

DATA ANALYTICS

HISTORICAL FACTS VS ENTERTAINMENT

Date: 12/11/2022

By: Ana de Ondarza



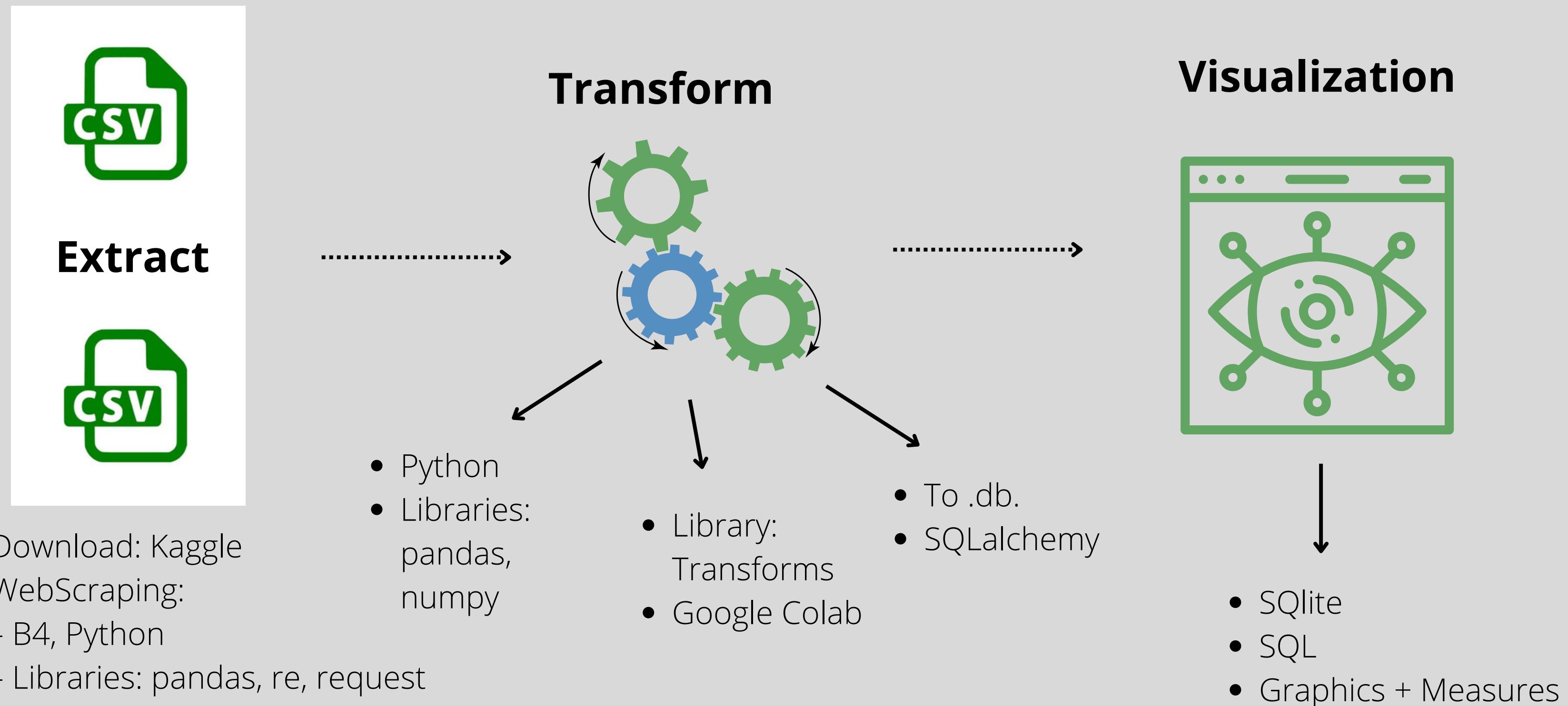
INDEX



- 1** **Movies DataSets**
From Kaggle
- 2** **Cleaning and wrangling CVS
(Movies, TV-Shows)**
*Tools: python + libraries:
pandas, numpy*
- 3** **Historical events CSV**
*Web Scraping, Tools:
python, BS4soup libraries:
request, re, pandas*
- 4** **Add sentiment**
*Tool: library
Transformers +
Google Colab*
- 5** **Transform to .db.**
*Tool: python,
libraries:
sqlalchemy, pandas*
- 6** **Analysis in PBI**
Tool: SQLite, SQL
- 7** **Analysis in PBI**
Measures + graphics
- 8** **Conclusions**
- 9** **Next Steps**



Project Architecture



MOVIES AND TV-SHOWS DATASETS

The image shows two side-by-side screenshots from the Kaggle platform. The left screenshot displays the 'Netflix popular movies dataset' by user 'itisnanarayan63'. It features a large thumbnail of the Netflix logo, followed by the dataset title, the author's name, and details like 'Updated a month ago', 'Usability 9.4', and '1 File (CSV)'. The right screenshot shows 'The Movies Dataset' by Rounak Banik, with a thumbnail of several movie stills, the dataset title, the author's name, and details like 'Updated 5 years ago', 'Usability 8.2', '239 MB', and '7 Files (CSV)'. Both screenshots include a small navigation bar at the bottom.

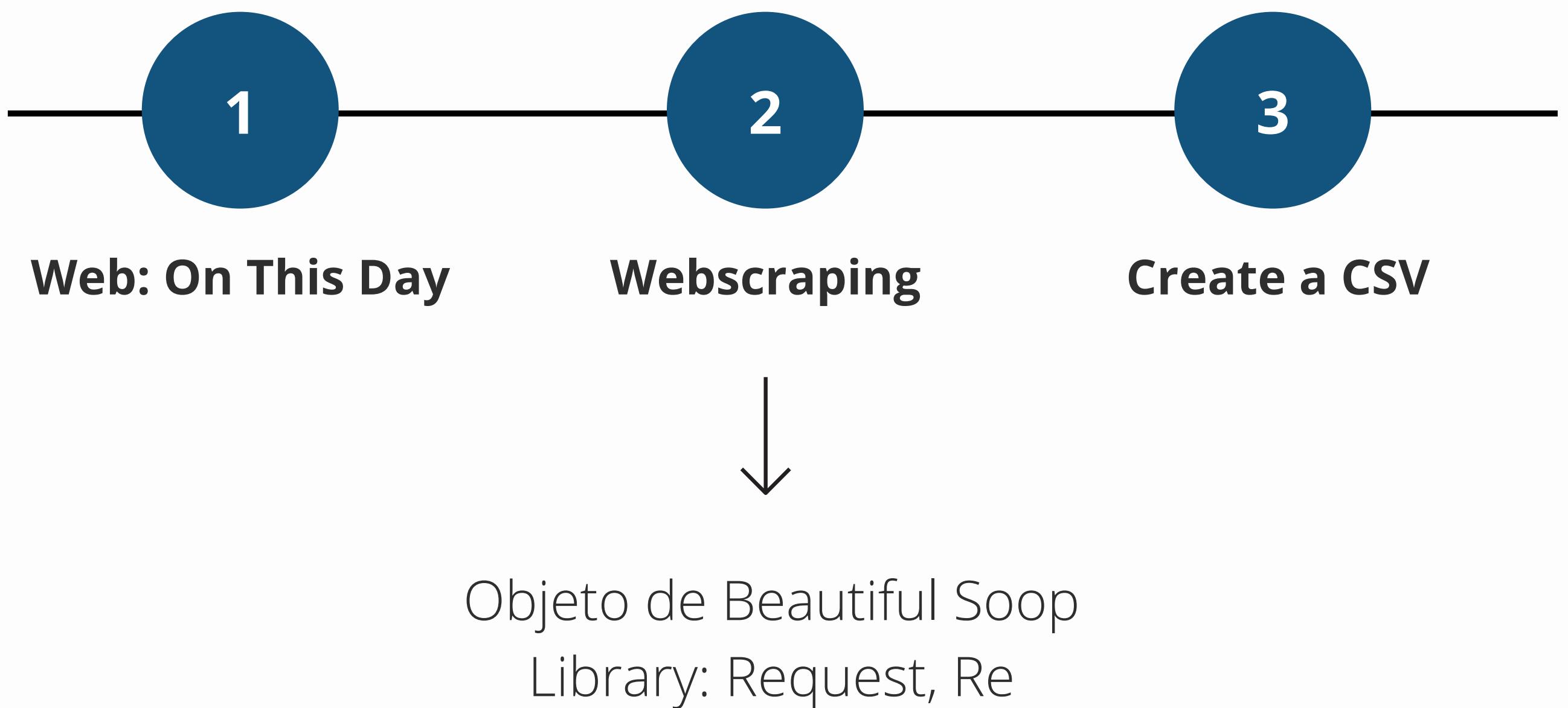
Kaggle Datasets:

1. Netflix
2. Disney
3. Amazon
4. Hulu

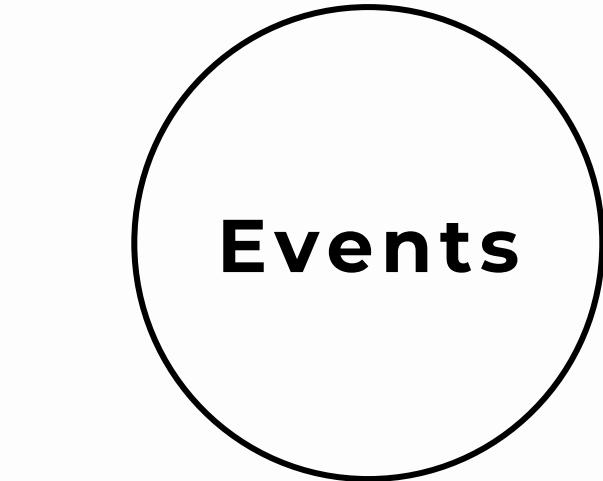
(1920-2022)

Tools: Python, libraries:
pandas and numpy

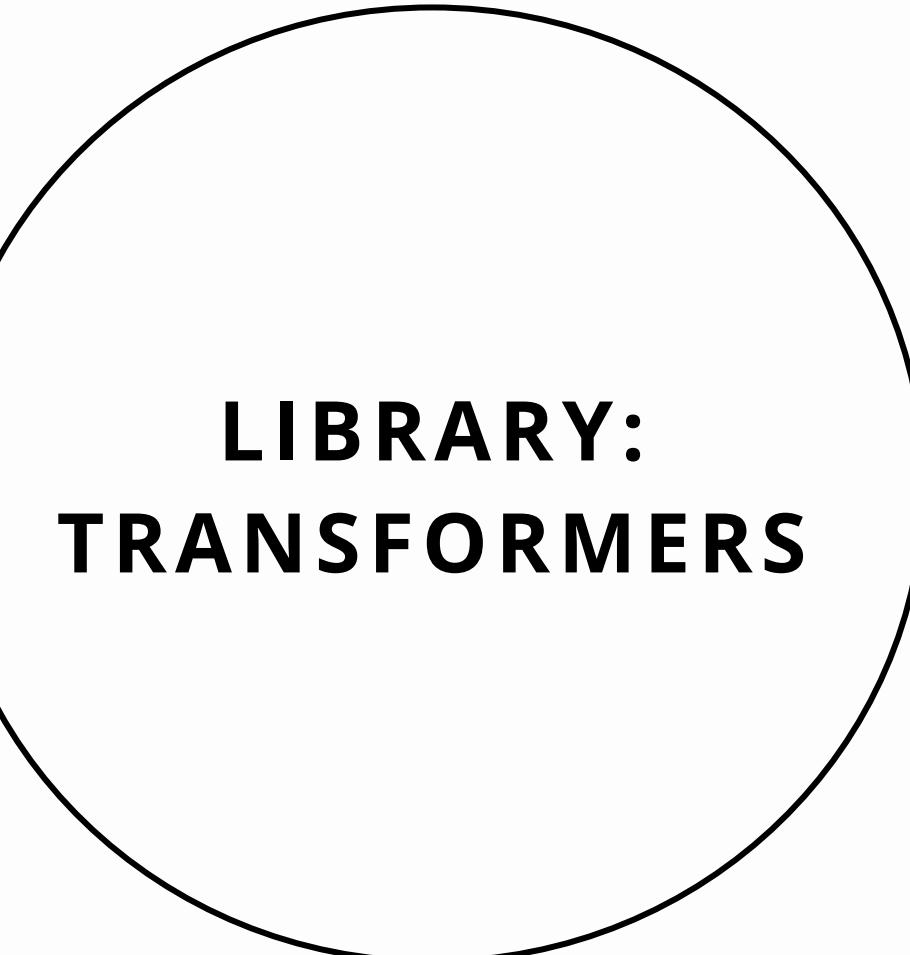
HISTORICAL EVENTS DATASET



ADD A SENTIMENT

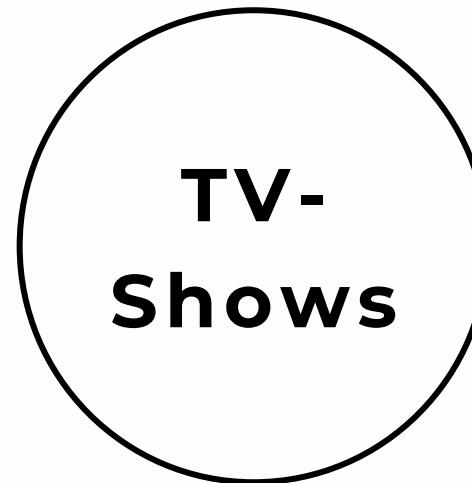


The principal column

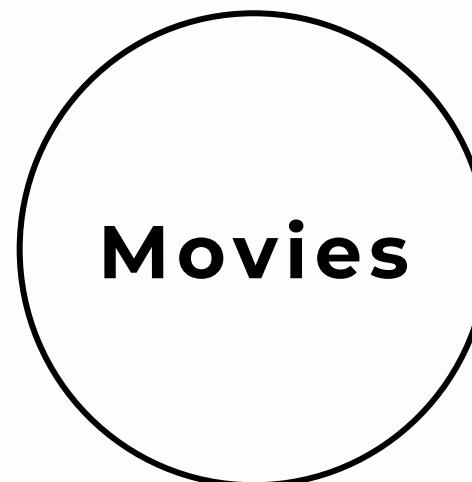


Objective: add a sentiment

Google Colab



Join all the columns in
one



Join all the columns in
one

SQL IN POWER BI

PHASE 1

TRANSFORM THE CSV TO .DB.

- Tools: Python,
libraries:
SQLalchemy and
Pandas

PHASE 2

SQLITE

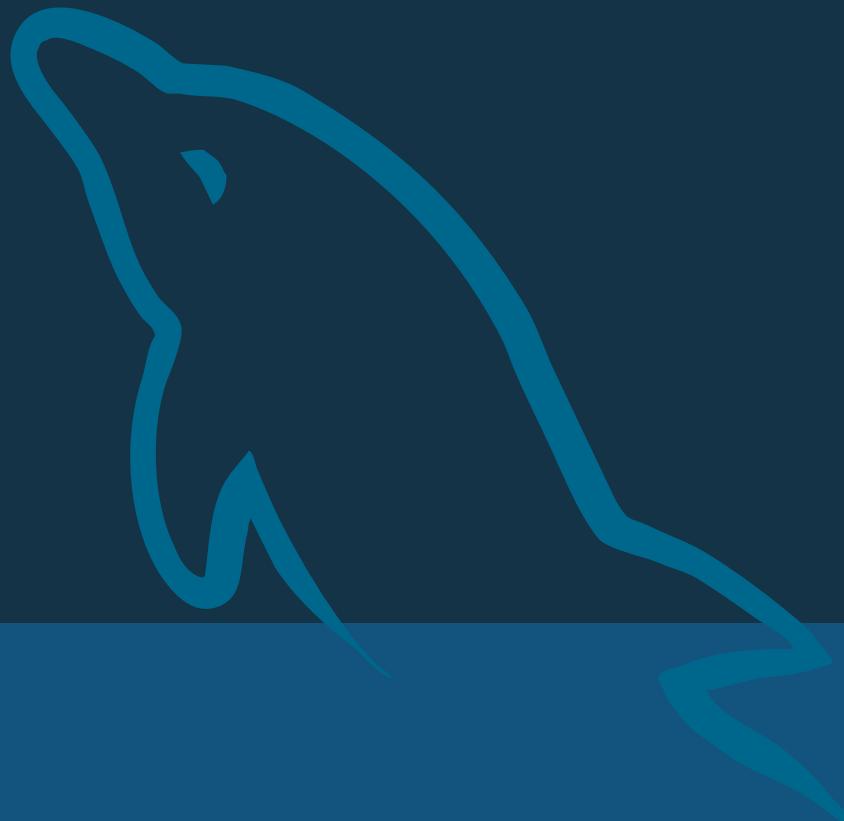
- Download SQLite
- ODBC connection
- Database = Path

PHASE 3

SQL QUERIES



Power BI



POWER BI ANALYSIS



NEXT STEPS



kaggle

- Direct connection to Kaggle's API
- Code refactoring

GRACIAS!!