Universidade de Brasília
Faculdade de Tecnologia

Avaliação de micro e pequenas empresas industriais do setor alimentício: aplicação da Análise Envoltória de Dados

Isabella Caciano Gomes Lacerda

TRABALHO DE GRADUAÇÃO

ENGENHARIA DE PRODUÇÃO

Brasília

2022

Universidade de Brasília
Faculdade de Tecnologia

Avaliação de micro e pequenas empresas industriais do setor alimentício: aplicação da Análise Envoltória de Dados

Isabella Caciano Gomes Lacerda

Trabalho de Graduação submetido como requisito parcial para obtenção do grau de Bacharel em Engenharia de Produção.

Orientador: Prof. Ph.D. Reinaldo Crispiniano Garcia

Brasília
2022

Lacerda, Isabella Caciano Gomes.
LZ123a Avaliação de micro e pequenas empresas industriais do setor alimentício: aplicação da Análise Envoltória de Dados / Isabella Caciano Gomes Lacerda; orientador Reinaldo Crispiniano Garcia. -- Brasília, 2022.
74 p.

Trabalho de Graduação em Engenharia de Produção -- Universidade de Brasília, 2022.

1. Análise Envoltória de Dados. 2. DEA. 3. Eficiência. 4. Micro e Pequenas Empresas. I. Garcia, Reinaldo Crispiniano, orient. II. Avaliação de micro e pequenas empresas industriais do setor alimentício: aplicação da Análise Envoltória de Dados

Universidade de Brasília
Faculdade de Tecnologia

Avaliação de micro e pequenas empresas industriais do setor alimentício: aplicação da Análise Envoltória de Dados

Isabella Caciano Gomes Lacerda

Trabalho de Graduação submetido como requisito parcial para obtenção do grau de Bacharel em Engenharia de Produção.

Brasília, 06 de outubro de 2022:

Prof. Ph.D. Reinaldo Crispiniano Garcia,
UnB/FT/EPR
Orientador

Prof. Ph.D. Annibal Affonso Neto,
UnB/FT/EPR
Examinador interno

Brasília
2022

Às constantes da minha vida: Kleyton, Lucinete e Dáfne.

Agradecimentos

Aos meus pais, Kleyton e Lucinete, que tanto sacrificaram para que eu pudesse me formar em engenharia na Universidade de Brasília. O empenho deles não se resume ao período de curso, desde a busca por um ensino de qualidade mais nova, pela constante busca por cursos à parte da minha formação básica, pelo café da manhã que antes das 5h estava pronto, pelas companhias cedo ou tarde no ponto de ônibus. Por todos os sacrifícios que eu não tenho conhecimento. Sou quem sou, porque vocês me deram a oportunidade de ser.

À minha irmã, Dáfne, que foi a primeira a pavimentar diversos dos caminhos que serviriam para que eu os percorresse. Pela conexão de quem vivenciou muito do mesmo e que será pra sempre só nossa, por todas as risadas, shows, construções e desconstruções, pelos momentos de descontração necessários que acontecem constantemente ao apenas nos cruzarmos.

Ao Romulo, por ser inspiração, pela companhia, crescimento e risadas constantes, pela celebração das pequenas e grandes vitórias, por todo auxílio e encorajamento.

A toda a minha família, tios, avós e primos que sempre me apoiaram, incentivaram e são exemplos de amor e dedicação em tudo que fazem.

Aos meus amigos do ensino fundamental, que permaneceram, aos do ensino médio, cujos laços se intensificaram, e aos que surgiram durante o período da faculdade, em especial às amizades improváveis. Vocês me inspiram sempre, seja no curtir ou no trabalhar. Obrigada por trazerem tanto equilíbrio.

À Universidade de Brasília, por permitir um ensino plural, que me permitiu conhecer um novo universo para além dos meus costumes, por todas as experiências que passei que me moldaram e me marcarão por toda a vida. Saio extremamente diferente de como entrei, com muito orgulho de quem tenho me tornado e com a certeza de que todas as conexões que fiz me trouxeram a esse ponto.

Aos professores do Departamento de Engenharia de Produção, em especial ao Prof. Reinaldo Garcia, por ser um exemplo de profissional e pelo constante incentivo a sonharmos chegar em lugares mais altos.

Muito obrigada a todos!

Resumo

O segmento de Micro, Pequenas e Médias Empresas é um foco importante das políticas públicas, responsável por significativo impacto social na geração de renda e emprego no Brasil, compõe cerca de 98,5% do universo de empresas formais no país e é responsável por mais de 29% do PIB. Estima-se que a baixa produtividade na América Latina se deve, principalmente, ao elevado volume de recursos alocado a firmas pequenas e de baixa produtividade, assim como à ausência de empresas com níveis médio e alto de produtividade. Este trabalho tem por objetivo avaliar a eficiência de micro e pequenas empresas do setor alimentício da indústria de transformação brasileira, trata-se de uma pesquisa de aplicação prática e abordagem quantitativa realizada sob à ótica da análise envoltória de dados. Verificou-se a homogeneidade da eficiência de 4047 empresas por meio de um recorte regional, apresentando regiões com maior concentração de empresas eficientes, como a região norte, e empresas com maiores discrepâncias nas variáveis utilizadas, como ocorreu com a região nordeste. Tais análises podem servir de suporte à tomada de decisão para investimentos no setor, com o intuito de projetar as empesas para o nível ótimo de funcionamento, com a alocação adequada de recursos. Também foram apresentados pontos de aprofundamento para elaboração de estudos futuros na área.

Palavras-chave: Análise Envoltória de Dados. DEA. Eficiência. Micro e Pequenas Empresas.

Abstract

The Micro, Small and Medium Enterprises segment is an important focus of public policies, responsible for a significant social impact on income and employment generation in Brazil, makes up about 98.5% of the universe of formal companies in the country and is responsible for more than 29% of GDP. It is estimated that the low productivity in Latin America is mainly due to the high volume of resources allocated to small and low productivity firms, as well as the absence of companies with medium and high levels of productivity. This work aims to evaluate the efficiency of micro and small companies in the food sector in the Brazilian manufacturing industry, it is a research of practical application and quantitative approach carried out from the perspective of data envelopment analysis. The efficiency homogeneity of 4047 companies was verified through a regional cut, presenting regions with a greater concentration of efficient companies, such as the northern region, and companies with greater discrepancies in the variables used, as occurred with the northeast region. Such analyzes can support decision-making for investments in the sector, with the aim of projecting companies to the optimal level of operation, with the proper allocation of resources. Deepening points were also presented for the elaboration of future research in the area.

Keywords: Data Envelopment Analysis. DEA. Efficiency. Micro and Small Enterprises.

Lista de ilustrações

Lista de tabelas

Lista de Quadros

Lista de abreviaturas e siglas

ABIA	Associação Brasileira da Indústria de Alimentos	20
CNI	Confederação Nacional da Indústria	24
CNPJ	Cadastro Nacional da Pessoa Jurídica	34
CTN	Código Tributário Nacional	34
DAU	Dívida Ativa da União	34
DEA	Data Envelopment Analysis	27
DMU	Decision Making Unit	27
IO	Input Oriented	30
MPEs	Micro e Pequenas Empresas	15
MPMEs	Micro, Pequenas e Médias Empresas	14
OO	Output Oriented	30
PGFN	Procuradoria Geral da Fazenda Nacional	34
PIB	Produto Interno Bruto	15
PL	Programação Linear	25
PMEs	Pequenas é Medias Empresas	15
PO	Pesquisa Operacional	16
RCE	Retorno Constante de Escala	29
RFB	Receita Federal do Brasil	34
RVE	Retornos Variáveis de Escala	31
TEA	Taxa de Empreendedorismo Inicial	14

Sumário

1 Introdução

A dinâmica econômica e o crescimento dos países em desenvolvimento são dependentes da criação de negócios que gerem emprego e renda para a população de forma sustentável, de modo a auxiliar esses países a atingirem níveis mais elevados de produção de bens e serviços e a ocuparem posições estratégicas na economia mundial (FERREIRA et al., 2012).

A análise do nível de atividade empreendedora em 54 países, o que representa 95% do PIB global e dois terços da população do planeta, demonstrou que a Taxa de Empreendedorismo em estágio inicial (TEA) do Brasil é de 14,89%, superando a média dos países participantes, 10,95% (MONITOR, 2011).

No entanto, analisando inovações aplicadas para os consumidores e o grau de concorrência, o Brasil está abaixo da média dos 54 países participantes, demonstrando que a inovação ainda está em estágio inicial no dia a dia das empresas e o empreendedor brasileiro é aquele com menor conteúdo inovador em seus negócios (MONITOR, 2011).

O ambiente político e econômico brasileiro levou a um alto nível de atividade comercial. Todavia, o Brasil tem baixa expectativas de crescimento e inovação, sugerindo que os empreendedores contribuem para a economia com base em seu alto nível de participação, em vez de qualquer impacto no nível individual. A baixa taxa de atividade empreendedora mostra que os funcionários não conseguem estimular o crescimento das empresas para as quais trabalham por meio da atividade empreendedora (BOSMA; KELLEY, 2019).

Com relação à produtividade na economia brasileira dos anos 1950 a 2009, apesar das profundas mudanças na estrutura econômica, política e produtiva que o país experimentou, a dinâmica da produtividade não mudou. Assim, embora houve profunda mudança em sua estrutura produtiva, passando de um país fundamentalmente agrícola para uma economia industrial, moderna e diversificada, o Brasil ainda se caracteriza por diferenças significativas de produtividade entre os diversos setores da economia, não apresentando convergência produtiva (SQUEFF; NOGUEIRA, 2013).

Diante disso, tem-se o segmento de Micro, Pequenas e Médias Empresas (MPMEs) como um foco importante das políticas públicas, responsável por significativo impacto social na geração de renda e emprego no Brasil. A Figura 4 apresenta a evolução da participação dos pequenos negócios do PIB brasileiro dos anos de 2009 a 2017.

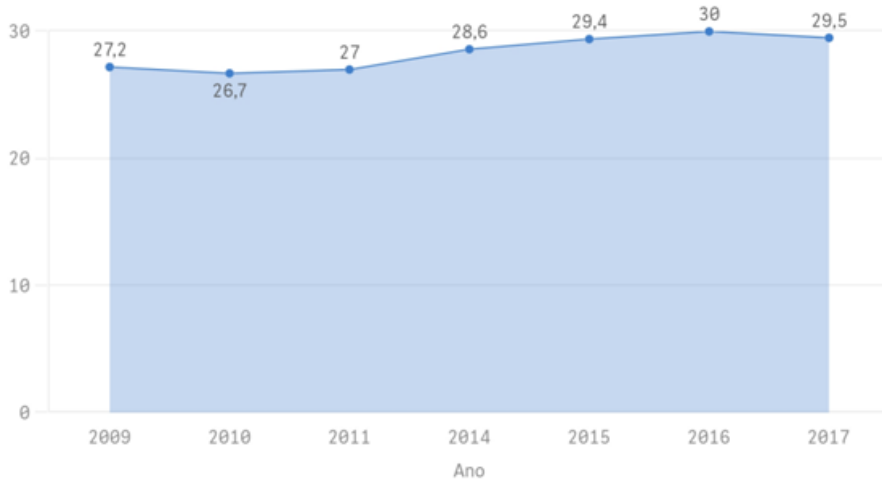


Figura 1 – Participação dos pequenos negócios no PIB (%)

Fonte: DataSebrae (2018).

Segundo estimativas do DataSebrae (2018), somente o segmento de Micro e Pequenas Empresas (MPEs) constitui cerca de 98,5% do universo de empresas formais no país e compôs mais de 29% do PIB, empregando cerca de 17,7 milhões de trabalhadores no ano de 2017.

1.1 Identificação da problemática

Há na contemporaneidade uma baixa produtividade atribuída à América Latina, que deve-se, principalmente, ao elevado volume de recursos alocados a firmas pequenas e de baixa produtividade, assim como à ausência de empresas com níveis médio e alto de produtividade (BONELLI; VELOSO; PINHEIRO, 2017).

Características do processo de produção, como o regime de competição entre indústrias, tornam o sistema de preços relativos inepto para que o mercado seja capaz de alocar recursos de forma eficaz ao nível das empresas, dos setores de atividade e da economia (BONELLI; VELOSO; PINHEIRO, 2017). Os determinantes internos mais relatados de produtividade de Pequenas e Medias Empresas (PMEs), segundo Marchese et al. (2019), são:

- Habilidades gerenciais e práticas de gestão- incluindo àquelas mais intimamente relacionadas à força de trabalho, como treinamento e gestão de recursos humanos;
- Tecnologias de informação, comunicação e digitalização- incluindo o uso de hardware, comércio eletrônico e programas que podem ajudar a profissionalizar a gestão de pequenas empresas;

- Redes de negócios- incluindo a participação em *clusters* e cadeias de suprimentos globais que ajudem as empresas a superarem as restrições relacionadas ao tamanho no que diz respeito ao acesso a recursos e mercados;
- Inovação- relacionada à introdução de novos produtos ou processos a nível da empresa, inclusive por meio de investimentos em pesquisa e desenvolvimento.

Com recursos limitados, é necessário entender qual a melhor proposta de alocação. Diante disso, faz-se necessária a revisão sistemática da literatura técnico-científica de produtividade de micro e pequenas empresas, com o intuito de aplicar uma metodologia para entendimento da eficiência industrial dessas, de modo a facilitar o processo de tomada de decisão de organizações na seleção das melhores ações relacionadas aos desafios dos negócios.

1.2 Objetivo geral

Este trabalho tem por objetivo avaliar a eficiência de micros e pequenas empresas do setor alimentício da indústria de transformação brasileira, por meio da Pesquisa Operacional (PO), em particular da análise envoltória de dados.

1.3 Objetivos específicos

A fim de alcançar o objetivo geral, foram definidos os seguintes objetivos específicos:

1. Analisar micro e pequenas empresas no contexto da indústria brasileira;
2. Identificar variáveis e modelagem ideal para o tema em questão;
3. Avaliar a eficiência das empresas estudadas;
4. Apresentar avaliação comparativa de resultados.

1.4 Estrutura do trabalho

Este trabalho é dividido em cinco capítulos, que traduzem toda a aplicação e estudos acerca do tema. O primeiro capítulo contextualiza os principais conceitos do tema abordado. O segundo capítulo consiste na fundamentação teórica sobre os conceitos de eficiência e análise de dados, a fim de compreender os principais conceitos da área para fundamentar a aplicação realizada. No terceiro capítulo, apresenta-se a metodologia utilizada nesta aplicação e como o projeto foi realizado, em etapas. O quarto capítulo aborda a aplicação do modelo em si e os resultados recebidos. No quinto capítulo apresenta-se a conclusão após a análise

dos resultados. Por fim, são apresentadas referências bibliográficas mencionadas ao longo do trabalho.

2 Referencial Teórico

A fundamentação teórica apresentada nesta seção aborda os conceitos elencados na problemática de pesquisa de modo a fundamentar o trabalho. Nesse sentido, serão abordados os tópicos de eficiência industrial e análise de produtividade. Em seguida, será aprofundado o tema de pesquisa operacional com os principais métodos e algoritmos para resolução de problemas.

2.1 Indústria brasileira

O setor industrial brasileiro é um dos maiores geradores de empregos no Brasil e fortalece todo o setor produtivo. Responsável por empregar 9,7 milhões de brasileiros, representa 20,4% do PIB e 69,2% das exportações brasileiras de bens e serviços, é responsável também por 69,2% do investimento empresarial em pesquisa e desenvolvimento e por 33% das arrecadações de tributos federais (CNI, 2022)

A indústria brasileira engloba uma série de atividades produtivas que impactam positivamente nos demais setores da economia, como o comércio e o agronegócio. É dividida em 3 perfis setoriais: indústria extrativa, indústria da construção e indústria da transformação. O comparativo entre segmentos da indústria no ano de 2021 no que se refere à participação no PIB da indústria é apresentado na Figura 2.

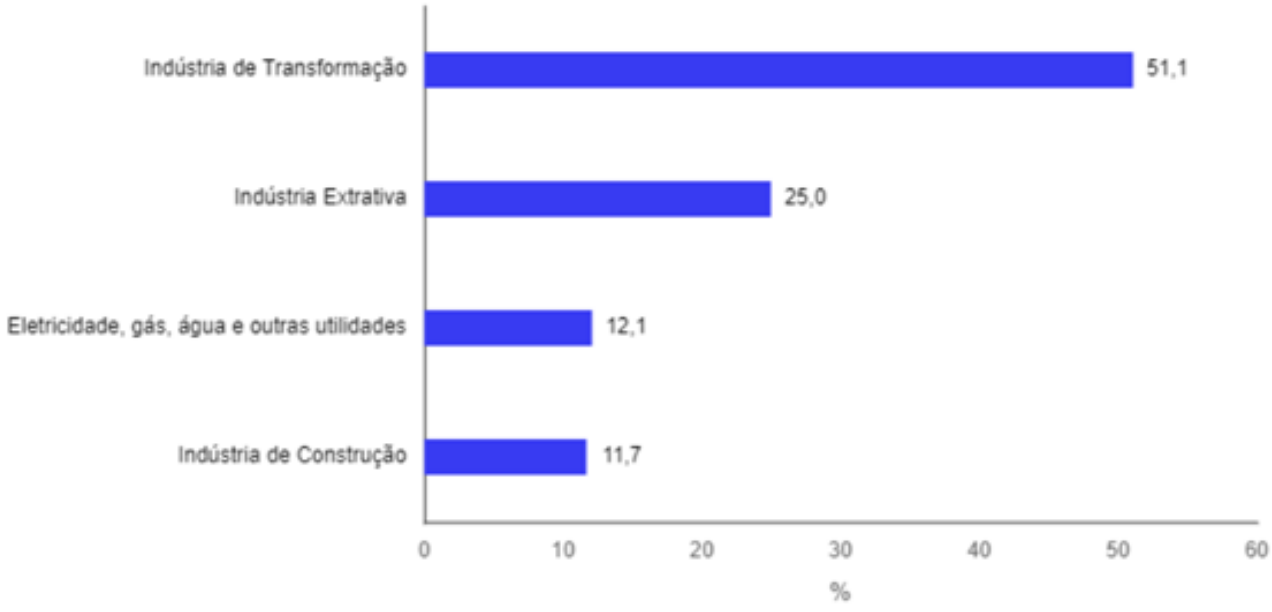


Figura 2 – Participação no PIB da indústria

Fonte: Portal da indústria (2021).

Dessa forma, é possível perceber a importância da indústria de transformação no cenário brasileiro, uma vez que compõe mais de 51% do total do PIB da indústria brasileira.

2.1.1 Indústria de transformação

A indústria de transformação é um seguimento de indústria que realiza a transformação de matéria-prima em um produto final ou intermediário que vai ser novamente modificado por outra indústria. Os materiais, substâncias e componentes usados por essas indústria são provenientes de produção agrícola, mineração, pesca, extração florestal e produtos de outras atividades industriais (CNI, 2022).

Desempenha um papel estratégico no fortalecimento de todo o setor produtivo e é responsável por fabricar a maioria dos produtos consumidos no dia a dia pelas famílias, além de produzir insumos, máquinas e equipamentos utilizados não só pela indústria como também pela agropecuária e pelo setor de serviços (CNI, 2022).

A indústria de transformação emprega 6,9 milhões de trabalhadores, sendo responsável por mais de 65% dos investimentos empresariais em pesquisa e desenvolvimento no país (CNI, 2022). Nesse sentido, verifica-se o número de estabelecimentos da indústria de transformação por porte, apresentado na Figura 3.

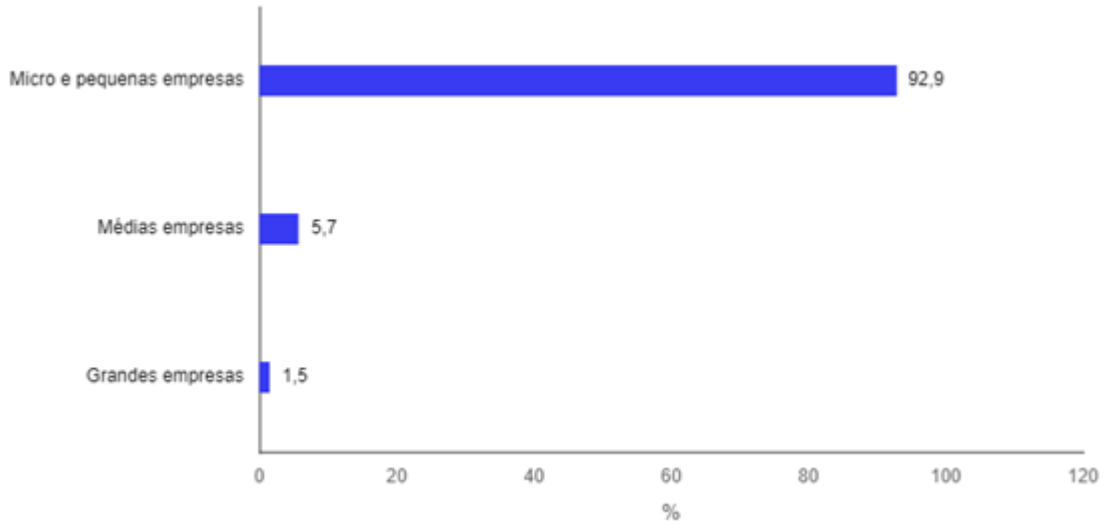


Figura 3 – Participação no total de estabelecimentos industriais - 2020 (%)

Fonte: CNI (2020).

Com mais de 90% dos estabelecimentos, o segmento de MPEs domina a indústria de transformação. No entanto, no que se refere à distribuição da produção da indústria de transformação por porte de estabelecimento, as MPEs são as que menos agregam valor, uma vez que grandes empresas são responsáveir por cerca de 75,6

2.1.2 Setor de alimentos

As indústrias de transformação concentraram 92,9% do faturamento das empresas industriais em 2020. O segmento de fabricação de produtos alimentícios ocupou a primeira posição no ranking de receita líquida de vendas, com 24,1% do faturamento da indústria brasileira. Foi o setor com maior ganho de participação de mercado, com incremento de 5,9 pontos percentuais entre os anos de 2011 a 2020, dos quais 3,6 foram relativos apenas ao período 2019-2020 (IBGE, 2020).

A indústria alimentícia abrange uma grande variedade de produtos e está intimamente relacionada à agricultura e pecuária, por serem os principais fornecedores de insumos desse setor. Devido a isso, a indústria tem produção sazonal relacionada à oferta sazonal de insumos. Relaciona-se, também, com canais de distribuição, indústrias de embalagens, máquinas e equipamentos, entre outros (VIANA, 2022).

É também muito importante na indústrias de transformação em termos de participação no PIB e criação de empregos. A indústria brasileira de alimentos faturou R$ 699,9 bilhões em 2019, representando 9,6% do PIB do Brasil naquele ano. Sua receita em 2020 foi de R$ 789,2 bilhões, equivalente a 10,5% do PIB do país, um aumento de 12,8% em relação à 2019 (ABIA, 2020).

A indústria de transformação de alimentos compreende o processamento e transformação de produtos da agricultura, pecuária e pesca em alimentos para uso humano e animal. Conforme dados da Associação Brasileira da Indústria de Alimentos (ABIA), o Brasil é o segundo maior exportador de alimentos industrializados do mundo, levando seus alimentos para 190 países. Dentre suas principais contribuições estão:

- 58% de toda a produção agropecuária é processada pela indústria de alimentos;
- 24% dos empregos da indústria de transformação brasileira são do setor de alimentos;
- Contribuiu com 63,7% da balança comercial do Brasil;
- Compôs 16% das exportações totais brasileiras em 2021.

Nesse sentido, a Figura 4 apresenta o desempenho das vendas reais, produção física e pessoal ocupado na indústria de transformação de alimentos de dezembro de 2018 a junho de 2022.

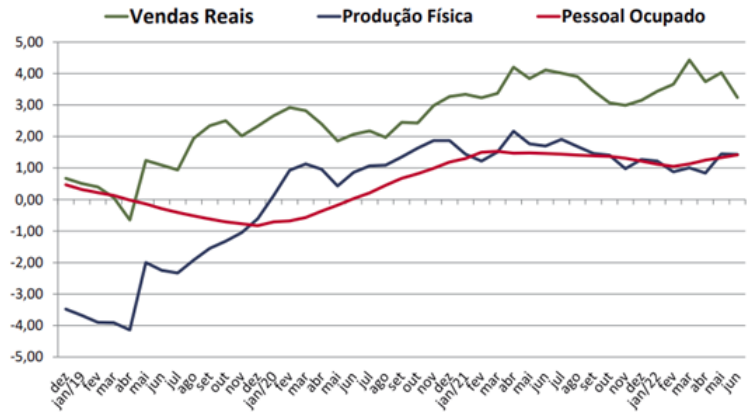


Figura 4 – Desempenho das vendas reais, produção física e pessoal ocupado

Fonte: ABIA (2022).

É possível perceber a tendência de crescimento do setor, ratificando sua importância para o desenvolvimento econômico brasileiro, bem como a necessidade de avaliação crítica com relação à eficiência das empresas desse universo.

2.1.3 Micro e Pequenas Empresas

As Micro e Pequenas Empresas podem ser definidas conforme dois critérios, faturamento ou número de funcionários. O critério de faturamento está previsto na Lei Complementar nº 123/2006, tal classificação determina o seguinte:

- Micro empresa: empresa que tem faturamento anual de até R$ 360 mil ou emprega até 9 pessoas no comércio e serviços ou 19 pessoas no setor industrial.
- Pequena empresa: empresa que tem faturamento anual de até R$ 4,8 milhões por ano ou emprega de 10 a 49 pessoas no comércio e serviços ou de 20 a 99 pessoas na indústria.

Por outro lado, utilizando a classificação por receita bruta anual, as empresas são definidas da seguinte maneira:

- Micro empresa: empresa com renda anual menor ou igual a R$ 360 mil
- Pequena empresa: empresa com renda anual maior que R$ 360 mil e menor ou igual a R$4,8 milhões

De acordo com o Portal da Indústria, em 2019, as MPEs representaram a maior parte dos empreendimentos no Brasil. Do total de 476.243 empresas, 71,7% eram consideradas

microempresas (até 9 empregados) e 22,6% eram consideradas pequenas empresas (de 10 a 49 empregados).

Com base nas empresas brasileiras com abertura em 2012 com dados até 2014, a taxa de sobrevivência dessas com dois anos de operação foi de 76,6% (SEBRAE, 2016). Essa taxa foi a maior taxa de sobrevivência para empresas nascidas em todo o período entre 2008 e 2012, como apresenta a Figura 5.

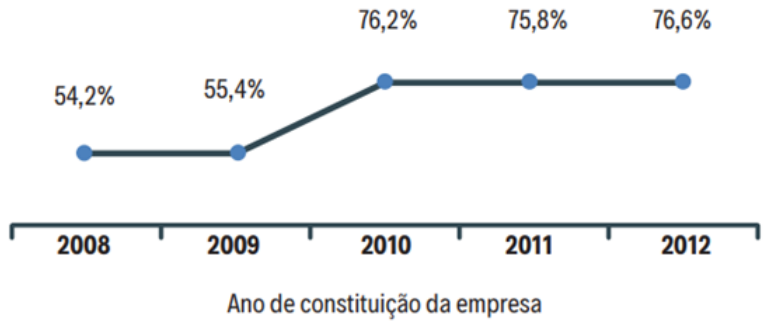


Figura 5 – Taxa de sobrevivência de empresas de dois anos: evolução no Brasil

Fonte: Sebrae (2016).

Como a taxa de mortalidade é complementar à taxa da sobrevivência, a taxa de mortalidade de empresas com até dois anos caiu de 45,8% para 23,4% em empresas fundadas em 2008 (SEBRAE, 2016). Com relação à um recorte por região, é possível visualizar a Figura 6.

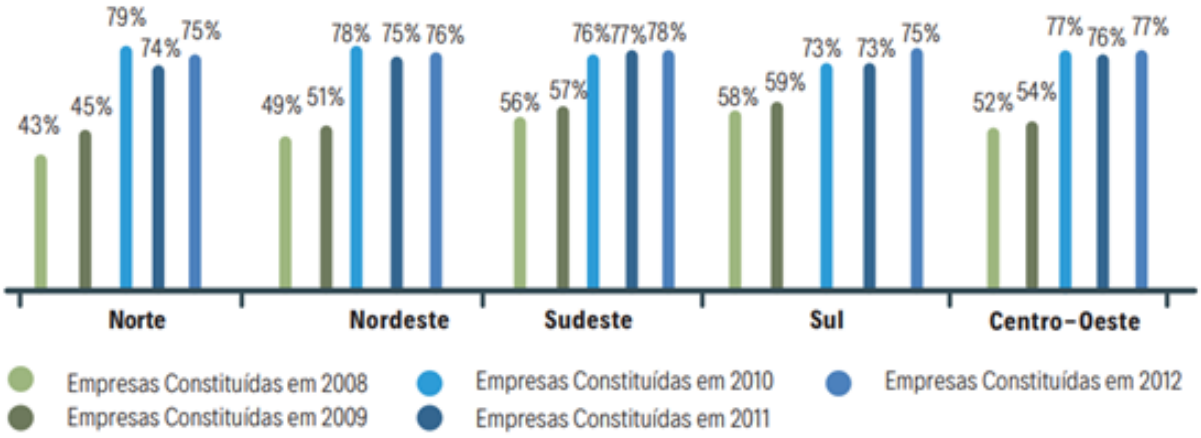


Figura 6 – Taxa de sobrevivência de empresas de dois anos, por região

Fonte: Sebrae (2016).

Observa-se que no ano de 2010 a taxa de sobrevivência para todas as regiões superou 70%. No entanto, tal aumento não é refletido diretamente na produtividade dessas empresas.

MPMEs enfrentam dificuldades com pouca capacitância de gestão e acesso à linhas de crédito e novos mercados, as levando a buscar soluções imediatas para seus problemas de fluxo de caixa e planejamento. Isto dificulta a definição de estratégias de crescimento de longo prazo, caracterizadas por inovações e mercados de maior valor agregado (CEPAL et al., 2018).

Verifica-se, assim, que a presença de múltiplas instituições que regulam o funcionamento dos mercados de bens, serviços e fatores, bem como a existência de retornos crescentes de escala em determinados setores de atividade, afetam a eficiência, seja no nível da empresa, no nível setorial ou mesmo em uma perspectiva macroeconômica. Desse modo, diferentes níveis de crescimento econômico, estruturas produtivas e sociais, assim como os vínculos que cada país cria com o resto do globo, podem influenciar na trajetória produtiva do país (CEPAL et al., 2018).

As MPEs podem desempenhar um papel importante no processo de desenvolvimento econômico no pais, visto que este é fortemente influenciado pelas estruturas produtivas locais. No contexto brasileiro, em que essas firmas dominam os setores industrial e comercial da economia, as MPEs são importantes para a promoção de uma melhor distribuição de renda, uma vez que grande parte emprega apenas proprietários e membros da família (GARCIA, 2007).

2.1.4 Eficiência industrial

A indústria se encontra em situação de retração e baixo dinamismo, tanto em termos de produtividade quanto de tecnologia. Entre os diversos fatores que têm sido discutidos que influenciam a queda contínua da participação dos itens industriais no produto total estão a baixa produtividade e competitividade da indústria brasileira e a necessidade de melhorar a taxa de investimento (BONELLI; VELOSO; PINHEIRO, 2017).

O estabelecimento de redes de cooperação entre empresas no segmento de MPMEs constitui um importante instrumento para estimular o crescimento destas empresas e fomentar processos de desenvolvimento local/regional, uma vez que limitações são agravadas na medida em que essas empresas operam de forma isolada (CEPAL et al., 2018).

HOUAISS (2020) define eficiência como a capacidade de realizar tarefas ou trabalhos de modo eficaz e com o mínimo de desperdício. Por outro lado, Coelli (2003) argumenta que existem diferentes tipos de eficiência, sendo elas:

- Eficiência Técnica: habilidade de atingir o máximo da produção dado um número de insumos;
- Eficiência de Escala: mede o grau de otimização do tamanho da operação e;

- Eficiência Alocativa: habilidade de seleção da combinação adequada de insumos, dados os preços e tecnologias disponíveis.

A Confederação Nacional da Indústria (CNI) estimou que, entre os anos de 2010 e 2016, a produtividade no trabalho na indústria brasileira cresceu 5,5%, enquanto a produtividade dos Estados Unidos cresceu 16,2%, e a da Argentina, 11,2%. Tal viés gera uma forte perda de competitividade da indústria brasileira tanto no mercado internacional quanto no mercado doméstico, que enfrenta a concorrência de produtores globais (CEPAL et al., 2018).

A produtividade agregada dos Estados Unidos é cerca de 6 vezes maior que a do Brasil, o que evidencia a grande distância do Brasil em relação à fronteira tecnológica. Embora a agropecuária seja o setor com maior crescimento da produtividade no Brasil nas últimas duas décadas, a produtividade do setor nos Estados Unidos ainda é cerca de 14 vezes maior que a brasileira (BONELLI; VELOSO; PINHEIRO, 2017).

À vista disso, Bonelli, Veloso e Pinheiro (2017) citam que a produtividade da indústria americana é 5,7 vezes maior que a do Brasil. Deve-se mencionar que na indústria de transformação a distância entre os países é ainda maior, chegando a 6,3.

Há elevado volume de recursos alocado a firmas pequenas e de baixa eficiência, assim como há ausência de empresas com níveis médio e alto de eficiência, o que contribui para a baixa produtividade na América Latina (BONELLI; VELOSO; PINHEIRO, 2017). A Figura 7 apresenta a produtividade relativa de MPMEs de países da América Latina e OCDE com relação à produtividade das grandes empresas desses países.

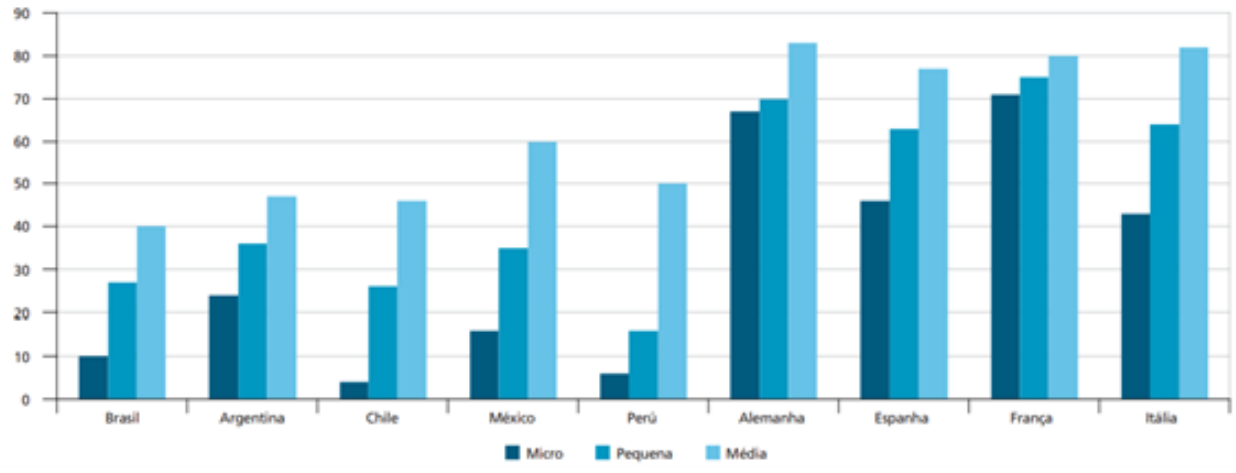


Figura 7 – Produtividade relativa em países selecionados da América Latina e OCDE (em % da produtividade das grandes empresas)

Fonte: Nogueira e Pereira (2015).

Diante disso, é possível observar que, embora presentes em maior número na economia, MPMEs no Brasil apresentaram menor produtividade com relação às grandes empresas,

demonstrando a necessidade de atenção à eficiência dessas.

2.2 Pesquisa Operacional

Considerada uma das primeiras abordagens científicas nas operações dentro das organizações, a pesquisa operacional teve sua origem nas ações militares nos primórdios da Segunda Guerra Mundial (HILLIER; LIEBERMAN, 2013). Possui seus fundamentos na matemática, lógica, estatística e ciência da computação, permitindo a resolução de problemas complexos nas organizações que abrangem muitas restrições, com o intuito de apoiar à tomada de decisão mais lógica e benéfica para as operações (GUPTA, 1992).

A pesquisa operacional gerou relevante impacto na melhoria da eficiência de inúmeras organizações pelo mundo, além de contribuir significativamente para o aumento da produtividade da economia de diversos países (HILLIER; LIEBERMAN, 2013).

A aplicação da pesquisa operacional para solução de um problema se dá em seis passos (HILLIER; LIEBERMAN, 2013):

1. Definição do problema e coleta de dados;
2. Construção de um modelo científico, tipicamente matemático, para representação do problema;
3. Desenvolvimento de um procedimento computacional para gerar soluções do problema modelado;
4. Teste das hipóteses levantadas;
5. Aplicação contínua do modelo;
6. Implementação do modelo.

Diversas técnicas de Pesquisa Operacional disponíveis são aplicadas de acordo com a particularidade de cada caso. Algumas destas técnicas são: Programação Linear, Análise de decisão, Simulação, PERT/CPM, Teoria das filas e Scheduling.

2.2.1 Programação linear

Problemas de Programação Linear (PL) consistem em funções matemáticas que são representadas por funções lineares. Sendo considerada uma função linear aquela que envolve apenas constantes e variáveis de primeira ordem. Neste caso, em geral são utilizadas variáveis de decisão contínuas, podendo assumir qualquer valor em um intervalo real (BELFIORE; FÁVERO, 2013).

Um problema de Programação Linear pode ser definido como um problema de maximização ou minimização de uma função linear sujeita a restrições lineares, as quais podem ser equações ou inequações. Dentre os principais termos adotados para a área, tem-se?

- Solução: Uma composição de valores para as variáveis de decisão;
- Solução viável: Uma solução que satisfaz todas as restrições;
- Região viável: A série de todas as soluções viáveis ao problema;
- Solução ótima: A solução viável com o maior valor para a função objetivo, em um problema de maximização, ou com o menor valor em um problema de minimização.

Além disso, conforme Lewis (2008) assume-se quatro premissas básicas em problemas de programação linear:

- Proporcionalidade: A contribuição de qualquer variável à função objetivo ou restrições é proporcional àquela variável; ou seja, não há descontos e nem economia de escala;
- Aditividade: Toda função em um modelo de programação linear é a soma das contribuições individuais das respectivas atividades;
- Divisibilidade: Os valores das variáveis de decisão podem ser fracionados;
- Certeza: Todos os parâmetros (coeficientes da função objetivo e restrições) do modelo são conhecidos com exatidão. Em aplicações reais, entretanto, coeficientes e parâmetros são muitas vezes resultado de estimativas e aproximações.

De modo geral, um problema de PL pode ser representado da seguinte forma (GOEMANS, 2015):

$$Maximizar\ ou\ minimizar\ z = c_0 + c_1x_1 + ... + c_nx_n$$

Sujeito a

$$ai_1x_1 + ai_2x_2 + ... + ai_nx_n \leq b_i \tag{2.1}$$

$$x_j \begin{cases} \leq 0 \\ \geq 0 \end{cases}$$

$$j = 1, ..., n;$$
$$i = 1, ..., m;$$

Uma das aplicações mais tradicionais para a programação linear é com relação à otimização da alocação de recursos: com uma quantidade limitada de recursos, um negócio visa o atingimento dos melhores resultados possíveis. Dentre as diversas metodologias que utilizam aplicações de PO, tem-se a Análise Envoltória de Dados, com foco na resolução de problemas de Programação Linear, tal método é detalhado no tópico a seguir.

2.2.2 Análise Envoltória de Dados

Há quatro métodos de descobrir, interpretar e comunicar padrões significativos em dados: a análise descritiva, que sintetiza os dados em gráficos; a análise preditiva, que utiliza de técnicas estatísticas e de aprendizado de máquina para analisar o desempenho histórico em um esforço para prever o futuro; a análise prescritiva, que recomenda decisões usando técnicas de otimização e; a análise decisiva, que apoia as decisões humanas com ferramentas visuais (ZHU, 2014).

A análise envoltória de dados se encontra na categoria de análise prescritiva, uma vez que tem por objetivo ser insumo para o processo de decisão de eficiência métrica em uma unidade tomadora de decisões, denominada *Decision Making Unit* (DMU) (ZHU, 2014).

Proposta por Charnes, Cooper e Rhodes (1978), a Análise Envoltória de Dados, do inglês *Data Envelopment Analysis* (DEA) , é uma técnica da Pesquisa Operacional baseada em uma programação matemática com o objetivo de fornecer uma avaliação de eficiência relativa para um grupo de unidades tomadoras de decisões, de modo a analisar as entradas (*inputs*) e saídas (*outputs*) das unidades.

A partir da publicação de 1978, houve um crescimento significativo no número de artigos relacionados à DEA por ano (EMROUZNEJAD; YANG, 2018). A Figura 8 apresenta a evolução de publicações dos anos de 1978 a 2016.

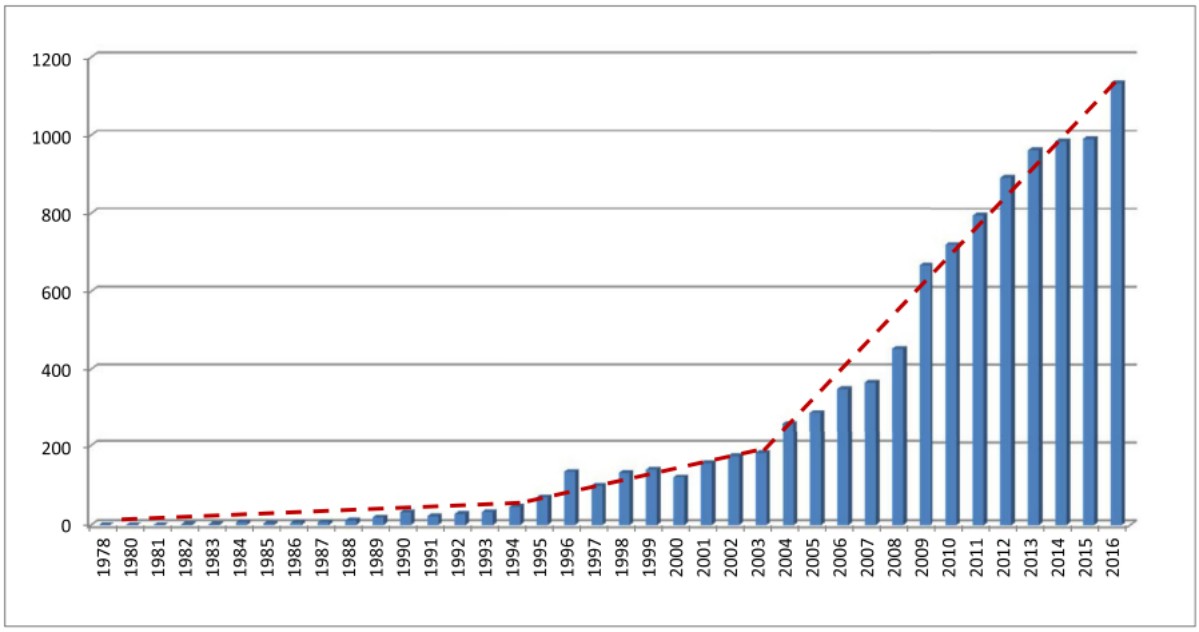


Figura 8 – Distribuição de artigos relacionados à DEA por ano (1978-2016)

Fonte: Emrouznejad e Yang (2018).

Nesse sentido, as principais áreas de atuação nas quais foram aplicadas a metodologia DEA nos anos de 2015 e 2016 são apresentadas no Quadro 1.

Quadro 1 – Campos de aplicação da metodologia DEA (2015-2016)

Posição	**Área**
1º	Agricultura
2º	Bancos
3º	Cadeia de suprimentos
4º	Transporte
5º	Políticas públicas

Fonte: Emrouznejad e Yang (2018).

A metodologia de DEA consiste em uma análise de operações orientada a dados que analisa várias métricas de desempenho, integra dados multidimensionais em um índice composto e recomenda direções para melhoria. O procedimento de comparar unidades produtoras para encontrar as eficientes e não eficientes, pode ser formulado como um problema de Programação Linear. A técnica para analisar a eficiência de n unidades produtoras é um conjunto de N problemas de otimização lineares para serem resolvidos (ZHU, 2014).

DEA é comumente denominada de benchmarking de equilíbrio, uma vez que considera diferentes métricas de eficiência em um modelo único, de modo a identificar quais as melhores práticas por meio de uma fronteira de eficiência (SHERMAN; ZHU, 2013).

Diante disso, tem-se que a eficiência de uma DMU é atingida apenas se nenhuma de suas métricas puder ser melhorada sem piorar algumas de suas outras métricas. Por outro lado, para a eficiência relativa, uma unidade deve ser avaliada como em plena eficiência com base nas evidências disponíveis, caso os desempenhos de outras unidades não apresentem pontos de melhoria para as métricas já adotadas (ZHU, 2014).

Desse modo, DEA tem por foco a eficiência relativa, uma vez que compara o desempenho de uma DMU com o de outras, a eficiência é calculada, portanto, a partir do desempenho de um conjunto de unidades tomadoras de decisão (ZHU, 2014).

Pode-se afirmar que uma DMU é eficiente quando não for mais possível (KOOPMANS, 1951):

- Aumentar a quantidade de qualquer um dos produtos por ela gerado sem, simultaneamente, ser necessário reduzir a quantidade de outro produto gerado ou aumentar as quantidades dos insumos consumidos;
- Diminuir a quantidade de qualquer um dos insumos por ela consumidos sem, simultaneamente, ser necessário aumentar a quantidade de outro insumo consumido ou diminuir as quantidades de produtos gerados.

2.2.3 Modelos de DEA

A primeira proposição de Charnes, Cooper e Rhodes (1978) foi designada CCR, devido ao nome dos autores e esse modelo matemático inicial foi construído para uma análise com Retornos Constantes de Escala (RCE), indicando que variações nos inputs produzem variações proporcionais nos outputs. O modelo pode ser representado por meio da seguinte equação:

Modelo RCE:

$$Max\ e_{j0},\ sujeito\ a\ e_j \leq 1 \tag{2.-3}$$

$$e_j = \frac{\sum_{r=1}^{s} u_r Y_{rj}}{\sum_{i=1}^{m} v_i X_{ij}} \tag{2.-3}$$

$$j = 1, ..., n;\ r = 1, ..., s;\ i = 1, ..., m.$$

Onde:

- X_{ij}: Inputs;
- Y_{rj}: Outputs;

- u_i: Peso atribuído à variável;
- v_r: Peso atribuído à variável;
- i: número de inputs;
- r: número de outputs.

O denominador da função objetivo do problema de otimização pode ser limitado a 1 de forma que o modelo possa ser transformado em um problema de PL. Para resolução, utiliza-se a seguinte linearização equivalente:

Orientado ao *Input*

$$Max \sum_{r=1}^{s} u_r Y_{rj} \tag{2.-4}$$

sujeito a:

$$\sum_{i=1}^{m} v_i X_{ij} = 1 \tag{2.-4}$$

$$\sum_{r=1}^{s} u_r Y_{rj} < \sum_{i=1}^{m} v_i X_{ij} \tag{2.-3}$$

$$j = 1, ..., n;\ r = 1, ..., s;\ i = 1, ..., m;\ v_i, u_r \geq 0$$

Assim, a eficiência de uma DMU pode ser maximizada por meio de duas abordagens. Quando maximizada a partir de uma redução nos níveis de *input*, mantendo o nível de output, tal análise é denominada Orientada ao Input- *Input Oriented* (IO). Quando maximizada a partir do aumento da produção, conservando os níveis de consumo dos insumos, é denominada Orientada ao *Output- Output Oriented* (OO) (CHARNES; COOPER; RHODES, 1978).

A Figura 9 apresenta o gráfico gerado para uma aplicação orientada ao *input* para o método de retornos constantes de escala.

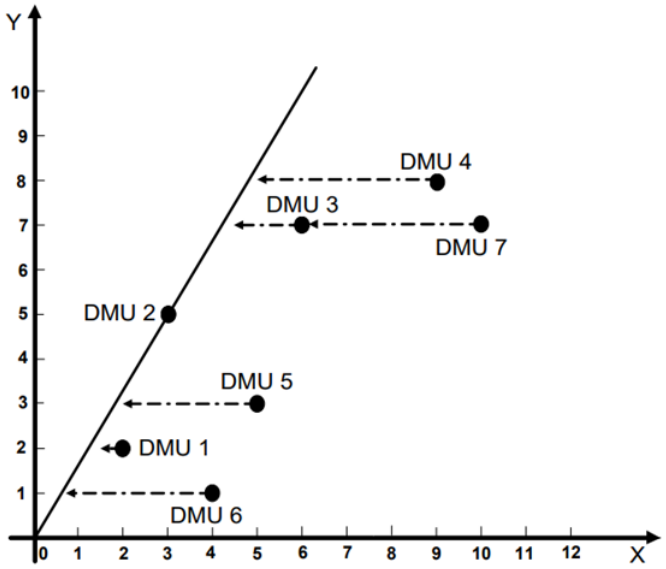


Figura 9 – Envoltória orientada ao input determinada pelo modelo CCR

Fonte: Souza (2008)

Por outro lado, o cálculo para a orientação ao *output*, dá-se da seguinte forma:

Orientado ao *Output*

$$Min\ \Theta_{CRS} \tag{2.-4}$$

sujeito a

$$\sum_{j=1}^{n} \Lambda_j X_{ij} \leq \Theta_{CRS} X_{ij_0} \quad i = 1,...,m; \tag{2.-4}$$

$$\sum_{j=1}^{n} \Lambda_j Y_{rj} \geq Y_{r0} \quad r = 1,...,s; \tag{2.-4}$$

$$\Lambda_j \geq 0 \quad j = 1,...,n; \tag{2.-4}$$

Banker, Charnes e Cooper (1984) desenvolveram um modelo para inclusão de Retornos Variáveis de Escala (RVE) que passou a chamar-se BCC. Nesse modelo, é estimada a eficiência técnica pura e a eficiência técnica de escala, na qual é possível o detalhamento das alterações de escala. O modelo pode ser representado por meio da seguinte equação:

Modelo RVE:

$$Min\ \Theta_{vrs} \tag{2.-4}$$

sujeito a

$$\sum_{j=1}^{n} \Lambda_j X_{ij} \leq \Theta_{vrs} X_{ij_0} \quad i = 1,...,m; \tag{2.-4}$$

$$\sum_{j=1}^{n} \Lambda_j Y_{rj} \geq Y_{r0} \quad r = 1,...,s; \tag{2.-4}$$

$$\sum_{j=1}^{n} \Lambda_j = 1 \tag{2.-4}$$

$$\Lambda_j \geq 0 \quad j = 1,...,n; \tag{2.-4}$$

A Figura 10 apresenta o gráfico gerado para uma aplicação orientada ao *input* para o método de retornos constantes de escala.

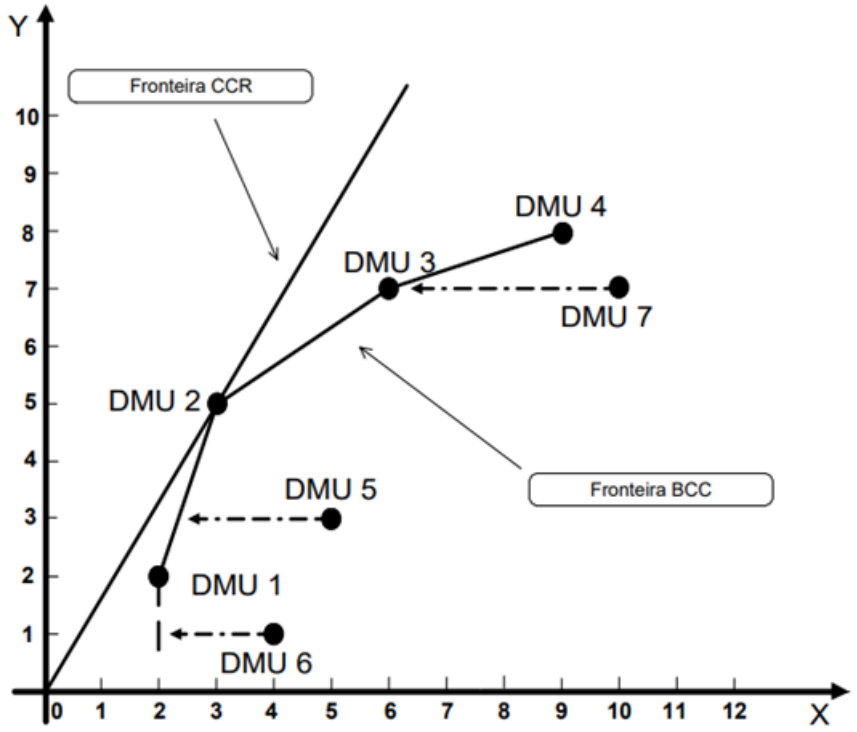


Figura 10 – Envoltória orientada ao input determinada pelos modelos CCR e BCC

Fonte: Souza (2008)

2.2.4 Fronteira de eficiência

DEA é aplicada sobre os dados de forma a construir uma fronteira de eficiência, formada pelas DMUs mais eficientes, ou seja, com a melhor relação entre insumo e produto. A posição das demais DMUs em relação a essa fronteira é denominada envoltória, já que a fronteira é criada de forma a envolver todas as DMUs, nenhuma DMU pode ficar além da

curva. O alvo de uma unidade produtora que não é eficiente se encontra na interseção com a fronteira partindo da origem.

Para definir a diferença entre aplicação da DEA e regressão linear, verifica-se que a regressão fornece um comportamento médio e a DEA se concentra nas melhores práticas que “flutuam” em cima de um conjunto de dados. Além de a regressão precisar especificar uma forma funcional entre a variável dependente e as variáveis independentes (ZHU, 2014).

A Figura 11 apresenta a diferença gráfica entre os resultados gerados pelas regressão linear (linha pontilhada) e aplicação da DEA (fronteira sobre a qual estão os pontos descriminados) para um conjunto de dados.

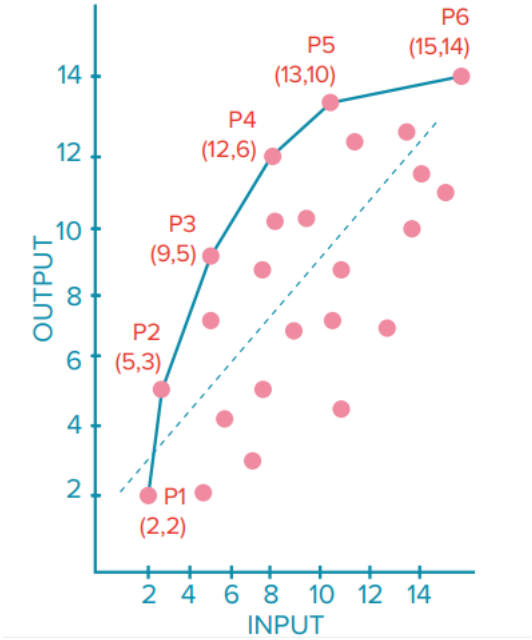


Figura 11 – Comparação entre DEA e regressão linear

Fonte: TCU (2019)

A eficiência é uma medida relativa na DEA e varia entre 0 e 1, DMUs mais eficientes são representadas pelo valor 1. DMUs que estão sobre a fronteira recebem a pontuação máxima. Para calcular a eficiência de DMUs que estão fora da fronteira, a DEA cria uma projeção de cada DMU ineficiente sobre a fronteira com base nas DMU que se situam sobre ela.

Diante disso, quando uma DMU é ineficiente, ou está operando abaixo da sua performance, indicando uma razão menor que 1, essa DMU será, portanto, envolta pelas DMU de melhores práticas (ZHU, 2014).

3 Metodologia

Este estudo é uma pesquisa aplicada, uma vez que possui como objetivo gerar conhecimento para aplicação prática e é direcionado à solução de problemas específicos (FONSECA, 2002). Sua abordagem é quantitativa, com o intuito de quantificar os dados e generalizar os resultados da amostra para a população alvo (MALHOTRA, 2001).

Além disso, possui viés exploratório, à medida em que busca ampliar o conhecimento a respeito de um assunto, e explicativo, com a identificação dos fatores que determinam ou contribuem para a ocorrência de fenômenos (GIL, 2008).

Esta seção dividiu-se na coleta dos dados e procedimentos metodológicos necessários para a aplicação de DEA.

3.1 Coleta dos dados

A coleta dos dados confiáveis é de suma importância para êxito do projeto, sendo a primeira etapa de aplicação do método definido. Foram utilizados dados públicos em consonância com a Lei de Acesso à Informação (Lei nº 12.527/2011), provenientes de duas fontes: Cadastro Nacional da Pessoa Jurídica (CNPJ) e dados abertos da Dívida Ativa da União (DAU).

O Cadastro Nacional da Pessoa Jurídica é um banco de dados gerenciado pela Secretaria Especial da Receita Federal do Brasil (RFB), que armazena informações cadastrais das pessoas jurídicas e outras entidades de interesse das administrações tributárias da União, dos Estados, do Distrito Federal e dos Municípios. É disponibilizado em formato de banco de dados SQL, as tabelas foram organizadas e armazenadas utilizado o software DB Browser, que utiliza a biblioteca SQLite.

Atendendo às melhores práticas de transparência ativa, a Procuradoria Geral da Fazenda Nacional (PGFN) publica trimestralmente a base completa dos créditos inscritos em dívida ativa, uma vez que tais débitos não estão cobertos por sigilo, conforme o Código Tributário Nacional (CTN). Os dados são disponibilizados por meio de arquivos no formato csv, divididos entre os estados e o Distrito Federal.

O banco de dados do CNPJ foi extraído em maio de 2022, portanto os dados apresentados representam a situação das empresas neste recorte temporal. Com relação à base da DAU, os dados referem-se aos créditos ativos no primeiro trimestre de 2022.

Com relação à base da RFB, a tabela com informações dos estabelecimentos expõe os dados de empresas por instalação, ou seja, pode incluir mais de um cadastro por empresa

devido ao número de filiais, constituindo uma relação de *n* para 1 com a empreza matriz. Os campos que a compõem são apresentados no Quadro 2.

Quadro 2 – Informações do estabelecimento

Campo	Descrição
CNPJ básico	Número base de inscrição no CNPJ (oito primeiros dígitos do CNPJ)
CNPJ ordem	Número do estabelecimento de inscrição no CNPJ (do nono até o décimo segundo dígito no CNPJ)
CNPJ DV	Dígito verificador do número de inscrição no CNPJ (dois últimos dígitos do CNPJ)
Identificador Matriz/Filial	Código do identificador Matriz/Filial: 1- Matriz 2- Filial
Nome fantasia	Corresponde ao nome fantasia
Situação cadastral	Código da situação cadastral: 01 – nula 2 – ativa 3 – suspensa 4 – inapta 08 – baixada
Data situação cadastral	Data do evento da situação cadastral
Motivo situação cadastral	Código do motivo da situação cadastral
Nome da cidade no exterior	Nome da cidade no exterior
País	Código do país
Data de início atividade	Data de início da atividade
CNAE fiscal principal	Código da atividade econômica principal do estabelecimento
CNAE fiscal secundária	Código da(s) atividade(s) econômica(s) secundárias do estabelecimento
Tipo de logradouro	Descrição do tipo de logradouro
Logradouro	Nome do logradouro onde se localiza o estabelecimento
Número	Número onde se localiza o estabelecimento. Quando não houver preenchimento do número haverá 'S/N'
Complemento	Complemento para o endereço de localização do estabelecimento
Bairro	Bairro onde se localiza o estabelecimento
CEP	Código de endereçamento postal referente ao logradouro no qual o estabelecimento está localizado
UF	Sigla da unidade da federação em que se encontra o estabelecimento
Município	Código do município de jurisdição onde se encontra o estabelecimento
Situação especial	Situação especial da empresa
Data da situação especial	Data que a empresa entrou em situação especial

Fonte: Receita Federal do Brasil (2022).

As informações básicas de todas as empresas cadastradas, como CNPJ, razão social,

capital social e também o porte da empresa, são apresentadas no Quadro 3, tais campos são importantes filtros para aplicação do escopo definido.

Quadro 3 – Informações da empresa.

Campo	Descrição
CNPJ básico	Número base de inscrição no CNPJ (oito primeiros dígitos do CNPJ)
Razão Social/ Nome empresarial	Nome empresarial da pessoa jurídica
Natureza jurídica	Código da natureza jurídica
Qualificação do responsável	Qualificação da pessoa física responsável pela empresa
Capital social da empresa	Capital social da empresa
Porte da empresa	Código do porte da empresa: 00- Não informado 01- Micro empresa 03- Empresa de pequeno porte 05- Demais
Ente federativo responsável	O ente federativo responsável é preenchido para os casos de órgãos e entidades do grupo de natureza jurídica 1xxx. Para as demais naturezas, este atributo fica em branco

Fonte: Receita Federal do Brasil (2022).

Informações referentes aos sócios que iniciaram a empresa, como nome, qualificação e data de entrada na sociedade, são apresentadas no Quadro 4.

Quadro 4 – Informações dos sócios

Campo	Descrição
CNPJ básico	Número base de inscrição no CNPJ (oito primeiros dígitos do CNPJ)
Identificador de sócio	Código do identificador de sócio 1 – pessoa jurídica 2 – pessoa física 3 – estrangeiro
Nome do sócio (no caso PF) ou razão social (no caso PJ)	Nome do sócio pessoa física ou a razão social e/ou nome empresarial da pessoa jurídica e/ou nome do sócio/razão social do sócio estrangeiro
CNPJ/CPF do sócio	CPF ou CNPJ do sócio (sócio estrangeiro não tem esta informação)
Qualificação do sócio	Código da qualificação do sócio
Data de entrada sociedade	Data de entrada na sociedade
País	Código país do sócio estrangeiro
Representante legal	Número do CPF do representante legal
Nome do representante	Nome do representante legal
Qualificação do representante legal	Código da qualificação do representante legal
Faixa etária	Código correspondente à faixa etária do sócio

Fonte: Receita Federal do Brasil (2022).

A estrutura da tabela com informações referentes ao código CNAE da empresa é exposta no Quadro 5, esta é a tabela pela qual serão definidos os dados do setor analisado, especificamente, o setor alimentício.

Quadro 5 – Informações de CNAE

Campo	Descrição
Código	Código da atividade econômica
Descrição	Nome da atividade econômica

Fonte: Receita Federal do Brasil (2022).

Com relação à base da PGFN, a publicação abrange o conjunto de informações sobre débitos com a Fazenda Nacional ou FGTS inscritos em Dívida Ativa, em todas as situações, incluindo seus devedores, na condição de devedor principal, corresponsável ou solidário. A estrutura é apresentada no Quadro 6.

Quadro 6 – Dados abertos da dívida ativa

Campo	Descrição
cpf_cnpj	Número identificador do contribuinte no cadastro de pessoas físicas ou no cadastro nacional de pessoas jurídicas
data_inscricao	Data em que o crédito foi insrcito em dívida ativa
entidade_responsavel	Indica se o débito de FGTS está sendo cobrado pela PGFN ou pela Caixa Econômica Federal
indicador_ajuizado	Indica se o crédito está sendo cobrado judicialmente
nome_devedor	Nome do devedor
numero_inscricao	Número da inscrição em dívida ativa
receita_principal	Receita do crédito que está sendo cobrado
situacao_inscricao	Situação da inscrição no sistema de controle de créditos
tipo_devedor	Indica se o devedor é principal (titular original da dívida) ou corresponsável (foi vinculado posteriormente à dívida)
tipo_pessoa	Indica se é uma pessoa física ou jurídica
tipo_situacao_inscricao	Indica se a inscrição está em cobrança (situação irregular), em benefício fiscal (em parcelamento ou moratória), em negociação, suspenso por decisão judicial, garantia (integralmente garantida)
uf_unidade_responsavel	Unidade federativa da unidade da PGFN responsável pela cobrança do devedor
unidade_inscricao	Indica a unidade da PGFN que realizou a inscrição em dívida ativa
unidade_responsavel	Unidade da PGFN responsável pelo acompanhamento do devedor
valor_consolidado	Valor do débito na data de extração, com acréscimos legais

Fonte: PGFN (2022)

3.1.1 Limpeza e manipulação dos dados

A partir do entendimento das informações das bases e tabelas, faz-se necessária a conexão entre elas por meio das chaves comuns a fim de gerar um conjunto único de dados com todas as informações que serão necessárias para a aplicação do modelo no projeto. Com relação à conexão entre as tabelas da RFB, os seguintes passos foram executados:

1. A tabela de informações de empresas é conectada com a tabela de informações de estabelecimentos utilizando o campo CNPJ básico como chave;
2. A tabela de informações de estabelecimentos é conectada com a tabela de informações do simples utilizando o campo CNPJ básico como chave;
3. A tabela de informações de estabelecimentos é conectada com a tabela de informações dos sócios utilizando o campo CNPJ básico como chave;
4. A tabela de informações de estabelecimentos é conectada com a tabela de informações de CNAE utilizando o código dos campos CNAE Fiscal Principal e Código, respectivamente.

Com relação à base de dados da PGFN, os seguintes passos foram executados:

1. As diferentes planilhas separadas por unidade da federação foram agrupadas;
2. Os dados avindos do conjunto único da RFB foram relacionados aos dados abertos da PGFN por meio do CNPJ, proveniente do campo "cnpj_basico" e "cpf_cnpj" das respectivas bases.

3.2 Aplicação de DEA

Definidas as fontes de dados, a aplicação de DEA deve ocorrer em três fases (GOLANY; ROLL, 1989):

1. Definição de DMUs a serem analisadas;
2. Seleção das variáveis (insumos e produtos) relevantes e apropriadas para estabelecer a eficiência relativa das DMUs;
3. Execução do modelo DEA em programas de computador.

3.2.1 Definição de DMUs

Para definição do universo de DMUs a ser analisado, é fundamental escolher uma amostra homogênea, a fim de não analisar motivos externos às variáveis que farão parte do estudo. Diante disso, às seguintes restrições foram atendidas:

- Empresas ativas, conforme classificação do SEBRAE, situadas com atividade econômica mercantil, ou seja, com fins lucrativos;
- Empresas com dados de porte igual a micro ou pequena, conforme classificação da Receita Federal, declarado no momento de abertura da empresa;
- Empresas que possuem dados sobre os sócios disponíveis;
- Apenas as matrizes foram selecionadas;
- Apenas os sócios que iniciaram a empresa foram selecionados;
- Empresas com data de abertura entre os anos de 2018 e 2019. Buscou-se utilizar um número limitado de anos a fim de os resultados não serem afetados por flutuações econômicas que pudessem interferir na eficiência calculada;
- Empresas do setor de produtos alimentícios dentro da indústria de transformação;
- Para os dados referentes à empresas inscritas em dívida ativa, só serão considerados os casos de empresas com dívida ativa na base disponibilizada no primeiro trimestre de 2022.

3.2.2 Seleção de variáveis

A partir da coleta dos dados, as variáveis que comporão a análise são selecionadas de duas formas: variáveis qualitativas, que serão utilizadas para categorização dos dados, e variáveis quantitativas, que serão utilizadas para execução do modelo DEA.

Para a escolha das variáveis é fundamental investigar a correlação entre elas, coeficientes de correlação podem ser utilizados para identificar e medir a relação presentes em variáveis. Assim, faz sentido a exclusão das variáveis que são fortemente correlacionadas.

A relação de dados final a ser utilizada para análise e aplicação da metodologia DEA, é apresentada no Quadro 7.

Quadro 7 – Relação final de dados utilizados

Variável	Nome	Descrição
cnpj	CNPJ	Número identificador do contribuinte no cadastro nacional de pessoas jurídicas
situação_cadastral	Situação cadastral	Código da situação cadastral: 01 – nula 2 – ativa 3 – suspensa 4 – inapta 08 – baixada
capital_social	Capital social	Capital social da empresa
opcao_simples	Simples Nacional	Aderência ao programa: 0- Não aderência 1- Aderência
num_socios	Número de sócios	Quantidade de sócios
num_empresas_ativas_tot	Empresas ativas	Quantidade de empresas ativas relacionadas aos sócios
num_empresas_encerradas_tot	Empresas encerradas	Quantidade de empresas encerradas relacionadas aos sócios
num_filiais	Número de filiais	Quantidade de filiais associadas à matriz
valor_consolidado	Valor da dívida	Valor do débito na data de extração, com acréscimos legais
regiao	Região	Região em que a empresa está localizada
num_cnae_sec	CNAE secundário	Quantidade de CNAEs secundários

Fonte: Autoria própria.

É importante ressaltar que as empresas foram separadas em regiões previamente à execução do modelo, visto que buscou-se eliminar variações decorrentes de impostos aplicados e influências da localização. Dentre as variáveis a serem analisadas no conjunto de dados resultante, tem-se a divisão entre qualitativas e quantitativas, a fim de entender quais de fato serão utilizadas para cálculo.

Variáveis quantitativas:

- Capital social: Valor estabelecido para a empresa no momento da abertura. É a quantia bruta que é investida, o montante necessário para iniciar as atividades de uma nova empresa;
- Número de sócios: O número de sócios declarados no momento de abertura da empresa,

sem descriminação da participação de cada um na sociedade;

- Número de empresas ativas: O número de empresas ativas associadas aos sócios, ou seja, caso o sócio esteja no quadro societário de outras empresas, verificam-se as empresas cujas situações encontram-se como 'ativa';
- Número de empresas encerradas: O número de empresas encerradas associadas aos sócios, ou seja, caso o sócio esteja no quadro societário de outras empresas, verifica-se as empresas cujas situações encontra-se como 'encerrada';
- Número de filiais: O número de filiais associadas à empresa matriz;
- Valor de dívida: O valor, em reais, cadastrado na base da dívida ativa, associado ao CNPJ da empresa;
- Número de CNAEs secundários: Quantidade de atividades econômicas secundárias exercidas em uma mesma unidade produtiva.

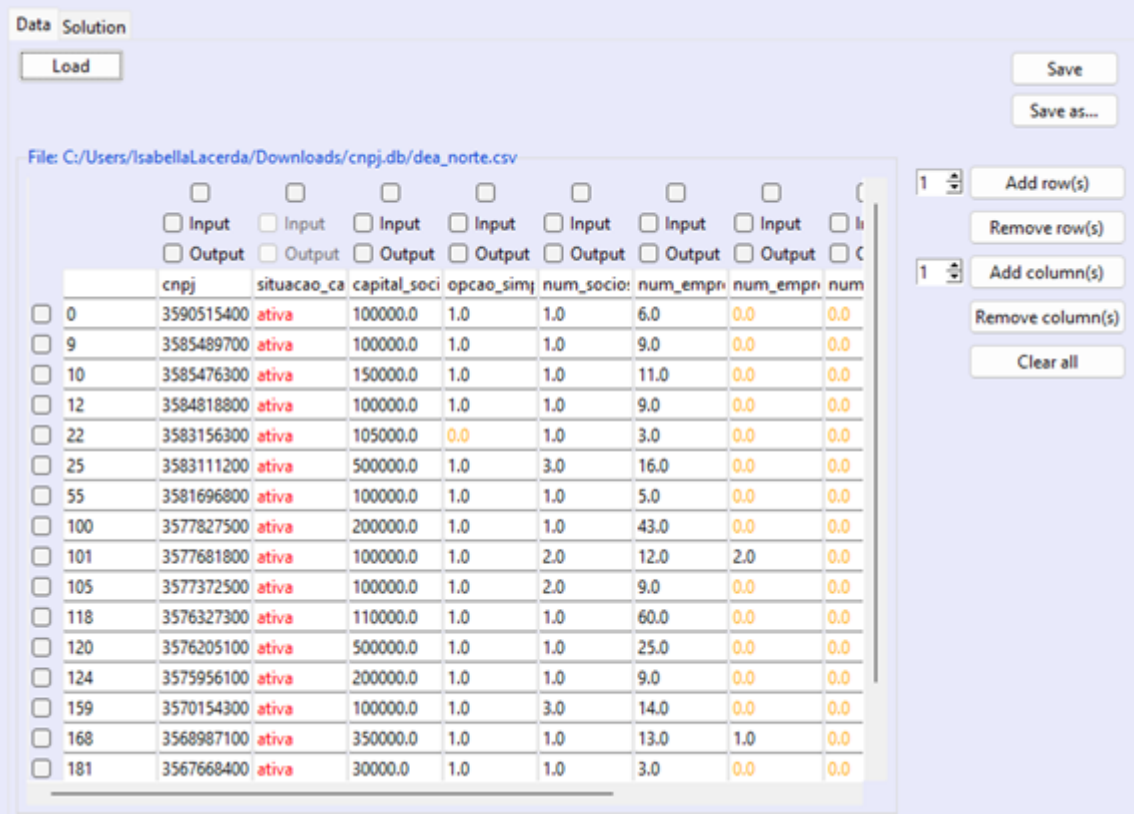
Variáveis qualitativas:

- Região: localização regional da empresa matricial;
- Aderência ao Simples: Indicador de aderência, ou não, ao programa do Simples nacional no momento de abertura da empresa.

3.2.3 Execução do modelo

O modelo foi executado por meio do pyDEA, um software de código aberto desenvolvido em python que permite que os dados sejam importados e exportados por meio de uma planilha Excel.

A planilha com os dados do Quadro 7 é carregada no software, os dados são carregados na tabela interna do programa como mostra, a Figura 12, nele são apresentados em vermelho os dados que não podem ser utilizados para cálculo, como é o caso das variáveis qualitativas.



		cnpj	situacao_ca	capital_soci	opcao_simj	num_socio:	num_empr	num_empr	num
☐	0	3590515400	ativa	100000.0	1.0	1.0	6.0	0.0	0.0
☐	9	3585489700	ativa	100000.0	1.0	1.0	9.0	0.0	0.0
☐	10	3585476300	ativa	150000.0	1.0	1.0	11.0	0.0	0.0
☐	12	3584818800	ativa	100000.0	1.0	1.0	9.0	0.0	0.0
☐	22	3583156300	ativa	105000.0	0.0	1.0	3.0	0.0	0.0
☐	25	3583111200	ativa	500000.0	1.0	3.0	16.0	0.0	0.0
☐	55	3581696800	ativa	100000.0	1.0	1.0	5.0	0.0	0.0
☐	100	3577827500	ativa	200000.0	1.0	1.0	43.0	0.0	0.0
☐	101	3577681800	ativa	100000.0	1.0	2.0	12.0	2.0	0.0
☐	105	3577372500	ativa	100000.0	1.0	2.0	9.0	0.0	0.0
☐	118	3576327300	ativa	110000.0	1.0	1.0	60.0	0.0	0.0
☐	120	3576205100	ativa	500000.0	1.0	1.0	25.0	0.0	0.0
☐	124	3575956100	ativa	200000.0	1.0	1.0	9.0	0.0	0.0
☐	159	3570154300	ativa	100000.0	1.0	3.0	14.0	0.0	0.0
☐	168	3568987100	ativa	350000.0	1.0	1.0	13.0	1.0	0.0
☐	181	3567668400	ativa	30000.0	1.0	1.0	3.0	0.0	0.0

Figura 12 – Interface do pyDEA

Fonte: Autoria própria.

Como é possível observar, a primeira coluna é reservada para os "nomes" das DMUs, colunas consecutivas podem ser usadas como entradas e saídas, é preciso selecionar a opção correspondente na interface apresentada.

Os dados de entrada são exibidos na guia *Data* na janela principal. Colunas de dados incompletas, ou aquelas com dados inválidos, como dados nulos ou negativos, não podem ser escolhidas. Os resultados são exibidos na guia *Solution* da janela principal, como pontuações de eficiência entre outros tipos de saída, como pares, contagens de pares, pesos e valores alvos.

A partir da orientação desejada da aplicação, as variáveis utilizadas no modelo devem ser categorizadas como *inputs* ou *outputs*, podendo ser realizadas algumas análises, utilizando retornos de escalas variáveis, constantes ou ambos, bem como a orientação e método. O programa ainda disponibiliza outras abordagens a serem aplicadas em conjunto, como exposto na Figura 13.

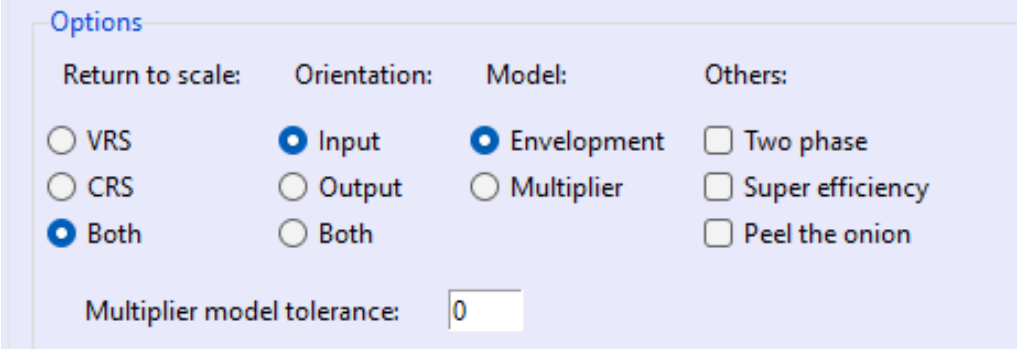


Figura 13 – Opções disponíveis no pyDEA

Fonte: pyDEA (2022)

3.3 Análise do modelo

Para análise do resultado do pyDEA, o Superset foi escolhido a fim de facilitar a interação por meio de filtros a partir das variáveis categóricas, com o intuito de gerar mais fluidez no processo de validação e análise dos dados. Foram analisados os dados produzidos a partir da análise do pyDEA, bem como análises do perfil das empresas selecionadas, geradas a partir das restrições comentadas na subseção 3.2.1 por meio das bases de dados utilizadas.

Apache Superset é uma ferramenta de *Business Intelligence* de código aberto e de fácil utilização, coleta e processa dados em grandes volumes para produzir resultados visuais, como gráficos e tabelas. Assim, a aplicação web permite que os usuários gerem *dashboards* e relatórios. A Figura 14 apresenta a interface do Superset com o exemplo de uma *dashboard*.

Figura 14 – Interface do Superset

Fonte: Superset (2022)

A interface genérica apresentada pela plataforma indica a possibilidade de criação de visualizações ricas. É um software repleto de opções que facilitam a exploração e visualização de dados por usuários de todas as habilidades, desde gráficos de linhas simples até gráficos geoespaciais altamente detalhados.

4 Resultados e Discussões

O capítulo de resultados abrange a aplicação do modelo nas etapas mencionadas na seção de metodologia, com a definição do problema da indústria analisada, a coleta e a análise dos dados, a criação e a aplicação do algoritmo, provendo conclusões sobre cada um dos resultados alcançados.

4.1 Definição de DMUs

Após aplicadas as restrições expostas na subseção 3.2.1, foram analisadas 4047 empresas distribuídas entre as regiões brasileiras, conforme apresenta a Tabela 1.

Tabela 1 – Quantidade de empresas analisadas

Região	**Quantidade de empresas**
Sudeste	1661
Sul	881
Nordeste	749
Centro-Oeste	495
Norte	261
Total	**4047**

Fonte: Autoria própria.

Embora o *benchmarking* comparativo seja realizado apenas entre DMUs de uma mesma região, como comentado na seção 3.3, foi possível analisar os perfis empresariais de cada região. Para verificação dos resultados, é importante explorar os perfis das empresas com relação às variáveis selecionadas.

Com relação a média do capital social, a região norte, apresentou o maior valor, cerca de R$ 173 mil reais, a região sul apresentou o menor valor, com cerca de R$ 77 mil reais, a Figura 15 mostra os resultados encontrados.

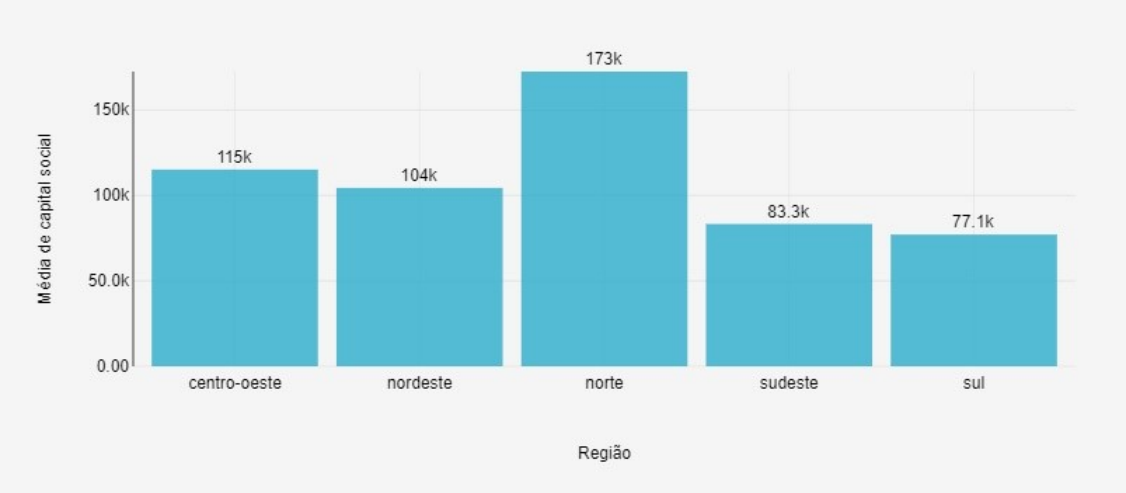


Figura 15 – Valor médio de capital social por região

Fonte: Autoria própria

.Já com relação ao número de sócios, todas as regiões apresentaram valores semelhantes, por todo o país a média foi menor que dois, nessa situação a região sudeste se destacou com o maior valor, como apresentado na Figura 16.

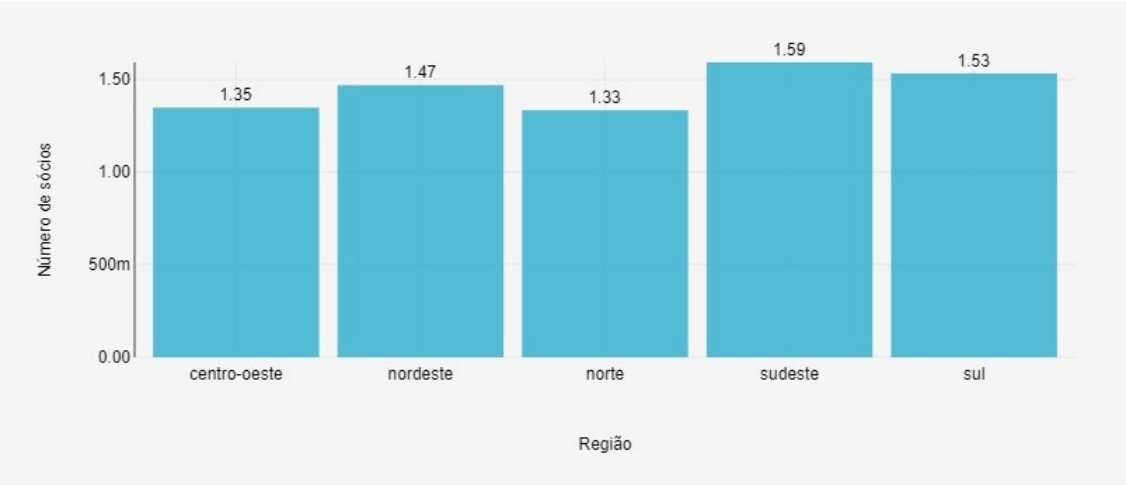


Figura 16 – Número médio de sócios por região

Fonte: Autoria própria

4.2 Seleção de variáveis

A análise de correlação foi implementada em linguagem Python, e retorna uma matriz simétrica quadrada de 7x7, na qual a diagonal principal assume valor 1. Essa matriz mostra os índices de correlação entre as variáveis e, se esse índice for igual ou superior a 0.85, foi considerado que essas variáveis possuem uma forte correlação entre elas, sendo portanto

descartadas do o modelo. Notou-se que, nenhuma das variáveis relacionadas possuía forte correlação, como apresenta a Figura 17.

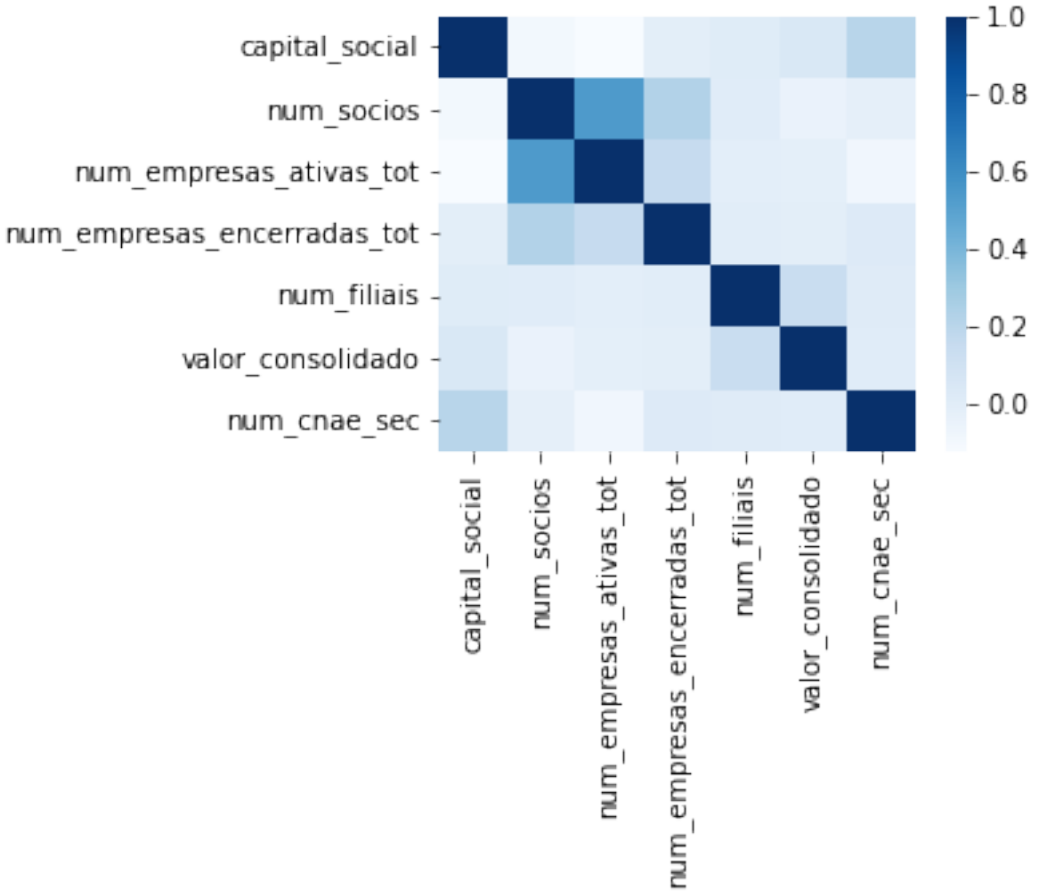


Figura 17 – Matriz de correlação

Fonte: Autoria própria

A matriz apresenta, respectivamente, os nomes das variáveis referentes a capital social, número de sócios, número de empresas ativas, número de empresas encerradas, número de filiais, valor da dívida ativa e número de CNAEs secundários.

A partir da seleção das variáveis, é necessário entender quais devemos manter, aumentar ou diminuir, uma vez que por meio desse entendimento são definidos os *inputs* e *outputs* do problema.

As variáveis número de sócios, número de empresas encerradas e valor da dívida são do tipo quanto maior, pior para a empresa. O objetivo da aplicação do DEA será, portanto, diminuí-las para maximização da eficiência.

Por outro lado, capital social, número de empresas ativas, número de filiais e número de CNAEs secundários são entendidas como variáveis do tipo quanto maior melhor para a empresa, geram maior confiança e promovem maior segurança para o estabelecimento.

Devem, portanto, ser mantidas ou maximizadas.

Diante disso, optou-se pelo modelo DEA com orientação ao *input*, quando a eficiência é maximizada a partir de uma redução nos níveis de *input*, mantendo os níveis de *output*. As variáveis número de sócios, número de empresas encerradas e valor da dívida foram definidas como os *inputs* da operação e as variáveis capital social, número de empresas ativas, número de filiais e número de CNAEs secundários foram definidas como os *outputs*. A Figura 18 mostra como são apresentadas tais variáveis no software pydea. A partir disso, é possível observar os resultados obtidos na seção seguinte

Figura 18 – Variáveis no pyDEA

Fonte: Autoria própria

4.3 Avaliação de DMUs

Como exposto previamente, as análises foram executadas por região, os valores médios para Retornos Variáveis de Escala (RVE), em que é estimada a eficiência técnica pura e a eficiência técnica de escala, e para Retornos Constantes de Escala (RCE), que indica que variações nos *inputs* produzem variações proporcionais nos *outputs* são apresentados na Tabela 2.

É importante ressaltar que os valores apresentados não denotam um ranking de eficiência entre as regiões, uma vez que os dados foram obtidos por meio de uma análise intra-regional. Todavia, apontam o *benchmarking* comparativo, sugerindo maior ou me-

nor homogeneidade nas operações de uma empresa. Por meio disso, ela pode ser, ou não, considerada eficiente dado o recorte regional.

Tabela 2 – Resultados obtidos por região

Região	RVE médio	RCE médio
norte	0.90	0.52
centro-oeste	0.86	0.47
sul	0.80	0.51
nordeste	0.80	0.35
sudeste	0.80	0.50

Fonte: Autoria própria.

É possível perceber que, em ambos os retornos de escala, a região norte obteve maior homogeneidade em valores de alta eficiência de suas empresas. Por outro lado, a região nordeste apresentou o menor resultado, com eficiência média de apenas 0,35 para a análise RCE, indicando que há, proporcionalmente, maior quantidade de empresas abaixo da eficiência esperada na região.

Os resultados da Tabela 3, apresentam os valores médios de *inputs* por região, tendo em vista que a aplicação foi orientada ao input, empresas que apresentassem menores valores de *inputs*, seriam mais eficientes.

Tabela 3 – Valores médios de *inputs*

Região	Valor médio de dívida	Número médio de sócios	Número médio de empresas encerradas
norte	4.70k	1.33	0.45
centro-oeste	9.94k	1.35	0.55
sul	8.22k	1.53	0.42
nordeste	8.07k	1.47	0.58
sudeste	11.5k	1.59	0.55

Fonte: Autoria própria.

Em um primeiro momento, é notório perceber que quanto maior os valores encontrados para os *inputs*, menor a eficiência relativa da DMU. No entanto, tal relação não é verdade para a análise dos valores de outputs, uma vez que o objetivo na aplicação de DEA era de manter seus níveis. Assim, não é possível fazer uma análise direta dos valores encontrados para RCE e RVE com relação aos valores médios de outputs. Os resultados médios de outputs por região, são apresentados na Tabela 4.

Embora as empresas da região norte tenham apresentado significativa concentração de altos valores de eficiência, é importante mencionar o baixo número de empresas analisadas

para essa região, correspondendo a apenas 6,45% da amostra, o menor número de empresas por região deste trabalho. É provável que, caso o número de empresas analisadas da região norte aumentasse, sua eficiência relativa diminuísse.

Tabela 4 – Valores médios de *outputs*

Região	Valor médio de capital social	Número médio de empresas ativas	Número médio de CNAEs secundários	Número médio de filiais
norte	173k	13.28	5.36	0.01
centro-oeste	115k	19.88	3.35	0.00
sul	77.1k	28.63	2.87	0.00
nordeste	104k	14.95	3.42	0.00
sudeste	83.3k	47.43	2.88	0.01

Fonte: Autoria própria.

Diante disso, a Figura 19 apresenta os principais resultados encontrados por região, atrelados às metodologias de cálculo da DEA. Nela são apresentados os valores RVE e RCE, separados por resultado. Um resultado RVE=1 ou RCE=1 indica que a DMU é eficiente para aquele método, já valores menores que um indicam ineficiência.

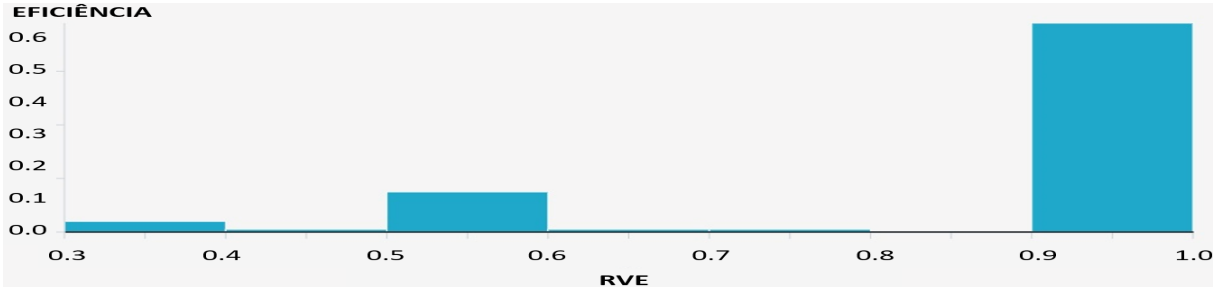
Região	RVE=1	RV1<1	RCE=1	RCE<1
norte	79%	21%	11%	89%
centro-oeste	72%	28%	6%	94%
nordeste	62%	39%	5%	95%
sudeste	52%	48%	3%	97%
sul	58%	42%	7%	93%

Figura 19 – Comparação entre regiões

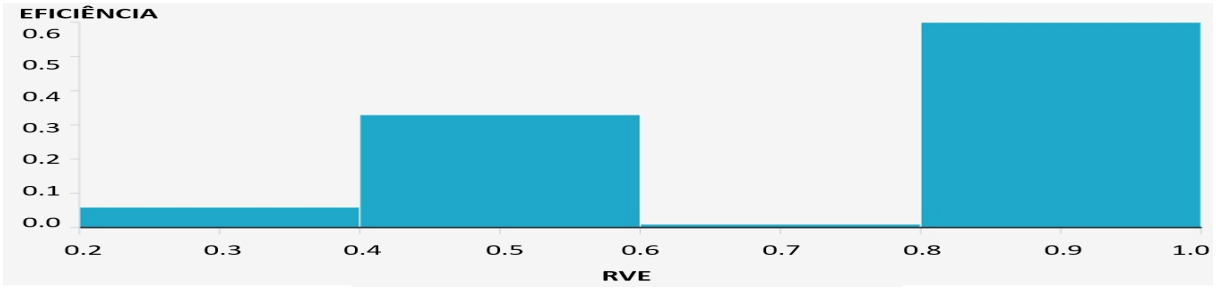
Fonte: Autoria própria

É possível perceber que apenas 11% das DMUs analisadas da região norte são plenamente eficientes. No geral, apenas 5% das empresas analisadas são eficientes, 216 empresas de uma amostra de 4047, que são as empresas cujo valor RCE é igual a 1.

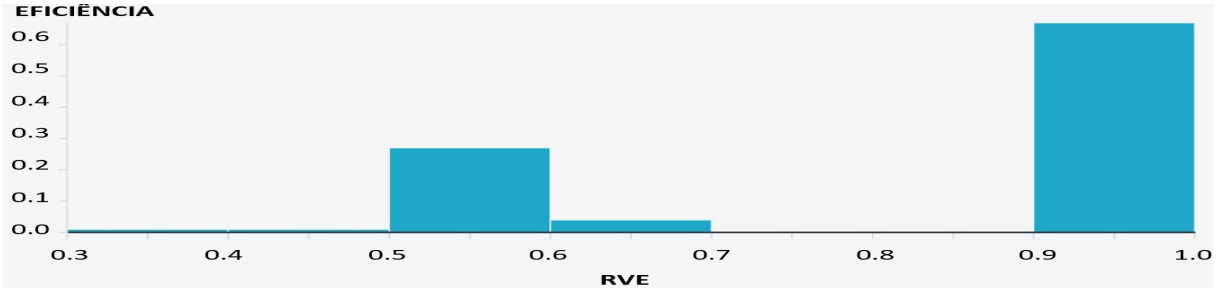
A distribuição dos valores com RVE por região são apresentadas na Figura 20 e a distribuição dos valores com RCE na Figura 21. Ambos os gráficos foram construídos como histogramas com 5 intervalos, o eixo x apresenta a concentração dos valores e o eixo y apresenta o valor encontrado para a eficiência das respectivas empresas.



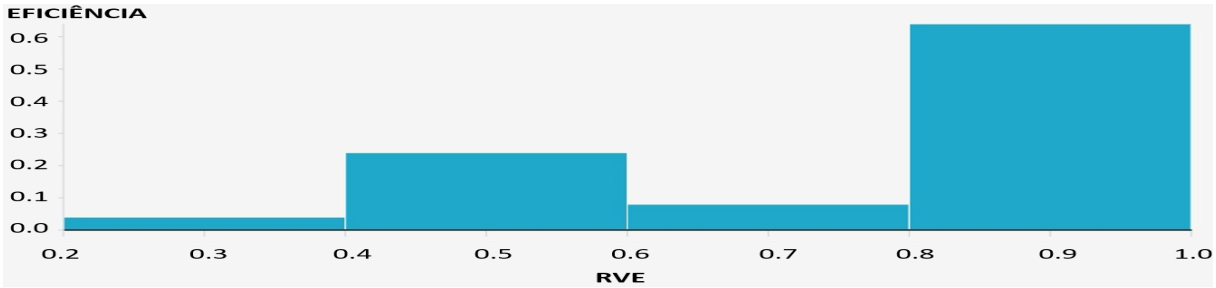
(a) Norte



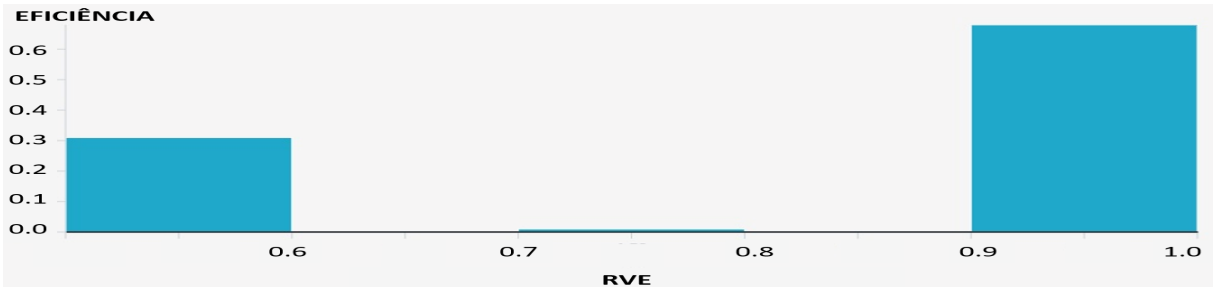
(b) Nordeste



(c) Sul



(d) Sudeste



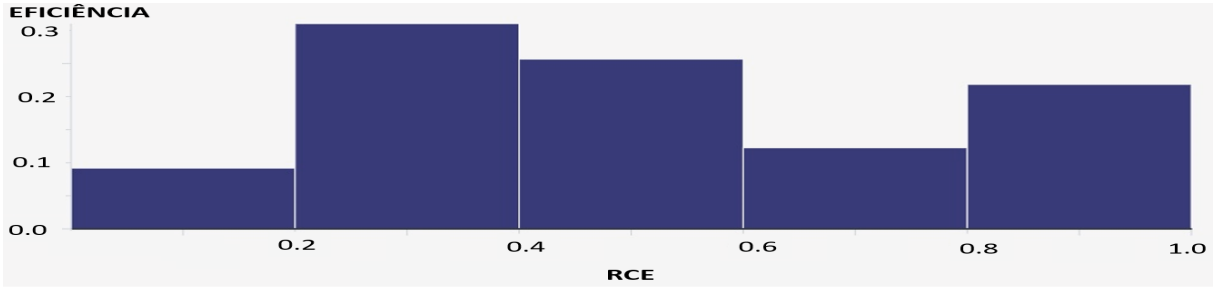
(e) Centro-oeste

Figura 20 – Distribuição de valores de RVE por região.

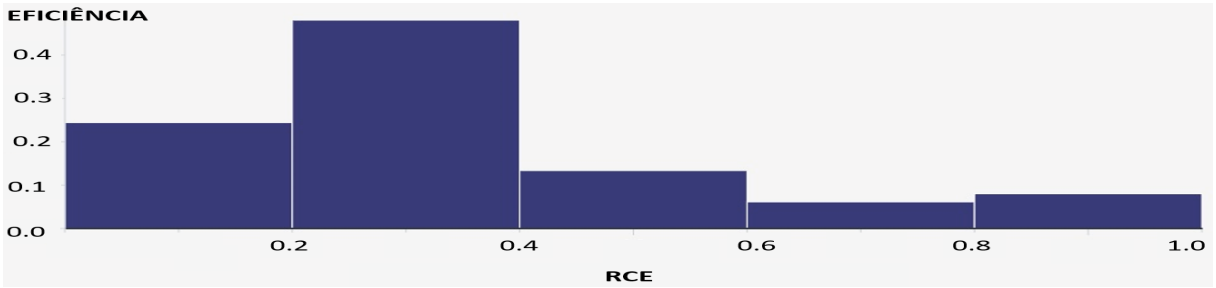
Fonte: Autoria própria

É possível perceber que, embora a região centro-oeste - Figura 20 (e) - apresente 28% das DMUs com RVE<1, portanto não eficientes, a concentração dos valores se distribui em empresas com RVE mais próximos de 0 e empresas com RVE mais próximos de 1. Isto demonstra, portanto, alta variabilidade na operação das empresas da região.

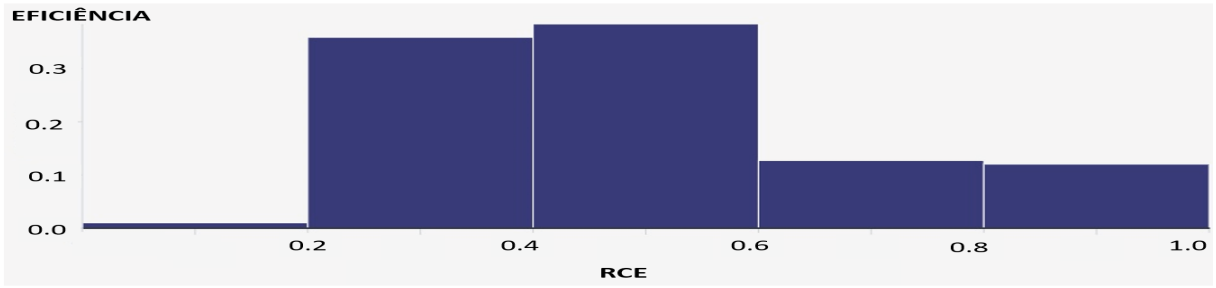
Por outro lado, as empresas do nordeste - Figura 20 (b) -, sul - Figura 20 (c) - e sudeste - Figura 20 (d) -, embora possuam significativa concentração de valores mais próximos de 1, possuem uma distribuição maior de empresas com valores médios de RVE próximos a 0,5. Isto demonstra menor variabilidade na operação das empresas, mas o maior valor de empresas concentradas nesta área do gráfico, promove a diminuição a eficiência média geral da região.



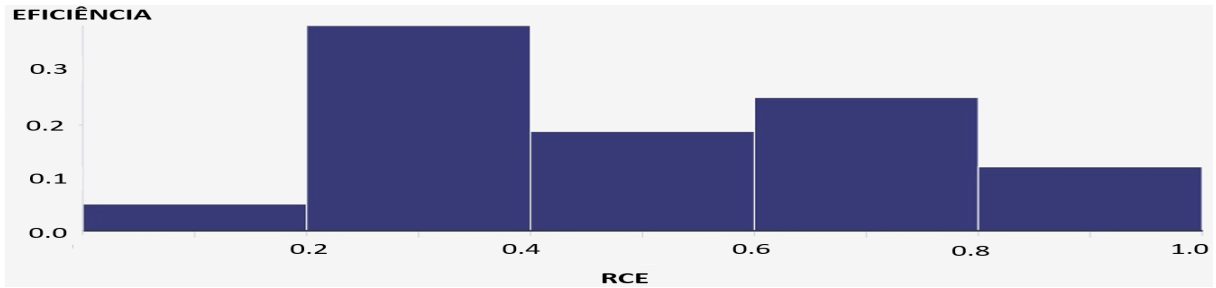
(a) Norte



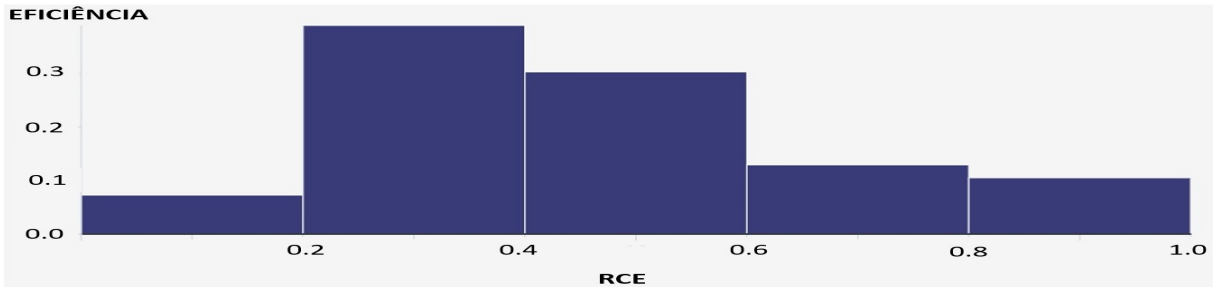
(b) Nordeste



(c) Sul



(d) Sudeste



(e) Centro-oeste

Figura 21 – Distribuição de valores de RCE por região.

Fonte: Autoria própria

Com relação à análise RCE, é possível perceber que as regiões norte - Figura 21 (a) - e sudeste - Figura 21 (d) - apresentaram valores mais próximos de 1 com relação à amostra analisada. A região sul - Figura 21 (c) - teve maior concentração de valores entre 0,4 e 0,6 de eficiência.

A região nordeste - Figura 21 (b) - apresentou uma proporção maior de DMUs com eficiências próximas a zero, comparada às outras regiões. Assim, embora não fosse a que apresentou distribuição concentrada mais próxima de 0 para a análise RVE, obteve altos valores para RVE do que comparados aos resultados com RCE.

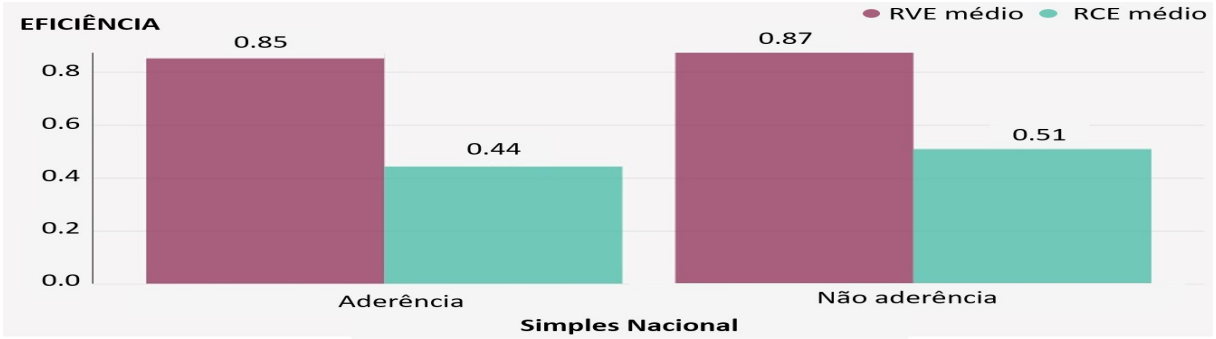
A Figura 22 apresenta um gráfico de eficiência RVE, em roxo, e eficiência RCE, em verde, com relação à aderência ao simples nacional no momento de abertura, à esquerda são apresentadas empresas que aderiram ao programa e à direita as que não aderiram.

(a) Norte

(b) Nordeste

(c) Sul

(d) Sudeste



(e) Centro-oeste

Figura 22 – Eficiência de acordo com aderência ao simples nacional.

Fonte: Autoria própria

O Simples Nacional é um programa da Receita Federal cujo objetivo é facilitar o recolhimento de contribuições das micro e pequenas empresas, propondo uma arrecadação direta de tributos, além de uma possível redução da carga tributária.

Com relação à aderência ao programa, uma das variáveis categóricas utilizadas no trabalho, não foram identificadas alterações significativas nas eficiências calculadas. Desse modo, a aderência não implica diretamente em alterações nos níveis de eficiência das empresas analisadas.

Este resultado deve ser avaliado mais profundamente em trabalhos futuros, uma vez que a implantação do Simples deveria acarretar em melhoria da eficiência das empresas. Uma possível análise a ser realizada é se as empresas que aderiram ao simples, em sua maioria, foram empresas de micro e pequeno porte. Estas podem ser empresas altamente ineficientes que aderiram ao simples até objetivando sua própria sobrevivência, mas permanecendo ainda como altamente ineficientes.

5 Conclusão

Com constantes tentativas do país para contornar a atual crise econômica, combater o desemprego e buscar um crescimento sustentável, tem-se como fundamental o estímulo e investimento em micro e pequenas empresas, tendo em vista a relevância destas para o desenvolvimento nacional. Além de as perspectivas para a indústria de alimentos brasileira para 2022 serem positivas, do ponto de vista das empresas do setor, a ABIA prevê um aumento de 2% nas vendas reais.

A produtividade é uma questão que vem ganhando importância no Brasil devido ao menor crescimento populacional, o que reduz a taxa de crescimento da força de trabalho. O aumento da produtividade do trabalho será essencial para um maior crescimento econômico no futuro (BONELLI; VELOSO; PINHEIRO, 2017).

Com base nos resultados obtidos, bem como nas análises apresentadas no presente trabalho, verifica-se que o objetivo geral de de avaliar a eficiência relativa entre micros e pequenas empresas do setor alimentício da indústria de transformação brasileira, foi alcançado.

Para isso, foi necessário realizar uma revisão de literatura, versando principalmente sobre conceitos e abordagens metodológicas de análise de eficiência e principais modelos. A partir dessa revisão, foi possível sistematizar os dados coletados e aplicar a metodologia para avaliação da eficiência das empresas.

Com relação à metodologia aplicada para avaliação da eficiência de empresas, tem se que, no modelo DEA, não são feitas suposições sobre formas de distribuição, ou seja, utiliza modelos não paramétricos, não exigem uma única forma funcional que determina como as entradas produzem saídas, mas permitem flexibilidade de DMUs individuais em suas configurações de produção.

Além disso, DEA tem a propriedade de invariância de unidades, o que significa que entradas e saídas podem ser medidas em unidades diferentes, facilitando a utilização de múltiplas variáveis para análise.

Tendo o setor alimentício como principal constituinte da receita da indústria de transformação, esse trabalho ratifica a importância de maior pesquisa e desenvolvimento em um dos principais pilares da economia brasileira, a indústria.

Dentre as limitações da pesquisa, bem como sugestões para trabalhos futuros, é notório que no presente estudo foram utilizados apenas dados públicos. Assim, a aplicação de informações de caráter privado, como número de trabalhadores e faturamento, agregariam significativamente para o estudo. Além disso, a utilização de quantidades de empresas

semelhantes por região poderia apresentar pontos de melhoria da eficiência mais claros, sendo possível um entendimento mais granular à nível das variáveis. Além de ser necessário avaliar o impacto da aderência ao Simples Nacional na eficiência das empresas

Além disso, é possível perceber a necessidade de realização de uma análise inter-regional, empresas que aparentaram eficiência neste trabalho, podem ser ineficientes com relação à empresas de outras regiões. Assim, é provável que, caso o número de empresas analisadas da região norte aumentasse, sua eficiência relativa diminuísse, devido ao maior desenvolvimento de outras regiões brasileiras.

Tem-se que todos os objetivos foram alcançados, visto que foi possível: (i) entender micro e pequenas empresas no contexto da indústria brasileira, por meio da revisão da literatura e entendimento dos efeitos que MPEs geram na sociedade e economia; (ii) identificar variáveis para análise e a escolha da modelagem ideal para o problema em questão, a partir das bases disponíveis, foram identificadas as variáveis quantitativas que possuíam maior relevância para a pesquisa; (iii) avaliar a eficiência das empresas estudadas, por meio da aplicação da metodologia DEA e; (iv) apresentar avaliação comparativa de resultados, utilizando o Superset, ferramenta de *business intelligence*.

Referências

ABIA. Relatório Anual da Associação Brasileira da Indústria de Alimentos. **São Paulo**, 2020. Citado na p. 20.

ABIA. Relatório Anual da Associação Brasileira da Indústria de Alimentos. **São Paulo**, 2022. Citado na p. 21.

BANKER, R. D.; CHARNES, A.; COOPER, W. W. Some models for estimating technical and scale inefficiencies in data envelopment analysis. **Management science**, INFORMS, v. 30, n. 9, p. 1078–1092, 1984. Citado na p. 31.

BELFIORE, P.; FÁVERO, L. P. **Pesquisa Operacional para cursos de Engenharia**. Elsevier Brasil, 2013. v. 1. Citado na p. 25.

BONELLI, R.; VELOSO, F.; PINHEIRO, A. C. Anatomia da produtividade no Brasil. **Rio de Janeiro: IBRE/FGV e Elsevier, Rio de Janeiro**, 2017. Citado nas pp. 15, 23, 24, 57.

BOSMA, N.; KELLEY, D. Global entrepreneurship monitor 2018/2019 global report. **Global Entrepreneurship Research Association (GERA)**, 2019. Citado na p. 14.

CEPAL, N. et al. **Avaliação de desempenho do Brasil mais produtivo**. CEPAL, 2018. Citado nas pp. 23, 24.

CHARNES, A.; COOPER, W. W.; RHODES, E. Measuring the efficiency of decision making units. **European journal of operational research**, Elsevier, v. 2, n. 6, p. 429–444, 1978. Citado nas pp. 27, 29, 30.

CNI. **Indústria de A-Z**. 2022. Disponível em: <https://www.portaldaindustria.com.br/industria-de-a-z/>. Citado nas pp. 18, 19.

CNI. Portal da Indústria, 2020. Disponível em: <https://industriabrasileira.portaldaindustria.com.br/grafico/total/producao/#/industria-transformacao>. Citado na p. 19.

COELLI, T. **A primer on efficiency measurement for utilities and transport regulators**. World Bank Publications, 2003. v. 953. Citado na p. 23.

DATASEBRAE. Indicadores de Empregados, 2018. Disponível em: <https://dataseb raeindicadores.sebrae.com.br/resources/sites/data-sebrae/data-sebrae.html#/Empregados>. Citado na p. 15.

EMROUZNEJAD, A.; YANG, G.-l. A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. **Socio-economic planning sciences**, Elsevier, v. 61, p. 4–8, 2018. Citado nas pp. 27, 28.

FERREIRA, L. F. F.; OLIVA, F. L.; SANTOS, S. A. d.; GRISI, C. C. d. H.; LIMA, A. C. Análise quantitativa sobre a mortalidade precoce de micro e pequenas empresas da cidade de São Paulo. **Gestão & Produção**, SciELO Brasil, v. 19, p. 811–823, 2012. Citado na p. 14.

FONSECA, J. J. S. da. **Apostila de metodologia da pesquisa cientıfica**. João José Saraiva da Fonseca, 2002. Citado na p. 34.

GARCIA, J. R. A importância dos instrumentos de apoio à inovação para micro e pequenas empresas para o desenvolvimento econômico. **Revista da FAE**, v. 10, n. 2, 2007. Citado na p. 23.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. Ediitora Atlas SA, 2008. Citado na p. 34.

GOEMANS, M. X. Lecture notes on Linear programming. **Massachusetts Institute of Technology**, 2015. Citado na p. 26.

GOLANY, B.; ROLL, Y. An application procedure for DEA. **Omega**, Elsevier, v. 17, n. 3, p. 237–250, 1989. Citado na p. 38.

GUPTA, R. **Operations research**. Krishna Prakashan Media, 1992. Citado na p. 25.

HILLIER, F. S.; LIEBERMAN, G. J. **Introdução à pesquisa operacional**. McGraw Hill Brasil, 2013. Citado na p. 25.

HOUAISS, A. Dicionário online de português. **São Paulo, out**, 2020. Citado na p. 23.

IBGE. Pesquisa Industrial Anual Empresa, 2020. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/periodicos/1719/pia_2020_v39_n1_empresa_informativo.pdf>. Citado na p. 20.

KOOPMANS, T. C. An analysis of production as an efficient combination of activities. **Activity analysis of production and allocation**, Wiley, 1951. Citado na p. 29.

LEWIS, C. Linear programming: theory and applications. **Whitman College Mathematics Department**, 2008. Citado na p. 26.

MALHOTRA, N. K. **Pesquisa de marketing-: uma orientação aplicada**. Bookman Editora, 2001. Citado na p. 34.

MARCHESE, M.; GIULIANI, E.; SALAZAR-ELENA, J. C.; STONE, I. Enhancing SME productivity. n. 16, 2019. DOI: https://doi.org/https://doi.org/10.1787/825bd8a8-en. Disponível em: <https://www.oecd-ilibrary.org/content/paper/825bd8a8-en>. Citado na p. 15.

MONITOR, G. E. Global entrepreneurship monitor. **Empreendedorismo no Brasil (Relatório Nacional). Curitiba: Instituto Brasileiro de Qualidade e Produtividade, Paraná**, 2011. Citado na p. 14.

NOGUEIRA, M. O.; PEREIRA, L. d. S. As Empresas de Pequeno Porte e a Produtividade Sistêmica da Economia Brasileira: obstáculo ou fator de crescimento? Instituto de Pesquisa Econômica Aplicada (Ipea), 2015. Citado na p. 24.

PGFN. Base de dados da dívida ativa da União, 2022. Disponível em: <https://www.gov.br/pgfn/pt-br/assuntos/divida-ativa-da-uniao/dados-abertos>. Citado na p. 37.

RECEITA FEDERAL DO BRASIL, R. .-. S. E. da. Base de dados do Cadastro Nacional da Pessoa Jurídica (CNPJ), 2022. Disponível em: <https://www.gov.br/receitafederal/pt-br/acesso-a-informacao/dados-abertos>. Citado nas pp. 35–37.

SEBRAE. Sobrevivência das empresas no Brasil, 2016. Disponível em: <https://www.sebrae.com.br/Sebrae/Portal%20Sebrae/Anexos/sobrevivencia-das-empresas-no-brasil-102016.pdf>. Citado na p. 22.

SHERMAN, H. D.; ZHU, J. Analyzing performance in service organizations. **MIT Sloan Management Review**, Massachusetts Institute of Technology, Cambridge, MA, v. 54, n. 4, p. 37, 2013. Citado na p. 28.

SOUZA, M. Uma abordagem bayesiana para o cálculo dos custos operacionais eficientes das distribuidoras de energia elétrica. **Rio de Janeiro: Departamento de Engenharia Elétrica, Pontificia Universidade Católica do Rio de Janeiro**, 2008. Citado nas pp. 31, 32.

SQUEFF, G. C.; NOGUEIRA, M. O. A heterogeneidade estrutural no Brasil de 1950 a 2009. CEPAL, 2013. Citado na p. 14.

TCU. Técnica de Análise Envoltória de Dados em auditoria. Tribunal de Contas da União, 2019. Citado na p. 33.

VIANA, F. L. E. Indústria de alimentos. Banco do Nordeste do Brasil, 2022. Citado na p. 20.

ZHU, J. **Data envelopment analysis: Let the data speak for themselves**. Joe Zhu, 2014. Citado nas pp. 27–29, 33.

Apêndices

APÊNDICE A – Códigos de programação

A.1 Coleta e tratamento dos dados

Código A.1 – Coleta e tratamento de dados em python

```
#!/usr/bin/env python
# coding: utf-8

# In[1]:


import sqlite3
import pandas as pd
import numpy as np
from datetime import datetime,date
import seaborn as sns
import matplotlib.pyplot as plt
from pulp import *
import warnings

warnings.filterwarnings("ignore")


# Conexão com o banco de dados

# In[2]:


def df_query_result(query):
    conn = sqlite3.connect('cnpj.db') # conecta com o banco de
       dados utilizado no estudo
    conn.row_factory = sqlite3.Row # método utilizado para que o o
       cur.fetchall retorne o nome das colunas consultadas
    cur = conn.cursor()
    cur.execute(query) # executa a query
    rows = cur.fetchall() #retorna o resultado da query, mas é
       necessário formatar
    data=[]
    try:
        for row in rows:
            data.append(dict(row))
        return pd.DataFrame(data) # cria um DataFrame com os
           resultados da query
    except:
```

```
36          return rows
37      cur.close_connection() # fecha a conexão com o banco de dados
38
39
40  # # Extração dos Dados
41
42  # ### Query para gerar uma base com os dados brutos que serão
        utilizados na análise
43  #
44  # #### Foram realizadas algumas transformações na própria base:
45  #
46  # - Somente foram selecionadas as empresas que possuem dados sobre
        os sócios disponíveis (t m muitos registros que não estão
        disponíveis);
47  # - Para a situação cadastral, foram consideradas ativas aquelas
        que possuem código = 02. Caso contrário, assumiu-se que
        estariam inativas (usei como base a classificação do SEBRAE pra
        isso, mas podemos mudar);
48  # - Criei um campo para avaliar a opção pelo Imposto Simples
        Nacional no momento da abertura da empresa (é diferente da
        avaliação que está presente na base);
49  # - Filtro das empresas cujo porte é 01 ou 03, ou seja, micro e
        pequenas empresas conforme classificação da Receita Federal;
50  # - Foram selecionadas somente as filiais (para avaliarmos as
        empresas, e não os estabelecimentos);
51  # - Foram selecionados somente os sócios que iniciaram a empresa
        (se algum sócio tiver entrado depois, ele não está nessa base);
52  # - Na condição "estabelecimento.cnae_fiscal like '10%'",
        realizou-se um filtro para selecionar somente as empresas do
        setor alimentício, dentro da indústria de transformação
53
54  # In[ ]:
55
56
57  query = '''
58          select
59              -- estabelecimento
60              estabelecimento.cnpj,
61              estabelecimento.cnpj_basico,
62              case when estabelecimento.situacao_cadastral = '02'
                    then 'ativa'
63              else 'encerrada' end as situacao_cadastral,
64              situacao_cadastral as situacao_cadastral_raw,
65              estabelecimento.data_situacao_cadastral,
66              estabelecimento.data_inicio_atividades,
67              estabelecimento.cnae_fiscal,
68              estabelecimento.cnae_fiscal_secundaria,
69              estabelecimento.uf,
70              estabelecimento.tipo_logradouro,
71              estabelecimento.nome_fantasia,
72
73              -- empresas
```

```
74              empresas.capital_social,
75              empresas.natureza_juridica,
76              empresas.qualificacao_responsavel,
77
78              -- simples
79              case
80                  when
81                      simples.data_opcao_simples <=
                           estabelecimento.data_inicio_atividades
82                      and simples.data_opcao_simples is not null
                           then '1'
83                  else
84                      '0' end as opcao_simples,
85              simples.data_opcao_simples,
86              simples.data_exclusao_simples,
87
88              --socios
89              socios.qualificacao_socio,
90              socios.data_entrada_sociedade,
91              socios.representante_legal,
92              socios.faixa_etaria,
93              socios.nome_socio,
94              socios.cnpj_cpf_socio
95
96          from estabelecimento
97
98          left join empresas on estabelecimento.cnpj_basico =
               empresas.cnpj_basico
99          inner join socios on estabelecimento.cnpj_basico =
               socios.cnpj_basico
100         left join simples on estabelecimento.cnpj_basico =
               simples.cnpj_basico
101         inner join cnae on estabelecimento.cnae_fiscal =
               cnae.codigo
102
103         where
104             estabelecimento.cnae_fiscal like '10%' and
105             porte_empresa in ('01', '03')
106             and cast(data_inicio_atividades as int) >= 20180101
107             and cast(data_inicio_atividades as int) < 20200101
108             and socios.nome_socio is not null
109             and socios.data_entrada_sociedade =
                   data_inicio_atividades
110             and estabelecimento.matriz_filial = '1'
111             and estabelecimento.cnpj_ordem = '0001'
112             and estabelecimento.situacao_cadastral = '02'
113             '''
114
115 df_base = df_query_result(query)
116
117
118 # # Visão Inicial da Base
```

```
119
120 # # Limpeza e Manipulação dos Dados
121
122 # ### Criação de coluna com o número de empresas ativas que os
    sócios possuíam ao abrir a empresa
123 #
124 # Vale ressaltar que será considerada a situação cadastral da
    empresa no momento da abertura da analisada. Ou seja, mesmo que
    a empresa hoje esteja encerrada, ela será considerada ativa se
    estivesse na época.
125
126 # In[ ]:
127
128
129 # Função para acrescentar uma coluna com o número de empresas
    ativas que o sócio possuia na época em que abriu a empresa
130
131 def num_empresas_ativas(df):
132     conn = sqlite3.connect('cnpj.db')
133     cur = conn.cursor()
134     num_empresas_ativas = []
135     for cnpj_cpf_socio, data_inicio_atividades in
136     zip( df['cnpj_cpf_socio'], df['data_inicio_atividades']):
137         query = f'''
138                 select
139                     socios.cnpj_cpf_socio,-- || socios.nome_socio,
140                     count(distinct socios.cnpj)
141                 from
142                     socios
143                 left join
144                     estabelecimento on socios.cnpj =
                        estabelecimento.cnpj
145                 where
146                     (socios.cnpj_cpf_socio = '{cnpj_cpf_socio}'
147
148                                         and estabelecimento.
149                                         data_inicio_atividades <
                                            {data_inicio_atividades}
150                     and estabelecimento.matriz_filial = '1')
151                     and (estabelecimento.situacao_cadastral = '02'
152                     or estabelecimento.situacao_cadastral != '02'
153                     and estabelecimento.data_situacao_cadastral <=
                        {data_inicio_atividades})
154                 group by 1
155         '''
156         cur.execute(query)
157         row = cur.fetchall()
158         if len(row)==0:
159             num_empresas_ativas.append(0)
160         else:
161             num_empresas_ativas.append(row[0][-1])
162
```

```
163     cur.close()
164     conn.close()
165 
166     df.loc[:,'num_empresas_ativas'] = num_empresas_ativas
167 
168     return df
169 
170 
171 # In[ ]:
172 
173 
174 df_base = num_empresas_ativas(df_base)
175 
176 
177 # ### Criação de coluna com o total de empresas ativas por CNPJ
        (somando todos os sócios)
178 
179 # In[ ]:
180 
181 
182 df_ativa = df_base.groupby('cnpj')['num_empresas_ativas'].
183 sum().reset_index().rename(columns={'num_empresas_ativas':
184                                     'num_empresas_ativas_tot'})
185 df_base = df_base.merge(df_ativa, on='cnpj')
186 
187 
188 # ### Criação de coluna com o número de empresas encerradas que os
        sócios possuíam ao abrir a empresa
189 #
190 
191 # In[ ]:
192 
193 
194 # Função para acrescentar uma coluna com o número de empresas
        encerradas que o sócio possui
195 
196 def num_empresas_encerradas(df):
197     conn = sqlite3.connect('cnpj.db')
198     cur = conn.cursor()
199     num_empresas_encerradas = []
200     for nome_socio, cnpj_cpf_socio,
201     data_inicio_atividades in zip(df['nome_socio'],
202 
203                                   df['cnpj_cpf_socio'],
                                          df['data_inicio_atividades']):
204         query = f'''
205                 select
206                     socios.cnpj_cpf_socio || socios.nome_socio,
207                     count(distinct socios.cnpj)
208                 from
209                     socios
210                 left join
```

```
                    estabelecimento on socios.cnpj =
                       estabelecimento.cnpj
                where
                    socios.cnpj_cpf_socio = '{cnpj_cpf_socio}'
                    and socios.nome_socio = '{nome_socio}'
                    and estabelecimento.
                    data_inicio_atividades <
                       {data_inicio_atividades}
                    and estabelecimento.matriz_filial = '1'
                    and estabelecimento.situacao_cadastral != '02'
                group by 1
        '''
        cur.execute(query)
        row = cur.fetchall()
        if len(row)==0:
            num_empresas_encerradas.append(0)
        else:
            num_empresas_encerradas.append(row[0][-1])

    cur.close()
    conn.close()

    df.loc[:,'num_empresas_encerradas'] = num_empresas_encerradas

    return df


# In[ ]:


df_base = num_empresas_encerradas(df_base)


# ### Criação de coluna com o total de empresas encerradas por CNPJ

# In[ ]:


df_encerrada = df_base.groupby('cnpj')['num_empresas_encerradas'].
sum().reset_index().rename(columns={'num_empresas_encerradas':
                                    'num_empresas_encerradas_tot'})
df_base = df_base.merge(df_encerrada, on='cnpj')


# ### Avaliação se abriu uma filial junto com a matriz

# In[ ]:


# Função para acrescentar uma coluna com o número de empresas
   encerradas que o sócio possui

```

```
def num_filiais(df):
    conn = sqlite3.connect('cnpj.db')
    cur = conn.cursor()
    num_filiais = []
    for cnpj_basico, data_inicio_atividades in
    zip(df['cnpj_basico'], df['data_inicio_atividades']):
        query = f'''
                select
                    count(distinct estabelecimento.cnpj)
                from
                    estabelecimento
                where
                    estabelecimento.cnpj_basico =
                    '{cnpj_basico}'
                    and estabelecimento.
                    data_inicio_atividades
                    = {data_inicio_atividades}
                    and estabelecimento.matriz_filial = '2'
        '''
        cur.execute(query)
        row = cur.fetchall()
        num_filiais.append(row[0][0])

    cur.close()

    conn.close()

    df.loc[:,'num_filiais'] = num_filiais

    return df


# ### Criação de coluna com o número de filiais abertas em
   conjunto com a matriz

# In[ ]:


df_base = num_filiais(df_base)

df_base.loc[:,'cnpj_basico'] = df_base['cnpj'].apply(lambda x:
   x[:8])


# ### Cópia do DataFrame, para não perder os dados iniciais

# In[ ]:


df = df_base.copy()


```

```
# <!-- ### Formatar as colunas data_situacao_cadastral e data_inicio_atividades como data (estão como texto) -->

# ### Contagem do número de CNAEs fiscais secundários por empresa

# In[ ]:


df.loc[:,'num_cnae_sec'] = np.where(df['cnae_fiscal_secundaria']=='',
                        0,
                        df['cnae_fiscal_secundaria'].apply(lambda x: len(x.split(','))))


# ### Cálculo do número de sócios por empresa

# In[ ]:


num_socios = df.groupby('cnpj')[['qualificacao_socio']].count().rename(columns={'qualificacao_socio':'num_socios'}).reset_index()

df = df.merge(num_socios, how='left')


# ### Selecionar somente um registro de cada empresa, criando colunas com a faixa etária do sócio mais velho (faixa_etaria_o) e do mais novo (faixa_etaria_y)

# In[ ]:


df_y = df.sort_values(['cnpj','faixa_etaria'], ascending=True).drop_duplicates('cnpj',keep='first').reset_index(drop=True)
df_y.rename(columns={'faixa_etaria':'faixa_etaria_y'},inplace=True)


# In[ ]:


df_o = df.sort_values(['cnpj','faixa_etaria'], ascending=False).drop_duplicates('cnpj',keep='first').reset_index(drop=True)
df_o.rename(columns={'faixa_etaria':'faixa_etaria_o'},inplace=True)


# In[ ]:


df = df_o.merge(df_y[['cnpj','faixa_etaria_y']], on='cnpj')


```

```
357 # ### Criar uma nova coluna com os estados agrupados por região
358
359 # In[ ]:
360
361
362 pd.read_html('https://www.estadosecapitaisdobrasil.com/')[0].head()
363
364
365 # In[ ]:
366
367
368 # Relação entre estados e regiões
369
370 tabela_estados = pd.read_html(
371     'https://www.estadosecapitaisdobrasil.com/')
372 [0][['Sigla','Região','Estado']]
373
374 df = pd.merge(df,tabela_estados, how='left', left_on='uf',
       right_on='Sigla')
375
376
377 # In[ ]:
378
379
380 df.rename(columns={'Região':'regiao',
       'Sigla':'sigla','Estado':'estado'}, inplace=True)
381
382 df.loc[:,'regiao'] = df['regiao'].apply(lambda x: str(x).lower())
383
384
385 # ### Remoção de outliers - Capital Social
386 #
387 # Ao analisar a variável capital social, observou-se que há alguns
       números muito discrepantes, que atrapalham na análise. Dessa
       forma, estes serão desconsiderados do estudo
388
389 # In[ ]:
390
391
392 # Serão considerados 99% dos registros -- ainda necessita de
       validação
393
394 df = df[df['capital_social'] <=
       df['capital_social'].quantile(.99)].reset_index(drop=True)
395
396
397 # # Dívida ativa
398 #
399 # Conexão entre banco e tabelas geradas trimestralmente pela PGFN
400
401 # In[ ]:
402
```

```
403
404 #read from csv
405 tabela_divida = pd.read_csv("divida_ativa_trim_unificada.csv")
406
407 #select the columns chosen for the study
408 colunas_divida=['CPF_CNPJ','VALOR_CONSOLIDADO']
409
410 df_divida=tabela_divida[colunas_divida]
411
412 #lower columns labels
413 df_divida.columns = [x.lower() for x in df_divida.columns]
414
415 #group by cnpj
416
417 df_divida=df_divida.groupby('cpf_cnpj')
418 ['valor_consolidado'].sum().reset_index()
419
420
421 #delete special carachter
422 df_divida['cpf_cnpj']=df_divida['cpf_cnpj'].str.replace("[./-]","")
423
424
425 # In[ ]:
426
427
428 #conectando a tabela DA com o banco de dados
429 df_modelo = pd.merge(df,df_divida, how='left', left_on='cnpj',
       right_on='cpf_cnpj')
430
431
432 # In[ ]:
433
434
435
436 #seleção das colunas
437 colunas_da =
438 ['cnpj','situacao_cadastral', 'capital_social',
439 'opcao_simples','data_inicio_atividades',
440  'num_socios',
       'num_empresas_ativas_tot','num_empresas_encerradas_tot',
441 'num_filiais','num_cnae_sec','valor_consolidado','regiao', 'uf']
442
443 df_final = df_modelo[colunas_da]
444
445 df_final=df_final.fillna(0)
446
447
448 # In[ ]:
449
450
451 df_final.to_csv('df_dea_v1.csv')
```

A.2 Desenvolvimento de conjunto de dados para pydea e superset

Código A.2 – Seleção das variáveis e geração de arquivos para pydea e superset em python

```
#!/usr/bin/env python
# coding: utf-8

# # Seleção de variáveis

# In[1]:


import sqlite3
import pandas as pd
import numpy as np
from datetime import datetime,date
import seaborn as sns
import matplotlib.pyplot as plt
from pulp import *
import warnings

warnings.filterwarnings("ignore")


# In[2]:


df = pd.read_csv('df_dea_v1.csv')


# In[3]:


corr=['capital_social','num_socios', 'num_empresas_ativas_tot',
   'num_empresas_encerradas_tot',
       'num_filiais', 'valor_consolidado','num_cnae_sec']
df_corr=df[corr]

fig, ax = plt.subplots()
fig.set_size_inches(4,3)


# Correlation
corr = df_corr.corr()
# Heatmap
sns.heatmap(corr, cmap="Blues")
plt.savefig('correlacao.png', bbox_inches='tight')
print(corr)


```

```
46 # In[4]:
47
48
49 estudo=['cnpj', 'situacao_cadastral', 'capital_social',
       'opcao_simples',
50         'num_socios', 'num_empresas_ativas_tot',
             'num_empresas_encerradas_tot',
51         'num_filiais', 'valor_consolidado', 'regiao',
             'num_cnae_sec']
52 df=df[estudo]
53
54
55 # In[5]:
56
57
58 get_ipython().run_cell_magic('time', '', "\n#Gerar arquivos para
       rodar no pydea\nregioes=df.regiao.unique()\n\nfor x in
       regioes:\n    df_regiao=df['regiao']==x\n
       df_final=df[df_regiao].to_csv('dea_{}.csv'.format(x))")
59
60
61 # In[6]:
62
63
64 #Gerar arquivo para o superset
65 df_superset = pd.read_excel('pydea_consolidado_v1.xlsx')
66 df_superset=df_superset.merge(df, how='left', on='cnpj')
67 df_superset.to_csv('superset.csv'.format(x))
68 df_superset.head()
69
70
71 # In[7]:
72
73
74 #analise_regioes
75 colunas_regiao=['opcao_simples','cnpj']
76 df_regiao=df_superset[colunas_regiao]
77 df_regiao=df_superset.groupby('regiao')
78 print (df_regiao)
79
80
81 # In[ ]:
```