



PROJETO FINAL DELOITTE DATA LITERACY

Squad 5
Março de 2022

AGENDA

- Visão Geral
- Análises
 - Base I – Births.csv – Análise em Python
 - Base I – Births.csv – Análise em PowerBI
 - Base II – Brasil.csv – Análise em Python
 - Base II – Brasil.csv – Análise em PowerBI



VISÃO GERAL

VISÃO GERAL

Equipe – Squad 5

Ana Paula Oliveira Alves

Gerente | Business Tax

Deloitte Touche Tohmatsu

+ 55 31 9 9845-5259

anapaalves@deloitte.com



Jaqueline de Souza Bandeira

Senior | TAX VAT

Deloitte Touche Tohmatsu

+ 55 21 3981-0546

jbandeira@deloitte.com



Keny Wassano

Senior | TAX Mobility, Payroll, Rewards

Deloitte Touche Tohmatsu

kwassano@deloitte.com

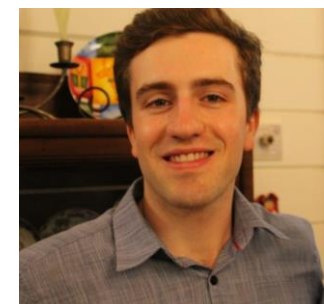


Lucas Horvath Campana

Pleno | Audit Delivery Center

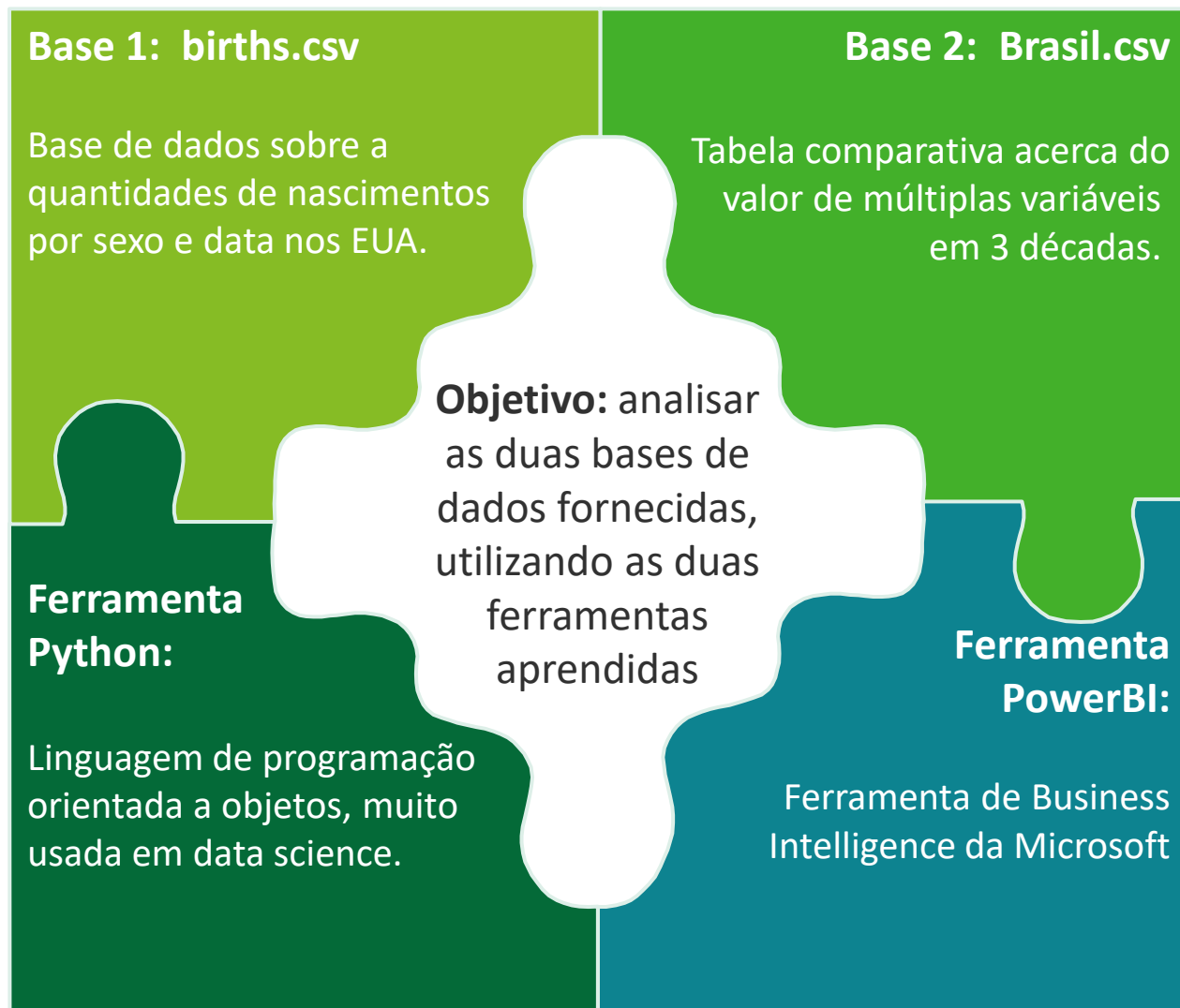
Deloitte Touche Tohmatsu

lucampana@deloitte.com



VISÃO GERAL

Projeto



Divisão das atividades

Base I – Births.csv

Atividade	Responsável
Análise em Python	Ana Alves
Análise em PowerBI	Jaqueline Bandeira

Base II – Brasil.csv

Análise em Python	Lucas Campana
Análise em PowerBI	Kenny Wassano

VISÃO GERAL

Kanban : Ferramenta de gestão utilizada - Zehnk

The screenshot displays the Zehnk Kanban board interface. The left sidebar contains the Zehnk logo, a 'Criar' button, and navigation links for 'Visão geral', 'Tarefa', 'Conteúdo', 'Conversa', 'Pesquisar workspaces...', and a list of workspaces including 'Consultoria Tributaria - RJ', 'SB_GCR_BR', 'TAX_BPS_RJ', and 'Sem coleção'. The main workspace is 'Data Literacy' (Sem coleção), dated 18/03/22. The board is organized into three columns: 'Não iniciada (1)', 'Em progresso (4)', and 'Concluída (4)'. The 'Não iniciada' column contains one task: 'Validar a documentação do processo' (ID 005, due 17/03/22). The 'Em progresso' column contains four tasks: 'Análise dos dados_BasePDF' (ID 007), 'Análise dos dados_BaseCSV' (ID 003), 'Elaborar os dashboards_BaseCSV' (ID 008), and 'Elaborar os dashboards_BasePDF' (ID 011). The 'Concluída' column contains four tasks: 'Limpar os arquivos_BaseCSV' (ID 002), 'Divisão das atividades' (ID 001), 'Limpar os arquivos_Base pdf' (ID 010), and 'Material suporte| Aulas PY' (ID 009). Each task card shows its ID, due date, progress status, and assigned users. The bottom of the board features a search bar for tasks and a 'Fale conosco' button.

ZEHNK

Data Literacy (Sem coleção)

18/03/22

TAREFAS ANEXOS CONVERSAÇÃO MAIS

Arquivadas Filtrar

Não iniciada (1)

Validar a documentação do processo

005 17/03/22 0/1 ✓ 3

Em progresso (4)

Análise dos dados_BasePDF

007 3 0/6 ✓ 1

Análise dos dados_BaseCSV

003 0/6 ✓

Elaborar os dashboards_BaseCSV

008 1 3/6 ✓ 2

Elaborar os dashboards_BasePDF

011

Concluída (4)

Limpar os arquivos_BaseCSV

002 6

Divisão das atividades

001 1

Limpar os arquivos_Base pdf

010

Material suporte| Aulas PY

009

Digite uma tarefa e aperte ENTER

Fale conosco

ANÁLISES

Base I – Births.csv – Análise em Python

ANÁLISES

Base I – Births.csv – Análise em Python

GitHub Link

CARGA DA BASE DE DADOS

	year	month	day	gender	births
0	1969	1	1.0	F	4046
1	1969	1	1.0	M	4440
2	1969	1	2.0	F	4454

LIMPEZA E PREPARAÇÃO DA
BASE

- Exclusão dos valores nulos;
- Exclusão de valores inválidos (ex.: '99' na coluna do dia);
- Exclusão de dias inexistentes (ex.: 31 de fevereiro).

REALIZAÇÃO DAS ANÁLISES

- 1) Qual é o sexo biológico que nasce mais a cada década passada?
- 2) Em quais dias da semana há mais nascimentos?
- 3) Há alguma relação entre o dia da semana e possíveis natalidades?
- 4) Relação de nascimentos por mês e por gênero.
- 5) Relação de nascimentos por mês e por década.

ANÁLISES

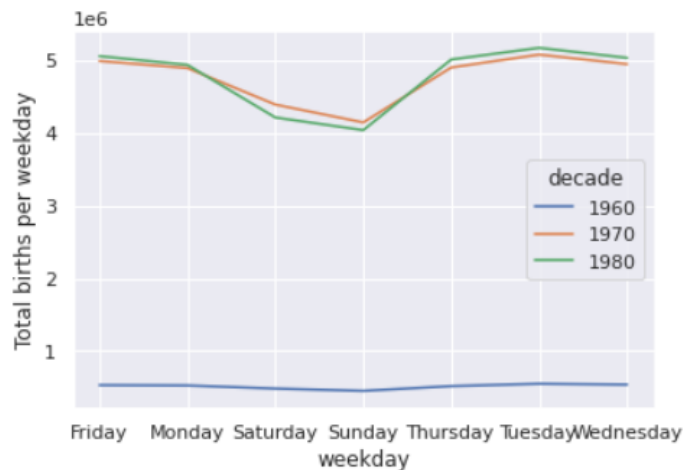
Base I – Births.csv – Análise em Python

Exemplo de análise efetuada

3) Há alguma relação entre o dia da semana e possíveis natalidades?

```
In [16]: df_auxiliar1['decade'] = 10 * (df_auxiliar1['year'] // 10)
df_auxiliar1.pivot_table('births', index='weekday', columns='decade', aggfunc='sum')
print(df_auxiliar1.head())
```

```
In [17]: sns.set()
df_auxiliar1_weekday = df_auxiliar1.pivot_table('births', index='weekday', columns='decade', aggfunc='sum')
df_auxiliar1_weekday.plot()
plt.ylabel("Total births per weekday")
plt.show()
```



ANÁLISES

Base I – Births.csv – Análise em Python

Principal dificuldade encontrada

EXPECTATIVA

Inclusão de dias da semana

In [10]:

```
df_auxiliar1 = pd.DataFrame(df_auxiliar, columns = ['year', 'month', 'day', 'gender', 'births', 'weekday'])

for i in range(0, len(df_auxiliar)):
    df_auxiliar1['weekday'].iloc[i] = (dt.date(df_auxiliar['year'].iloc[i], df_auxiliar['month'].iloc[i], df_auxiliar['day'].iloc[i].astype(int)).strftime('%A'))
df_auxiliar1
```

ANÁLISES

Base I – Births.csv – Análise em Python

Principal dificuldade encontrada

REALIDADE

```
In [9]: # Criação do dicionário que indica quantos dias tem cada mês
dict_help = {1:31,
2:28,
3:31,
4:30,
5:31,
6:30,
7:31,
8:31,
9:30,
10:31,
11:30,
12:31}

df_teste = df_cleanBase
# Criação de um novo dataframe, em que os dias que não são válidos para cada mês não são incluídos
df_auxiliar = pd.DataFrame()
print('Aguarde estamos processando...')
# Para cada linha do dataframe
for i in range(0, len(df_teste)):
    print(f"iteração {i} de {len(df_teste)} - {(i/len(df_teste))*100} %")
    # Trago o mês daquela linha
    int_month = df_teste.iloc[[i]]['month'][i]
    # Trago o dia daquela linha
    int_day = df_teste.iloc[[i]]['day'][i]
    # Se aquele dia é menor ou igual do que o dia máximo daquele mês convencionado
    if int_day <= dict_help[int_month]:
        # Aquela linha é adicionada ao novo dataframe, pois aquela data é válida
        df_auxiliar = df_auxiliar.append(df_teste.iloc[i])

# convertendo para inteiro os valores armazenados
df_auxiliar['year'] = df_auxiliar['year'].astype(int)
df_auxiliar['month'] = df_auxiliar['month'].astype(int)
df_auxiliar['day'] = df_auxiliar['day'].astype(int)
df_auxiliar
```

ANÁLISES

Base I – Births.csv – Análise em PowerBI

ANÁLISES

PERCEPÇÃO (DIFICULDADES versus ASSERTIVIDADE)

ASSERTIVIDADE

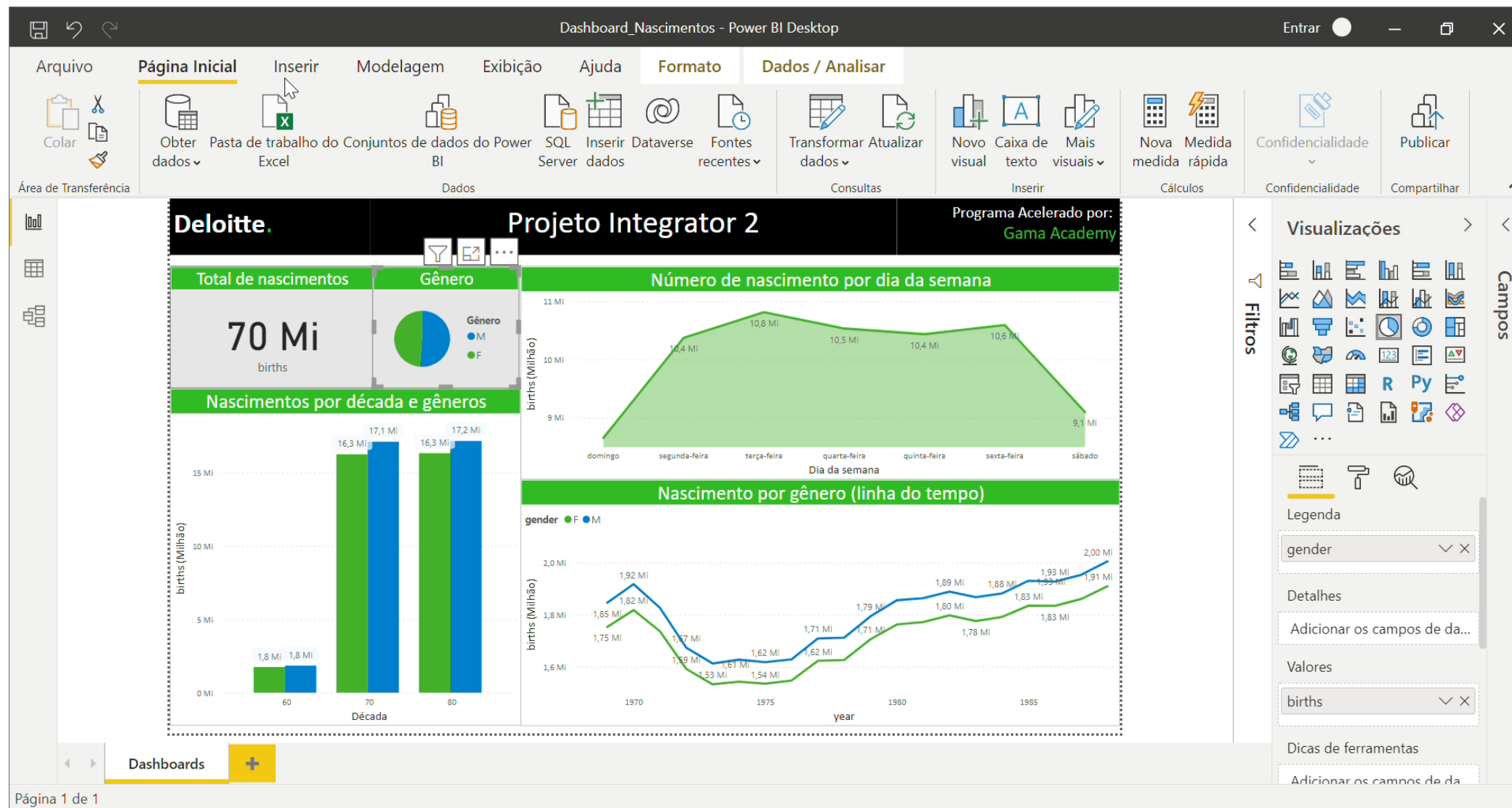
- As aulas de *powerbi* foram bem completas, logo não encontramos dificuldades na elaboração dos dashboards.

DIFICULDADES

- Como o arquivo para elaboração dos dashboards foi retirado da internet para análises com python, ao efetuarmos as limpezas dos dados e gerarmos os dashboards perdemos a visualização histórica dos anos (de 39 anos para 19 anos após limpeza);
- Criar um arquivo .json com a paleta de cores da Deloitte.

ANÁLISES

Dahsboards – vídeo demonstrativo

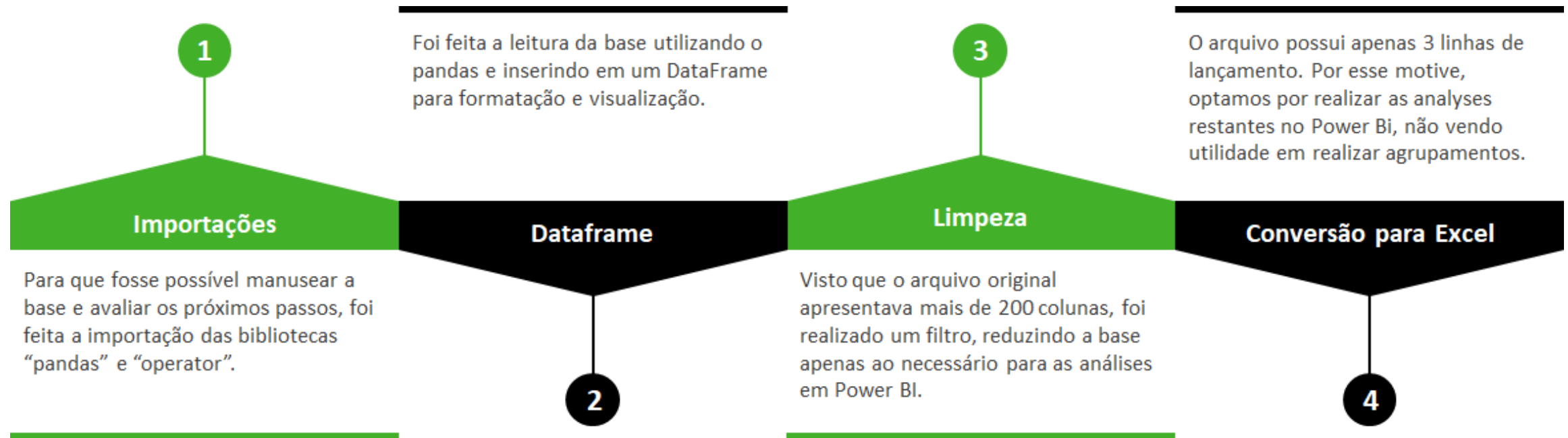


ANÁLISES

Base II – Brasil.csv – Análise em Python

ANÁLISES

Base II – Brasil.csv – Análise em Python



ANÁLISES

Base II – Brasil.csv – Análise em Python

1

Importações

```
1 # Importação das Bibliotecas
2 ✓ from operator import index
3 import pandas as pd
4
5 # Lendo o arquivo csv e armazenando uma
  variável
```

2

DataFrame

```
5 # Lendo o arquivo csv e armazenando uma
  variável
6
7 brasil_base = pd.read_csv('brasil.csv')
8
9 # Transformando arquivo em DataFrame
10
11 brasil_df = pd.DataFrame(brasil_base)
12
```

3

Limpeza

```
9 # Transformando arquivo em DataFrame
10
11 brasil_df = pd.DataFrame(brasil_base)
12
13 # Reduzindo o número de colunas para
  aquelas essenciais à análise
14
15 brasil_reuzido = brasil_df[['ano',
16 'expectativa_vida', 'fecundidade_total',
17 'mortalidade_1', 'mortalidade_5',
18 'prob_sobrevivencia_40',
19 'prob_sobrevivencia_60',
20 'prop_pobreza_extrema', 'renda_pc',
21 'populacao_urbana', 'populacao_rural',
22 'indice_escolaridade', 'idhm']]
```

4

Conversão para Excel

```
17 # Salvando o arquivo limpo em formato
  Excel
18
19 brasil_reuzido.to_excel('brasil_reuzido.
  xlsx')
20
21 print(brasil_reuzido)
22
```

ANÁLISES

Base II – Brasil.csv – Análise em PowerBI

ANÁLISES

Base II – Brasil.csv – Análise em PowerBI

Deloitte.

Projeto Integrator 2

Análise da variação da expectativa de vida em 3 décadas

Programa Acelerado por:

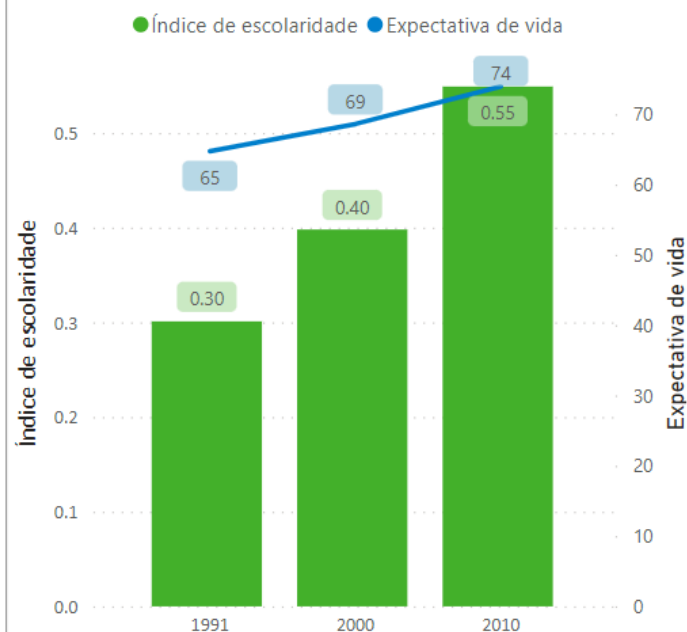
Gama Academy

Expectativa de Vida (Anos)

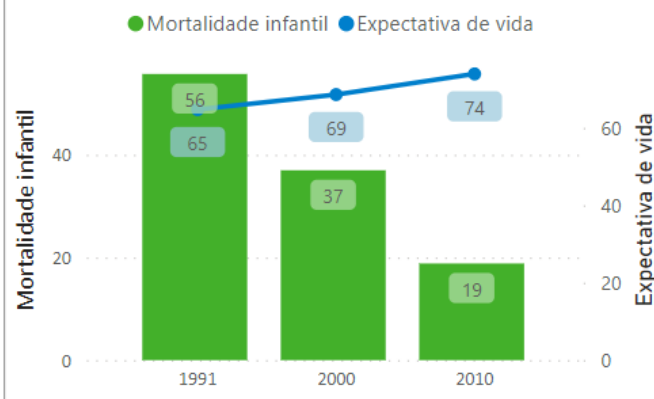
69.09

Average of expectativa_vida

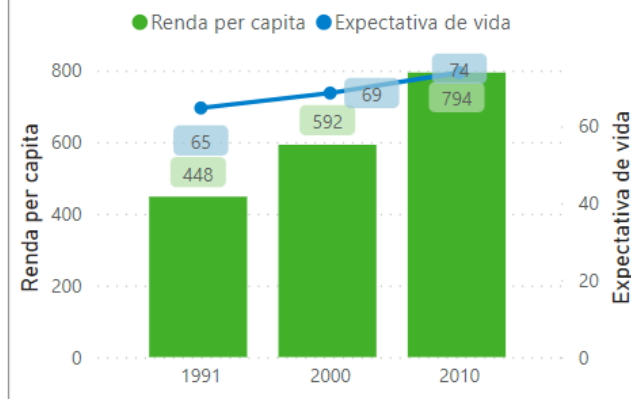
Expectativa de vida por escolaridade



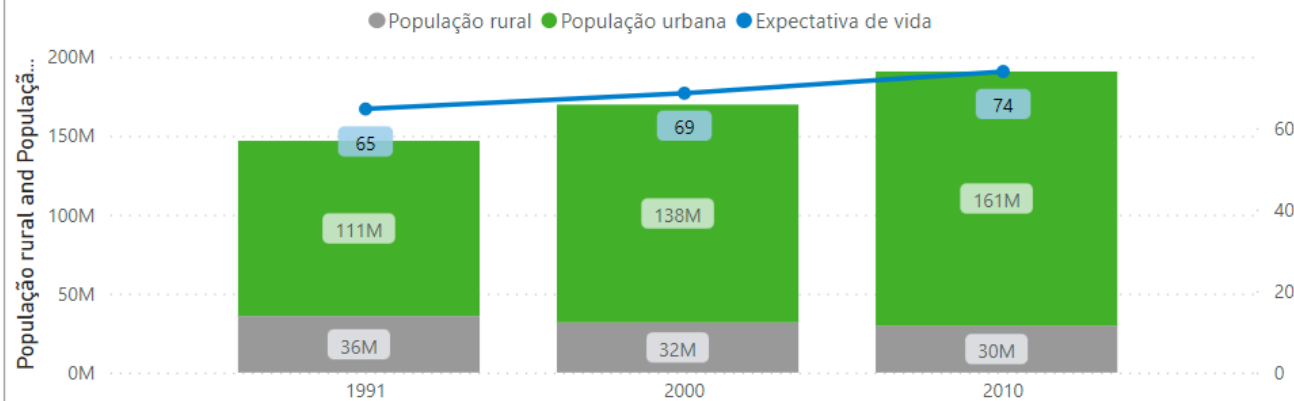
Taxa de mortalidade infantil em relação à expectativa de vida



Renda per capita em relação à expectativa de vida

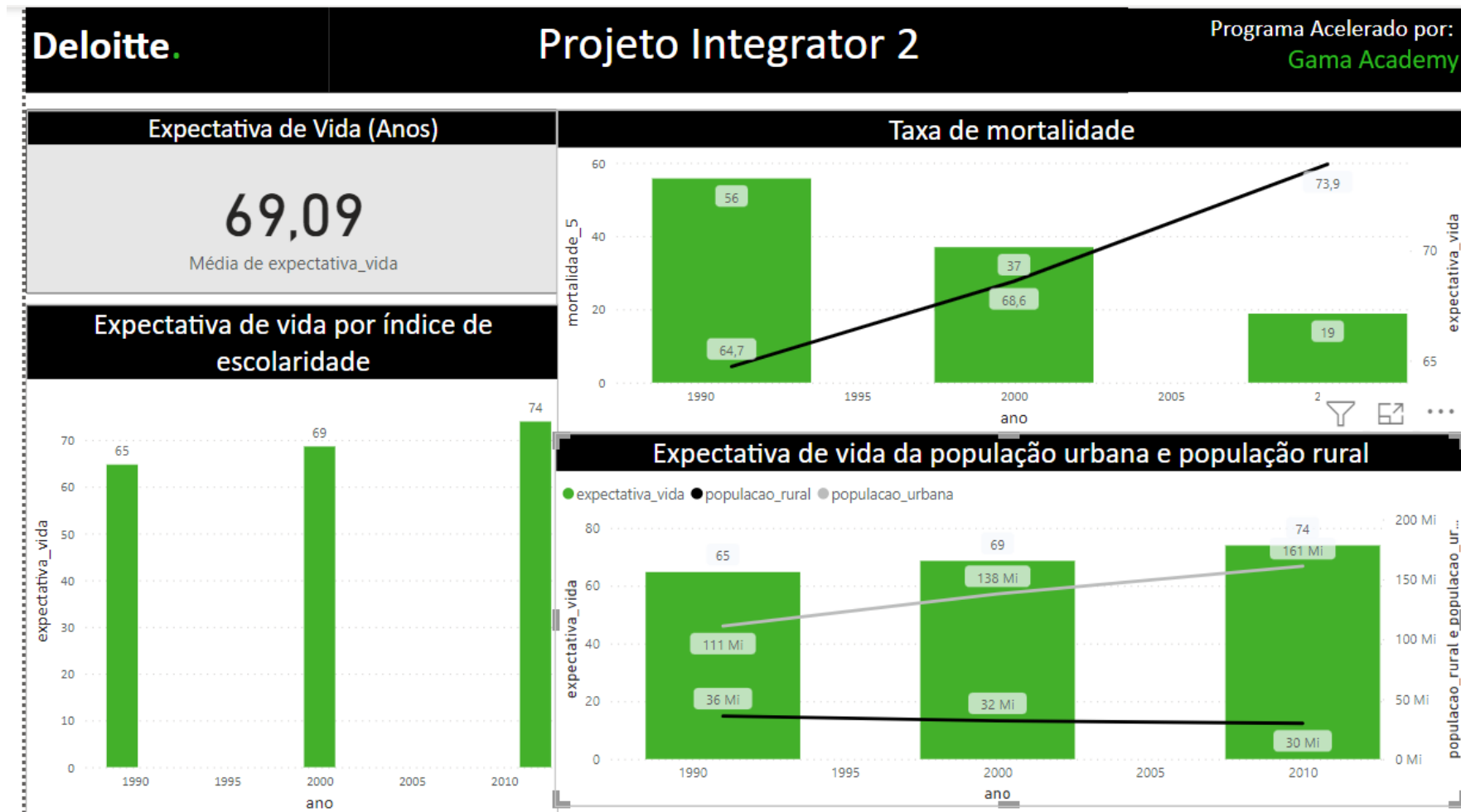


Proporção da população em relação à expectativa de vida



ANÁLISES

Base II – Brasil.csv – Análise em PowerBI



OBRIGADO!