

The data analysis and modelling were done with different datasets. Below, we explain the content of each column, separated by the different dataframe names/files.

Raw data (data_PR.csv)

CNPJ_BASICO:

Company identification number (CNPJ), each company has a 14 number that uniquely identifies the company and its branch/headquarters. The “CNPJ_BASICO” identifies the company and corresponds to the first 8 digits.

CNPJ_ORDEM

Branch/headquarter identification number from Company identification number. Headquarters are assigned the number 0001, while branches are given the number associated with the branch (4 digits).

IDENTIFICADOR_MATRIZ/FILIAL

Identifies if the company is the headquarter (1) or a branch (2)

NOME_FANTASIA

The company's trade name

SITUACAO_CADASTRAL

The current company status can be:

- 01 – Null (the company is not considered valid)
- 2 – Active
- 3 – Suspended
- 4 – Inapt
- 8 - Defunct

DATA_SITUACAO_CADASTRAL

Last date in which the company status changed.

DATA_DE_INICIO_ATIVIDADE

Company's creation date

CNAE_FISCAL_PRINCIPAL

The company's economical activity is given by a CNAE code. The number is structured such that the first 2 digits describe 21 general classes of economical activities.

UF

State in which the company is located

MUNICIPIO

Town code in which the company is located

CORREIO_ELETRONICO

Email used to register the company

NATUREZA_JURIDICA

The company's legal entity type. It is given by a 4-number code

CAPITAL_SOCIAL_DA_EMPRESA

Self-declared company share capital in BRL (R\$)

PORTE_DA_EMPRESA

The company size can be:

00 – Not informed

01 – Microbusiness

03 – Small

05 – Other

Clean data**base_cnpj**

Same as CNPJ_BASICO

company_start_date

Company's creation date

company_status

Same as SITUACAO_CADASTRAL

company_status_date

Last date in which the company status changed.

town_name

Town's name in which the company is located

economical_sector

Company economic sector (agriculture, industry and service)

entity_type

Legal entity type. We consider the most common ones (Empresário Individual, Sociedade Empresaria Limitada, Sociedade Simples Limitada, Sociedade Anonima Fechada)¹ and group all others into the “other” category.

share_capital

Self-declared company share capital in BRL (R\$)

company_size

Company size (Not informed, Microbusiness, Small, Other)

Headquarters

Is it the company headquarters: (1) yes, (0) no

branch_number

Branch number of the company (the headquarters is 1)

branches_state

Number of branches in PR state

email_provider

Email provider

municipios.csv

ano

year

sigla_uf

state

id_municipio

Town's code using IBGE number

Populacao

Number of habitants in each year

municipio_code.csv

NAME_MUNICIPIO

Town's name

CODE_SEF-MG

Internal government code

MUNICIPIO

Town's code using the same number as the MUNICIPIO column