PROOF OF WORKING

AND

PROOF OF UNDERSTANDING

**Student: Tosca Voogd**

**Student number: 500753992**

**Professor: Rob Loke**

**Master Digital Driven Business**

**Amsterdam University of Applied Sciences (AUAS)**

**SQL**

The sql files that belong to this are:

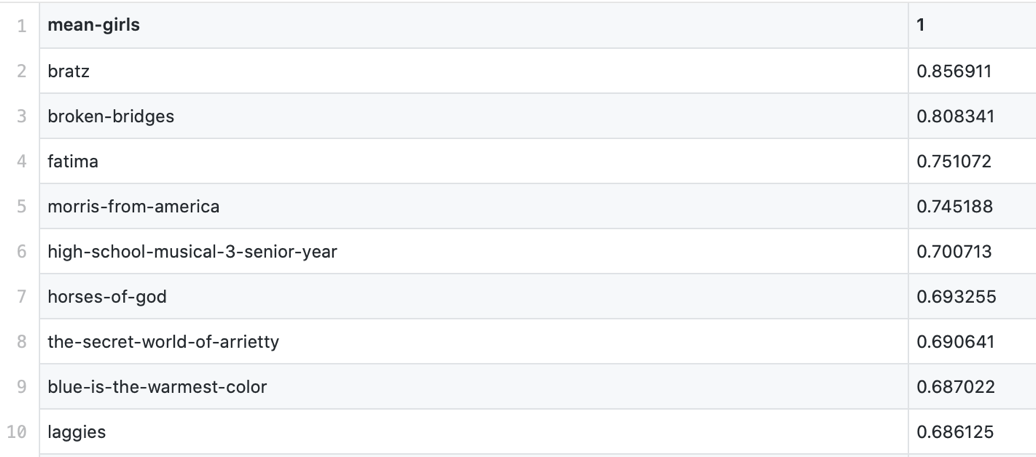
* Summary.sql
* title.sql
* Starring.sql

The output that belongs to this are:

* top50recommendationssummary.csv
* top50recommendationtitle.csv
* top50recommendationsstarring.csv

My favorite movie from the past are Mean Girls and Legally Blonde. The output for Mean Girls was better, so I chose to go for Mean girls.

I put comments in before the commands in the sql files to tell what I am doing. For all the recommender systems, I set the threshold at 0.01. I played with this, but if I wanted to get 50 recommendations, the threshold must be set at 0.01.

**Output: top 50 recommendations** 

Above, a screenshot of the top 10 recommendations is shown. In the whole dataset, the top 50 recommendations can be found.

Originally, Mean Girls is a movie for teen girls. That is why the results of the output look very logical to me. Names like: high school musical and bratz are also movies in the same genre.

**PYTHON**

The python file that belong here is:

* recommender.py

The output file that belongs here is:

* recommendationbasedonmetascore.csv

Also, for the Python part, I chose the movie Mean Girls.

**Output: top 50 recommendations**

Afbeelding met tafel

Automatisch gegenereerde beschrijving

Afbeelding met tekst

Automatisch gegenereerde beschrijving

This output is based on the metascore the authors gave on my favorite movie. To build this, Python is used. I did not have any experience with Python so, it took some effort to understand this. But thanks to Google and classmates who helped me out and explained it, I manage to get the outcome and understand the steps I made.

In the output you can see the recommended movies based on higher metascores that the same authors gave to other movies than my favorite movie Mean Girls. Here, the movie does not have to look like my favorite movie, but it is just based on the metascores the authors gave. This is also the reason why the movies in the recommender do not look like the movie I chose. It is just movies that the authors preferred over Mean Girls.

The steps I did:

1. Import the dataset with all the reviews.
2. Made a subset of my favorite movie Mean Girls.
3. Printed the subset so I saw how many reviews there are on Mean Girls.
4. Then with the data.columns.tolist(), I created the same variables for the subset as the userRevies.csv
5. Then I created the recommender dataframe (called recommendation). We want to make a RS based on the Author and the Metascore he or she gave. So first, I used the iterrows() to loop over the subset and create the filters author and metascore.
6. The filters 1 and 2 are made so the recommender dataframe has the movies in it where the same authors gave a higher metascore than on Mean Girls.
7. I made a new dataframe (possible\_recommendations) with the outcome of filter1 & filter2 and added the relative increase and abstract increase of the metascores.
8. After that, I appended the possible\_recommendations to the recommendation (final recommender dataframe).
9. In the end, I made a top 50 of the recommendations and copied this to a new csv file: recommendationsbasedonmetascore.csv.