

Projeto NEFESH

A3 Data - Challenge Women

```
In [1]: from pyathena import connect
import pandas as pd
conn = connect(s3_staging_dir='s3://nefesh-raw-data/athena/', region='us-east-1')
pd.set_option('display.float_format', lambda x: '%.2f' % x)
```

Pergunta 1

Número de indústrias ativas por mês/ano entre 2010 - 2021, discriminado por MEI ou Simples, em cada município brasileiro.

```
In [2]: sql_1 = """
WITH cte_industria AS (
    SELECT distinct cnpj_basico, cnpj_ordem, cnpj_dv
    FROM nefesh_trusted.dm_cnpj_por_cnae
    WHERE lower(desc_cnae) LIKE '%fabrica%' or lower(desc_cnae) LIKE '%servico%'
)
SELECT
    COUNT(*) as qtde_industrias_ativas,
    substr(CAST(dt_inicio_atividade as varchar),1,6) as ano_mes,
    CASE
        WHEN opcao_simples='S' THEN 'simples'
        WHEN opcao_mei='S' THEN 'mei'
    END as simples_mei,
    desc_municipio
FROM nefesh_trusted.dm_empresa as dm
INNER JOIN cte_industria as cte
    ON (dm.cnpj_basico=cte.cnpj_basico AND dm.cnpj_ordem=cte.cnpj_ordem)
WHERE
    situacao_cadastral = 2
    and dt_inicio_atividade BETWEEN 20100101 AND 20210831
    and (opcao_simples='S' or opcao_mei='S')
GROUP BY
    substr(CAST(dt_inicio_atividade as varchar),1,6),
    CASE
        WHEN opcao_simples='S' THEN 'simples'
        WHEN opcao_mei='S' THEN 'mei'
    END,
    desc_municipio
ORDER BY
    substr(CAST(dt_inicio_atividade as varchar),1,6) desc,
    desc_municipio asc
"""
df_1 = pd.read_sql(sql_1, conn)
df_1
```

Out [2]:

	qtde_industrias_ativas	ano_mes	simples_mei	desc_municipio
0	1	202108	simples	ABADIA DE GOIAS
1	1	202108	simples	ABADIANIA
2	1	202108	simples	ABAETE
3	1	202108	simples	ABAETETUBA
4	2	202108	simples	ABRE CAMPO
...
280080	4	201001	simples	VITORIA DA CONQUISTA
280081	2	201001	simples	VITORIA DE SANTO ANTAO
280082	1	201001	simples	VITORINO
280083	4	201001	simples	VOLTA REDONDA
280084	3	201001	simples	XANXERE

280085 rows × 4 columns

Pergunta 2

Número de comércios que fecharam por mês/ano entre 2010 - 2021, discriminado por MEI ou Simples, em cada município brasileiro.

```
In [3]: sql_2 = """
WITH cte_comercio AS (
    SELECT distinct cnpj_basico, cnpj_ordem, cnpj_dv
    FROM nefesh_trusted.dm_cnpj_por_cnae
    WHERE lower(desc_cnae) LIKE '%comercio%' or lower(desc_cnae) LI
)
SELECT
    COUNT(*) as qtde_comercios_fechados,
    substr(CAST(dt_situacao_cadastral as varchar),1,6) as ano_mes,
    CASE
        WHEN opcao_simples='S' THEN 'simples'
        WHEN opcao_mei='S' THEN 'mei'
    END as simples_mei,
    desc_municipio
FROM nefesh_trusted.dm_empresa as dm
INNER JOIN cte_comercio as cte
    ON (dm.cnpj_basico=cte.cnpj_basico AND dm.cnpj_ordem=cte.cnpj_o
WHERE
    situacao_cadastral = 8
    and dt_situacao_cadastral BETWEEN 20100101 AND 20210831
    and (opcao_simples='S' or opcao_mei='S')
GROUP BY
    substr(CAST(dt_situacao_cadastral as varchar),1,6),
    CASE
```

```

        WHEN opcao_simples='S' THEN 'simples'
        WHEN opcao_mei='S' THEN 'mei'
    END,
    desc_municipio
ORDER BY
    substr(CAST(dt_situacao_cadastral as varchar),1,6) desc,
    desc_municipio asc
"""
df_2 = pd.read_sql(sql_2, conn)
df_2

```

Out[3]:

	qtde_comercios_fechados	ano_mes	simples_mei	desc_municipio
0	1	202108	simples	AGUDOS DO SUL
1	1	202108	simples	ALFREDO CHAVES
2	1	202108	simples	ALVORADA
3	1	202108	simples	AMERICANA
4	1	202108	simples	AMPARO
...
45900	1	201001	simples	VALE DO SOL
45901	2	201001	simples	VERANOPOLIS
45902	1	201001	simples	VICOSA
45903	2	201001	simples	VILA VELHA
45904	1	201001	simples	XANXERE

45905 rows × 4 columns

Pergunta 3

Número de CNPJ novos por mês/ano entre 2010 - 2021, discriminado por MEI ou Simples, em cada município brasileiro.

```

In [4]: sql_3 = """
SELECT
    COUNT(*) as qtde_cnpj_novos,
    substr(CAST(dt_inicio_atividade as varchar),1,6) as ano_mes,
    CASE
        WHEN opcao_simples='S' THEN 'simples'

```

```

        WHEN opcao_mei='S' THEN 'mei'
    END as simples_mei,
    desc_municipio
FROM nefesh_trusted.dm_empresa
WHERE
    situacao_cadastral = 2
    and dt_inicio_atividade BETWEEN 20100101 AND 20210831
    and (opcao_simples='S' or opcao_mei='S')
GROUP BY
    substr(CAST(dt_inicio_atividade as varchar),1,6),
    CASE
        WHEN opcao_simples='S' THEN 'simples'
        WHEN opcao_mei='S' THEN 'mei'
    END,
    desc_municipio
ORDER BY
    substr(CAST(dt_inicio_atividade as varchar),1,6) desc,
    desc_municipio asc
''''''
df_3 = pd.read_sql(sql_3, conn)
df_3

```

Out[4]:

	qtde_cnpj_novos	ano_mes	simples_mei	desc_municipio
0	11	202108	simples	ABADIA DE GOIAS
1	2	202108	simples	ABADIA DOS DOURADOS
2	4	202108	simples	ABADIANIA
3	16	202108	simples	ABAETE
4	21	202108	simples	ABAETETUBA
...
582881	13	201001	simples	XANXERE
582882	2	201001	simples	XAPURI
582883	5	201001	simples	XAXIM
582884	1	201001	simples	XINGUARA
582885	1	201001	simples	XIQUE-XIQUE

582886 rows × 4 columns

Pergunta 4

Qual o número de CNPJ que surgiram do grupo de educação superior, entre 2015 - 2021, discriminado por ano, em cada estado brasileiro?

```
In [5]: sql_4 = """
WITH cte_ensino_superior AS (
    SELECT cnpj_basico, cnpj_ordem, cnpj_dv
    FROM nefesh_trusted.dm_cnpj_por_cnae
    WHERE lower(desc_cnae) LIKE '%educação superior%' or lower(desc
)
SELECT
    COUNT(*) AS qtde_cnpj_educacao_superior,
    substr(CAST(dt_inicio_atividade as varchar),1,4) as ano,
    uf
FROM nefesh_trusted.dm_empresa_cnpj
```

```

FROM nefesh_trusted.dm_empresa as dm
INNER JOIN cte_ensino_superior as cte
    ON (dm.cnpj_basico=cte.cnpj_basico AND dm.cnpj_ordem=cte.cnpj_o
WHERE
    dt_inicio_atividade BETWEEN 20150101 AND 20210831
GROUP BY
    substr(CAST(dt_inicio_atividade as varchar),1,4),
    uf
ORDER BY
    substr(CAST(dt_inicio_atividade as varchar),1,4) desc,
    uf asc
"""
df_4 = pd.read_sql(sql_4, conn)
df_4

```

Out[5]:

	qtde_cnpj_educacao_superior	ano	uf
0	17	2021	AC
1	19	2021	AL
2	37	2021	AM
3	4	2021	AP
4	82	2021	BA
...
188	85	2015	RS
189	31	2015	SC
190	16	2015	SE
191	319	2015	SP
192	28	2015	TO

193 rows × 3 columns

Pergunta 5

Qual a classe de CNAE com maior capital social médio no Brasil no último ano?

```

In [6]: sql_5 = """
        SELECT
            AVG(capital_social) as media_capital_social,
            cn.cnae,
            cn.desc_cnae,
            substr(CAST(cn.dt_inicio_atividade as varchar), 1, 4) as ano

```

```

        substr(CAST(emp.dt_inicio_atividade as varchar),1,4) as ano
FROM nefesh_trusted.dm_cnpj_por_cnae as cn
INNER JOIN nefesh_trusted.dm_empresa as emp
    ON (cn.cnpj_basico=emp.cnpj_basico AND cn.cnpj_ordem=emp.cn
WHERE
    substr(CAST(emp.dt_inicio_atividade as varchar),1,4)='2020'
GROUP By
    cn.cnae, cn.desc_cnae, substr(CAST(emp.dt_inicio_atividade
ORDER BY AVG(capital_social) DESC
"""
df_5 = pd.read_sql(sql_5, conn)
df_5

```

Out[6]:

	media_capital_social	cnae	desc_cnae	ano
0	42184951432.92	6422100	Bancos múltiplos, com carteira comercial	2020
1	35441956461.62	6423900	Caixas econômicas	2020
2	17272286737.30	600001	Extração de petróleo e gás natural	2020
3	13001771852.84	710301	Extração de minério de ferro	2020
4	12653929673.09	1921700	Fabricação de produtos do refino de petróleo	2020
...
1309	0.00	210102	Cultivo de acácia-negra	2020
1310	0.00	112102	Cultivo de juta	2020
1311	0.00	6470101	Fundos de investimento, exceto previdenciários...	2020
1312	0.00	159804	Criação de bicho-da-seda	2020
1313	0.00	6499901	Clubes de investimento	2020

1314 rows × 4 columns

Pergunta 6

Qual a média do capital social das empresas de pequeno porte por natureza jurídica no último ano?

```

In [7]: sql_6 = """
        SELECT
            AVG(capital_social) as media_capital_social,
            natureza_juridica,
            desc_natureza_juridica,
            substr(CAST(dt_inicio_atividade as varchar),1,4) as ano
        FROM nefesh_trusted.dm_empresa

```

```

FROM nefesh_trusted.ann_empresa
WHERE
    substr(CAST(dt_inicio_atividade as varchar),1,4)='2020'
    AND porte_empresa=3
GROUP By
    natureza_juridica, desc_natureza_juridica, substr(CAST(dt_i
ORDER BY AVG(capital_social) DESC
"""
df_6 = pd.read_sql(sql_6, conn)
df_6

```

Out[7]:

	media_capital_social	natureza_juridica	desc_natureza_juridica	ano
0	9902060.00	2054	Sociedade Anônima Fechada	2020
1	1001980.00	2070	Sociedade Empresária em Nome Coletivo	2020
2	258749.81	2062	Sociedade Empresária Limitada	2020
3	240020.81	2305	Empresa Individual de Responsabilidade Limitad...	2020
4	163421.71	2313	Empresa Individual de Responsabilidade Limitad...	2020
5	91605.84	2240	Sociedade Simples Limitada	2020
6	54735.14	2135	Empresário (Individual)	2020
7	37842.76	2321	Sociedade Unipessoal de Advocacia	2020
8	24988.87	2232	Sociedade Simples Pura	2020
9	20000.00	2330	Cooperativas de Consumo	2020

In []: