

SI 618 Project Proposal 2

Ping-Lun Lai (alanlai)

1. Summarize and motivate your proposed project

The project will focus on analyzing the K-pop idols and music. In recent years, the Hallyu (Korean Wave) has a significant influence in Asia. As a K-pop lover, I have listened and watched plenty of K-pop music video from YouTube. Thus, I am interested in interpreting the growth of the K-pop music.

2. Datasets

The dataset I'm going to use is from Kaggle ([K-Pop Database \(1992-2020\) | Kaggle](#)), which include all kinds of information about K-pop idols and groups such as Name, Date of Birth, Group Name and Gender. Besides, because there is no information about music videos' views in the 'kpop_music_videos.csv' file, I'll use the YouTube API to collect the views data.

3. Analysis

Group Analysis

a. Boy Groups and Girl Groups Amount

For each year, I'll calculate the total amount of the new debut groups and plot these. The x-axis will be the year and the y-axis will be the amount. This might help me visualized the time when the K-pop starts flourishing.

b. Country Variance Analysis

For all the groups in the dataset, I'll calculate the percentage of the members who are not Korean, count the average per year, and plot them to see the trend.

c. Group Size Analysis

I'll group the boy and girl groups separately by their debut year and check the average member amount. By plotting this, it can show how the audience like the size of group since companies will always try to pander to the demand.

Music Videos Analysis

- d. For each boy groups and boy solo artists, count the percentage of the growth of the music videos and get the average of them in different years. Then use a linear regression to predict the growth percentage of them.
- e. For each girl groups and girl solo artists, do the same analysis above.
- f. To see how the album type (Major, Minor, Japanese, English, etc.) affect the views count, group all the data by type and year then plot the average growth percentage.
- g. Group these music videos by the release date (3 months), count the average videos length and plot them to see the pattern of the music length that audience tend to like. Use a linear regression to predict the music videos length in the future (2020-05-22 ~).
- h. Create a scatter plot of the debut year and the total average views count for groups that have at least one music videos over 100,000,000 to see the distribution of these famous K-pop groups.