

IMDB DATASET - EXPORTAÇÃO DE DADOS RELACIONAIS PARA DADOS NÃO RELACIONAIS

INTEGRANTES:

Jéssica Nascimento Pessoa da Silva - 01717533

João Guilherme Caetano dos Santos - 01481382

João Victor Mendonça da Silva - 01480878

Paulo Vinícius Feliciano de Souza - 01618133

Rosilene da Silva Lima - 01619051

TURMA: 3NA

FACULDADE MAURÍCIO DE NASSAU - GRAÇAS - PE -RECIFE

1. Ferramentas necessárias para executar o projeto:.....	3
---	----------

1. Introdução

Nesse projeto iremos abordar diversas ferramentas para que seja garantida a comunicação de bancos de dados SQL e NoSQL, utilizando de diversos artefatos para alcançar esse objetivo, nesse momento utilizamos de uma dataset disponibilizada pelo site [Kaggle](#), o qual optamos por utilizar o dataset referente ao top 250 filmes do IMDB.

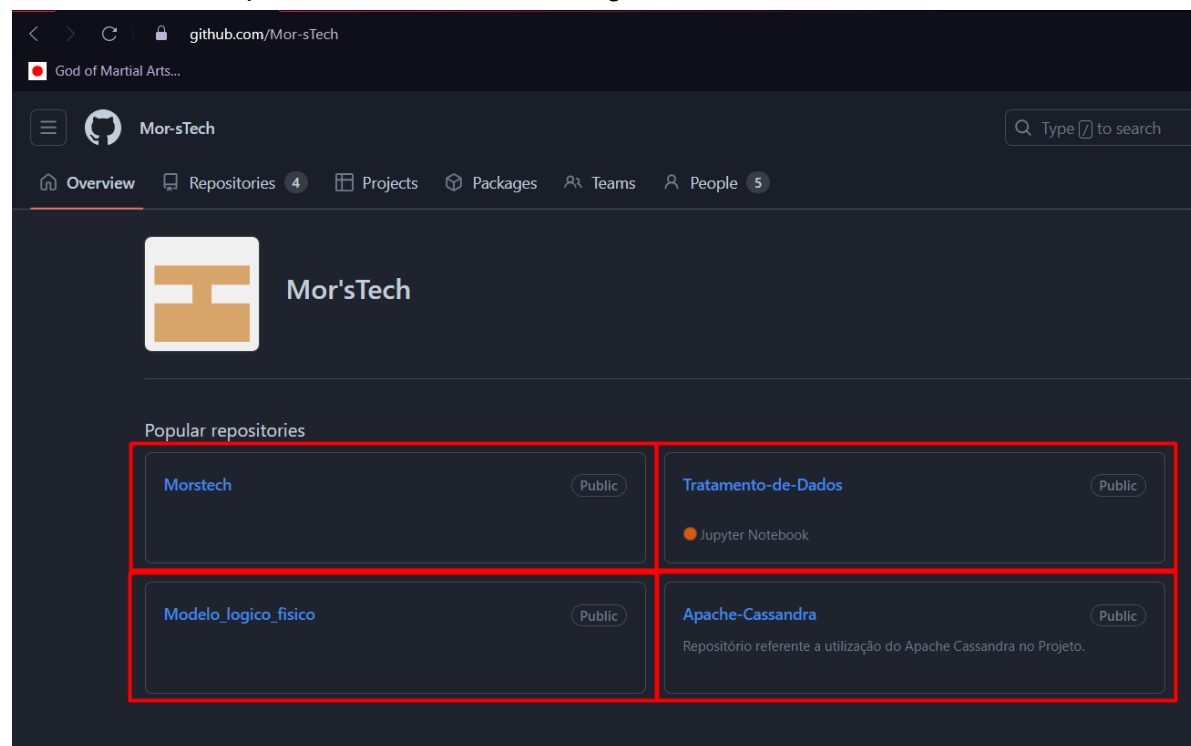
2. Ferramentas necessárias para executar o projeto:

- GIT
- VSCode
- Docker Desktop
- MYSQL Workbench
- PYTHON

Toda essa composição de ferramentas se torna parte crucial para a execução do projeto.

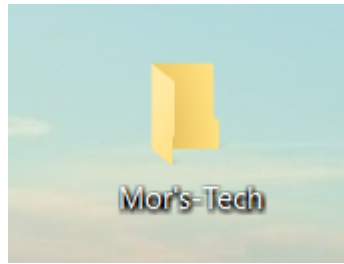
3. Passo a passo para utilização do projeto

- 3.1. Primeiro passo: utilize do artefato GIT para clonar todos os repositórios necessários, os repositórios se encontram no qithub da [Mor's Tech](#)

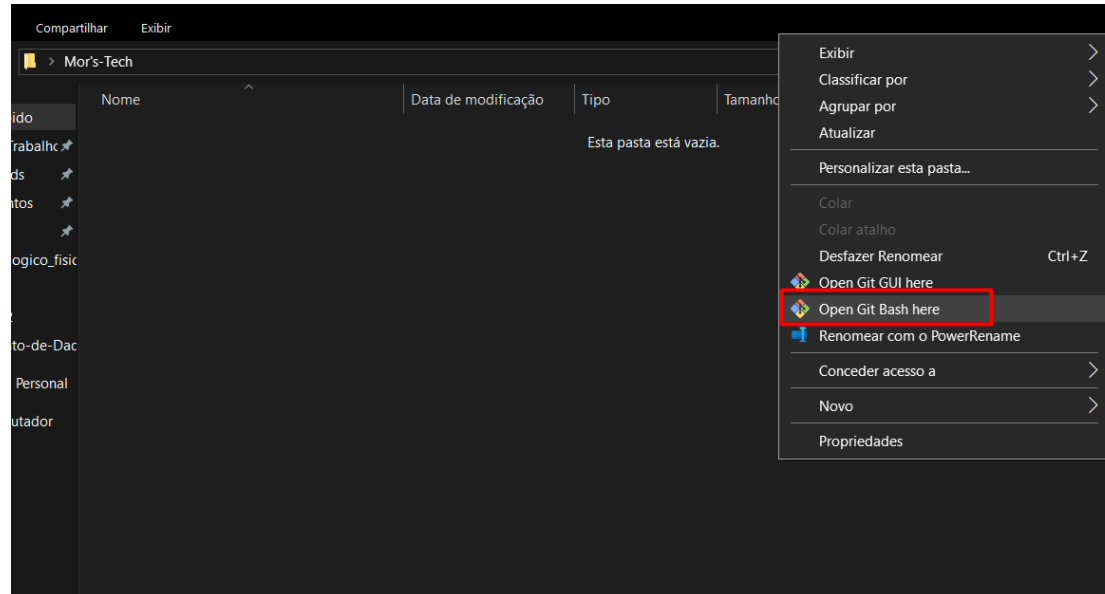


- 3.2. Após encontrar os repositórios devemos utilizar o clone dos repositórios, da seguinte maneira:

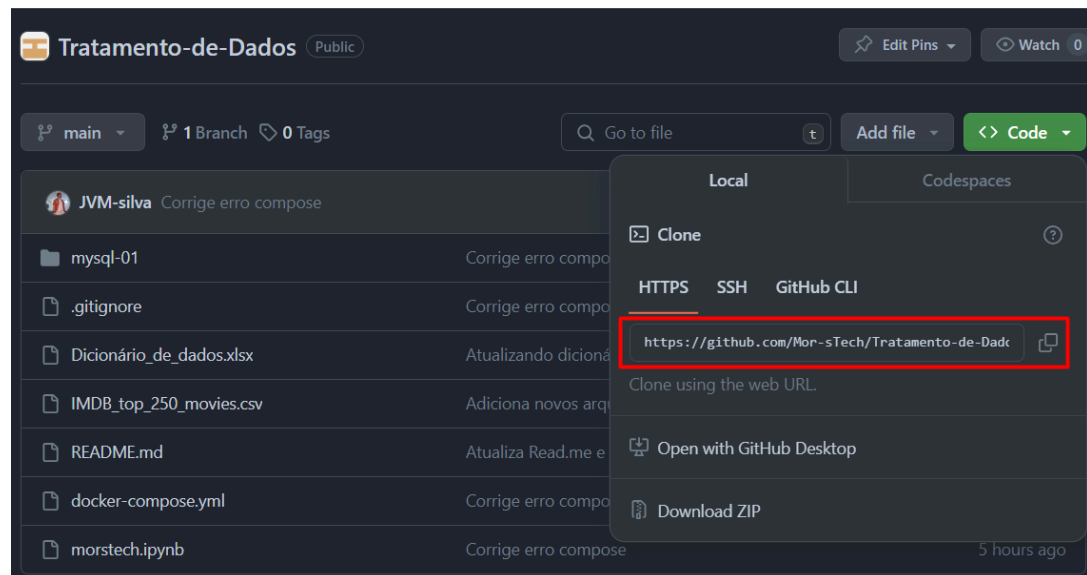
- ### 3.2.1. Criamos uma pasta nova, para manter os arquivos organizados:



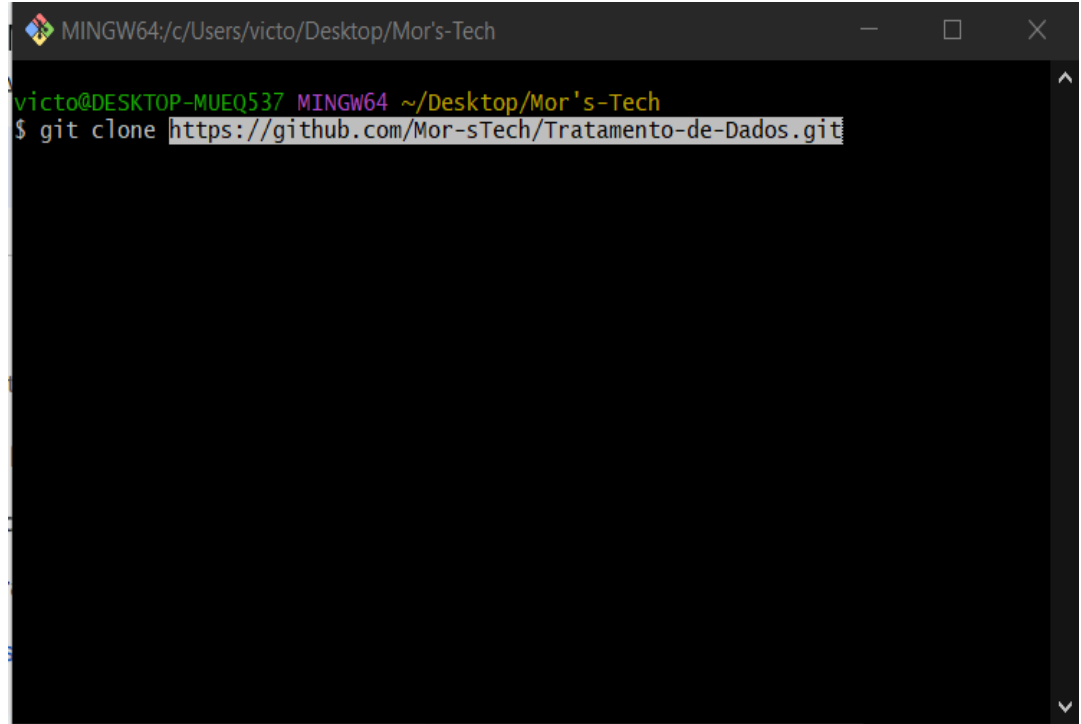
3.2.2. Pressione botão direito e acesse a opção git bash here



3.2.3. Retornamos para o repositório e copiamos o HTTPS ao pressionar o ícone “code” e copiando a área destacada:



- 3.2.4. Já estando dentro do git que abriu mais cedo utilizamos o código “git clone ‘endereço do repositório’ ”:

A terminal window with a dark background. The title bar shows the path "MINGW64:/c/Users/victo/Desktop/Mor's-Tech". The prompt is "victo@DESKTOP-MUEQ537 MINGW64 ~/Desktop/Mor's-Tech". The command entered is "\$ git clone https://github.com/Mor-sTech/Tratamento-de-Dados.git".

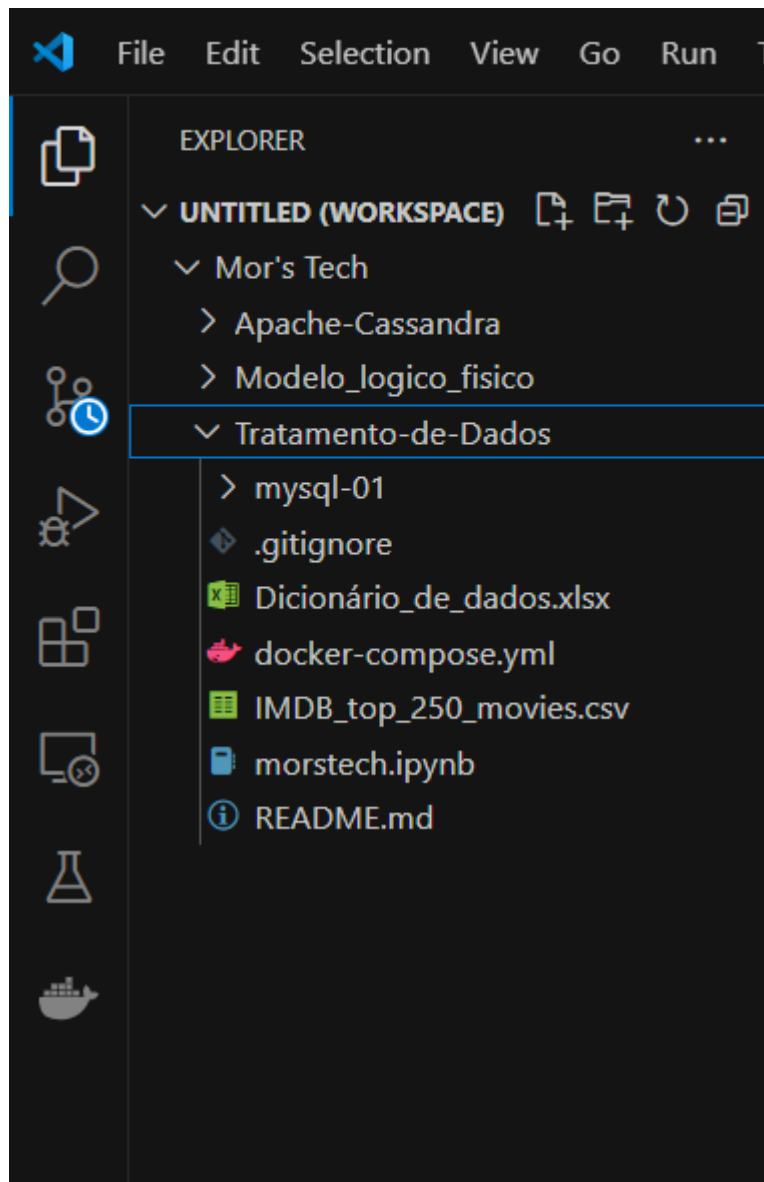
```
MINGW64:/c/Users/victo/Desktop/Mor's-Tech
victo@DESKTOP-MUEQ537 MINGW64 ~/Desktop/Mor's-Tech
$ git clone https://github.com/Mor-sTech/Tratamento-de-Dados.git
```

- 3.2.5. Repetimos para os outros repositórios.

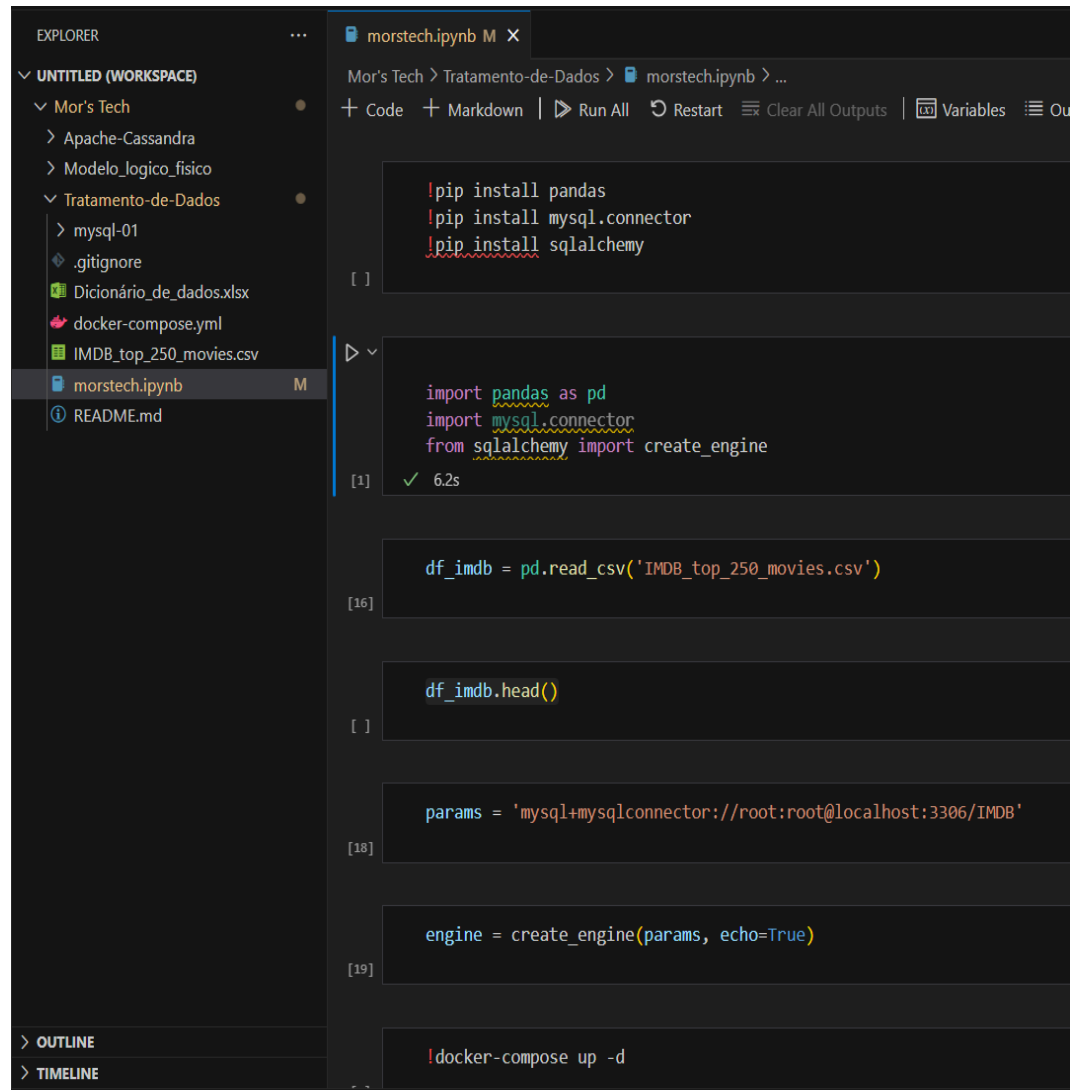
- 3.2.6. Você terá as seguintes pastas:

Nome	Data de modificação	Tipo	Tamanho
Apache-Cassandra	13/05/2024 23:14	Pasta de arquivos	
Modelo_logico_fisico	13/05/2024 20:53	Pasta de arquivos	
Tratamento-de-Dados	13/05/2024 17:48	Pasta de arquivos	

- 3.3. Acesse seu VSCode e siga os seguintes passos:
- 3.3.1. Abra a pasta onde se encontram os repositórios dentro do VSCode



3.3.2. Abra o arquivo “morstech.ipynb”



```
!pip install pandas
!pip install mysql.connector
!pip install sqlalchemy

import pandas as pd
import mysql.connector
from sqlalchemy import create_engine

df_imdb = pd.read_csv('IMDB_top_250_movies.csv')

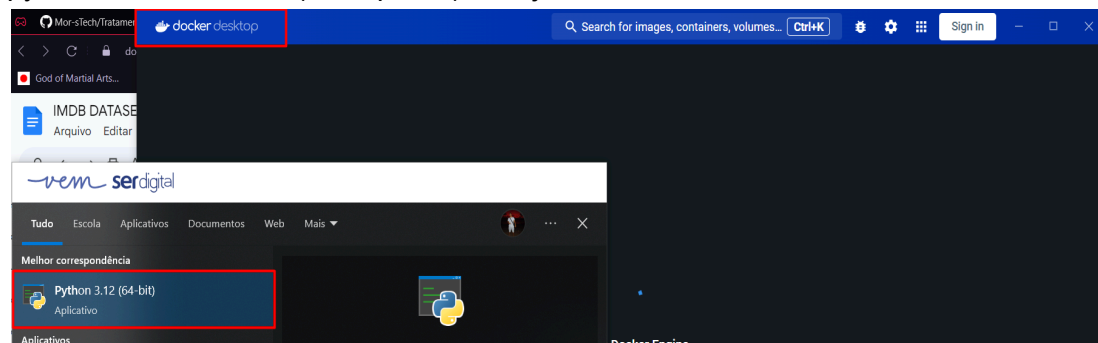
df_imdb.head()

params = 'mysql:mysqlconnector://root:root@localhost:3306/IMDB'

engine = create_engine(params, echo=True)

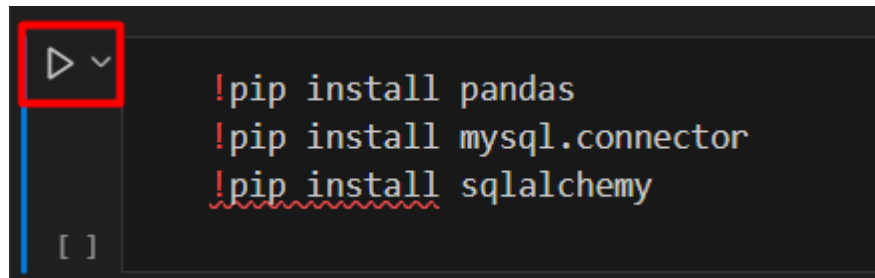
!docker-compose up -d
```

3.3.3. Você irá se deparar com diversas funções, imports e outros comandos a mais, nessa parte é crucial que você já tenha instalado o python, versão 3.12.x (ou superior), esteja com o docker aberto.



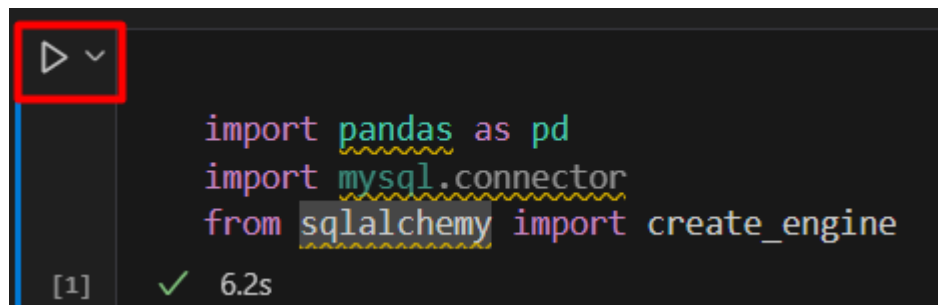
3.3.4. Execute as partes que serão indicadas pressionando o seguinte ponto:

First: Instale as bibliotecas



```
!pip install pandas
!pip install mysql.connector
!pip install sqlalchemy
```

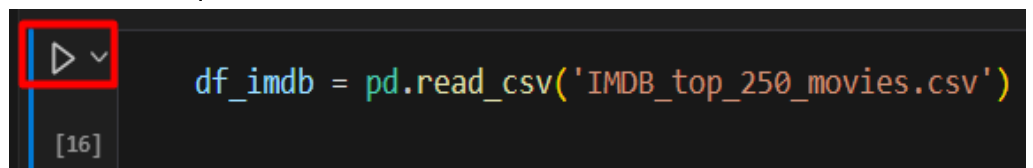
Next: Importe as bibliotecas



```
import pandas as pd
import mysql.connector
from sqlalchemy import create_engine
```

[1] ✓ 6.2s

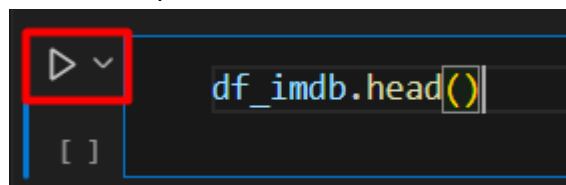
Next: Lê os arquivos do csv



```
df_imdb = pd.read_csv('IMDB_top_250_movies.csv')
```

[16]

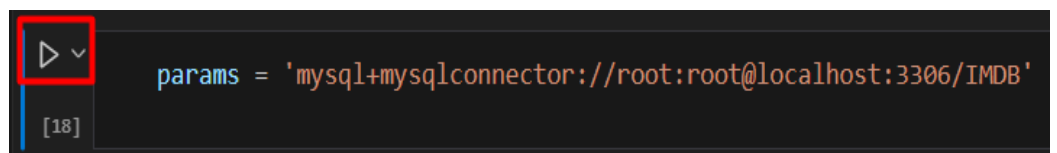
OBS: Esse próximo mostra como ficou seu banco



```
df_imdb.head()
```

[]

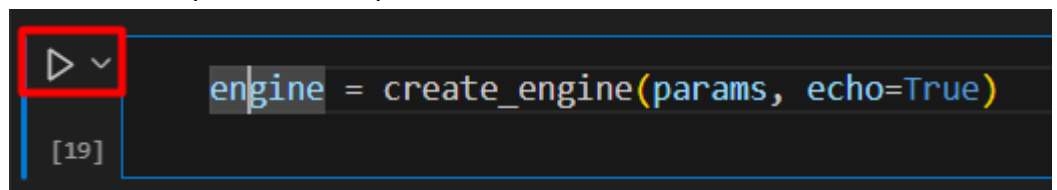
Next: Adiciona os parâmetros para conexão do banco de dados
MYSQL



```
params = 'mysql+mysqlconnector://root:root@localhost:3306/IMDB'
```

[18]

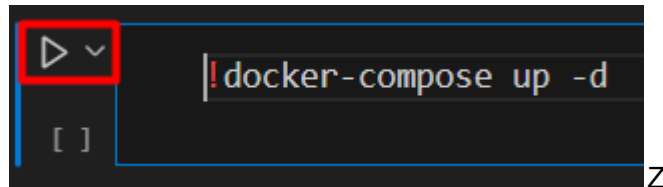
Next: Starta o processo dos parâmetros



```
engine = create_engine(params, echo=True)
```

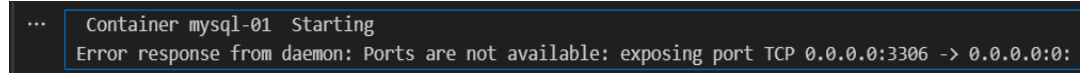
[19]

Next: Inicia o contêiner docker



```
!docker-compose up -d
```

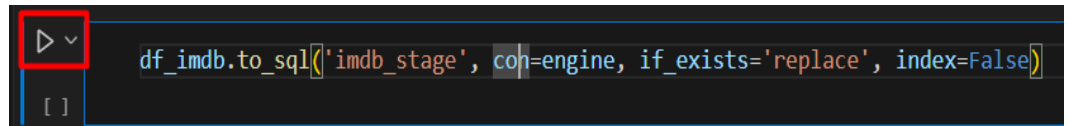
OBS: Caso apareça o seguinte erro, finalize as instâncias que estejam utilizando a porta 3306, como por exemplo "msqld.exe"



```
... Container mysql-01 Starting  
Error response from daemon: Ports are not available: exposing port TCP 0.0.0.0:3306 -> 0.0.0.0:0:
```

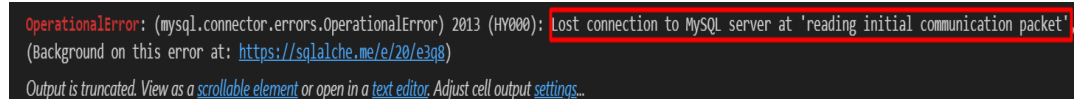
OBS: aguarde aproximadamente 1 minuto para que o contêiner possa subir o banco de dados e então siga o próximo passo:

Next: Envia os dados do csv para a tabela "imdb_stage"



```
df_imdb.to_sql('imdb_stage', con=engine, if_exists='replace', index=False)
```

Caso ocorra este erro, abra o MYSQL Workbench e volte para o VSCode para rodar novamente essa função



```
OperationalError: (mysql.connector.errors.OperationalError) 2013 (HY000): Lost connection to MySQL server at 'reading initial communication packet'  
(Background on this error at: https://sqlalche.me/e/20/e3q8)  
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

3.3.5.