**GitHub Documentation - Movie recommandation - Data Breakers**

**Access the GitHub**: <https://github.com/Sebastiao199/Project3MRS>

**Deliverables**

* A notebook containing the data exploration and cleaning and visualizations. You will **explain your cleaning choices and exploration findings in a document of your choice.**
* Present the exploratory analysis and the relevant KPIs
* A notebook for the Recommendation System step with the source code and your comments

**Project 3 description:**

The first part of the project

The second part of the project (“Machine Learning” folder) is a machine learning algorithm that returns a list of recommended movies based on movie names chosen by a user. These recommendations are integrated into a dashboarding tool from the command line ("input", streamlit, or other).

**Description of the 9 folders on GitHub:**

1. FinalColabFiles

**What is it?** Those are the final Google Colaboratory files that reply to our questions.

**Link**: <https://github.com/Sebastiao199/Project3MRS/tree/main/FinalColabFiles>

1. Machine Learning

**What is it?** Those are the code for the algorithms, the recommendation system and the input file.

**Link**: <https://github.com/Sebastiao199/Project3MRS/tree/main/MachineLearning>

1. Exploratory Data Analysis

**What is it?** This is our initial analysis on the different datasets.

**Link** : <https://github.com/Sebastiao199/Project3MRS/tree/main/FirstExploratoryDataAnalysis>

1. OtherTables rename > LinkingTables

**What is it?** Those are the tables we created to go from one file to another.

**Link**: <https://github.com/Sebastiao199/Project3MRS/tree/main/OtherTables>

1. .ipynb\_checkpoints > new name “Streamlit support files” : ask to Cristina

**What is it?** Support files from Streamlit

**Link:** <https://github.com/Sebastiao199/Project3MRS/tree/main/.ipynb_checkpoints>

1. TablesForStreamlit

**What is it?** Those are the final tables we used for Streamlit.

**Link**: <https://github.com/Sebastiao199/Project3MRS/tree/main/TablesForStreamlit>

1. OtherScripts > rename “Old”

**What is it?** We used these files before and kept them if necessary.

**Link:** <https://github.com/Sebastiao199/Project3MRS/tree/main/OtherScripts>

1. Pictures

**What is it?** This is the folder with the pictures we used for Streamlit.

Link: <https://github.com/Sebastiao199/Project3MRS/tree/main/Pictures>

1. StreamlitFinal.py

**What is it?** This is the final code for the Streamlit.

**Link** : <https://github.com/Sebastiao199/Project3MRS/blob/main/StreamlitFinal.py>

Contact

Jdsds

:

**RT - Kids movies & Clean dataset for join IMDB (Kaggle) : the kids**

the kids

**3) RT - Kids movies & Clean dataset for join IMDB (Kaggle)**

Goal:

Find the best movies for kids based on the top studios

IN:

-rotten\_tomatoes\_movies.csv

OUT:

- kids\_movies.csv (movies for kids based on top studios and rating-only Fresh/Certified Fresh movies)

- out4.csv (new column ‘dir\_mov’ for join with IMDB)

**4) RT-IMDB merge + Directors.ipynb (Collab)**

Goal:

Join RT and IMDb tables // Get director’s table for streamlit

IN:

- out4.csv

- tprn\_10

OUT:

- imdb\_rt\_ratings.csv

- directors\_rt\_imdb1.csv

**5) Kids movies - IMDb and RT.ipynb (Collab)**

Goal:

Join rating with kids table

IN:

- imdb\_rt\_ratings.csv

- kids\_movies.csv

OUT:

- kids\_imdb\_rt1.csv

**6) Viz - ratings .ipynb**

Goal: Statistics on ratings

IN:

- out4.csv

- tprn\_10.csv

- directors\_rt\_imdb1.csv

OUT (plots):

- IMDB ratings distribution

- Evolution of movies’ ratings

- RT vs IMDB ratings