

Jupyter Notebook

João Pedro V. Pinheiro

`joaopedro.pinheiro88@gmail.com`

`20-Mai-2022`

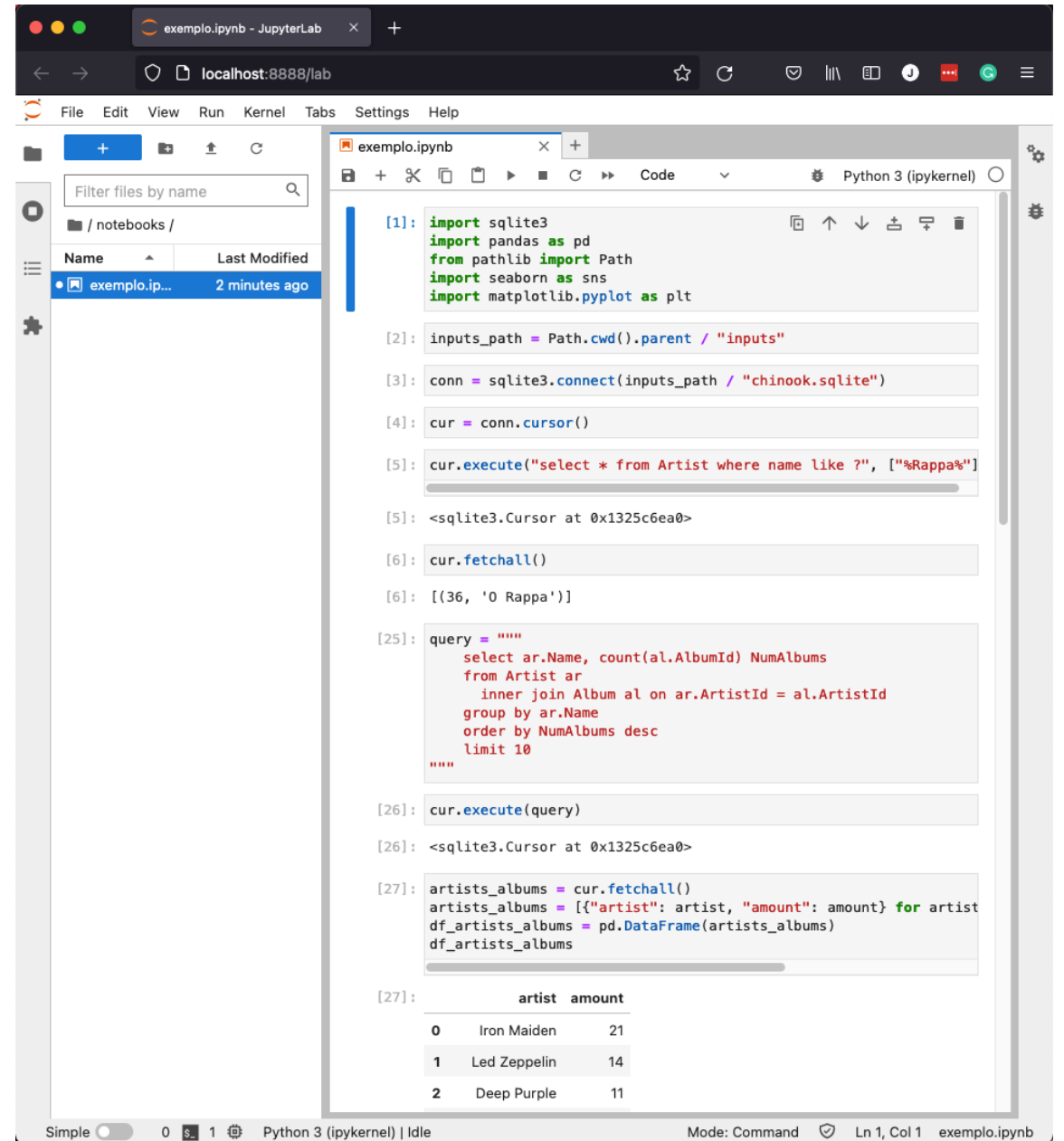
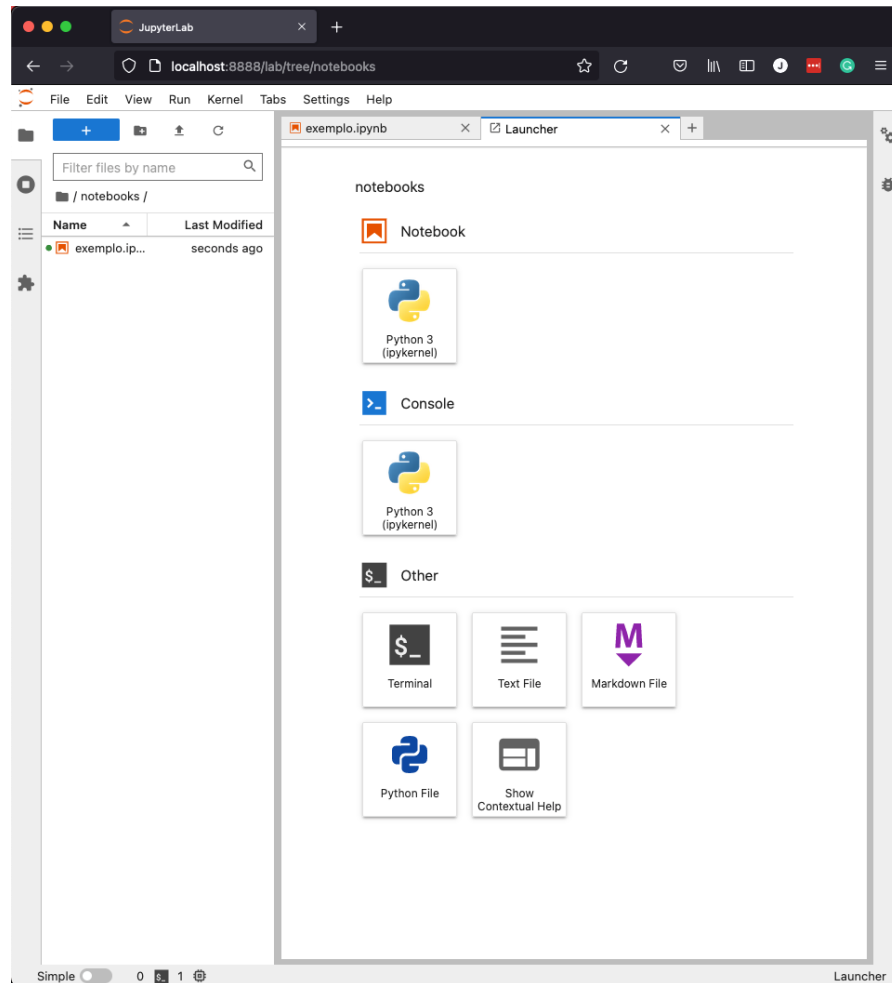
Jupyter

Definição e Origem

- Jupyter é projeto sem fins lucrativos e *open-source*
- Proveniente do Projeto IPython em 2014
 - interpretador interativo para várias linguagens de programação
 - utilizado como *kernel* (núcleo do SO) no projeto Jupyter
 - atender as demandas das áreas de Ciência de Dados e Computação Científica
- Diferentemente do seu antecessor, Jupyter Notebook (ou JupyterLab) é executado diretamente do navegador
- Seu desenvolvimento favorece a escrita e execução de pequenos trechos de código
 - essa característica acelera o desenvolvimento e facilita o teste de hipóteses

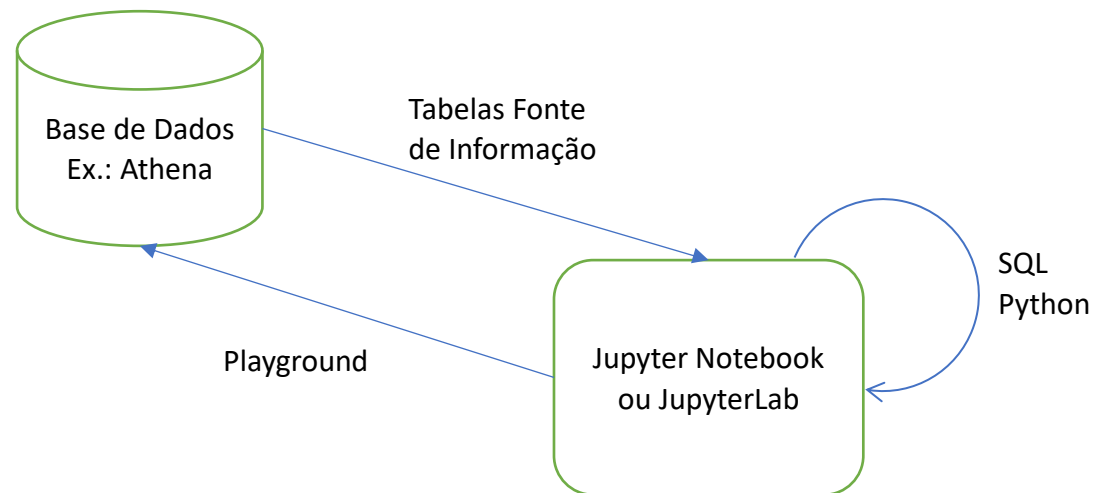
Jupyter

Interface web interativa



Airflow

Arquitetura



Jupyter

Arquitetura

- A arquitetura proposta foi baseada principalmente na apresentação do último encontro
- Com SQL e Python há possibilidade de exploração e visualização da informação
 - configuração para conexão com a base do Amazon Athena
 - <https://www.cdata.com/kb/tech/athena-odbc-python-linux.rst>
 - bibliotecas de visualização e manipulação dos dados
 - Pandas (<https://pandas.pydata.org/>)
 - Seaborn (<https://seaborn.pydata.org/>)
 - ...

Jupyter

Exemplo

- Carregamento das bibliotecas

```
[1]: import sqlite3
import pandas as pd
from pathlib import Path
import seaborn as sns
import matplotlib.pyplot as plt
```

- Consulta ao banco de dados

```
[25]: query = """
      select ar.Name, count(al.AlbumId) NumAlbums
      from Artist ar
      inner join Album al on ar.ArtistId = al.ArtistId
      group by ar.Name
      order by NumAlbums desc
      limit 10
      """
```

```
[26]: cur.execute(query)
```

Jupyter

Exemplo

■ Visualização dos dados

```
[27]: artists_albums = cur.fetchall()
artists_albums = [{"artist": artist, "amount": amount} for artist, amount in artists_albums]
df_artists_albums = pd.DataFrame(artists_albums)
df_artists_albums
```

```
[27]:
```

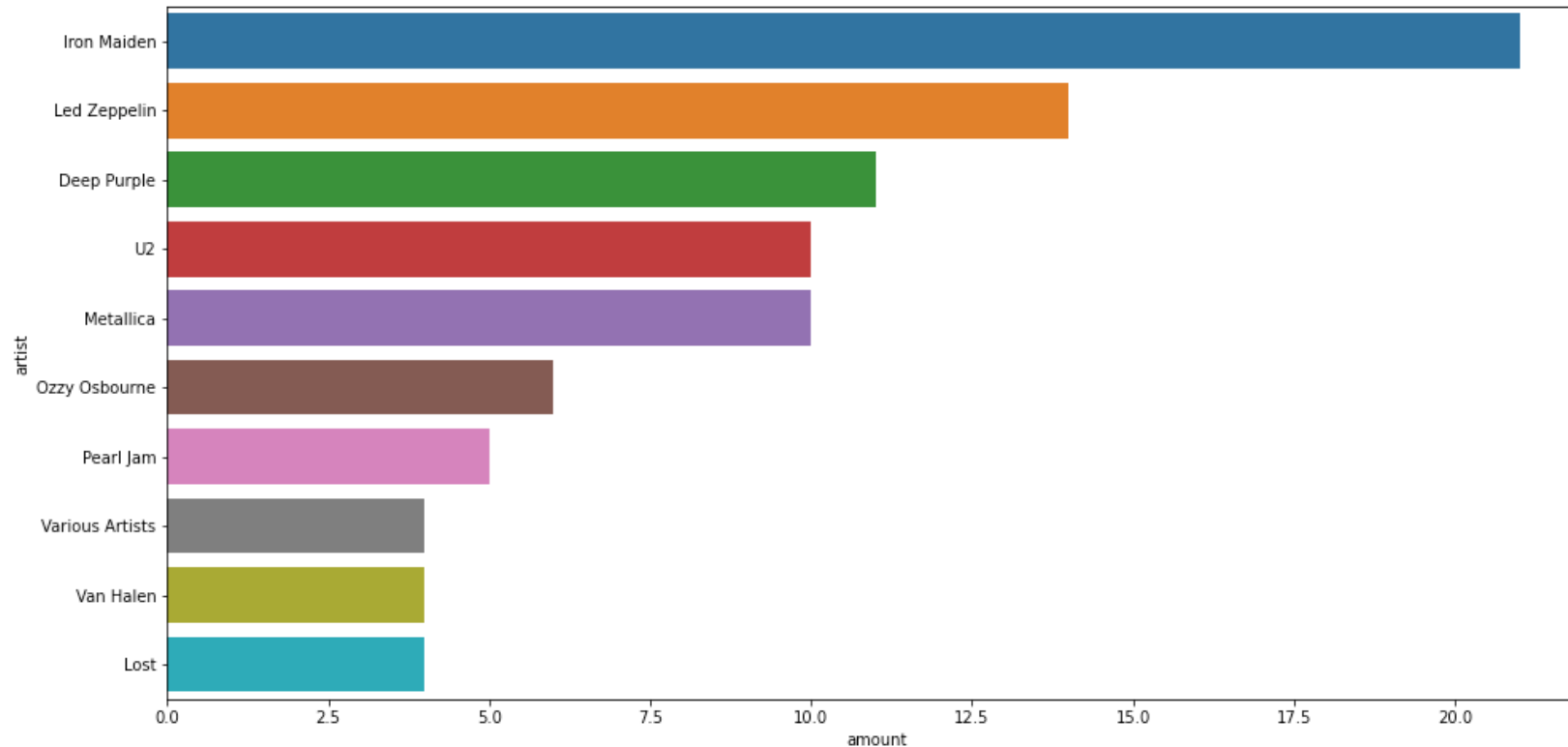
	artist	amount
0	Iron Maiden	21
1	Led Zeppelin	14
2	Deep Purple	11
3	U2	10
4	Metallica	10
5	Ozzy Osbourne	6
6	Pearl Jam	5
7	Various Artists	4
8	Van Halen	4
9	Lost	4

Jupyter

Exemplo

■ Visualização dos dados

```
[28]: plt.figure(figsize=(16, 8))  
sns.barplot(x="amount", y="artist", data=df_artists_albums)  
plt.close(2)
```



Obrigado!

João Pedro V. Pinheiro

`joaopedro.pinheiro88@gmail.com`

`29-Abr-2022`