**Final Project Report**

Shujie Li

**Project Code**

<https://github.com/lishujie18/507_final_project.git>

**Required Python Packages**

BeautifulSoup, requests, sqlite3, pandas, plotly

**Data sources**

1. Billboard website: [https://www.billboard.com/](https://www.billboard.com/Billboard)

* Billboard website contains various music information like charts, news, video, photos.
* Data format: HTML
* Access technique: Beautiful Soup HTML parser
* Caching is used when access the data.
* Data summary: Billboard data include top 3 popular charts (other charts are limited to premier user), and user can choose a chart to look at, then the system would print top 20 items on the chart.

The Billboard data are showed on the right.

Each billboard chart is a chart instant which has chart name, chart URL and detail ranking information.

And each ranking item on the chart is also an item instance which contain name, info, rank.

Text

Description automatically generatedText

Description automatically generated

Text

Description automatically generated

1. YouTube API: <https://developers.google.com/youtube/v3/docs/search/list>

* YouTube API can be used to search for videos matching specific search terms, topics, locations, publication dates, and much more.
* Data format: json
* Access technique: API request.
* Caching is used when access the data.
* Data summary: once user chooses a ranking item on a specific billboard chart, he can get top 20 view counts video related to that ranking item.

The YouTube data are showed on the right.

Text

Description automatically generatedEach video item is a video instance which contain name, videoed, view, counts like counts, and dislike counts

Graphical user interface, text

Description automatically generated with medium confidence

Text

Description automatically generated

1. Data retrieving with cache

Text

Description automatically generated

**Database**

1. Database schema:

Graphical user interface, text, application

Description automatically generatedCREATE TABLE IF NOT EXISTS "{name}" (

"Id" INTEGER PRIMARY KEY AUTOINCREMENT UNIQUE,

"VideoName" TEXT NOT NULL,

"VideoId" TEXT NOT NULL,

"Views" INTEGER NOT NULL,

"Likes" INTEGER NOT NULL,

"Dislikes" INTEGER NOT NULL)

\* table name is search term

(the name of the ranking item)

1. Database data:

The image on the right shows the data collected in the database.

**Interaction and Presentation**

1. Interaction:

The system will first show top 3 popular billboard chart, and user can choose on chart to look at. Then the system will display the ranking information about the chart (top 20 items), And user can choose an item to see the detail information.

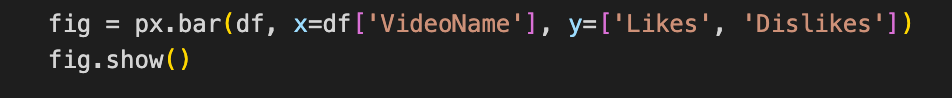
The detail information is the top 20 video information about the item, including the views, likes, dislikes of each video.

1. Presentation:

Graphical user interface

Description automatically generateda stack bar plot shows the like and dislike counts for each video.

1. Technology used: Plotly



**Demo Video**