

# Detalhes da análise - PARTE 1

## Tabela base1

Total linhas: 476.615

Total linhas únicas (cpf\_anonimo, primeiro\_nome): 479.164

Contém a renda por indivíduo

Percebi que há duplicidade nas informações de salários para um indivíduo. Alguns exemplos são mostrados na figura abaixo:

```
WITH duplicados AS (
  SELECT cpf_anonimo, primeiro_nome
  FROM base1
  GROUP BY cpf_anonimo, primeiro_nome
  HAVING COUNT(*) > 1 )
SELECT b.* FROM base1 b
JOIN duplicados d
ON b.cpf_anonimo = d.cpf_anonimo
AND b.primeiro_nome = d.primeiro_nome
ORDER BY b.cpf_anonimo, b.primeiro_nome;
```

A-Z cpf anonimo	A-Z primeiro nome	A-Z renda
***.009.243-**	Isabella	3283
***.009.243-**	Isabella	5750
***.010.375-**	Tomas	1995
***.010.375-**	Tomas	6999
***.010.591-**	Luiz Gustavo	561
***.010.591-**	Luiz Gustavo	8839
***.012.587-**	Diogo	6648
***.012.587-**	Diogo	4871
***.014.473-**	Nick	5625

Uma prática recomendada ao coletar dados é sempre incluir um carimbo de data e hora. Dado que não tenho acesso à data de coleta, eu optei por excluir essas linhas duvidosas, visto que representam menos de 0,1% do total de linhas.

## Tabela base2

Total linhas: 660.912

Total linhas únicas (cpf\_anonimo, segundo\_nome): 660.141

A tabela base2 apresenta cpf\_anonimo, ultimo\_nome, uf e orgsup\_lotacao\_instituidor\_pensao.

A última coluna refere-se a qual ministério origina-se o benefício. Aqui percebo o mesmo problema com relação a duplicações e não há nenhum modo lógico de decidir entre um ministério ou outro, como observa-se na tabela abaixo o caso de Alves Vieira que ora pode pertencer ao ministério da Educação, ora pode pertencer ao ministério da Economia. Optei por remover essas informações duvidosas, visto que representavam menos de 0,03% das linhas.

```

WITH duplicados AS (
    SELECT cpf_anonimo, ultimo_nome from base2
    GROUP BY cpf_anonimo, ultimo_nome having
    COUNT(*) > 1 )
SELECT b.* from base2 b
JOIN duplicados d on b.cpf_anonimo =
d.cpf_anonimo
    AND b.ultimo_nome = d.ultimo_nome
ORDER BY b.cpf_anonimo, b.ultimo_nome;

```

Az cpf anonimo	Az ultimo nome	Az uf	Az orgsup lotacao instituidor pensao
***.001.217-**	Alves Vieira	RR	Minist�rio da Educa�o
***.001.217-**	Alves Vieira	PE	Minist�rio da Economia
***.002.489-**	Amorim	AL	Minist�rio da Economia
***.002.489-**	Amorim	AL	Minist�rio da Economia
***.004.588-**	Araujo Rofrano	MT	Minist�rio da Economia
***.004.588-**	Araujo Rofrano	PB	Minist�rio da Economia
***.006.712-**	Arantes Nazario	RO	Sem informa�o

## Encoding

O arquivos vieram em encodings windows-1251. Isso impactou a base2.

## Tabela base3

Total linhas: 45.648

Total linhas  nicas (cpf): 45.648

Essa tabela cont m cpf, nome\_completo, email, uf e dt\_nascimento.

An lise a partir das tr s tabelas

A partir da observa o dos dados nas tr s tabelas:

		# de linhas distintas	
base1	cpf x renda	479.164	
base2	cpf x origem_renda x uf	660.912	
base3	cpf x uf x dt_nascimento	45.648	

## Correla o das tr s tabelas

Ao implementar a query abaixo, mesmo tendo eliminado duplicatas nas tabelas individuais, ainda percebemos que alguns indiv duos est o duplicados. Um exemplo   Lucca que tem duas linhas com sobrenomes diferentes e por isso foi identificado como  nico na primeira an lise.

Geração da tabela fulldataTotal de linhas: 17.419

Linhas únicas:

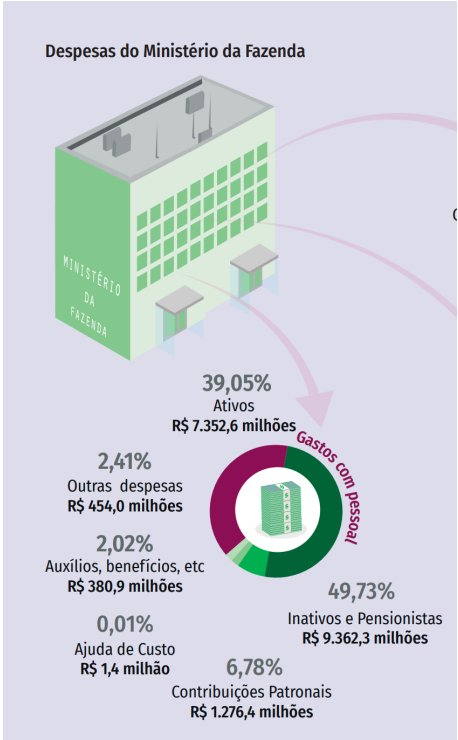
```
drop table fulldata;
create table fulldata as
with base1_distinct as (
    select
        cpf_anonimo,
        primeiro_nome,
        renda
    from
        base1
    group by
        cpf_anonimo,
        primeiro_nome,
        renda
    having
        count(*) = 1
base2_distinct as (
    select
        cpf_anonimo,
        ultimo_nome
    from
        base2
    group by
        cpf_anonimo,
        ultimo_nome
    having
        count(*) = 1
)
select distinct on (b3.cpf)
    b1.cpf_anonimo as cpf_anonimo_base1,
    b1.primeiro_nome,
    b1.renda,
    b2.ultimo_nome,
    b3.cpf,
    b3.nome_completo,
    b3.email,
    b3.uf,
    b3.dt_nascimento
from
    base3 b3
inner join
    base1_distinct b1
on
    substring(b1.cpf_anonimo, 5, 3) = substring(b3.cpf, 4, 3)
and substring(b1.cpf_anonimo, 9, 3) = substring(b3.cpf, 7, 3)
and b1.primeiro_nome = split_part(b3.nome_completo, ' ', 1) -- correlaciona com base3
```

```
inner join
  base2_distinct b2
on
  b1.cpf_anonimo = b2.cpf_anonimo
order by
  b3.cpf, b1.primeiro_nome, b2.ultimo_nome;
```

**As análises são exploratórias e estão descritas em detalhes no README.md do repositório em : <https://github.com/esthercamilo/pan>**

### Estimativa dos gastos públicos com base em relatórios anuais

Para a presente análise extraí alguns dados de relatórios do TCU apenas para fins de estimativa. Com disponibilidade de tempo, o melhor modo a se obter dados precisos e atualizados é através da API do governo federal portal da transparência.

 <p><b>Despesas do Ministério da Fazenda</b></p> <p><b>39,05%</b> Ativos R\$ 7.352,6 milhões</p> <p><b>2,41%</b> Outras despesas R\$ 454,0 milhões</p> <p><b>2,02%</b> Auxílios, benefícios, etc R\$ 380,9 milhões</p> <p><b>0,01%</b> Ajuda de Custo R\$ 1,4 milhão</p> <p><b>6,78%</b> Contribuições Patronais R\$ 1.276,4 milhões</p> <p><b>49,73%</b> Inativos e Pensionistas R\$ 9.362,3 milhões</p> <p><b>Gastos com pessoal</b></p>	<p>Estima-se que 31 bilhões de reais sejam gastos com pensionistas somente para o ministério da Fazenda (atual ministério da Economia).</p> <p>Na presente base, a renda movimentada é de aproximadamente 17 bilhões, ou seja, mesmo para uma das maiores bases do banco ainda há potencial de expansão.</p> <p>Pensionistas do ministério da saúde</p>

## Detalhes da análise - PARTE 2

IDs são únicos

Dados

- Custo médio por minuto de atendimento: R\$9,00/min . contato
- Custo de manutenção sistêmica: R\$ 600 a cada 400 contatos (deve ser rateado proporcionalmente entre os produtos)
- TMA (tempo médio de atendimento) dos contatos feitos em Banking (voz e chat): 360 segundos / contato
- CR (contact rate)(1) - Conta corrente: 6%
- CR (contact rate)(1) - Conta corrente + Cartão: 15%
- Receita líquida gerada por cliente – Conta: R\$3/mês
- Receita líquida gerada por cliente – Conta + Cartão: R\$25/mês

### Responder às questões

**Qual o custo mensal da operação de Atendimento desse Banco, no segmento de Banking?**

```
{
  "canal": {
    "Chat": 1000,
    "Voz": 500
  },
  "segmento": {
    "Conta + Cartão": 1200,
    "Conta": 300
  },
  "motivo": {
    "Cartão fraudado": 514,
    "PIX inoperante": 452,
    "Acesso ao app": 196,
    "Cartão perdido": 191,
    "Cartão roubado": 147
  }
}
```

### **Análise para os dados de 1500 coletados no mês de dezembro de 2023**

Custo com atendimento:  $6 \cdot 9 \cdot 1500 = \text{R\$ } 81000$

Total de tentativas para 300 cc de sucessos:  $300/0,06 = 5000$

Total de tentativas para 1200 cc + cartão:  $1200/0,15 = 8000$

Manutenção sistêmica para 13000 contatos =  $(600/400) \cdot 13000 = \text{R\$ } 19500$

Receita mensal gerada para 300 cc =  $3 \cdot 300 = \text{R\$ } 900$

Receita mensal gerada para 1200 cc + cartão =  $25 \cdot 1200 = \text{R\$ } 30000$

Custos: R\$ 100500,00

Receita: R\$ 30900,00

#### **Principais ações urgentes para mitigação dos prejuízos**

1. Capacitar a equipe de atendimento para reduzir o tempo de atendimento
2. Identificar os motivos dos contatos e implementar ações preventivas.
3. Oferecer suportes múltiplos (chats inteligentes, redes sociais)

#### **Mais ações para melhoria contínua e aumento do lucro**

1. Educação do cliente
2. Aprimorar a experiência do cliente - colher feedback
3. Renegociar contratos com fornecedores. Diminuir custos com manutenção sistêmica.
4. Otimizar a carga de trabalho dos agentes.

**Qual a rentabilidade mensal de cada uma das operações do Banco, considerando a receita gerada e os custos em questão? (não é necessário considerar nenhuma informação diferente das já fornecidas)**

Rentabilidade cc = **-20300**

Receita mensal gerada para 300 cc =  $3 \cdot 300 = \text{R\$ } 900,00$

Custo com atendimento:  $6 \cdot 9 \cdot 300 = \text{R\$ } 16200,00$

Total de tentativas para 300 cc de sucessos:  $300/0,06 = 5000$

Rentabilidade cc + cartão = **-42800**

Receita mensal gerada para 1200 cc + cartão =  $25 \cdot 1200 = \text{R\$ } 30000$

Custo com atendimento:  $6 \cdot 9 \cdot 1200 = \text{R\$ } 64800$

Total de tentativas para 1200 cc + cartão:  $1200/0,15 = 8000$

## Links

<https://2cyr.com/decode/?lang=en>

## Slides

[https://docs.google.com/presentation/d/1dtY1Q0AFogYQcKOdsx\\_1vfYqsNciw6l1L0UVOhf025o/edit?usp=sharing](https://docs.google.com/presentation/d/1dtY1Q0AFogYQcKOdsx_1vfYqsNciw6l1L0UVOhf025o/edit?usp=sharing)

## Este documento

<https://docs.google.com/document/d/1NFJQTJs7HWv-L0BOd8aqx0suuemOgdkmPSgNKBCXi8Q/edit?usp=sharing>

## Gihub

<https://github.com/esthercamilo/pan>