SI507 Final Project Data Checkpoint

Chao-Yuan Cheng (crisscy)

Project Code

- https://github.com/CrissCheng/si507-final-project

Data Sources

1. list of country names and alpha2 code
    1. url: https://laendercode.net/en/2-letter-list.html
    2. format: html
    3. method of access: scraping, caching is used.
    4. summary of data:
        1. # of records available: 252
        2. # of records retrieved: 252
        3. description of records:
            1. Country name
            2. Country alpha2 code
        4. evidence of caching: Around line 179 of final.py, there's a function called 'scrape_request' where I use caching to scrape the website.
2. Musixmatch api
    1. url: https://developer.musixmatch.com/
    2. format: Json
    3. method of access: api call (request), caching is used.
    4. summary of data:
        1. # of records available: billions ? The api provides all songs from the history of time and there are different kinds of data available like lyrics, country charts, etc.
        2. # of records retrieved: 252*5*2 = 2520. For each country, it retrieves top 5 songs and the lyrics of those songs so it calls the api twice.
        3. description of records:
            1. track name
            2. album name
            3. artist name
            4. lyrics
            5. lyrics language
        4. evidence of caching: Around line 161 of final.py, there's a function called 'request_musixmatch' where I use caching to call the api.
3. Youtube data api
    1. url: https://developers.google.com/youtube/v3/getting-started
    2. format: Json
    3. method of access: api call (request), caching is used.
    4. summary of data:
        1. # of records available: quadrillion ? All the data in the youtube history
        2. # of records retrieved: 252*5*2 = 2520. For each country, it retrieves top 5 songs and the lyrics of those songs so it calls the api twice.
        3. description of records:
            1. youtube video id
            2. youtube video title
            3. video like counts
            4. video like counts
            5. video dislike counts
            6. video comment counts

4. evidence of caching: Around line 135 of final.py, there’s a function called ‘request_youtube’ where I use caching to call the api.

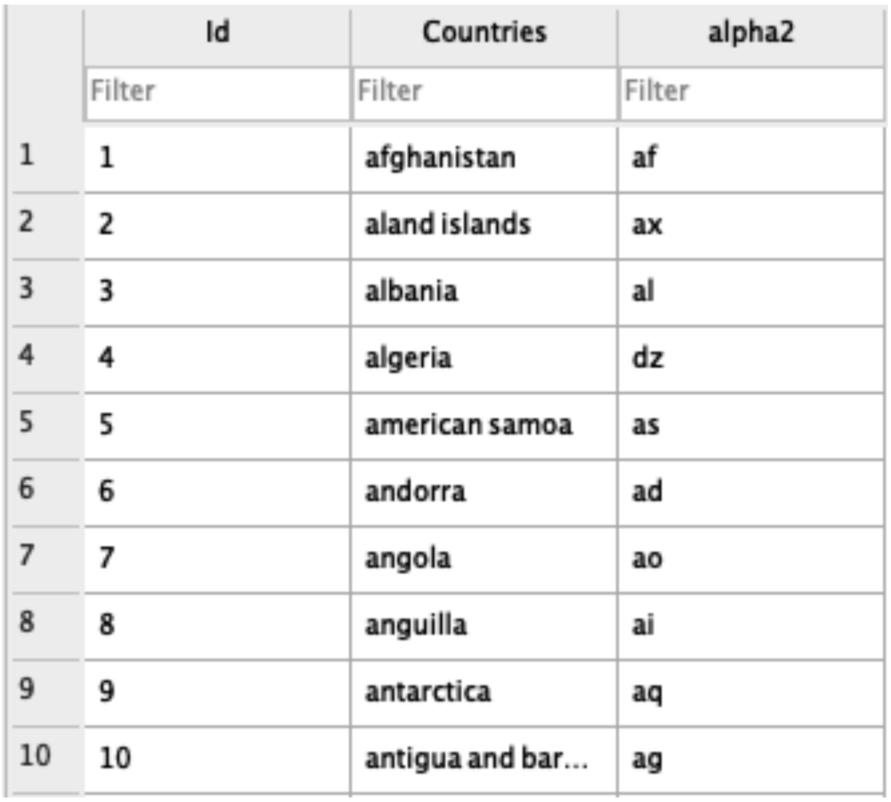
Database

1. Database schema
    1. Countries table

        ```
        CREATE TABLE IF NOT EXISTS "Countries" (
        'Id' INTEGER PRIMARY KEY AUTOINCREMENT,
        'Countries' TEXT NOT NULL,
        'alpha2' TEXT NOT NULL)
        ```

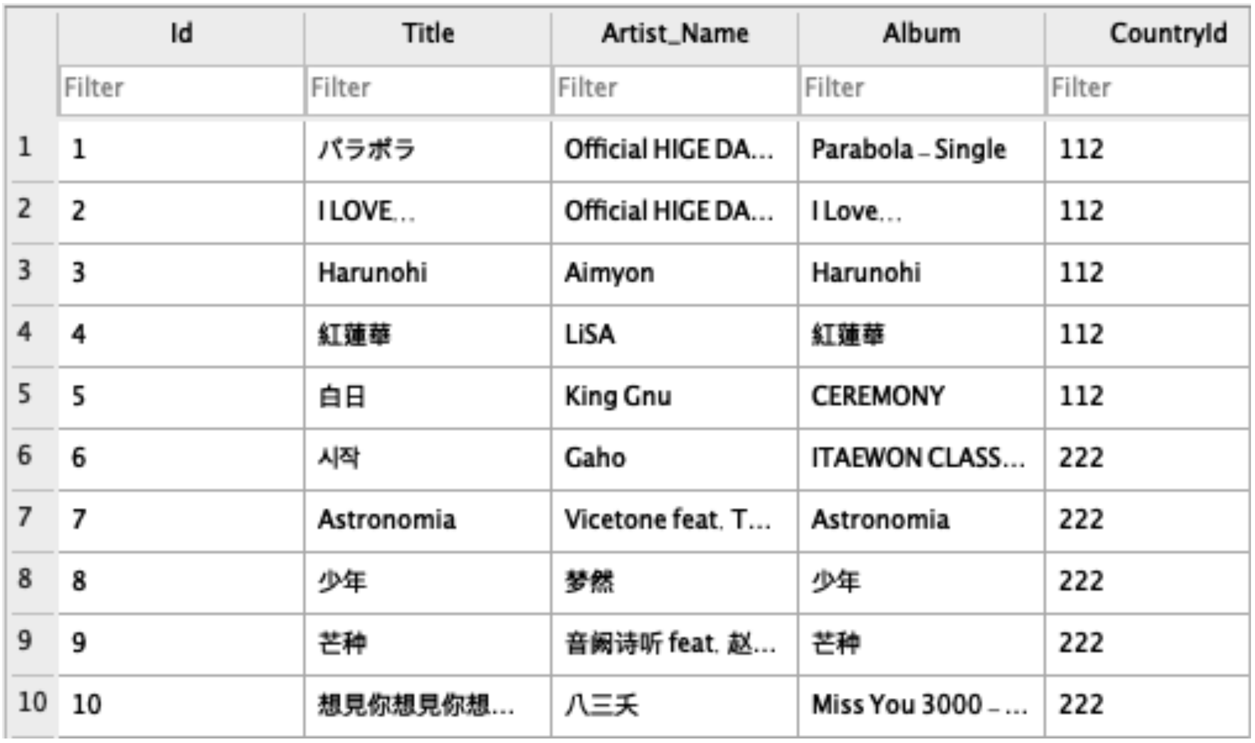
    2. Videos table

        ```
        CREATE TABLE IF NOT EXISTS "Videos"(
        'Id' INTEGER PRIMARY KEY AUTOINCREMENT,
        'Title' TEXT NOT NULL,
        'Artist_Name' TEXT NOT NULL,
        'Album' TEXT NOT NULL,
        'CountryId' INTEGER NOT NULL,
        'Lyrics' TEXT NOT NULL,
        'Url' TEXT NOT NULL,
        'Views' INTEGER NOT NULL,
        'Likes' INTEGER NOT NULL,
        'Dislikes' INTEGER NOT NULL,
        'commentCount' INTEGER NOT NULL
        ```
2. foreign key - primary key relations: Videos CountryId is corresponded to Countries Id
3. Screenshot
    1. Countries



	Id	Countries	alpha2
	Filter	Filter	Filter
1	1	afghanistan	af
2	2	aland islands	ax
3	3	albania	al
4	4	algeria	dz
5	5	american samoa	as
6	6	andorra	ad
7	7	angola	ao
8	8	anguilla	ai
9	9	antarctica	aq
10	10	antigua and bar...	ag

2. Videos



	Id	Title	Artist_Name	Album	CountryId
	Filter	Filter	Filter	Filter	Filter
1	1	パラボラ	Official HIGE DA...	Parabola – Single	112
2	2	I LOVE...	Official HIGE DA...	I Love...	112
3	3	Harunohi	Aimyon	Harunohi	112
4	4	紅蓮華	LiSA	紅蓮華	112
5	5	白日	King Gnu	CEREMONY	112
6	6	시작	Gaho	ITAEWON CLASS...	222
7	7	Astronomia	Vicetone feat. T...	Astronomia	222
8	8	少年	梦然	少年	222
9	9	芒种	音阙诗听 feat. 赵...	芒种	222
10	10	想見你想見你想...	八三夭	Miss You 3000 – ...	222

Interaction and Presentation Plans

1. High-level plan: I'm going to let users choose the country that they are interested in, and show the top 5 songs currently in that country. The user can further choose the details of the song that they like. These details include the song's view counts on youtube, compared with dislikes, and comments. It can also show the lyrics of the song.
2. Technologies: Flask and Plotly. I don't intend to use command line prompts.