OBJETIVO

O Índice de Desenvolvimento humano é uma medida resumida do progresso a longo prazo em três dimensões básicas do desenvolvimento humano: renda, educação e saúde. O objetivo da criação do IDH foi oferecer um contraponto a outro indicador utilizado, o Produto Interno Bruto (PIB) per capita, que considera apenas a dimensão econômica do desenvolvimento.

As Faixas do IDH são:

Abaixo de 0,499 – Muito Baixo

Entre 0,500 e 0,599 – Baixo

Entre 0,600 e 0,699 – Médio

Entre 0,700 e 0,799 – Alto

Acima de 0,800 – Muito alto

O objetivo desse trabalho é verificar quais indicadores estão diretamente ligados ao IDH e que tipos de medidas podem ser executadas com o intuito de elevar esse índice.

Quais são as 30 cidades com o maior IDH? E o que elas têm em comum?

Quais são as 30 cidades com o menor IDH? E o que elas têm em comum?

O aumento nos impostos arrecadados pode aumentar o IDH?

O aumento das despesas municipais pode aumentar o IDH?

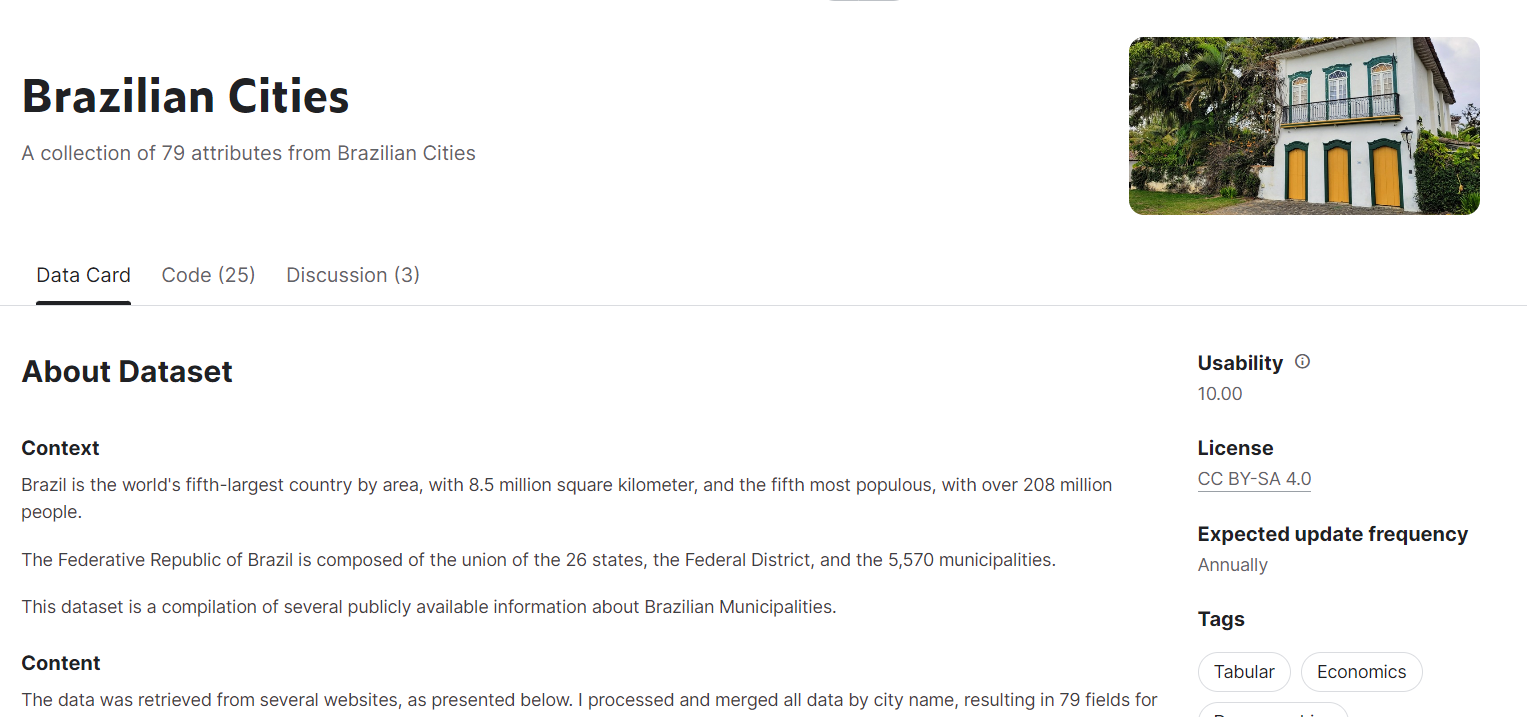
DETALHAMENTO

1. Busca pelos dados

Para realização desse estudo, foi necessário realizar as buscas pelos dados que poderiam responder a essas perguntas, para tanto, foi encontrado no repositório de dados abertos da kaggle, um dataset com as informações necessárias para atingir o objetivo deste projeto:

<https://www.kaggle.com/datasets/crisparada/brazilian-cities>

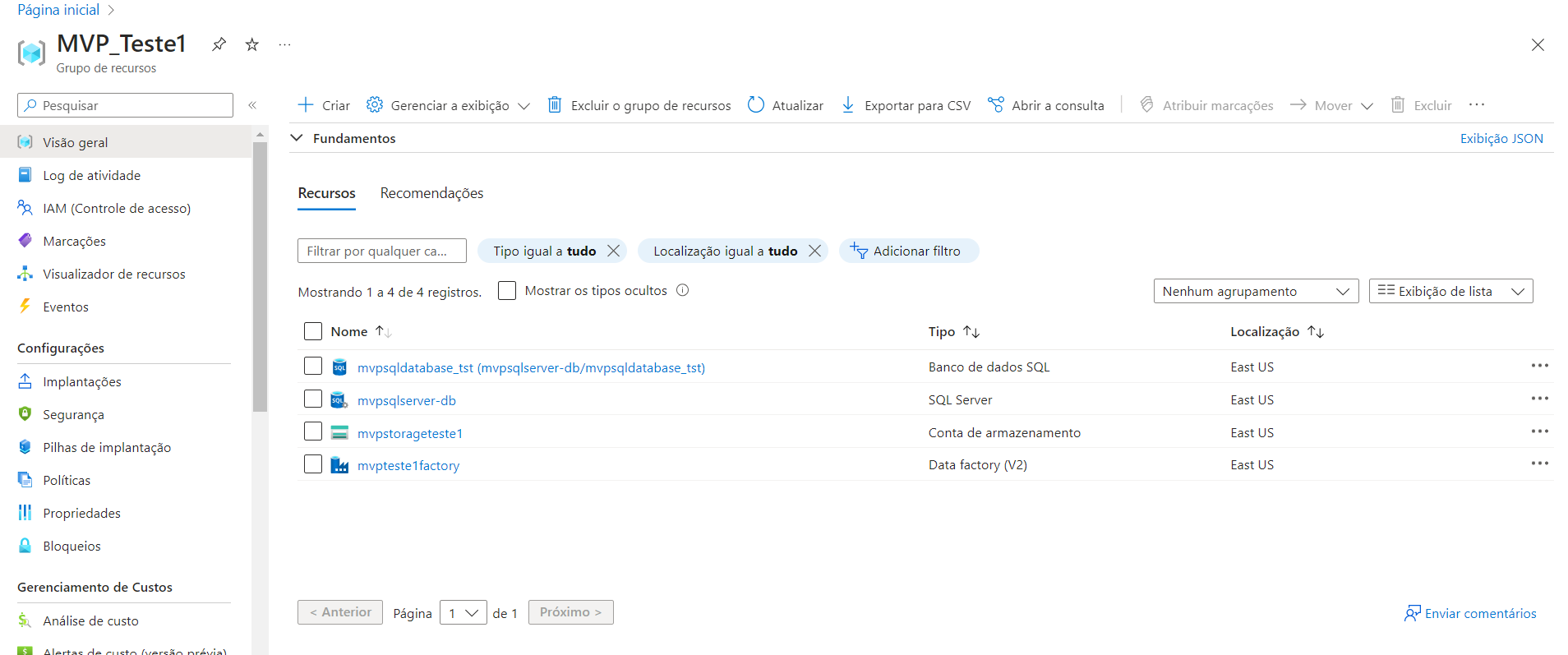
Esse dataset já estava pronto, com os dados retirados de diversos sites, conforme apresentado no catálogo de dados.



1. Coleta dos dados

A coleta dos dados foi realizada por intermédio do download dos dados e estes foram armazenados na nuvem, para tanto, foi necessário realizar a criação de uma plataforma na nuvem para armazenamento dos dados e posterior tratamento destes.

Neste passo, foi criada uma conta na plataforma da Microsoft Azure, inicialmente foi necessário realizar a criação de um Grupo de Recursos, onde os recursos criados seriam armazenados dentro deste.



Após a criação deste grupo de recursos, foi necessário realizar a criação de um storage, onde seriam efetivamente carregados os dados.

Foi realizada a criação de um SQL server e de um Banco de Dados.

1. Modelagem de Dados

Foi idealizado inicialmente a idéia de ter somente uma tabela, tendo em vista que o dataset que seria utilizado possuía todas as informações necessárias, entretanto, ao se importar o Dataset no Dataflow do Azure Data Factory, foi verificado que os atributos não eram exibidos no select individualmente, pois a quantidade de atributos excedia o número de 50, o que acarretava em exibir todos os atributos juntos, portanto, foi realizado o particionamento deste dataset em dois.

Para realizar o particionamento dos datasets, foi utilizado o Virtual Studio Code e realizado por intermédio da linguagem Python a divisão do DataSet em dois, conforme pode ser observado na figura a seguir:

XXXXXXXXXXX