

# Sobrevivência empresarial em Pernambuco com base no PIB e nos dados abertos da RFB: uma análise em dois períodos sociopolítico econômicos da história do Brasil

Elisângela P.M. Santana<sup>1</sup>, Edivaldo R.S. Júnior<sup>2</sup>, Cleide Lins de Almeida<sup>3</sup>

<sup>1</sup>Banco do Nordeste do Brasil (BNB) - Recife, PE, Brazil

<sup>2</sup>Seção Especial de Tecnologia e Segurança da Informação - Recife, PE, Brazil

<sup>3</sup>Inspetoria do Porto de Suape - Recife, PE, Brazil

elisangelapms@bnb.gov.br, edivaldo.santana@rfb.gov.br,  
cleide.almeida@rfb.gov.br

**Abstract.** *In this work, the survival of companies in the state of Pernambuco will be analyzed, Pernambuco is the main economic hub in Brazilian northeast region, two different sociopolitical periods in Brazil's modern history will be considered and correlated with the Gross Domestic Product (GDP). Open data from the Brazilian Federal Revenue Service (RFB) will be used to calculate the survival rate of these companies and define their profile. This research methodology from Sebrae and OCDE is used and adapted for our proposal. The results achieved indicated a higher incidence of the number of companies founded in times of greater economic instability, low survival rate of companies and retail business as the main economic activity in the state.*

**Keywords:** *Business survival; Survival rate; Mortality rate; Gross Domestic Product; Corporate legal nature; Economic activity*

**Resumo.** *Neste trabalho será analisado a sobrevivência das empresas no estado de Pernambuco, principal polo da economia do nordeste brasileiro, em dois períodos sociopolítico econômicos da história recente do Brasil, correlacionando com o Produto Interno Bruto (PIB). Será utilizado os dados abertos da Receita Federal do Brasil (RFB) para calcular a taxa de sobrevivência das empresas e caracterizar o seu perfil, a partir da metodologia utilizada pelo Sebrae e OCDE, adaptada ao objetivo proposto. Os resultados alcançados indicaram uma maior incidência do número de empresas abertas em momentos de maior instabilidade econômica, baixa taxa de sobrevivência das empresas e comércio varejista como principal atividade explorada no estado.*

**Palavras-chave:** *Sobrevivência empresarial; Taxa de sobrevivência; Taxa de mortalidade; Produto Interno Bruto; Natureza jurídica empresarial; Atividade Econômica*

## 1. Introdução

A procura de estatísticas sobre demografia empresarial cresceu e desenvolveu-se consideravelmente nos últimos anos. Dados sobre nascimentos e mortes de empresas, sua expectativa de vida e o importante papel que desempenham no crescimento econômico e na produtividade, bem como as informações que eles fornecem para lidar com questões

sociodemográficas, são cada vez mais solicitadas pelos formuladores de políticas (OCDE, 2006).

Em síntese, nos últimos 60 anos, diversos fatos definitivamente marcaram essa instabilidade no cenário político-econômico do país tais como: o golpe militar de 1964; o "milagre econômico" da metade da década de 60 até o final da década de 70; a chamada "década perdida" nos anos 80; a redemocratização, com a promulgação da constituição cidadã de 1988; e, finalmente em 1994, o Plano real que combateu a hiperinflação galopante (GIAMBIAGI et. al., 2011).

O Produto Interno Bruto (PIB) é uma das diversas variáveis que podem interferir na sobrevivência das empresas. Além do PIB, pode-se citar outras variáveis macroeconômicas tais como: taxa de inflação, taxa de juros, desemprego, Índice de Desenvolvimento Humano (IDH), etc. Conforme Mankiw (2015) o PIB geralmente é considerado o melhor indicador para refletir o desempenho de uma economia.

A análise de sobrevivência é uma área da estatística que teve origem na área médica e trata de estimar o tempo de vida até a ocorrência de um evento de interesse, que pode ser a morte, a cura ou a recidiva de doença (COLOSIMO, 2006). Todavia este termo tem sido usado para explicar outros fenômenos sociais. O estudo em lide está assentado na análise da sobrevivência de empresas dentro de um cenário mais extenso do que o tradicionalmente é tratado, conforme será visto com relação a metodologia adotada.

O estudo de sobrevivência das empresas interessa a diversos *stakeholders* a exemplo do Serviço Brasileiro de Apoio às Micro e Pequenas Empresas<sup>1</sup> (Sebrae) que tem o foco no empreendedorismo, a Receita Federal do Brasil<sup>2</sup> (RFB) que também tem interesse na saúde das unidades produtivas para fins de arrecadação e instituições de financiamento de crédito para empresas, tal como o Banco do Nordeste do Brasil<sup>3</sup> (BNB) entre outros. O Sebrae tem publicado estudos sobre sobrevivência empresarial desde 2011, nos quais utiliza uma metodologia que emprega o processamento e análise das bases de dados disponibilizadas pela RFB. Esta metodologia basicamente utiliza um espaço temporal de 2 anos para o cálculo da taxa de sobrevivência (SEBRAE, 2016).

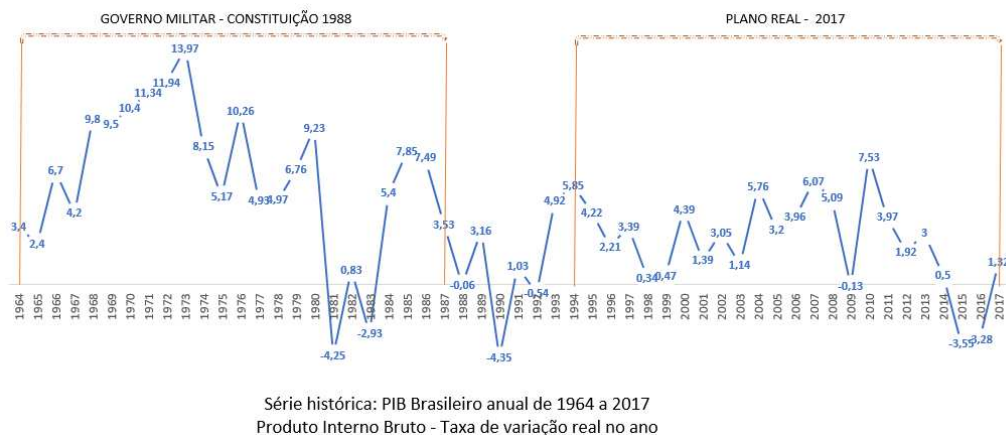
Diante deste contexto, delimitou-se no espaço e tempo o estudo da sobrevivência das empresas a população representada pelas empresas estabelecidas dentro do estado de Pernambuco; em dois períodos distintos, sendo o primeiro, de 1964 a 1987; o segundo, de 1994 a 2017. O primeiro período abrange o início do governo militar e se estende até antes da promulgação da constituição de 1988. O segundo período começa em 1994, cujo marco foi o Plano Real, e finda em 2017, totalizando 23 anos em cada período. De acordo com a Figura 01 tem-se uma síntese destes momentos. Analisando a série histórica do PIB, observa-se que no primeiro período o PIB acumulado ficou próximo a 150%. No segundo intervalo o PIB ficou em torno de 60%.

---

<sup>1</sup> <https://sebrae.com.br/sites/PortalSebrae>

<sup>2</sup> <https://www.gov.br/receitafederal/pt-br>

<sup>3</sup> <https://www.bnb.gov.br/>



**Figura 01. Série histórica do PIB. Fonte: autores.**

O microambiente formado por forças como: fornecedores, distribuidores, clientes e outros *stakeholders* podem ser, de certa forma, controlados pelas empresas. Por outro lado, os componentes do microambiente organizacional são forças que podem trazer transtornos e provocar mudanças de magnitude tamanha que as empresas podem sucumbir se não se adaptarem às mudanças ou se não as antevêm e tomarem medidas corretivas, (CHIAVENATO, 2022). O autor deixa claro que a capacidade de tomar boas decisões é a chave para um desempenho gerencial bem-sucedido (MCGUIGAN, 2022), e para tomar boas decisões é necessário informação e desta extrair conhecimento para produzir resultados positivos (CHIAVENATO, 2022).

As forças macroeconômicas compõem um ambiente hostil e o seu monitoramento contínuo é essencial para um bom desempenho do negócio (GOMES, 2022). Neste íterim, o PIB representa o maior nível de agregação econômica de um país, a atividade econômica global. Contudo, componentes específicos do PIB, a exemplo do Produto Interno Bruto dos municípios (PIBM), podem fornecer informações mais interessantes (MCGUIGAN, 2022).

Como dizia Chico Science, o artista criador do movimento Manguebeat, "modernizar o passado é uma evolução musical" (SCIENCE, 1994), ou seja, olhar o passado sobre uma nova perspectiva não deixa de ser uma inovação. Sendo assim, a principal contribuição deste trabalho é agregar ao universo da Ciência de Dados um estudo histórico sobre o tema sobrevivência de empresas. Não é tarefa simples mirar a taxa de sobrevivência comparando dois contextos históricos tão ímpares, ainda mais sob a ótica de uma variável macroeconômica que possui uma grande complexidade como é o caso do PIB.

O objetivo geral deste artigo consiste em fazer uma análise dos dados nos dois períodos em tela, com a finalidade de identificar qual a possível relação entre a nossa variável alvo, que é a taxa de sobrevivência das empresas, e o PIB.

Os objetivos específicos são:

- Caracterizar o perfil das empresas sobreviventes nos dois períodos;
- Analisar o crescimento/decrescimento do número de empresas ano a ano;

- Discutir se o PIB teria alguma influência no comportamento da taxa de sobrevivência das empresas estabelecidas no estado de Pernambuco nos dois períodos em análise.

O presente trabalho é uma pesquisa do tipo exploratória descritiva, com abordagem quantitativa, usando técnicas da Ciência de Dados.

Este trabalho está organizado em seis seções. A primeira seção contém a introdução sobre o tema proposto. Na segunda seção tem-se o referencial teórico, subdividido em três subseções: a primeira e segunda, na qual são descritos os conceitos de PIB e taxa de sobrevivência; a terceira descreve o cenário sociopolítico econômico dos períodos analisados. A terceira seção traz a metodologia utilizada na pesquisa, a qual subdivide-se em quatro subseções, onde é discutida a compreensão do negócio, a coleta dos dados, preparação dos dados e processamento/tratamento/limpeza dos dados. A quarta seção descreve a análise de dados estudados. Na quinta seção são apresentados os resultados e as discussões. E, por fim, a sexta seção com as considerações finais e trabalhos futuros.

## **2. Referencial Teórico**

Nesta seção serão abordados os principais conceitos contidos no tema do presente estudo, com o objetivo de apresentá-los de forma pormenorizada.

### **2.1 O Produto Interno Bruto e a taxa de sobrevivência**

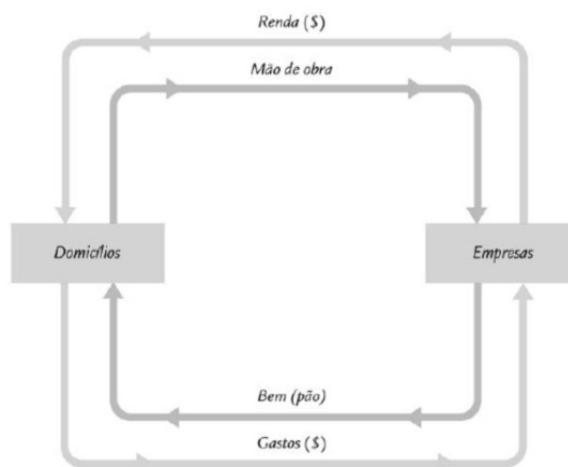
#### **2.1.1 O Produto Interno Bruto**

O produto interno bruto (PIB) de um país é um indicador econômico que mede a produção de bens e serviços finais em determinado período, que geralmente pode ser anual, semestral ou trimestral. O PIB juntamente com a taxa de inflação e a taxa de desemprego constituem os melhores indicadores de desempenho de uma economia (MANKIW, 2015).

A atividade econômica das empresas, que são as unidades produtivas do sistema econômico, está totalmente imbricada com o comportamento das famílias. Na opinião de Mankiw (2015) uma economia com grande produção de bens e serviços pode suprir melhor as demandas das famílias, das empresas e do governo.

Uma maneira de explicar o PIB é, conforme a Figura 02 observar o fluxo de moeda corrente na economia. Supondo uma economia bem simples em que as famílias entregam mão de obra para empresas, e as empresas produzem os bens que serão vendidos às famílias.

A renda obtida com a venda dos bens é utilizada para remunerar com o salário a mão de obra dos funcionários, e outra parte da renda é o lucro que remunera os proprietários das empresas. Os funcionários e os proprietários das empresas fazem parte das famílias.



**Figura 02. Ciclo da Economia. Fonte: Mankiw (2015)**

Sendo assim, tem-se um fluxo de moeda partindo das famílias para as empresas e outro em sentido inverso. O PIB pode ser calculado de duas maneiras: considerando a renda gerada pelos salários e lucros que partem das empresas para as famílias, o que se pode observar na parte superior da Figura 02; ou considerando o total do gasto para produzir os bens na sua metade inferior.

### 2.1.2 A taxa de sobrevivência

Em síntese, para fins de contagem, a sobrevivência de uma empresa “A” nascida no ano XX é verificada com relação a sua situação ao final do ano XX+1. Para tal são verificados os seguintes critérios: a empresa encontra-se ativa em termos de volume de negócios e/ou emprego; a empresa alterou o seu quadro societário (foi adquirida por outro proprietário) (OECD, 2007).

Os casos de fusão, cisão ou incorporação não são considerados sobrevivência, pois estes fatos envolvem outras empresas estabelecidas antes do ano XX+1, e, portanto, a empresa não é considerada como tendo sobrevivido (OECD, 2007). A Organização para a Cooperação e Desenvolvimento Econômico<sup>4</sup> (OCDE) utiliza períodos de 1 até 5 anos para fazer a verificação da situação de sobrevivência. Essa sistematização é seguida pelos países membros desta organização.

Para avaliar as chances de sobrevivência das empresas, um período de cinco anos tem sido normalmente tomado como referencial de comparação na literatura (BRIX e GROTS, 2006). O autor deixa claro que as empresas têm menos chance de sobreviver nos primeiros cinco anos, conforme a maioria dos estudos empíricos sobre sobrevivência empresarial.

De acordo com Brix e Grots (2006, p.9):

A pressão competitiva entre as empresas jovens, que, devido à sua idade, geralmente ainda são pequenas e têm pouca experiência de mercado, é

<sup>4</sup> <https://www.oecd.org/>

especialmente forte em grandes grupos de novas empresas. Parece razoável supor que novas empresas lutam por fatias de mercado principalmente com outras empresas jovens e menos com empresas já estabelecidas no mercado.

A contagem de empresas sobreviventes é feita da seguinte forma, considerando a informação da Figura 03 (OECD,2007):

A – A empresa nasceu ou já existia no ano XX e não se encontra ativa no ano XX+1;

B – A empresa nasceu/existia no ano XX e encontra-se ativa em XX+1;

C – A empresa nasceu/existia em XX, estava inativa em XX+1 e após XX+1 foi reativada;

D – A empresa nasceu/existia em XX, estava ativa em XX+1 e após XX+1 foi reativada.

Por conseguinte, para fins de contagem, as empresas B e D contariam como ativas para a o decurso de XX até XX+1. As empresas A e C contam como inativas.

Production of survival data for xx+1

Outcome	Births or survivals in year xx	Active enterprises in year xx+1	Enterprises taken over by a new enterprise
A	✓		
B	✓	✓	
C	✓		✓
D	✓	✓	✓

**Figura 03. Geração de dados de sobrevivência. Fonte: (OECD, 2007)**

## 2.2 O cenário sociopolítico econômico nos períodos analisados

### 2.2.1 De 1964 a 1987

Em 1963, o Brasil apresentava uma inflação bastante elevada de 83,2% a.a. O diagnóstico feito pela equipe econômica foi de em excesso de demanda, e a solução proposta e executada foram políticas econômicas ortodoxas, no âmbito de política fiscal contracionista, política monetária com restrição de crédito e redução do salário real (GREMAUD; VASCONCELOS; JUNIOR, 2009).

Mesmo com as medidas contracionistas tomadas, o Produto Interno Bruto (PIB) brasileiro teve um sensível crescimento no período: 3,4%, 2,4%, 6,7% e 4,2% respectivamente entre 1964 a 1967. As reformas saneadoras feitas no período e os financiamentos externos feitos de forma vultosa, com um investimento maciço no setor secundário, fez o famoso “milagre brasileiro” (GIAMBIGI, et al, 2011).

Com o foco no crescimento econômico, os números atingidos foram surpreendentes: 9,5%, 10,4%, 11,3%, 12,1% e 14,0%. Assim, entre os anos 1968 e 1974 a economia brasileira deu um grande salto de tamanho, com grandes investimentos públicos infra estruturantes, nova dinâmica da construção civil e exportações.

Com a crise econômica que se abateu sobre o mundo, a partir de 1973, a crise do petróleo, as estratégias de crescimento econômico do Brasil perduraram. Esta etapa de “crescimento forçado” da economia pode até, de forma aparente, ter sido uma boa decisão

estratégica de crescimento de nações, mas teve um preço futuro bastante indigesto e com consequências bastante decepcionantes. No entanto, nesse período o Brasil continuou a crescer: 5,21%, 9,87%, 4,6%, 4,8% e 7,2%, respectivamente entre os anos de 1975 a 1979 (GIAMBIAGI; SCHWARTAMAN, 2014).

Os anos 80, último capítulo do governo militar, com João Baptista de Figueiredo, trouxe grande tensão e desalento para a econômica brasileira. O Brasil não resistiu à segunda grande crise do petróleo (1979) e o ortodoxismo tomou conta do cenário econômico. O ajustamento voluntário realizado trouxe uma queda expressiva no PIB e marcou o fim da expansão econômica (MENDES, 2014).

Com o colapso do papel dinâmico do estado na economia e os ajustes macroeconômicos realizados no início dos anos 80, o Brasil ao invés de expandir seu PIB, entra em uma fase de estagnação econômica e índices de inflação nunca vistos.

O endividamento do setor público e o pagamento dos juros da dívida externa comprometeu o crescimento econômico. A rolagem da dívida teve que ser realizada cada vez mais no curto prazo pois a credibilidade de que o governo brasileiro pudesse honrar seus compromissos reduziu drasticamente, desestabilizando internamente a economia (MENDES, 2014).

Em 1980 o Brasil cresce 9,2%, vindo em 1981 a drástica queda do PIB para -4,5% e em 1982 um pífio crescimento de 0,83%. Em 1983 o PIB volta a encolher (-2,3%), acompanhado de crescimento de 5,4% e 7,85% respectivamente entre 1984 e 1985.

Em 1986 o Brasil vivenciou, no governo Sarney, um dos mais ousados planos econômicos da história, o Plano Cruzado. Neste ano o Brasil cresceu 7,5% acompanhado de um crescimento de 3,5% em 1987, ano do Plano Bresser.

### **2.2.2 De 1988 a 2017**

Após uma série de planos econômicos iniciados em 1986 com o Plano Cruzado, no final de 1993, começa a ser implementado um dos mais emblemáticos e engenhosos planos econômicos no Brasil, o Plano Real (GREMAUD et. al., 2009).

Apesar da estabilização dos preços, a oscilação do PIB foi bem inferior à do período analisado anteriormente (1964 a 1987). A inflação começou a ceder e, em raros momentos, superou os dois dígitos.

Em 1994 o PIB cresceu 5,33%, bem acima dos próximos anos, que tiveram reduções gradativas, tendo o ano de 1995 reduzido para 4,42% e, logo em seguida, 2,21% em 1996.

Neste novo cenário econômico e político as variáveis macroeconômicas, quando analisadas em seu conjunto, como inflação, produção, juros e desemprego, não tiveram o comportamento desejado. Entre 1997 e 2001 o Brasil continuou com um baixo crescimento relativo 3,4%, 0,34%, 0,47%, 0,41% e 1,39%, respectivamente (MENDES, 2014).

O controle da inflação com a nova política monetária, cambial e fiscal tiveram reflexo no crescimento econômico, até então adotado. O fim do grande crescimento do período militar, baseado em endividamento externo e inversões de capital realizadas pelo Estado, deixou o Brasil com uma dívida externa de grande magnitude, resultando em desconfiança nos mercados internos e externos.

Entre 2001 e 2003, os últimos anos do Governo de Fernando Henrique Cardoso, o ‘Pai do Real’ não conseguiu a popularidade suficiente para eleger seu sucessor. Lula assumiu a presidência até 2011, no entanto, o PIB continuou oscilante no período, tendo conseguido atingir crescimento bem mais expressivos. Entre 2003 e 2010 quando cresceu, respectivamente 1,14%, 5,76%, 3,2%, 3,96%, 6,07%, 5,04%, -0,13% e 7,53%.

Mesmo com oscilações, tanto a inflação como o PIB, tiveram desempenhos satisfatórios no governo Lula, no entanto, no governo Dilma Rousseff, entre 2011 e 2016 o PIB teve um desempenho sofrível de 3,97%, 1,92%, 3,0%, 0,50%, -3,77% e -3,6%.

Em 2017, após o golpe de estado no governo Temer, o PIB teve um fraco desempenho de 1,3%. O período de estabilização econômica, teve como foco principal o combate à inflação, foi um período marcado pela estabilidade dos preços, porém foi acompanhado de um PIB oscilante e extremamente inferior ao ritmo de crescimento que o país estava acostumado nos anos 60 e 70.

### **3. Metodologia**

É fundamental para o processo de análise de dados obedecer a uma metodologia com passos bem definidos e sistemáticos a fim de extrair conhecimento necessário para resolução do problema que permita confiabilidade no conhecimento produzido. (PROVOST; FAWCETT, 2016)

Para a finalidade deste estudo, foi implementada a metodologia conforme as etapas abaixo:

- Compreensão do negócio
- Compreensão dos dados
- Preparação dos dados
- Agregação dos dados
- Análise e visualização dos dados
- Interpretação do resultado

#### **3.1 Compreensão do negócio**



Para realizar o experimento utilizou-se dados de domínio público coletados no site da RFB<sup>5</sup> e Instituto Brasileiro de Geografia e Estatística<sup>6</sup> (IBGE), sobre os quais foram aplicadas a presente metodologia a fim de atingir os objetivos propostos.

A importância de mensurar a taxa de sobrevivência para as empresas usando período de até cinco anos de atividade está relacionada com o objetivo do estudo. Se o propósito é o de direcionar as empresas jovens a melhor se posicionarem no mercado, adotar um parâmetro de tempo de atividade de até cinco anos é crucial para tal fim (SEBRAE, 2016).

No Brasil, o nascimento de uma empresa ocorre quando a empresa é registrada por uma Junta Comercial<sup>7</sup> em uma das unidades da federação. Logo após o ato do registro, ele é comunicado eletronicamente a RFB, ocasião em que a empresa será incluída no banco de dados Cadastro Nacional da Pessoa Jurídica<sup>8</sup> (CNPJ). Neste banco de dados existe um atributo chamado Situação Cadastral, o qual informa se a empresa se encontra ativa ou não. Se a empresa entregou as suas declarações a RFB para o ano corrente, é um indício de que ela está operando. Depois de dois anos sem prestar as declarações, o atributo Situação Cadastral é alterado para outro valor que informará o tipo de inatividade (nula, suspensão, inapta ou baixada).

Como foi visto no referencial teórico, normalmente é usado um período de até 5 anos para verificação da sobrevivência da empresa. A taxa de sobrevivência é o percentual de empresas ativas com relação ao total de empresas existentes no banco de dados CNPJ, dentro do período de XX até XX+N anos.

Na metodologia adotada para este estudo, não foi levado em conta um período de até 5 anos para fins de contagem, conforme recomendado em (SEBRAE, 2016) e (OECD, 2007), pois o objetivo em questão não é o de analisar a sobrevivência para fins de acompanhamento e consultoria dos pequenos negócios.

Para atingir o objetivo, que é o de contar as empresas sobreviventes em um período extenso, considerou-se o nascimento no ano XX e a verificação da situação de atividade/inatividade em um momento atual. Como consequência desse método, pode ocorrer de uma empresa ter sua situação alterada de ativa para inativa, e de inativa para ativa por diversas vezes, durante o período avaliado. Considerar-se-á a situação final da empresa ao final do período em questão.

### **3.2 Compreensão dos dados**

A etapa de compreensão do negócio envolveu o alinhamento das expectativas e formulação do problema a ser resolvido. Nesta etapa de compressão dos dados foram definidas quais informações, dentre as disponíveis no site da RFB, seriam necessárias para realizar a análise descritiva e exploratória dos dados.

A fim de cumprir o requisito de confiabilidade dos dados foram utilizadas informações oficiais divulgadas pelo governo. A Tabela 01 mostra estas informações

---

<sup>5</sup> <https://www.gov.br/receitafederal/pt-br/assuntos/orientacao-tributaria/cadastros/consultas/dados-publicos-cnpj>

<sup>6</sup> <https://www.ibge.gov.br/>

<sup>7</sup> <https://portal.jucepe.pe.gov.br/>

<sup>8</sup> <https://www.gov.br/receitafederal/pt-br/servicos/cadastro/cnpj>

extraídas do site da RFB, os dados abertos de CNPJ das EMPRESAS, ESTABELECIMENTO, SIMPLES NACIONAL e tabelas auxiliares.

**Tabela 01 – Informações extraídos do site RFB. Fontes: autores.**

Dados Abertos Sítio RFB Data Base 13-03-2022	Extração em 01-05-2022
Dados Abertos CNPJ EMPRESA 01 a 10	K3241.K03200Y0.D20312.EMPRESV a Y9
Dados Abertos CNPJ ESTABELECIMENTO 01 a 10	K3241.K03200Y0.D20312.ESTABELE a Y9
Informações sobre o Simples Nacional/MEI	F.K03200\$W.SIMPLES.CSV.D20312
Tabela de atributo CNAE	F.K03200\$Z.D20312.CNAECSV
Tabela de motivo da situação cadastral	F.K03200\$Z.D20312.MOTICSV
Tabela de atributo Município	F.K03200\$Z.D20312.MUNICCSV
Tabela de atributo Natureza Jurídica	F.K03200\$Z.D20312.NATJUCSV

Para análise do PIB foi realizada a extração dos dados de PIB do Brasil do site IBGE dos anos 2002 a 2019.

- PIB Brasil – base de dados 2002-2009
- PIB Brasil – base de dados 2010-2019

Para processamento dos dados utilizamos computador AMD *Ryzen 5 3350G* com *Radeon veja Graphics 3.60 GHz*, 16GB de RAM, *Windows 10 Pro 64* e para manipulação dos dados utilizamos *Excel 365*, *Jupyter Notebook Versão 6.4.5* rodando *Python Versão 3.9.7*, *Google Colab* e *VS Studio Code Versão 1.67.2*, *Windows Power Shell*

Foi realizado um estudo dos dados RFB a partir do dicionário de dados e leiaute dos arquivos disponibilizado no site e no manual do usuário CNPJ, documento interno de acesso dos colaboradores da instituição, que tem como objetivo explicar a estrutura de dados disponibilizada. A Tabela 02 mostra as colunas elencadas como dados importantes de entrada para o estudo.

**Tabela 02 – Estrutura dos Datasets da RFB. Fonte: autores.**

ESTABELECIMENTOS				EMPRESAS	
FEATURES DO DATASET	ESCOLHIDAS	FEATURES DO DATASET	ESCOLHIDAS	FEATURES DO DATASET	ESCOLHIDAS
CNPJ_BASICO	X	NUMERO		CNPJ BASICO	X
CNPJ_ORDEM		COMPL		RAZÃO SOCIAL	X
CNPJ_DV		BAIRRO		NATJU	X
IDE_MAT_FIL	X	CEP		QUALIF	
NOME_FANT	X	UF	X	CAP_SOCIAL	X
SIT_CADASTRAL	X	MUNICIPIO	X	PORTE	X
DATA_SIT_CADAS	X	DDD_1		ENTE_FED	
MOT_SIT_CADAS	X	TELEFONE_1			

NOME_DA_CID		DDD_2	
PAIS		TELEFONE_2	
DATA_INI_ATV	X	DDD_DO_FAX	
CNAE_FISCAL	X	FAX	
CNAE_FISCAL_SEC		CORREIO_ELET	
TIPO_DE_LOG		SITUACAO_ESPECIAL	

PIB POR MUNICIPIO	
FEATURES	ESCOLHIDAS
ANO	X
CÓDIGO DO MUNICÍPIO	X
PIB	X
PIB PER CAPITA	X

### 3.3 Preparação dos Dados

A terceira etapa de preparação de dados envolveu maior número de horas trabalhadas, tanto em função da quantidade de dados, que totalizou 29 arquivos com tamanho total de 16,4GB, quanto da exigência de hardware para processá-los. Nesta etapa elaborou-se 3 scripts *Jupyter Notebook*<sup>9</sup>, com intuito de viabilizar sua análise e exploração.

*Script 01-TratamentoInicial.ipynb*: A principal tarefa deste *Jupyter Notebook* é de coleta dos dados. Em função do tamanho dos arquivos, e exigência de efetuar *upload* dos dados para processamento no *Google Colab*<sup>10</sup>, optou-se por executar este script localmente inicialmente utilizando *Jupyter Notebook*, todavia, não foi possível finalizar a tarefa dado travamento recorrente do sistema em meio ao processamento. Após diversas tentativas recorreu-se ao *VStudio* rodando *Python 3.9.7*. para finalização da tarefa.

Foi utilizado um ambiente virtual isolado do *Python*<sup>11</sup> e bibliotecas *Matplotlib*<sup>12</sup>, *Numpy*<sup>13</sup> e *Pandas*<sup>14</sup>, por serem suficientes para o tratamento dos dados previstos.

Para cada arquivo baixado no site RFB foi realizado o carregamento no *VStudio*<sup>15</sup> a fim de obter uma base consolidada por conjunto de dados (empresas e estabelecimento) que permitisse mineração sem precisar reprocessá-las, utilizado o *Pandas*.

Após importação das tabelas empresas (empresas0 a empresas9), agrupou-se os dados para gerar um único arquivo empresasTratada por meio do comando *append*. Foram validadas a integridade do arquivo empresasTratada comparando a soma dos registros de cada arquivo com a arquivo final. Este processo foi realizado para os 10 arquivos de dados abertos CNPJ empresa e os 10 arquivos de dados abertos CNPJ estabelecimento, gerando o *dataset* estabTratada, e totalizando a importação de 49.151.491 registros de pessoas jurídicas e 51.860.331 registros de estabelecimentos, visto tratar-se da contabilização de matriz (49.151.491) e filial (27.088.40). A Figura 04 detalha a estrutura dos *datasets* empresaTratada e estabTratada.

<sup>9</sup>

<https://jupyter.org/>

<sup>10</sup> <https://colab.research.google.com/>

<sup>11</sup> <https://www.python.org/>

<sup>12</sup> <https://matplotlib.org/>

<sup>13</sup> <https://numpy.org/>

<sup>14</sup> <https://pandas.pydata.org/>

<sup>15</sup> <https://visualstudio.microsoft.com/pt-br/>

<pre> estabTratada.info() ✓ 0.7s  &lt;class 'pandas.core.frame.DataFrame'&gt; Int64Index: 51860331 entries, 0 to 4753433 Data columns (total 12 columns): #   Column      Dtype ---  --- 0   CNPJ_BASICO  int64 1   CNPJ_ORDEM   int64 2   CNPJ_DV      int64 3   IDE_MAT_FIL  int64 4   SIT_CADASTRAL int64 5   DATA_SIT_CADAS int64 6   MOT_SIT_CADAS int64 7   DATA_INI_ATV int64 8   CNAE_FISCAL  int64 9   CNAE_FISCAL_SEC object 10  UF           object 11  MUNICIPIO    int64 dtypes: int64(10), object(2) memory usage: 5.0+ GB </pre>	<pre> empresaTratada.info() ✓ 0.1s  &lt;class 'pandas.core.frame.DataFrame'&gt; RangeIndex: 49151491 entries, 0 to 49151490 Data columns (total 5 columns): #   Column      Dtype ---  --- 0   CNPJ BASICO  int64 1   NATJU        int64 2   QUALIF       int64 3   CAP_SOCIAL   object 4   PORTE        float64 dtypes: float64(1), int64(3), object(1) memory usage: 1.8+ GB </pre>
---	---

**Figura 04. Estrutura dos *datasets* empresaTratada e estabTratada. Fonte: autores.**

Além dos dados das empresas e estabelecimento, foram importadas as tabelas complementares de CNAE, natureza jurídica, motivo de situação cadastral e município.

Nesta etapa foi possível efetuar as seguintes tarefas:

- Unificar os registros de empresa e estabelecimento em um único arquivo
- Identificar existência de valores ausentes/nulos
- Conhecer a estrutura do *dataset* para mineração dos dados.

*Script 02-Agregar.ipynb*: A tarefa do *Jupyter notebook* agregar é unificar as tabelas empresas, estabelecimento e simples em um único arquivo. Seu processamento objetivou efetuar as subtarefas abaixo:

- Importar *dataset* empresas, estabelecimento e simples previamente tratados a fim de produzir um único arquivo empresasBrasil com 51.860.364 registros e 16 colunas conforme Figuras 05 e 06;
- Unificar os nomes das colunas utilizadas para cruzamento dos dados;
- Verificar valores faltantes dos *datasets*;
- Formatar dados das colunas dos *datasets* de acordo com nossa necessidade.

<pre>#VERIFICAR VALORES FALTANTES print(pd.isnull(empresasBrasil).sum())</pre> <p>✓ 20.8s</p> <table> <tr><td>CNPJ_BASICO</td><td>0</td></tr> <tr><td>NATJU</td><td>7840</td></tr> <tr><td>CAP_SOCIAL</td><td>7840</td></tr> <tr><td>PORTE</td><td>71027</td></tr> <tr><td>CNPJ_ORDEM</td><td>19</td></tr> <tr><td>CNPJ_DV</td><td>19</td></tr> <tr><td>IDE_MAT_FIL</td><td>19</td></tr> <tr><td>SIT_CADASTRAL</td><td>19</td></tr> <tr><td>DATA_SIT_CADAS</td><td>19</td></tr> <tr><td>MOT_SIT_CADAS</td><td>19</td></tr> <tr><td>DATA_INI_ATV</td><td>19</td></tr> <tr><td>CNAE_FISCAL</td><td>19</td></tr> <tr><td>UF</td><td>19</td></tr> <tr><td>MUNICIPIO</td><td>19</td></tr> <tr><td>OP_SIMPLES</td><td>19900293</td></tr> <tr><td>MEI</td><td>19900293</td></tr> </table> <p>dtype: int64</p>	CNPJ_BASICO	0	NATJU	7840	CAP_SOCIAL	7840	PORTE	71027	CNPJ_ORDEM	19	CNPJ_DV	19	IDE_MAT_FIL	19	SIT_CADASTRAL	19	DATA_SIT_CADAS	19	MOT_SIT_CADAS	19	DATA_INI_ATV	19	CNAE_FISCAL	19	UF	19	MUNICIPIO	19	OP_SIMPLES	19900293	MEI	19900293	<pre>empresasBrasil.info()</pre> <p>✓ 0.7s</p> <pre>&lt;class 'pandas.core.frame.DataFrame'&gt; Int64Index: 51860364 entries, 0 to 51860363 Data columns (total 16 columns): #   Column          Dtype ---  - 0   CNPJ_BASICO      int64 1   NATJU            float64 2   CAP_SOCIAL       float64 3   PORTE            float64 4   CNPJ_ORDEM       float64 5   CNPJ_DV          float64 6   IDE_MAT_FIL      float64 7   SIT_CADASTRAL    float64 8   DATA_SIT_CADAS  float64 9   MOT_SIT_CADAS    float64 10  DATA_INI_ATV    float64 11  CNAE_FISCAL       float64 12  UF               object 13  MUNICIPIO        float64 14  OP_SIMPLES       object 15  MEI              object dtypes: float64(12), int64(1), object(3) memory usage: 6.6+ GB</pre>
CNPJ_BASICO	0																																
NATJU	7840																																
CAP_SOCIAL	7840																																
PORTE	71027																																
CNPJ_ORDEM	19																																
CNPJ_DV	19																																
IDE_MAT_FIL	19																																
SIT_CADASTRAL	19																																
DATA_SIT_CADAS	19																																
MOT_SIT_CADAS	19																																
DATA_INI_ATV	19																																
CNAE_FISCAL	19																																
UF	19																																
MUNICIPIO	19																																
OP_SIMPLES	19900293																																
MEI	19900293																																

**Figura 05. Estrutura do *dataset* empresasBrasil. Fonte: autores.**

```
# VISUALIZAÇÃO DO DATASET EMPRESAS BRASIL
empresasBrasil.head(2)
```

CNPJ_BASICO	NATJU	CAP_SOCIAL	PORTE	CNPJ_ORDEM	CNPJ_DV	IDE_MAT_FIL	SIT_CADASTRAL	DATA_SIT_CADAS	MOT_SIT_CADAS	DATA_INI_ATV	CNAE_FISCAL	UF	MUNICIPIO	OP_SIMPLES	MEI
41273603	2062.0	10000.0	1.0	1.0	44.0	1.0	2.0	20210318.0	0.0	20210318.0	4789099.0	PR	7535.0	S	N
41273604	2062.0	10000.0	1.0	1.0	99.0	1.0	2.0	20210318.0	0.0	20210318.0	8592901.0	RS	8625.0	S	N

**Figura 06. Amostra do *dataset* Empresas Brasil. Fonte: autores.**

*Script* 03-TratamentoFinal.ipynb: A tarefa do Jupyter notebook 03 é o tratamento do *dataset* empresasBrasil para extração e visualização dos dados de Pernambuco. Iniciou-se com carregamento das bibliotecas básicas e configuração das variáveis de ambiente para melhor visualização e análise dos dados.

De acordo com o escopo do trabalho, foram extraídos os dados das empresas localizadas em Pernambuco, criando o *dataset* empresasPe mostrado na Figura 07, totalizando 1.582.849 registros de pessoas jurídicas, equivalente 3,05% das empresas do Brasil e distribuídas em 185 municípios

```
empresasPE.info()
✓ 0.3s

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1582849 entries, 16 to 51860349
Data columns (total 16 columns):
#   Column          Non-Null Count  Dtype
---  -
0   CNPJ_BASICO      1582849 non-null int64
1   NATJU            1582569 non-null float64
2   CAP_SOCIAL       1582569 non-null float64
3   PORTE            1582568 non-null float64
4   CNPJ_ORDEM       1582849 non-null float64
5   CNPJ_DV          1582849 non-null float64
6   IDE_MAT_FIL      1582849 non-null float64
7   SIT_CADASTRAL    1582849 non-null float64
8   DATA_SIT_CADAS  1582849 non-null float64
9   MOT_SIT_CADAS    1582849 non-null float64
10  DATA_INI_ATV    1582849 non-null float64
11  CNAE_FISCAL      1582849 non-null float64
12  UF               1582849 non-null object
13  MUNICIPIO        1582849 non-null float64
14  OP_SIMPLES       966044 non-null object
15  MEI              966044 non-null object
dtypes: float64(12), int64(1), object(3)
memory usage: 205.3+ MB
```

**Figura 07. Detalhes do *dataset* EmpresasPe. Fonte: autores.**

Optou-se por excluir as empresas identificadas como Microempreendedor Individual (MEI) no estado de PE, pois sua criação sucede o período de análise proposto para o trabalho. Considerando o objetivo de realizar uma análise influenciada por fatores econômicos, também foram excluídos os CNAES ligados a organizações políticas, atividades não informadas e empresas cujas natureza jurídica não estão associadas a instituições com fins lucrativos.

Para alcançar o objetivo analítico foi extraído da data de início da atividade, e data da situação fiscal as informações de ano usando o comando *slice* do *Python*. Identificou-se que alguns registros de data de ano da situação cadastral, coluna adicionada ao *dataset* empresasPe encontrava-se com valor zero (2110 registros). Neste caso optou-se por excluí-los visto tratar-se de uma quantidade irrelevante para o número de registro total analisados 984.085.

Foi criado o *dataset* df\_FinalPe, no qual excluiu-se as colunas irrelevantes para o estudo, mantendo-se apenas as colunas necessárias para realizar os *insights* de nossa análise, conforme Figura 08.

```
df_FinalPe.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 984085 entries, 0 to 987745
Data columns (total 16 columns):
 #   Column          Non-Null Count  Dtype  
---  -
 0   NOME             984085 non-null object  
 1   MUNICIPIO        984085 non-null int64  
 2   UF               984085 non-null object  
 3   DESC_NATJU       984085 non-null object  
 4   NATJU            984085 non-null int32  
 5   PORTE            984085 non-null object  
 6   IDE_MAT_FIL      984085 non-null int32  
 7   SIT_CADASTRAL    984085 non-null object  
 8   MOT_SIT_CADAS    984085 non-null int32  
 9   ANO_SIT_CADAS    984085 non-null int32  
10   ANO_INI_ATV      984085 non-null int32  
11   CNAE_FISCAL      984085 non-null int32  
12   N_MUNIC          984085 non-null int32  
13   DESC_CNAE        984085 non-null object  
14   LATITUDE         984085 non-null object  
15   LONGITUDE        984085 non-null object  
dtypes: int32(7), int64(1), object(8)
memory usage: 101.4+ MB
```

**Figura 08. Detalhes do dataset df\_FinalPe. Fonte: autores.**

Por fim, a coluna SIT\_CADASTRAL do dataset df\_FinalPE foi transformada em um indicador que informa se a empresa está ativa ou não. Seguiu-se a nomenclatura utilizada para RFB conforme tabela 03, todavia adaptando para o artigo.

**Tabela 03 – Situação Cadastral. Fonte: autores.**

SITUAÇÃO CADASTRAL		
CODIGO	RFB	ARTIGO
1	NULA	INATIVA
2	ATIVA	ATIVA
3	SUSPENSA	INATIVA
4	INAPTA	INATIVA
8	BAIXADA	INATIVA

O processamento do *script* 03 objetivou efetuar as subtarefas abaixo, para deixar o dataset df\_FinalPe pronto para a próxima etapa:

- Extrair dados do estado de Pernambuco;
- Identificação de valores ausentes/nulos;
- Converter tipos de dados;
- Formatar atributos do datasets de acordo com a necessidade;
- Criar coluna ANO\_SIT\_CADAS e ANO\_INI\_ATIV

- Excluir registros e colunas desnecessários;
- Transformar a coluna SIT\_CADASTRAL.

### 3.4 Agregação dos dados

Esta etapa foi realizada usando o software Microsoft Excel<sup>16</sup>, com o qual foi criada a planilha PibBrasil, mostrada na Figura 09. Para gerar esta planilha, os seguintes passos foram executados:

- Foram inseridos nas colunas ANO e TAXA\_PIB\_BRASIL da planilha os valores do ano e taxa de variação do PIB, extraídos da série histórica do site do IBGE;
- Foi importado para a planilha o *dataset* df\_FinalPe, a fim de computar os valores de TOT\_EMP e TOT\_ATIV por ano;
- Computado os valores de TOT\_EMP através da fórmula de somatório das empresas que iniciaram atividades no ano XX;
- Computado os valores de TOT\_ATIV através da fórmula de somatório das empresas que iniciaram atividades no ano XX e sobreviveram até o ano XX+N, onde XX+N é o ano atual;
- Calculada a Taxa\_Sobrev através da fórmula  $TOT\_ATIV / TOT\_EMP$ .

ANO	Taxa_PIB_Brasil(%)	TOT_EMP	TOT_ATIV	Taxa_Sobrev(%)
1967	4,2	1329	110	8,28
1968	9,8	1083	173	15,97

Figura 09. Detalhes da planilha PibBrasil. Fonte: autores.

A planilha PibBrasil gerada nesta etapa será importada no *script* 04–Análise, a fim de produzir a entrada para a análise de dados proposta neste estudo.

## 4. Análise dos dados

Para realizar a análise dos dados foi criado o último *script* denominado 04–Análise com o objetivo de fazer análise exploratória dos dados. O foco desta análise consistiu em caracterizar o perfil das empresas sobreviventes nos anos de 1964 a 1987 e de 1994 a 2017, doravante denominados período 01 e período 02, analisar o crescimento/decrescimento do número de empresas ano a ano e discutir o comportamento da taxa de sobrevivência em relação ao PIB nos períodos supracitados.

A fim de diminuir a complexidade da análise dos dados optou-se por aplicar filtro no *dataset* df\_FinalPE, a fim de separá-lo por período de estudo, período 01 e período 02, criando *datasets* df\_Per01 e df\_Per02 respectivamente. Com o objetivo de analisar o crescimento/decrescimento do número de empresas ano a ano de cada período de estudo, plotou-se o histograma da coluna “ANO\_INI\_ATV” mostrado na Figura 10, correspondente ao ano do início de atividade da empresa, para verificar sua distribuição ano a ano, o que permitiu extrair informações importantes.

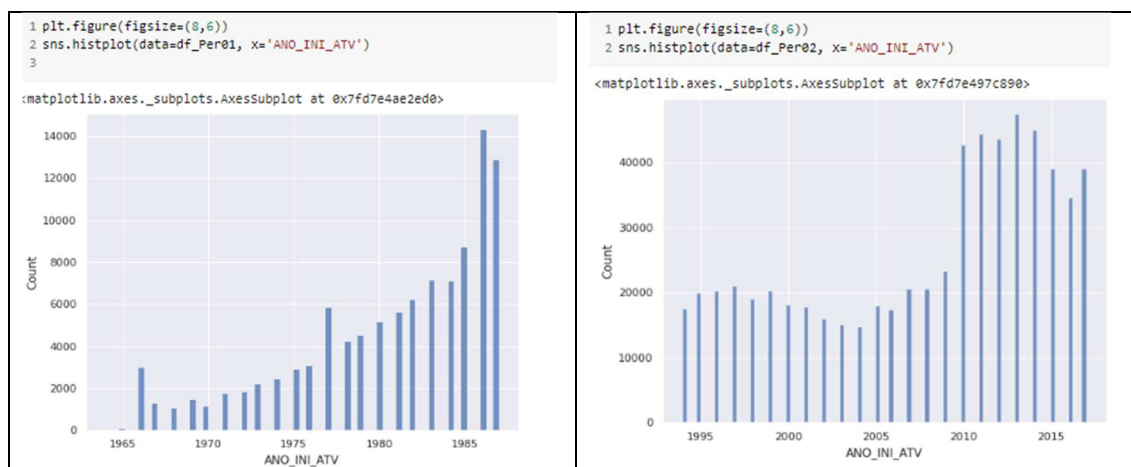
No período 01 observou-se um crescimento significativo da atividade empreendedora ano após ano, notadamente a partir da década de 80 tendo o ano de 1986

<sup>16</sup> <https://support.microsoft.com/pt-br/excel>



a maior volume de empresas abertas do período analisado, 14.337, contra 8.747 do ano anterior. A média de empresas abertas neste período foi de 4357 anualmente, todavia, 75% destas foram abertas após 1982. A quantidade total de empresas abertas em PE no período 01 foi de 104.583. O desvio padrão foi de 3.714 e variância de 13800908.505 o que significa que os dados no conjunto fornecido estão muito mais espalhados do que o valor médio.

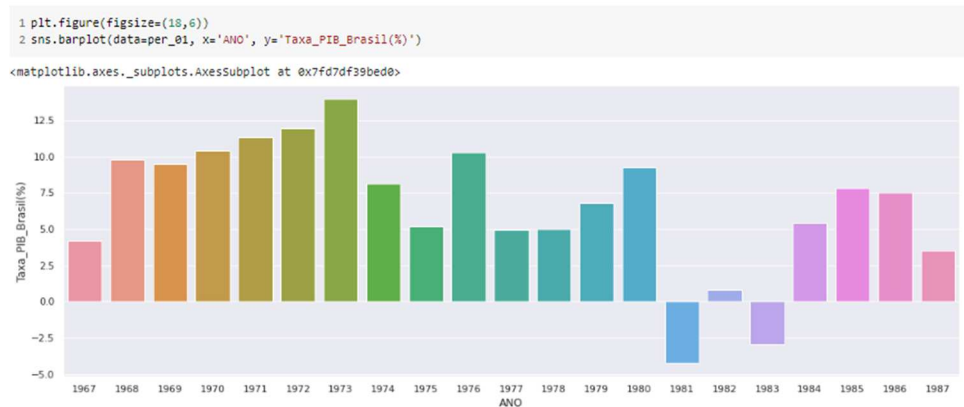
No período 02 a quantidade de empresas abertas em PE foi de 633.623. A série apresenta um salto a partir do ano 2010 com a abertura de 42.661 empresas contra 23.270 do ano anterior 2009, tendência esta que se mantém até o fim do período estudado. Neste período a média de empresas abertas anualmente foi de 26.400, com 75% destas abertas a partir do ano de 2013, o que demonstra uma concentração significativa no fim da série estudada. O desvio padrão no período 02 foi de 11.575 com alta variância de 134002282.3894.



**Figura 10. Evolução do nascimento das empresas. Fonte: autores.**

Para análise da taxa de variação do PIB, importamos o *dataset* PibBrasil, separando-o por período, sendo o período 01 representado pelo *dataset* per\_01 e período 02 pelo *dataset* per\_02.

Quando analisado o PIB do período 01, conforme Figura 11, a partir dos anos 1980 observou-se uma instabilidade econômica, caracterizado pela queda acentuada do PIB, que teve seu menor valor em 1981 com queda de -4,25, contra 0,83, e -2,93 dos anos 1982, 1983 respectivamente.



**Figura 11. PIB do Brasil do Período 01(de 1964 a 1987) ano a ano. Fonte: autores.**

Em média o PIB do período 01 foi de 6,5971%, com desvio padrão de 4,63 como mostrado nas estatísticas do *dataset*, Figura 12. Neste período o PIB apresentou crescimento acumulado de 138,54% e uma taxa de sobrevivência acumulada de 225,37%. A variância deste indicador retrata pouca variação em relação à média, 21,47.

```
1 per_01.describe()
```

	ANO	Taxa_PIB_Brasil(%)	TOT_EMP	TOT_ATIV	Taxa_Sobrev(%)
count	21.0000	21.0000	21.0000	21.0000	21.0000
mean	1977.0000	6.5971	4831.6190	461.1905	10.7319
std	6.2048	4.6338	3700.1422	283.3935	2.7691
min	1967.0000	-4.2500	1083.0000	110.0000	6.4906
25%	1972.0000	4.9300	1840.0000	216.0000	9.2541
50%	1977.0000	7.4900	4283.0000	381.0000	10.6075
75%	1982.0000	9.8000	6243.0000	665.0000	11.8614
max	1987.0000	13.9700	14337.0000	1065.0000	18.3362

**Figura 12. Estatísticas do PIB do Brasil do Período 01(de 1964 a 1987) ano a ano. Fonte: autores.**

Quanto ao período 02 o PIB apresentou-se tímido, com média de 2,57 %. Três anos apresentaram períodos de decrescimento de -0,13 em 2009, -3,55 em 2015 e -3,28 em 2016 que influenciaram de forma significativa o resultado do período, vide Figura 13.

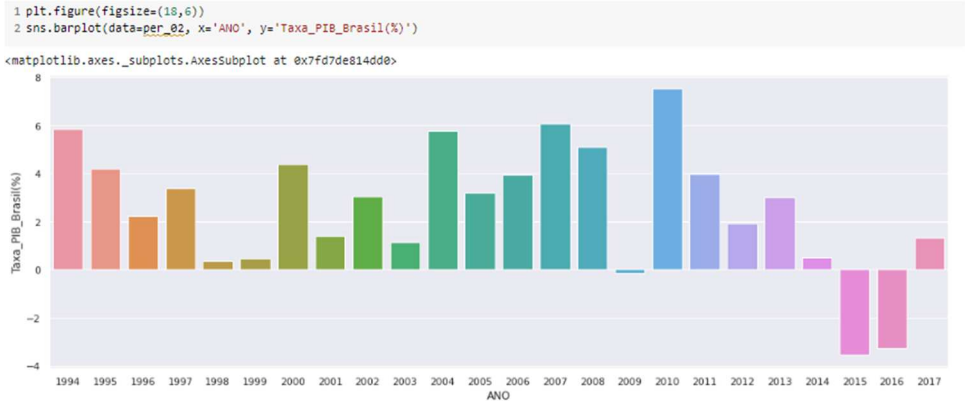


Figura 13. PIB do Brasil do Período 02(de 1994 a 2017) ano a ano. Fonte: autores.

As estatísticas do *dataset* per\_02 apresentado na Figura 14 mostraram um desvio padrão da Taxa PIB Brasil de 2,74, variação máxima de 7,53% e mínima de -3,55. Neste período a variância foi de 61,80 maior que o período 01. Neste período o PIB acumulado foi de 30,04%.

```
1 per_02.describe()
```

	ANO	Taxa_PIB_Brasil(%)	TOT_EMP	TOT_ATIV	Taxa_Sobrev(%)
count	24.0000	24.0000	24.0000	24.0000	24.0000
mean	2005.5000	2.5754	26400.9583	5893.1250	21.8768
std	7.0711	2.7484	11575.9355	3124.7303	5.4810
min	1994.0000	-3.5500	14754.0000	2161.0000	13.3785
25%	1999.7500	0.9800	17848.5000	3507.2500	19.3683
50%	2005.5000	3.0250	20330.0000	4501.0000	20.9928
75%	2011.2500	4.2625	38937.5000	8784.5000	25.6146
max	2017.0000	7.5300	47387.0000	13204.0000	33.9269

Figura 14. Estatísticas do PIB do Brasil do Período 02 (de 1994 a 2017) ano a ano. Fonte: autores.

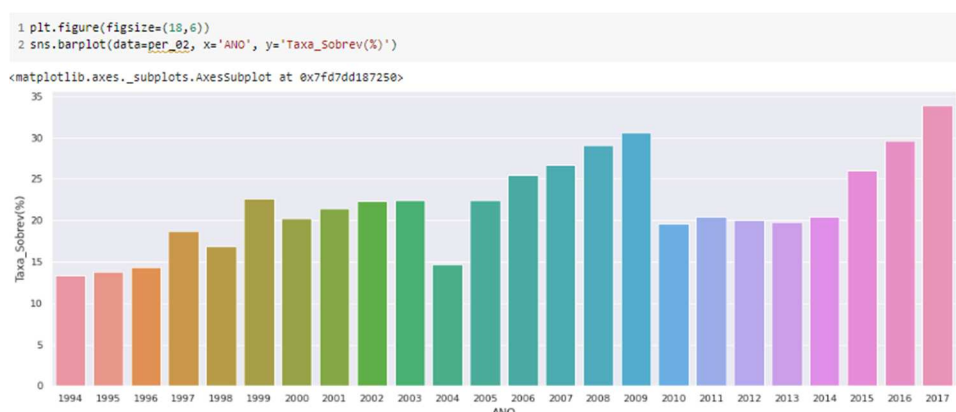
Ao analisar taxa de sobrevivência das empresas no período 01 conforme escopo definido neste trabalho, observou-se uma taxa média de 10,73% com desvio de 2,76 e variância de 7,66, o que representa pouca variação desta variável em relação à média. A

taxa de sobrevivência acumulada do período 01 foi de 225,37%. Abaixo na Figura 15 foi plotado a taxa de sobrevivência para melhor visualização.



**Figura 15. Taxa de sobrevivência das empresas no Brasil do período 01 (de 1964 a 1987) ano a ano.**  
Fonte: autores.

No período 02 a taxa de sobrevivência das empresas, plotada na Figura 16 apresentou taxa média de 21,87% com desvio de 5,48. A taxa mínima ocorreu em 1994, com 13,37% e máxima em 2017 com 33,92%. A variância auferida foi de 30,04 e taxa de sobrevivência acumulado de 525,04%.

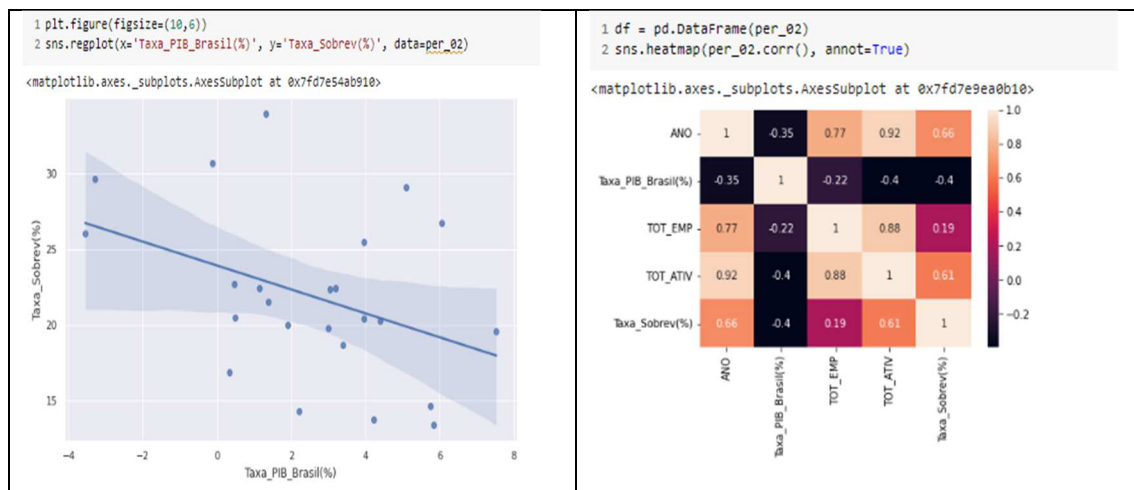


**Figura 16. Taxa de sobrevivência das empresas do Brasil do período 02 (de 1994 a 2017) ano a ano.**  
Fonte: autores.

Quando correlacionados a taxa de sobrevivência das empresas de PE e taxa de PIB Brasil no período 01, Figura 17, observou-se que os dados se apresentam dispersos com fraca correlação positiva de 0,38 entre estas variáveis.

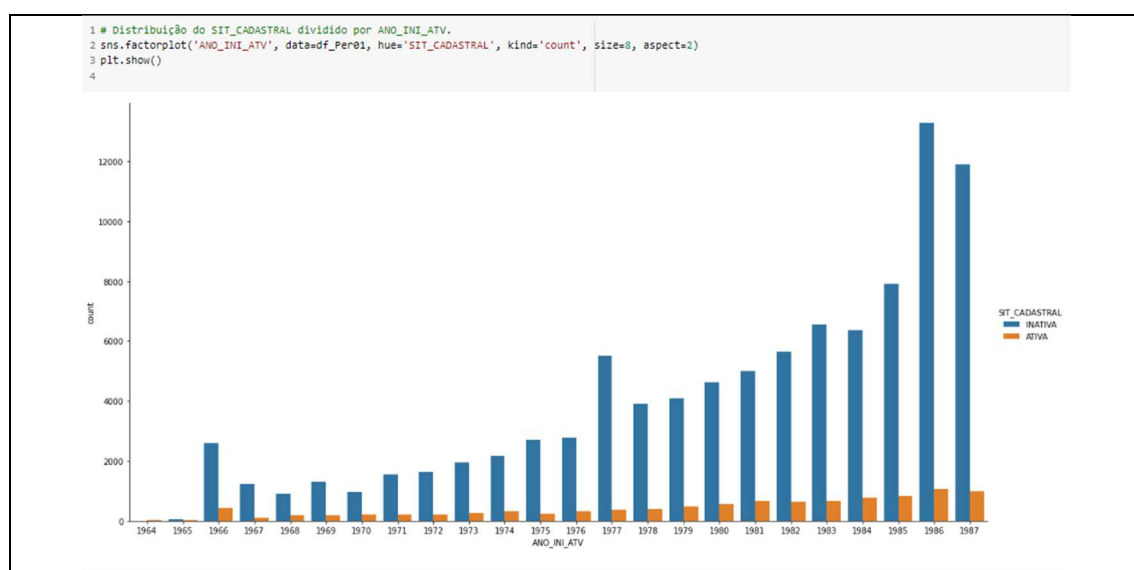
**Figura 17. Correlação Taxa de Sobrevivência das empresas de PE x PIB Brasil no período 01 (de 1964 a 1987). Fonte: autores.**

Fazendo-se a mesma análise para o período 02, Figura 18 os dados mantem-se dispersos, todavia com uma fraca correlação negativa de  $-0,4$ .

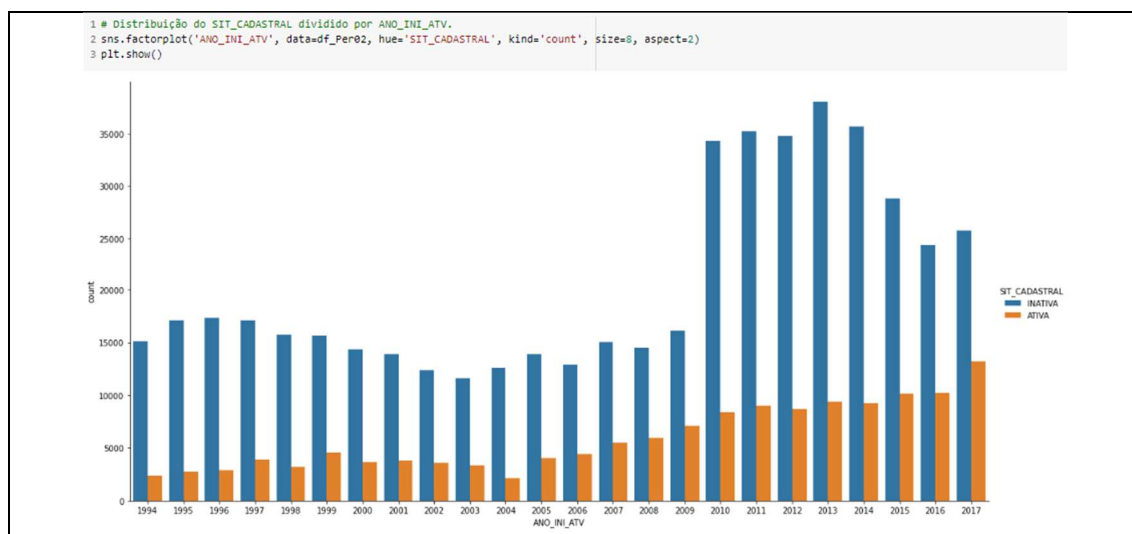


**Figura 18. Correlação Taxa de Sobrevivência das empresas de PE x PIB Brasil no período 02 (de 1994 a 2017). Fonte: autores.**

Na análise da situação cadastral ano a ano no período 01, a Figura 19 mostrou comparativamente quantidade de empresas cuja situação encontra-se ativa ou inativa no ano XX, representando um crescimento contínuo até o fim do período, sobretudo a partir de 1977 de 5489 no ano, contra 2781 do ano anterior, 1976. No período 02, Figura 20 percebe-se um crescimento de empresas ATIVAS comparativamente as inativas, de forma mais representativa a partir de 2010.



**Figura 19. Empresas ATIVAS x INATIVAS de PE do período 01. Fonte: autores.**



**Figura 20. Empresas ATIVAS x INATIVAS de PE do período 02. Fonte: autores.**

Ao caracterizar o perfil das empresas de PE, optou-se por unificar as atividades por setor segundo a nomenclatura CNAE 2.0<sup>17</sup> definida pelo IBGE. No período 01, visto na Figura 21 parte superior, a atividade de comercio varejista se sobressai, representando 55,15% do total de empresas deste período, e taxa de sobrevivência de 25,45%, seguida pela atividade de organização associativa com 3,45% e 9,91% respectivamente do total de empresas e empresas ativas. O comercio por atacado representa 5,78% do total de empresas e 5,58% das ativas.

No período 02, visto na figura 21 parte inferior, comercio varejista representa 36,56% das empresas abertas, e 27,29% das empresas ativas cuja abertura deu-se de 1988 a 2017. Educação, atividades de atenção à saúde e alimentação ganha, relevância,

<sup>17</sup> <https://cnae.ibge.gov.br>

representando respectivamente 2,91%, 1,64% e 6,07% das empresas abertas neste período e 4,19%, 3,73% e 3,30% das empresas sobreviventes.

SETOR DE ATIVIDADE - PERÍODO 01				%	
Setor	ATIVA	INATIVA	TOTAL	ATIVA	TOTAL
COMÉRCIO VAREJISTA	2.816	58.216	61.032	25,45%	55,15%
ATIVIDADES DE ORGANIZAÇÕES ASSOCIATIVAS	1.096	2.725	3.821	9,91%	3,45%
SERVIÇOS PARA EDIFÍCIOS E ATIVIDADES PAISAGÍSTICAS	1.037	162	1.199	9,37%	1,08%
COMÉRCIO POR ATACADO, EXCETO VEÍCULOS AUTOMOTORES E MOTOCICLETAS	617	5.780	6.397	5,58%	5,78%
CONSTRUÇÃO DE EDIFÍCIOS	383	1.026	1.409	3,46%	1,27%
COMÉRCIO E REPARAÇÃO DE VEÍCULOS AUTOMOTORES E MOTOCICLETAS	345	2.708	3.053	3,12%	2,76%
EDUCAÇÃO	316	928	1.244	2,86%	1,12%
FABRICAÇÃO DE PRODUTOS ALIMENTÍCIOS	299	2.029	2.328	2,70%	2,10%
ATIVIDADES ESPORTIVAS E DE RECREAÇÃO E LAZER	272	1.199	1.471	2,46%	1,33%
ATIVIDADES JURÍDICAS, DE CONTABILIDADE E DE AUDITORIA	271	204	475	2,45%	0,43%
SETOR DE ATIVIDADE - PERÍODO 02				%	
Setor	ATIVA	INATIVA	TOTAL	ATIVA	TOTAL
COMÉRCIO VAREJISTA	44.207	224.398	268.605	27,29%	36,56%
ATIVIDADES DE ORGANIZAÇÕES ASSOCIATIVAS	13.026	16.001	29.027	8,04%	3,95%
COMÉRCIO POR ATACADO, EXCETO VEÍCULOS AUTOMOTORES E MOTOCICLETAS	10.406	28.725	39.131	6,42%	5,33%
EDUCAÇÃO	6.784	14.582	21.366	4,19%	2,91%
ATIVIDADES DE ATENÇÃO À SAÚDE HUMANA	6.043	5.975	12.018	3,73%	1,64%
COMÉRCIO E REPARAÇÃO DE VEÍCULOS AUTOMOTORES E MOTOCICLETAS	6.001	22.113	28.114	3,71%	3,83%
SERVIÇOS PARA EDIFÍCIOS E ATIVIDADES PAISAGÍSTICAS	5.488	2.567	8.055	3,39%	1,10%
ALIMENTAÇÃO	5.352	39.217	44.569	3,30%	6,07%
SERVIÇOS DE ESCRITÓRIO, DE APOIO ADMINISTRATIVO E OUTROS SERVIÇOS PRESTADOS PRINCIPALMENTE ÀS EMPRESAS	5.130	17.160	22.290	3,17%	3,03%
CONFECÇÃO DE ARTIGOS DO VESTUÁRIO E ACESSÓRIOS	4.318	20.603	24.921	2,67%	3,39%

**Figura 21. Distribuição da taxa de sobrevivência por setor econômico. Fonte: autores.**

Como visto na Figura 22 na sua parte superior, quanto a natureza jurídica do período 01 apesar do tipo empresário individual representar um quantitativo maior de empresas abertas no período, o tipo a sociedade empresarial limitada apresentou maior longevidade representando 40,3% das empresas abertas neste período e que se apresentam como ativas na atualidade. No período 02 a representatividade dos dois maiores tipos empresariais ativos permanece as mesmas como mostra a Figura 22 na parte inferior, todavia em percentuais diferentes, sociedade empresarial representando 27,2% do total de empresas e 38,4% das ativas, enquanto o tipo empresário individual representa 60,5% das empresas abertas e 33,5% das ativas.

NATUREZA JURÍDICA - PERÍODO 01				%	
Natureza Jurídica	ATIVA	INATIVA	TOTAL	ATIVA	TOTAL
Sociedade Empresária Limitada	4.458	27.937	32.395	40,3%	29,3%
Empresário \Individual)	2.103	61.460	63.563	19,0%	57,4%
Associação Privada	1.563	4.054	5.617	14,1%	5,1%
Condomínio Edilício	1.021	7	1.028	9,2%	0,9%
Sociedade Anônima Fechada	627	1.716	2.343	5,7%	2,1%
Sociedade de Economia Mista	264	94	358	2,4%	0,3%
Empresa Individual de Responsabilidade Limitada \de Natureza Empresária)	260	119	379	2,3%	0,3%
Serviço Notarial e Registral \Cartório)	238	64	302	2,2%	0,3%
Sociedade Anônima Aberta	184	580	764	1,7%	0,7%
Sociedade Simples Limitada	154	796	950	1,4%	0,9%
NATUREZA JURÍDICA - PERÍODO 02				%	
Natureza Jurídica	ATIVA	INATIVA	TOTAL	ATIVA	TOTAL
Sociedade Empresária Limitada	62.128	137.885	200.013	38,4%	27,2%
Empresário \Individual)	54.327	390.443	444.770	33,5%	60,5%
Associação Privada	17.209	20.749	37.958	10,6%	5,2%
Empresa Individual de Responsabilidade Limitada \de Natureza Empresária)	15.264	11.431	26.695	9,4%	3,6%
Condomínio Edilício	4.863	25	4.888	3,0%	0,7%
Sociedade Anônima Fechada	2.623	2.797	5.420	1,6%	0,7%
Sociedade Simples Limitada	934	1.978	2.912	0,6%	0,4%
Sociedade Simples Pura	902	373	1.275	0,6%	0,2%
Sociedade Anônima Aberta	874	404	1.278	0,5%	0,2%
Sociedade em Conta de Participação	766	152	918	0,5%	0,1%



**Figura 22. Distribuição da taxa de sobrevivência por natureza jurídica. Fonte: autores.**

Por fim a Figura 23 mostra análise do porte das empresas sobreviventes nos dois períodos, tendo o porte 5 uma representatividade maior nos anos analisados, de 5359 empresas ativas do período 01 e 4386 do período 02.

PORTE PERÍODO 01 ▾ QTD		PORTE PERÍODO 02 ▾ QTD	
1	23286	1	26711
ATIVA	3818	ATIVA	4649
INATIVA	19468	INATIVA	22062
3	1507	3	1471
ATIVA	969	ATIVA	921
INATIVA	538	INATIVA	550
5	79790	5	62027
ATIVA	5359	ATIVA	4386
INATIVA	74431	INATIVA	57641
Total Geral	104583	Total Geral	90209

**Figura 23. Distribuição da taxa de sobrevivência por porte da empresa. Fonte: autores.**

## 5. Resultados e Discussão

A presente análise de dados foi efetuada para avaliar o perfil das empresas sobreviventes em períodos históricos distintos (de 1964 a 1987 e de 1988 a 2017), analisar o crescimento/decrescimento do número de empresas ano a ano e discutir se o PIB teria alguma influência no comportamento da taxa de sobrevivência das empresas estabelecidas no estado de Pernambuco nos dois períodos em análise.

Com os resultados encontrados, foi possível se chegar às seguintes conclusões sobre as questões inicialmente interpeladas:

### 1- Qual o perfil das empresas sobreviventes nos dois períodos?

As empresas de comércio varejista compõem a grande maioria das pessoas jurídicas abertas no país nos dois períodos analisados. Verificamos no período 02 um incremento muito significativo do quantitativo de empresas abertas, sobretudo após redemocratização e abertura de mercado, todavia, esta atividade permanece relevante no mercado. A atividade de comércio varejista compreende revenda de bens de consumo novos e usados para o consumidor final em pequenas quantidades representando o último elo da cadeia de distribuição. Em segundo lugar encontra-se as atividades de organização associativa que compreende organizações cujos interesses dos membros são o desenvolvimento e prosperidade de empresas ou de ramos comerciais específicos, atividades das organizações, federações e confederações empresariais e patronais, nos níveis nacional, estadual e municipal, centradas na representação e na comunicação, e as atividades das câmaras de comércio e das corporações e organismos similares. Apesar de não representar o segundo lugar no quantitativo total de empresas abertas, esta apresentou o segundo maior número de empresas sobreviventes nos dois períodos apresentados. Importante



atentar que a atividade de alimentação apesar de ocupar a segunda posição de número de empresas abertas nos dois períodos, com 7,02% período 01 e 6,07 no período 02, a longevidade é pequena relativamente a outras atividades, não estando entre as 10 mais do período 01, e ocupando a 8 posição de empresas ativas no período 02, mais recente, todavia com apenas 3,3% das empresas ativas abertas naquele período.

## 2 - Analisar o crescimento/decrescimento do número de empresas ano a ano.

Observou-se que o no período 01 o número de empresas abertas anualmente cresceu de forma contínua e gradativa, mesmo em períodos em que o PIB sofreu variação negativa. Apesar de não apresentar uma correlação taxativa, foi possível verificar um acréscimo da quantidade de empresas abertas em anos de maior crise econômica, o que consequentemente reverbera e aumento da taxa de desemprego. Diante desta afirmação, intuiu-se tratar-se do empreendedorismo por necessidade para muitos brasileiros que enfrentam o desemprego e buscam uma outra fonte de renda.

No período 01 este movimento ficou mais evidente em meados de 1977 quando tivemos um acréscimo de empresas abertas, 5870 na base da Receita Federal frente 3111 do ano anterior. Importante observar que o PIB em 1976 teve variação de 10,26% frente 4,96% de 1977 e 4,97% em 1978.

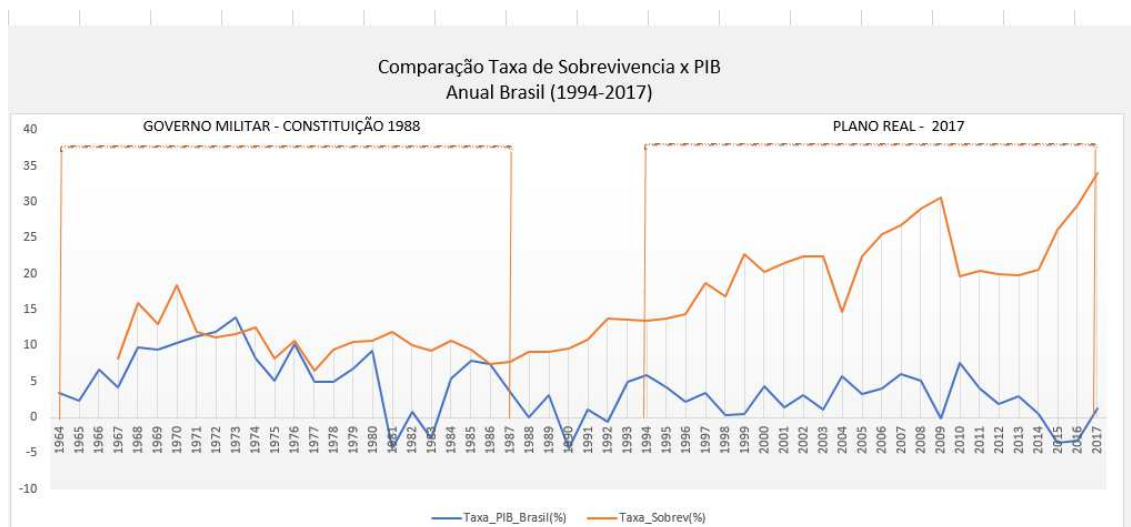
No período 02, verificou-se uma explosão do número de empresas abertas, relativamente ao período 01. O momento se apresenta com maior abertura de mercado, há maior incentivo a empreender com a redemocratização e universalização da informação por meio da internet o que propiciou amadurecimento da atividade empreendedora, além de maior instabilidade econômica.

Notadamente após de 2008, maior crise econômica contemporânea ocorrida nos EUA e que reverberaram por todo o globo, observamos que em PE ocorreu a partir de 2010 o dobro do número de empresas abertas anualmente, 2009 foram 23270 empresas e 2010 subiu para 42661 empresas abertas e cujo PIB de 2009 pairou em -0,13%. Enquanto a média do PIB no período de 2009 a 2017 girou em torno de 1,25% a.a a taxa de abertura de empresas ficou em 4,8% por ano.

## 3- Discutir se o PIB teria alguma influência no comportamento da taxa de sobrevivência das empresas estabelecidas no estado de Pernambuco nos dois períodos em análise.

Em relação ao terceiro objetivo não foi possível mensurar a influência da taxa de variação do PIB Brasil na sobrevivência das empresas durante os períodos analisados, pois os dados de situação cadastral das empresas disponíveis para análise refletem apenas o momento atual destas.

Foi possível mensurar a taxa de sobrevivência de longo prazo, entre a data de abertura da empresa e situação atual, o que representa a longevidade da atividade empresarial no Brasil, conforme mostra a Figura 24.



**Figura 24. Taxa de Sobrevivência x PIB Brasil. Fonte: autores.**

## 6. Considerações finais e trabalhos futuros

Este trabalho apresentou uma análise exploratória dos dados abertos da base CNPJ e série histórica do PIB, com a finalidade de caracterizar a taxa de sobrevivência das empresas em dois períodos distintos da História do Brasil, além de mapear as características de natureza jurídica, porte e atividade econômica destas empresas em confronto com a taxa de sobrevivência.

O objetivo foi alcançado em parte, em face dos dados disponíveis para análise não possuírem os requisitos necessários para estabelecer a influência do PIB ano na situação cadastral das empresas, todavia, foi possível analisar o perfil das empresas sobreviventes, cuja abertura deu-se nos anos escolhidos e que permanecem ativas, apresentando uma longevidade significativa. Foi verificado ainda uma incidência maior do número de empresas abertas em momentos de maior instabilidade econômica, todavia, dados internos das empresas tais como balanços, faturamento e quantidade de empregados, chamadas variáveis internas das empresas, poderiam dar maiores subsídios para uma análise mais completa. O período de verificação pode ser diminuído para o recomendado pela OECD, de forma que seja subdividido em vários períodos curtos.

Como trabalhos futuros, pode-se melhorar a metodologia de verificação da sobrevivência com o acréscimo de outros dados oriundos da RFB, tais como dados das declarações entregues, do Simples Nacional e da arrecadação. A presente

metodologia só verificou a situação de ativa/inativa a partir do valor final do campo “Situação Cadastral”.

Outra sugestão, seria acrescentar à análise os dados macroeconômicos (IDH, taxa de inflação, taxa de juros etc.) e dados econômicos dos municípios de localização das empresas, pois estas variáveis econômicas afetam sobremaneira o ambiente externo das empresas.

O período de verificação poderia ser diminuído para o recomendado pela OECD, de forma que o período longo seja subdividido em vários períodos curtos. Como foi visto no referencial teórico, os períodos de até 5 anos são mais adequados quando se analisa os micros e pequenos empreendimentos que constituem a maioria no Brasil.

## **Referências**

BRIX, U.; GROTZ, R. Regional Patterns and Determinants of New Firm Formation and Survival in Western Germany. IAB Discussion Paper, nº 5/2006.

CHIAVENATO, Idalberto. Comportamento Organizacional: A Dinâmica do Sucesso das Organizações. São Paulo: Grupo GEN, 2021.

GIAMBIAGI, Fábio et. al. (org.). Economia Brasileira Contemporânea [1945-2010]. 2ª edição. Rio de Janeiro: Elsevier, 2011.

GIAMBIAGI, Fábio; SCHWARTSMAN, Alexandre. Complacência. Rio de Janeiro: Elsevier, 2014.

GOMES, Elisabeth; BRAGA, Fabiane. Inteligência Competitiva Tempos Big Data. Rio de Janeiro: Editora Alta Books, 2017.

GREMAUD, Amaury et. al. Economia Brasileira Contemporânea. São Paulo: Atlas, 2009.

GREMAUD, Amaury P.; VASCONCELLOS, Marco Antonio Sandoval D.; JR., Rudinei T. Economia Brasileira Contemporânea. 8ª edição. São Paulo: Grupo GEN, 2016.

LOPES, Gesiel Rios et al. [Introdução à Análise Exploratória de Dados com Python](#). In: SOCIEDADE BRASILEIRA DE COMPUTAÇÃO (Teresina). Minicursos ERCAS & ENUCMPI. Porto Alegre: Sociedade Brasileira de Computação, 2019.

MCGUIGAN, James R.; MOYER, R C.; HARRIS, Frederick H. de B. Economia de Empresas: Aplicações, estratégia e táticas - Tradução da 13ª edição norte-americana. São Paulo: Cengage Learning Brasil, 2016.

MENDES, Marcos. Porque o Brasil cresce pouco?: Desigualdade, democracia e baixo crescimento no país do futuro. 1ª ed. Rio de Janeiro: Elsevier, 2014.

MARTIKAINEN, Jani. Data Mining in Tax Administration: Using Analytics to Enhance Tax Compliance. Information Systems Science Master's thesis (2012). Department of Information and Service Economy - Aalto University - School of Business.

NASCIMENTO, Hugo A. D. do; FERREIRA, Cristiane B. R. [Uma introdução à visualização de informações](#). Visualidades, Goiânia, v. 9, n. 2, p. 13-43, dez. 2011.

OECD – ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. SDBS Business Demography Indicators. Paris: OECD, 2016. Disponível em: <[http://stats.oecd.org/Index.aspx?DataSetCode=SDBS\\_BDI](http://stats.oecd.org/Index.aspx?DataSetCode=SDBS_BDI)>.

PROVOST, Foster, FAWCETT, Tom. Data Science para negócios [recurso eletrônico] – Rio de Janeiro, RJ: Alta Books, 2016

RAMOS, Carlos A. Economia da Felicidade: rumo a uma nova medição de prosperidade das nações. Rio de Janeiro: Editora Alta Books, 2021.

Receita Federal do Brasil. Cadeia de Valor da Receita Federal. Receita Federal do Brasil, 2021. <https://www.gov.br/receitafederal/pt-br/aceso-a-informacao/institucional/cadeia-de-valor-1>. Acesso em 19/06/2022.

Sobrevivência das empresas no Brasil. / Marco Aurélio Bedê (Coord.) – Brasília: Sebrae, 2016.

#### BASES CONJ RECEITA FEDERAL:

Cadastro Nacional de Pessoas Jurídicas - CNPJ — Português (Brasil) <<https://www.gov.br/receitafederal/pt-br/assuntos/orientacao-tributaria/cadastros/consultas/dados-publicos-cnpj>>

#### PRODUTO INTERNO BRUTO POR MUNICÍPIO:

[https://ftp.ibge.gov.br/Pib\\_Municipios/2019/base/base\\_de\\_dados\\_2010\\_2019\\_xls.zip](https://ftp.ibge.gov.br/Pib_Municipios/2019/base/base_de_dados_2010_2019_xls.zip)

#### ARRECADÇÃO POR MUNICÍPIO (2019) :

[https://www.gov.br/receitafederal/pt-br/aceso-a-informacao/dados-abertos/receitadata/arrecadacao/copy\\_of\\_arrecadacao-das-receitas-administradas-pela-rfb-por-municipio/arrecadacao-por-municipios](https://www.gov.br/receitafederal/pt-br/aceso-a-informacao/dados-abertos/receitadata/arrecadacao/copy_of_arrecadacao-das-receitas-administradas-pela-rfb-por-municipio/arrecadacao-por-municipios)

Brasil. Decreto n. 8.777, de 11 de maio de 2016. Institui a Política de Dados Abertos do Poder Executivo Federal. Diário Oficial da União, Brasília, 12/05/2016. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2016/decreto/D8777.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/D8777.htm). Acesso em 19/06/2022.

Tribunal de Contas da União. 5 motivos para a abertura de dados na administração pública. Brasília: TCU; 2015. Disponível em: <http://portal.tcu.gov.br/biblioteca-digital/cinco-motivos-para-a-abertura-dedados-na-administracao-publica.htm>. Acesso em 19/06/2022.

