

Objetivo

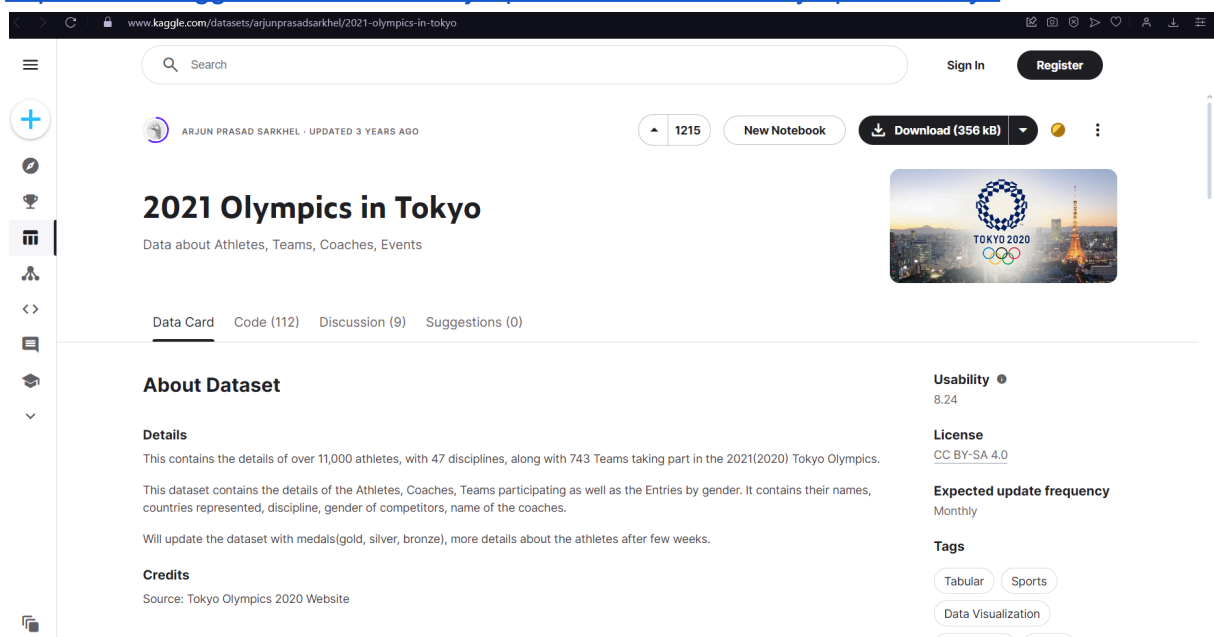
A partir de dados sobre as olimpíadas de tokyo, analisar:

1. Listar em ordem decrescente os países com mais medalhas.
2. Média de inscrições por gênero em cada disciplina

Detalhamento

Busca pelos Dados

<https://www.kaggle.com/datasets/arjunprasadsarkhel/2021-olympics-in-tokyo>



The screenshot shows the Kaggle dataset page for '2021 Olympics in Tokyo' by Arjun Prasad Sarkhel. The page includes a search bar, a sidebar with navigation icons, and a main content area. The dataset is described as 'Data about Athletes, Teams, Coaches, Events'. It has 1215 code files, 9 discussions, and 0 suggestions. The dataset is updated 3 years ago. The 'About Dataset' section provides details about the data, including the number of athletes, disciplines, and teams. The 'Credits' section mentions the source as the Tokyo Olympics 2020 Website. The 'Usability' score is 8.24, and the license is CC BY-SA 4.0. The 'Expected update frequency' is monthly. The 'Tags' section includes 'Tabular', 'Sports', and 'Data Visualization'.

Coleta

Dados baixados para a máquina local e inseridos manualmente em um azure blob storage.

Microsoft Azure | Data Factory | tokyo-olympic-dfmlp | Pesquisar alocador e documentação

Experiência de versão prévia | Desligado

Atividades

- Pesquisar atividades
- Mover e transformar
 - Copiar dados
 - Fluxo de dados
- Synapse
- Azure Data Explorer
- Azure Functions
- Serviço em Lote
- Databricks
- Data Lake Analytics
- Geral
- HDInsight
- Iteração e condicionais
- Machine Learning
- Power Query

```

graph LR
    A[Copiar dados - Athletes] --> B[Copiar dados - Coaches]
    B --> C[Copiar dados - EntriesGender]
    C --> D[Copiar dados - Medals]
    D --> E[Copiar dados - Teams]
  
```

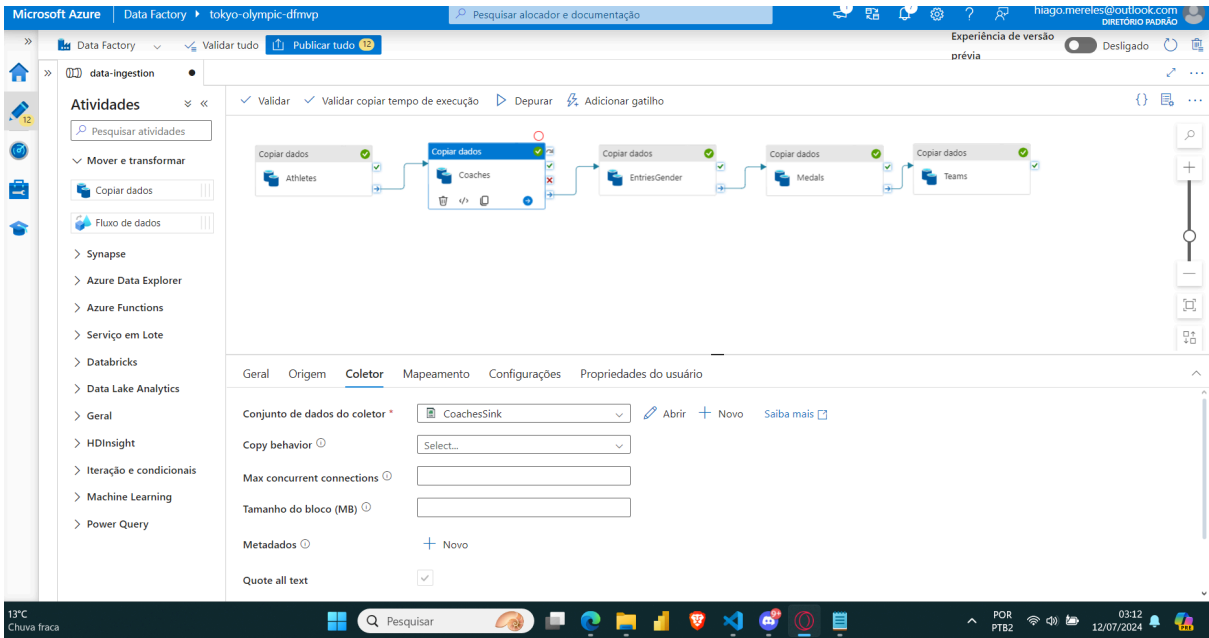
Parâmetros Variáveis Configurações **Saída**

ID de execução de pipeline: bbb9a74d-d2ef-42f4-b462-dda964078294 🔄 ⌚ 🛑 **Status do pipeline:** Bem-sucedido Monitorar nas Métricas do Azure 🔍 Exportar para CSV 📄

All status ▾
Mostrando 1 – 5 de 5 itens

Nome da atividade ↕	Status da atividade 1	Tipo de atividade ↕	Início da execução ↑	Duração ↕	Runtime de integração	Propriedades do usuário	ID de execução
Teams	Bem-sucedido	Copiar dados	7/12/2024, 4:55:05 AM	13s	AutoResolveIntegrator	a3bcc7dd-753f-	731d4ba4-8c07-
Medals	Bem-sucedido	Copiar dados	7/12/2024, 4:54:53 AM	12s	AutoResolveIntegrator	731d4ba4-8c07-	721ca326-3581-
EntriesGender	Bem-sucedido	Copiar dados	7/12/2024, 4:54:29 AM	22s	AutoResolveIntegrator	721ca326-3581-	

Depois de executar a ingestão de dados realizei as configurações para extrair os dados da fonte, no caso os arquivos .csv:
Athletes, Coaches, EntriesGender, Medals e Teams da pasta raw-data do azure blob.



Toda a transformação de dados e análises citadas no início do documento foram realizadas no Databricks:

https://colab.research.google.com/drive/1ctne1jixVEjKwVp_JjLSpHS-YKAdjEE?usp=sharing

Dados transformados

Página inicial > tokyoolympicdatamp > Contêineres >

tokyo-olympic-data Contêiner

Pesquisar

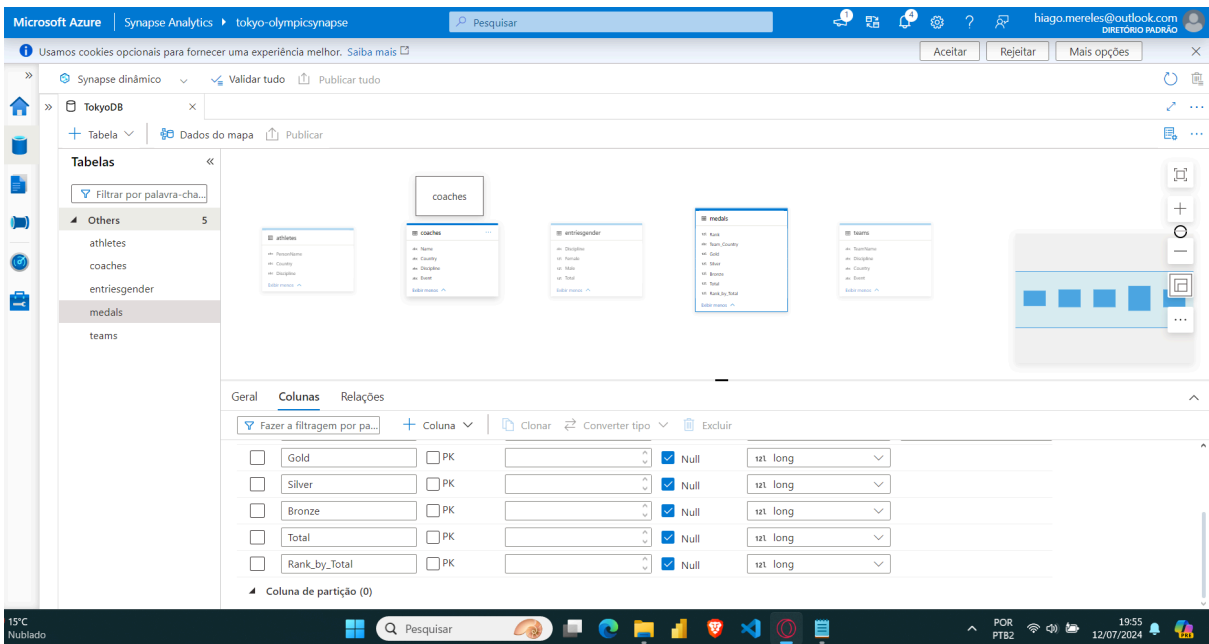
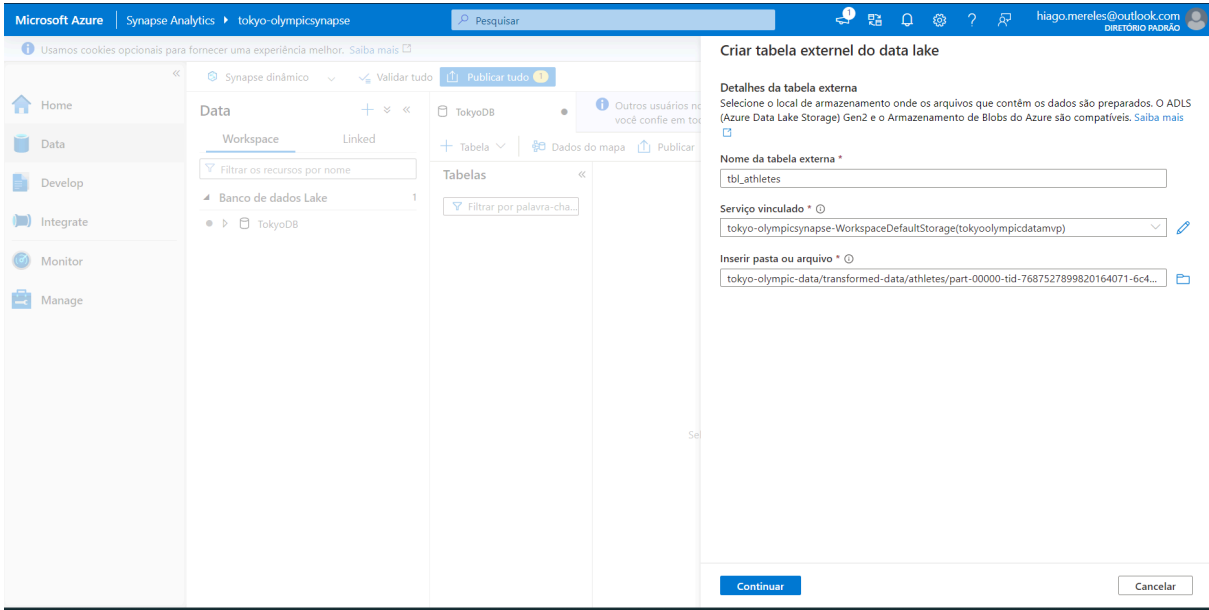
Carregar + Adicionar Diretório Atualizar | Renomear Excluir Alterar a camada Adquirir concessão Interromper concessão Enviar comentários

Método de autenticação: Chave de acesso (Alternar para conta de usuário do Microsoft Entra)
Local: tokyo-olympic-data / transformed-data

Pesquisar blobs por prefixo (diferenciar maiúsculas de minúsculas) ☐ Mostrar os objetos que foram excluídos

Nome	Modificado	Camada de acesso	Status do arquivo	Tipo de blob	Tamanho	Estado de concessão
<input type="checkbox"/> [-]						...
<input type="checkbox"/> athletes					-	...
<input type="checkbox"/> coaches					-	...
<input type="checkbox"/> entriesgender					-	...
<input type="checkbox"/> medals					-	...
<input type="checkbox"/> teams					-	...

e para concluir, podemos o Azure Synapse Analytics para realizar algumas análises utilizando python ou SQL como nos exemplos abaixo:



Microsoft Azure

Synapse Analytics

tokyo-olympicssynapse

Pesquisar

hiago.mereles@outlook.com

DIRETORIO PADRAO

Usamos cookies opcionais para fornecer uma experiência melhor. Saiba mais

Aceitar

Rejeitar

Mais opções

»

Synapse dinâmico

Validar tudo

Publicar tudo

»

Workspace

Linked

Filtrar os recursos por nome

Banco de dados Lake

TokyoDB

Tabelas

Athletes

Colunas

TokyoDB

SQL script 1

Executar

Desfazer

Publicar

Plano de consulta

Conectar-se com

Interno

Usar o banco de dados

TokyoDB

1

2

3

4

SELECT TOP (100) [PersonName]

,[Country]

,[Discipline]

FROM [TokyoDB].[dbo].[Athletes]

Resultados

Mensagens

Exibir

Tabela

Gráfico

Exportar os resultados

Pesquisar

PersonName

Country

Discipline

AALERUD Katrine

Norway

Cycling Road

ABAD Nestor

Spain

Artistic Gymnastics

ABAGNALE Giovanni

Italy

Rowing

00:00:00 Consulta executada com êxito.

Propriedades

Geral

Relacionados (0)

Nome

SQL script 1

Descrição

Tipo

.sql script

Tamanho

87 bytes

Configurações de resultados por consulta

Primeiras 5000 linhas (padrão)

Todas as linhas

15°C

Nublado

Pesquisar

POR

PTB2

19:49

12/07/2024