

- **fl\_matriz**: boolean value, `true` if the CNPJ corresponds to the "matriz".
- **de\_natureza\_juridica**: character, juridic nature of the company.
- **sg\_uf**: character, initials of the name of the state of the company.
- **natureza\_juridica\_macro**: a macro legal category for "natureza jurídica".
- **de\_ramo**: description of a macro activity category/branch based on the CNAE code.
- **setor**: character, economic sector of the company (Industry, Services, Agrobusiness, Etc.)
- **idade\_empresa\_anos**: numeric value, age of the company.
- **idade\_emp\_cat**: character, age of the company by ranges
- **fl\_me**: boolean value, true if the company has the term 'ME' in the end of its legal name.
- **fl\_sa**: boolean value, true if the company has the term 'SA' in the end of its legal name.
- **fl\_epp**: boolean value, true if the company has the term 'EPP' in the end of its legal name.
- **fl\_mei**: boolean value, true if the company has the term 'MEI' in the end of its legal name.
- **fl ltda**: boolean value, true if the company has the term 'LTDA' in the end of its legal name.
- **dt\_situacao**: date when the "de\_situacao" was registered by the IRS.
- **fl\_st\_especial**: boolean value, true if `dt_situacao_especial` is not null. If it is not null means that some extraordinary situation is identified by the IRS (*ESPOLIO DE EMPRESARIO EMPRESA INDIVIDUAL OU EIRELI, FALIDO, EM LIQUIDACAO, LIQUIDACAO JUDICIAL, LIQUIDACAO EXTRA JUDICIAL, REGISTRO NA JUNTA COMERCIAL EM ANDAMENTO, EM LIQUID EXTRA JUDICIAL, RECUPERACAO JUDICIAL, INTERVENCAO*)
- **fl\_email**: boolean value, true if the cnpj has an email registered at the IRF database.
- **fl\_telefone**: boolean value, true if the company has a phone number registered in IRS database.
- **fl\_rm**: character, true if the company address is located in an metropolitan area.
- **nm\_divisao**: character, name description of the primary economic activity of the company
- **nm\_segmento**: character, name description of the primary economic activity cluster of the company
- **fl\_spa**: boolean, true if the company has it's own fuel station
- **fl\_antt**: boolean, true if the company is certified by ANTT for transportation purposes
- **fl\_veiculo**: boolean, true if the company owns at least one vehicle
- **vl\_total\_tancagem**: numeric, total capacity of fuel storage of the company
- **vl\_total\_veiculos\_antt**: integer, total number of vehicles of the company
- **vl\_total\_veiculos\_leves**: integer, total number of light weight vehicles of the company
- **vl\_total\_veiculos\_pesados**: integer, total number of heavy weight vehicles of the company
- **fl\_optante\_simples**: boolean, true if the company is taxed according to SIMPLES regime
- **qt\_art**: integer, number of constructions(buildings) authorized by an architect (ART - Anotação de Responsabilidade Técnica)
- **vl\_total\_veiculos\_pesados\_grupo**: numeric, economic value of the heavy weight vehicles of the economic group of the company
- **vl\_total\_veiculos\_leves\_grupo**: numeric, economic value of the light weight vehicles of the economic group of the company
- **vl\_total\_tancagem\_grupo**: numeric, total capacity of fuel storage of the economic group of the company
- **vl\_total\_veiculos\_antt\_grupo**: numeric, total number of vehicles of the economic group of the company that are registered with ANTT certification
- **vl\_potenc\_cons\_oleo\_gas**: numeric, potential consumption of oil and gas
- **fl\_optante\_simei**: boolean, true if the company is taxed as an MEI - Individual Micro Entrepreneur (Micro Empreendedor Individual)

- **sg\_uf\_matriz**:character, state of the main office location
- **de\_saude\_tributaria**:character, indicator of health tax status, Green if all tax are OK, Red if none are OK
- **de\_saude\_rescencia**:character, indicates time of update of the most lagged input of the indicator.
- **nu\_meses\_rescencia**:character, number of months since the last update of the most lagged input from saude\_tributaria
- **de\_nivel\_atividade**:character, probability of being operating, ALTA high probability, BAIXA, low probability.
- **de\_indicador\_telefone**:character, propability to reach the company using the phone numbers provided by Neoway
- **fl\_simples\_irregular**: boolean, true if the company is taxed under the SIMPLES regime and has a impeditive CNAE (economic activities that are not allowed to be taxed according to the simples regime) revenue above the limit of the regime.
- **vl\_frota**: numeric, aggregate value of all vehicles of the company
- **empsetorcensitariofaixarendapopulacao**: numeric, average income from a sample of residents in a census unity (smallest territory area according to IBGE)
- **nm\_meso\_regiao**:character, name of the meso region where the company is located.
- **nm\_micro\_regiao**:character, name of the micro region where the company is located.
- **fl\_passivel\_iss**:boolean, true if the company performs any activity that is taxable under ISS - Tax under services.
- **qt\_socios**: integer value, quantity of partners/shareholders of the cnpj
- **qt\_socios\_pf**: integer value, quantity of partners/shareholders of the cnpj that are persons
- **qt\_socios\_pj**: integer value, quantity of partners/shareholders of the cnpj that are companies
- **idade\_media\_socios**: numeric value, average age of the partners of the cnpj(considers only the cpf partners)
- **idade\_maxima\_socios**: same as above but with max metric
- **idade\_minima\_socios**: same as above but wiht min metric
- **qt\_socios\_st\_regular**: integer value, quantity of partners with regular situation under IRS - Receita Federal
- **qt\_socios\_st\_suspensa**: integer value, quantity of partners with suspended situation under IRS - Receit Federal
- **qt\_socios\_masculino**: integer value, quantity of male partners
- **qt\_socios\_feminino**: integer value, quantity of female partners (womens)
- **qt\_socios\_pep**: integer value, quantity of partners of the cnpj that are politically exposed
- **qt\_alteracao\_socio\_total**: integer value, quantity of changes in the QSA
- **qt\_alteracao\_socio\_90d**: integer value, quantity of changes in the QSA in the last 90 days
- **qt\_alteracao\_socio\_180d**: integer value, quantity of changes in the QSA in the last 180 days
- **qt\_alteracao\_socio\_365d**: integer value, quantity of changes in the QSA in the last 365 days
- **qt\_socios\_pj\_ativos**: integer value, quantity of PJ partners that are ATIVA in the RF.
- **qt\_socios\_pj\_nulos**: integer value, quantity of PJ partners that are NULA in the RF.
- **qt\_socios\_pj\_baixados**: integer value, quantity of PJ partners that are BAIXADA in the RF.
- **qt\_socios\_pj\_suspensos**: integer value, quantity of PJ partners that are SUSPENSA in the RF.
- **qt\_socios\_pj\_inaptos**: integer value, quantity of PJ partners that are INAPTA in the RF.
- **vl\_idade\_media\_socios\_pj**: float value, avarage age of the PJ partners.
- **vl\_idade\_maxima\_socios\_pj**: float value, maximum age of the PJ partners.
- **vl\_idade\_minima\_socios\_pj**: float value, minimum age of the PJ partners.

- **qt\_coligados:** integer value, quantity of connected companies
- **qt\_socios\_coligados:** integer value, sum of partners from connected companies
- **qt\_coligados\_matriz:** integer value, quantity of connected companies that are a matrix company
- **qt\_coligados\_ativo:** integer value, quantity of connected companies that are considered as active according to the IRS - receita federal
- **qt\_coligados\_baixada:** integer value, quantity of connected companies that are considered as closed according to the IRS
- **qt\_coligados\_inapta:** integer value, quantity of connected companies that are considered as inapt/unfit according to the IRS
- **qt\_coligados\_suspensa:** integer value, quantity of connected companies that are considered suspended according to the IRS
- **qt\_coligados\_nula:** integer value, quantity of connected companies that are considered as nule companies according to the IRS
- **idade\_media\_coligadas:** float, average age (months) of connected companies
- **idade\_maxima\_coligadas:** float, maximum age (months) of connected companies
- **idade\_minima\_coligadas:** float, minimum age (months) of connected companies
- **coligada\_mais\_nova\_ativa:** float, age (months) of youngest coligate that is active
- **coligada\_mais\_antiga\_ativa:** float, age (months) of oldest coligate that is active
- **idade\_media\_coligadas\_ativas:** float, average age (months) of active connected companies
- **coligada\_mais\_nova\_baixada:** float, age (months) of yougest closed coligate
- **coligada\_mais\_antiga\_baixada:** float, age (months) of oldest closed coligate
- **idade\_media\_coligadas\_baixadas:** float, average (months) of closed coligate
- **qt\_coligados\_sa:** integer value, quantity of connected companies that has the term SA in the end of the business name
- **qt\_coligados\_me:** integer value, quantity of connected companies that has the term ME in the end of the business name
- **qt\_coligados\_mei:** integer value, quantity of connected companies that has the term MEI in the end of the business name
- **qt\_coligados ltda:** integer value, quantity of connected companies that has the term LTDA in the end of the business name
- **qt\_coligados\_epp:** integer value, quantity of connected companies that has the term EPP in the end of the business name
- **qt\_coligados\_norte:** integer value, quantity of connected companies that are located in the north region of the country
- **qt\_coligados\_sul:** integer value, quantity of connected companies that are located in the south region of the country
- **qt\_coligados\_nordeste:** integer value, quantity of connected companies that are located in the northeast region of the country
- **qt\_coligados\_centro:** integer value, quantity of connected companies that are located in the center-west (centro-oeste) region of the country
- **qt\_coligados\_sudeste:** integer value, quantity of connected companies that are located in the southeast of the country
- **qt\_coligados\_exterior:** integer value, quantity of connected companies that are located abroad the country

- **qt\_ufs\_coligados:** integer value, quantity of distinct states where the connected companies are located
- **qt\_regioes\_coligados:** integer value, quantity of distinct ufs where the connected companies are located
- **qt\_ramos\_coligados:** integer value, quantity of distinct economic branches from connected companies
- **qt\_coligados\_industria:** integer value, quantity of connected companies from the industry sector
- **qt\_coligados\_agropecuaria:** integer value, quantity of connected companies from the agrobusiness sector
- **qt\_coligados\_comercio:** integer value, quantity of connected companies
- **qt\_coligados\_servico:** integer value, quantity of connected companies from the service sector
- **qt\_coligados\_ccivil:** integer value, quantity of connected companies from the construction sector
- **qt\_funcionarios\_coligados:** integer value, quantity of employees from the connected companies
- **qt\_funcionarios\_coligados\_gp:** integer value, quantity of employees from connected companies and their subsidiaries
- **media\_funcionarios\_coligados\_gp:** numeric value, average of employees from the connected companies and their subsidiaries
- **max\_funcionarios\_coligados\_gp:** numeric value, maximum number of employees from connected companies and their subsidiaries
- **min\_funcionarios\_coligados\_gp:** numeric value, minimum number of employees from connected companies and their subsidiaries
- **vl\_folha\_coligados:** numeric value, sum of payroll value from connected companies
- **media\_vl\_folha\_coligados:** numeric value, average of payroll value from connected companies
- **max\_vl\_folha\_coligados:** numeric value, maximum of payroll value from connected companies
- **min\_vl\_folha\_coligados:** numeric value, minimum of payroll value from connected companies
- **vl\_folha\_coligados\_gp:** numeric value, sum of payroll value from connected companies and their subsidiaries
- **media\_vl\_folha\_coligados\_gp:** numeric value, average of payroll value from connected companies and their subsidiaries
- **max\_vl\_folha\_coligados\_gp:** numeric value, maximum of payroll value from connected companies and their subsidiaries
- **min\_vl\_folha\_coligados\_gp:** numeric value, minimum of payroll value from connected companies and their subsidiaries
- **faturamento\_est\_coligados:** numeric value, sum of estimated revenue from connected companies
- **media\_faturamento\_est\_coligados:** numeric value, average of estimated revenue from connected companies
- **max\_faturamento\_est\_coligados:** numeric value, maximum value of estimated revenue from connected companies
- **min\_faturamento\_est\_coligados:** numeric value, minimum value of estimated revenue from connected companies
- **faturamento\_est\_coligados\_gp:** numeric value, sum of estimated revenue from connected companies and their subsidiaries
- **media\_faturamento\_est\_coligados\_gp:** numeric value, average of estimated revenue from connected companies and their subsidiaries
- **max\_faturamento\_est\_coligados\_gp:** numeric value, maximum value of estimated revenue from connected companies and their subsidiaries
- **min\_faturamento\_est\_coligados\_gp:** numeric value, minimum value of estimated revenue from

connected companies and their subsidiaries

- **total\_filiais\_coligados**: integer value, quantity of subsidiaries from connected companies
- **media\_filiais\_coligados**: numeric value, average of subsidiaries from connected companies
- **max\_filiais\_coligados**: numeric value, maximum number of subsidiaries from connected companies
- **min\_filiais\_coligados**: numeric value, minimum number of subsidiaries from connected companies
- **qt\_coligados\_atividade\_alto**: integer value, quantity of connected companies with estimated activity level classifier equal to 'ALTO'
- **qt\_coligados\_atividade\_medio**: integer value, quantity of connected companies with estimated activity level classifier equal to 'Medio'
- **qt\_coligados\_atividade\_baixo**: integer value, quantity of connected companies with estimated activity level classifier equal to 'Baixo'
- **qt\_coligados\_atividade\_mt\_baixo**: integer value, quantity of connected companies with estimated activity level classifier equal to 'Muito Baixo'
- **qt\_coligados\_atividade\_inativo**: integer value, quantity of connected companies with estimated activity level classifier equal to 'INATIVO'
- **qt\_coligadas**: number of linked companies, i.e., companies where one of the main owners (referenced CJNPJ) has participation.
- **sum\_faturamento\_estimado\_coligadas**: the sum of all *coligadas* estimated
- **de\_faixa\_faturamento\_estimado**: character value, class of the estimated revenue of the company
- **de\_faixa\_faturamento\_estimado\_grupo**: character value, class of the sum of estimated revenue for the matrix company and branches
- **vl\_faturamento\_estimado\_aux**: numeric value, value of the estimated revenue
- **vl\_faturamento\_estimado\_grupo\_aux**: numeric value, sum of the estimated revenue for the matrix company and the branches
- **qt\_ex\_funcionarios\_cnseg**: integer value, number of employees that have left the company.
- **qt\_funcionarios\_grupo\_cnseg**: integer value, quantity of active employees considering the branches and matrix company.
- **percent\_func\_genero\_masc**: numeric value, share of male employees in the company
- **percent\_func\_genero\_fem**: numeric value, share of female employees in the company
- **idade\_ate\_18**: integer value, quantity of employees under age of 18 years old.
- **idade\_de\_19\_a\_23**: integer value, quantity of employees with age between 19 and 23 years old.
- **idade\_de\_24\_a\_28**: integer value, quantity of employees with age between 24 and 28 years old.
- **idade\_de\_29\_a\_33**: integer value, quantity of employees with age between 29 and 33 years old.
- **idade\_de\_34\_a\_38**: integer value, quantity of employees with age between 34 and 38 years old.
- **idade\_de\_39\_a\_43**: integer value, quantity of employees with age between 39 and 43 years old.
- **idade\_de\_44\_a\_48**: integer value, quantity of employees with age between 44 and 48 years old.
- **idade\_de\_49\_a\_53**: integer value, quantity of employees with age between 49 and 53 years old.
- **idade\_de\_54\_a\_58**: integer value, quantity of employees with age between 54 and 58 years old.
- **idade\_acima\_de\_58**: integer value, quantity of employees with age above
- **grau\_instrucao\_macro\_analfabeto**: integer value, quantity of employees in positions with no education level required.
- **grau\_instrucao\_macro\_escolaridade\_fundamental**: integer value, quantity of employees in positions with required fundamental level (complete and incomplete)
- **grau\_instrucao\_macro\_escolaridade\_media**: integer value, sum of grau\_instrucao\_medio\_incompl and

grau\_instrucao\_medio\_compl

- **grau\_instrucao\_macro\_escolaridade\_superior**: integer value, sum of grau\_instrucao\_sup\_incompl, grau\_instrucao\_sup\_compl, grau\_instrucao\_mestrado and grau\_instrucao\_doutorado
- **grau\_instrucao\_macro\_desconhecido**: integer value, same as grau\_instrucao\_desconhecido
- **total**: integer value, actual number of employees of the company
- **meses\_ultima\_contratacao**: numeric, months since the last hire.
- **qt\_admitidos\_12meses**: integer, quantity of employees hired in the last 12 months
- **qt\_desligados\_12meses**: integer, quantity of employees dismissed in the last 12 months
- **qt\_desligados**: integer, quantity of employees dismissed by the company
- **qt\_admitidos**: integer, quantity of employees hired by the company
- **media\_meses\_servicos\_all**: numeric, average number of months worked by all the employees of the company (active and dismissed)
- **max\_meses\_servicos\_all**: numeric, maximum number of months worked by all the employees of the company (active and dismissed)
- **min\_meses\_servicos\_all**: numeric, minimum number of months worked by all the employees of the company (active and dismissed)
- **media\_meses\_servicos**: numeric, average number of months worked by the active employees of the company
- **max\_meses\_servicos**: numeric, maximum number of months worked by the active employees of the company
- **min\_meses\_servicos**: numeric, minimum number of months worked by the active employees of the company
- **qt\_funcionarios**: integer, quantity of actual employees
- **qt\_funcionarios\_12meses**: integer, quantity of employees active 12 months before
- **qt\_funcionarios\_24meses**: integer, quantity of employees active 24 months before
- **tx\_crescimento\_12meses**: integer, growth of employees in relation to the number of employees 12 months before
- **tx\_crescimento\_24meses**: integer, growth of employees in relation to the number of employees 24 months before
- **tx\_rotatividade**: integer, measures the overall stability of employees considering hiring and dismissals and the actual number of employees.
- **qt\_filiais**: self-explanatory.