SidCritique Project 2019

Data

Mickael ANJORAN

Alexis DUSART

Pierrick HERVE

Lina NICOLAIEFF

Raphael NJAMO KAPDEM

Camille ROGER

Robin VAYSSE

Adil ZOUITINE

CSV

CSV files are encoded in UTF8. The separator is the character "§". They can be easily used with the Pandas library.

Exemple for read a CSV file

```
import pandas as pd
ratings = pd.read_csv("ratings_V2.csv", sep="§", nrows=10,
        engine="python", encoding="utf-8")
```

CSV dictionary

Name	Type	Description	Source
USER			
user_id	Integer	User id	Sens Critique
gender_id	Integer	Sex of the user, probably: - 1 : Man - 2 : Woman	Sens Critique
date_create	Text	Creation date of the user account Format: AAAA-MM-DD	Sens Critique
date_birth	Numeric (Date)	Birth date of the user (AAAA-MM-DD)	Sens Critique

Name	Type	Description	Source
PRODUCT			
product_id	Integer	Product id	Sens Critique
title	Text	Title of the product	Sens Critique
storyline	Text	Brief summary of the scenario of the product	Sens Critique

date_release	Text	Release date of the product Format: AAAA-MM-DD	Sens Critique
year_of_production	Text	Production year of the product Format: AAAA	
subtype_id	Integer	Type of the product (probably: - 1 = Movie - 2 = Book - 4 = Serie	Sens Critique
authors	Text (list)	Author of the product	Sens Critique
genres	Text (list)	Gender of the product (eg: Thriller, Drama…)	Sens Critique
actors	Text (list)	Actors who played in the movie /serie	Sens Critique
actors_imdb	Text (list)	Actors who played in the movie /serie	IMDb
categories_imdb	Text (list)	Gender of the product (only for movies)	IMDb
id_JP	Integer	Id of the movie in the database of JPBoxoffice (if exists)	JP Boxoffice
code_country_imdb	Text (list)	Country code like Fr, En… representing the country of production of the movie	IMDb
director_imdb	Text (list)	Realisator(s) of the movie	IMDb
id_movie_imdb	Integer	ID of the movie in the database of IMDb	IMDb
imdb_note	Float	Average note on IMDb	IMDb
imdb_vote	integer	The number of users who voted on IMDb	IMDb
image_format	Text	Aspect ratio of movie's images (describes the proportional relationship between its width and its height)	IMDb
language_vo	Text	Original language of the movie	IMDb

long_summary	Text	More than 239 characters	IMDb
music_composer_imdb	Text (list)	Name(s) of the music composer(s) of the movie	IMDb
poster_url	Text	Url that refers to the poster of the movie	IMDb
producers_imdb	Text (list)	Producers of the movie	IMDb
name_companies_imdb	Text (list)	Names of the production companies	IMDb
production_country_imdb	Text (list)	Countries of production of the movie	IMDb
duration	integer	Duration of the movie in minutes	IMDb
short_summary	Text	Outlines describe the story (no longer than 239 characters)	IMDb
sound_format	Text	Audio format of the movie's sound	IMDb
synopsis_imdb	Text	Description of the entire plot of the title, including spoilers, so users who haven't seen a movie can read everything about the title. (exclusively details of the story of the film, with no commentary)	IMDb
imdb_top_movie	integer	The film's ranking in the top 250 of IMDb	IMDb
title_imdb	Text	Title of the movie found on IMDb	IMDb
writers_imdb	Text (list)	List of the script writers of the movies	IMDb
budget	Integer	Budget for the realisation (movies only)	IMDb and JP boxoffice
currency	Text	Currency type associated to budget, ('€', '$'...) (movies only)	IMDb and JP boxoffice

entries_fr	Integer	Total number of entries in France (movies only)	JP boxoffice
entries_paris	Integer	Total number of entries in France (movies only)	JP boxoffice
rentability	Integer	Percentage of rentability based on the budget (can be over 100) (movies only)	JP boxoffice
start_fr	Integer	Number of entries in France during the release day (movies only)	JP boxoffice
start_paris	Integer	Number of entries in Paris during the release day (movies only)	JP boxoffice

Name	Type	Description	Source
REVIEW			
review_id	Integer	Review id	Sens Critique
user_id	Integer	User id	Sens Critique
product_id	Integer	Product id	Sens Critique
title_review	Text	Title of the review	Sens Critique
content_review	Text	Review content of the product by the user	Sens Critique
date_review	Text	Date of the review Format: AAAA-MM-DD	Sens Critique

Name	Type	Description	Source
Wishes			
wish_id	Integer	Review id	Sens Critique
user_id	Integer	User id	Sens Critique
product_id	Integer	Product id	Sens Critique
date_wish	Text	Date of the review Format: AAAA-MM-DD	Sens Critique

Name	Type	Description	Source
RATING			
rating_id	Integer	User id	Sens Critique
user_id	Integer	Product id	Sens Critique
product_id	Integer	Rate gives by the user for the product	Sens Critique
rating	Integer		Sens Critique
date_rating	Text	Date of the rating Format: AAAA-MM-DD	Sens Critique

Name	Type	Description	Source
WEATHER			
date_weather	Text	Date of the measured weather Format: AAAA-MM-DD	Auto increment
temperature_min_*city*	Integer	Minimum temperature measured for a single day	Dark sky website
temperature_max_*city*	Float	Maximum temperature measured for a single day	Dark sky website
humidity_*city*	Float	Humidity for a single day	Dark sky website
visibility_*city*	Float	Visibility for a single days	Dark sky website
desc_weather_*city*	Text	Short description of weather for a single day	Dark sky website

Instead of "*city*", we can found "paris", "toulouse"," lyon", "lille", "marseille" or "bordeaux"

Database

The database runs on sqlite3.

Use the package sqlite3 to interact with it with python3.

Example of use

```python
import sqlite3
databaseFile = V2.db'
conn = sqlite3.connect(databaseFile)
cursor = conn.cursor()
cursor.execute('SELECT id_category, name_category FROM category')
rows = cursor.fetchall()
for row in rows:
    print(row)
cursor.close()
conn.close()
```

Database dictionary

Name	Type	Description	Source
USER			
id_user	Integer	User id	Sens Critique
sex	Integer	Sex of the user : - 0 : Man - 1 : Woman	Sens Critique
date_create	Numeric (Date)	Creation date of the user account (AAAA-MM-DD)	Sens Critique
date_birth	Numeric (Date)	Birth date of the user (AAAA-MM-DD)	Sens Critique

Name	Type	Description	Source
PRODUCT			
id_product	Integer	Product id	Sens Critique
title_product	Text	Title of the product	Sens Critique
pitch	Text	Brief summary of the scenario of the product	Sens Critique
date_release	Numeric (Date)	Release date of the product (AAAA-MM-DD)	Sens Critique
year_production	Numeric (Date)	Production year of the product (AAAA)	
type_product	Integer	Type of the product : - 1 = Film - 2 = Livre - 4 = Série	Sens Critique
duration	integer?	Duration of the movie in	IMDb

		minutes	
color	Text		IMDb
image_format	Text	Aspect ratio of movie's images (describes the proportional relationship between its width and its height)	IMDb
sound_format	Text	Audio format of the movie's sound	IMDb
award	Integer	Number of awards the movie won	IMDb
imdb_note	Real (float)	Rating given by IMDb	IMDb
imdb_top_movie	integer	The film's ranking in the top 250 of IMDb	IMDb
imdb_vote	integer	The number of users who voted on IMDb	IMDb
image_format	Text	Video image quality	IMDb
language_vo	Text	Original language of the movie	IMDb
long_summary	Text	More than 239 characters	IMDb
short_summary	Text	Outlines describe the story (no longer than 239 characters)	IMDb
synopsis	Text	Description of the entire plot of the title, including spoilers, so users who haven't seen a movie can read everything about the title. (exclusively details of the story of the film, with no commentary)	IMDb
sound_format			IMDb
poster_url	url	Url that refers to the poster of the movie	IMDb
budget	Integer	Budget for the realisation (movies only)	IMDb et JP boxoffice

currency	Text	Currency type associated to budget, (€, $...) (movies only)	IMDb et JP boxoffice
entries_fr	Integer	Total number of entries in France (movies only)	JP boxoffice
entries_paris	Integer	Total number of entries in France (movies only)	JP boxoffice
rentability	Integer	Percantage of rentability based on the budget (can be over 100) (movies only)	JP boxoffice
start_fr	Integer	Number of entries in France during the release day (movies only)	JP boxoffice
start_paris	Integer	Number of entries in Paris during the release day (movies only)	JP boxoffice

Name	Type	Description	Source
CATEGORY			
id_category	Integer	Category id	Sens Critique
name_category	Text	Category of the product, e.g. Drame, Comédie	Sens Critique (IMDb?)

Name	Type	Description	Source
PERSON			
id_person	Integer	Person id	Auto increment
name_person	Text	Person name	Sens Critique et IMDb

Name	Type	Description	Source
REVIEW			
id_user	Integer	User id	Sens Critique
id_product	Integer	Product id	Sens Critique
text_review	Text	Review content of the product by the user	Sens Critique
date_review	Numeric (Date)	Date of the review (AAAA-MM-DD)	Sens Critique
title_review	Text	Title of the review	Sens Critique

Name	Type	Description	Source
RATING			
id_user	Integer	User id	Sens Critique
id_product	Integer	Product id	Sens Critique
rate	Integer	Rate gives by the user for the product	Sens Critique
date_rating	Numeric (date)	Date of the rating (AAAA-MM-DD)	Sens Critique

Name	Type	Description	Source
PARTICIPATE			
id_product	Integer	Product id	Sens Critique
id_person	Integer	Person id	Auto increment
duty	Text	Job of the person: - 0 : Author - 1 : Actor - Music composer - Director - Producer - Writer	Sens Critique et IMDb

Name	Type	Description	Source
WEATHER			
id_city	Integer	City id: - 0 : Paris - 1 : Toulouse - 2 : Lyon - 3 : Lille - 4 : Marseille - 5 : Bordeaux	Auto increment
date_weather	Numeric (Date)	Date of the measured weather (AAAA-MM-DD)	Dark sky website
temperature_min	Float	Minimum temperature measured for a single day	Dark sky website
temperature_max	Float	Maximum temperature measured for a single day	Dark sky website
humidity	Float	Humidity for a single day	Dark sky website
visibility	Float	Visibility for a single days	Dark sky website
desc_weather	Text	Short description of weather for a single day	Dark sky website

Name	Type	Description	Source
COUNTRIES			
id_country	Integer	Country id	Auto increment
name_country	Text	Country name	IMDb
code_country	Text	Country code like Fr, En…	IMDb

Name	Type	Description	Source
COMPANIES			
id_company	Integer	Company id	Auto increment
name_company	Text	Name of the production company that participated in the film.	IMDb

Remarque :

La table WEATHER contient les données climatiques journalières des six grandes villes de France à savoir Paris,Toulouse,Lyon,Bordeaux,Marseille et Lille. Ces données sont entre le 08 juillet 2004 inclus et le 28 novembre 2018 inclus. Les requêtes, effectuées sur des dates non incluses dans cet intervalle de temps, ou sur une ville autre que celle citée, ne retourneront aucun résultat. Si jamais vous avez besoin de données climatiques à d'autres dates ou d'autres villes que celles fournies dans la table WEATHER, vous pouvez les avoir sur l'un des sites web : https://darksky.net, https://www.wunderground.com ou http://www.meteofrance.com.