

City Advisor

by Antonio Saccoman

August 2020

1. Introduction. Description of the problem and discussion of the background.

The business case for the analysis being developed here consists of advisory value for a wealthy client looking for consultation over the potential purchase of a second home in one of the largest metropolitan areas around the world.

Client is aiming at a purchase in central areas, also to maximize the number of available venues within short distance. He/she also does not express preference for a specific city, but wishes to extend the search to multiple cities in different countries. For more details about client's assumptions, please see Methodology chapter.

Data Science tools will be applied to 1) select most central areas of a given list of cities; 2) retrieve venues info for all the candidate areas; 3) analyse venues in terms of category frequency per area; 4) narrow the area search progressively via clustering and ranking functions; 5) provide final suggestions to the client in the form of best 3 areas where to look for a property.

2. Data description and sources

Analysis as introduced above will require multiple types of data; see below.

a) Initial boroughs list for each city will be retrieved ('scraped') from the relevant city's Wikipedia page info (e.g. initial boroughs list); via BeautifulSoup.
See example of query output below. Also, further below example of additional data such as Quality of Life index from Numbeo website.

https://it.wikipedia.org/wiki/Municipi_di_Milano#Schema_delle_zone_di_Milano

...

	#	Denominazione	Superficie(km ²)	Abitanti(31.12.2018)	Densità(ab/km ²)
0	Municipio 1	Centro storico	967	98 531	10 189
1	Municipio 2	Stazione Centrale, Gorla, Turro, Greco, Cresce...	1258	162 090	12 884
2	Municipio 3	Città Studi, Lambrate, Venezia	1423	144 110	10 127
3	Municipio 4	Vittoria, Forlanini	2095	161 551	7 711
4	Municipio 5	Vigentino, Chiaravalle, Gratosoglio	2987	126 089	4 221

https://www.numbeo.com/quality-of-life/rankings_current.jsp

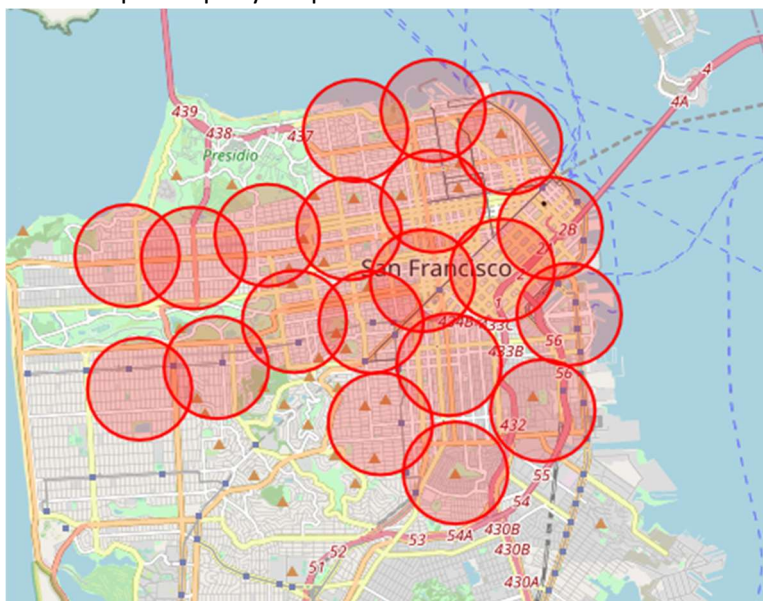
...

	Rank	City	Quality of Life Index	Purchasing Power Index	Safety Index	Health Care Index	Cost of Living Index	Property Price to Income Ratio	Traffic Commute Time Index	Pollution Index	Climate Index
0	NaN	Adelaide, Australia	207.83	108.87	71.37	80.91	71.20	4.27	24.13	17.94	94.96
1	NaN	Canberra, Australia	206.65	103.63	78.93	81.44	78.38	5.21	24.05	14.01	82.72
2	NaN	Raleigh, NC, United States	202.53	129.80	66.17	75.62	70.25	2.82	32.41	21.95	83.88
3	NaN	Wellington, New Zealand	200.21	97.75	70.07	74.90	71.02	6.36	27.74	13.66	97.68
4	NaN	Madison, WI, United States	197.34	119.50	68.89	78.89	67.15	3.43	23.73	17.51	51.64

b) City areas geographical coordinates; via Nominatim.
See example of query output below.

	Borough	City+	Latitude	Longitude
62	Hammersmith	London, United Kingdom	51.492038	-0.223640
63	Bond street	London, United Kingdom	51.514299	-0.149002
64	Kennington	London, United Kingdom	51.488286	-0.105883
65	Battersea	London, United Kingdom	51.470793	-0.172214
66	Chelsea	London, United Kingdom	51.487542	-0.168220
67	Holborn	London, United Kingdom	51.517934	-0.119528
68	Shoreditch	London, United Kingdom	51.526669	-0.079893

c) Maps of areas in a given city; via Folium.
See example of query output below.



d) Venues info for each area, such as geo coordinates and category; via Foursquare API.
See example of query output below; also mapping via Folium further below.

	Borough	City+	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
4250	6th Arrondissement	Paris, France	48.850433	2.332951	Place Saint-Sulpice	48.850823	2.333323	Plaza
4251	6th Arrondissement	Paris, France	48.850433	2.332951	Pierre Hermé	48.851532	2.332766	Pastry Shop
4252	6th Arrondissement	Paris, France	48.850433	2.332951	Evi Evane	48.851886	2.334288	Greek Restaurant
4253	6th Arrondissement	Paris, France	48.850433	2.332951	L'Avant-Comptoir du Marché	48.851781	2.335429	Bistro
4254	6th Arrondissement	Paris, France	48.850433	2.332951	Muji	48.851272	2.335605	Miscellaneous Shop

