

# RECEITA FEDERAL BRASILEIRA

RIO DE JANEIRO | NOVEMBRO 2024

## 1. Introdução

Este relatório apresenta uma análise detalhada dos metadados associados aos conjuntos de dados utilizados em nossos projetos recentes.

Metadados, que são informações sobre os dados, desempenham um papel crucial na organização, acesso e interpretação eficaz das informações. Através da avaliação dos metadados, buscamos entender a estrutura, o contexto e a origem dos dados, facilitando a gestão e a utilização adequada das informações em nossos processos analíticos.

Este documento fornece uma visão abrangente dos principais elementos dos metadados, incluindo descrições, formatos e relações entre os dados, oferecendo uma base sólida para futuras análises, utilizações e otimizações.

**Objetivo:** O Cadastro Nacional da Pessoa Jurídica (CNPJ) 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.

**Histórico:** Buscando aprimorar o processo de mapeamento de estabelecimentos, foi desenvolvido o app “Sistema Web Empresas”, e para isso, é preciso acessar a base de CNPJs da RFB, disponibilizando as empresas ativas por CNAE, UF e município.

**Sigla:** RFB.

**Temática:** Empresas cadastradas na Receita Federal Brasileira.

**Situação do Processo:** Ativo.

## 2. Descrição das Fontes de Dados

- 2.1. Fontes: Site disponibilizado pela RFB (<https://200.152.38.155/CNPJ/>).
- 2.2. Forma obtenção: Extração via robô (python).
- 2.3. Periodicidade: Trimestral.
- 2.4. Início da série: Julho de 2022.
- 2.5. Instituição Responsável: Secretaria Especial da Receita Federal do Brasil (RFB).
- 2.6. Área interna Responsável: Coleta não tradicional – SPDO.
- 2.7. Unidade(s) de Análise: Dados cadastrais.
- 2.8. Técnica de Investigação: Censitária.
- 2.9. População Alvo: CNPJs em território brasileiro, focado nas empresas ativas.
- 2.10. Abrangência Geográfica: Nacional.

## 3. Tratamento dos Dados

3.1. Processos de Limpeza: Para trabalhar com uma base de empresas ativas, foram filtrados apenas os registros com STATUS ativo no momento da coleta.

3.2. Transformações: Para consolidar as 10 tabelas obtidas na extração, é realizada a seguinte *Query* no SQL:

```
“
CREATE TABLE mvp_cons AS(
SELECT
concat(e.cnpj,"", e.cnpj_ordem,"", e.cnpj_dv) as cnpj,
e.nome as nome_fantasia,
emp.razao as razao_social,
e.matriz_filial,
emp.porte,
emp.capital,
e.situacao,
e.cnae_fiscal,
concat(e.cnae_fiscal, ' - ', c.descricao) as cnae_descr,
e.cnae_secundario,
e.logradouro,
e.numero,
e.complemento,
e.bairro,
e.cep,
e.uf,
m.descricao as municipio,
e.ddd_1,
e.telefone_1,
e.ddd_2,
e.telefone_2,
```

```

e.email
FROM
    public.estabelecimentos e
    LEFT JOIN public.municipios m ON e.municipio = m.codigo
    LEFT JOIN public.empresas emp ON e.cnpj = emp.cnpj
    LEFT JOIN public.cnaes c ON e.cnae_fiscal = c.codigo
WHERE
    e.situacao='2'
);

```

3.3. Padronização: Não foi realizado nenhum tipo de padronização.

3.4. Metodologia: A base é disponibilizada mensalmente no site da RFB, considerando as alterações cadastrais, como endereço, telefone, CNAE ou até mesmo o status da empresa, além de incluir novos CNPJs que foram cadastrados.

3.5. Época da Coleta: Trimestral (Janeiro – Abril – Julho – Outubro).

3.6. Tempo Previsto entre o Início da Coleta e a Liberação dos Dados: 5 dias.

## 4. Qualidade dos Dados

4.1. Dicionário de Variáveis:

4.1.1. CNAEs

Variável	Tipo de Variável	Descrição	Constraints
CÓDIGO	string	CÓDIGO DA ATIVIDADE ECONÔMICA	PK
DESCRIÇÃO	string	NOME DA ATIVIDADE ECONÔMICA	NOT NULL

#### 4.1.2. Dados simples

Variável	Tipo de Variável	Descrição	Constraints
CNPJ BÁSICO	string	NÚMERO BASE DE INSCRIÇÃO NO CNPJ (OITO PRIMEIROS DÍGITOS DO CNPJ).	NOT NULL
OPÇÃO PELO SIMPLES	string	INDICADOR DA EXISTÊNCIA DA OPÇÃO PELO SIMPLES. S - SIM N - NÃO EM BRANCO – OUTROS	NOT NULL
DATA DE OPÇÃO PELO SIMPLES	date	DATA DE OPÇÃO PELO SIMPLES	NOT NULL
DATA DE EXCLUSÃO DO SIMPLES	date	DATA DE EXCLUSÃO DO SIMPLES	-
OPÇÃO PELO MEI	string	INDICADOR DA EXISTÊNCIA DA OPÇÃO PELO MEI S - SIM N - NÃO EM BRANCO - OUTROS	NOT NULL
DATA DE OPÇÃO PELO MEI	date	DATA DE OPÇÃO PELO MEI	-
DATA DE EXCLUSÃO DO MEI	date	DATA DE EXCLUSÃO DO MEI	-

#### 4.1.3. Empresas

Variável	Tipo de Variável	Descrição	Constraints
CNPJ BÁSICO	string	NÚMERO BASE DE INSCRIÇÃO NO CNPJ (OITO PRIMEIROS DÍGITOS DO CNPJ).	NOT NULL
RAZÃO SOCIAL / NOME EMPRESARIAL	string	NOME EMPRESARIAL DA PESSOA JURÍDICA	-
NATUREZA JURÍDICA	integer	CÓDIGO DA NATUREZA JURÍDICA	NOT NULL
QUALIFICAÇÃO DO RESPONSÁVEL	integer	QUALIFICAÇÃO DA PESSOA FÍSICA RESPONSÁVEL PELA EMPRESA	NOT NULL
CAPITAL SOCIAL DA EMPRESA	integer	CAPITAL SOCIAL DA EMPRESA	NOT NULL
PORTE DA EMPRESA	integer	CÓDIGO DO PORTE DA EMPRESA: 00 – NÃO INFORMADO 01 - MICRO EMPRESA 03 - EMPRESA DE PEQUENO PORTE 05 - DEMAIS	-
ENTE FEDERATIVO RESPONSÁVEL	integer	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.	-

#### 4.1.4. Estabelecimentos

Variável	Tipo de Variável	Descrição	Constraints
CNPJ BÁSICO	string	NÚMERO BASE DE INSCRIÇÃO NO CNPJ (OITO PRIMEIROS DÍGITOS DO CNPJ).	NOT NULL
CNPJ ORDEM	string	NÚMERO DO ESTABELECIMENTO DE INSCRIÇÃO NO CNPJ (DO NONO ATÉ O DÉCIMO SEGUNDO DÍGITO DO CNPJ).	NOT NULL
CNPJ DV	string	DÍGITO VERIFICADOR DO NÚMERO DE INSCRIÇÃO NO CNPJ (DOIS ÚLTIMOS DÍGITOS DO CNPJ).	NOT NULL
IDENTIFICADOR MATRIZ/FILIAL	integer	CÓDIGO DO IDENTIFICADOR MATRIZ/FILIAL: 1 – MATRIZ 2 – FILIAL	NOT NULL
NOME FANTASIA	string	CORRESPONDE AO NOME FANTASIA	-
SITUAÇÃO CADASTRAL	integer	CÓDIGO DA SITUAÇÃO CADASTRAL: 01 – NULA 2 – ATIVA 3 – SUSPENSA 4 – INAPTA 08 – BAIXADA	NOT NULL
DATA SITUAÇÃO CADASTRAL	date	DATA DO EVENTO DA SITUAÇÃO CADASTRAL	-
MOTIVO SITUAÇÃO CADASTRAL	string	CÓDIGO DO MOTIVO DA SITUAÇÃO CADASTRAL	NOT NULL
NOME DA CIDADE NO EXTERIOR	string	NOME DA CIDADE NO EXTERIOR	-
PAIS	integer	CÓDIGO DO PAIS	-
DATA DE INÍCIO ATIVIDADE	date	DATA DE INÍCIO DA ATIVIDADE	NOT NULL
CNAE FISCAL PRINCIPAL	string	CÓDIGO DA ATIVIDADE ECONÔMICA PRINCIPAL DO ESTABELECIMENTO	NOT NULL
CNAE FISCAL SECUNDÁRIA	string	CÓDIGO DA(S) ATIVIDADE(S)	-

		ECONÔMICA(S) SECUNDÁRIA(S) DO ESTABELECIMENTO	
TIPO DE LOGRADOURO	string	DESCRIÇÃO DO TIPO DE LOGRADOURO	-
LOGRADOURO	string	NOME DO LOGRADOURO ONDE SE LOCALIZA O ESTABELECIMENTO.	-
NÚMERO	string	NÚMERO ONDE SE LOCALIZA O ESTABELECIMENTO. QUANDO NÃO HOVER PREENCHIMENTO DO NÚMERO HAVERÁ 'S/N'.	NOT NULL
COMPLEMENTO	string	COMPLEMENTO PARA O ENDEREÇO DE LOCALIZAÇÃO DO ESTABELECIMENTO	-
BAIRRO	string	BAIRRO ONDE SE LOCALIZA O ESTABELECIMENTO	-
CEP	string	CÓDIGO DE ENDEREÇAMENTO POSTAL REFERENTE AO LOGRADOURO NO QUAL O ESTABELECIMENTO ESTA LOCALIZADO	-
UF	string	SIGLA DA UNIDADE DA FEDERAÇÃO EM QUE SE ENCONTRA O ESTABELECIMENTO	NOT NULL
MUNICÍPIO	integer	CÓDIGO DO MUNICÍPIO DE JURISDIÇÃO ONDE SE ENCONTRA O ESTABELECIMENTO	NOT NULL
DDD 1	string	CONTÉM O DDD 1	-
TELEFONE 1	string	CONTÉM O NÚMERO DO TELEFONE 1	NOT NULL
DDD 2	string	CONTÉM O DDD 2	-
TELEFONE 2	string	CONTÉM O NÚMERO DO TELEFONE 2	-
DDD DO FAX	string	CONTÉM O DDD DO FAX	-
FAX	string	CONTÉM O NÚMERO DO FAX	-
CORREIO ELETRÔNICO	string	CONTÉM O E-MAIL DO CONTRIBUINTE	-
SITUAÇÃO ESPECIAL	string	SITUAÇÃO ESPECIAL DA EMPRESA	-



DATA DA SITUAÇÃO ESPECIAL	date	DATA EM QUE A EMPRESA ENTROU EM SITUAÇÃO ESPECIAL	-
---------------------------	------	---	---

#### 4.1.5. Motivo cadastral

Variável	Tipo de Variável	Descrição	Constraints
CÓDIGO	integer	CÓDIGO DO MOTIVO CADASTRAL	PK
DESCRIÇÃO	string	NOME DO MOTIVO CADASTRAL	NOT NULL

#### 4.1.6. Municípios

Variável	Tipo de Variável	Descrição	Constraints
CÓDIGO	integer	CÓDIGO DO MUNICÍPIO	PK
DESCRIÇÃO	string	NOME DO MUNICÍPIO	NOT NULL

#### 4.1.7. Naturezas

Variável	Tipo de Variável	Descrição	Constraints
CÓDIGO	integer	CÓDIGO DA NATUREZA JURÍDICA	PK
DESCRIÇÃO	string	NOME DA NATUREZA JURÍDICA	NOT NULL

#### 4.1.8. Países

Variável	Tipo de Variável	Descrição	Constraints
CÓDIGO	integer	CÓDIGO DO PAÍS	PK
DESCRIÇÃO	string	NOME DO PAÍS	NOT NULL

#### 4.1.9. Qualificações

Variável	Tipo de Variável	Descrição	Constraints
CÓDIGO	integer	CÓDIGO DA QUALIFICAÇÃO DO SÓCIO	PK
DESCRIÇÃO	string	NOME DA QUALIFICAÇÃO DO SÓCIO	NOT NULL

#### 4.1.10. CNAEs

Variável	Tipo de Variável	Descrição	Constraints
CNPJ BÁSICO	string	NÚMERO BASE DE INSCRIÇÃO NO CNPJ (CADASTRO NACIONAL DA PESSOA JURÍDICA).	NOT NULL
IDENTIFICADOR DE SÓCIO	integer	CÓDIGO DO IDENTIFICADOR DE SÓCIO 1 – PESSOA JURÍDICA 2 – PESSOA FÍSICA 3 – ESTRANGEIRO	NOT NULL
NOME DO SÓCIO (NO CASO PF) OU RAZÃO SOCIAL (NO CASO PJ)	string	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	string	CPF OU CNPJ DO SÓCIO (SÓCIO ESTRANGEIRO NÃO TEM ESTA INFORMAÇÃO).	-
QUALIFICAÇÃO DO SÓCIO	integer	CÓDIGO DA QUALIFICAÇÃO DO SÓCIO	NOT NULL
DATA DE ENTRADA SOCIEDADE	date	DATA DE ENTRADA NA SOCIEDADE	NOT NULL
PAIS	integer	CÓDIGO PAÍS DO SÓCIO ESTRANGEIRO	-
REPRESENTANTE LEGAL	string	NÚMERO DO CPF DO REPRESENTANTE LEGAL	NOT NULL
NOME DO REPRESENTANTE	string	NOME DO REPRESENTANTE LEGAL	-
QUALIFICAÇÃO DO REPRESENTANTE LEGAL	integer	CÓDIGO DA QUALIFICAÇÃO DO REPRESENTANTE LEGAL	NOT NULL
FAIXA ETÁRIA	integer	CÓDIGO CORRESPONDENTE À FAIXA ETÁRIA DO SÓCIO	NOT NULL

#### 4.1.11.Consolidada (Tabela utilizada pelos usuários)

Variável	Tipo de Variável	Descrição	Constraints
cnpj	string	NUMERO DE INSCRICAO NO CNPJ (DOZE PRIMEIROS DIGITOS)	NOT NULL
nome_fantasia	string	CORRESPONDE AO NOME FANTASIA	-
razao_social	string	NOME EMPRESARIAL DA PESSOA JURIDICA	NOT NULL
matriz_filial	integer	CODIGO DO IDENTIFICADOR MATRIZ/FILIAL: 1 - MATRIZ 2 - FILIAL	NOT NULL
porte	integer	CODIGO DO PORTE DA EMPRESA: 00 - NAO INFORMADO 01 - MICRO EMPRESA 03 - EMPRESA DE PEQUENO PORTE 05 - DEMAIS	NOT NULL
situacao	integer	CODIGO DA SITUACAO CADASTRAL: 2 - ATIVA	NOT NULL
cnae_fiscal	integer	CODIGO DA ATIVIDADE ECONOMICA PRINCIPAL DO ESTABELECIMENTO	NOT NULL
cnae_descr	string	CNAE + DESCRICAO	NOT NULL
cnae_secundario	string	CODIGO DA(S) ATIVIDADE(S) ECONOMICA(S) SECUNDARIA(S) DO ESTABELECIMENTO	-
logradouro	string	NOME DO LOGRADOURO ONDE SE LOCALIZA O ESTABELECIMENTO.	-
numero	integer	NUMERO ONDE SE LOCALIZA O ESTABELECIMENTO. QUANDO NAO HOUVER PREENCHIMENTO DO NUMERO HAVERA 'S/N'.	-
complemento	string	COMPLEMENTO PARA O ENDEREÇO DE LOCALIZACAO DO ESTABELECIMENTO	-
bairro	string	BAIRRO ONDE SE LOCALIZA O ESTABELECIMENTO.	-
cep	string	CODIGO DE ENDERECAMENTO POSTAL REFERENTE AO LOGRADOURO NO QUAL O	-

		ESTABELECIMENTO ESTA LOCALIZADO	
uf	string	SIGLA DA UNIDADE DA FEDERACAO EM QUE SE ENCONTRA O ESTABELECIMENTO	-
municipio	string	CODIGO DO MUNICIPIO DE JURISDICAÇÃO ONDE SE ENCONTRA O ESTABELECIMENTO	-
ddd_1	string	CONTEM O DDD 1	-
telefone_1	string	CONTEM O NUMERO DO TELEFONE	-
ddd_2	string	CONTEM O DDD 2	-
telefone_2	string	CONTEM O NUMERO DO TELEFONE 2	-
email	string	CONTEM O E-MAIL DO CONTRIBUINTE	-

#### 4.2. Métricas:

4.2.1. Quantidade de registros: 24.861.758.

4.2.2. Referência: Agosto de 2024.

4.2.3. Última atualização: 14 de agosto de 2024.

4.2.4. Problemas Identificados: Nenhum problema foi identificado até o momento.

## 5. Disponibilização dos dados

5.1. Armazenamento e Formatado: Estrutura de banco de dados organizada no *PgAdmin* (cada usuário precisa dessa estrutura armazenada em seu computador, atualmente).

5.2. Método de acesso aos dados: Aplicativo instalado na máquina de cada usuário.

5.3. Interface de Acesso: Aplicativo em *RShiny*.

5.4. Funcionalidades: Consulta por CNAE, UF e município, auxiliando o mapeamento de novas empresas.

5.5. Nível de Divulgação: Estadual e municipal, a depender da consulta realizada.

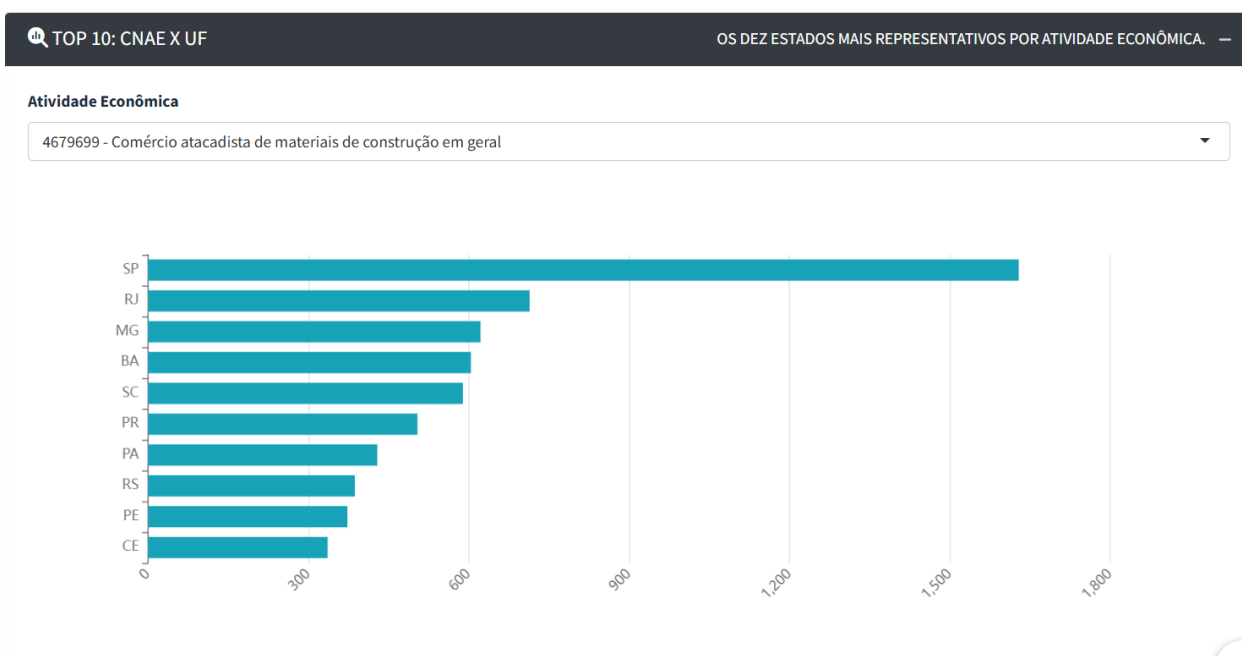
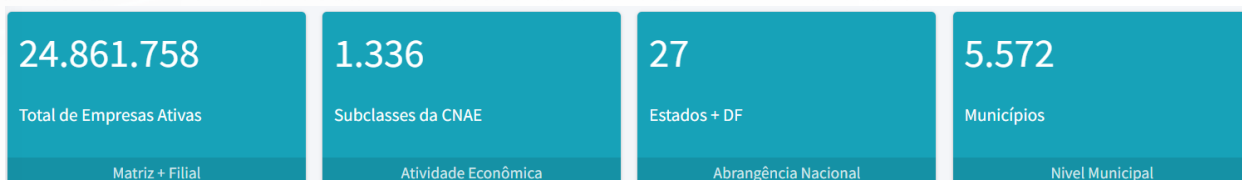
5.6. Área responsável pelo acesso: Gestão de Dados – SPDO.

## 6. Overview dos Dados

### 6.1. Demonstração dos Dados:

cnpj	nome_fantasia	razao_social	matriz_filial	porte	capital	situacao	cnae_fiscal	cnae_descr
54187028000109		54.187.028 LORENA CRISTINE COSTA DA SILVA	1	1	5000	2	4744099	4744099 - Comércio varejista de materiais de construção em geral
<b>cnae_secundario</b> 4321500								
<b>logradouro</b> COSMORAMA								
<b>numero</b> 900								
<b>complemento</b> QUADRA B;LOTE 21;CASA 130								
<b>bairro</b> COSMORAMA								
<b>cep</b> 26582020								
<b>uf</b> RJ								
<b>municipio</b> MESQUITA								
<b>ddd_1</b> 21								
<b>telefone_1</b> 83256225								
<b>ddd_2</b>								
<b>telefone_2</b>								
<b>email</b> KECONCERTINA@GMAIL.COM								

## 6.2. Estatísticas Descritivas:



CNAE X CIDADES

OS MUNICÍPIOS MAIS REPRESENTATIVOS POR ATIVIDADE ECONÔMICA. —

UF

RJ

cnae_descr	uf	municipio	↓ count
4679699 - Comércio atacadista de materiais de construção em geral	RJ	RIO DE JANEIRO	278
4679699 - Comércio atacadista de materiais de construção em geral	RJ	DUQUE DE CAXIAS	44
4679699 - Comércio atacadista de materiais de construção em geral	RJ	NITEROI	40
4679699 - Comércio atacadista de materiais de construção em geral	RJ	SAO GONCALO	32
4679699 - Comércio atacadista de materiais de construção em geral	RJ	SEROPEDICA	22
4679699 - Comércio atacadista de materiais de construção em geral	RJ	NOVA IGUACU	20
4679699 - Comércio atacadista de materiais de construção em geral	RJ	ARARUAMA	19
4679699 - Comércio atacadista de materiais de construção em geral	RJ	SAO JOAO DE MERITI	18

+





INSTITUTO  
BRASILEIRO  
DE ECONOMIA

 [portalibre.fgv.br](http://portalibre.fgv.br)