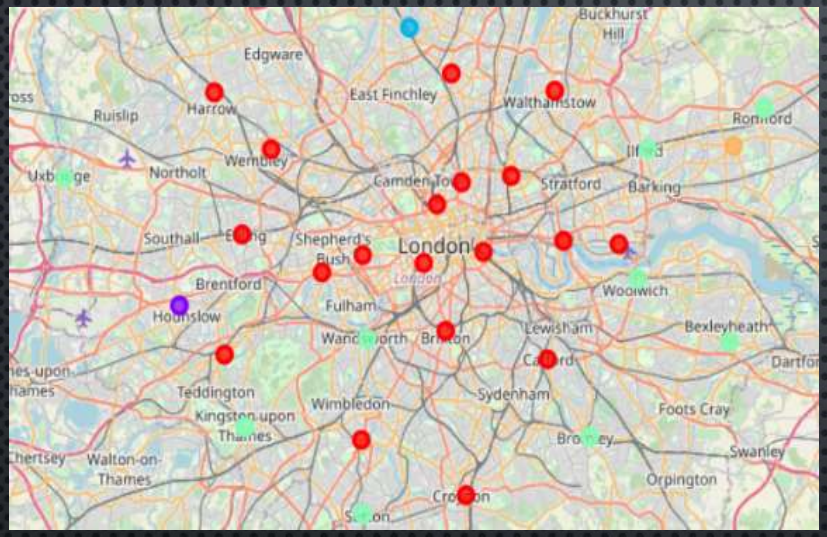


	Borough	Cluster Labels	Population	Population Density	Average Age	Employment Rate (%)	Crime Rate	House Price	Median Income
5	Camden	0	242500	111.3	36.4	69.2	123.5	700000	£43,750
11	Hammersmith and Fulham	0	185300	113	35.7	77.5	113.2	730000	£43,820
17	Islington	0	231200	155.6	34.8	72.6	121.2	583000	£39,790
20	Lambeth	0	328900	122.7	34.5	78.5	104.6	450000	£38,490
26	Southwark	0	314300	108.9	34.4	74.2	100.6	475000	£37,100
30	Wandsworth	0	321000	93.7	35.0	78.8	72.6	557000	£47,480

	Borough	Cluster Labels	Population	Population Density	Average Age	Employment Rate (%)	Crime Rate	House Price	Median Income
18	Kensington and Chelsea	4	159000	131.1	39.3	68.2	120.9	1200000	£55,620
31	Westminster	4	242100	112.7	37.7	65.6	212.4	920000	£47,510

- MAIN AIM: GUARANTEE HIGH-END CUSTOMER BOROUGH PROFILE, AS REQUIRED BY THE STAKEHOLDER
- ✓ SPECIAL FOCUS PUT IN CLUSTERS WITH ‚HIGH MEDIAN INCOME‘
- ✓ CENTRAL LOCATION OF BOROUGHS ALSO FAVORED

CLUSTERING BY VENUES



	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	C
3	Brent	0	Coffee Shop	Hotel	Sporting Goods Shop	Bar	
5	Camden	0	Café	Hotel	Coffee Shop	Pub	
6	Croydon	0	Coffee Shop	Pub	Portuguese Restaurant	Burrito Place	
7	Ealing	0	Coffee Shop	Clothing Store	Italian Restaurant	Pub	
10	Hackney	0	Pub	Café	Coffee Shop	Bakery	
11	Hammersmith and Fulham	0	Pub	Café	Indian Restaurant	Coffee Shop	Ita
12	Haringey	0	Fast Food Restaurant	Park	Portuguese Restaurant	Gym / Fitness Center	
13	Harrow	0	Indie Movie Theater	Indian Restaurant	Coffee Shop	Supermarket	

- MAIN AIM: GUARANTEE THE SELECTION OF SUITABLE BOROUGHS FOR THE COMMERCIAL SUCCESS OF THE COFFEE SHOP CHAIN

✓ FOCUS ON CLUSTERS WITH HIGH PRESENCE OF COFFEE SHOPS AND CAFÉS

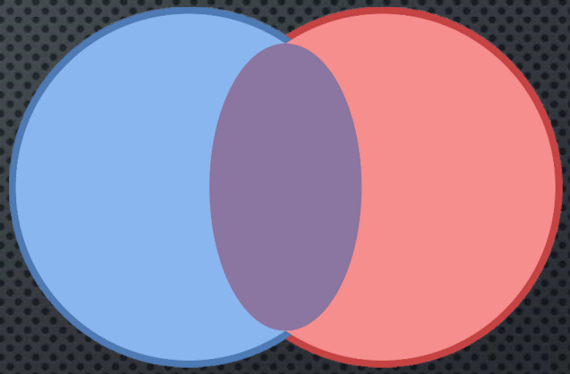
✓ ATTENTION TO COMPLIMENTARY SOURCE OF TARGETED CUSTOMERS SUCH AS HOTELS

LOW ARI SCORE: INTERSECTION REQUIRED

```
from sklearn.metrics.cluster import adjusted_rand_score
```

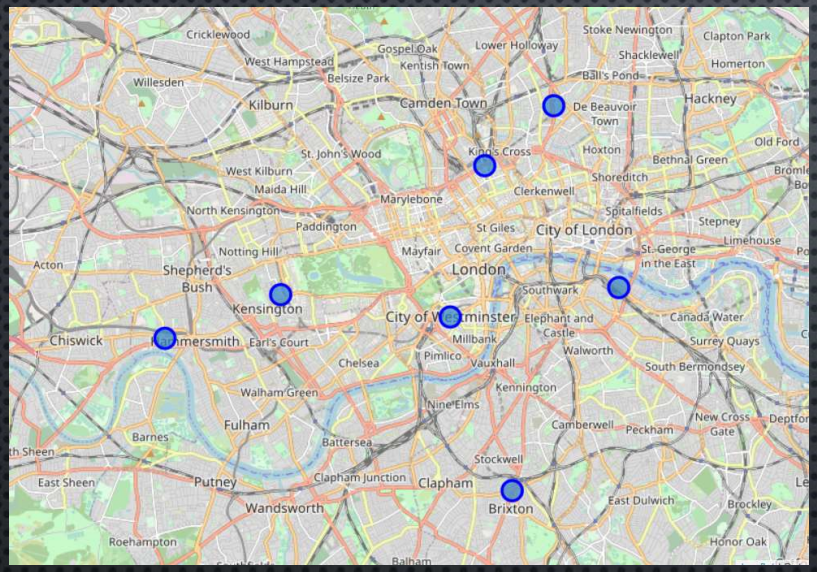
```
adjusted_rand_score(kmeans_venues.labels_, kmeans_dem.labels_)
```

```
-0.0217877094972067
```



- A LOW ARI SCORE SUGGEST INDEPENDENT LOGICS UNDER THE 2 CLUSTERING LOGICS EMPOYED
- ✓ BOTH LOGICS SHOULD BE COMBINED TO FULFILL THE REQUIREMENTS OF OUR STAKEHOLDER
- ✓ INTERSECTION OF THE BOROUGHS SELECTED AFTER CLUSTERING SELECTION WILL BE APPLIED

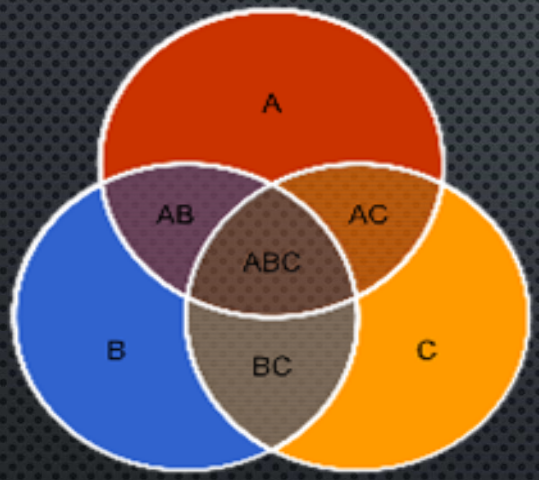
FINAL RESULTS



- CAMDEN
- HAMMERSMITH AND FULHAM
- ISLINGTON
- KENSINGTON AND CHELSEA
- LAMBETH
- SOUTHWARK
- WESTMINSTER

- THE INTERSECTION OF CLUSTERING RESULTS PROVIDED A SHORT LIST OF 7 BOROUGHS
- ✓ THOSE BOROUGHS FULFILL BOTH, THE MARKET POTENTIAL AND CUSTOMER PROFILE REQUIREMENTS.

CONCLUSION & DISCUSSION



- ✓ VERY OFTEN, WE WOULD NEED TO COMBINE DIFFERENT PROCEDURES TO ATTEND OUR STAKEHOLDER'S REQUIREMENTS
- ✓ THE INTERSECTION OF RESULTS IS A GREAT AND ESCALABLE TOOL TO SATISFY MULTIPLE REQUIREMENTS
- ✓ IN A REAL-LIFE SCENARIO, CLOSE PARTICIPATION OF OUR STAKEHOLDERS COULD HELP US FURTHER REFINE OUR RESULTS

IBM

coursera

THANKS FOR YOUR ATTENTION