



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE ENSINO SUPERIOR DO SERIDÓ
DEPARTAMENTO DE COMPUTAÇÃO E TECNOLOGIA - DCT
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

HUGO RAFAEL DE MEDEIROS FERNANDES

**Mineração de Dados aplicada ao Terceiro Setor do
Município de Caicó/RN**

CAICÓ-RN

2022

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE ENSINO SUPERIOR DO SERIDÓ
DEPARTAMENTO DE COMPUTAÇÃO E TECNOLOGIA - DCT
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

HUGO RAFAEL DE MEDEIROS FERNANDES

**Mineração de Dados aplicada ao Terceiro Setor do
Município de Caicó/RN**

Trabalho de Conclusão de Curso apresentado ao
Programa de Graduação em Sistema de Informa-
ção, como parte dos requisitos para obtenção do
título de Bacharel em Sistemas de Informação da
Universidade Federal do Rio Grande do Norte.

Orientador: Professor Dr. Almir Miranda Ferreira

CAICÓ-RN

2022

HUGO RAFAEL DE MEDEIROS FERNANDES

Mineração de Dados aplicada ao Terceiro Setor do Município de Caicó/RN

Trabalho de Conclusão de Curso apresentado ao
Programa de Graduação em Sistema de Informa-
ção, como parte dos requisitos para obtenção do
título de Bacharel em Sistemas de Informação da
Universidade Federal do Rio Grande do Norte.

Aprovado em 07 de janeiro de 2022.

Comissão Examinadora

Professor Dr. Almir Miranda Ferreira
Orientador – UFRN

Professor Dr. Gilson Gomes da Silva
Examinador Interno – UFRN

Professor Me. Humberto Rabelo
Examinador Interno – UFRN

Agradecimentos

Agradeço a Deus por me abençoar com saúde e motivação para concluir este desafio. Agradeço à minha família que sempre acreditou em mim. Agradeço aos amigos que sempre estiveram do meu lado. Agradeço aos professores e funcionários da UFRN que me ajudaram nesta jornada, em especial meu orientador Professor Almir, que me apoiou neste trabalho. Por fim, agradeço a sociedade que me permitiu estudar em uma instituição pública na minha cidade natal.

Espero profundamente conseguir retribuir todo o auxílio que recebi.

Resumo

O presente trabalho através de pesquisa realizada em 2021 com indivíduos residentes no município de Caicó/RN, buscou identificar perfis dos colaboradores (doadores e voluntários) de ONGs e instituições do terceiro setor da região, com o propósito de auxiliar essas entidades na tomada de decisão de estratégias para captação de recursos, além de observar o panorama do terceiro setor no município.

Para isso, o estudo foi dividido em três etapas, sendo a primeira a coleta dos dados, que foi realizada com o auxílio da plataforma de formulários do Google Forms, em seguida os dados foram analisados e tratados com a ajuda da linguagem de programação Python. Por último, para a aquisição dos resultados foram utilizadas técnicas de mineração de dados, que permitiram traçar perfis dos colaboradores, assim como obter informações acerca do terceiro setor no município. O trabalho também buscou o apoio de conceitos de estatística descritiva e inferencial para construção e estratificação da amostra populacional utilizada na pesquisa, que com isso permitiram alcançar um grau de confiança relevante nos resultados.

Palavras-Chave: ONGs; Mineração de Dados; Estatística; ; Python

Abstract

The present work, through research carried out in 2021 with individuals residing in the municipality of Caicó/RN, sought to identify profiles of collaborators (donors and volunteers) of NGOs and third sector institutions in the region, with the purpose of assisting these entities in decision making. strategies for raising funds, in addition to observing the panorama of the third sector in the municipality.

For this, the study was divided into three stages, the first being data collection, which was carried out with the help of the Google Forms platform, then the data were analyzed and processed with the help of the Python programming language. Finally, for the acquisition of results, data mining techniques were used, which made it possible to draw profiles of employees, as well as obtain information about the third sector in the municipality. The work also sought the support of descriptive and inferential statistics concepts for the construction and stratification of the population sample used in the research, which thus allowed reaching a relevant degree of confidence in the results.

Keywords: NGOs; Data Mining; Statistic; Python

Lista de Figuras

2.1	Métodos de Amostragem	14
2.2	Determinação do valor crítico através do intervalo de confiança	16
2.3	Áreas sob a curva da distribuição normal padrão	16
3.1	Modelo de Treinamento e Teste	20
3.2	Exemplos de Conjuntos de Treinamento e Teste	20
3.3	Casos de aplicação do SVM	21
3.4	Caso de aplicação de Regressão Linear	21
3.5	Dados agrupados em três clusters	22
4.1	Setores Censitários de Caicó/RN	25
4.2	Quem ajuda mais e menos o terceiro setor	31
4.3	Efeitos da Pandemia nas Contribuições	31
4.4	Popularidade das Instituições da Região	32
4.5	Avaliação das Entidades da Região	32
4.6	Proporção de Usuários por Entidades	33
4.7	Popularidade das Doações	36
4.8	Bairros que ajudam mais e menos	37
4.9	Análise de Agrupamentos	37

Lista de Tabelas

3.1	Métodos, Tarefas e Técnicas da Mineração de Dados	18
4.1	Tamanho da Amostra	24
4.2	Estimativa da Amostra Mínima por Bairros de Caicó/RN	26
4.3	Atributos Brutos da Base de Dados	28
4.4	Binarização das Questões P9, P11, P20 e P21	29
4.5	Principais Resultados Gerais	30
4.6	Perfis por Entidades	34
4.7	Bairros que avaliaram mal as entidades	35
4.8	Doações realizadas no último ano	36
4.9	Principais Características dos Grupos	38

Sumário

Capítulo 1:	Introdução	9
1.1	Contextualização e Problema	9
1.2	Objetivos da Pesquisa	11
1.2.1	Objetivo Geral	11
1.2.2	Objetivos Específicos	11
1.3	Delimitação do Estudo	11
1.4	Trabalhos Relacionados	12
1.5	Motivação e Justificativa	12
Capítulo 2:	Estatística	13
2.1	Técnica de Amostragem	14
2.1.1	Amostragem Estratificada	14
2.2	Tamanho da Amostra	15
Capítulo 3:	Mineração de Dados	17
3.1	Métodos, Tarefas e Técnicas da Mineração de Dados	18
3.1.1	Descrição	19
3.1.2	Predição: Classificação e Regressão	19
3.1.3	Agrupamento (Análise de Clusterização)	22
3.1.4	Regras de Associação (ou Mineração de Relações)	22
Capítulo 4:	Metodologia	24
4.1	Coleta dos Dados	24
4.2	Análise e Tratamento dos Dados	27
4.3	Apresentação dos Resultados	28
Capítulo 5:	Conclusão	39
Referências Bibliográficas		41
Apêndice A: Questionário		44
Apêndice B: Modelo do Convite da Pesquisa		53

1 Introdução

1.1 Contextualização e Problema

As organizações não governamentais (ONGs) são entidades privadas da sociedade civil, sem fins lucrativos, cujo propósito é defender e promover causas de diversos tipos, sejam direitos humanos, animais, indígenas, luta contra o racismo, meio ambiente, entre muitos outros, como cita Blume e Marmentini (2017).

No entanto, de acordo com Silveira (2006), no Brasil a base jurídica mais sólida de que trata das ONGs é a Lei 9.790/1999, sendo que essa lei se refere às organizações da sociedade civil de interesse público (OSCIPs), mas não chega a mencionar o termo organização não governamental. Mais recentemente, a Lei 13.019/2014, conhecida como o Marco Regulatório das Organizações da Sociedade Civil (MROSC) estabeleceu as regras para parcerias entre poder público e as chamadas organizações da sociedade civil (OSCs). Legalmente, portanto, o termo ONG não existe no Brasil. Mesmo assim, ele é utilizado corriqueiramente pela população, imprensa e governo. Blume e Marmentini (2017) lembram que há outras nomenclaturas que também geram confusão quando queremos nos referir às ONGs, que são elas: instituto, entidade, organização sem fins lucrativos (OSFL), organização filantrópica, OSCIP, dentre outras.

Conforme explica Tozzi (2015), essas organizações são parte do terceiro setor, grupo que abarca todas as entidades sem fins lucrativos. Bem como Castro (2017) acrescenta, que a expressão terceiro setor é resultado de uma divisão criada pelo governo dos Estados Unidos, em que o primeiro setor é constituído pelo Estado, o segundo setor pelos entes privados que buscam fins lucrativos – ou seja, o mercado – e o terceiro setor é formado pelas organizações privadas sem fins lucrativos prestadoras de serviços públicos (popularmente chamadas de ONG's).

O termo ONG foi usado pela primeira vez em uma resolução do Conselho Econômico e Social da Organização das Nações Unidas (ONU) em 1950, após a Segunda Guerra Mundial. Já no Brasil, o surgimento das ONGs está diretamente ligado aos movimentos sociais que se organizaram no período da Ditadura Militar, principalmente a partir da década de 1970. Dessa forma, as ONGs surgiram para suprir demandas que não eram atendidas de modo satisfatório pelos Estados, lidando com questões que todos os governos não queriam ou não eram capazes de tratar, segundo o trabalho de Ferreira (2005b).

Em 2010, haviam 290 mil fundações privadas e associações sem fins lucrativos no Brasil, voltadas, predominantemente, à religião (28,5%), segundo a Associação Brasileira de Organizações Não Governamentais (ABONG). Ainda segundo a pesquisa ABONG (2010), dessas instituições, 72,2% (210,0 mil) não possuíam sequer um empregado formalizado, apoiando-se em trabalho voluntário e prestação de serviços autônomos, sendo intensa a presença feminina (62,9%). Já em 2021 com dados da plataforma Mapa das Organizações da Sociedade Civil - IPEA (2021a), mostram que atualmente existem 781.921 OSCs em atividade no país.

Assim, como aborda Blume e Marmentini (2017), uma das maiores dificuldades enfrentadas por essas instituições são os desafios para se manter, como não possuem fins lucrativos, as ONGs precisam buscar formas alternativas para continuar em atividade. Para isso, contam com o apoio financeiro de outras entidades privadas (do terceiro setor ou do empresariado), de pessoas físicas (cidadãos conscientes e engajados) e até mesmo do governo, que pode fornecer recursos e apoio para atividades em que ambos possam unir forças.

Segundo dados da pesquisa Foundation (2018), o Índice de Solidariedade Mundial (World Giving Index), que mede o nível de solidariedade e engajamento social da população dos 144 países pesquisados, o Brasil figura na posição 122^o levando em consideração 3 quesitos: Ajudar um Estranho (105^o), Doar dinheiro para uma ONG (112^o) e se voluntariar para uma causa social (109^o). Complementa Borges (2017), que em todos os quesitos, em termos proporcionais, o Brasil não figurou nem entre os 10 primeiros países, ficando atrás até de países com PIB mais baixo.

Já conforme o relatório Giving (2020), o retrato da doação no Brasil, 78% dos brasileiros doaram dinheiro nos últimos 12 meses, sendo predominantemente apoiando organizações religiosas (49%), onde o valor médio anual gira em torno de R\$617,00. Ainda de acordo com o relatório, os brasileiros tentem a doar mais dinheiro (65%) do que realizar trabalho voluntário (53%). Finalmente, o relatório ainda destaca que mais de 80% dos brasileiros concordam que as organizações sociais têm um impacto positivo na sociedade e que empresas também têm um importante papel de apoiar a comunidade onde atuam.

Nesse contexto, a situação das ONGs se agrava com a chegada do novo coronavírus ao Brasil, como citam Gaglioni (2021) e Prete (2021), que nos primeiros três meses de pandemia, a resposta solidária contou com mais de R\$ 6 bilhões em doações de pessoas físicas e jurídicas, mas que agora o volume de dinheiro da filantropia reduziu drasticamente, chegando a valores de R\$ 19 milhões entre janeiro e fevereiro de 2021.

Se tratando da realidade no Município de Caicó, segundo IPEA (2021b) os setores das ONGs mais presentes são os relacionados a igrejas (ou organizações religiosas), a assistência aos animais de rua, auxílio aos dependentes químicos, e cuidados com pessoas idosas e portadoras de necessidades especiais. Em entrevistas feitas nos meses de Novembro e Dezembro de 2020, com dirigentes (diretores), colaboradores (doadores e voluntários)

e parceiros (funcionários) de ONGs que prestam serviços na região, foi percebido uma enorme carência de recursos e dados, muitas vezes não sendo capazes de manter os serviços mais básicos, e nem mesmo de saber identificar um simples perfil dos seus colaboradores.

Assim, diante do exposto, percebe-se a dificuldade que essas entidades enfrentam para se manterem, e da carência de dados sobre seus colaboradores. Para buscar entender melhor esses problemas, este trabalho busca responder às seguintes questões de pesquisa:

Questão de Pesquisa (Q1): Os resultados da análise dos dados coletados ajuda na tomada de decisão do trabalho desenvolvido pelas ONGs da Cidade de Caicó/RN? Esses dados podem ajudar na captação de recursos para essas entidades?

Questão de Pesquisa (Q2): É possível traçar perfis de quem está disposto a colaborar, de alguma forma, com instituições beneficentes no município de Caicó/RN? É possível identificar o padrão de renda dessas pessoas, os setores da cidade onde residem, sua escolaridade, gênero, faixa etária, como costumam ou preferem ajudar, quais instituições conhecem e como as avaliam, dentre outras informações?

Questão de Pesquisa (Q3): Os dados coletados podem descrever um panorama ou retrato do Terceiro Setor no Município de Caicó/RN?

1.2 Objetivos da Pesquisa

1.2.1 Objetivo Geral

Auxiliar, por meio dos resultados obtidos, nas estratégias de captação de recursos desenvolvidas pelas ONGs da cidade de Caicó/RN. Ajudando assim, na tomada de decisão de futuras campanhas dessas instituições.

1.2.2 Objetivos Específicos

- Identificar perfis dos colaboradores (doadores e voluntários) de ONGs no Município de Caicó/RN, como por exemplo, sua faixa de renda, o setor da cidade onde residem, sua ocupação, escolaridade, as instituições que conhecem, dentre outras características que permitam descrever esses indivíduos em grupos;
- Buscar informações acerca dos doadores e entidades do terceiro setor do município de Caicó/RN, obtendo dessa forma, um retrato dos principais aspectos do setor no município.

1.3 Delimitação do Estudo

Este trabalho se concentra em coletar e analisar dados, de forma descritiva e inferencial, sobre o terceiro setor no Município de Caicó/RN com o auxílio de ferramentas

de tecnologia da informação e métodos estatísticas.

1.4 Trabalhos Relacionados

Buscando entender melhor o cenário proposto na delimitação do estudo, procurou-se trabalhos relacionados ao tema.

Tendo isso em mente, nos meses que seguiram a pesquisa para este trabalho foram levantados alguns estudos relacionados aos termos ONGs e terceiro setor, delimitando a região, os quais vale aqui destacar, como por exemplo: O trabalho de Ferreira (2019), que desenvolve um Sistema de Informação Gerencial (SIG) para uma ONG de proteção Animal na cidade de Caicó/RN. Também como o projeto de Germano (2018), que constroi um ambiente de jogos acessíveis para alunos da Associação de Pais e Amigos dos Excepcionais (APAE) de Caicó/RN. Ainda o trabalho de Silva (2015), que investiga a percepção dos alunos concluintes do curso de ciências contábeis a respeito da contabilidade aplicada ao terceiro setor. Como também, o trabalho de Medeiros (2017), que apresenta as dificuldades para se captar recursos no terceiro setor usando como estudo de casos uma Instituição Filantrópica da cidade de Currais Novos/RN, bem como o trabalho de Júnior e Figueiredo (2019), que trás uma análise da Instituição Cáritas Diocesana de Caicó/RN. Sem esquecer do projeto de Oliveira (2019), que implementa um programa de Português para refugiados Venezuelanos na ONG SOS Aldeias Infantis de Caicó/RN, nem o trabalho de Araújo (2018), que analisa a necessidade de fiscalização de ONGs pelo governo em busca de garantir a dignidade da pessoa humana. Por fim o trabalho de Maia (2012), que realiza uma pesquisa sobre a Banda de Música Taiyo Ongakutai, fundada em Caicó/RN e pertencente a ONG Brasil Soka Gakkai Internacional, e sua importância para a comunidade.

Todos os estudos citados delimitam suas pesquisas a uma entidade em particular ou a uma área do conhecimento, entretanto nunca mostrando um panorama do setor. Assim o presente trabalho busca contribuir nessa discursão.

1.5 Motivação e Justificativa

A principal motivação para realização deste trabalho é o forte caráter social que o estudo possui e os possíveis benefícios que possa trazer para as ONGs da região. Como pôde ser visto até aqui, o terceiro setor sofre de inúmeras carências, sejam de dados como também de recursos. Se este estudo for capaz de gerar informações que possam ser úteis para essas entidades, que desenvolvem importantes serviços para a comunidade, então este trabalho alcançou seus objetivos.

Outra motivação para este trabalho, é a aplicação dos conhecimentos adquiridos durante o curso de Sistemas de Informação em um estudo desse porte.

2 Estatística

Neste capítulo é apresentada a base teórica sobre Estatística que foi utilizada para construção deste trabalho.

Conceitos Iniciais

Conforme explica Fernandes (1999), a estatística tem como objetivo fornecer informações (conhecimento) utilizando quantidades numéricas. Assim, a estatística divide o estudo e a análise dos dados em:

1. Obtenção dos dados
2. Descrição, classificação e apresentação dos dados
3. Conclusões a tirar dos dados

Ainda segundo a autora, a 2^a fase é normalmente conhecida por Estatística Descritiva e a 3^a por Estatística Inferencial.

Como Ferreira (2005a) argumenta, a estatística descritiva tem como objetivo a descrição dos dados, sejam eles de uma amostra ou de uma população. Enquanto, a estatística inferencial preocupa-se com o raciocínio necessário para, a partir dos dados, se obter conclusões gerais, uma afirmação acerca de uma população com base numa amostra.

Vale nesse ponto, diferenciar alguns termos frequentemente adotados no vocabulário da estatística. Consoante Assis, Souza e Dias (2019), amostra é um grupo de elementos ou sujeitos selecionados a partir de um grupo maior chamado de população, portanto sendo a amostra um subconjunto da população em estudo.

Outros termos recorrentes são censo e amostragem. É razoável pensar que ao realizar uma pesquisa, se deseja entrevistar todos os membros de uma população, afim de alcançar uma maior precisão. Esse procedimento de coleta de dados de toda a população é chamado de censo. No entanto, na maioria das pesquisas é absolutamente inviável estudar todos os sujeitos de uma população. Ao invés disso, o pesquisador estuda apenas uma amostra. Para isso, se vale de um procedimento chamado amostragem, no qual se constitui essa amostra da população.

2.1 Técnica de Amostragem

Defende Santos (2018) que ao optar por um estudo por amostragem há que definir quais as unidades estatísticas que vão pertencer à amostra, quantas unidades devem ser observadas, como vão ser selecionadas e qual seu grau de precisão pretendido. Assume-se que quanto maior for a dimensão da amostra, melhores serão os resultados obtidos, dessa maneira busca-se o equilíbrio entre precisão de resultados e custo. O uso da técnica de amostragem adequada permite aumentar a precisão sem aumentar os custos.

Figura 2.1: Métodos de Amostragem

Não probabilísticos (empíricos)	Amostragem por conveniência
	Amostragem por julgamento
	Amostragem por quotas
Probabilísticos (aleatórios)	Amostragem aleatória simples
	Amostragem sistemática
	Amostragem estratificada
	Amostragem por conglomerados

Fonte: SANTOS, 2018

A amostragem probabilística a qual é aquela em que todos os membros da população têm uma probabilidade conhecida de serem selecionados para a amostra. Já a amostragem não probabilística, não emprega a teoria da probabilidade para determinação dos sujeitos que farão parte da amostra. A amostragem não probabilística é menos precisa, contudo pode ser útil quando os custos são inconvenientes, como expressa Assis, Souza e Dias (2019).

2.1.1 Amostragem Estratificada

Dentre as técnicas de amostragem, destaca-se a amostragem por estratificação, que segundo Fernandes (1999) consiste em dividir-se as unidades, donde se vai retirar a amostra, por grupos, conhecidos por estratos. Esses grupos são escolhidos por algum interesse especial que se deseja observar ou pelo fato de que dentro de cada grupo as unidades são semelhantes.

Ainda de acordo com a autora, neste tipo de amostra, as unidades da população não têm todas a mesma oportunidade de serem escolhidas. Alguns estratos da população

podem ser representados por excesso.

2.2 Tamanho da Amostra

Definido o método que serão coletados os dados se faz necessário determinar o tamanho da amostra. Triola (2017) cita a Equação 2.1 capaz de encontrar o tamanho da amostra, levando em consideração o tamanho da população, o grau de confiança, a margem de erro, entre outros critérios.

Complementa o autor, que esta equação é recomendada para populações finitas e quando se desconhece medidas de tendência central e dispersão, como média, desvio padrão e variância da população pesquisada.

$$n = \frac{N \cdot p \cdot q \cdot Z_{\alpha/2}^2}{p \cdot q \cdot Z_{\alpha/2}^2 + (N - 1) \cdot E^2} \quad (2.1)$$

Onde :

n = Número de indivíduos na amostra.

N = A estimativa da população.

p^1 = Proporção populacional de indivíduos que pertencem a categoria estudada.

q = Proporção populacional de indivíduos que não pertencem a categoria estudada.

$Z_{\alpha/2}$ = O valor crítico (ou z-score) que é um multiplicador do desvio padrão que corresponde ao intervalo de confiança em uma distribuição normal padrão.

E = A margem de erro.

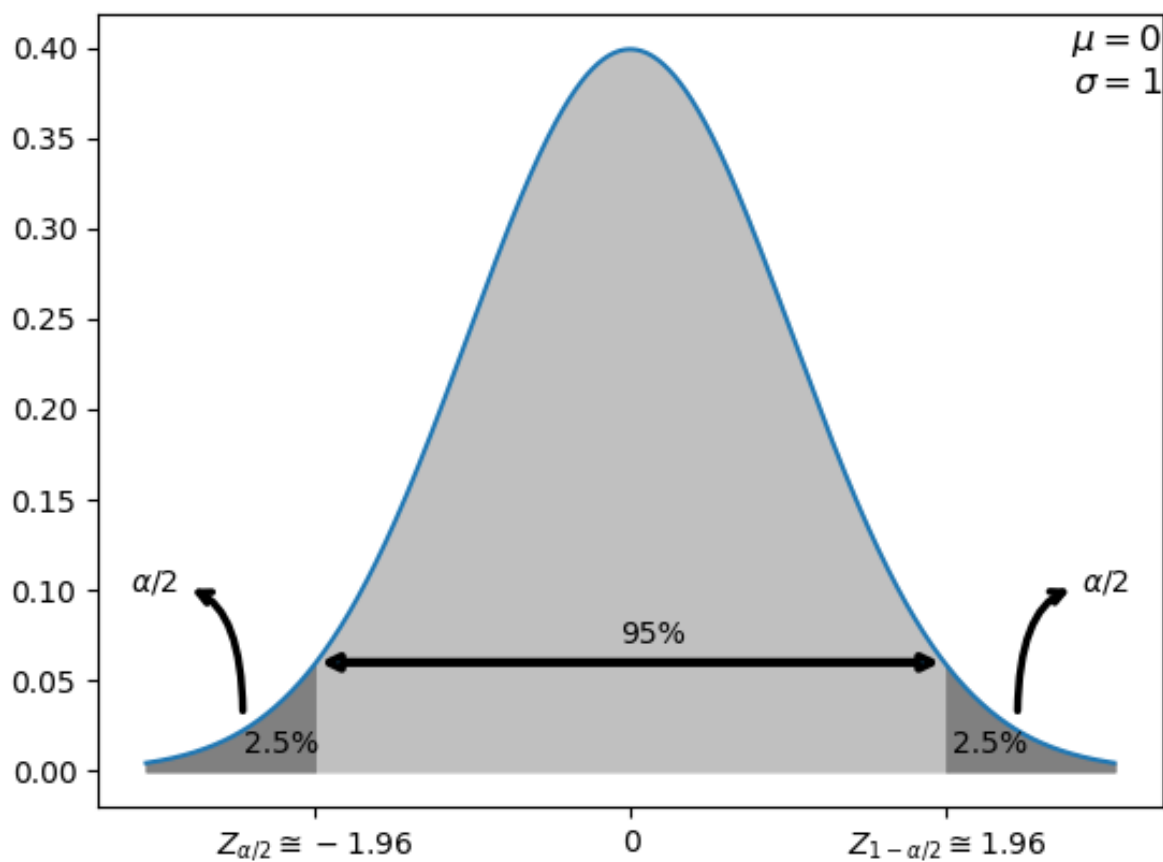
Outro aspecto relevante que vale citar, é em relação ao valor crítico ($Z_{\alpha/2}$). Segundo Freire (2021), para uma população que segue a distribuição normal padrão, dado um valor α ($0 \leq \alpha \leq 1$) se obtém o intervalo ($Z_{\alpha/2}, Z_{1-\alpha/2}$) que conterà com probabilidade $(1 - \alpha)$ o valor de um elemento extraído aleatoriamente dessa população. Assim, para $\alpha = 5\%$, $Z_{0.05/2} = Z_{0.025} \cong -1.96$. Como a curva normal padrão é simétrica em torno de 0, a área sob a curva acima de $Z_{1-\alpha/2} \cong 1.96$, conforme é mostrado na Figura 2.2: Determinação do valor crítico através do intervalo de confiança.

A Figura 2.3: Áreas sob a curva da distribuição normal padrão mostra alguns outros exemplos de intervalos de confiança.

Diante do exposto, os conceitos sobre estatística aqui apresentados foram utilizados neste estudo, principalmente para construção da amostra com o objetivo de alcançar um grau de confiança que permitisse inferir conclusões acerca dos resultados obtidos.

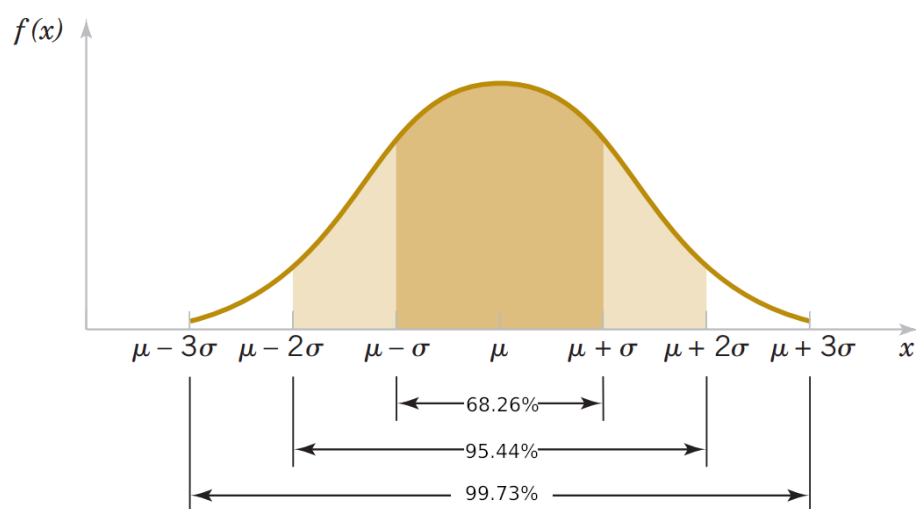
¹Quando não se conhece esse valor, utiliza-se 0.5 tanto para p como q . A razão para atribuir esse valor é que o máximo possível do produto $p \cdot q$ é 0.25, que ocorre quando $p = 0.5$ e $q = 0.5$

Figura 2.2: Determinação do valor crítico através do intervalo de confiança



Fonte: PRÓPRIA, 2021

Figura 2.3: Áreas sob a curva da distribuição normal padrão



Fonte: ZIBETTI, 2021

3 Mineração de Dados

Neste capítulo é apresentada a base teórica sobre Mineração de Dados que foi utilizada para construção deste trabalho.

Conceitos Iniciais

Conforme Amo (2004), mineração de dados é um ramo da computação multidisciplinar, incluindo tecnologias de banco de dados, inteligência artificial, aprendizado de máquina, redes neurais, estatística, reconhecimento de padrões, dentre outras, que teve início nos anos 80, quando os profissionais começaram a se preocupar com os grandes volumes de dados dentro das empresas. Naquela época, o trabalho consistia essencialmente em extrair informações dessas bases de dados da maneira mais automatizada possível. Hoje, no entanto, consiste sobretudo na análise dos dados após a extração, permitindo levantar as necessidades reais e hipotéticas úteis no processo decisório.

Ainda segundo a autora, vale destacar a diferença entre Mineração de Dados e Knowledge Discovery in Database¹ (KDD), que para muitos pesquisadores são considerados sinônimos. Em suma, KDD é um processo mais abrangente consistindo das seguintes etapas:

1. *Limpeza dos dados*: etapa onde são eliminados os ruídos e dados inconsistentes.
2. *Integração dos dados*: etapa onde se permite combinar diferentes repositórios.
3. *Seleção*: etapa onde são selecionados atributos de interesse.
4. *Transformação dos dados*: etapa onde os dados são transformados ou normalizados em um formato adequado à aplicação de técnicas e algoritmos de mineração, reduzir redundâncias, ajudar na integração e desempenho.
5. *Mineração de dados*: etapa onde são aplicadas as técnicas e algoritmos.
6. *Avaliação ou Pós-processamento*: etapa onde são identificados os padrões de interesse.
7. *Visualização dos resultados*: etapa onde são apresentados o conhecimento extraído.

¹Descoberta de Conhecimento em Banco de Dados, em tradução livre

Cada etapa citada é um campo de pesquisa com técnicas, algoritmos e métodos próprios. Como Camilo e Silva (2009) explicam, só o processo de preparação dos dados para a mineração, também chamado de pré-processamento, engloba as etapas de 1 a 4, processo esse que na maioria dos projetos de mineração, compreende até 80% de todo o trabalho.

3.1 Métodos, Tarefas e Técnicas da Mineração de Dados

O objetivo dessa seção é apresentar as principais abordagens utilizadas na mineração de dados, focalizando nas utilizadas neste trabalho, facilitando sua compreensão, percebendo o que cada uma faz e em quais situações aplicá-las.

Para Costa *et al.* (2012) e Camilo e Silva (2009) se categoriza as técnicas de mineração de dados conforme seus métodos e tarefas, como é apresentado na Tabela 3.1: Métodos, Tarefas e Técnicas da Mineração de Dados:

Tabela 3.1: Métodos, Tarefas e Técnicas da Mineração de Dados

Métodos	Tarefas	Principais Técnicas
Aprendizado Supervisionado	Predição Classificação Regressão	Árvore de Decisão Redes Neurais SVM Redes Bayesianas Regressão Linear
Aprendizado Não-Supervisionado	Descrição Agrupamentos Regras de Associação	Análise Exploratória K-Means Apriori

Fonte: PRÓPRIA, 2021

Como cita Camilo e Silva (2009), os métodos de mineração de dados são tradicionalmente divididos em aprendizado supervisionado e não-supervisionado, embora o limite dessa divisão seja tênue, sendo que já existem variações, como abordagens semi-supervisionadas. Ainda de acordo com o autor, a principal diferença entre os dois métodos é que no aprendizado não-supervisionado não é necessário uma pré-categorização dos dados, atributo, rótulo ou classe alvo, enquanto que no aprendizado supervisionado os dados possuem um atributo pré-definido sendo categorizados em torno dele. Vale ainda destacar que todas as técnicas possuem uma maneira de avaliar sua confiança.

Por fim, durante o trabalho de mineração, as diversas técnicas devem ser testadas, combinadas, comparadas e avaliadas afim de que a melhor abordagem seja utilizada.

Daqui em diante, serão apresentados alguns métodos e técnicas de mineração de dados.

3.1.1 Descrição

Camilo e Silva (2009) detalha a descrição como uma tarefa muito utilizada em conjunto com a técnica de análise exploratória, que busca resumir as principais características dos dados, frequentemente com auxílio de métodos visuais, visando maximizar a obtenção de informações ocultas.

Geralmente essa é uma das primeiras abordagens utilizados no processo de obtenção de informações úteis em bases de dados.

3.1.2 Predição: Classificação e Regressão

De acordo com Costa *et al.* (2012), as tarefas de Classificação e Regressão são espécies da tarefa de Predição, cuja meta é desenvolver modelos que façam inferências sobre aspectos específicos dos dados, sendo um modelo preditivo compreendido como uma função $f(X, \beta) \approx Y$, onde X é um conjunto de variáveis preditoras, β são parâmetros desconhecidos e Y é a variável preditiva. Dessa forma, busca-se estimar o valor de Y por meio da descoberta de β utilizando-se X .

Os autores ainda destacam que a tarefa de predição ocorre de maneira supervisionada, sendo fundamental definir um rótulo ou atributo (variável preditiva) para o processo de aprendizagem do modelo, também chamado de treinamento, onde são separados uma proporção dos dados a fim de que o modelo possa "treinar" antes de ser apresentado aos dados os quais se deseja inferir. Esse processo pode ser visto na Figura 3.1: Modelo de Treinamento e Teste e complementado na Figura 3.2: Exemplos de Conjuntos de Treinamento e Teste.

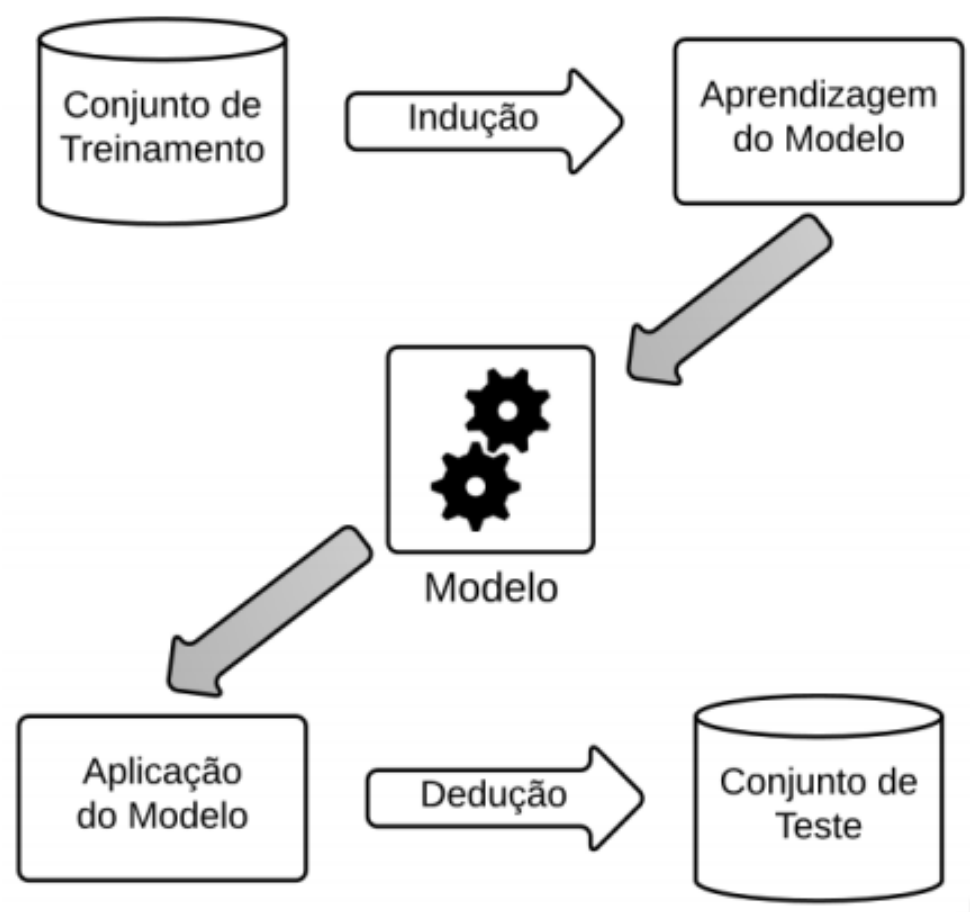
Diante do exposto, vale aqui discutir algumas técnicas de classificação e regressão. Como cita Costa *et al.* (2012), na classificação, as técnicas mais utilizadas são Árvores de Decisão e Support Vector Machine (SVM)², já na regressão, as técnicas mais populares são Regressão Linear, Redes Neurais e SVM.

Support Vector Machine (SVM)

O SVM é uma técnica supervisionado que utiliza um hiperplano, geralmente bidimensional, como separador de classes em uma base de dados. Este hiperplano (linhas tracejadas) é descoberto na fase de treinamento do modelo, sendo seu objetivo encontrar qual o melhor hiperplano que maximize a distância entre as instâncias das classes vizinhas, permitindo assim definir o limite entre os diversos registros afim de que seja possível diferenciá-los, como explica Han, Pei e Kamber (2006). A técnica é apresentada na Figura 3.3: Casos de aplicação do SVM.

²Máquina de Vetores de Suporte, em tradução livre

Figura 3.1: Modelo de Treinamento e Teste



Fonte: COSTA *et al.*, 2012

Figura 3.2: Exemplos de Conjuntos de Treinamento e Teste

Atributo 1	Atributo 2	Atributo 3	Classe	Atributo 1	Atributo 2	Atributo 3	Classe
0,5	Falso	Azul	Sim	0,9	Falso	Branco	?
0,3	Falso	Branco	Não	0,2	Verdadeiro	Azul	?
0,8	Verdadeiro	Azul	Não	0,1	Verdadeiro	Verde	?
0,6	Falso	Verde	Sim	0,3	Falso	Verde	?

Conjunto de Treinamento

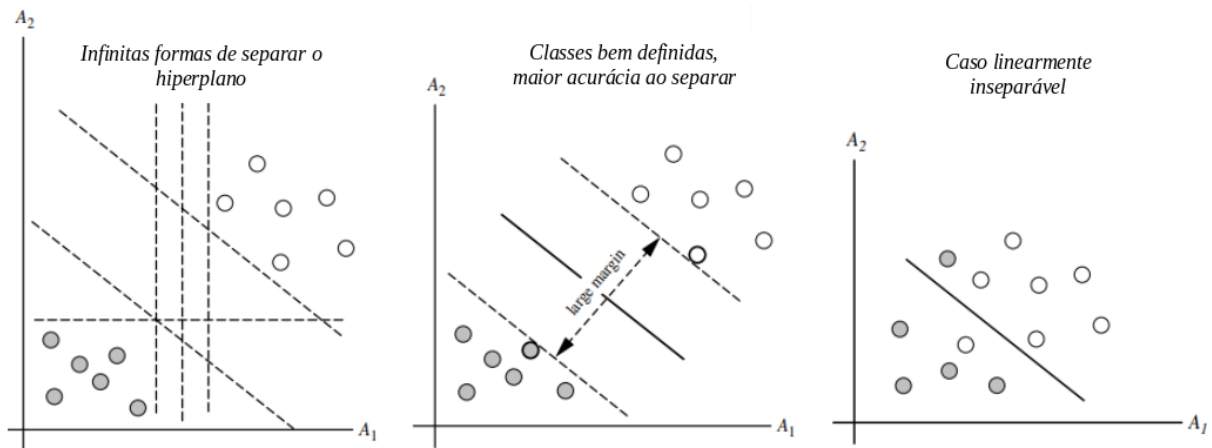
Conjunto de Teste

Fonte: COSTA *et al.*, 2012

Regressão Linear

Segundo Han, Pei e Kamber (2006), a regressão linear é uma técnica supervisionada que segue a equação $y = b + wx$, onde y como uma variável preditiva constante é modelado em função linear de x , a única variável preditora. Ainda também, b e w como sendo coeficientes de regressão, que podem ser encontrados pelo método dos mínimos quadrados, aquele no qual minimiza o erro entre os dados atuais e a estimativa da melhor reta para representa os dados, como é vista nas Equação 3.1 e Equação 3.2.

Figura 3.3: Casos de aplicação do SVM



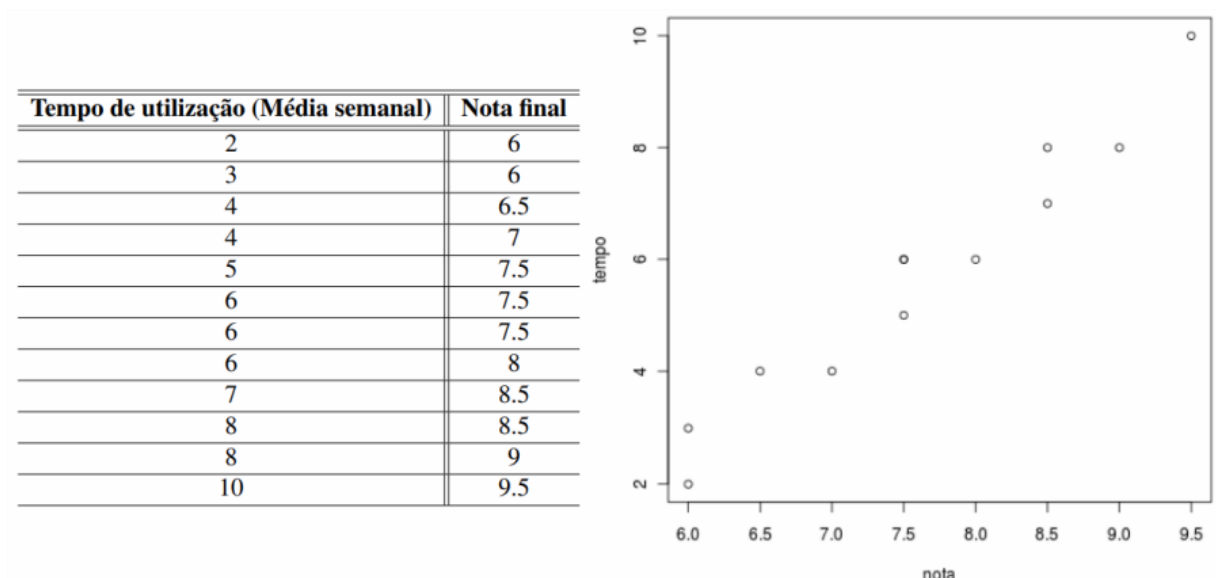
Fonte: HAN; PEI; KAMBER, 2006

$$w = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (3.1)$$

$$b = \bar{y} - w\bar{x} \quad (3.2)$$

A Figura 3.4: Caso de aplicação de Regressão Linear, trás um exemplo de regressão linear aplicado a uma turma de alunos de matemática, em que se relaciona o tempo de utilização do tutor e as notas finais dos estudantes. Analisando o gráfico é possível verificar um padrão linear dos dados.

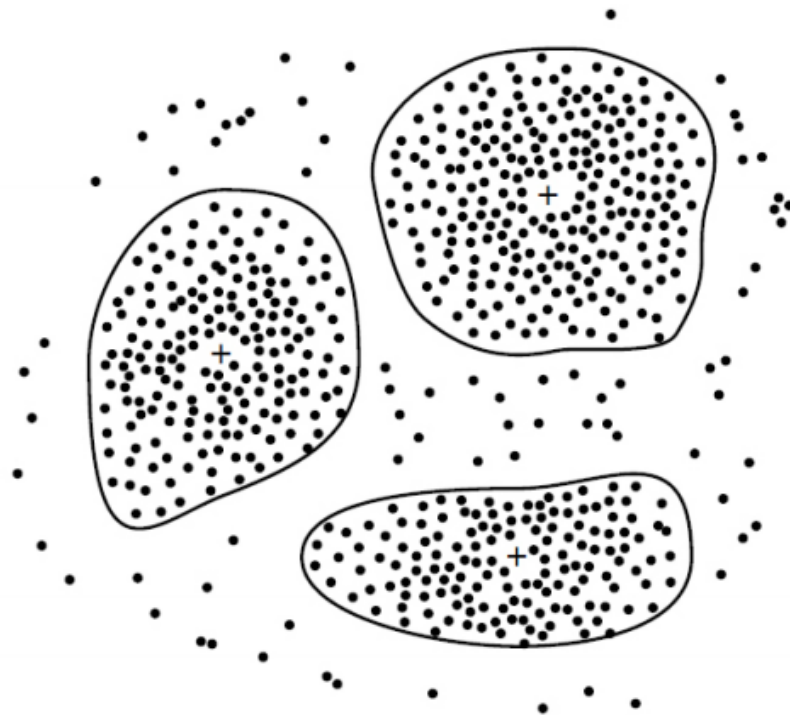
Figura 3.4: Caso de aplicação de Regressão Linear

Fonte: COSTA *et al.*, 2012

3.1.3 Agrupamento (Análise de Clusterização)

Para Camilo e Silva (2009), a análise de clusterização visa identificar os dados similares e os agrupar em coleções (ou *clusters*). Esta tarefa difere da classificação pois não necessita que os registros sejam previamente categorizados, sendo assim um método de aprendizado não-supervisionado. Além do mais, a análise de agrupamentos não tem pretensão de classificar, estimar ou prever, apenas identificar os grupos de dados similares, conforme a Figura 3.5: Dados agrupados em três clusters.

Figura 3.5: Dados agrupados em três clusters



Fonte: CAMILO; SILVA, 2009

Uma técnica bastante utilizado para esta tarefa é o *K-means*. Seu funcionamento consiste em estimar K pontos (ou centroides), marcados com um "+", e a partir daí calcular os dados próximos (ou vizinhos) daquela região. Ao longo da execução, os valores dos K pontos são constantemente atualizados até o critério de convergência ser alcançado e os agrupamentos ficarem bem definidos.

3.1.4 Regras de Associação (ou Mineração de Relações)

Segundo Amo (2004), regras de associação são padrões no formato $X \rightarrow Y$, onde X e Y são conjuntos de registros na base de dados. O objetivo dessa tarefa é buscar associações entre atributos, comportamentos ou demais padrões, afim de identificar uma relação entre os mesmos. Um exemplo clássico da aplicação dessa tarefa são os sistemas

de recomendações presentes em diversos serviços de compras online, nos quais mostram produtos que a maioria dos clientes tendem a adquirir em conjunto.

Como a autora explica, a tarefa de associação pode ser sequencial ou não, sendo que a primeira leva em consideração a ordem cronológica dos itens. Assim, por exemplo, um cliente que compra um carro, tempos depois poderá comprar pneus ou realizar manutenções de seu automóvel, demonstrando uma relação sequencial entre o primeiro e o segundo item adquirido.

Vale ressaltar que regras de associação atribuem duas variáveis para cada relação encontrada, chamadas de *suporte* e *confiança*, que juntas medem o grau de certeza de uma associação. O suporte (*supp*) é a probabilidade $P(X \cup Y)$, no qual representa a porcentagem de transações que a regra $X \cup Y$ se verifica nos dados, como pode ser visto na Equação 3.3. Já a confiança (*conf*) é a probabilidade condicional $P(X \rightarrow Y)$, isto é, a porcentagem de transações contendo os itens de X que também contenha os itens de Y , conforme mostrado na Equação 3.4.

$$supp(X \cup Y) = \frac{\text{Número de Transações com } X \text{ e } Y}{\text{Número de Transações}} \quad (3.3)$$

$$conf(X \rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)} \quad (3.4)$$

Dessa forma, uma regra de associação é dita relevante se $conf \geq \alpha$ e $supp \geq \beta$, em que α e β são respectivamente um grau mínimo de confiança e suporte especificados pelo usuário.

Em síntese, esse processo exige bastante poder computacional, porque dependendo muito do tamanho e das inúmeras relações presentes na base de dados, sendo uma das técnicas mais utilizadas a chamada Apriori.

4 Metodologia

A metodologia para construção deste trabalho foi dividida em três partes: Coleta dos Dados, Análise e Tratamento dos Dados, e Apresentação dos Resultados. As etapas seguiram sequencialmente ao longo do estudo. Por fim, em Capítulo 5: Conclusão são dadas considerações finais sobre o estudo.

4.1 Coleta dos Dados

Tendo delimitado o estudo a população do Município de Caicó/RN, foi necessário descobrir o tamanho dessa população. De acordo com o IBGE (2021), estima-se que a cidade de Caicó conta com 68.343 habitantes, sendo 62.602 (91,6%) na zona urbana e 5.741 (8,4%) na zona rural. Sabendo disso, e com o propósito de se aplicar técnicas estatísticas e mineração de dados, foi adotada dentre as formas de amostragem probabilística existentes, o método da amostragem por estratificação, que assim garante uma maior puerização da amostra pelos 30 bairros da zona urbana. Para este estudo foi preferido não incluir a zona rural, em virtude da extensa quantidade de sítios, povoados e distritos.

Outro aspecto de suma importância que vale mencionar, é sobre o tamanho da amostra. Utilizando o método de Triola (2017) citado na Equação 2.1 foi encontrado o número base de 96 indivíduos para compor o tamanho mínimo recomendado da amostra. Os valores inseridos para alcançar essa amostra podem ser encontrados na Tabela 4.1: Tamanho da Amostra.

Tabela 4.1: Tamanho da Amostra

População da Zona Urbana de Caicó/RN	62.602 habitantes
Tamanho Mínimo da Amostra	96 indivíduos
Intervalo de Confiança	95%
Margem de Erro	10%

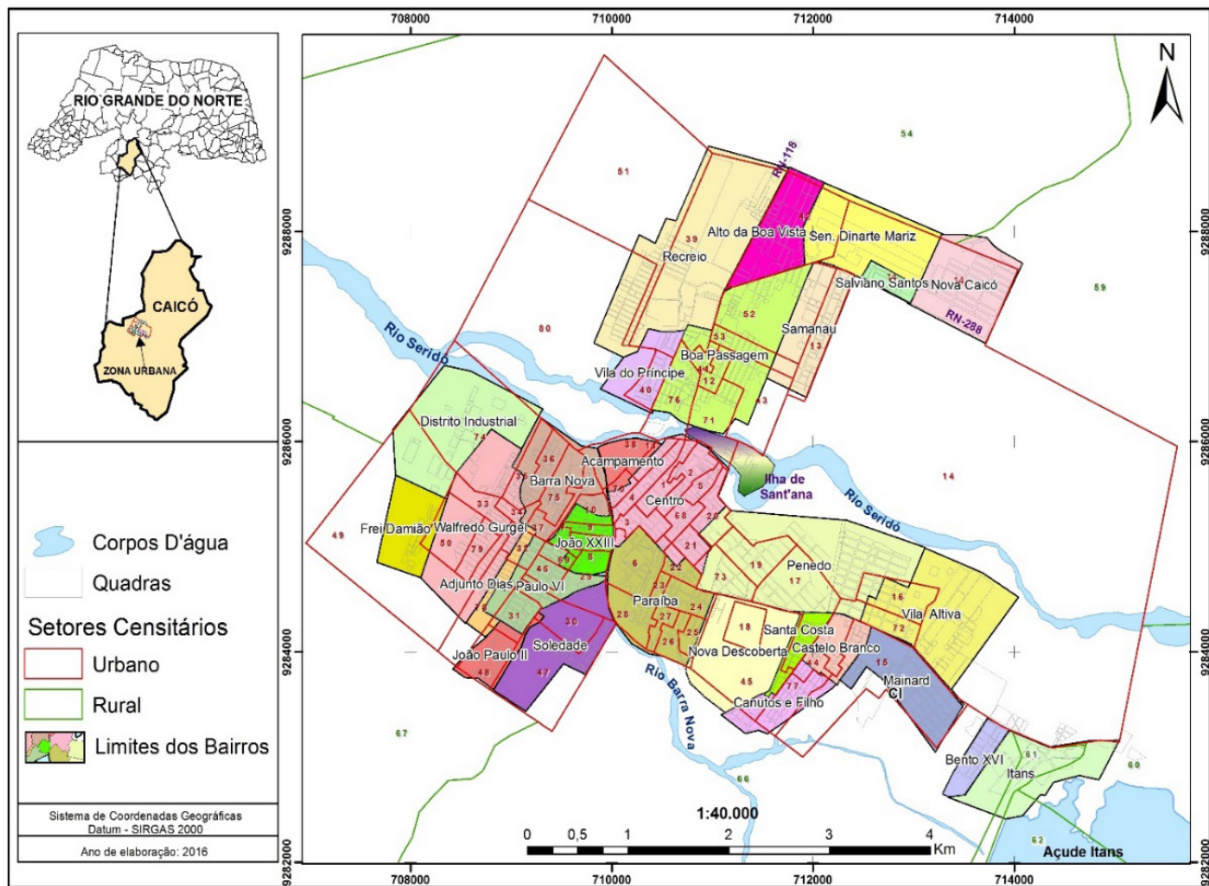
Fonte: PRÓPRIA, 2021

$$n = \frac{N \cdot p \cdot q \cdot Z_{\alpha/2}^2}{p \cdot q \cdot Z_{\alpha/2}^2 + (N - 1) \cdot E^2}$$

$$n = \frac{62602 \cdot 0,5 \cdot 0,5 \cdot 1,96^2}{0,5 \cdot 0,5 \cdot 1,96^2 + (62602 - 1) \cdot 0,1^2} \cong 96$$

Para divisão da amostra pelos bairros do município foi utilizado a base de dados do Cadastro Nacional de Endereços para Fins Estatísticos IBGE e CNEFE (2010), que foi atualizada pela última vez no Censo (2010). Como explica Queiroz, Moraes e Aloufa (2018), o município de Caicó não se encontra oficialmente delimitado em bairros, embora existam os decretos de criação nos quais constam a sua identificação. Dessa forma, o Censo 2010 optou por utilizar uma metodologia de dividir o município em 80 setores censitários, dos quais 12 estão inseridos na zona rural e os demais na zona urbana.

Figura 4.1: Setores Censitários de Caicó/RN



Fonte: Queiroz, Moraes e Aloufa (2018)

Assim, utilizando os dados do CNEFE foi possível fazer uma estimativa da população por bairros da cidade e puerizar a amostra, como é mostrado na Tabela 4.2: Estimativa da Amostra Mínima por Bairros de Caicó/RN.

Tabela 4.2: Estimativa da Amostra Mínima por Bairros de Caicó/RN

Zona	Bairros	Número de Endereços	Amostra Mínima por Bairro ¹	Indivíduos Participantes
Centro	Centro	3902 (16,15%)	$96 \cdot 16,15\% \cong 16$	24
	Acampamento	555 (2,30%)	$96 \cdot 2,30\% \cong 3$	3
Zona Norte	Boa Passagem	2273 (9,41%)	$96 \cdot 9,41\% \cong 10$	11
	Recreio	872 (3,61%)	$96 \cdot 3,61\% \cong 4$	5
	Nova Caicó	683 (2,83%)	$96 \cdot 2,83\% \cong 3$	3
	Alto da Boa Vista	579 (2,40%)	$96 \cdot 2,40\% \cong 3$	5
	Samanaú	514 (2,13%)	$96 \cdot 2,13\% \cong 3$	3
	Vila do Príncipe	386 (1,60%)	$96 \cdot 1,60\% \cong 2$	3
	Darci Fonseca	168 (0,70%)	$96 \cdot 0,70\% \cong 1$	2
	Salviano Santos	148 (0,61%)	$96 \cdot 0,61\% \cong 1$	1
	Serrote Branco ²	306 (1,27%)	$96 \cdot 1,27\% \cong 2$	3
Zona Sul	Paraíba	2562 (10,60%)	$96 \cdot 10,60\% \cong 11$	12
	Soledade	525 (2,17%)	$96 \cdot 2,17\% \cong 3$	3
Zona Oeste	Walfredo Gurgel	1744 (7,22%)	$96 \cdot 7,22\% \cong 7$	9
	Barra Nova	1600 (6,62%)	$96 \cdot 6,62\% \cong 7$	8
	Paulo VI	1253 (5,19%)	$96 \cdot 5,19\% \cong 5$	7
	João XXIII	1171 (4,85%)	$96 \cdot 4,85\% \cong 5$	11
	Adjunto Dias	732 (3,03%)	$96 \cdot 3,03\% \cong 3$	3
	João Paulo II	335 (1,39%)	$96 \cdot 1,39\% \cong 2$	2
	Frei Damião	92 (0,38%)	$96 \cdot 0,38\% \cong 1$	1
Zona Leste	Penedo	1119 (4,63%)	$96 \cdot 4,63\% \cong 5$	8
	Castelo Branco	700 (2,90%)	$96 \cdot 2,90\% \cong 3$	6
	Nova Descoberta	631 (2,61%)	$96 \cdot 2,61\% \cong 3$	4
	Canuto e Filhos	396 (1,64%)	$96 \cdot 1,64\% \cong 2$	4
	Vila Altiva	275 (1,14%)	$96 \cdot 1,14\% \cong 2$	3
	Maynard	243 (1,01%)	$96 \cdot 1,01\% \cong 1$	4
	Itans	219 (0,91%)	$96 \cdot 0,91\% \cong 1$	2
	IPE	88 (0,36%)	$96 \cdot 0,36\% \cong 1$	2
	Santa Costa	61 (0,25%)	$96 \cdot 0,25\% \cong 1$	1
	Bento XVI	29 (0,12%)	$96 \cdot 0,12\% \cong 1$	1
Zona Rural	Diversos Sítios	-	-	2
TOTAL		24161 (100%)	112 indivíduos	156

Fonte: PRÓPRIA, 2021

Dessa maneira, sabendo quantos indivíduos precisariam ser convidados por localidades do município foi aplicado o Apêndice A: Questionário, para coleta dos dados necessários. Vale destacar que os participantes foram todos voluntários do estudo, podendo a qualquer momento recusar participar, e que seus dados foram coletados de forma anonimizada, onde nenhuma das informações são consideradas sensíveis, de acordo com LGPD (2018). Outra informação de suma importância, é que quaisquer dados de menores de 18 anos não foram utilizados neste estudo.

¹Os valores foram arredondados para cima.

²No Censo 2010 não existia o bairro Serrote Branco. Dessa forma, para inclui-lo nesse estudo a estimativa de endereços foi realizada com a ajuda da secretaria de tributação do município que levantou o número de IPTUs nessa localidade.

Por fim, a pesquisa será realizada na modalidade online, no qual os participantes acessam um formulário online e respondem aos itens listados. A construção do formulário foi com o auxílio da ferramenta Google Forms. Já os convites para participar da pesquisa serão impressos e entregues em vias públicas da cidade de Caicó/RN pelo método de panfletagem, onde no Apêndice B: Modelo do Convite da Pesquisa consistirá o link para o questionário online, como também os termos gerais da pesquisa. Além disso o questionário utiliza uma abordagem do tipo estimulada, no qual as questões são de múltipla escolha.

4.2 Análise e Tratamento dos Dados

Nesta fase, todo o processo de análise e tratamento dos dados foi realizado com o auxílio da linguagem de programação Python em sua versão 3.7, e das bibliotecas Pandas e SKlearn, que possui diversas técnicas de mineração de dados, além da Matplotlib em sua versão 3.4.2, que conta com um vasto ferramental para visualização dos resultados.

Os arquivos de código e demais recursos utilizados neste projeto podem ser encontrados no repositório do GitHub, acessando aqui.

Dito isso, os dados coletados que compõe a base de dados que foi estudada nesta pesquisa, contém em sua forma bruta 35 atributos, cada um correspondendo alguma questão presente no formulário online apresentado aos participantes do estudo. Todos os atributos podem ser vistos na Tabela 4.3: Atributos Brutos da Base de Dados.

Outro aspecto que vale destacar, é que algumas respostas de participantes foram removidas do estudo, em especial respostas da questão P2 (relativa a idade), no qual os participantes alegaram serem menores de 18 anos, assim como respostas da questão P7 (relativa ao bairro), onde foi marcada a opção "Zona Rural".

Estando ciente do formato bruto da base de dados, foi iniciado o processo de pré-processamento. A primeira ação realizada foi a remoção dos atributos "time" e "aceitar", tendo em vista que esses atributos não serão utilizados.

Em seguida, questões onde os participantes não responderam, por ser permitido deixar em branco, foram compreendidas como abstenções. Assim, nesses dados faltosos foi atribuído o valor "Abstenção".

As questões P9, P11, P20 e P21 permitem múltiplas opções de respostas. Para tratar isso, foi decidido binarizar essas questões, como pode ser visto na Tabela 4.4: Binarização das Questões P9, P11, P20 e P21. O processo de binarização divide o atributo selecionado em múltiplos atributos de acordo com a quantidade de opções de resposta, adicionando o valor "1" a opção de resposta escolhida e "0" quando contrário. Após esse processo, os atributos referentes as questões citadas acima foram substituídos pelos novos atributos binarizados. Dessa forma, a base de dados foi atualizada de 35 para 66 atributos.

Por último, especificamente o algoritmo de análises de agrupamentos necessitou que os dados fossem convertidos para valores números e posteriormente submetidos ao processo

Tabela 4.3: Atributos Brutos da Base de Dados

Questão do Formulário	Nome do Atributo	Tipo	Escala
Automático	time	qualitativo	nominal
P1	sexo	qualitativo	nominal
P2	idade	qualitativo	ordinal
P3	filhos	qualitativo	nominal
P4	estadoCivil	qualitativo	nominal
P5	escolaridade	qualitativo	ordinal
P6	ocupacao	qualitativo	nominal
P7	bairro	qualitativo	nominal
P8	apae	qualitativo	ordinal
	acapam	qualitativo	ordinal
	aldeiasSOS	qualitativo	ordinal
	abrigo	qualitativo	ordinal
	caritas	qualitativo	ordinal
	fazendaEsperanca	qualitativo	ordinal
	cvv	qualitativo	ordinal
	risoterapia	qualitativo	ordinal
	hemocentro	qualitativo	ordinal
P9	usuario	qualitativo	nominal
P10	ajudaMais	qualitativo	nominal
P10	ajudaMenos	qualitativo	nominal
P11	midias	qualitativo	nominal
P12	dinheiro	qualitativo	ordinal
P13	voluntario	qualitativo	ordinal
P14	alimento	qualitativo	ordinal
P15	roupas	qualitativo	ordinal
P16	higiene	qualitativo	ordinal
P17	racao	qualitativo	ordinal
P18	brinquedos	qualitativo	ordinal
P19	sangue	quantitativo	ordinal
P20	preferencias	qualitativo	nominal
P21	meses	qualitativo	ordinal
P22	pandemia	qualitativo	ordinal
P23	fe	quantitativo	ordinal
P24	renda	qualitativo	ordinal
Termo de Consentimento	aceitar	qualitativo	nominal

Fonte: PRÓPRIA, 2021

de normalização, fazendo com que todos os atributos tivessem a mesma relevância para a análise do algoritmo de agrupamentos. Esse pré-processamento somente foi necessário nessa tarefa.

4.3 Apresentação dos Resultados

A terceira parte consiste em apresentar os resultados e tentar responder as questões de pesquisa.

Esta etapa da pesquisa foram gerados aproximadamente 470 resultados entre gráficos e tabelas dos dados coletados. Esses resultados foram divididos em 5 grupos: Resultados Gerais, Resultados por Entidades, Resultados por Doações, Resultado da Análise de Agrupamentos e Resultados da Análise de Associação.

Tabela 4.4: Binarização das Questões P9, P11, P20 e P21

Questão do Formulário	Nome do Atributo	Valor
P9	APAE Caicó	0 ou 1
	Acapam	0 ou 1
	Aldeias Infantis SOS	0 ou 1
	Abrigo Pedro Gurgel	0 ou 1
	Cáritas Diocesana	0 ou 1
	Fazenda da Esperança	0 ou 1
	CVV Caicó	0 ou 1
	Risoterapia	0 ou 1
	Hemocentro Caicó	0 ou 1
P11	Rádio	0 ou 1
	Blogs e Sites	0 ou 1
	Televisão	0 ou 1
	Jornais e Revistas	0 ou 1
	Carros de Som	0 ou 1
	Igreja	0 ou 1
	Facebook	0 ou 1
	Instagram	0 ou 1
P20	Whatsapp	0 ou 1
	Recolher em casa	0 ou 1
	Transferência ou Depósito	0 ou 1
	Correspondência	0 ou 1
	Pix	0 ou 1
	Troco Solidário	0 ou 1
	Rifas e Sorteios	0 ou 1
P21	Dízimo	0 ou 1
	Janeiro	0 ou 1
	Fevereiro	0 ou 1
	Março	0 ou 1
	Abril	0 ou 1
	Maio	0 ou 1
	Junho	0 ou 1
	Julho	0 ou 1
	Agosto	0 ou 1
	Setembro	0 ou 1
	Outubro	0 ou 1
	Novembro	0 ou 1
	Dezembro	0 ou 1

Fonte: PRÓPRIA, 2021

Resultados Gerais

Os resultados gerais é uma visão supeficial da amostra. Na Tabela 4.5: Principais Resultados Gerais é mostrado os valores mais escolhidos pelos participantes do estudo. Embora supeficial é possível destacar algumas características, em primeiro lugar os meses preferidos para colaboração pelos participantes, Dezembro (final do ano, natal e 13^o salário) e Julho (festa de santana). Ainda se verifica a importância do rádio para a região. E por último a adoção do Pix como principal meio de colaboração.

Além disso, foi perguntado aos participantes da pesquisa, quem eles acreditavam ajudar mais e menos o terceiro setor no município de Caicó/RN. Quase unanimemente

³Os resultados são apresentados aqui, como também em outras tabelas do gênero no formato "absoluto (relativo%)"

Tabela 4.5: Principais Resultados Gerais

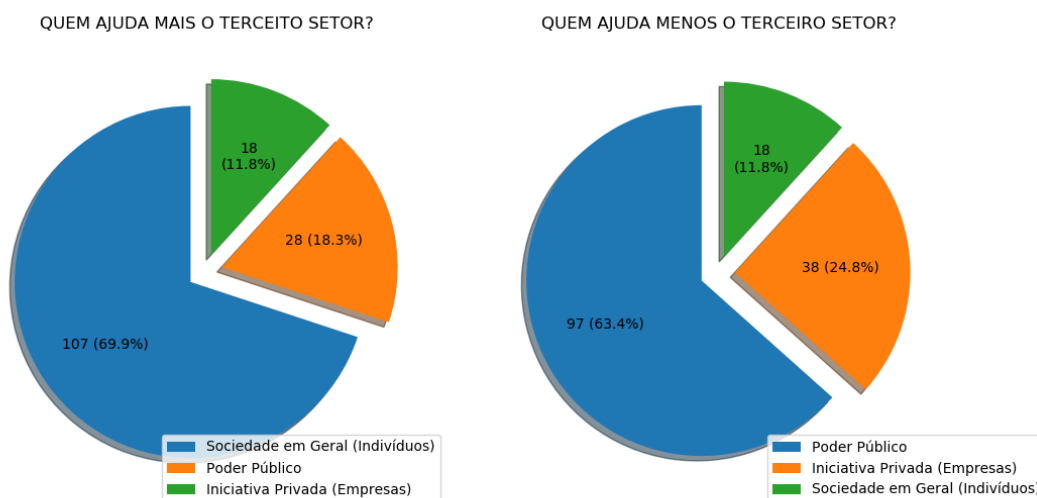
Atributos	Principais Resultados ³
Idade	De 18 a 30 anos : 53 (34.6%) De 31 a 40 anos : 44 (28.8%) De 41 a 50 anos : 32 (20.9%)
Gênero	Masculino : 63 (41.2%) Feminino : 90 (58.8%)
Escolaridade	Ensino Superior Completo : 80 (52.3%) Ensino Médio Completo : 39 (25.5%) Ensino Superior Incompleto : 20 (13.1%)
Ocupação	Servidor Público : 63 (41.2%) Empregado na Inic. Privada : 27 (17.6%) Autônomo : 26 (17%)
Estado Civil	Solteiro : 78 (51%) Casado(a) : 47 (30.7%)
Filhos	Sem filhos : 77 (50.3%) Filhos Biológicos : 72 (47.1%)
Bairros	Centro : 24 (15.7%) Paraíba : 12 (7.8%) Boa Passagem : 11 (7.2%)
Faixa de Renda Domiciliar	De 1 a 3 SM : 68 (44.4%) De 3 a 5 SM : 35 (22.9%) Até 1 SM : 34 (22.2%)
Nível de Espiritualidade	Muito Alto : 47 (30.7%) Médio : 46 (30.1%) Alto : 33 (21.6%)
Meios de Comunicação Preferidos	Rádio : 104 (23.3%) Instagram : 91 (20.4%) Blogs e Sites : 89 (20%) WhatsApp : 51 (11.4%)
Meios de Colaboração Preferidos	Pix : 60 (23.8%) Recolher em casa : 58 (23%) Transferência ou Depósito : 40 (15.9%) Troco Solidário : 34 (13.5%)
Meses Preferidos para Colaboração	Dezembro : 108 (18.1%) Julho : 61 (10.2%) Janeiro : 52 (8.7%)

Fonte: PRÓPRIA, 2021

os membros escolheram a sociedade civil como quem ajuda mais e o poder público como quem ajuda menos. Os demais resultados podem ser vistos na Figura 4.2: Quem ajuda mais e menos o terceiro setor.

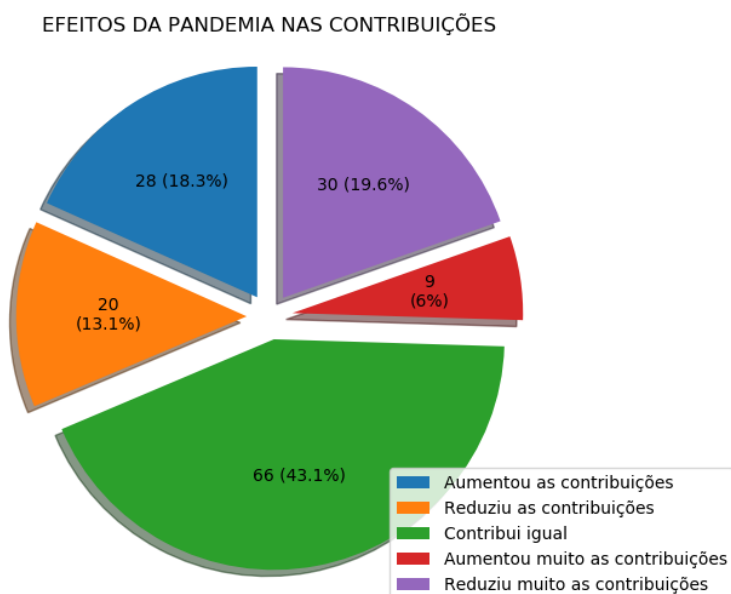
Observou-se também que com a pandemia de covid-19, a maioria dos participantes da pesquisa, 43,1% não alterou seu padrão de contribuições em relação ao ano anterior. Embora a quantidade de pessoas que admitiu reduzi suas contribuições foi de 32,7%, e aqueles que aumentaram atingiu 24,3%. Os valores podem ser vistos em detalhes na Figura 4.3: Efeitos da Pandemia nas Contribuições.

Figura 4.2: Quem ajuda mais e menos o terceiro setor



Fonte: PRÓPRIA, 2021

Figura 4.3: Efeitos da Pandemia nas Contribuições



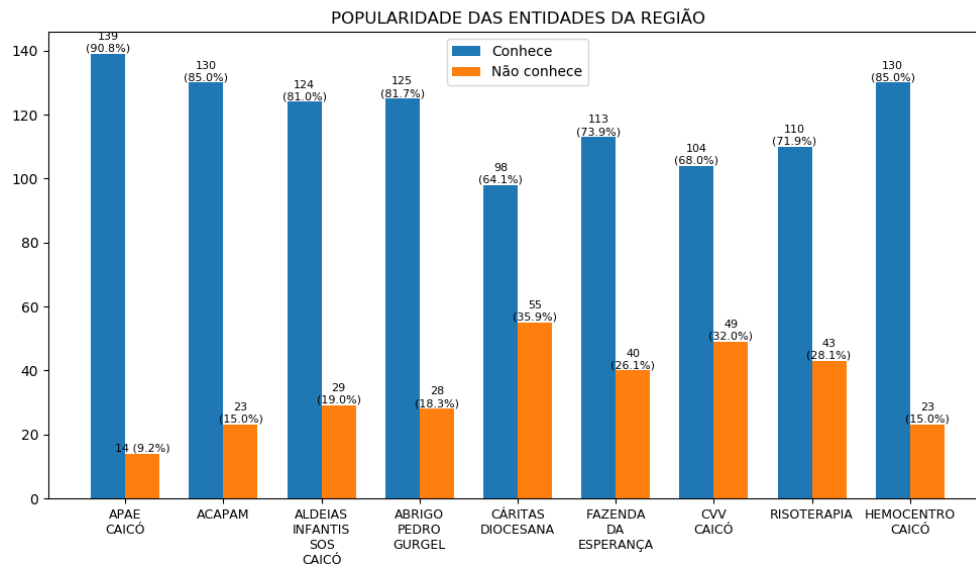
Fonte: PRÓPRIA, 2021

Resultados por Entidades

Outra análise que foi realizada diz respeito as entidades que prestam serviços sociais a comunidade, muitas vezes chamadas de ONGs. Assim, foram escolhidas alguns instituições da região por sua relevância na cidade de Caicó/RN e por contemplar diversas áreas de atuação. Todas essas entidades desenvolvem seus serviços sem fins lucrativos, e a

maioria se mantem com recursos advindos de doações, uma exceção é o Hemocentro Caicó que se trata de um órgão público. Dito isso, a Figura 4.4: Popularidade das Instituições da Região mostra as entidades escolhidas neste trabalho e sua relevância na cidade.

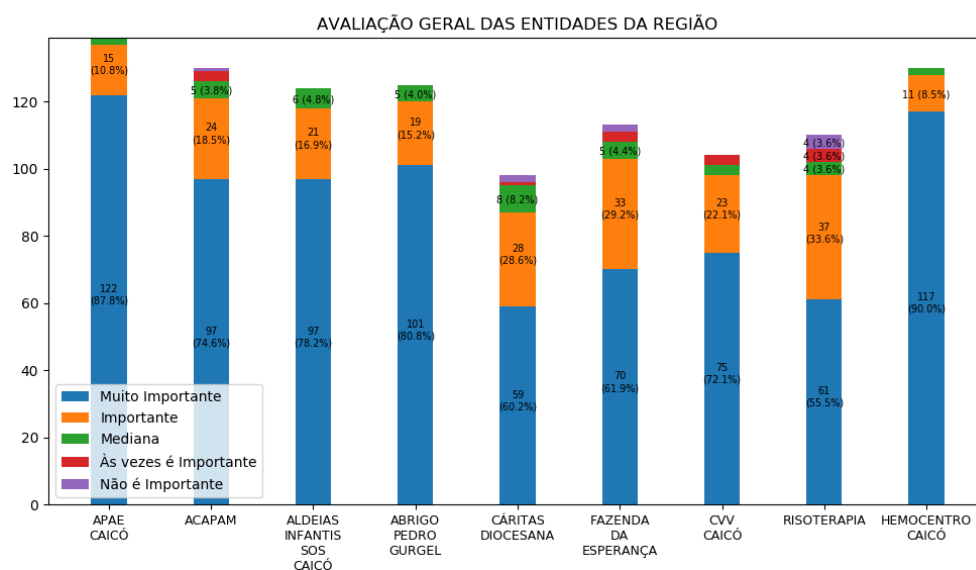
Figura 4.4: Popularidade das Instituições da Região



Fonte: PRÓPRIA, 2021

Em conformidade com o exposto anteriormente, a Figura 4.5: Avaliação das Entidades da Região procura transparecer a percepção dos participantes a respeito dessas instituições.

Figura 4.5: Avaliação das Entidades da Região

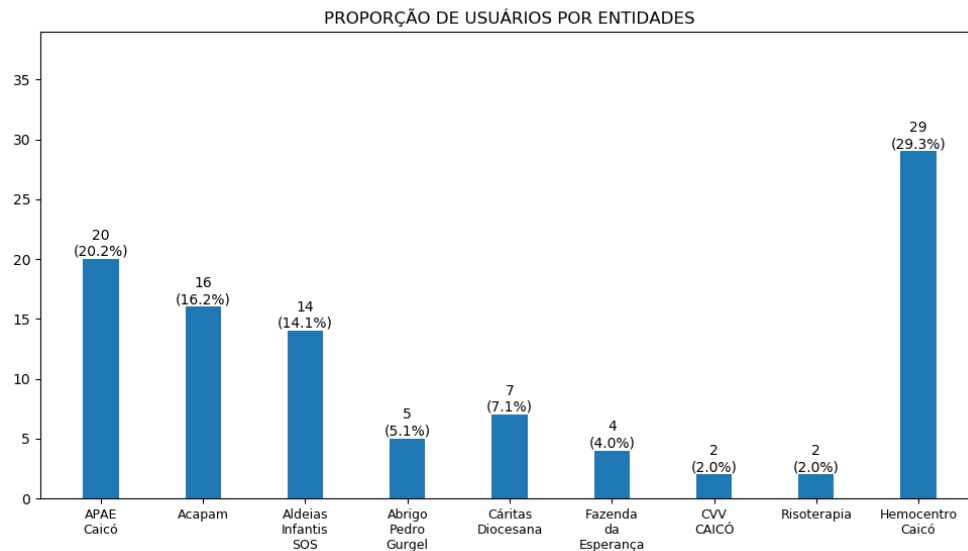


Fonte: PRÓPRIA, 2021

Além disso, a Figura 4.6: Proporção de Usuários por Entidades destaca a quantidade

de participantes que utilizam os serviços dessas instituições.

Figura 4.6: Proporção de Usuários por Entidades



Fonte: PRÓPRIA, 2021

Ainda nos resultados das entidades, a Tabela 4.6: Perfis por Entidades tenta diferenciar as características dos participantes que avaliaram bem e mal as instituições. Aqui, foi considerado uma avaliação benéfica os valores Muito Importante e Importante, as demais opções compõem uma avaliação ruim.

Por último, no geral as entidades escolhidas foram muito bem avaliadas pelos participantes da pesquisa. Além de que se mostraram bastante populares na região. Como forma de verificar os setores de discordância da média, a Tabela 4.7: Bairros que avaliaram mal as entidades ressaltam as regiões da cidade de Caicó/RN que as instituições não tiveram uma boa aprovação de acordo com os participantes da pesquisa. Na tabela é mostrado o bairro e a quantidade de pessoas que avaliaram ruim.

Resultados por Doações

Seguindo com a visualização dos resultados, foram feitas várias análises em relação as doações da amostra. Na Figura 4.7: Popularidade das Doações é mostrado o perfil de contribuição dos participantes da pesquisa.

Ainda em relação as doações, buscou-se encontrar as principais características dos participantes pelo tipo de doação. A Tabela 4.8: Doações realizadas no último ano ilustra os valores de contribuição mais escolhidos pelos participantes, assim como alguns aspectos daqueles que contribuem mais e menos. Essa tabela considera como indivíduos que ajudam mais, aqueles que contribuem com qualquer valor diferente de 0 (Zero).

Ao final da análise dos perfis de doações, mostrou-se que os atributos idade e filhos têm um peso relevante para diferenciar quem costuma ajudar mais ou menos o terceiro

Tabela 4.6: Perfis por Entidades

Entidades	Perfil daqueles que avaliam bem	Perfil daqueles que avaliam mal
APAE Caicó	Servidor Público : 61 (44.5%) Filhos biológicos : 67 (48.9%)	Solteiro : 2 (100%) Desempregado : 1 (50%) Sem filhos : 2 (100%)
Acapam	Servidor Público : 54 (44.6%)	Autônomo : 3 (33.3%) Não usam WhatsApp e Igreja
Aldeias Infantis SOS	De 18 a 30 anos : 42 (35.6%)	De 31 a 40 anos : 3 (50%) Sem filhos : 5 (83.3%)
Abrigo Pedro Gurgel	Feminino : 73 (60.8%) Filhos biológicos : 61 (50.8%) Servidor Público : 54 (45%)	Masculino : 3 (60%) De 18 a 30 anos : 5 (100%) Autônomo : 3 (60%)
Cáritas Diocesana	De 31 a 40 anos : 28 (32.2%)	Utilizam pouco Rádio e Igreja De 18 a 30 anos : 7 (63.6%)
Fazenda da Esperança	Solteiro : 54 (52.4%) Servidor Público : 48 (46.6%)	Casado : 5 (50%) Fé muito baixa : 4 (40%)
CVV Caicó	Feminino : 61 (62.2%) Solteiro : 55 (56.1%) Servidor Público 48 (49%)	Masculino : 4 (66.7%) Casado : 3 (50%)
Risoterapia	Feminino : 60 (61.2%) De 18 a 30 anos : 41 (41.8%)	Masculino : 7 (58.3%) De 31 a 40 anos : 5 (41.7%)
Hemocentro Caicó	De 1 a 3 SM : 61 (47.7%)	Ens. Sup. Completo : 2 (100%) Não utilizam TV, Jornais, Igrejas, Facebook e WhatsApp

Fonte: PRÓPRIA, 2021

setor, sendo pessoas mais jovens e sem filhos contribuindo menos na maioria dos casos. Também se observou que na maioria dos casos, a doação é uma exceção, sendo a regra o indivíduo não ajudar ou ajudar com pequenos valores. Ainda, a Figura 4.8: Bairros que ajudam mais e menos mostra um extrato dos setores da cidade que contribuem muito e pouco com as entidades citadas. Essa figura considera como indivíduos que ajudam mais, aqueles que contribuem com qualquer valor diferente de 0 (Zero) em pelo menos 3 modalidades de doação.

Resultados da Análise de Agrupamentos

Nesta seção, a amostra dos dados foi apresentada ao algoritmo de agrupamentos em 3 situações, nas quais dividiu a amostra em 8, 5 e 4 grupos. A situação que obteve os melhores resultados foi com 4 grupos. A representação gráfica dos grupos ou perfis podem ser vistos na Figura 4.9: Análise de Agrupamentos.

Os grupos 0 e 1 ficaram relativamente próximos, o grupo 2 está um pouco mais isolado, e o grupo 3 só tem um elemento bem distante dos demais, compreendido aqui como um Outlier (um elemento atípico). Na Tabela 4.9: Principais Características dos Grupos é mostrado esses perfis em detalhes.

Comentando um pouco sobre os perfis gerados pela análise de agrupamentos, chamou atenção o grupo 2 por seu perfil, em maioria, formado por homens de meia idade, com

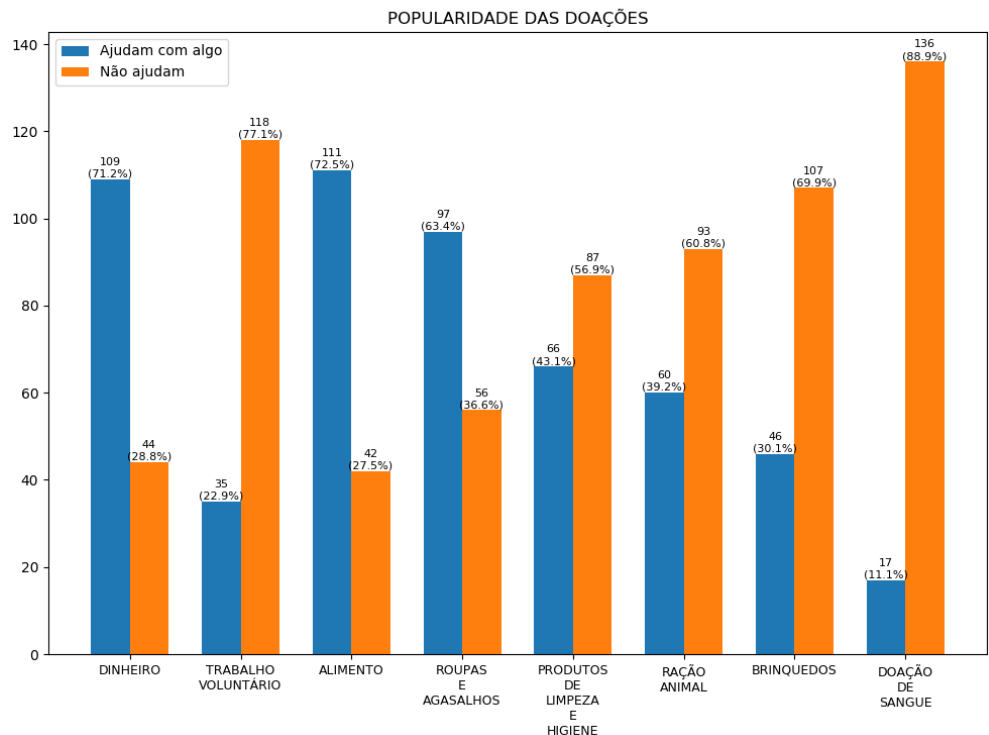
Tabela 4.7: Bairros que avaliaram mal as entidades

Entidades	Bairros que avaliaram mal
APAE Caicó	Walfredo Gurgel : 1 Centro : 1
Acapam	Barra Nova : 1 Centro : 1 Penedo : 1 João XXIII : 1 Acampamento : 1 Vila Altiva : 2 Walfredo Gurgel : 1 Castelo Branco : 1
Aldeias Infantis SOS	Nova Descoberta : 1 Walfredo Gurgel : 2 Centro : 1 Boa Passagem : 1 Serrote Branco : 1
Abrigo Pedro Gurgel	João XXIII : 2 Walfredo Gurgel : 1 Paulo VI : 1 Centro : 1
Cáritas Diocesana	Paraíba : 1 Darci Fonseca : 1 João XXIII : 1 Centro : 2 Acampamento : 1 Vila Altiva : 2 Barra Nova : 1 Maynard : 1 Serrote Branco : 1
Fazenda da Esperança	Paraíba : 1 Paulo VI : 1 Vila Altiva : 1 Barra Nova : 1 João XXIII : 1 Vila do Príncipe : 2 Centro : 1 Boa Passagem : 1 Serrote Branco : 1
CVV Caicó	Vila do Príncipe : 2 Vila Altiva : 1 Barra Nova : 1 Centro : 1 Serrote Branco : 1
Risoterapia	Penedo : 3 Centro : 1 Serrote Branco : 2 Barra Nova : 2 Vila do Príncipe : 3 Darci Fonseca : 1 Nova Descoberta : 1
Hemocentro Caicó	Walfredo Gurgel : 1 Penedo : 1

Fonte: PRÓPRIA, 2021

uma maior renda domiciliar e valores mais elevados nas doações. O grupo 1 se destacou como o perfil mais jovem e com níveis de espiritualidade menores. O grupo 0 manteve índices bem próximos da média geral da amostra, sua principal característica foi um forte desconhecimento do trabalho realizado pelas entidades da região. O grupo 3 por possuir apenas um membro, foi considerado um caso atípico, no qual não se pode inferir muitas conclusões.

Figura 4.7: Popularidade das Doações



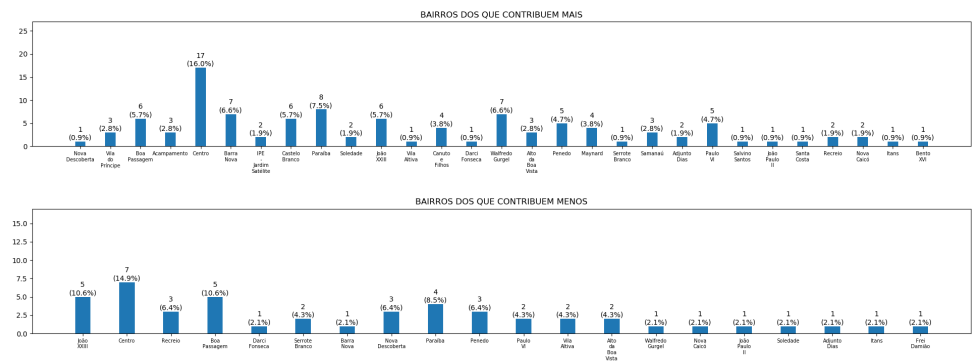
Fonte: PRÓPRIA, 2021

Tabela 4.8: Doações realizadas no último ano

Doações	Principais Contribuições	Perfil daqueles que contribuem mais	Perfil daqueles que contribuem menos
Dinheiro	Entre R\$1 e R\$100 65 (42.5%)	Ens. Sup. Completo : 64(58%) Servidor Público : 50(45.9%)	De 18 a 30 anos : 23(52.3%) Masculino : 22(50%)
Trabalho Voluntário	Nenhum : 118(77%) Entre 1 a 5 : 26(17%)	Filhos biológicos : 18(51.4%) De 31 a 40 anos : 14(40%)	Sem filhos : 62(52.5%) De 18 a 30 anos : 45(38.1%)
Alimentos	De 1 a 5 kg : 63(41%) 0kg : 42(27.5%)	De 1 a 3 SM : 52(46.8%) Filhos biológicos : 58(52.3%)	Solteiro : 28(66.7%) Sem filhos : 28(66.7%)
Ração Animal	0kg : 93 (60.8%) De 1 a 5 kg : 39(25%)	De 31 a 40 anos : 18(30%) Ens. Sup. Completo : 30(50%)	De 18 a 30 anos : 38(40.9%) De 1 a 3 SM : 47(50.5%)
Roupas e Agasalhos	Nenhuma : 56(36%) Mais de 20 : 27(17%)	Feminino : 69(71.1%) Filhos biológicos : 55(56.7%)	Masculino : 35(62.5%) Sem filhos : 38(67.9%)
Produtos de Limpeza e Higiene	Nenhum : 87(56%) Entre 1 e 5 : 37(24%)	De 41 a 50 anos : 22(33.3%) Feminino : 46(69.7%) Filhos biológicos : 39(59.1%)	De 18 a 30 anos : 39(44.8%) Masculino : 43(49.4%) Sem filhos : 53(60.9%)
Brinquedos	Nenhum : 107(69%) Entre 1 e 5 : 28(18%)	Feminino : 31(67.4%) Filhos biológicos : 31(67.4%)	Masculino : 48(44.9%) Sem filhos : 64(59.8%)
Doação de Sangue	Nenhuma : 136(88%) Uma : 8(5%)	Masculino : 12(70.6%) De 31 a 40 anos : 10(58.8%)	Feminino : 85(62.5%) De 18 a 30 anos : 48(35.3%)

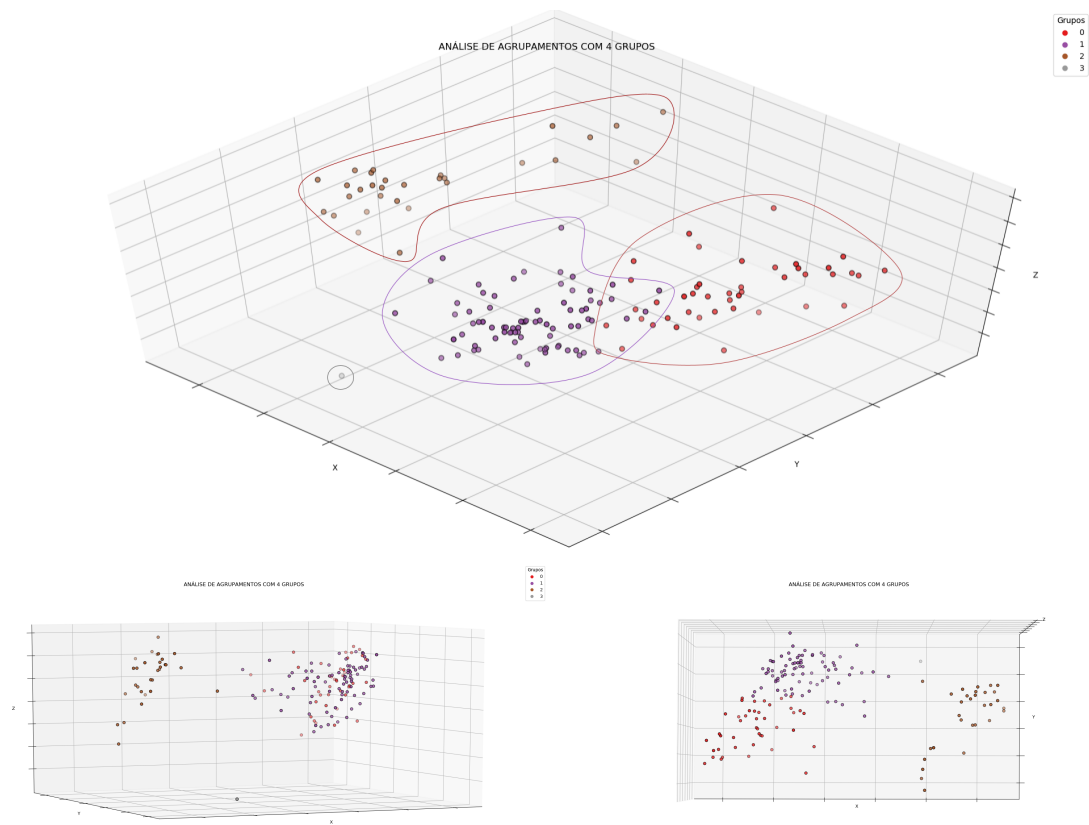
Fonte: PRÓPRIA, 2021

Figura 4.8: Bairros que ajudam mais e menos



Fonte: PRÓPRIA, 2021

Figura 4.9: Análise de Agrupamentos



Fonte: PRÓPRIA, 2021

Resultados da Análise de Associação

Por fim, a última análise realizada foi utilizando o algoritmo Apriori, conhecido por extrair regras de associação em bases de dados.

A técnica foi executada diversas vezes com parâmetros diferentes, no entanto as melhores associações que o algoritmo retornou foram em relação aos usuários das entidades da região, como por exemplo, participantes que não são usuários de duas entidades,

Tabela 4.9: Principais Características dos Grupos

Grupos	Principais Características
Grupo 0 (Comuns)	<p>De 41 a 50 anos : 12 (27.9%) e De 31 a 40 anos : 10 (23.3%) Filhos biológicos : 26 (60.5%) Ensino Médio Completo : 18 (41.9%) De 1 a 3 SM : 17 (39.5%) Nível de espiritualidade alto a muito alto : 26 (60.5%) Preferem Rádio e Instagram Preferem entregar suas contribuições de casa, pelo Pix e por Dízimo Preferem colaborar em Dezembro : 26 (36.6%) Concentrados nos bairros Centro e Castelo Branco Costumam ajudar com Dinheiro, Alimento e Roupas Costumam ser usuários das Aldeias Infantis SOS e Hemocentro Caicó A maioria não conhecem o trabalho da Cáritas, CVV Caicó e Risoterapia</p>
Grupo 1 (Jovens)	<p>De 18 a 30 anos : 38 (48.1%) Sem filhos : 49 (62%) Ensino Superior Completo : 42 (53.2%) De 1 a 3 SM : 42 (53.2%) Nível de espiritualidade médio a muito baixo : 47 (59.5%) Preferem Rádio, Blogs/Sites e Instagram Preferem entregar suas contribuições pelo Pix, de casa e por Troco Solidário Preferem colaborar em Dezembro : 53 (31.7%) Concentrados nos bairros Centro, Paraíba e João XXIII Costumam ajudar com Dinheiro e Alimento Costumam ser usuários do Hemocentro, APAE Caicó e Acapam</p>
Grupo 2 (Ricos)	<p>De 31 a 40 anos : 11 (36.7%) e De 41 a 50 anos : 8 (26.7%) Sem filhos : 15 (50%) e Filhos biológicos : 15 (50%) Ensino Superior Completo : 24 (80%) De 3 a 5 SM : 13 (43.3%) e De 5 a 10 SM : 9 (30%) Nível de espiritualidade alto a muito alto : 21 (70%) Preferem Rádio e Blogs/Sites Preferem entregar suas contrib. de casa, pelo Pix e Transferências/Depósitos Colaboram em todos os meses do ano Concentrados nos bairros Centro e Boa Passagem Costumam ajudar com Dinheiro, Alimento e Roupas Costumam ser usuários do Hemocentro Caicó e APAE Caicó Único grupo que possui mais homens : 16 (53,3%)</p>
Grupo 3 (Outlier)	<p>De 18 a 30 anos Filhos biológicos Ensino Superior Completo De 5 a 10 SM Nível de espiritualidade muito alto Prefere Facebook, Instagram e Igreja Prefere entregar suas contribuições pelo Pix e Rifas/Sorteios Prefere colaborar em Abril, Junho, Setembro e Novembro Mora no bairro Centro Costumam ajudar com Dinheiro, Alimento, Roupas e Brinquedos Usuária das entidades, exceto da Acapam e Aldeias Infantis SOS</p>

Fonte: PRÓPRIA, 2021

tendem a também não ser da terceira. Isso ocorre em 100% dos casos entre os usuários das instituições CVV Caicó, Risoterapia e Cáritas Diocesana. Da mesma forma, ocorreu variações entre os usuários acima citados e usuários das entidades Abrigo Pedro Gurgel e Fazendo da Esperança, nesse caso com taxas de confiança que variam de 95% a 99%.

5 Conclusão

Nesta seção serão apresentadas considerações finais sobre o estudo, como também sugestões de trabalhos futuros.

Como primeira consideração sobre este estudo, vale destacar o número de participantes da pesquisa. No período de realização da coleta dos dados, foi percebida uma baixa adesão dos convidados. Os motivos disso são diversos: a quantidade de questões, receio do convidado, desinteresse, falta de conhecimento em informática, dentre outros. Para suprir essa dificuldade, foi elevada a margem de erro da pesquisa de inicialmente 4% para 10%, fazendo com isso que o tamanho mínimo da amostra fosse reduzido. Uma sugestão para trabalhos futuros seria buscar métodos para tentar reduzir essa margem de erro, e mesmo assim mantendo uma boa taxa de adesão dos convidados.

Outro ponto importante, consiste nos dados estatísticos do IBGE sobre a população de Caicó, que foram utilizados neste estudo nas tarefas de construção do tamanho da amostra e estratificação da mesma. Esses dados são do Censo de 2010, uma atualização com dados mais recentes seria bem vinda.

Ainda em relação a coleta de dados, uma sugestão de trabalho futuro seria tentar coletar dados de repositórios abertos ou de redes sociais através de ferramentas de mercado, como também coletar dados em outro época diferente da pandemia de covid-19 e comparar os resultados.

Ao longo do trabalho também foi sugerido a utilização de cotas por idade, sexo e renda, no entanto, por causa da baixa adesão de participantes, foi mantida a estratégia inicial de estratificar os participantes apenas por bairros da cidade.

A respeito da parte técnica de programação dos algoritmos de análise dos dados, todo o processo foi realizado através da linguagem python. Entretanto, existem diversas outras soluções concorrentes. Uma muito popular é o software WEKA, uma sugestão de trabalho futuro seria aplicar essas outras soluções e comparar os resultados.

Sobre a forma de apresentação dos resultados, foi utilizada a biblioteca de visualização Matplotlib, que auxiliou na construção dos gráficos. Embora essa ferramenta supra o propósito do estudo, um passo além seria a construção de uma ferramenta que permita a visualização dos resultados em tempo real, como por exemplo um site disponível ao público, onde o usuário pudesse dinamicamente filtrar e visualizar os resultados de acordo com sua interação.

Por último, tratando especificamente das questões de pesquisa propostas inicialmente. A questão Q1 perguntava se os resultados do estudo eram capazes de ajudar na tomada de decisão e captação de recursos pelas entidades. De fato um estudo posterior poderia verificar sua eficácia, entretanto com os dados apresentados ações iniciais poderiam ser tomadas, como por exemplo melhorar a avaliação das entidades nos bairros que tiveram má aprovação, intensificando campanhas nessas localidades. Como também, adotar mais os meios de comunicação e colaboração preferidos pelos participantes do estudo. Observar os meses nos quais a ajuda é mais presente também parece ser uma boa estratégia.

A questão Q2 buscava traçar perfis dos colaboradores das entidades da região. Logo, foram traçados diversos perfis por instituição, por padrão de doação e por agrupamentos. Assim, cada ONG pode verificar o perfil do seu público e focar suas ações nos grupos que deseja alcançar.

Já a questão Q3 queria saber se os dados coletados permitiriam descrever um retrato do terceiro setor no município de Caicó/RN. Como pôde ser vista na seção de resultados, embora as análises tivessem propósitos específicos, a maioria vieram acompanhadas de observações gerais do terceiro setor.

Em suma, os resultados apresentados atendem em grande parte o proposto nas questões de pesquisa.

Referências Bibliográficas

ABONG. **Números e dados das Fundações e Associações Privadas Sem Fins Lucrativos no Brasil – Pesquisa FASFIL 2010 (lançada em dezembro de 2012)**. 2010. Disponível em: <<http://encurtador.com.br/tFT39>>.

AMO, S. de. **Técnicas de Mineração de Dados**. Faculdade de Computação, Universidade de Uberlândia, 2004. Disponível em: <<http://files.sistemas2012.webnode.com.br/200000095-bf367bfb43/Tecnicas%20de%20Minera%C3%A7%C3%A3o%20de%20Dados.pdf>>.

ARAÚJO, H. S. D. O desafio da captação de recursos no terceiro setor: o caso da instituição casa do pobre. Departamento de Direito - Universidade Federal do Rio Grande do Norte, Caicó/RN, <https://monografias.ufrn.br/jspui/handle/123456789/7702>, 2018.

ASSIS, J. P. de; SOUZA, R. P. de; DIAS, C. T. dos S. **Glossário de Estatística**. UFESA, 2019. Disponível em: <https://livraria.ufersa.edu.br/wp-content/uploads/sites/165/2019/07/Glossario-de-Estatistica_2019-1.pdf>.

BLUME, B. A.; MARMENTINI, G. **O que são ONGs?** 2017. Disponível em: <<https://www.politize.com.br/ong-o-que-e/>>.

BORGES, L. **Por que a cultura de doação no Brasil ainda é atrasada?** 2017. Disponível em: <<https://www.politize.com.br/cultura-de-doacao-no-brasil/>>.

CAMILO, C. O.; SILVA, J. C. da. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. Instituto de Informática, Universidade Federal de Goiás, 2009. Disponível em: <https://rozero.webcindario.com/disciplinas/fbmg/dm/RT-INF_001-09.pdf>.

CASTRO, L. **Terceiro Setor: 4 pontos para entender**. 2017. Disponível em: <<https://www.politize.com.br/terceiro-setor-o-que-e/>>.

CENSO. **IBGE - Censo2010**. 2010. Disponível em: <<https://censo2010.ibge.gov.br/>>.

COSTA, E.; BAKER, R. S.; AMORIM, L.; MAGALHÃES, J.; MARINHO, T. **Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações**. Anais da Jornada de Atualização em Informática na Educação, 2012. Disponível em: <<https://www.br-ie.org/pub/index.php/pie/article/view/2341>>.

FERNANDES, E. M. da G. **Estatística Aplicada**. 1999. Disponível em: <<https://biblioteca.isced.ac.mz/bitstream/123456789/528/1/Bibliot%20Virtual%20%20Estatistica%20Aplicada.pdf>>.

FERREIRA, A. C. de S. Modelagem e projeto de um sistema de gerenciamento para uma organização de proteção animal. Departamento de Computação e Tecnologia - Universidade Federal do Rio Grande do Norte, Caicó/RN, <https://monografias.ufrn.br/jspui/handle/123456789/10535>, 2019.

FERREIRA, P. L. **Estatística Descritiva e Inferencial: Breves Notas**. 2005. Disponível em: <<https://estudogeral.uc.pt/bitstream/10316/9961/1/AP200501.pdf>>.

FERREIRA, V. C. P. Ongs no brasil: um estudo sobre suas características e fatores que têm induzido seu crescimento. Fundação Getúlio Vargas, Rio de Janeiro/RJ, <http://bibliotecadigital.fgv.br/dspace/handle/10438/3266>, 2005.

FOUNDATION, C. A. **World Giving Index**. 2018. Disponível em: <https://www.idis.org.br/wp-content/uploads/2018/11/WGI2018_REPORT.pdf>.

FREIRE, S. M. **Bioestatística Básica**. 2021. Disponível em: <http://www.lampada.uerj.br/arquivosdb/_book/bioestatisticaBasica.html>.

GAGLIONI, C. **A queda das doações na pandemia. E os caminhos para ajudar**. 2021. Disponível em: <<https://www.nexojornal.com.br/expresso/2021/03/23/A-queda-das-doa%C3%A7%C3%B5es-na-pandemia.-E-os-caminhos-para-ajudar>>.

GERMANO, E. V. de A. Jogos acessíveis: proposta de portal para alunos com necessidades educacionais especiais. Departamento de Computação e Tecnologia - Universidade Federal do Rio Grande do Norte, Caicó/RN, <https://monografias.ufrn.br/jspui/handle/123456789/8394>, 2018.

GIVING, B. **Brasil Giving 2020: Um retrato da Doação no Brasil**. 2020. Disponível em: <https://www.idis.org.br/wp-content/uploads/2020/07/CAF_BrazilGiving2020.pdf>.

HAN, J.; PEI, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. [S.l.]: Morgan Kaufmann, 2006. v. 2 ed.

IBGE. **Cidades e Estados**. 2021. Disponível em: <<https://www.ibge.gov.br/cidades-e-estados/rn/caico.html>>.

IBGE; CNEFE. **Cadastro Nacional de Endereços**. 2010. Disponível em: <<https://censo2010.ibge.gov.br/cnefe/>>.

IPEA. **Mapa das Organizações da Sociedade Civil**. 2021. Disponível em: <<https://mapaosci.ipea.gov.br/resultado-consulta.html>>.

IPEA. **Mapa das Organizações da Sociedade Civil (Resumo das OSC de Caicó/RN)**. 2021. Disponível em: <<https://mapaosci.ipea.gov.br/mapa/2402006>>.

JÚNIOR, A. G. de A.; FIGUEIREDO, N. M. de M. Terceiro setor e estado mínimo: Breve análise sobre a instituição Cáritas diocesana de Caicó, no contexto neoliberal. 16º Congresso Brasileiro de Assistentes Sociais, Brasília/DF, <https://broseguini.bonino.com.br/ojs/index.php/CBAS/article/view/1754/1714>, 2019.

LGPD, C. N. **Lei 13.709 - Lei Geral de Proteção de Dados Pessoais (LGPD)**. 2018. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm>.

MAIA, A. F. O desafio da captação de recursos no terceiro setor: o caso da instituição casa do pobre. Escola de Música, Universidade Federal do Rio Grande do Norte, Natal/RN, <https://monografias.ufrn.br/jspui/handle/123456789/1576>, 2012.

MEDEIROS, A. A. S. O desafio da captação de recursos no terceiro setor: o caso da instituição casa do pobre. Departamento de Ciências Sociais e Humanas - Universidade Federal do Rio Grande do Norte, Currais Novos/RN, <https://monografias.ufrn.br/jspui/handle/123456789/4236>, 2017.

OLIVEIRA, M. R. de. Implantando plac em caicÓ/rn: O ensino de língua portuguesa para refugiados venezuelanos. Revista Eletrônica Científica Ensino Interdisciplinar, Mossoró/RN, <http://natal.uern.br/periodicos/index.php/RECEI/article/view/1699/2274>, 2019.

PRETE, R. L. **O Assunto 420: Na pior hora, doações despencam.** 2021. Disponível em: <<https://g1.globo.com/podcast/o-assunto/noticia/2021/03/29/o-assunto-420-na-pior-hora-doacoes-despencam.ghtml>>.

QUEIROZ, L. M. N. de; MORAIS, I. R.; ALOUFA, M. A. Expansão urbana e vulnerabilidade socioeconômica: Carto(grafias) da cidade. Desenvolvimento em Questão - Revista Unijui, <https://revistas.unijui.edu.br/index.php/desenvolvimentoemquestao/article/view/7470/6001>, 2018.

SANTOS, C. M. L. da Silva Alfonso dos. **Estatística Descritiva: Manual de Auto-Aprendizagem.** Edições Sílabo, 2018. Disponível em: <<https://static.fnac-static.com/multimedia/PT/pdf/9789726189688.pdf>>.

SILVA, J. C. dos S. Contabilidade no terceiro setor: uma análise da percepção dos alunos de ciências contábeis do centro de ensino superior do seridó-ceres- caicó/rn. Departamento de Ciências Exatas e Aplicadas - Universidade Federal do Rio Grande do Norte, Caicó/RN, <https://monografias.ufrn.br/jspui/handle/123456789/1899>, 2015.

SILVEIRA, R. V. ConstituiÇÃo jurídica das organizaÇÕes nÃo-governamentais. UNIVATES, 2006.

TOZZI, J. A. **SOS da ONG.** [S.l.]: Gente, São Paulo/SP, 2015.

TRIOLA, M. F. **Introdução à Estatística.** [S.l.]: LTC, Rio de Janeiro, 2017. v. 12 ed.

ZIBETTI, A. **Probabilidade para Engenharias utilizando o Rstudio.** 2021. Disponível em: <<https://www.inf.ufsc.br/~andre.zibetti/probabilidade/normal.html#>>.

Apêndice A: Questionário

Retrato do Terceiro Setor no Município de Caicó/RN

Este é um convite para participar da pesquisa: "Retrato do Terceiro Setor no Município de Caicó/RN", sob responsabilidade das seguintes pessoas: Hugo Rafael de M. Fernandes (pesquisador) e Almir Miranda Ferreira (orientador), e desenvolvida pela UFRN.

Este estudo busca entender o que a população de Caicó/RN pensa a respeito do Terceiro Setor, além de tentar traçar um perfil do colaborador de instituições que prestam serviços filantrópicos e beneficentes no município.

OBS: TODAS AS RESPOSTAS DESTE ESTUDO SERÃO ANÔNIMAS E VOLUNTÁRIAS.

***Obrigatório**

1. P1. Qual seu Sexo? *

Marcar apenas uma oval.

☐ Masculino

☐ Feminino

2. P2. Qual sua Idade? *

Marcar apenas uma oval.

☐ Menor de 18 anos

☐ De 18 a 30 anos

☐ De 31 a 40 anos

☐ De 41 a 50 anos

☐ De 51 a 60 anos

☐ Acima dos 60 anos

3. P3. Você tem filhos(as)? *

Marcar apenas uma oval.

☐ Não tenho filhos(as)

☐ Somente filhos(as) biológicos

☐ Somente filhos(as) adotados/criação

☐ Tenho filhos(as) biológicos e adotados/criação

4. P4. Qual seu Estado Civil *

Marcar apenas uma oval.

- ☐ Solteiro
- ☐ Casado
- ☐ União Estável
- ☐ Divorciado
- ☐ Viúvo

5. P5. Qual sua Escolaridade? *

Marcar apenas uma oval.

- ☐ Ensino Fundamental Incompleto
- ☐ Ensino Fundamental Completo
- ☐ Ensino Médio Incompleto
- ☐ Ensino Médio Completo
- ☐ Ensino Superior Incompleto
- ☐ Ensino Superior Completo

6. P6. Qual sua ocupação? *

Marcar apenas uma oval.

- ☐ Servidor Público
- ☐ Empregado na Iniciativa Privada
- ☐ Autônomo ou Profissional Liberal
- ☐ Aposentado ou Pensionista
- ☐ Estudante
- ☐ Do Lar
- ☐ Desempregado

7. P7. Onde você mora em Caicó/RN? *

Marcar apenas uma oval.

- ☐ Acampamento
- ☐ Adjunto Dias
- ☐ Alto da Boa Vista
- ☐ Barra Nova
- ☐ Bento XVI
- ☐ Boa Passagem
- ☐ Canuto e Filhos
- ☐ Castelo Branco
- ☐ Centro
- ☐ Darci Fonseca
- ☐ Frei Damião
- ☐ IPE - Jardim Satélite
- ☐ Itans
- ☐ João XXIII
- ☐ João Paulo II
- ☐ Maynard
- ☐ Nova Caicó
- ☐ Nova Descoberta
- ☐ Paulo VI
- ☐ Paraíba
- ☐ Penedo
- ☐ Recreio
- ☐ Samanaú
- ☐ Santa Costa
- ☐ Salvino Santos
- ☐ Serrote Branco
- ☐ Soledade
- ☐ Vila do Príncipe
- ☐ Vila Altiva
- ☐ Walfredo Gurgel
- ☐ Zona Rural

8. P8. Quais destas entidades você conhece o trabalho e como as avalia?

Você só precisa avaliar as entidades que conhece o trabalho.

Marcar apenas uma oval por linha.

	Muito Importante	Importante	Mediana	Às vezes é Importante	Não é Importante
APAE Caicó	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Acapam	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Aldeias Infantis SOS	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Abrigo Pedro Gurgel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cáritas Diocesana	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fazenda da Esperança	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Centro de Valorização da Vida - CVV	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Risoterapia	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hemocentro Caicó	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

9. P9. Você já precisou dos serviços de alguma dessas entidades? De quais?

Você pode marcar mais de uma opção, como também deixar em branco.

Marque todas que se aplicam.

- ☐ APAE Caicó
- ☐ Acapam
- ☐ Aldeias Infantis SOS
- ☐ Abrigo Pedro Gurgel
- ☐ Cáritas Diocesana
- ☐ Fazenda da Esperança
- ☐ Centro de Valorização da Vida - CVV
- ☐ Risoterapia
- ☐ Hemocentro Caicó

10. P10. Quem você acha que financia o trabalho dessas instituições? *

Você deve marcar quem ajuda mais e quem ajuda menos.

Marcar apenas uma oval por linha.

	Poder Público	Iniciativa Privada (Empresas)	Sociedade em Geral (Indivíduos)
Ajuda Mais	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ajuda Menos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

11. P11. Como você fica sabendo do trabalho dessas entidades? *

Você pode marcar mais de uma opção.

Marque todas que se aplicam.

- ☐ Rádio
- ☐ Blogs e Sites
- ☐ Televisão
- ☐ Jornais e Revistas
- ☐ Carros de Som
- ☐ Igreja
- ☐ Facebook
- ☐ Instagram
- ☐ Whatsapp

12. P12. Quanto Dinheiro você doou no último ano? *

Marcar apenas uma oval.

- ☐ R\$ 0,00 (Zero)
- ☐ Entre R\$1,00 e R\$100,00
- ☐ Entre R\$101,00 e R\$200,00
- ☐ Entre R\$201,00 e R\$300,00
- ☐ Entre R\$301,00 e R\$400,00
- ☐ Entre R\$401,00 e R\$500,00
- ☐ Entre R\$501,00 e R\$600,00
- ☐ Mais de R\$600,00

13. P13. Quantas vezes você prestou Trabalho Voluntário no último ano? *

Marcar apenas uma oval.

- ☐ 0 (Zero)
- ☐ Entre 1 e 5
- ☐ Entre 6 e 10
- ☐ Entre 11 e 15
- ☐ Entre 16 e 20
- ☐ Mais de 20

14. P14. Quantos quilos de Alimento você doou no último ano? *

Marcar apenas uma oval.

- ☐ 0 kg (Zero)
- ☐ Entre 1kg e 5kg
- ☐ Entre 6kg e 10kg
- ☐ Entre 11kg e 15kg
- ☐ Entre 16kg e 20kg
- ☐ Mais de 20kg

15. P15. Quantas peças de Roupas e Agasalhos você doou no último ano? *

Marcar apenas uma oval.

- ☐ 0 (Zero)
- ☐ Entre 1 e 5
- ☐ Entre 6 e 10
- ☐ Entre 11 e 15
- ☐ Entre 16 e 20
- ☐ Mais de 20

16. P16. Quantos Produtos de Limpeza e Higiene você doou no último ano? *

Marcar apenas uma oval.

- ☐ 0 (Zero)
- ☐ Entre 1 e 5
- ☐ Entre 6 e 10
- ☐ Entre 11 e 15
- ☐ Entre 16 e 20
- ☐ Mais de 20

17. P17. Quantos quilos de Ração Animal você doou no último ano? *

Marcar apenas uma oval.

- ☐ 0 kg (Zero)
- ☐ Entre 1kg e 5kg
- ☐ Entre 6kg e 10kg
- ☐ Entre 11kg e 15kg
- ☐ Entre 16kg e 20kg
- ☐ Mais de 20kg

18. P18. Quantos Brinquedos você doou no último ano? *

Marcar apenas uma oval.

- ☐ 0 (Zero)
- ☐ Entre 1 e 5
- ☐ Entre 6 e 10
- ☐ Entre 11 e 15
- ☐ Entre 16 e 20
- ☐ Mais de 20

19. P19. Quantas Doações de Sangue você fez no último ano? *

Marcar apenas uma oval.

- ☐ 0 (Zero)
- ☐ 1 (Uma)
- ☐ 2 (Duas)
- ☐ 3 (Três)
- ☐ 4 (Quatro)

20. P20. Como você prefere entregar suas contribuições? *

Você pode marcar mais de uma opção.

Marque todas que se aplicam.

- ☐ Prefiro que alguém venha recolher na minha casa
- ☐ Prefiro por Transferência ou Depósito Bancário
- ☐ Prefiro por Correspondência
- ☐ Prefiro por Pix
- ☐ Prefiro por Caixinha de Troco Solidário
- ☐ Prefiro por Rifas e Sorteios
- ☐ Prefiro por Dízimo

21. P21. Quais meses do ano você costuma colaborar? *

Você pode marcar mais de uma opção.

Marque todas que se aplicam.

- ☐ Janeiro
- ☐ Fevereiro
- ☐ Março
- ☐ Abril
- ☐ Maio
- ☐ Junho
- ☐ Julho
- ☐ Agosto
- ☐ Setembro
- ☐ Outubro
- ☐ Novembro
- ☐ Dezembro

22. P22. Antes da pandemia, quanto você contribuía comparado com agora? *

Marcar apenas uma oval.

- ☐ Muito menos
☐ Menos
☐ Mesma quantidade
☐ Mais
☐ Muito mais

23. P23. Qual seu nível de espiritualidade? *

Marcar apenas uma oval.

	1	2	3	4	5	
Pouco Espiritualizado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Bastante Espiritualizado

24. P24. Qual sua faixa de renda domiciliar? *

Incluindo você e seus familiares.

Marcar apenas uma oval.

- ☐ Até R\$ 1.100,00 (até 1 s.m.)
☐ De R\$ 1.100,01 a R\$ 3.300,00 (de 1 a 3 s.m.)
☐ De R\$ 3.300,01 a R\$ 5.500,00 (de 3 a 5 s.m.)
☐ De R\$ 5.500,01 a R\$ 11.000,00 (de 5 a 10 s.m.)
☐ Mais que R\$ 11.000,01 (mais que 10 s.m.)

25. Declaro ter lido e compreendido este documento, bem como as informações apresentadas. Foi-me garantida a possibilidade de, em qualquer momento, recusar participar deste estudo. Dessa forma, aceito participar e permito a utilização dos dados de forma voluntária, confiando em que apenas serão utilizados para esta investigação e nas garantias de confidencialidade e anonimato que me são dadas pelos pesquisadores. *

Para mais informações acesse: bit.ly/perfilcaico-termos

Marque todas que se aplicam.

☐ Aceitar

Apêndice B: Modelo do Convite da Pesquisa

Este é um convite para participar da pesquisa: "**Retrato do Terceiro Setor no Município de Caicó/RN**", sob responsabilidade das seguintes pessoas: Hugo Rafael de Medeiros Fernandes (pesquisador) e Almir Miranda Ferreira (orientador), e desenvolvida pela UFRN.

Este estudo busca entender o que a população de Caicó/RN pensa a respeito do Terceiro Setor, além de tentar traçar um perfil do colaborador de instituições que prestam serviços filantrópicos e beneficentes no município.

Acesse o link: **bit.ly/perfilcaico**
e dê sua contribuição. Agradecemos sua participação!

OBS: TODAS AS RESPOSTAS DESTE ESTUDO SERÃO ANÔNIMAS E VOLUNTÁRIAS.
Para mais informações acesse: **bit.ly/perfilcaico-termos**

Modelo do Convite da Pesquisa
(Formato Panfleto)