DSCI 510: Principles of Programming Informatics
Professor: Gleb Satyukov
Final Project Report
Jing Chen-1338977098

Analyzing Genre Dominance and Release Timing in East Asian Music Project

1. Introduction

Every year, East Asia publish tens of thousands of songs. Therefore, I am curious about the current dominant song category. Additionally, I aim to explore the relationship between artist followers and top track popularity. To conduct this investigation, I've obtained two datasets from Kaggle, a popular platform for data science. Utilizing Python, I plan to employ code for reading and processing these datasets comprehensively. This analytical approach will unveil insights into the evolving musical landscape of East Asia and shed light on the dynamics between release dates and the ascent of top tracks in popularity.

2. Data and Data Processing

i. Data Collection

For the project, I use two main sets of data: one is about artists, called east_asia_top_artists, and the other is about music tracks, named east_asia_top_tracks. These files contain a lot of useful information. The artist data shows the names of singers or bands from East Asia, how many people follow them on Spotify, their music styles, and how popular their most famous songs are, which shown in the Figure 1. Also, the track data in the Figure 2 gives details about their hit songs, like when each song came out and how much people like it. Both of two datasets have 700 rows in total.

To get these files, I first had to sign up on a website called Kaggle, where people share a lot of different data. After signing up, Kaggle gave me a special username and a key to get into their system. With this key, I could go to a specific page on Kaggle (here: https://www.kaggle.com/datasets/crxxom/spotify-popular-east-asian-artists-and-tracks) and download the data I needed for my project. This made it easy to get all the information quickly and start working on understanding what makes these artists and songs so popular.

	artist_name	popularity	followers	artist_link	genres	top_track	top_track_album	top_track_popularit
0	BTS	88	67507448	https://open.spotify.com/artist/3Nrfpe0tUJi4K4DXYWgMUX	['k-pop', 'k-pop boy group', 'pop']	Left and Right (Feat. Jung Kook of BTS)	CHARLIE	88.
1	BLACKPINK	82	43385244	https://open.spotify.com/artist/41MozSoPIsD1dJM0CLPjZF	['k-pop', 'k-pop girl group', 'pop']	Shut Down	BORN PINK	79.
2	TWICE	79	18278225	https://open.spotify.com/artist/7n2Ycct7Beij7Dj7meI4X0	['k-pop', 'k-pop girl group', 'pop']	MOONLIGHT SUNRISE	READY TO BE	75.
3	j-hope	70	15407840	https://open.spotify.com/artist/0b1slQumIAsNbqAolCiSpy	['k-pop', 'k-rap']	on the street (with J. Cole)	on the street (with J. Cole)	83.
4	V	66	14088149	https://open.spotify.com/artist/3JsHnjpbhX4SnySpvpa9DK	['j-division', 'korean ost']	Christmas Tree	Our Beloved Summer (Original Television Soundtrack), Pt. 5	84.
5	Stray Kids	82	11990935	https://open.spotify.com/artist/2dlgFjaIVxs4ThymZ67YCE	['k-pop', 'k-pop boy group', 'pop']	S-Class	5-STAR	89.
6	SEVENTEEN	78	9731900	https://open.spotify.com/artist/7nqOGRx0Xj7N2JYbgNEjYH	['k-pop', 'k-pop boy group']	Super	SEVENTEEN 10th Mini Album 'FML'	87.
7	EXO	75	9701818	https://open.spotify.com/artist/3cjEqqelV9zb4BYE3qDQ4O	['k-pop', 'k-pop boy group']	Cream Soda	EXIST – The 7th Album	83.
8	Jung Kook	84	8656152	https://open.spotify.com/artist/6HaGTQPmzraVmaVxvz6EUc	['k-pop']	Seven (feat. Latto) (Explicit Ver.)	Seven (feat. Latto)	97.

Figure 1. east_asia_top_artists dataset

	song_name	album_name	album_link	artist_name	popularity	release_date
0	Cupid - Twin Ver.	The Beginning: Cupid	https://open.spotify.com/album/5letLUZIFsQikJYShfGNs4	FIFTY FIFTY	98	2023-02-24
1	Seven (feat. Latto) (Explicit Ver.)	Seven (feat. Latto)	https://open.spotify.com/album/53985D8g3JcGBoULSOYYKX	Jung Kook	97	2023-07-14
2	Like Crazy	FACE	https://open.spotify.com/album/4xc3Lc9yASZgEJGH7acWMB	Jimin	96	2023-03-24
3	MONEY	LALISA	https://open.spotify.com/album/4ASxFYWyk2216OloHoaSh8	LISA	96	2021-09-10
4	Take Two	Take Two	https://open.spotify.com/album/3jeQDa9OFZ6GndLindHx3k	BTS	95	2023-06-09
5	Like Crazy	Like Crazy (Remixes)	https://open.spotify.com/album/639nejcoHHwxJCKqr35ww2	Jimin	93	2023-03-26
6	Super Shy	NewJeans 'Super Shy'	https://open.spotify.com/album/5V729UqvhwNOcMejx0m55I	NewJeans	93	2023-07-07
7	Seven (feat. Latto) (Clean Ver.)	Seven (feat. Latto)	https://open.spotify.com/album/53985D8g3JcGBoULSOYYKX	Jung Kook	93	2023-07-14

Figure 2. east_asia_top_tracks dataset

ii. **Data Cleaning**

Since there are some information which are irrelevant to my project, I refined the raw data by selecting key columns such as “artist_name”, “followers”, “top_track”, “top_track_popularity”, “top_track_release_date”, “genres”, and “query_genre”. When I processed the code without definition formula, I found that there are four artists were missing value: Yui Aragaki, French Kiss, Rino Sashiara, Eriko Tamura. To address this, I tried to the artist data within the track data. However, Yui Aragaki had the track information available, and other three artists did not have which meant that their records could not contribute to a precise analysis. Therefore, I decided to exclude these incomplete entries from the dataset. I utilized libraries such as pandas for data manipulation and JSON for data storage. It allowed for an efficient workflow, where the necessary data was loaded, filtered, and then saved in a JSON format for further use. (Figure 3)


```
0:
  artist_name:            "BTS"
  followers:              67507448
  top_track:              "Left and Right (Feat. Jung Kook of BTS)"
  top_track_popularity:   88
  top_track_release_date: "2022-10-06"
  genres:                 "['k-pop', 'k-pop boy group', 'pop']"
  query_genre:            "j-pop"
1:
  artist_name:            "BLACKPINK"
  followers:              43385244
  top_track:              "Shut Down"
  top_track_popularity:   79
  top_track_release_date: "2022-09-15"
  genres:                 "['k-pop', 'k-pop girl group', 'pop']"
  query_genre:            "j-pop"
2:
  artist_name:            "TWICE"
  followers:              18278225
  top_track:              "MOONLIGHT SUNRISE"
  top_track_popularity:   75
  top_track_release_date: "2023-03-10"
  genres:                 "['k-pop', 'k-pop girl group', 'pop']"
  query_genre:            "j-pop"
```


Figure 3. Cleaned data in JSON format

3. Analysis

In order to find out the most popular type of song, the first step is to load the data and identifying the top 50 tracks based on their popularity scores. Also, I sorted all the tracks by the column top_track_popularity in descending order. This means that the tracks with the highest popularity are placed at the top of the DataFrame. After sorting, the function then selects the top 50 tracks by using the head (50) method,

which takes the first 50 rows of the DataFrame. These rows represent the 50 most popular tracks according to the dataset. This filtering was crucial to pinpoint the tracks that have been most impactful to listeners. Once these tracks were selected, the next step was to standardize the genre labels. Recognizing the array of genre descriptors used on Spotify, which can often include sub-genres and overlapping terms, I implemented a normalization process that consolidated genre variations into broader, recognizable categories, specifically targeting “"k-pop” and “j-pop”. I attached the result of analysis step below in Figure 4.

Figure 4. Analyzed Data in JSON format

4. Results

At first, I created a bar chart categorize by genres, and I set x label as ‘Count Type of Genres’, y label as ‘Type of Genres’, and the title as ‘Distribution of Genres’. (Figure 5) It is obvious that many types of genres contain two kinds instead of one, and the count type of k-pop and j-pop were less than actual number.

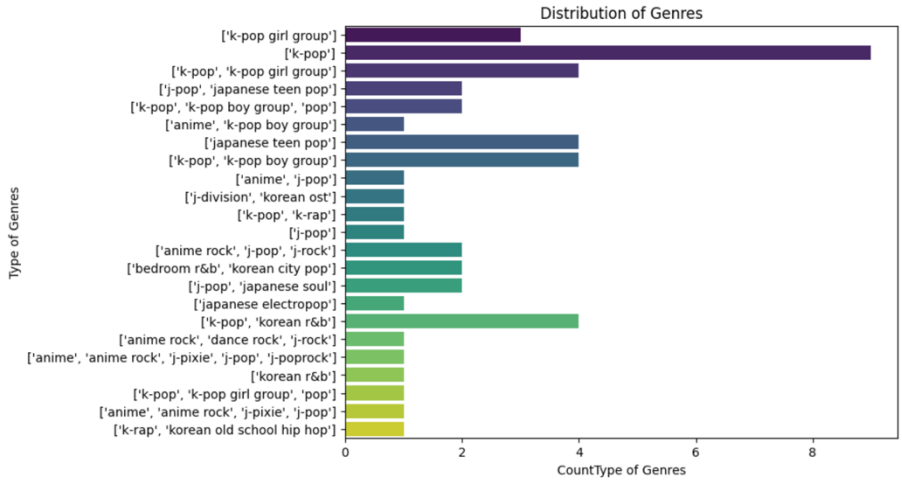

Figure 5. Distribution of Genres

We could tell that k-pop is the most popular type of song in East Asia. However, there are many songs have two types of genre classification, I need to separate the type in each genre. Except for k-pop, j-pop seems like the second popular type of song, and I will count the type of song to make sure how many k-pop and j-pop song list in the top 50. As the result, there are 29 songs belongs to k-pop genre, 10 songs belong to j-pop genres, and rest of 11 songs belong to other type of genres. Based on the result, K-pop is the most popular genre of music in East Asia, followed by J-pop as the second most popular The visualization result is attached in the Figure 6.

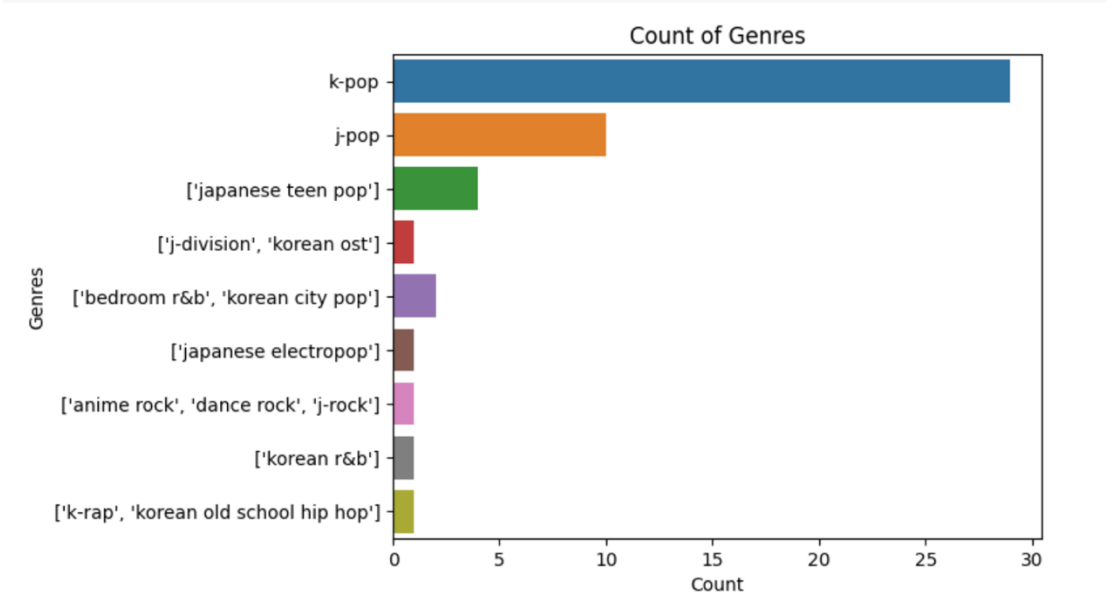

Figure 6. Count of Genres Bar Chart

I created a scatter plot to analyze the relationship between the number of artist followers and the popularity of their top tracks, setting “followers” as the x-axis, “top_track_popularity” as the y-axis, and “genres” as the hue. However, I found no correlation between the number of artist followers and the popularity of their top tracks

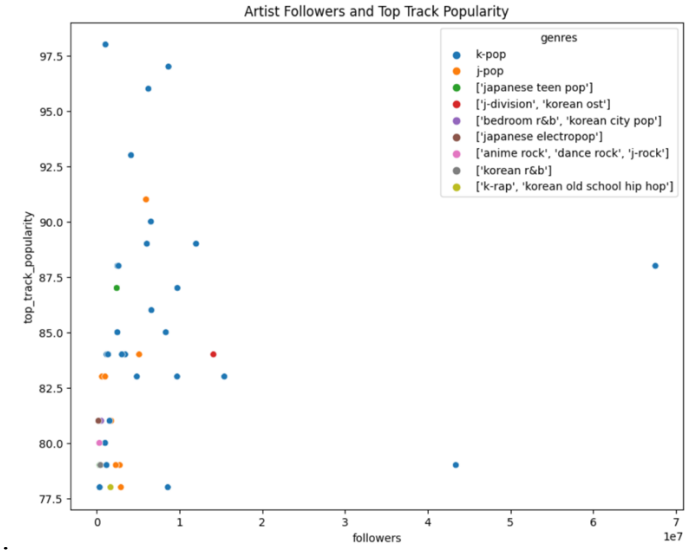

Figure 8. Count of Genres Bar Chart

5. Future Work

With additional time, I would deeper the project in temporal analysis to track the evolution and endurance of music trends. This would involve scrutinizing data over extended periods to discern long-term patterns in genre popularity and artist success, as well as seasonal variations in listening habits. Understanding these temporal dynamics could inform predictions about future industry trends and guide artists and producers in making strategic decisions about releases and marketing efforts, ultimately enabling a more proactive rather than reactive stance in a rapidly changing music landscape.