DS532 Final Project Proposal Summer 2022: Sorting and Ranking Top TikTok Influencers Using Different Popularity Indexes

Eric McDermott, Simi Rajeev, Hai Jing(Jane) Tu

June 17, 2022

1 Project Summary

Our final project plan is to design a web app that allows users to query a Kaggle dataset on the top 1000 influencers on TikTok in the year 2020. The dataset is in a .csv format and can be downloaded from Kaggle.

Our database application will provide query functions that allow users to sort and update the popularity of top TikTok influencers using different measures, including subscribers count, average Views, average Likes, average number of Comments, and average Shares.

We will use Django and MySQL to implement the project.

2 Project Description

1. **Team**: We have three team members working on this project: Eric McDermott, Simi Rajeev, and Hai Jing(Jane) Tu.

2. **Objectives**: Our objectives include creating a relational database based on our dataset, providing sorting and ranking service using the database, and creating an interactive interface for users to interact with and update data in our database.

   Specifically, we will allow users to sort the data based on popularity and find out how Tiktok influencers are popular in different ways, in-

cluding most shared, most commented, most liked, etc.

In addition, we plan to allow users to update the database with the reason they believe that makes the TikTok influencer popular, for example, being funny, motivating, touching, surprising, creative, or being controversial, etc.

3. **Usefulness**: TikTok influencers are popular in different ways. They are known for different speciaties, such as skills, creativity, knowledge, or opinions. It would be interesting to see which influencers receive the most likes, and whether those who are most liked are also those who are most commented or most shared.

   Allowing users to sort and rank influencers with different indexes of popularity is useful for Tiktok Analytics and social media studies. Popular apps that allow analytics on Tiktok include tick Tik Meter, Tikbuddy, etc. Our app will be unique because it allows users to identify the reason for the popularity of influencers, and this will make the search more informative.

4. **Dataset**:

   The dataset origin is /urlhttps://hypeauditor.com/top-tiktok/. The title of the data is "Top 1000 TikTok Influencers Ranking." It contains 1000 records of TikTok influencers who are top ranked in the year 2020. There are 11 columns in the .csv file, including popularity measures and influencer identity columns. Figure 1 shows the table heading after loading it in `Python`.

```
In [14]: df = pd.read_csv('/Users/Htu1/Documents/IUB/DSCI-D532/Final Project/tiktok_top_1000.csv')
         df.head(2)
```

Out[14]:

	Country	Rank	Account	Title	Link	Subscribers count	Views avg.	Likes avg.	Comments avg.	Shares avg.	Scraped
0	All	1	billieeilish	BILLIE EILISH	https://www.tiktok.com/@billieeilish	41600000.0	95000000.0	18100000.0	151000.0	54800.0	2022-02-07 16:54:35.641971
1	All	2	badbunny	Bad Bunny	https://www.tiktok.com/@badbunny	13400000.0	33300000.0	7300000.0	101900.0	106800.0	2022-02-07 16:54:35.641971

Figure 1: Dataset Loaded in Python

Summary stats of the dataset show that there are **1000 records** in the table, but 2 of them have the same `Account` name. Therefore,

there are only **999 unique accounts** in this dataset.

It can be seen from the output summary that there are **two null values** in the `Title` field. Figure 2 shows part of the output summary below:

```
In [7]: df.isnull().sum()

Out[7]: Country              0
        Rank                 0
        Account              0
        Title                2
        Link                 0
        Subscribers count    0
        Views avg.           0
        Likes avg.           0
        Comments avg.        0
        Shares avg.          0
        Scraped              0
        dtype: int64

In [8]: len(df. index)

Out[8]: 1000

In [10]: df.Account.value_counts()

Out[10]: kimandnorth         2
         billieeilish        1
         mikasalamanca       1
         jaccleexs           1
         justthenobodys      1
                            ..
         notshanebennett     1
         39jeshi             1
         emillyvickof        1
         haileybieber        1
         davidgetial         1
         Name: Account, Length: 999, dtype: int64
```

Figure 2: Dataset Stats in Python

5. **Description of the functionalities**:

   Basic Functions: We allow users to read, sort, and update the data on top ranked TikTok influencers. Users will be asked to identify the reason why the top ranked influencers is popular when they update the dataset. Their answer will be recorded in a new column.

   Advanced Functions: We will provide visualization of popularity distribution and find out how to make it available through the webapp. We also hope to find out what kind of influencers are most liked, compared to wht kind of influencers tend to be most commented or shared.

6. **Communication and Sharing**: We will have weekly meeting at **7 pm on every Thursday** using Zoom. Our group files will be uploaded to github @T3T4. We will collaborate in the Github folder https://github.com/T3T4/DS532-Summer2022.

7. **Milestones**:

   Week 6 - project description (Group work finalized by Jane)

   Week 8 – database (Group work will be finalized by Eric)

   Week 10 - web app mock (Group work will be finalized by Simi)

   Week 11 - short video presentation (Group work will be finalized by Jane)

   Week 12 - full demo (Group work to be finalized by ALL)