



Instituto Infnet

Clusterizando os Baixos do Rio de Janeiro e Similares

Introdução

A cidade do Rio de Janeiro possui diversas maneiras de entretenimento, entre elas, a reunião de amigos em volta de uma mesa de bar ou restaurante. Esta atividade deixa sua marca e disposição em um determinado espaço, dentre os quais ganha notoriedade o **Baixo Gávea**.

O sucesso do Baixo Gávea fez com que diversos outros empresários construíssem Baixos pela cidade.

Baixos são áreas especializadas pelo agrupamento de bares e serviços noturnos, formando um tipo de “distrito de diversões e lazer”, que converge para uma forma espacial complexa, de grande expressão no bairro e adjacências, aumentando a circulação individual por este espaço.



Introdução

Neste projeto, tentaremos compreender a diversidade dos Baixos, utilizando a **API do Google Places** e o aprendizado de máquina não supervisionado do **K-means**.

Nossa **hipótese** é que alguns baixos possuem uma proximidade espacial que não se refere à área da cidade que ocupam, mas sim ao estilo dos estabelecimentos circundantes.

Estes dados podem ajudar tanto urbanistas quanto políticos a entender melhor os espaços delimitados como baixos (ou similares), buscando estabelecer melhores políticas públicas e infraestrutura.

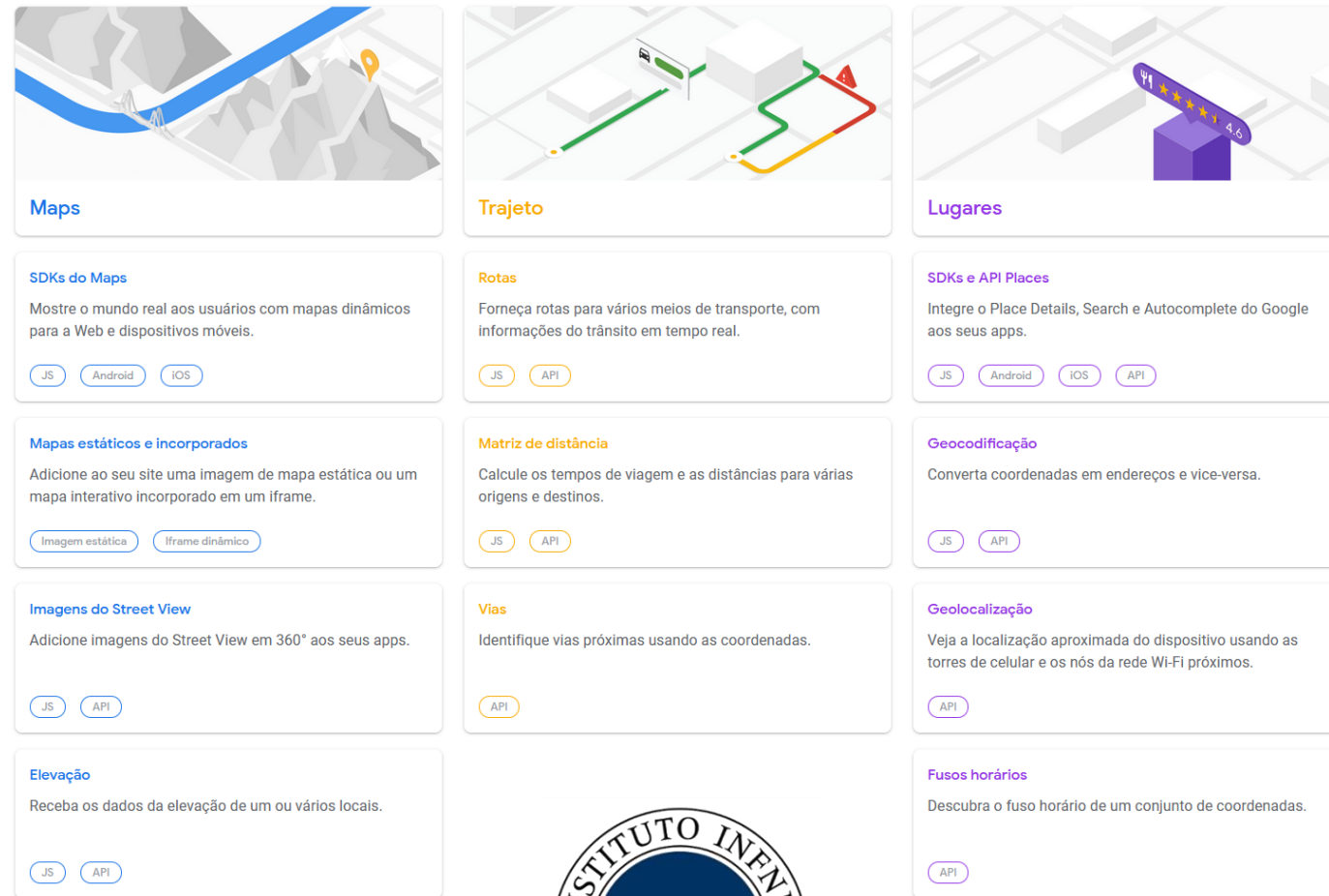


Google Places API



Google Places API

- Dados atualizados diariamente com a base interativa do Google Maps Platform.
- Volume mensal de 100 mil requests grátis.
- O Google dá US\$ 300 de bônus por 90 dias para o caso de você ultrapassar esse volume.
- Facilidade de configuração, necessitando apenas sua chave do API para realizar o request.



Maps	Trajeto	Lugares
SDKs do Maps Mostre o mundo real aos usuários com mapas dinâmicos para a Web e dispositivos móveis. JS Android iOS	Rotas Forneça rotas para vários meios de transporte, com informações do trânsito em tempo real. JS API	SDKs e API Places Integre o Place Details, Search e Autocomplete do Google aos seus apps. JS Android iOS API
Mapas estáticos e incorporados Adicione ao seu site uma imagem de mapa estática ou um mapa interativo incorporado em um iframe. Imagem estática Iframe dinâmico	Matriz de distância Calcule os tempos de viagem e as distâncias para várias origens e destinos. JS API	Geocodificação Converta coordenadas em endereços e vice-versa. JS API
Imagens do Street View Adicione imagens do Street View em 360° aos seus apps. JS API	Vias Identifique vias próximas usando as coordenadas. API	Geolocalização Veja a localização aproximada do dispositivo usando as torres de celular e os nós da rede Wi-Fi próximos. API
Elevação Receba os dados da elevação de um ou vários locais. JS API		Fusos horários Descubra o fuso horário de um conjunto de coordenadas. API

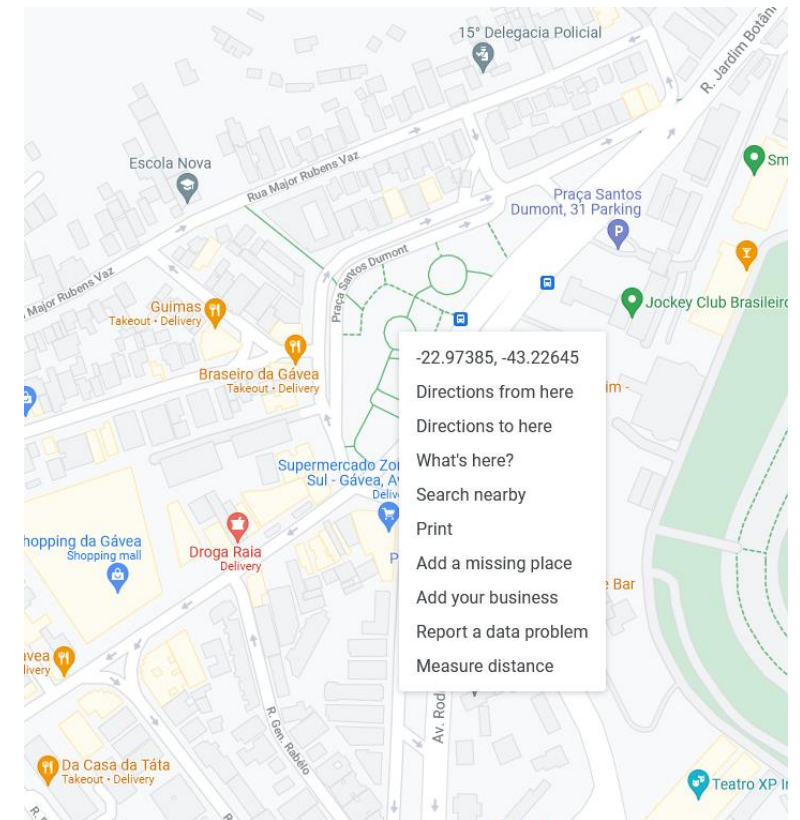
https://maps.googleapis.com/maps/api/place/nearbysearch/json?location=-33.8670522,151.1957362&radius=500&types=food&name=harbour&key=YOUR_API_KEY



Dados

- Latitude e Longitude dos baixos retirados do Google Maps.
- DataFrame (DF) feito com uma lista dos dados retirados.
- Mapa de visualização feito com o Folium.

	neighbourhood	latitude	longitude
0	Baixo Gávea	-22.973781	-43.226453
1	Baixo Leblon	-22.985447	-43.227879
2	Baixo Botafogo	-22.951085	-43.184303
3	Baixo Tijuca	-22.919139	-43.234393
4	Baixo Méier	-22.900954	-43.279693
5	Baixo Parque Leopoldina	-22.885182	-43.454211
6	Baixo Jabour	-22.882916	-43.491181
7	Baixo Marechal	-22.863508	-43.372683
8	Baixo Madureira	-22.876430	-43.335195
9	Baixo Valqueire	-22.879312	-43.366461
10	Baixo São João de Meriti	-22.779338	-43.368417
11	Baixo Vista Alegre	-22.830709	-43.319064



Dados

- Baseado na tabela total de tipos de estabelecimentos do Google Places API, produzimos uma tabela com meta-categorias.

arts	['art gallery', 'movie rental', 'movie theater', 'museum']
education	['book store', 'library', 'primary school', 'school', 'secondary school', 'university']
finance	['accounting', 'atm', 'bank']
food	['bakery', 'cafe', 'meal delivery', 'meal takeaway', 'restaurant', 'supermarket']
health	['dentist', 'doctor', 'drugstore', 'funeral home', 'gym', 'hospital', 'pharmacy', 'physiotherapist', 'spa', 'veterinary care']
nightlife	['bar', 'casino', 'night club', 'liquor store']
outdoors	['amusement park', 'aquarium', 'bowling alley', 'campground', 'city hall', 'park', 'rv park', 'stadium', 'tourist attraction', 'zoo']
religion	['cemetery', 'church', 'hindu temple', 'mosque', 'synagogue']
service	['beauty salon', 'electrician', 'florist', 'gas station', 'hair care', 'insurance agency', 'laundry', 'lawyer', 'locksmith', 'lodging', 'moving company', 'painter', 'parking', 'plumber', 'real estate agency', 'roofing contractor', 'storage', 'travel agency']
shop	['clothing store', 'convenience store', 'department store', 'electronics store', 'furniture store', 'hardware store', 'home goods store', 'jewelry store', 'pet store', 'shoe store', 'shopping mall', 'store']
state	['courthouse', 'embassy', 'fire station', 'local government office', 'police', 'post office']
transport	['airport', 'bus station', 'bicycle store', 'car dealer', 'car rental', 'car repair', 'car wash', 'light rail station', 'subway station', 'taxi stand', 'train station', 'transit station']



Dados

- Coleta de dados com o Nearby Search do Google Places API.
- DF feito com dados retirados do Json resultante.

	baixo	name	latitude	longitude	types	price level	rating	total ratings
0	Baixo Gávea	Rio de Janeiro	-22.906847	-43.172897	['locality', 'political']	none	none	none
1	Baixo Gávea	Gavea quarto casal 1	-22.974366	-43.228875	['lodging', 'point_of_interest', 'establishment']	none	none	none
2	Baixo Gávea	Jockey Club Brasileiro	-22.973768	-43.224891	['point_of_interest', 'establishment']	none	4.6	6318
3	Baixo Gávea	Nossa Senhora da Conceição	-22.975556	-43.227222	['church', 'place_of_worship', 'point_of_inter...']	none	4.8	130
4	Baixo Gávea	Bacalhau do Rei	-22.975188	-43.226976	['restaurant', 'food', 'point_of_interest', 'e...']	2	4.4	474
...
2635	Baixo Taquara	Tienes Planos	-22.919589	-43.387489	['electronics_store', 'home_goods_store', 'poi...']	none	4.8	4
2636	Baixo Taquara	CELTIC BATERIAS	-22.916717	-43.385830	['car_repair', 'point_of_interest', 'store', 'l...']	none	4.4	94
2637	Baixo Taquara	Ramell	-22.916231	-43.387250	['point_of_interest', 'clothing_store', 'store...']	none	none	none
2638	Baixo Taquara	Thati Ganimi Biju- Atacado e Varejo	-22.916223	-43.387251	['point_of_interest', 'clothing_store', 'store...']	none	5	2
2639	Baixo Taquara	Gustosita	-22.917136	-43.389098	['restaurant', 'food', 'point_of_interest', 'e...']	2	4.6	184

2640 rows × 8 columns



Dados

- Após limpeza e melhor análise, fizemos um “one-hot encoding” (Obrigado Ciça).

	baixo	name	latitude	longitude	arts	education	finance	food	health	nightlife	outdoors	religion	service	shop	state	transport
0	Baixo Gávea	Gavea quarto casal 1	-22.974366	-43.228875	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
1	Baixo Gávea	Nossa Senhora da Conceição	-22.975556	-43.227222	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
2	Baixo Gávea	Bacalhau do Rei	-22.975188	-43.226976	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	Baixo Gávea	La Pastasciutta	-22.974908	-43.228766	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	Baixo Gávea	Soraia Cals Escritório de Arte	-22.975135	-43.227552	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...
1662	Baixo Taquara	Tienes Pianos	-22.919589	-43.387489	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
1663	Baixo Taquara	CELTIC BATERIAS	-22.916717	-43.385830	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0
1664	Baixo Taquara	Ramell	-22.916231	-43.387250	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
1665	Baixo Taquara	Thati Ganimi Biju- Atacado e Varejo	-22.916223	-43.387251	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
1666	Baixo Taquara	Gustosita	-22.917136	-43.389098	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

1667 rows × 16 columns



Clusterização

- Os dados foram agrupados por Baixos, antes de sua normalização.

```
[ ] # Group by baixo, calculate the mean
    clust_df = df.groupby('baixo').mean().reset_index()
    print(f'Dataframe shape: {clust_df.shape[0]} rows, {clust_df.shape[1]} columns.')
    clust_df.head(10)
```

Dataframe shape: 44 rows, 15 columns.

	baixo	latitude	longitude	arts	education	finance	food	health	nightlife	outdoors	religion	service	shop	state	transport
0	Arco do Teles	-22.903096	-43.175991	0.035714	0.071429	0.000000	0.285714	0.035714	0.035714	0.107143	0.071429	0.250000	0.285714	0.035714	0.035714
1	Baixo Afonso Pena	-22.918338	-43.218250	0.000000	0.060606	0.151515	0.181818	0.212121	0.060606	0.000000	0.060606	0.333333	0.212121	0.030303	0.030303
2	Baixo Anil	-22.951819	-43.338602	0.000000	0.093023	0.023256	0.116279	0.093023	0.046512	0.023256	0.046512	0.209302	0.441860	0.139535	0.139535
3	Baixo Bangu	-22.875012	-43.463241	0.000000	0.024390	0.048780	0.024390	0.219512	0.000000	0.000000	0.073171	0.097561	0.634146	0.048780	0.048780
4	Baixo Barra da Tijuca	-23.010328	-43.304743	0.000000	0.000000	0.068182	0.136364	0.113636	0.090909	0.000000	0.000000	0.204545	0.590909	0.022727	0.022727
5	Baixo Botafogo	-22.950916	-43.184058	0.000000	0.000000	0.025641	0.153846	0.102564	0.076923	0.000000	0.025641	0.435897	0.256410	0.076923	0.076923
6	Baixo Cachambi	-22.887484	-43.276361	0.000000	0.043478	0.086957	0.195652	0.065217	0.086957	0.000000	0.065217	0.108696	0.434783	0.065217	0.065217
7	Baixo Campo Grande	-22.901103	-43.566737	0.000000	0.117647	0.029412	0.117647	0.205882	0.000000	0.000000	0.088235	0.176471	0.235294	0.058824	0.058824
8	Baixo Copacabana	-22.975924	-43.189638	0.000000	0.020000	0.040000	0.180000	0.140000	0.000000	0.000000	0.000000	0.500000	0.220000	0.000000	0.000000
9	Baixo Del Castilho	-22.878380	-43.272105	0.000000	0.020000	0.060000	0.060000	0.040000	0.040000	0.000000	0.000000	0.160000	0.660000	0.020000	0.020000



Clusterização

- Rankeamento das 5 atividades mais realizadas em cada baixo.

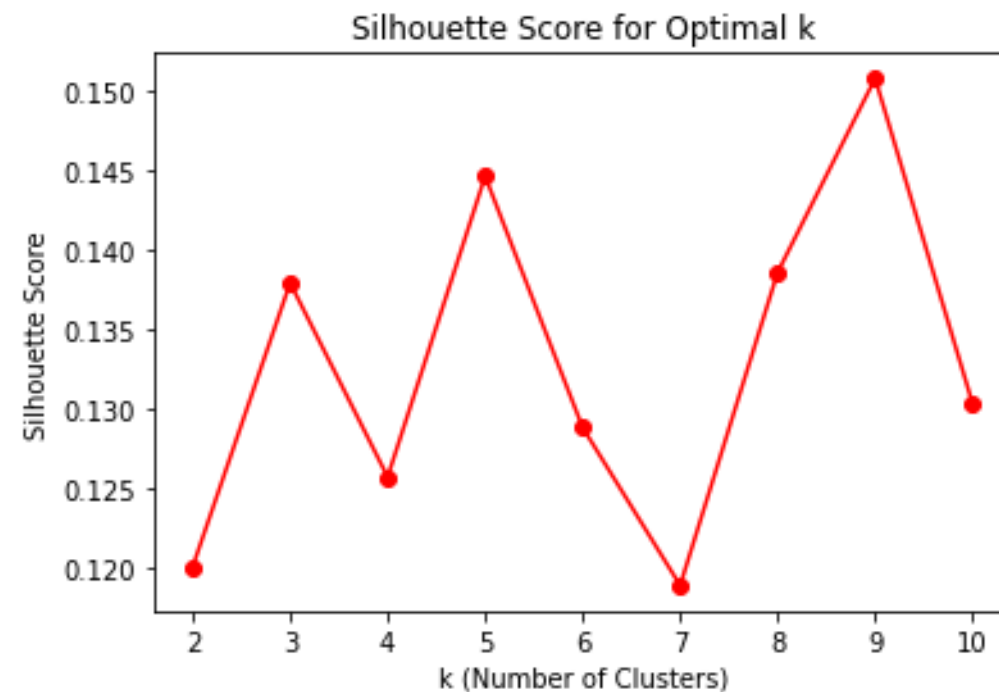
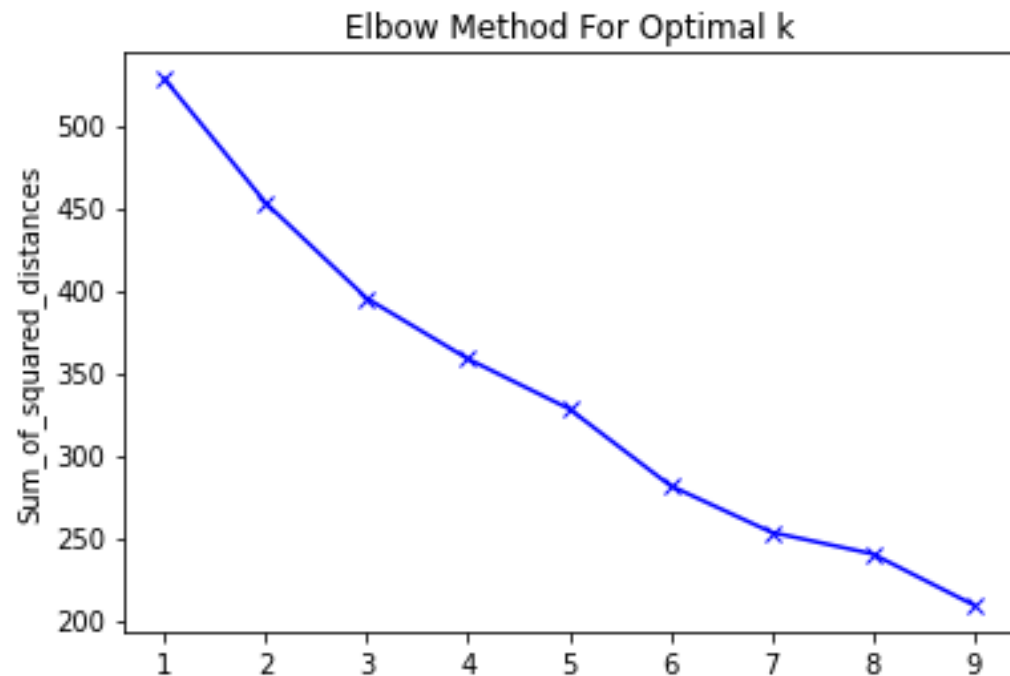
	baixo	1st Common Category	2nd Common Category	3rd Common Category	4th Common Category	5th Common Category
0	Arco do Teles	shop	food	service	outdoors	religion
1	Baixo Afonso Pena	service	shop	health	food	finance
2	Baixo Anil	shop	service	transport	state	food
3	Baixo Bangu	shop	health	service	religion	transport
4	Baixo Barra da Tijuca	shop	service	food	health	nightlife
5	Baixo Botafogo	service	shop	food	health	transport
6	Baixo Cachambi	shop	food	service	nightlife	finance
7	Baixo Campo Grande	shop	health	service	food	education
8	Baixo Copacabana	service	shop	food	health	finance
9	Baixo Del Castilho	shop	service	food	finance	nightlife

	1st Common Category	2nd Common Category	3rd Common Category	4th Common Category	5th Common Category
count	44	44	44	44	44
unique	4	8	9	11	11
top	shop	service	food	health	nightlife
freq	31	13	17	10	11



Clusterização

- Determinação do número ótimo de clusters com Elbow Method e Silhouette Score.



Clusterização

- Utilização do K-Means e incorporação no DF agrupado e normalizado.

```
[ ] k = 9 # number of clusters
```

```
# Run K-Means algorithm, random_state=0 to generate reproducible results
model = KMeans(n_clusters=k, random_state=0).fit(dfnorm)
model.labels_ # Cluster labels generated for each row
```

```
array([8, 5, 3, 1, 0, 2, 3, 1, 2, 0, 2, 2, 1, 4, 0, 6, 2, 0, 0, 2, 1, 6,
       6, 1, 3, 0, 1, 6, 0, 5, 3, 0, 5, 1, 3, 0, 0, 3, 3, 0, 2, 0, 7, 5],
      dtype=int32)
```

```
[ ] clust_df['cluster'] = model.labels_
clust_df.head()
```

	baixo	latitude	longitude	arts	education	finance	food	health	nightlife	outdoors	religion	service	shop	state	transport	cluster
0	Arco do Teles	-22.903096	-43.175991	0.035714	0.071429	0.000000	0.285714	0.035714	0.035714	0.107143	0.071429	0.250000	0.285714	0.035714	0.035714	8
1	Baixo Afonso Pena	-22.918338	-43.218250	0.000000	0.060606	0.151515	0.181818	0.212121	0.060606	0.000000	0.060606	0.333333	0.212121	0.030303	0.030303	5
2	Baixo Anil	-22.951819	-43.338602	0.000000	0.093023	0.023256	0.116279	0.093023	0.046512	0.023256	0.046512	0.209302	0.441860	0.139535	0.139535	3
3	Baixo Bangu	-22.875012	-43.463241	0.000000	0.024390	0.048780	0.024390	0.219512	0.000000	0.000000	0.073171	0.097561	0.634146	0.048780	0.048780	1
4	Baixo Barra da Tijuca	-23.010328	-43.304743	0.000000	0.000000	0.068182	0.136364	0.113636	0.090909	0.000000	0.000000	0.204545	0.590909	0.022727	0.022727	0



Resultados

- Gerado um mapa de visualização com o Folium.
- Clusters filtrados para maior análise como o exemplo a seguir.

```
[ ] # Filter only Cluster 0
cluster0 = df_clust_check.loc[df_clust_check['cluster'] == 0].reset_index(drop=True)
cluster0
```

	baixo	latitude	longitude	arts	education	finance	food	health	nightlife	outdoors	religion	service	shop	state	transport	cluster	1st Common Category	2nd Common Category	3rd Common Category	4th Common Category	5th Common Category
0	Baixo Barra da Tijuca	-23.010328	-43.304743	0.0	0.000000	0.068182	0.136364	0.113636	0.090909	0.000000	0.000000	0.204545	0.590909	0.022727	0.022727	0	shop	service	food	health	nightlife
1	Baixo Del Castilho	-22.878380	-43.272105	0.0	0.020000	0.060000	0.060000	0.040000	0.040000	0.000000	0.000000	0.160000	0.660000	0.020000	0.020000	0	shop	service	food	finance	nightlife
2	Baixo Icarai	-22.904395	-43.102328	0.0	0.048780	0.000000	0.365854	0.219512	0.146341	0.000000	0.024390	0.146341	0.365854	0.000000	0.000000	0	shop	food	health	service	nightlife
3	Baixo Jabour	-22.882517	-43.491425	0.0	0.135135	0.027027	0.135135	0.054054	0.108108	0.027027	0.000000	0.135135	0.405405	0.027027	0.027027	0	shop	service	food	education	nightlife
4	Baixo Largo do Machado	-22.931066	-43.178865	0.0	0.050000	0.000000	0.175000	0.150000	0.025000	0.050000	0.025000	0.300000	0.450000	0.000000	0.000000	0	shop	service	food	health	outdoors
5	Baixo Parque Leopoldina	-22.884759	-43.454293	0.0	0.000000	0.025641	0.282051	0.153846	0.051282	0.025641	0.051282	0.179487	0.230769	0.076923	0.076923	0	food	shop	service	health	transport
6	Baixo Recreio dos Bandeirantes	-23.017800	-43.459553	0.0	0.088235	0.029412	0.176471	0.147059	0.147059	0.000000	0.058824	0.264706	0.323529	0.029412	0.029412	0	shop	service	food	nightlife	health
7	Baixo São Francisco (Niterói)	-22.920091	-43.093453	0.0	0.025641	0.025641	0.205128	0.179487	0.076923	0.000000	0.000000	0.282051	0.333333	0.025641	0.025641	0	shop	service	food	health	nightlife
8	Baixo Taquara	-22.917606	-43.386641	0.0	0.044444	0.022222	0.288889	0.066667	0.066667	0.000000	0.000000	0.133333	0.466667	0.022222	0.022222	0	shop	food	service	nightlife	health
9	Baixo Tijuca	-22.919346	-43.234227	0.0	0.000000	0.029412	0.294118	0.117647	0.264706	0.000000	0.000000	0.176471	0.352941	0.058824	0.058824	0	shop	food	nightlife	service	health
10	Baixo Vargem Grande	-22.984529	-43.497433	0.0	0.027778	0.000000	0.333333	0.027778	0.138889	0.027778	0.000000	0.222222	0.333333	0.055556	0.055556	0	shop	food	service	nightlife	transport
11	Baixo Vista Alegre	-22.830937	-43.319161	0.0	0.055556	0.138889	0.305556	0.138889	0.083333	0.000000	0.055556	0.027778	0.416667	0.027778	0.027778	0	shop	food	health	finance	nightlife





Instituto Infnet

GitHub: alexro6/Mid-term-Project

Muito obrigado!