Logotipo, nome da empresa

Descrição gerada automaticamente

Pós-graduação

Cientista de Dados – 2021

Cientista de Dados nas Organizações

Coleta e armazenamento de dados

Atividade de Aprofundamento Trilha 6

“Conhecendo Banco de Dados **NoSQL**”

Aluno: Denyson Tomaz de Lima

Matr. 92174337

# Índice

[1. Índice 2](#_Toc89679011)

[2. Introdução 3](#_Toc89679012)

[3. Tarefa 3](#_Toc89679013)

[4. Resultado: 9](#_Toc89679014)

# Introdução

*Trabalho referente a trilha 6:*

Agora já conhecemos os bancos de dados NoSQL e as suas características. Vamos então trabalhar estes conceitos na nossa última atividade de aprofundamento. Selecione um banco de dados NoSQL de sua preferência. Crie uma instância de teste e faça a inserção de dados a partir de um arquivo csv.

Utilize a base de dados World Happiness disponível no site do Kaggle:

<https://www.kaggle.com/ajaypalsinghlo/world-happiness-report-2021>

Apresente um notebook com o código utilizado e uma evidência do cadastro realizado.

Obs.: Alguns bancos de dados possuem área de sandbox disponível em nuvem. Se você tiver alguma dificuldade em baixar o banco, utilize a opção em nuvem.

Todos os arquivos estão disponíveis no github

Site: <https://github.com/DenysonLima/ColArmDados>

# Tarefa

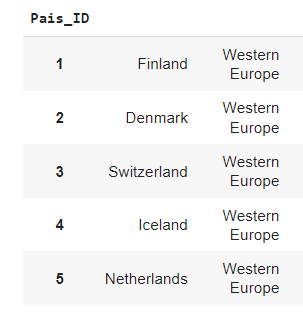
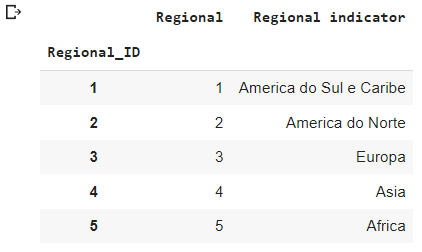
Através do site do Kaggle, foi feito o download do arquivo:

world-happiness-report-2021.csv.

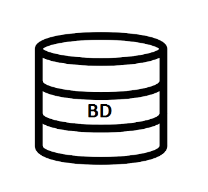
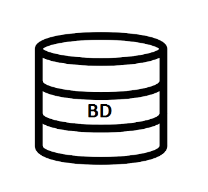
Este arquivo serviu de base do projeto desta tarefa para ser carregado no Banco de dados NoSQL no site: <https://sandbox.neo4j.com/>.

Antes de inserir o arquivo no Banco de dados no Neo4j, foi feito uma preparação e atualização do Banco de dados world-happiness-report-2021.csv, anexando duas novas colunas.

No arquivo world-happiness-report-2021.csv foi acrescentado uma coluna que se refere ao continente e a região na qual apelidamos de Regional\_ID. Esta coluna contém um número de 1 a 6 que representa a região do continente, identificado como:



**Regional\_ID Regional**



***Arquivo: world-happiness-report-2021.csv Arquivo: regional\_ID.csv***

Foi criado um arquivo chamado de regional\_ID que contém a associação das regiões continentais (1 a 6).

O arquivo world-happiness-report-2021.csv foi preparado para receber a nova coluna (Regional\_ID) identificando a região. Foi aplicado o programa PYTHON para realizar esta inserção da nova coluna, veja abaixo o programa em PYTHON.

**Iniciação:**

import pandas as pd

import numpy as np

mundo\_feliz = pd.read\_csv('world-happiness-report-2021.csv')

**Criando uma nova coluna que reflete a da tabela da região continental (America, Europa, Asia, Africa e Estados Independentes)**

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'Latin America and Caribbean', 'Regional\_ID'] = '1'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'North America and ANZ', 'Regional\_ID'] = '2'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'Central and Eastern Europe', 'Regional\_ID'] = '3'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'Western Europe', 'Regional\_ID'] = '3'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'East Asia', 'Regional\_ID'] = '4'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'Southeast Asia', 'Regional\_ID'] = '4'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'South Asia', 'Regional\_ID'] = '4'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'Sub-Saharan Africa', 'Regional\_ID'] = '5'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'Middle East and North Africa', 'Regional\_ID'] = '5'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'Commonwealth of Independent States', 'Regional\_ID'] = '6'

mundo\_feliz.index = mundo\_feliz.index + 1

mundo\_feliz.index.names = ['Pais\_ID']

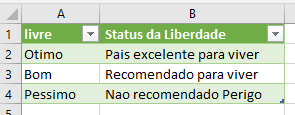
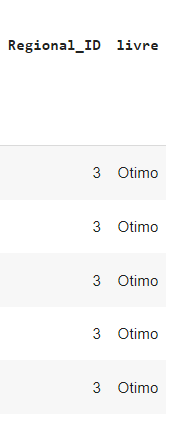
mundo\_feliz.to\_csv('mundo\_feliz\_2021.csv')

mundo\_feliz.head()

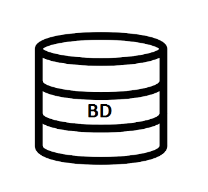
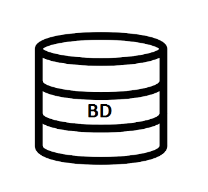
**Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente**

No arquivo world-happiness-report-2021.csv também foi acrescentado outra coluna que se refere uma indexação de liberdade de livre escolha do povo. Foi mapeado como: Péssimo, Bom e Ótimo, apenas para manipular o Banco de dados, sem objetivo de realizar uma pesquisa série com dados relevante. O intuito modesto tarefa é de apenas manipular e visualizar o Banco e dados através do NoSQL, aqui representado pelo Neo4j.



**livre livre**



***Arquivo: world-happiness-report-2021.csv Arquivo: livre.csv***

O arquivo world-happiness-report-2021.csv revisão A foi preparado para receber também uma nova coluna (livre) identificando a satisfação da liberdade do povo. Foi aplicado o programa PYTHON para realizar esta inserção da nova coluna, veja abaixo o programa em PYTHON.

**Criando uma coluna que reflete a da tabela da satisfação de liberdade de uma região continental (America, Europa, Asia, Africa e Estados Independentes)**

mundo\_feliz['livre'] = None

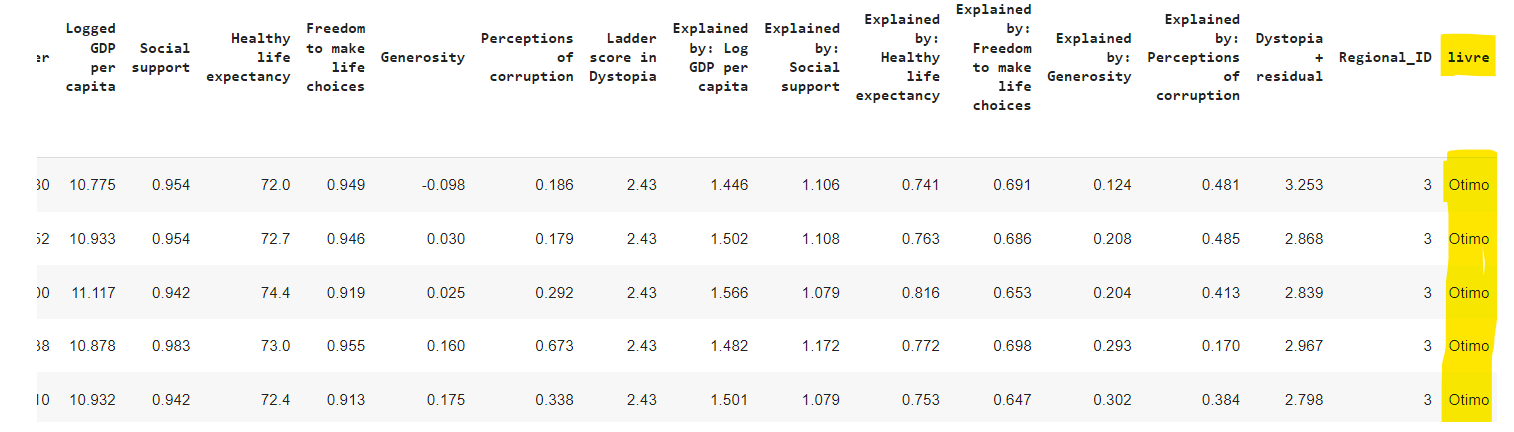
mundo\_feliz.loc[mundo\_feliz['Explained by: Freedom to make life choices'] < 0.3, 'livre'] =  'Pessimo'

mundo\_feliz.loc[(mundo\_feliz['Explained by: Freedom to make life choices'] >= 0.3) & (mundo\_feliz['Explained by: Freedom to make life choices'] < 0.6), 'livre'] = 'Bom'

mundo\_feliz.loc[mundo\_feliz['Explained by: Freedom to make life choices'] >= 0.6, 'livre'] =  'Otimo'

mundo\_feliz.to\_csv('mundo\_feliz\_2021.csv')

mundo\_feliz.head()



Após a preparação dos arquivos e criação de dois novos podemos carregar no site <https://sandbox.neo4j.com/> e manipulá-los. Veja o comando no site. Siga os passos abaixo:

1. Conecte ao site <https://sandbox.neo4j.com/> com sua chave e senha.
2. Crie um projeto do tipo blank
3. Abra o projeto criado
4. No comando abaixo digite a linha de carga (load) para realizar o download dos arquivos csv. (Nota: os arquivos deverão estar disponibilizados na web. Neste trabalho utilizamos a facilidade do sistema GitHub).

 4.1 Carregando o arquivo mundo\_feliz\_2021.csv (World Happiness Report)

***Comando:***

LOAD CSV WITH HEADERS FROM "https://raw.githubusercontent.com/DenysonLima/ColArmDados/main/mundo\_feliz\_2021.csv" AS row

CREATE (w:WorldHappinessReport)

SET w = row

 4.2 Carregando o arquivo livre.csv

***Comando:***

LOAD CSV WITH HEADERS FROM

"https://raw.githubusercontent.com/DenysonLima/ColArmDados/main/livre.csv" AS row

CREATE (l:livre)

SET l = row

4.3 Carregando o arquivo regional\_ID.csv

***Comando:***

LOAD CSV WITH HEADERS FROM

"https://raw.githubusercontent.com/DenysonLima/ColArmDados/main/regional\_ID.csv" AS row

CREATE (r:Regional)

SET r = row

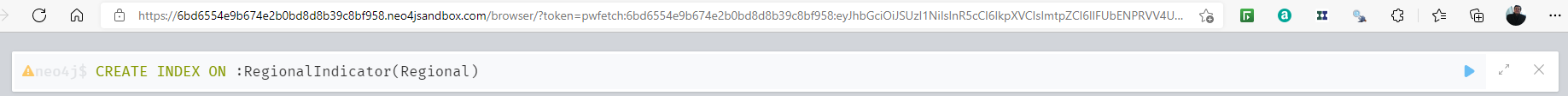
1. Após o carregamento dos três arquivos, devemos indexar todos eles, veja abaixo o comando:

***Comando:***

CREATE INDEX ON :WorldHappinessReport(Pais\_ID)

CREATE INDEX ON :WorldHappinessReport(livre)

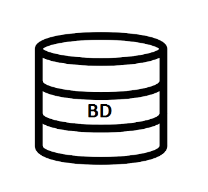
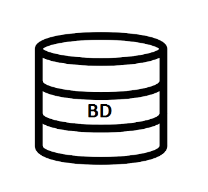
CREATE INDEX ON :RegionalIndicator(Regional)



1. Depois de carregado e indexado devemos realizar a operação do tipo *“Join”* ou relacionamentos entre os arquivos identificando os ID´s comuns entre os arquivos.



**Regional\_ID Regional**



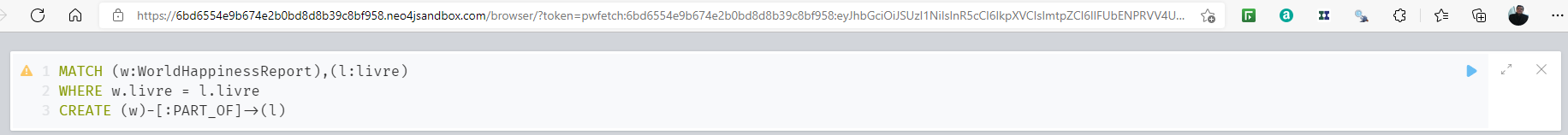
***Arquivo: world-happiness-report-2021.csv Arquivo: regional\_ID.csv***

***Comando:***

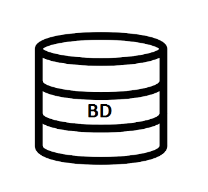
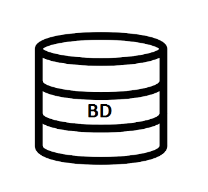
MATCH (w:WorldHappinessReport),(r:Regional)

WHERE w.Regional\_ID = r.Regional

CREATE (w)-[:PART\_OF]->(r)



**livre livre**



***Arquivo: world-happiness-report-2021.csv Arquivo: livre.csv***

***Comando:***

MATCH (w:WorldHappinessReport),(l:livre)

WHERE w.livre = l.livre

CREATE (w)-[:PART\_OF]->(l)

# Resultado:

Após o carregamento dos arquivos através do comando LOAD CSV WITH HEADERS FROM <https://www...) e a criação da indexação dos arquivos pertinentes através do comando CREATE INDEX ON <nome do arquivo.csv> e finalmente elaborado todos os relacionamentos entre os CSV´s, teremos o seguinte resultado da visualização.

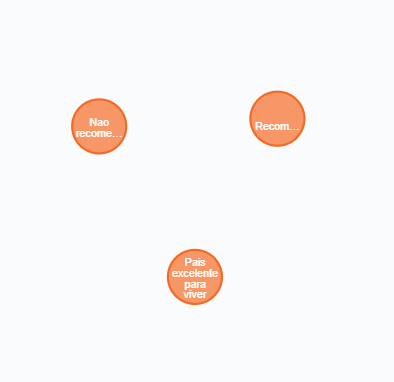
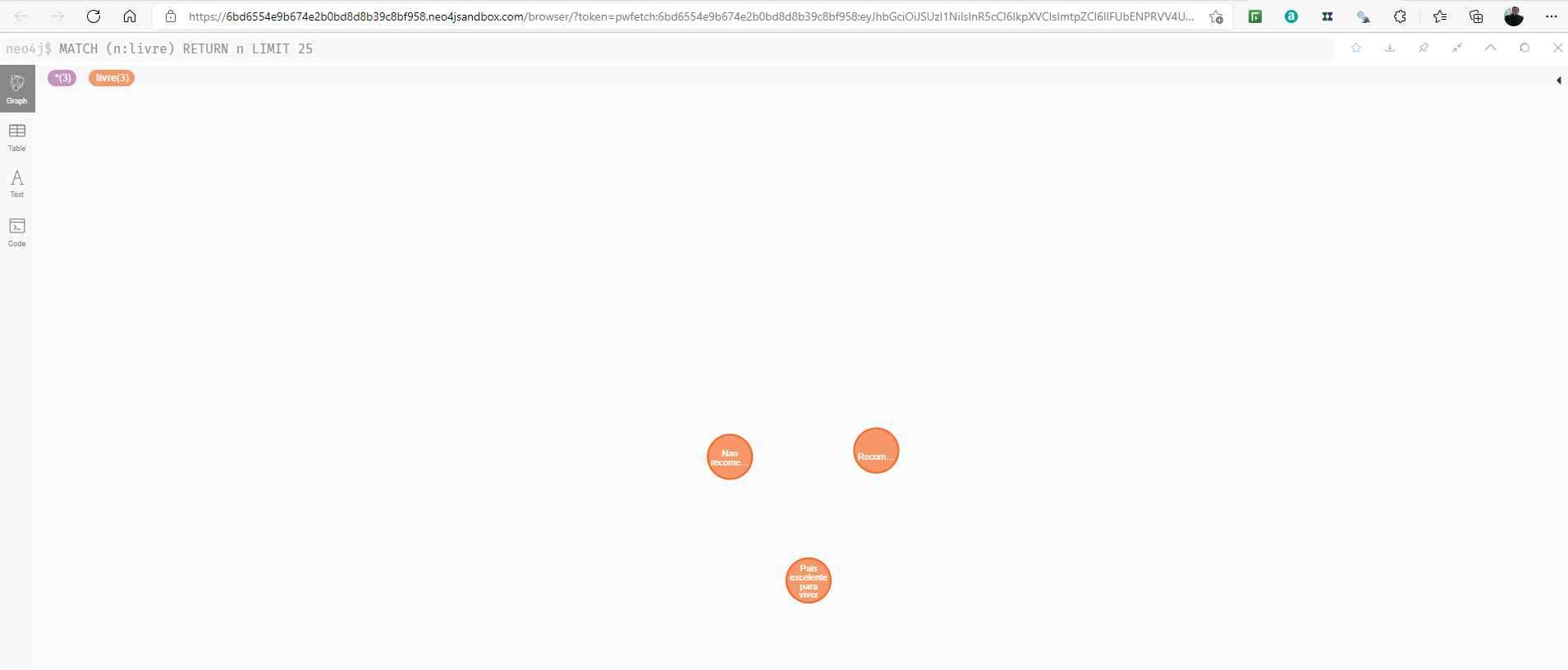
Uma imagem contendo Círculo

Descrição gerada automaticamente

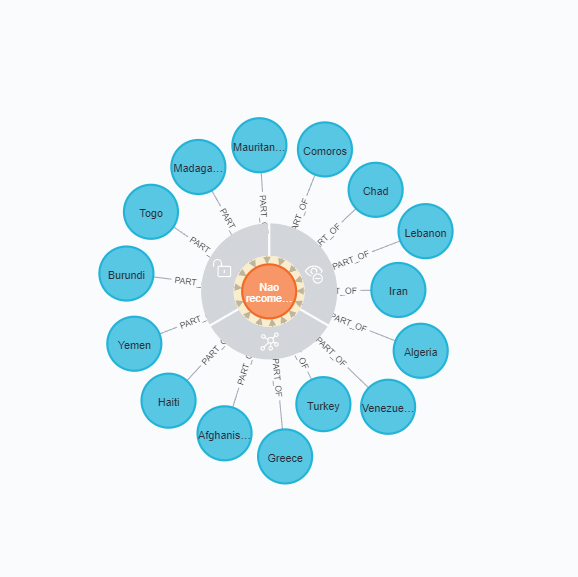
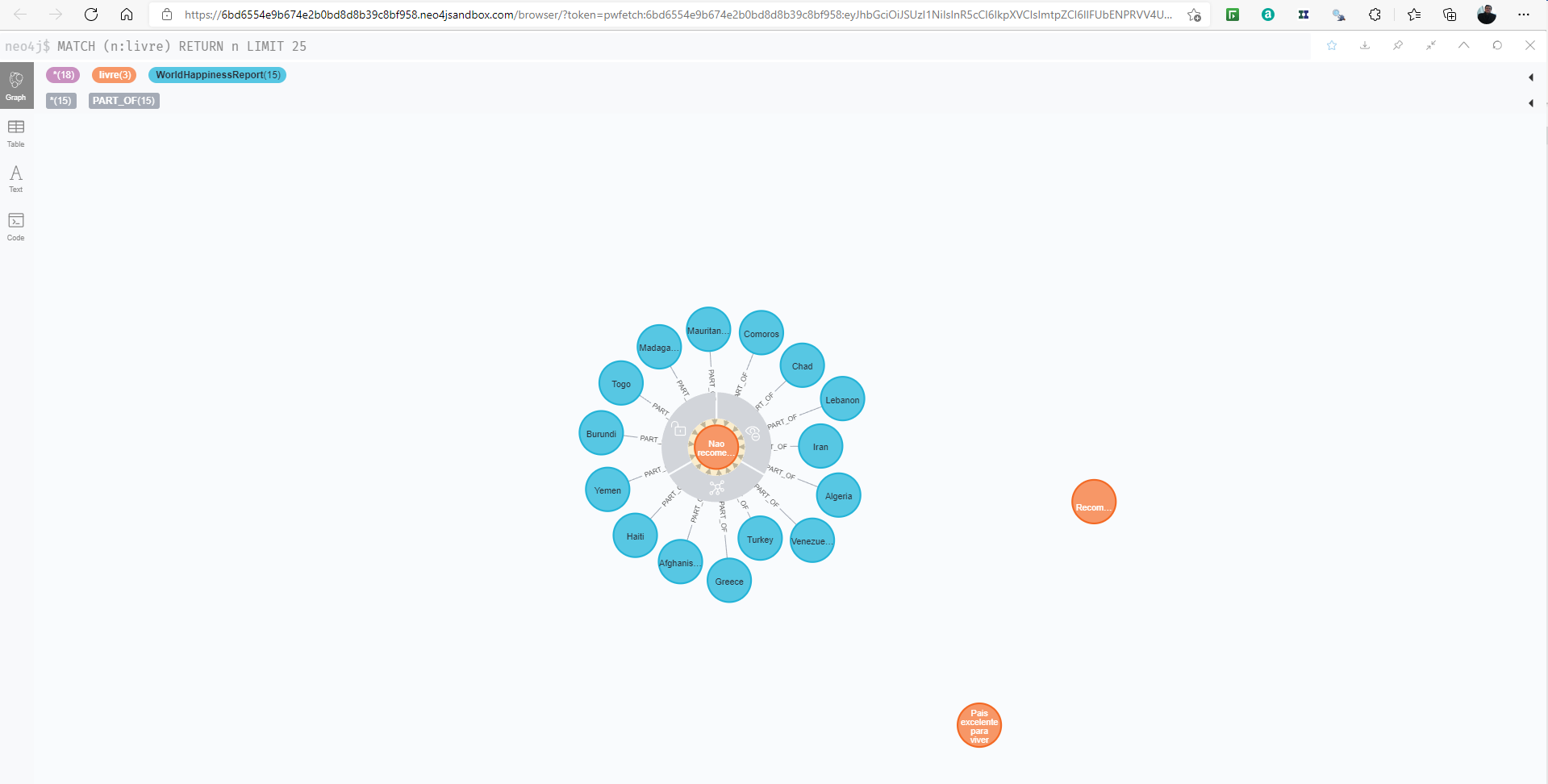
Gráfico, Padrão do plano de fundo

Descrição gerada automaticamente

Todos os países. (zoom) (mundo\_feliz\_2021.csv)



Categoria Liberdade para fazer escolhas (livre.csv)



Países Não recomendado para viver conforme coluna no

arquivo mundo\_feliz\_2021.csv

Gráfico, Gráfico de bolhas

Descrição gerada automaticamente

Países agrupado conforme o relacionamento regional.

Adotando a regra elaborado no PYTHON:

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'Latin America and Caribbean', 'Regional\_ID'] = '1'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'North America and ANZ', 'Regional\_ID'] = '2'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'Central and Eastern Europe', 'Regional\_ID'] = '3'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'Western Europe', 'Regional\_ID'] = '3'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'East Asia', 'Regional\_ID'] = '4'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'Southeast Asia', 'Regional\_ID'] = '4'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'South Asia', 'Regional\_ID'] = '4'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'Sub-Saharan Africa', 'Regional\_ID'] = '5'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'Middle East and North Africa', 'Regional\_ID'] = '5'

mundo\_feliz.loc[mundo\_feliz['Regional indicator'] == 'Commonwealth of Independent States', 'Regional\_ID'] = '6'

