Final Project Proposal

# **Introduction**

This project focuses on the analysis of QQ Music and Spotify, one of the largest music streaming platforms in China, and the, largest music platform. The project aims to explore key trends in music consumption, including user preferences, sentiment analysis of song comments, and geographic distribution of music plays. Additionally, this project seeks to answer questions like: What are the most popular types of music over the past 20 years on QQ Music and Spotify? How do user sentiments reflect their engagement with music? How does QQ Music's data compare to Spotify's international data in terms of user engagement and genre preferences?

# **Datasets**

For this analysis, we will collect data from two sources: QQ Music and Spotify. The QQ Music data will be obtained by using the QQ Music API. This data will include song collection information, user comments, artists, song categories, and play counts over the past 20 years. The dataset size is expected to be substantial, potentially containing thousands of songs with detailed metadata such as song name, artist, genre, play count, comment count, and user sentiment scores. For Spotify, we will retrieve data from a publicly available dataset on Kaggle, which includes song plays, user engagement, and global trends. The Spotify dataset will allow us to compare global music preferences with Chinese trends.

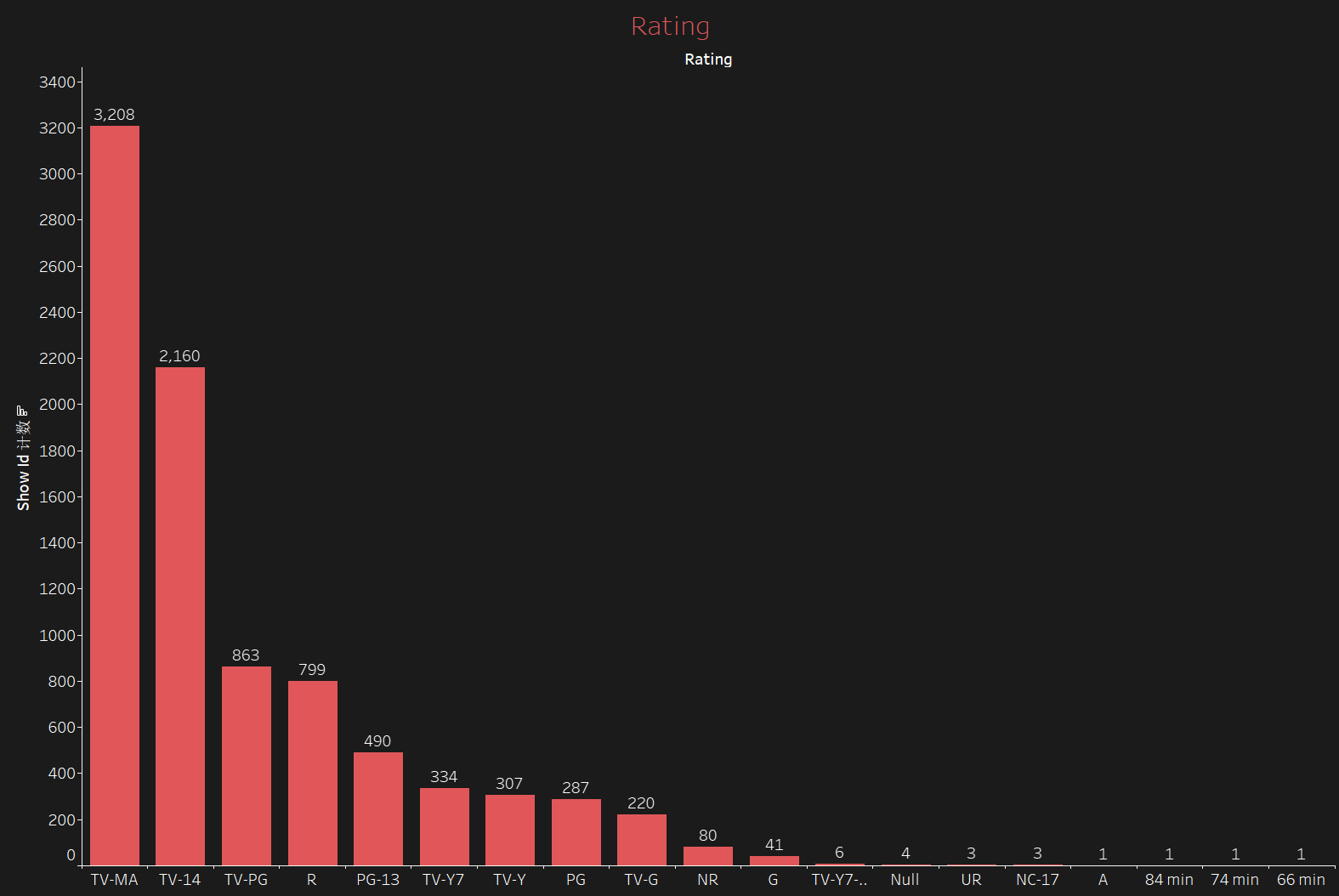
# **Analysis and Visualization Methods**

The analysis will be conducted using Python with libraries such as pandas, Translator and seaborn for data processing and visualization. Text mining techniques will be used to perform sentiment analysis on user comments. Tableau will be employed for creating interactive visualizations. The visualizations will include heatmaps to show the distribution of song plays over time, word clouds for displaying popular keywords in comments, pie charts to show genre distribution, and trend lines to depict changes in song popularity over time. Geographic maps will be used to illustrate the distribution of song plays across regions.

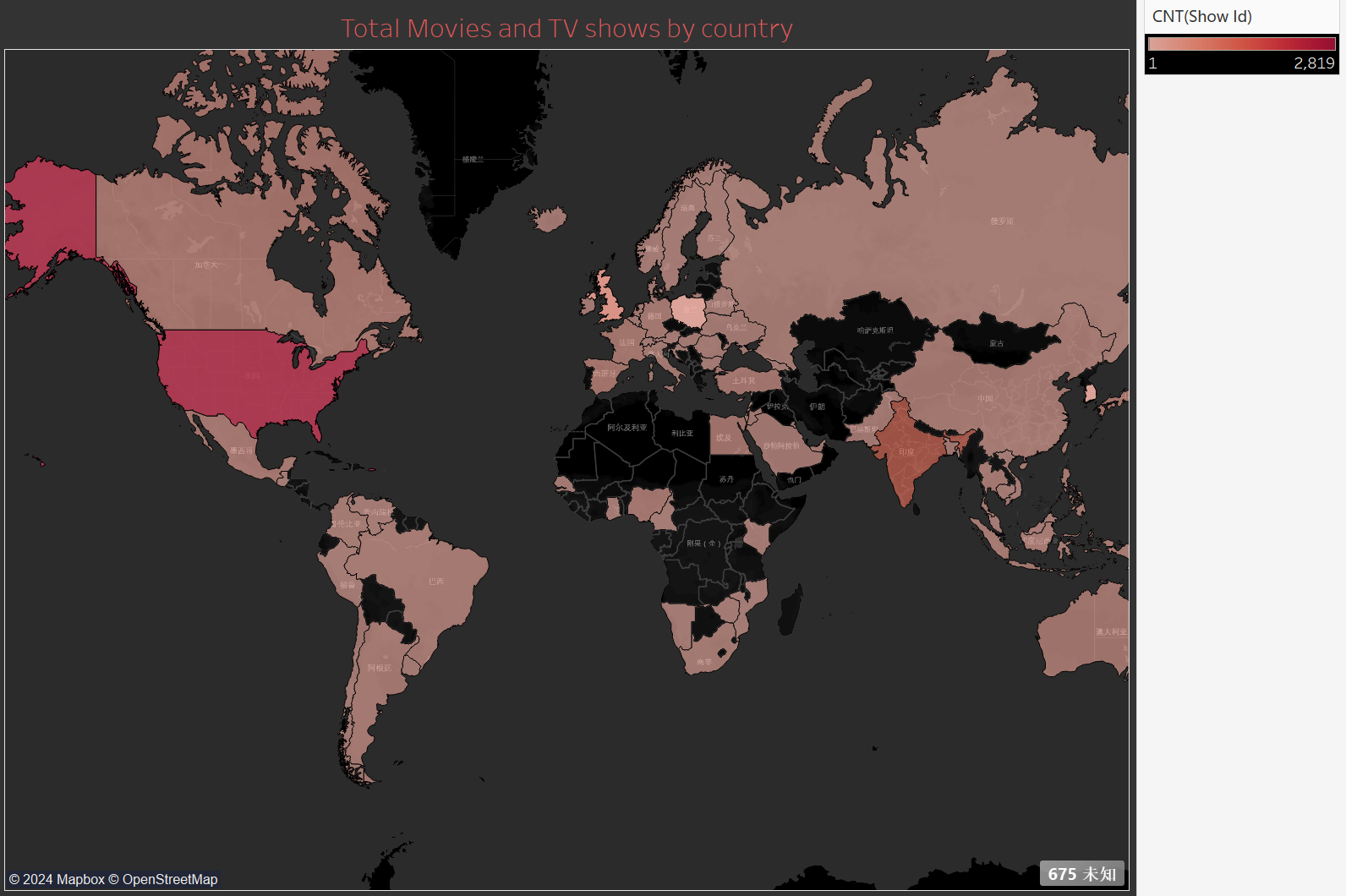
# **Sketches or References**

The visualizations will be designed based on the following ideas:

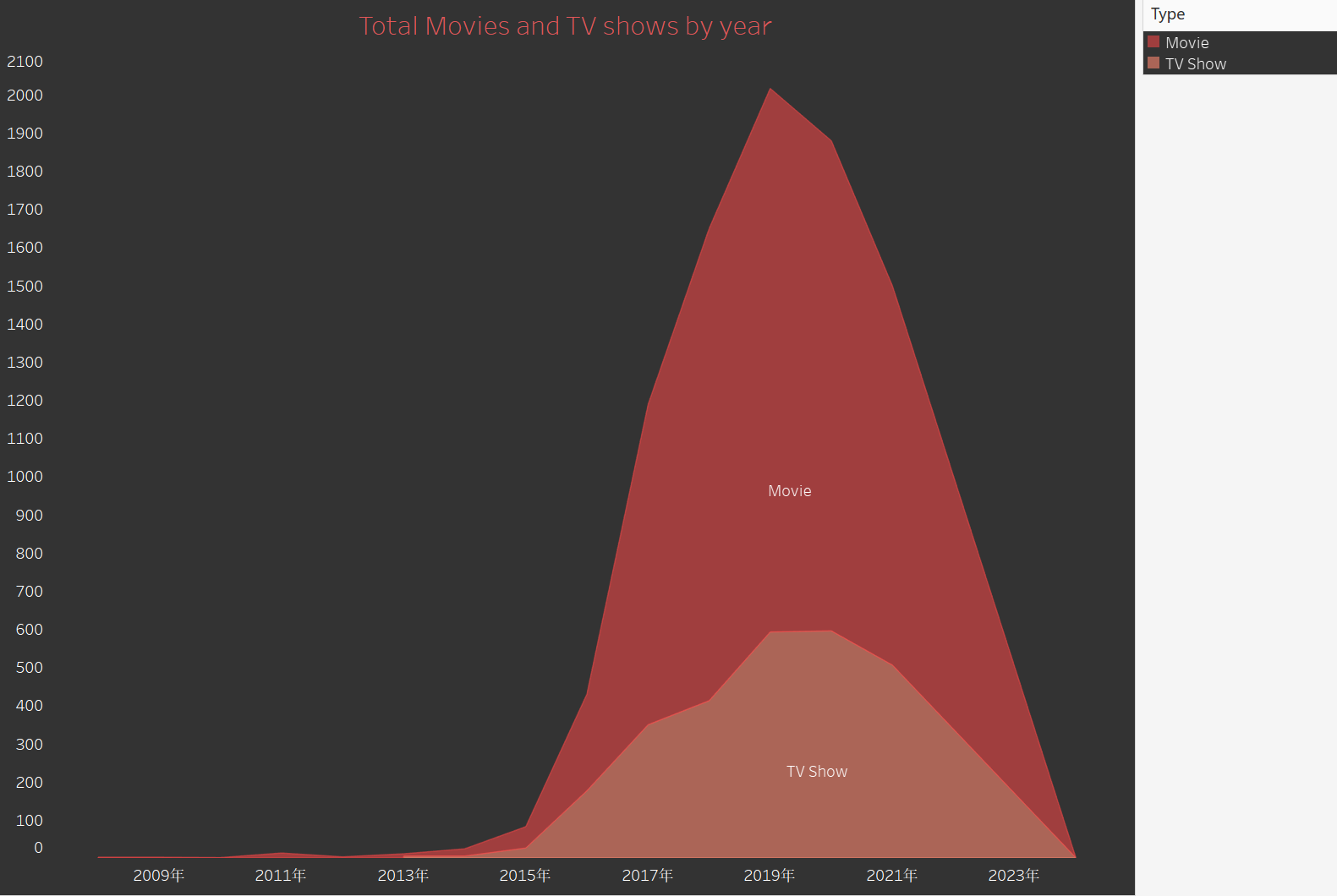
Barchart:



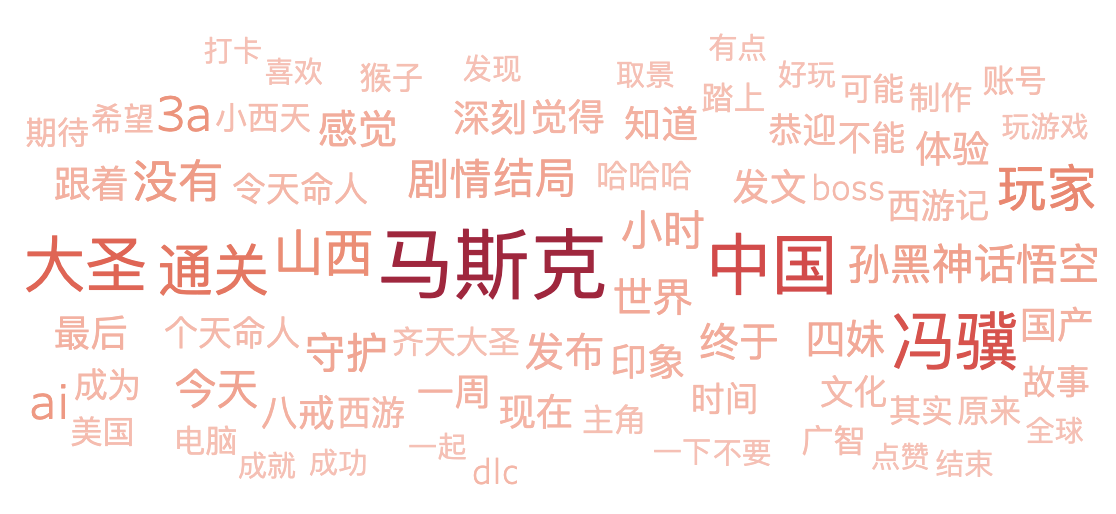
Heat Map:



Trend Line:



Word Cloud:



Reference: <https://www.dnainfo.com/new-york/20160229/bushwick/live-music-heatmap-shows-which-genres-are-being-performed-where>   
- A word cloud that represents the most frequent words in user comments and the sentiments they reflect, answering 'What are users' common themes and emotions when engaging with different songs?'

Reference: <https://www.uxforthemasses.com/word-clouds/>   
- A geographic map to display the distribution of song plays across different regions in China and globally, answering 'How do user preferences vary by location?'

Reference: <https://www.reddit.com/r/geography/comments/151yxz3/im_trying_to_collect_a_song_from_every_country/>

- A pie chart that visualizes the distribution of genres in QQ Music and Spotify, providing insight into the diversity of music preferences on each platform.

Reference: <https://trackify.am/tools/spotify-pie-charts>

- A trend chart showing the growth or decline of song play counts over time to capture the temporal trends in user engagement.

Reference: <https://www.highcharts.com/demo/highcharts/line-chart>

Reference: <https://www.kaggle.com/discussions/general/438299>

Reference: <https://www.kaggle.com/datasets/meeratif/spotify-most-streamed-artists-of-all-time>

Reference: <https://www.kaggle.com/datasets/meeratif/spotify-most-streamed-songs-of-all-time/data>

Reference: <https://www.kaggle.com/datasets/meeratif/spotify-top-artists-by-monthly-listeners>

# **Roles and Responsibilities**

Each group member will be responsible for distinct tasks:  
 -Miantong Zhang: Data collection and scraping from QQ Music API and Data cleaning.  
 - Lisha Qu: Data collection and scraping from Spotify Music Data cleaning and preprocessing using Python.

**Deliverables for the Interim Presentation**

For the interim presentation, we will deliver:  
 - A cleaned and preprocessed dataset from QQ Music and Spotify.  
 - Build up simple visualizations with accompanying analysis of trends, statistics, and geographic distributions.

# **Timeline and Milestones from Week 5 to Week 7**

Week 5: Data collection completed and initial data cleaning.  
Week 6: Sentiment analysis on QQ Music comments and comparison with Spotify trends.  
Week 7: Refinement of visualizations and further analysis of user engagement across platforms. Finalization of all visualizations and preparation of the final report.