Recommendation System for Airbnb listings

Ioannis Petromichelakis

Applicant for The Data Incubator Fellowship program

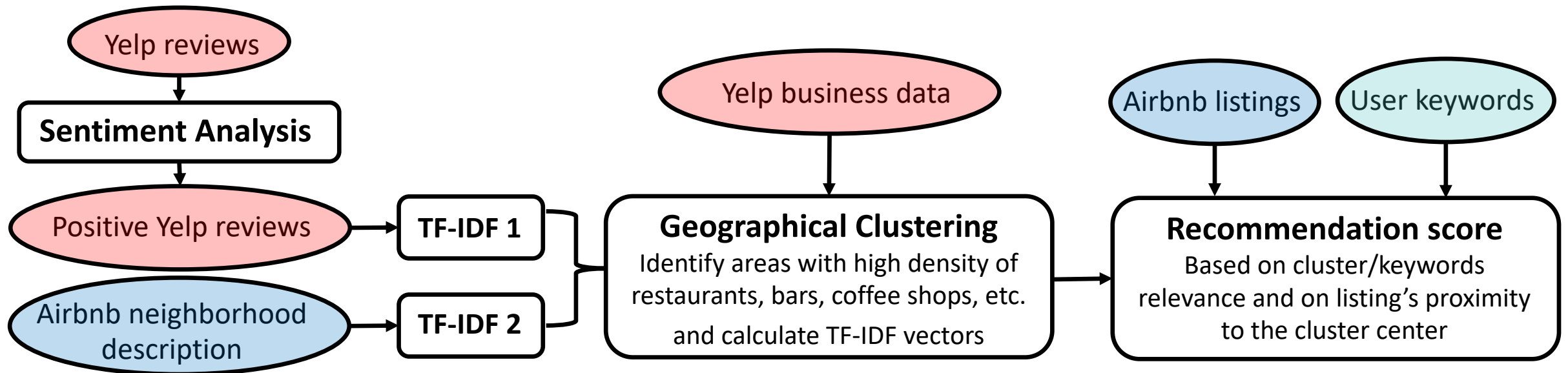The Data Incubator
A Pragmatic Institute Company

Outline

Motivation, assumptions and objectives

- Yelp Challenge Dataset (5.5GB) from yelp.com
- Airbnb Listings in Toronto and Montreal (200MB) from insideairbnb.com
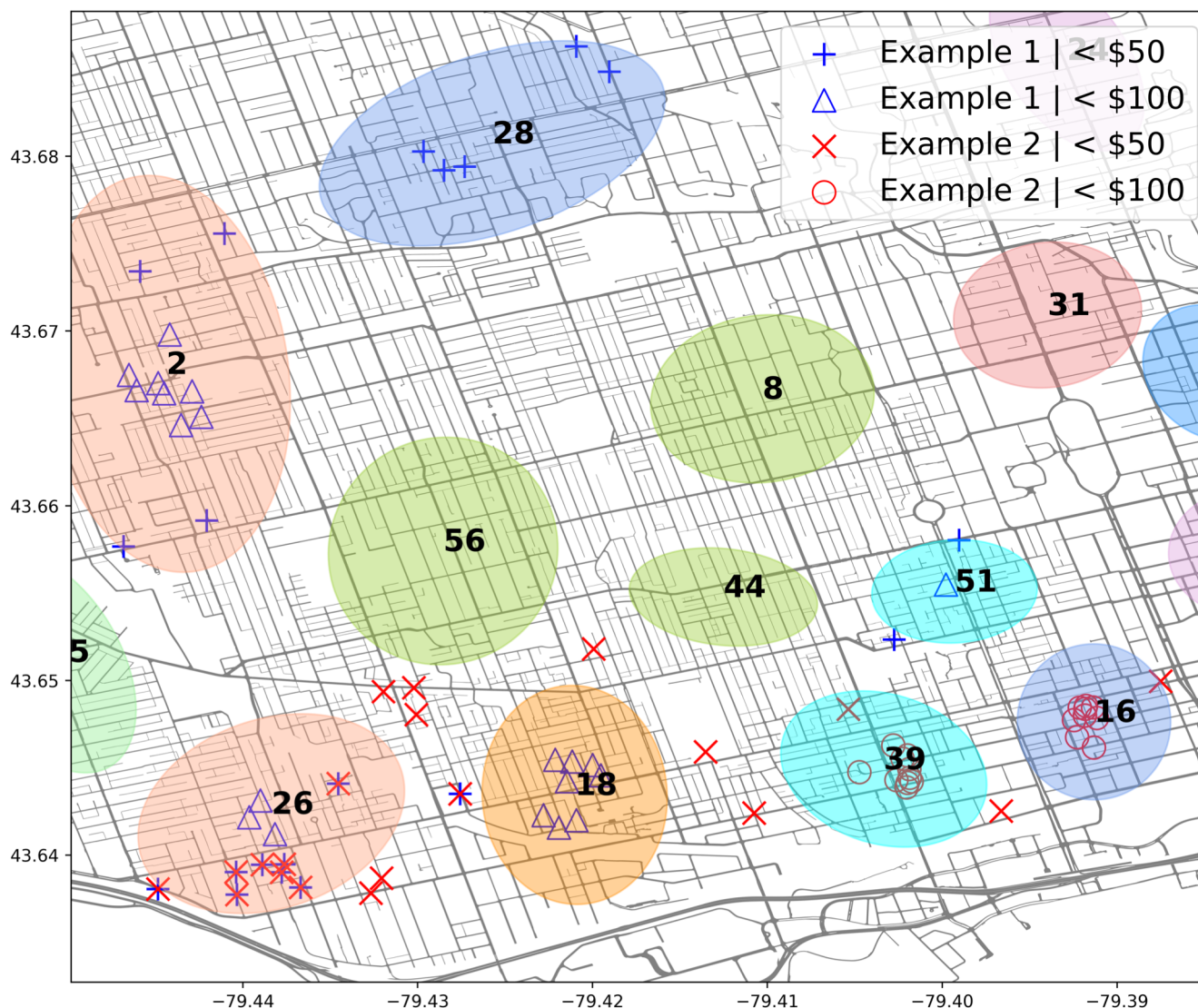
Data:

- Yelp Challenge Dataset (5.5GB) from yelp.com
- Airbnb Listings in Toronto and Montreal (200MB) from insideairbnb.com

Results – Toronto

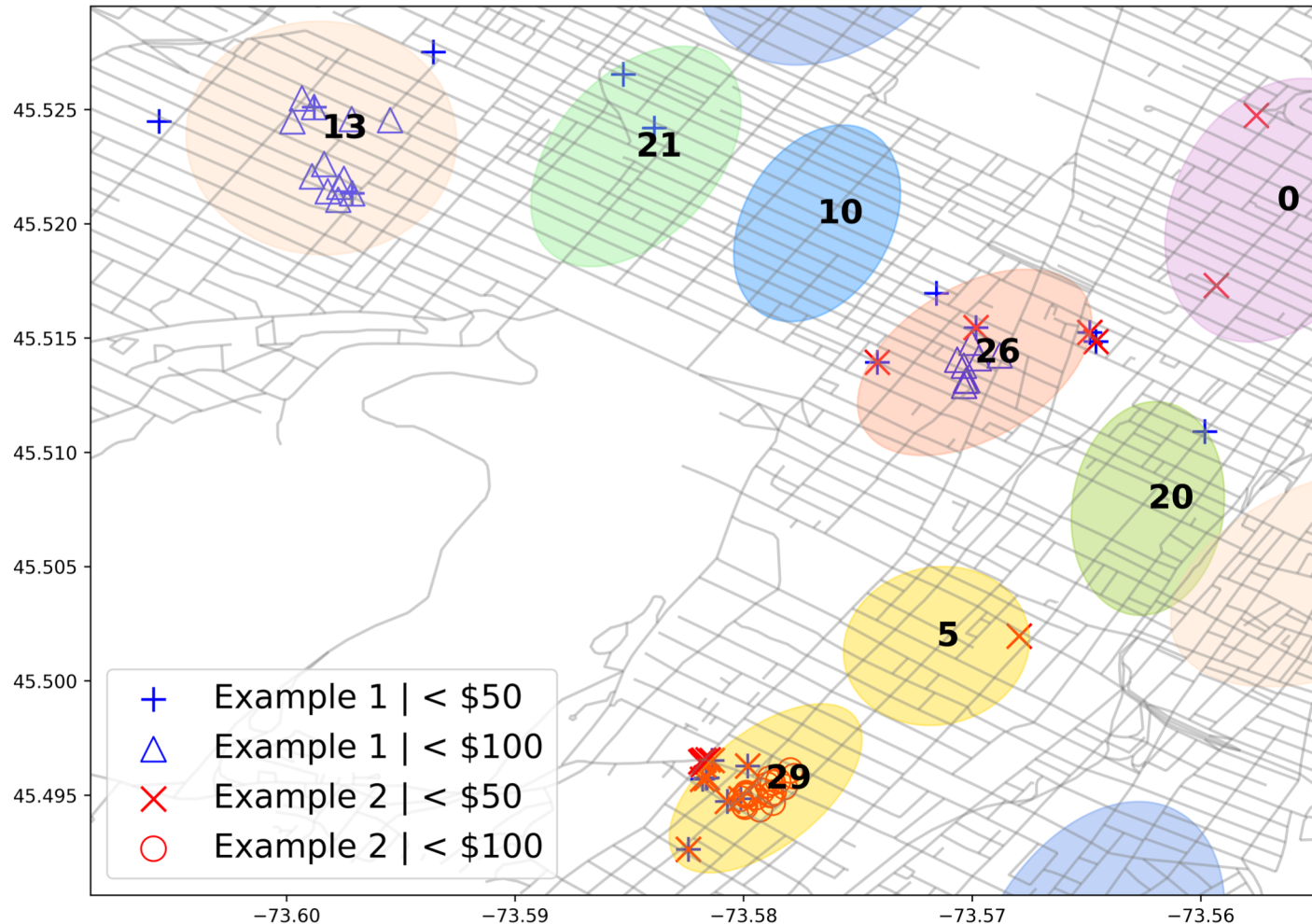Example 1: 'hipster', 'hippy, art,… Example 2: clubs, nightlife,…

Cluster	Scr 1	Scr 2	# listings	Av. Price	Neighborhoods	
2	6.6	2.3	279	$101.33	Dovercourt-Junction	69%
					Corso Italia-Davenport	21%
16	2.8	5.0	794	$204.44	Waterfront Communities	94%
18	6.8	3.8	487	$177.35	Little Portugal	37%
					Niagara	33%
					Trinity-Bellwoods	19%
26	6.1	4.2	319	$125.96	Roncesvalles	40%
					South Parkdale	38%
					Little Portugal	18%
28	4.1	1.6	125	$119.88	Wychwood	53%
					Humewood-Cedarvale	22%
39	3.5	5.1	769	$199.43	Waterfront Communities	68%
					Niagara	20%
44	3.1	3.7	174	$141.52	Trinity-Bellwoods	53%
					Palmerston-Little Italy	38%
51	6.4	1.9	250	$122.03	Kensington-Chinatown	98%

Results – Montreal

Cluster	Scr 1	Scr 2	# listings	Av. price	Neighborhood	
0	4.5	**7.4**	787	$109.20	Ville-Marie	100%
5	5.8	**8.8**	325	$186.20	Ville-Marie	76%
					Le Plateau	24%
13	**10.3**	5.3	603	$122.17	Le Plateau	77%
					Outremont	14%
20	**8.0**	6.4	453	$158.07	Ville-Marie	100%
21	**6.6**	5.5	691	$115.44	Le Plateau	99%
25	5.3	**7.0**	420	$187.64	Ville-Marie	98%
26	**8.3**	**8.2**	823	$159.81	Le Plateau	70%
					Ville-Marie	30%
29	**6.4**	**9.3**	728	$125.02	Ville-Marie	99%

Outline