



Arquivos de entrada (neste trabalho)

1. Relatório (resultado antes do Balancete):

Relatorio_Enriquecido_Balancete (3).xlsx

Aba: Enriquecido_Balancete

Colunas (ordem típica):

- CNPJ (*formatado, ex.: 00.073.041/0001-08*)
- Nome do fundo
- CDA_Protocolo
- CDA_Competencia
- Balancete_Protocolo (*muitas linhas com “Não possui”*)
- Balancete_Competencia (*muitas linhas com “Não possui”*)

2. Protocolos da CVM (texto em uma coluna com blocos repetidos):

Protocolo Balancete CVM 062025.xlsx

Estrutura típica de um **bloco** (linhas do arquivo):

```
Protocolo de Confirmação
Status:
...
Informe:
Balancete
...
Competência:
06/01/2025
...
Participante:
NOME DO FUNDO
(46.655.287/0001-99)
...
Nº Protocolo:
SCW198675035
```

⚠ Observações importantes:

- O arquivo traz **CPF** do usuário também entre parênteses em outras linhas.
Por isso, o CNPJ deve ser capturado APENAS no bloco após Participante: .
- Nº Protocolo varia (pode aparecer N° , N° , No , N o , etc.).
O regex precisa aceitar esses símbolos.

- Competência vem como 06/01/2025 , mas no relatório você quer 06/2025 (**MM/AAAA**).