

Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Eduardo Gomes Freire

Geomarketing para uma cooperativa de Crédito

Curitiba

2024

Eduardo Gomes Freire

Geomarketing para uma cooperativa de Crédito

Monografia apresentada ao Programa de Especialização em *Data Science* e *Big Data* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Dr. Paulo Ricardo Lisboa de Almeida

Curitiba
2024

Geomarketing para uma Cooperativa de Crédito

Geomarketing for a Credit Cooperative

Eduardo Gomes Freire¹, Paulo Ricardo Lisboa de Almeida²

¹Aluno do programa de Especialização em Data Science & Big Data, edugfreire@hotmail.com

²Professor do Departamento de Informática - DINF/UFPR, paulo@inf.ufpr.br

Este estudo tem como objetivo identificar municípios no estado do Paraná, Brasil, que apresentam maior similaridade para a instalação de agências de cooperativas de crédito, utilizando análise de similaridade com aqueles que já possuem tais unidades de negócio. A similaridade foi determinada utilizando o algoritmo 1-NN com distância euclidiana. Os dados utilizados incluem o número de profissionais de saúde (médicos, enfermeiros, dentistas, farmacêuticos e estabelecimentos de saúde), juntamente com outras métricas como o Índice de Desenvolvimento Humano Municipal (IDHM), a população total distribuída em três faixas etárias (menores de 15 anos, entre 15 e 65 anos e maiores de 65 anos) e o número total de agências bancárias, provenientes da Base de Dados do Estado - BDEweb/IPARDES. Os resultados indicam que o algoritmo é eficiente em identificar similaridades entre os municípios. Observou-se que as métricas relacionadas aos profissionais de saúde, juntamente com outras variáveis, contribuem significativamente para essa proximidade. Esta análise ressalta o potencial de municípios sem agências existentes, validando o uso desse algoritmo para identificar localidades promissoras para a expansão das cooperativas de crédito.

Palavras-chave: Geomarketing, Cooperativa de Crédito, Abertura de Agência, Euclidiana, K-Vizinhos Mais Próximos, Métricas de Distância

This study aims to identify municipalities in the state of Paraná, Brazil, that exhibit the greatest similarity for the establishment of credit cooperative agencies, using similarity analysis with those that already have such business units. Similarity was determined using the 1-NN algorithm with Euclidean distance. The data used includes the number of healthcare professionals (doctors, nurses, dentists, pharmacists, and healthcare establishments), along with other metrics such as the Municipal Human Development Index (IDHM), the total population distributed across three age groups (under 15 years old, between 15 and 65 years old, and over 65 years old), and the total number of bank branches, sourced from the State Database - BDEweb/IPARDES. The results indicate that the algorithm is efficient in identifying similarities among municipalities. It was observed that the metrics related to healthcare professionals, along with other variables, significantly contribute to this proximity. This analysis highlights the potential of municipalities without existing agencies, validating the use of this algorithm to identify promising locations for the expansion of credit cooperatives.

Keywords: Geomarketing, Credit Cooperative, Branch Opening, Euclidean, KNN, Distance Metrics

1. Introdução

Geomarketing é uma abordagem que utiliza inteligência sobre localização para identificar oportunidades de negócios, direcionar campanhas de marketing, segmentar públicos-alvo e atrair clientes, resultando em aumento de lucros. A ideia é que, ao entender padrões geográficos, as empresas possam tomar decisões estratégicas e eficazes [1, 2, 3].

Este estudo tem como objetivo identificar municípios adequados para a instalação de novas agências bancárias ("*unidades de negócios*") da Unicred do Brasil, com foco na área da saúde. Isto porque os profissi-

onais de saúde representam uma parte significativa do setor e têm grande potencial para serviços financeiros.

O Sistema Unicred foi fundado há 35 anos e conta hoje com 30 cooperativas associadas, aproximadamente 356 unidades de negócios em 24 estados brasileiros, e mais de 299 mil cooperados [4]. No estado do Paraná, o Sistema Unicred está presente em 16 municípios: Apucarana, Araçongas, Cascavel, Cianorte, Curitiba, Foz do Iguaçu, Francisco Beltrão, Guarapuava, Londrina, Maringá, Medianeira, Paranaíba, Pato Branco, Ponta Grossa, São José dos Pinhais e Toledo, que são o foco deste estudo.

O desenvolvimento do estudo buscou responder as seguintes perguntas de pesquisa:

1. Os dados públicos disponíveis utilizados são adequados para o problema em questão?
2. Como o algoritmo dos k-vizinhos mais próximos pode identificar similaridades e diferenças entre municípios?

O restante do trabalho está subdividido nas seguintes seções: na Seção 2, é abordado o Estado da Arte relacionado a problemas de geomarketing com foco agências bancárias. Na Seção 3, são apresentados os Materiais e Métodos, que detalham os dados utilizados no estudo. A Seção 4 expõe como os dados foram tratados para possibilitar a realização do estudo. Na Seção 5, são discutidos os resultados obtidos. A Seção 6 é dedicada à discussão do trabalho, enquanto a Seção 7 apresenta a conclusão do estudo.

2. Estado da Arte

O geomarketing tem sido amplamente empregado em várias áreas, especialmente no setor bancário, devido à sua capacidade de analisar dados geográficos para tomar decisões estratégicas comerciais para instalação de novas agências. Sicsú e Crocco (2003) conduziram um estudo pioneiro para identificar as variáveis mais determinantes na distribuição de agências bancárias no Brasil. Eles destacaram que critérios tradicionais de localização industrial não explicam completamente essa distribuição, enfatizando a influência de fatores como PIB e concentração de renda regional, que afetam a demanda por serviços bancários [5].

Em 2012, o Banco do Brasil utilizou técnicas de geomarketing para identificar locais estratégicos para abrir novas agências, levando em consideração variáveis como o fluxo de pedestres, a proximidade de pontos de ônibus e a cobertura de mercado das agências existentes [6].

Ferreira (2015) explorou conceitos avançados de geomarketing para analisar a localização e atratividade das agências bancárias, utilizando modelos como a área de influência e o modelo de Huff para prever o potencial de mercado das agências com base em características demográficas e geográficas das regiões [7].

Araujo (2016) desenvolveu um estudo em Brasília que integrava múltiplos fatores, como acessibilidade, perfil socioeconômico dos clientes, concorrência local e custos operacionais, para apoiar decisões estratégicas na abertura de novas agências bancárias [8].

Recentemente, um estudo realizado em Monte Carmelo-MG por Thaynara & Luziane (2023) analisou a distribuição de serviços comerciais na cidade, destacando a localização central das agências bancárias devido à concentração de atividades econômicas e fluxo de pessoas [9].

Este estudo contribui para o campo ao aplicar princípios de geomarketing para identificar municípios no estado do Paraná adequados para a instalação de agências bancárias. Além das métricas tradicionais, como Índice de Desenvolvimento Humano Municipal (IDHM) e densidade populacional, este trabalho incorpora dados específicos sobre profissionais de saúde, oferecendo uma perspectiva única sobre como essas variáveis podem influenciar a similaridade com municípios que já possuem agências bancárias.

3. Material e métodos

Para este estudo, utilizou-se a Base de Dados do Estado (BDEweb/IPARDES), um sistema de informações estatísticas aberto que compila dados do estado do Paraná e seus 399 municípios [10]. As variáveis selecionadas são referentes ao ano de 2022 e incluem dados de profissionais da saúde (médicos, dentistas, farmacêuticos, enfermeiros e número de estabelecimentos de saúde), demografia (população total dividida em três grupos etários: menores de 15 anos, entre 15 e 64 anos, e maiores de 65 anos), características municipais (quantidade total de residências) e o total de agências bancárias. O Produto Interno Bruto per capita municipal (PIBM) foi utilizado como uma medida de contexto econômico, referente ao ano de 2021.

As informações sobre a localização das agências da Unicred do Brasil nos municípios do Paraná foram obtidas através do site oficial da instituição [11].

4. Pré-Processamento

Para garantir a consistência e comparabilidade dos dados, todas as variáveis foram normalizadas, exceto o IDHM por já estar em escala de 0 a 1. A normalização por mil habitantes foi aplicada dividindo cada variável pela população total do município multiplicada por 1000. Tabela do passo a passo do pré processamento 1

Esta abordagem possibilitou a padronização das variáveis para cada mil habitantes, facilitando a comparação entre os municípios, conforme ilustrado na Tabela 1. Após a normalização por 1.000 habitantes, foi aplicada a técnica de Min-Max scale para ajustar os valores

Dados brutos								
Município	Médicos	Dentistas	Enfermeiros	Farmacêuticos	idade<14	idade>=15 <65	idade>=65	IDHM
Apucarana	492	173	260	49	24622	89372	16140	0.748
União da Vitória	152	70	88	31	10759	38140	6134	0.740

Dados por mil habitantes								
Município	Médicos	Dentistas	Enfermeiros	Farmacêuticos	idade<14	idade>=15 <65	idade>=65	IDHM
Apucarana	3.78	1.33	2.0	0.38	189.2	686.77	124.03	0.748
União da Vitória	2.76	1.27	1.6	0.56	195.5	693.04	111.46	0.740

Min-Max scale								
Município	Médicos	Dentistas	Enfermeiros	Farmacêuticos	idade<14	idade>=15 <65	idade>=65	IDHM
Apucarana	0.51	0.72	0.47	0.19	0.37	0.54	0.57	0.748
União da Vitória	0.37	0.69	0.36	0.29	0.44	0.63	0.47	0.740

Dados com distância euclidean									
Município	Médicos	Dentistas	Enfermeiros	Farmacêuticos	idade<14	idade>=15 <65	idade>=65	IDHM	Distância
Apucarana	0.51	0.72	0.47	0.19	0.37	0.54	0.57	0.748	
União da Vitória	0.37	0.69	0.36	0.29	0.44	0.63	0.47	0.740	0.256

Tabela 1: Pré-processamento

das variáveis para um intervalo entre 0 e 1, Equação 1 e Tabela 1. Esta etapa foi essencial para evitar distorções causadas por diferenças nas escalas das variáveis [12].

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

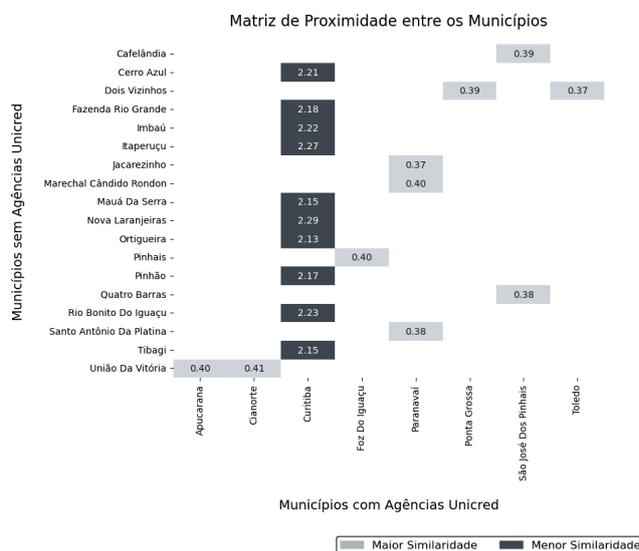


Figura 1: Matriz de distância entre os municípios

Para medir a similaridade entre municípios, foi utilizada a distância euclidiana, uma medida de distância aplicável a espaços com qualquer número de dimensões. Essa métrica é eficaz para comparar a similaridade ou dissimilaridade entre conjuntos de dados com múltiplas variáveis. Tabela 1 A Equação 2 ilustra o cálculo da distância euclidiana. [13].

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (2)$$

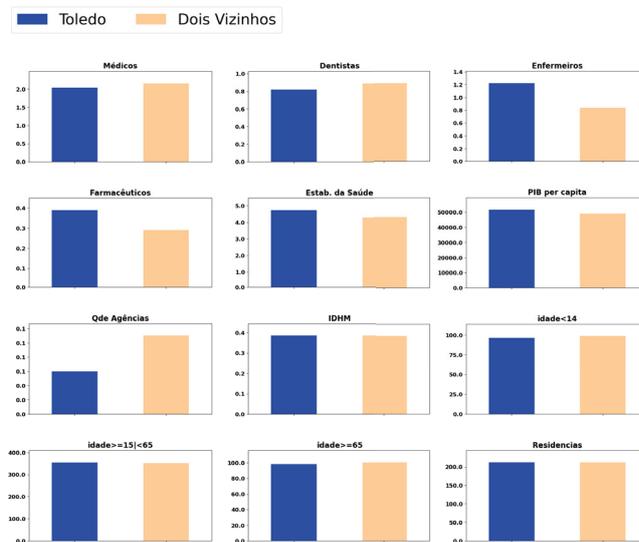


Figura 2: Comparativo Toledo e Dois Vizinhos

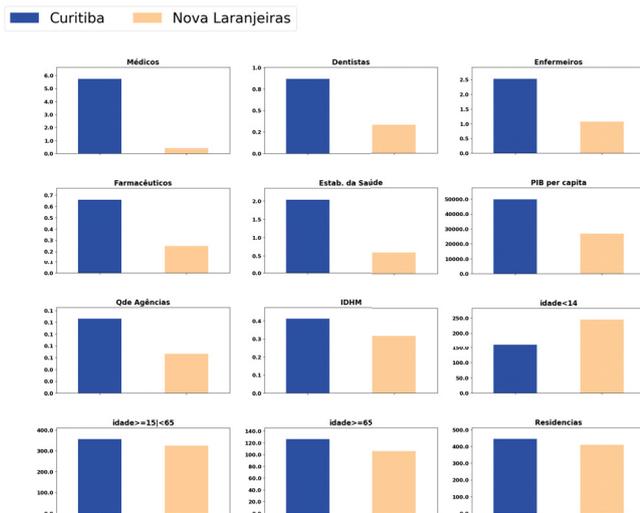


Figura 3: Comparativo Curitiba e Nova Laranjeiras

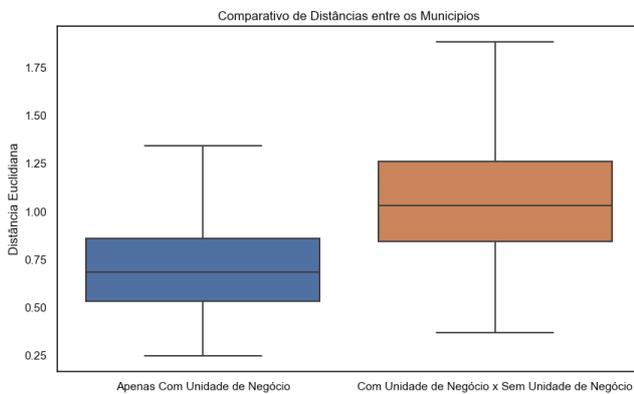


Figura 4: Boxplot Comparativo

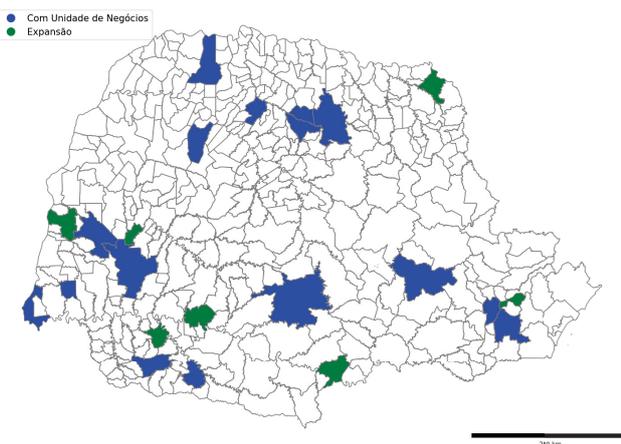


Figura 5: Posição dos municípios no estado do Paraná

Para visualizar a distribuição geográfica e identificar os municípios mais próximos e mais distantes no estado do Paraná, utilizou-se a biblioteca matplotlib [14]

e geopandas [15], em conjunto com a malha municipal disponibilizada pelo IBGE [16].

5. Resultado

A matriz de distância entre municípios foi construída utilizando a distância euclidiana, considerando vinte valores de proximidade: os dez mais próximos e os dez mais distantes (Figura 1). Foi selecionado um par com maior similaridade (menor distância euclidiana) e outro com menor similaridade (maior distância euclidiana). O eixo y representa os municípios sem unidades de negócios, enquanto o eixo x representa os municípios com unidades de negócios.

Entre os dez municípios mais próximos, destaca-se o par Dois Vizinhos e Toledo, com uma distância euclidiana de 0.37. A análise detalhada desses municípios está apresentada na Figura 2, que mostra os valores por mil habitantes das variáveis analisadas.

As variáveis que mais contribuíram para a similaridade entre Toledo e Dois Vizinhos foram o Índice de Desenvolvimento Humano Municipal (IDHM), a distribuição etária da população, o número de residências, e os profissionais de saúde, como médicos e dentistas, além dos estabelecimentos de saúde. Esses fatores apresentaram valores próximos, resultando em uma menor distância euclidiana 2) e demonstrando consistência nos dados entre os municípios.

No outro extremo, Curitiba e Nova Laranjeiras apresentaram uma distância euclidiana de 2.29 (Figura 1). A análise desses municípios (Figura 3) revela que as variáveis relacionadas à saúde, PIB per capita e quantidade de agências apresentaram valores mais discrepantes, justificando a maior distância euclidiana.

Com base nos resultados obtidos, foi possível identificar potenciais municípios para abertura de novas unidades de negócios, que é o objetivo deste estudo. No mapa do estado do Paraná, (Figura 5), os municípios que já possuem unidades de negócios estão destacados em azul, enquanto que os municípios recomendados para abertura de novas unidades de negócios estão em verde. Estes municípios são: Cafelândia, Cerro Azul, Jacarezinho, Marechal Cândido Rondon, Pinhais, Quatro Barras, Santo Antônio da Platina e União da Vitória. 5)

6. Discussão

Para avaliar a adequação dos dados públicos disponíveis para o problema em questão, foi realizada uma

comparação entre os municípios, o primeiro grupo inclui os 16 municípios com unidades de negócios, enquanto que o segundo grupo é composto pelos municípios com unidades de negócios versus os 383 municípios sem unidade de negócios como apresentado na Figura 4. Observou-se que a proximidade entre os municípios do primeiro grupo tende a ser menor em comparação com o segundo grupo, indicando que as variáveis utilizadas são adequadas para o estudo. De fato, o estudo conseguiu identificar 8 municípios com potencial para a abertura de novas unidades de negócios.

Para compreender a similaridade e dissimilaridade entre os municípios utilizando o algoritmo 1-NN foi analisada a semelhança entre os valores das variáveis comparadas. Quanto mais próximas as variáveis entre os municípios, menor é a distância euclidiana. Em contrapartida, quando há discrepâncias significativas entre os valores das variáveis comparadas, a distância entre os municípios tende a ser maior.

7. Conclusão

Com base nos resultados apresentados foi possível utilizar a base de dados disponíveis para identificar locais para abertura de novas unidades de negócios no estado do Paraná. Embora os dados utilizados sejam suficientes para a compreensão do problema, é importante considerar a inclusão de outras variáveis para uma análise mais abrangente. Variáveis adicionais, como concorrência, taxa de juros, taxa de serviços, quantidade de estabelecimentos de ensino na área da saúde, número de alunos e mais profissionais da área de saúde, poderiam enriquecer a análise. Além disso, sugerimos a utilização de outras métricas de distância, como a distância de Manhattan entre outras, para entender a dinâmica e comparar com os resultados apresentados neste trabalho. Isso pode oferecer uma perspectiva mais completa e detalhada sobre a similaridade entre os municípios, aprimorando a precisão das conclusões.

8. Agradecimentos

Gostaria de expressar meus agradecimentos ao meu orientador, Dr. Paulo Ricardo Lisboa de Almeida, pela orientação e apoio durante este estudo. Agradeço, em especial, à minha esposa Priscila Tremarin, e ao meu filho, Guilherme Tremarin Freire, pelo suporte e incentivo ao longo deste processo.

Referências

- [1] Churchill, Jr. G., Marketing - Criando Valor Para os Clientes. 3° ed. São Paulo: Saraiva. 2013.
- [2] Figueredo, W. A. Geomarketing aplicado à instituições educativas. Encontro da associação Nacional de Pós-graduação e Pesquisa em Administração, 30. Rio de Janeiro: ANPAD, 2006.
- [3] Sebra, A. L. C. Geomarketing: Estudo de áreas de influência de um restaurante de healthy food em Natal-RN. Trabalho de Conclusão de Curso (Graduação em Administração) - Universidade Federal do Rio Grande do Norte, Natal. 2014.
- [4] site institucional Unicred do Brasil maio, 2024, url <https://www.unicred.com.br/institucional/sistema-unicred/a-unicred> Acesso em: 02/05/2024.
- [5] Sicsú, J., Cocco, M. Em busca de uma teoria da localização das agências bancárias: Algumas evidências do caso brasileiro. Revista Economia, Rio de Janeiro, v. 4, n. 1, p. 85-112, 2003.
- [6] Silva, F. C., Schmitz, L. K. Geomarketing aplicado à instalação de novas agências do Banco do Brasil em Curitiba, 2012.
- [7] Ferreira, I. M. G., Ferreira, A. S. Análise de uma rede de Agências Bancárias, 2015.
- [8] Araújo, L. B., Granemann, S., R. Localização de Agências Bancárias, 2016.
- [9] Thaynara, F. V, Luziane, R. I. Análise Especial de Geomarketing para setores de comércio e serviços da cidade de Monte Carmelo em: https://www.docentes.univasf.edu.br/carlos.freitas/geometria_analitica/distancia.php. Acesso em: 08/05/2024.
- [10] Iparde - Instituto Paranaense de Desenvolvimento Econômico e Social: url <http://www.ipardes.gov.br/imp/index.php>. Acesso em: 01/03/2024.
- [11] site institucional Unicred do Brasil maio, 2024), url <https://www.unicred.com.br/atendimento/encontre-uma-agencia> Acesso em: 02/05/2024.
- [12] Carvalho, A. C. P. E, Menezes, A. G., BONIDIA, R. P. Ciência de dados, fundamentos e aplicação. Rio de Janeiro, 2024.
- [13] Hair Jr., J. F., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L. Análise Multivariada de Dados, 2009.
- [14] Hunter, J. D. "Matplotlib: A 2D Graphics Environment", Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, 2007.
- [15] Jacob, W., et al. GeoPandas: Python tools for geographic data. Journal of Open Source Software, vol. 3, no. 24, p. 540, 2018.
- [16] Instituto Brasileiro de Geografia e Estatística (IBGE). Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html?edicao=36516&t=acesso-ao-produto>. Acesso em: 02/05/2024.