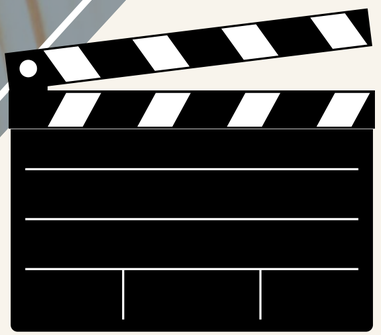
DATA ANALYTICS

HISTORICAL FACTS VS ENTERTAIMENT

Date: 12/11/2022

By: Ana de Ondarza



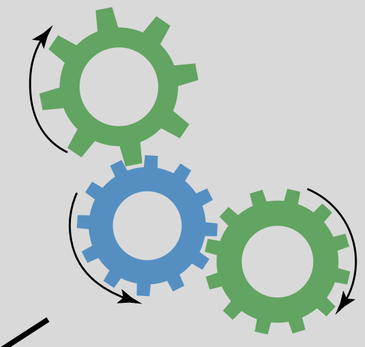
INDEX

Project Architecture

Extract

- Download: Kaggle
- WebScraping:
  - B4, Python
  - Libraries: pandas, re, request

Transform

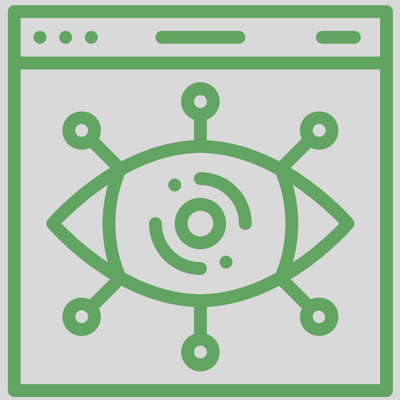


- Python
- Libraries: pandas, numpy

- Library: Transforms
- Google Colab

- To .db.
- SQLalchemy

Visualization



- SQlite
- SQL
- Graphics + Measures

MOVIES AND TV-SHOWS DATASETS

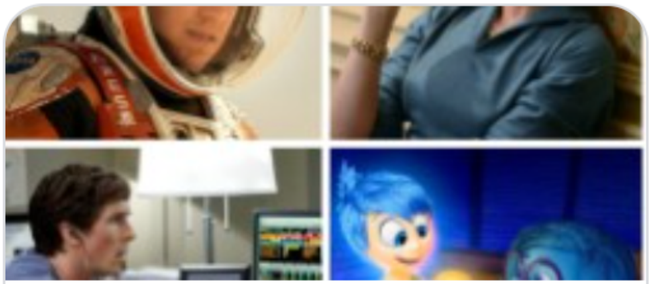
Netflix popular movies dataset

itisnarayan63 · Updated a month ago

Usability **9.4** · 1 MB

1 File (CSV)

72



The Movies Dataset

Rounak Banik · Updated 5 years ago

Usability 8.2 · 239 MB

7 Files (CSV)

3018

Kaggle Datasets:

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

(1920-2022)

Tools: Python, libraries: pandas and numpy

HISTORICAL EVENTS DATASET

1 **Web: On This Day**

2 **Webscraping**

3 **Create a CSV**

↓

Objeto de Beautiful Soop
Library: Request, Re

ADD A SENTIMENT

Events

The principal column

LIBRARY: TRANSFORMERS

Objective: add a sentiment

TV-Shows

Join all the columns in one

Movies

Join all the columns in one

Google Colab

SQL IN POWER BI

PHASE 1

TRANSFORM THE CSV TO .DB.

- Tools: Python, libraries: SQLalchemy and Pandas

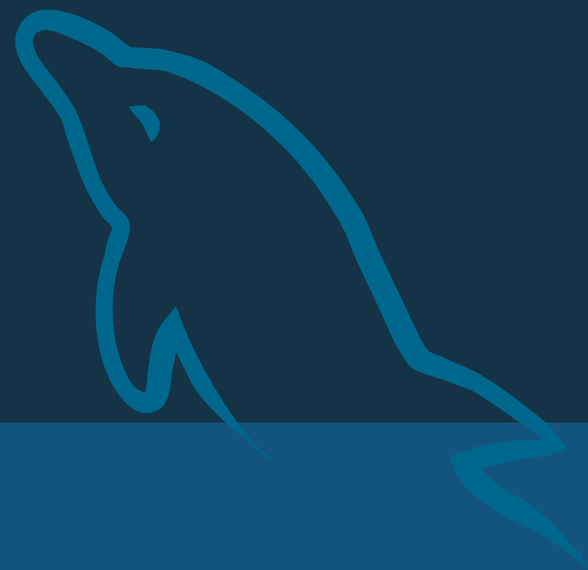
PHASE 2

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!!