# IBM Coursera Capstone

Starting a new business with Data Science

# Where to start a new business?

Location is decisive for success

Similar neighborhoods and clients should produce similar results

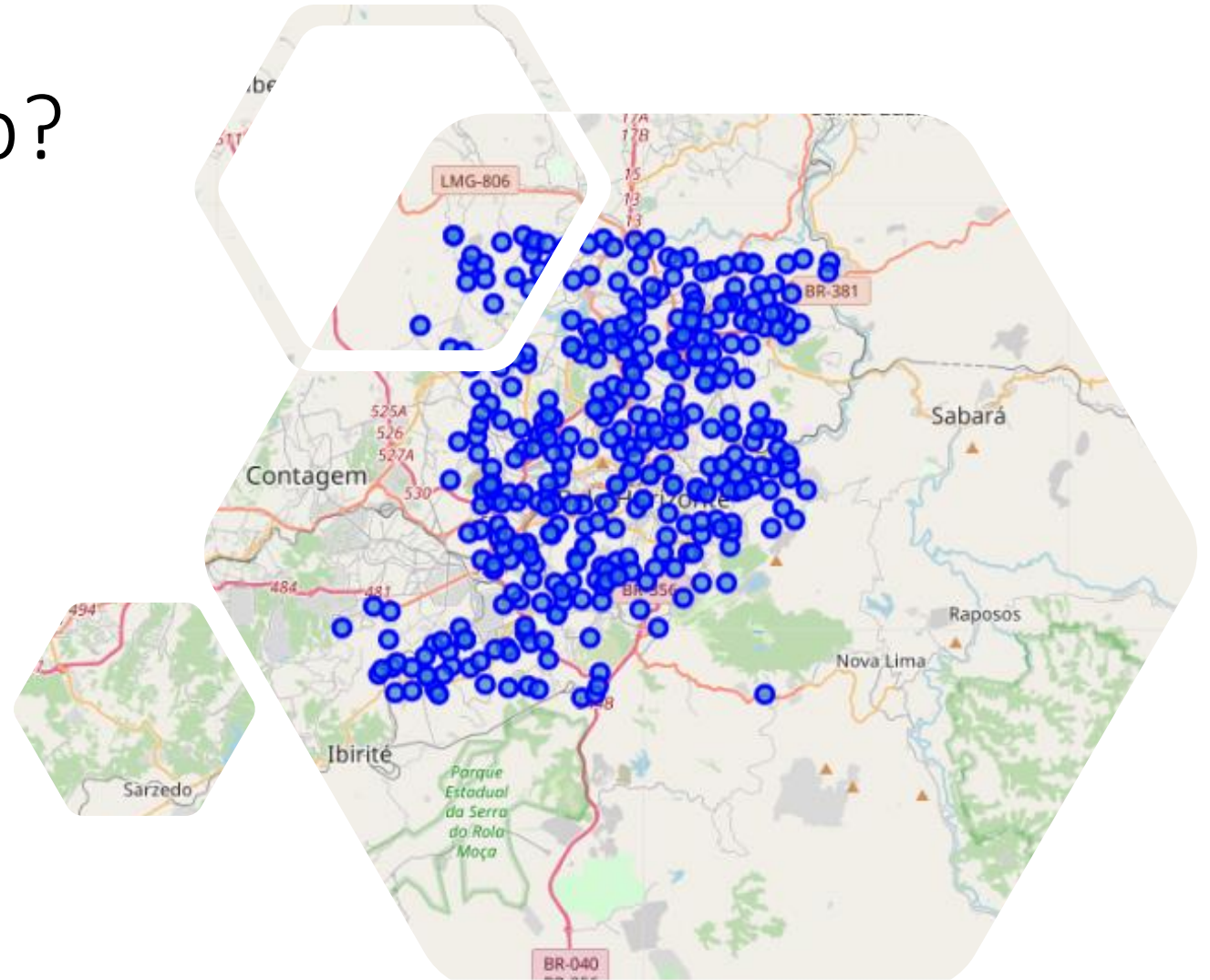Higher population density means more potential customers

Successful businesses shows the interest of a region

Given a Pet Store in Buritis, Belo Horizonte, where should a new store be opened?

# Can Data Science help?

- Belo Horizonte has over 300 neighborhoods

- More than 7864 venues on Foursquare

- Only 98 Pet Stores

# Finding similar neighborhoods

K-means clustering can be used to group neighborhoods with similar venue types
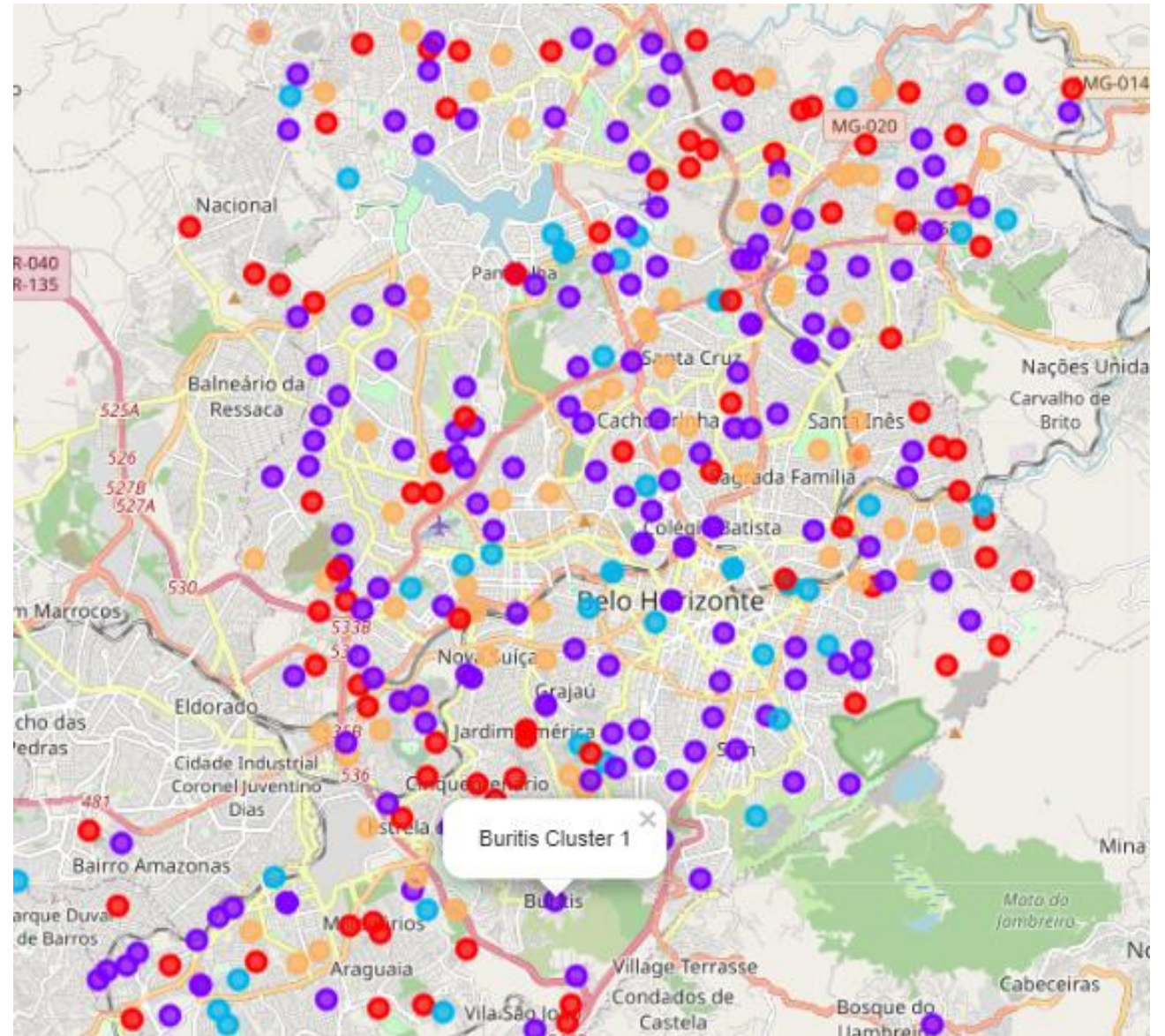
| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Aarão Reis | Ice Cream Shop | Bar | Bakery | Burger Joint | Pub | Park | Gym | Gym / Fitness Center | Supermarket | Restaurant |
| 1 | Acaba Mundo | Plaza | Racecourse | Food Truck | Comfort Food Restaurant | Snack Place | Women's Store | Fast Food Restaurant | Event Space | Fair | Farm |
| 2 | Acaiaca | Pizza Place | Bakery | Pharmacy | Juice Bar | IT Services | Gym / Fitness Center | Motorcycle Shop | Soccer Field | Snack Place | Fruit & Vegetable Store |
| 3 | Ademar Maldonado | Plaza | Bar | Bakery | Fast Food Restaurant | Sushi Restaurant | Ice Cream Shop | Pizza Place | Pet Store | Paper / Office Supplies Store | Soccer Field |
| 4 | Aeroporto | Office | Boutique | Food Truck | Water Park | Gym / Fitness Center | Event Space | Fair | Farm | Farmers Market | Fast Food Restaurant |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 359 | Xangri-lá | Gym / Fitness Center | Mineiro Restaurant | Burger Joint | Plaza | Furniture / Home Store | Restaurant | Bakery | Pharmacy | Snack Place | Fast Food Restaurant |
| 360 | Xodó-Marize | Hot Dog Joint | Bakery | Food Truck | Ice Cream Shop | Plaza | River | Fast Food Restaurant | Fish & Chips Shop | Event Service | Event Space |
| 361 | Álvaro Camargos | Women's Store | Brazilian Restaurant | Bar | Comfort Food Restaurant | Pet Store | Dessert Shop | Fish Market | Farm | Farmers Market | Fast Food Restaurant |
| 362 | Ápia | BBQ Joint | Buffet | Dive Bar | Music Venue | Park | Food Court | Food Stand | Food & Drink Shop | Food | Event Space |
| 363 | Átila de Paiva | Pizza Place | Plaza | Sushi Restaurant | BBQ Joint | Fast Food Restaurant | Bakery | Bar | Ice Cream Shop | Big Box Store | Soccer Field |

# Choosing k with the Elbow Method
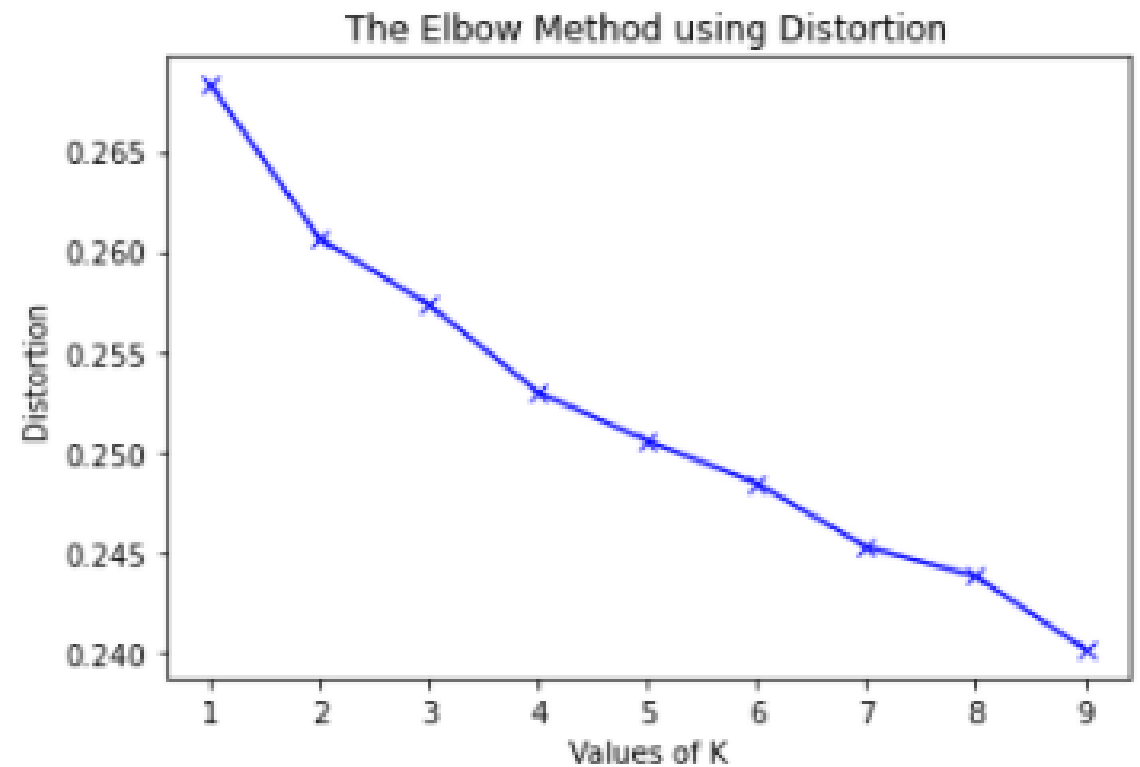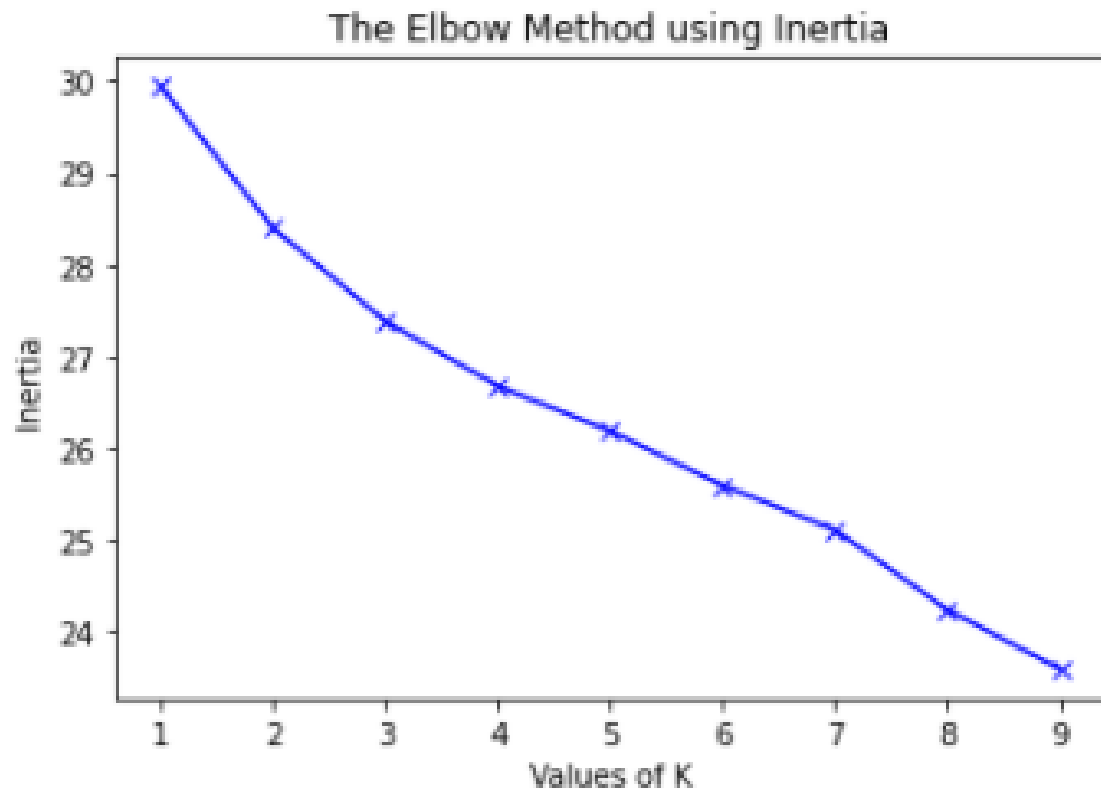
# Clusters results

# Cluster results

With a k=5 Buritis was placed in cluster 1, with a total 166 neighborhoods

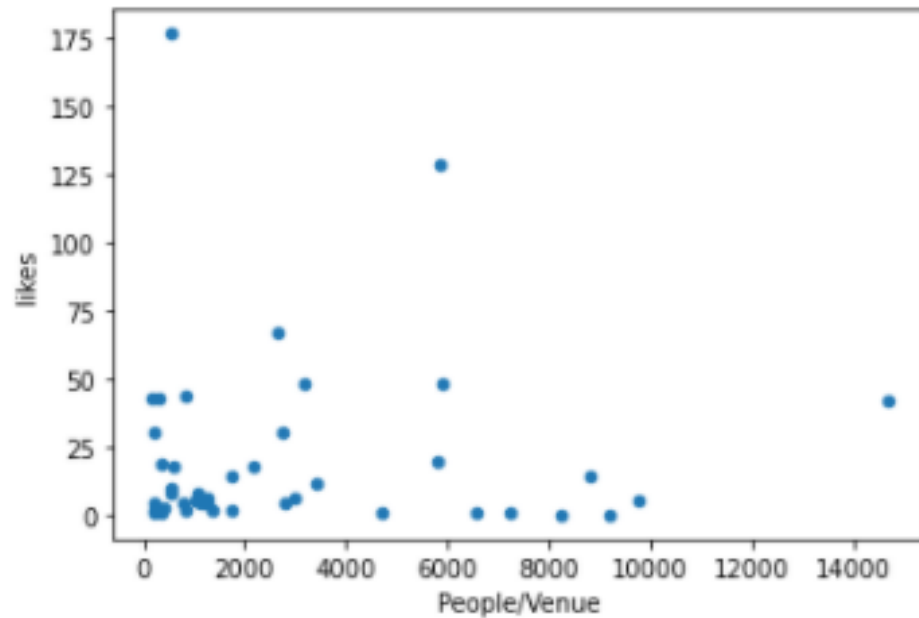| Neighborhood | Size(km²) | latitude | longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Buritis | 3,823\n | -19.976579 | -43.967416 | 1.0 | Gym / Fitness Center | Pharmacy | Pet Store | Park | Ice Cream Shop | Fruit & Vegetable Store | Fast Food Restaurant | Supermarket |
| Serra | 1,536\n | -19.936805 | -43.921403 | 1.0 | Pharmacy | Restaurant | Brazilian Restaurant | Bar | Bakery | Supermarket | Dance Studio | Gym / Fitness Center |
| Sion | 1,195\n | -19.949362 | -43.932587 | 1.0 | Pizza Place | Japanese Restaurant | Gym / Fitness Center | Bar | Brazilian Restaurant | Fruit & Vegetable Store | Plaza | Bakery |
| Pindorama | 1,605\n | -19.922732 | -43.945095 | 1.0 | Bar | Restaurant | Brazilian Restaurant | Café | Gym / Fitness Center | Hotel | Burger Joint | Juice Bar |
| Gutierrez | 1,246\n | -19.934260 | -43.957157 | 1.0 | Gym / Fitness Center | Bakery | Snack Place | Brazilian Restaurant | Pharmacy | Bar | Ice Cream Shop | Food Truck |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Maravilha | 0,004\n | -19.936128 | -44.017068 | 1.0 | Gymnastics Gym | Ice Cream Shop | Soccer Field | Mineiro Restaurant | Supermarket | Event Space | Seafood Restaurant | Electronics Store |
| Xangri-lá | 0,526\n | -19.838104 | -44.018154 | 1.0 | Gym / Fitness Center | Mineiro Restaurant | Burger Joint | Plaza | Furniture / Home Store | Restaurant | Bakery | Pharmacy |
| Campus UFMG | 3,85\n | -19.867975 | -43.964427 | 1.0 | College Cafeteria | Paper / Office Supplies Store | Snack Place | Plaza | Restaurant | Garden | Fast Food Restaurant | Event Service |
| Lagoa da Pampulha | 3,138\n | -19.853383 | -43.974785 | 1.0 | Gym / Fitness Center | Racecourse | Snack Place | Art Museum | Nightclub | Juice Bar | Water Park | Pool |
| Sumaré | 0,149\n | -19.887587 | -43.965484 | 1.0 | Clothing Store | Coffee Shop | Pizza Place | Shoe Store | Chocolate Shop | Steakhouse | Restaurant | Italian Restaurant |

Pet Stores
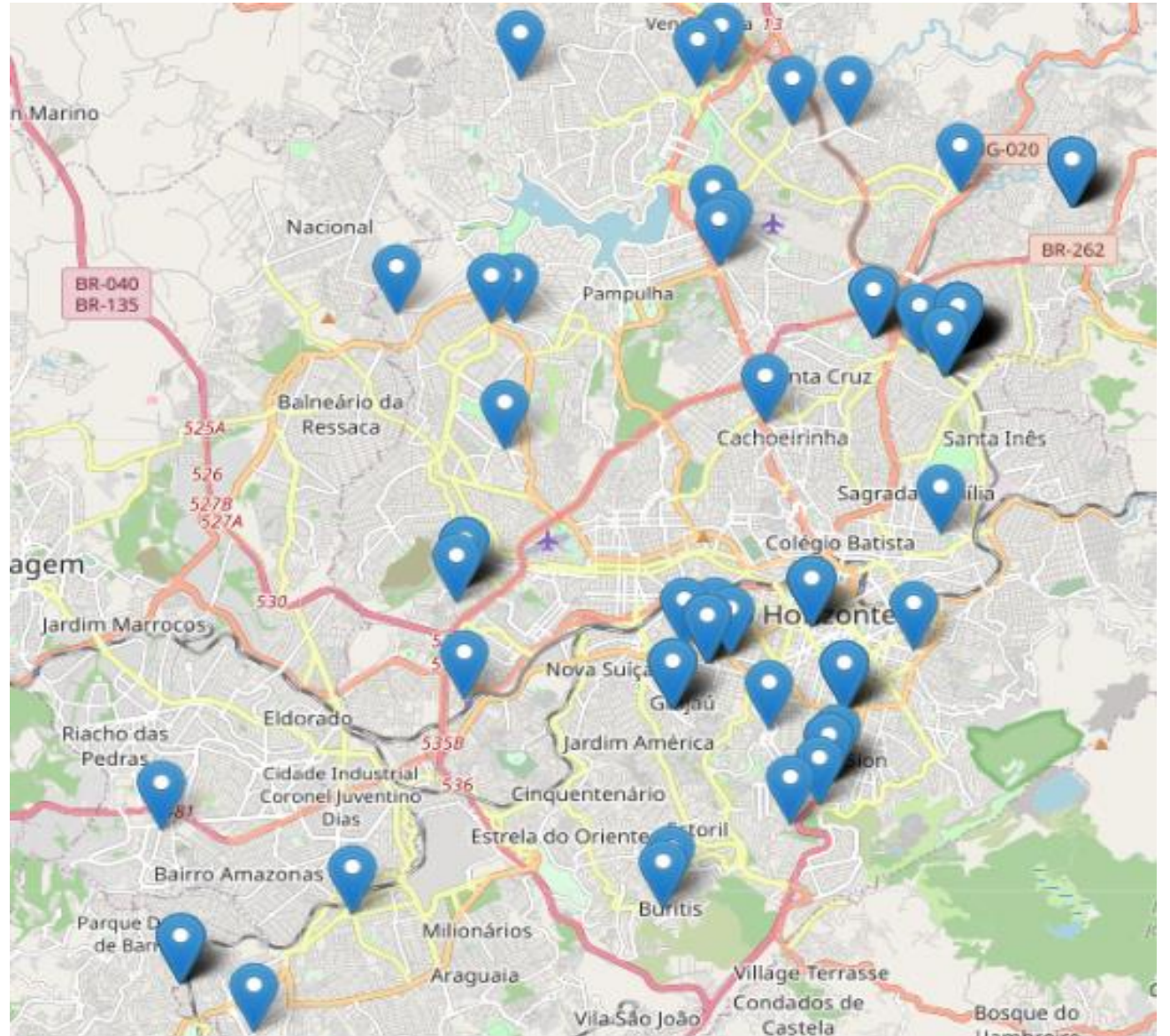
# Likes/person and People/Venue

Both those indicators must be maximized



| | Neighborhood | likes | Venue | Population | latitude | longitude | People/Venue | likes/person |
|---|---|---|---|---|---|---|---|---|
| 28 | Savassi | 48.0 | 2 | 11772 | -19.937246 | -43.935582 | 5886.000000 | 0.004077 |
| 16 | Gutierrez | 129.0 | 3 | 17507 | -19.934260 | -43.957157 | 5835.666667 | 0.007368 |
| 9 | Coração de Jesus | 20.0 | 1 | 5812 | -19.945874 | -43.951229 | 5812.000000 | 0.003441 |
| 17 | Horto | 12.0 | 1 | 3429 | -19.910141 | -43.917867 | 3429.000000 | 0.003500 |
| 30 | São Pedro | 48.0 | 2 | 6328 | -19.943598 | -43.937020 | 3164.000000 | 0.007585 |
| 29 | São Lucas | 6.0 | 1 | 2987 | -19.931156 | -43.921262 | 2987.000000 | 0.002009 |
| 7 | Campo Alegre | 4.0 | 1 | 2776 | -19.831924 | -43.947484 | 2776.000000 | 0.001441 |
| 22 | Nova Granada (ZS) | 30.0 | 2 | 5516 | -19.941317 | -43.969090 | 2758.000000 | 0.005439 |
| 33 | Vila Barragem Santa Lúcia | 67.0 | 3 | 7999 | -19.954831 | -43.943795 | 2666.333333 | 0.008376 |
| 23 | Palmares (ZL) | 18.0 | 2 | 4364 | -19.872856 | -43.930000 | 2182.000000 | 0.004125 |
| 20 | Minas Brasil | 14.0 | 2 | 3511 | -19.922732 | -43.945095 | 1755.500000 | 0.003987 |
| 35 | Vila Oeste | 2.0 | 1 | 1339 | -19.936669 | -44.001984 | 1339.000000 | 0.001494 |
| 0 | Acaiaca | 4.0 | 2 | 2565 | -19.844662 | -43.894791 | 1282.500000 | 0.001559 |
| 39 | Vila São João Batista | 6.0 | 2 | 2486 | -19.821674 | -43.962003 | 1243.000000 | 0.002414 |
| 42 | Átila de Paiva | 4.0 | 1 | 1207 | -19.978006 | -44.028716 | 1207.000000 | 0.003314 |
| 3 | Belmonte | 4.0 | 2 | 2236 | -19.846745 | -43.900156 | 1118.000000 | 0.001789 |
| 12 | Fernão Dias | 8.0 | 4 | 4190 | -19.873126 | -43.917893 | 1047.500000 | 0.001909 |
| 34 | Vila Califórnia | 5.0 | 3 | 3100 | -19.918988 | -44.008290 | 1033.333333 | 0.001613 |
| 2 | Barroca | 44.0 | 4 | 3311 | -19.931433 | -43.963553 | 827.750000 | 0.013289 |
| 40 | Vila Trinta e Um de Março | 4.0 | 2 | 1525 | -19.920415 | -44.000970 | 762.500000 | 0.002623 |
| 24 | Palmares (ZN) | 18.0 | 2 | 1157 | -19.872856 | -43.930000 | 578.500000 | 0.015557 |
| 25 | Penha | 8.0 | 4 | 2226 | -19.875560 | -43.912904 | 556.500000 | 0.003594 |
| 10 | Dom Joaquim | 10.0 | 6 | 3279 | -19.878130 | -43.918880 | 546.500000 | 0.003050 |
| 11 | Estrela | 177.0 | 3 | 1591 | -19.949646 | -43.940475 | 530.333333 | 0.111251 |
| 38 | Vila São Gabriel | 3.0 | 1 | 420 | -19.846534 | -43.912311 | 420.000000 | 0.007143 |
| 8 | Conjunto Lagoa | 19.0 | 2 | 662 | -19.867652 | -43.997944 | 331.000000 | 0.028701 |
| 13 | Flamengo | 1.0 | 1 | 328 | -19.964009 | -44.056256 | 328.000000 | 0.003049 |
| 32 | Vila Aeroporto Jaraguá | 43.0 | 2 | 618 | -19.856431 | -43.958784 | 309.000000 | 0.069579 |
| 21 | Nova Granada (ZO) | 30.0 | 2 | 463 | -19.941317 | -43.969090 | 231.500000 | 0.064795 |
| 15 | Grotinha | 4.0 | 2 | 447 | -19.850518 | -43.892311 | 223.500000 | 0.008949 |
| 5 | CDI Jatobá | 1.0 | 1 | 204 | -19.997196 | -44.040677 | 204.000000 | 0.004902 |
| 37 | Vila Rica | 2.0 | 1 | 200 | -19.855422 | -43.953409 | 200.000000 | 0.010000 |
| 41 | Vila de Sá | 43.0 | 7 | 971 | -19.877371 | -43.919882 | 138.714286 | 0.044284 |

Top 5 neighborhoods
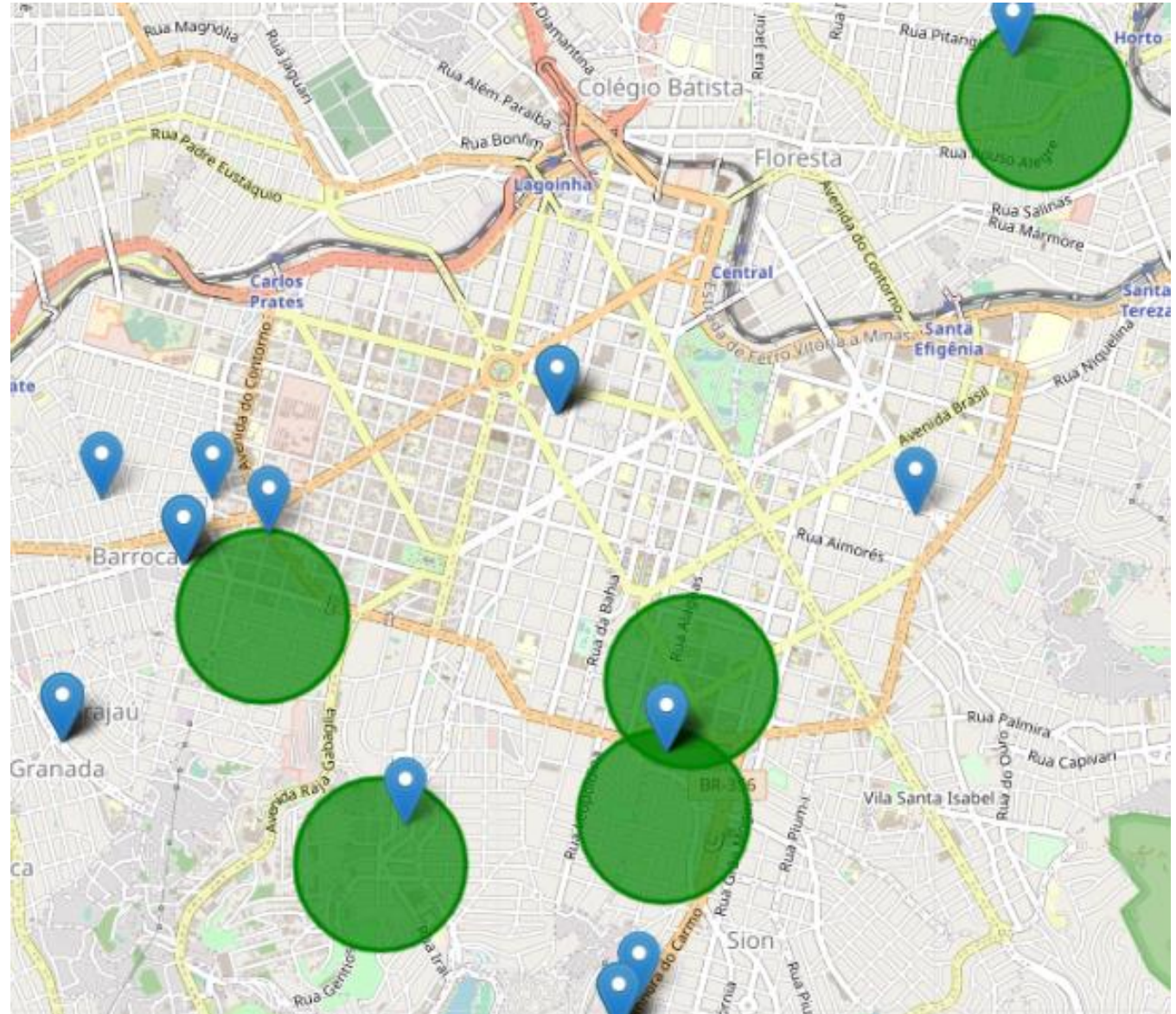
# Conclusion

The number of "likes", albeit sufficient for a basic analysis, is not a good measure of popularity.
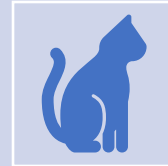
Using machine learning to find the best area to open a new store should be possible if more data about the pet stores are gathered

Given prior knowledge about the city and some of its neighborhoods, the results observed in this work are realistic and would not surprise a resident of the city.
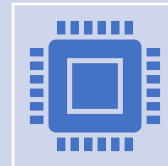
The basis of the work could be used to develop an algorithm to find the best place to open a franchise, given a list of other places where success was found.