Proposal

2025.01.29
We created this process book and found the dataset, here is the link:
https://www.kaggle.com/datasets/nelgiriyewithana/most-streamed-spotify-songs-2024

Project Abstract: This project looks at Spotify streaming data to find music trends and improve user experience. We will use the "Most Streamed Spotify Songs 2024" dataset from Kaggle. This dataset includes details like song name, artist, number of streams, release date, and audio features (such as danceability, energy, and mood). Our goal is to find the most popular artists and songs, study trends in music characteristics, and see how different song features relate to streaming numbers. We will also create an interactive dashboard to display these insights. This will help users find popular music and give the music industry useful information about what people like to listen to. Some challenges include working with a large dataset and making sure the analysis is accurate. In the end, we will build a tool that makes exploring music trends easy and fun.

2025.02.06
We finalized two foundational documents: the Team Agreement and the Project Plan. These documents will serve as guiding frameworks to ensure effective collaboration and project management throughout our work.
We also planned our next meeting at reading week.

Team Agreement

Team Agreement

1. Communication

- **Primary Channel**: Use **WeChat** for daily discussions and updates.
- **Meeting Schedule**: Weekly team meetings **after lecture** to review progress and discuss blockers. Notes will be documented in a **shared Google Doc** (Google Docs' history will track changes).
- **Urgent Matters**: Use **phone calls** for immediate concerns.

2. Task Management

- **Tracking Progress**: Use a **Progress Book** to document progress, deadlines, and challenges.
- **Updates**: Team members must log their task progress in the **Progress Book** after significant updates.

3. Code Guidelines

- Branch naming convention: feature/[ticket-number]-description (e.g., feature/123-login-page)
- All complex code requires inline comments explaining the logic. Functions longer than 20 lines must have detailed docstrings. Comment on any workarounds or technical debt.

4. Version Control

- Main branch protected - no direct commits
- Create feature branches from latest main
- Commit messages must be descriptive and reference ticket numbers

5. Quality Standards

- **Code & Documentation**:
    - Code must follow the agreed-upon **style guide**.
    - **Google Docs will track changes** to ensure updates are recorded.
- **Testing & Deployment**:
    - No failing tests in the `main` branch.
    - All major updates must be **reviewed and tested** before merging.

6. Conflict Resolution

- If disagreements arise, the team will **discuss the issue in the next weekly meeting**.
- If urgent, an **immediate discussion via phone call** will be scheduled.
- If unresolved, the team will seek advice from the **project instructor/mentor**.

7. Tasks Assign

	Haochen Ding	Yinuo Yang	Xinyue Li
1. Data Collection & Cleaning	Data exploration, normalization, and feature engineering	Document dataset details, source information, and limitations	Import & preprocess dataset (handle missing values, duplicates, data types)
2. Data Analysis	Statistical analysis on trends (e.g., most streamed songs, artists)	Correlation analysis between song features and streaming numbers	Write up analysis summaries and insights
3. Data Visualization	Generate static charts	Ensure visualizations are user-friendly & write explanations	Create interactive visualizations
4. Website Development	Structure setup Frontend & Backend		
5. Presentation Preparation	Create one part of the vis. presentation	Create one part of the vis. presentation	Create one part of the vis. presentation
6. Project Management	Manage GitHub repo, version control Ensure code quality, review PRs Track progress in Progress Book, update meeting notes		

Team Members

- **Haochen Ding**
- **Yinuo Yang**
- **Xinyue LI**

Agreement Date: 2025.2.5
Signed:

- **Haochen Ding**
- **Yinuo Yang**
- **Xinyue Li**

Project Plan

Spotify Data Visualization Project Plan

1. Basic Information

- **Project Title**: Exploring Spotify Streaming Trends: Popularity, Features, and Listener Preferences
- **Team Name**: D3 Dominators
- **Team Members & Email**:
  - **Haochen Ding** – haochen.ding@mail.utoronto.ca
  - **Yinuo Yang** – yinuo.yang@mail.utoronto.ca
  - **Xinyue Li** – hecateyue.li@mail.utoronto.ca

---

2. Background and Motivation

With millions of users worldwide, **Spotify** provides rich insights into music preferences and trends. Our team is interested in how different musical characteristics influence a song's popularity. By analyzing Spotify streaming data, we aim to:

- Identify **top artists and songs** based on streaming numbers.
- Examine how features like **danceability, energy, and tempo** correlate with a song's success.
- Provide an **interactive visualization** that helps users explore these trends.

Our motivation stems from an interest in **data-driven decision-making** in the music industry and how insights from big data can **shape music recommendations and marketing strategies**.

---

3. Related Work

We were inspired by:

- **Spotify's Year in Review** – A popular feature that provides users with personalized streaming insights.
- **Previous Kaggle Projects** – Several analyses of Spotify datasets inspired our focus on feature-based trend analysis.

- **Class Discussions** – The importance of **data storytelling** in visualizations helped us decide to build an interactive dashboard.

---

4. Data

- **Data Source**: Kaggle dataset – "Most Streamed Spotify Songs 2024" (https://www.kaggle.com/datasets/nelgiriyewithana/most-streamed-spotify-songs-2024)
- **Data Collection**:
  - The dataset includes song **name, artist, number of streams, release date,** and **audio features** like danceability, energy, tempo, valence, and loudness.
  - If necessary, we may use **Spotify's API** to fetch additional data.

---

5. Data Cleanup

Expected Cleanup Tasks

- Handling **missing values** (e.g., removing incomplete records).
- Converting **release dates** into usable formats.
- Standardizing **numeric features** (e.g., normalizing audio features between 0-1).
- Store cleaned data in **a structured JSON/CSV format** for easy integration into the web app.

Derived Quantities

- **Artist Popularity Score**: Total streams per artist.
- **Feature Impact Analysis**: Correlation between song attributes (danceability, energy) and streaming numbers.

Map

Potential Audiences:

1. **Music Enthusiasts and Spotify Users** – People who love discovering trends and exploring new music.
2. **Music Industry Professionals** – Artists, producers, and marketers who want to understand what makes a song popular.
3. **Data Analysts and Researchers** – People interested in music data for academic research or analytics.

Target Audience Selection: Music Industry Professionals

- **What do they know?**
  - They have a solid understanding of music production and market trends but may not be familiar with advanced data visualization or statistical methods.
- **What are their interests?**
  - Identifying what makes a song popular, understanding listener preferences, and predicting future music trends.
- **What visualization literacy do they have?**
  - Basic to moderate understanding of visualizations such as bar charts, line charts, and scatter plots, but prefer clear and intuitive dashboards over complex statistical graphs.
- **At what level of detail will you present information to them?**
  - Provide a high-level summary with the option to dive deeper into specific trends and song features, keeping technical jargon minimal.

Interesting Questions for Your Audience (At least 10 Questions)

1. Which platform provides the highest reach for top-ranked songs?
2. What is the relationship between TikTok views and Spotify streams?
3. Do songs with high Shazam counts have better streaming performance?
4. How does a song's playlist count affect its overall popularity?
5. Which artists consistently appear in the top ranks across all platforms?
6. Are songs released in specific months more successful?
7. What is the average track score for the top 50 songs?
8. How does YouTube engagement (likes and views) relate to Spotify playlist reach?
9. Do explicit tracks perform better or worse on different platforms?
10. What is the distribution of all-time rank among the top 100 songs?
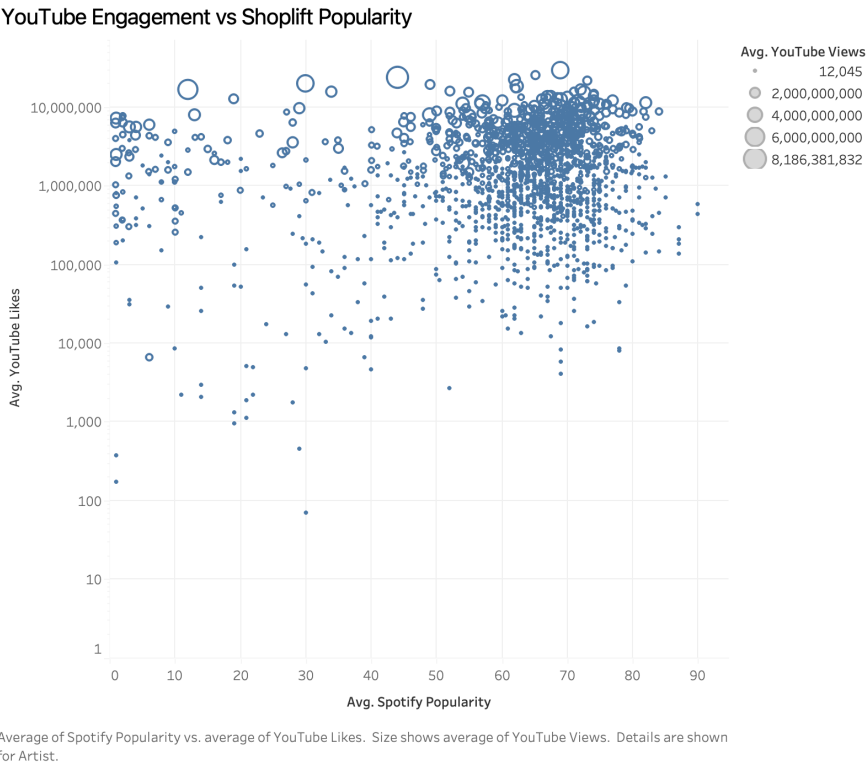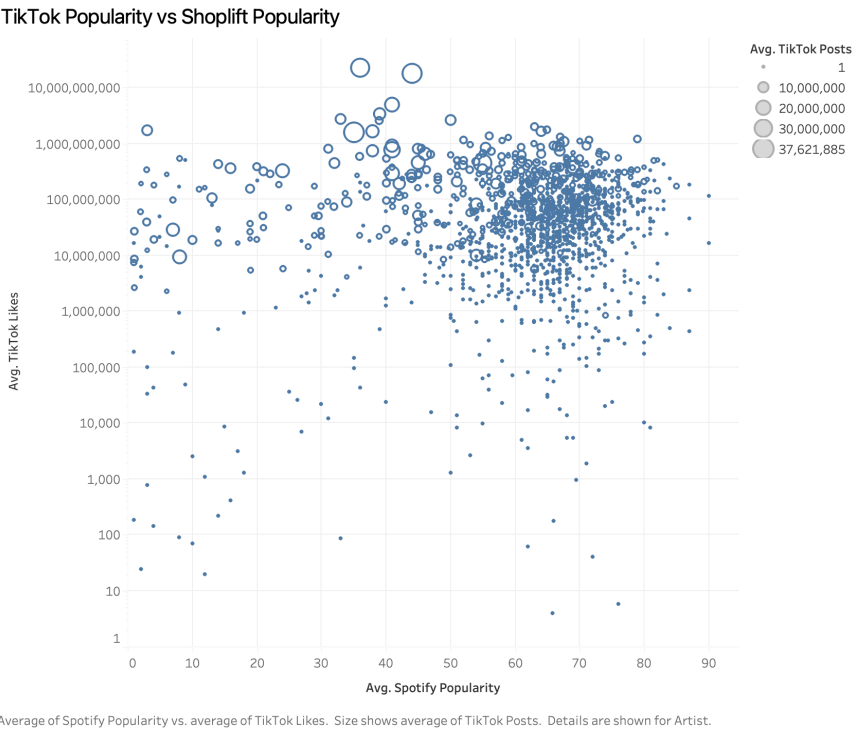
Dataset Attribute Breakdown and Data Types

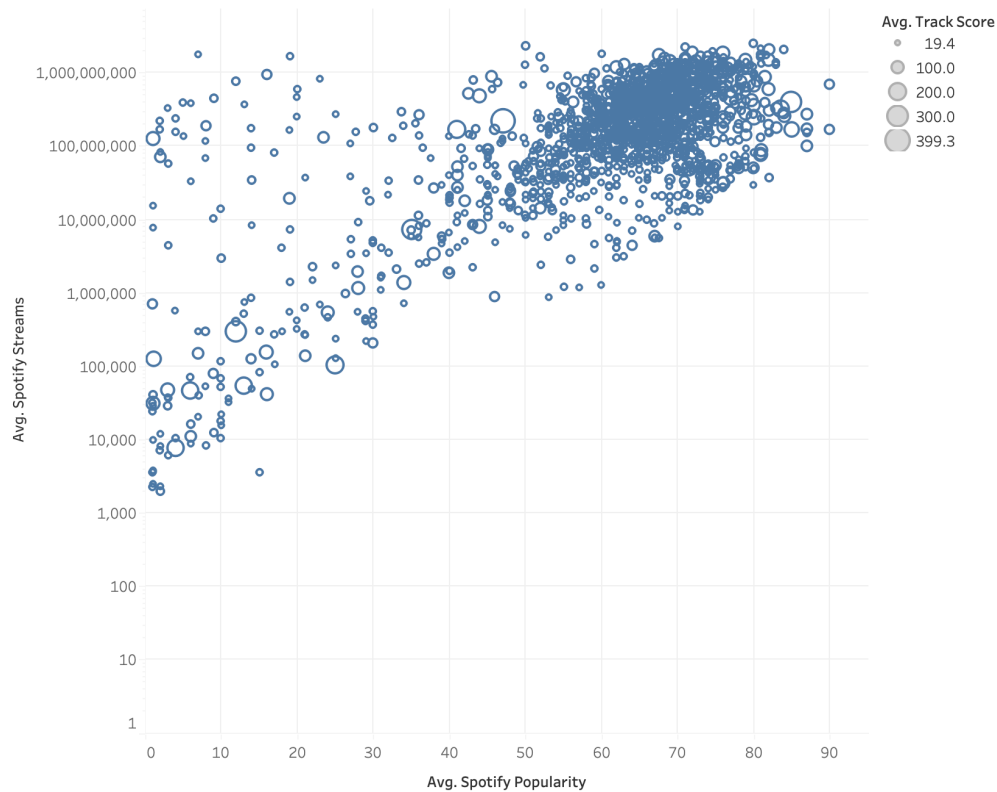| Attribute	Description	Data Type
Track Name	Name of the song	Categorical
Album Name	Name of the album the song belongs to	Categorical
Artist	Name of the artist(s) of the song	Categorical
Release Date	Date when the song was released	Ordinal (Date)
ISRC	International Standard Recording Code	Categorical
All Time Rank	Ranking of the song based on all-time popularity	Ordinal
Track Score	Score assigned to the track based on various factors	Quantitative
Spotify Streams	Total number of streams on Spotify	Quantitative
Spotify Playlist Count	Number of Spotify playlists the song is included in	Quantitative
Spotify Playlist Reach	Reach of the song across Spotify playlists	Quantitative
Spotify Popularity	Popularity score of the song on Spotify	Quantitative
YouTube Views	Total views of the song's official video on YouTube	Quantitative
YouTube Likes	Total likes on the song's official video on YouTube	Quantitative
TikTok Posts	Number of TikTok posts featuring the song	Quantitative
TikTok Likes	Total likes on TikTok posts featuring the song	Quantitative
TikTok Views	Total views on TikTok posts featuring the song	Quantitative
YouTube Playlist Reach	Reach of the song across YouTube playlists	Quantitative
Apple Music Playlist Count	Number of Apple Music playlists the song is included in	Quantitative

AirPlay Spins	Number of times the song has been played on radio stations	Quantitative
SiriusXM Spins	Number of times the song has been played on SiriusXM	Quantitative
Deezer Playlist Count	Number of Deezer playlists the song is included in	Quantitative
Deezer Playlist Reach	Reach of the song across Deezer playlists	Quantitative
Amazon Playlist Count	Number of Amazon Music playlists the song is included in	Quantitative
Pandora Streams	Total number of streams on Pandora	Quantitative
Pandora Track Stations	Number of Pandora stations featuring the song	Quantitative
Soundcloud Streams	Total number of streams on Soundcloud	Quantitative
Shazam Counts	Total number of times the song has been Shazamed	Quantitative
TIDAL Popularity	Popularity score of the song on TIDAL	Quantitative
Explicit Track	Indicates whether the song contains explicit content	Categorical (Boolean)

Tableaus

Xinyue Li

TikTok Popularity vs Shoplift Popularity


Average of Spotify Popularity vs. average of TikTok Likes. Size shows average of TikTok Posts. Details are shown for Artist.

YouTube Engagement vs Shoplift Popularity

Average of Spotify Popularity vs. average of YouTube Likes. Size shows average of YouTube Views. Details are shown for Artist.

Spotify Streams vs Shoplift Popularity

Average of Spotify Popularity vs. average of Spotify Streams. Size shows average of Track Score. Details are shown for Artist.

The questions answered in Tableau focused more on platform-specific engagement metrics, such as comparing TikTok likes, YouTube views, and Spotify streams in relation to Shoplift popularity. While these questions offered valuable insights into the relationship between platform engagement and overall popularity, they differed from some of the broader, trend-based questions the team initially proposed, such as analyzing patterns across genres, release dates, or explicit content. The shift occurred because platform-based metrics were more measurable and visually impactful in Tableau. These questions also provided clearer, more actionable insights for the target audience—music industry professionals. While broader questions remain valuable for identifying long-term trends, the Tableau visualizations helped answer specific, data-driven questions more effectively. Thus, the selected questions were a better fit for the available dataset and visualization capabilities.
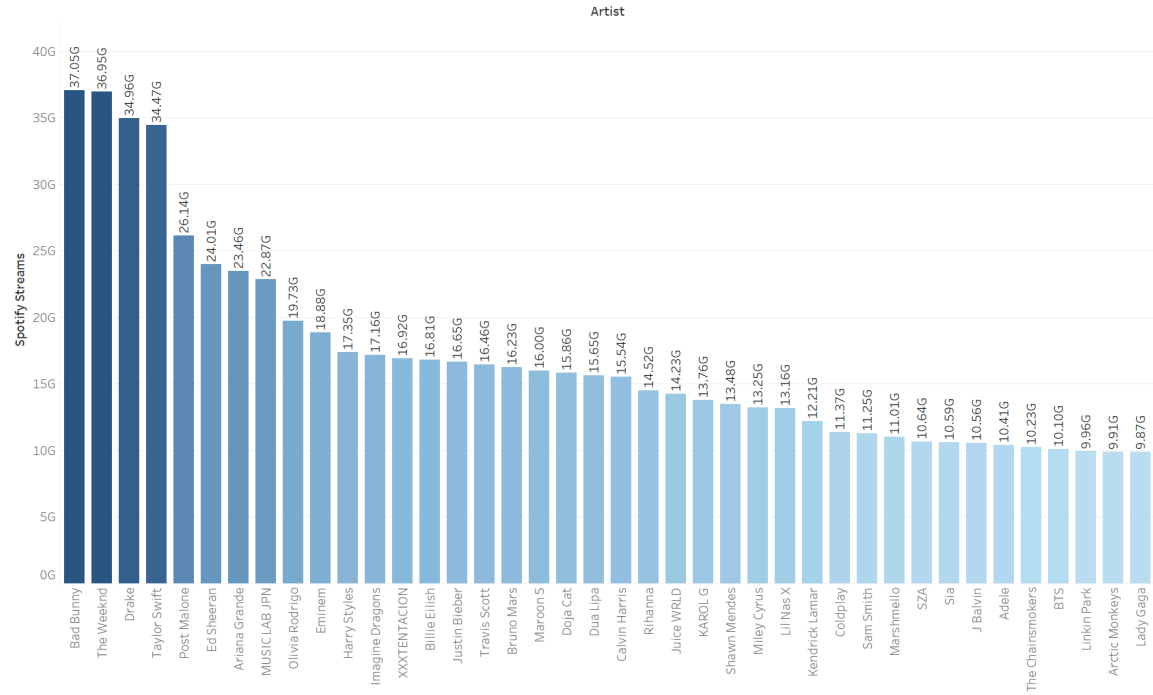
Haochen Ding:

Top 40 Total Spotify Streams Artist
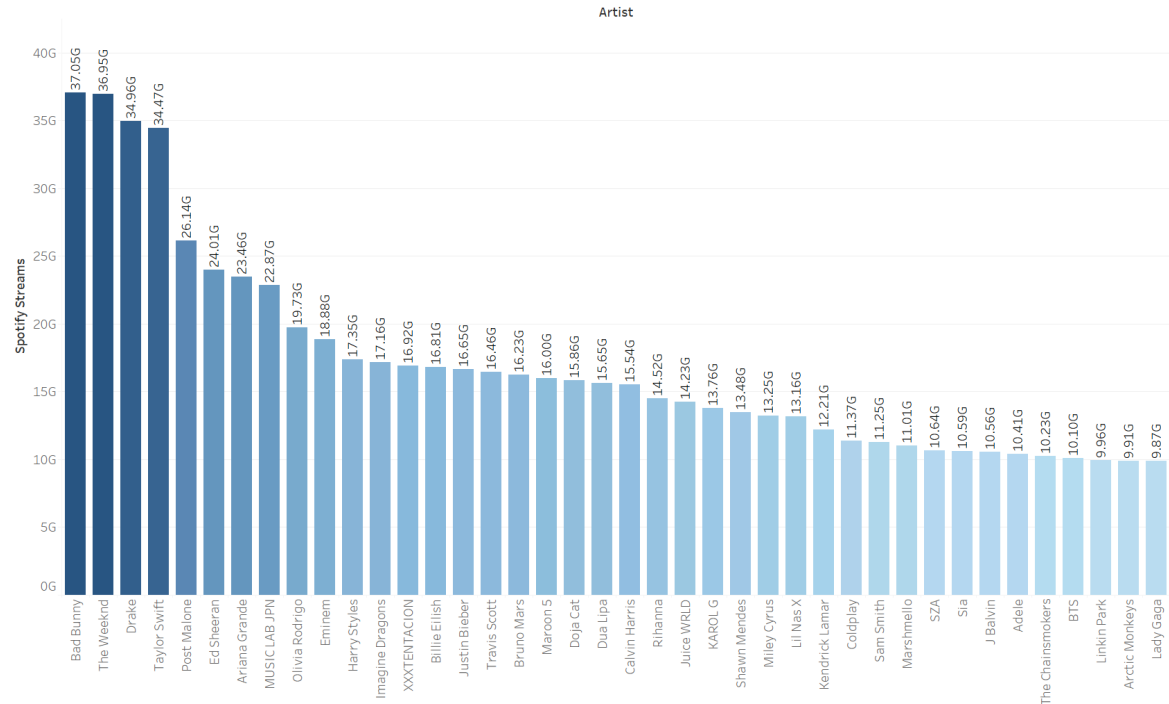
Artist

Sum of Spotify Streams for each Artist. Color shows sum of Spotify Streams. The marks are labeled by sum of Spotify Streams. The view is filtered on Artist, which keeps 40 of 2,000 members.
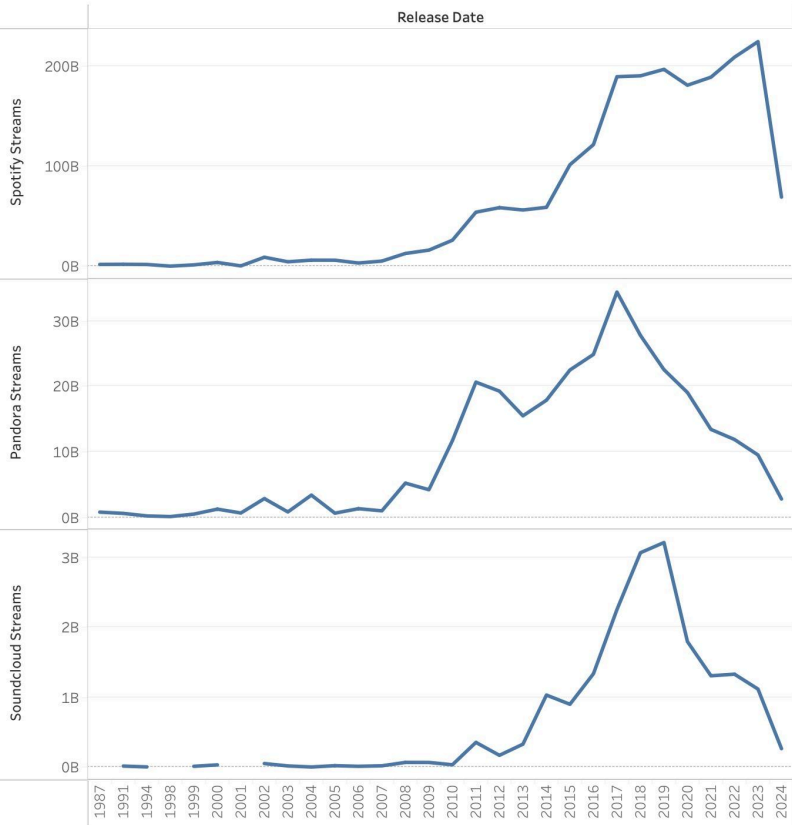
Top 40 Total Spotify Streams Artist

Artist

Sum of Spotify Streams for each Artist. Color shows sum of Spotify Streams. The marks are labeled by sum of Spotify Streams. The view is filtered on Artist, which keeps 40 of 2,000 members.

Songs' Playlist Reach vs. Spotify Streams
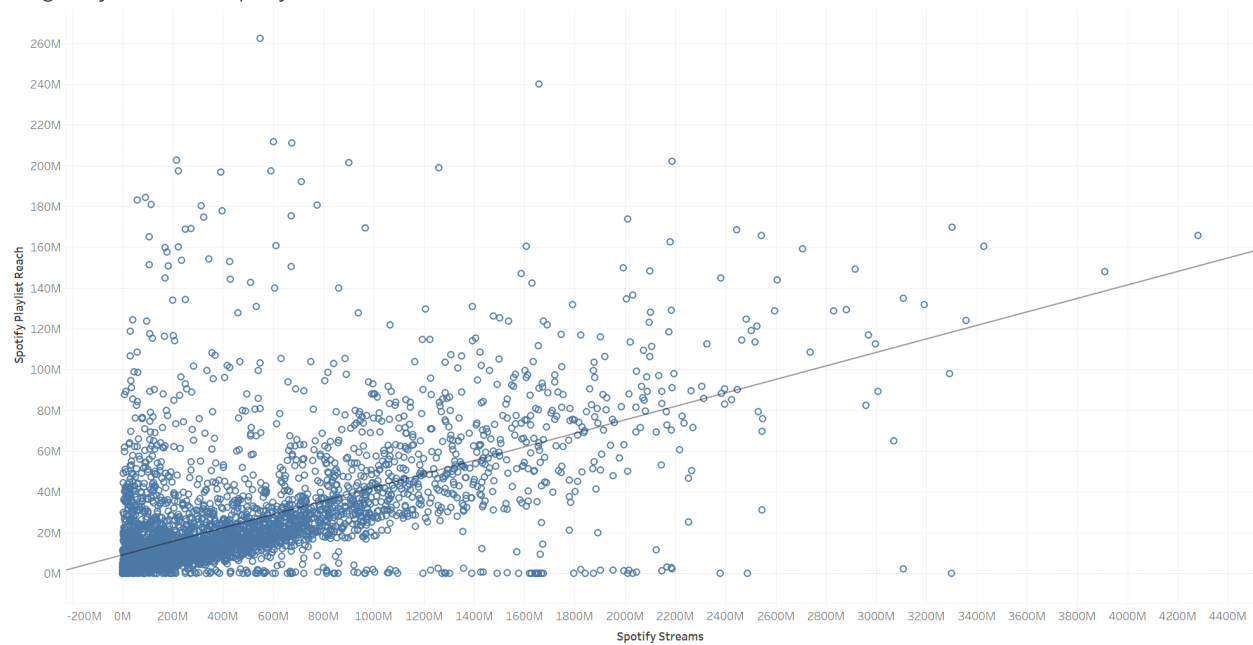
Spotify Streams vs. Spotify Playlist Reach.

The questions we answered in Tableau were different from our team's original questions because of the data we had, the types of charts we could make, and what would be most useful for our audience. At first, we wanted to compare multiple streaming and social media platforms, look at how TikTok views affect streams, and see if explicit songs perform better. But in my Tableau visualizations, I focused on simpler trends that we could actually analyze, like the most-streamed artists and tracks and the connection between playlist reach and Spotify streams. We decided to create clear and useful visualizations that music industry professionals could easily understand. Even though we didn't answer all of our original questions, some of them are still important and could be explored in the future with more data.
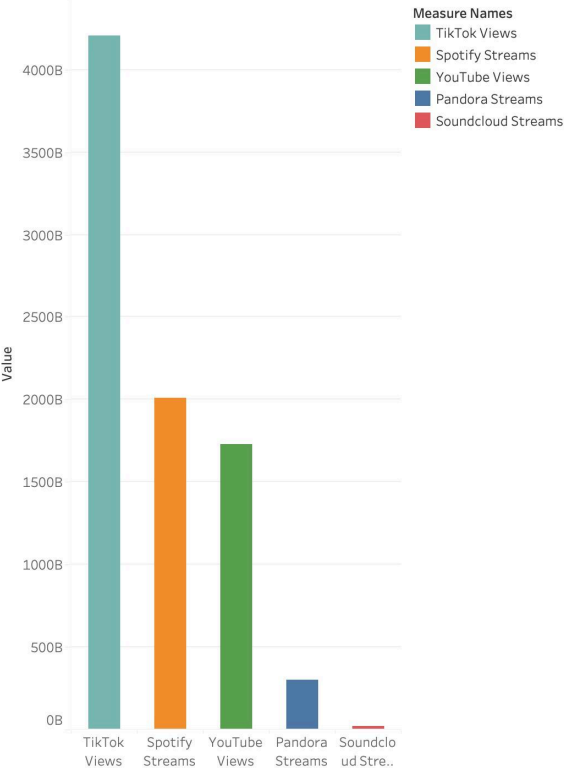
Yinuo Yang:

Sheet 1

Release Date

The trends of sum of Spotify Streams, sum of Pandora Streams and sum of Soundcloud Streams for Release Date Year.

Sheet 2

Measure Names
- TikTok Views
- Spotify Streams
- YouTube Views
- Pandora Streams
- Soundcloud Streams

Pandora Streams, Soundcloud Streams, Spotify Streams, TikTok Views and YouTube Views. Color shows details about Pandora Streams, Soundcloud Streams, Spotify Streams, TikTok Views and YouTube Views. Details are shown for Split Words.

Sheet 3

Spotify Streams
523,954,769 — 9B
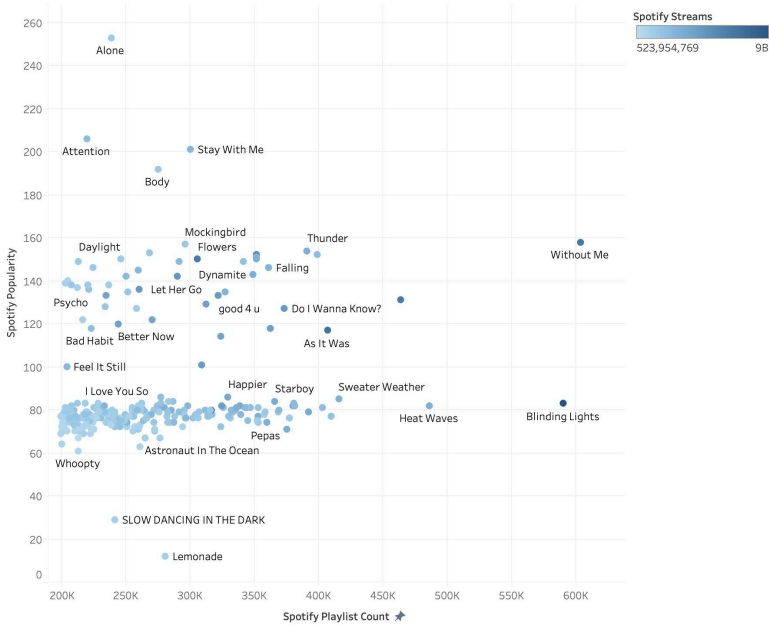
Sum of Spotify Playlist Count vs. sum of Spotify Popularity. Color shows sum of Spotify Streams. The marks are labeled by Track. The view is filtered on sum of Spotify Playlist Count and sum of Spotify Popularity. The sum of Spotify Playlist Count filter includes values greater than or equal to 200,000. The sum of Spotify Popularity filter keeps non-Null values only.

The questions answered in Tableau focused on platform-based engagement metrics, such as comparing TikTok views, YouTube views, and Spotify streams in relation to Spotify popularity. These insights helped illustrate the relationship between playlist exposure, platform reach, and overall song success. However, some of the broader trend-based questions the team initially proposed—such as analyzing explicit content performance, seasonal trends in releases, or artist consistency across platforms—were not directly addressed. This shift occurred because platform-driven metrics were more quantifiable and visually impactful in Tableau, making them more suitable for immediate insights. Additionally, these questions aligned better with our target audience of music industry professionals, offering clear, data-driven takeaways. While the broader questions remain useful for long-term industry analysis, the Tableau visualizations provided more specific, actionable insights, making them a better fit for our dataset and visualization capabilities.

Sketches

1. Which platform provides the highest reach for top-ranked songs?
2. What is the relationship between TikTok views and Spotify streams?
3. Do songs with high Shazam counts have better streaming performance?
4. How does a song's playlist count affect its overall popularity?
5. Which artists consistently appear in the top ranks across all platforms?
6. Are songs released in specific months more successful?
7. What is the average track score for the top 50 songs?
8. How does YouTube engagement (likes and views) relate to Spotify playlist reach?
9. Do explicit tracks perform better or worse on different platforms?
10. What is the distribution of all-time rank among the top 100 songs?

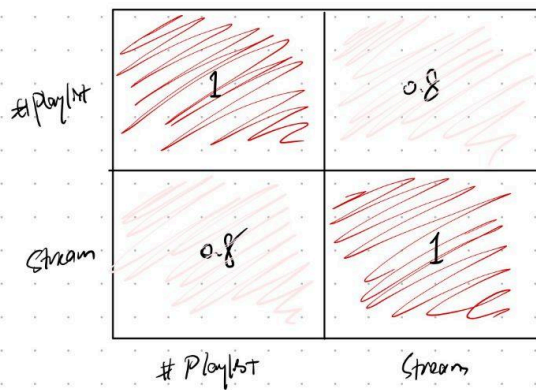| Sketch ID	Question ID	Author	# vote
P1	Q4	Haochen Ding	3
P2	Q4	Haochen Ding	0
P3	Q3	Haochen Ding	0
P4	Q6	Haochen Ding	0
P5	Q1	Haochen Ding	1
P6	Q1	Yinuo Yang	0
P7	Q2	Yinuo Yang	2
P8	Q9	Yinuo Yang	0
P9	Q6	Yinuo Yang	3
P10	Q5	Yinuo Yang	1
P11	Q8	Xinyue Li	0
P12	Q5	Xinyue Li	3
P13	Q6	Xinyue Li	0
P14	Q10	Xinyue Li	0
P15	Q7	Xinyue Li	2

Haochen Ding

Spotify playlist counts vs. Spotify Streams Q4
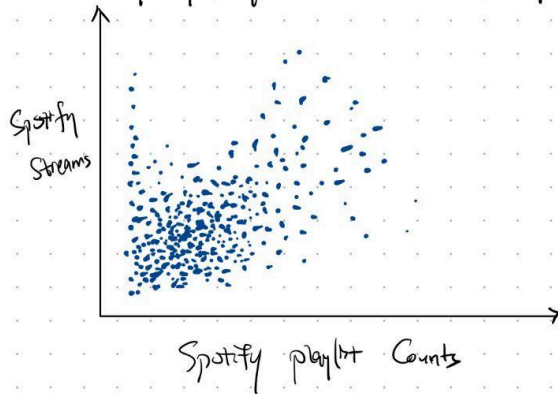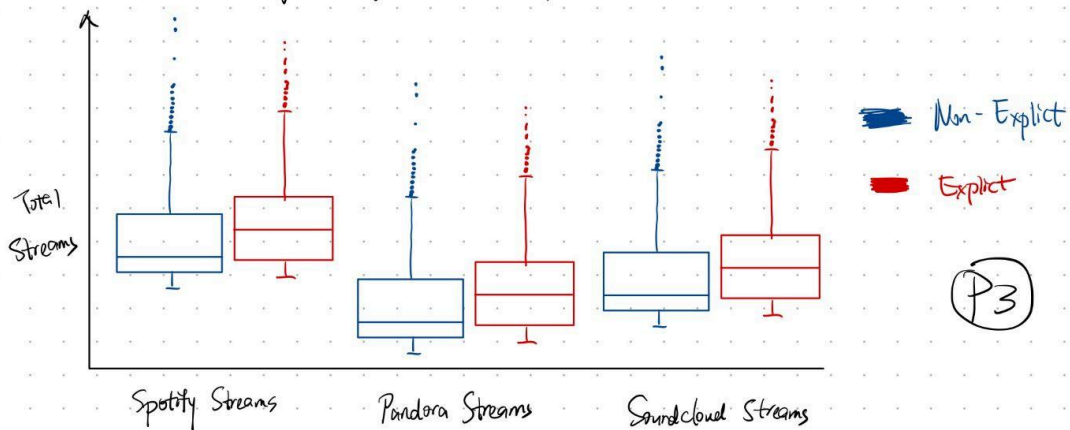

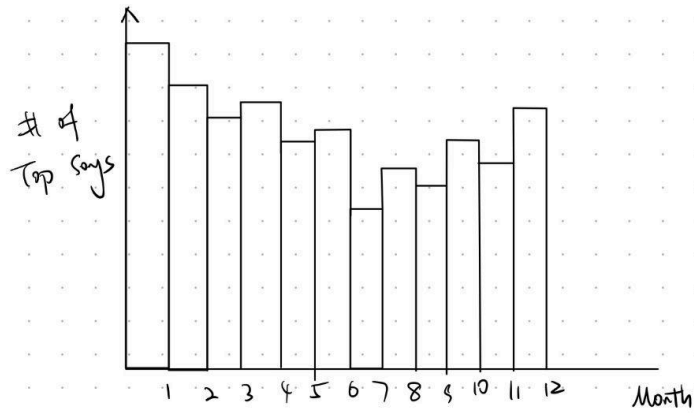P1

Correlation between Playlist Count and Streams Q4

P2

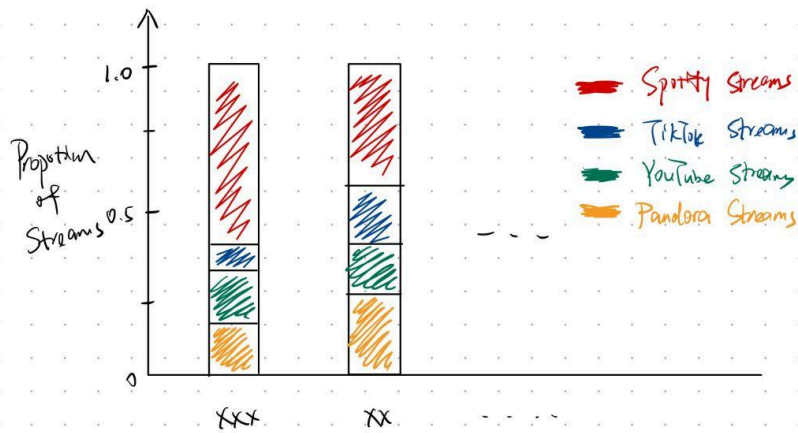Total streams for Explict / Non-explict Track Q3

Non-Explict
Explict

P3

Distribution of # of Top songs in different month   Q6

# of Top songs

1 2 3 4 5 6 7 8 9 10 11 12   Month

(P4)

Platform Contribution to Top songs   (Modify Q1: Which platforms contribute most to a song's success?)

Proportion of Streams

1.0

0.5

0

━━ Spotify Streams
━━ TikTok Streams
━━ YouTube Streams
━━ Pandora Streams

(P5)

XXX   XX   - - - -

Yinuo Yang

Q1: Multi-Platform Popularity Comparison

Legend:
- Spotify
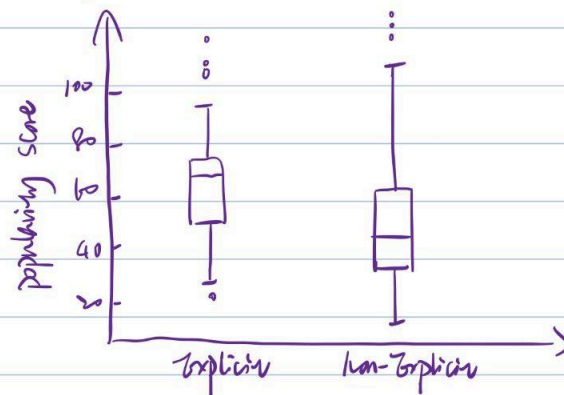- Youtube
- TikTok
- Apple Music
- Other
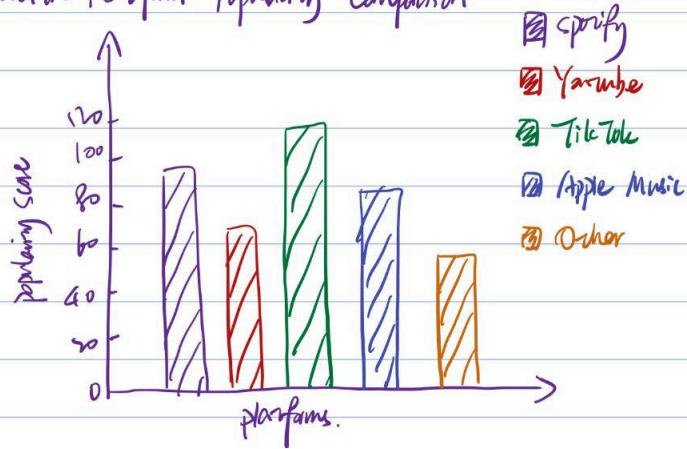
(P6)

Q2: TikTok Views vs. Spotify Streams.

(P7)
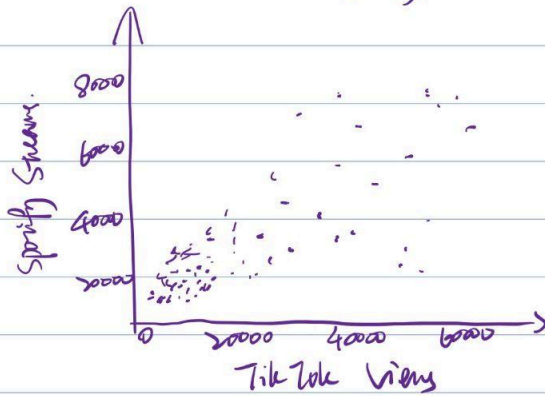
Q9: Explicit Content Performance

(P8)

Q6: Seasonal Song Popularity Trend

P9

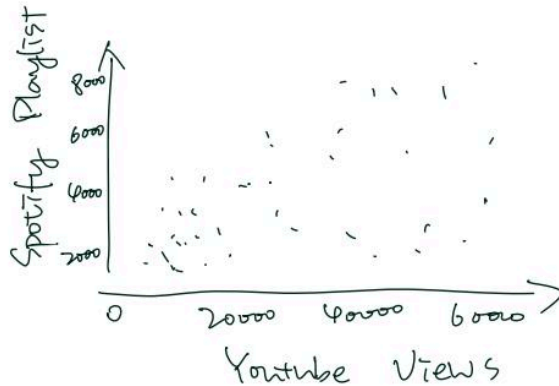## Q5: Artist Consistency Across Platforms.

Artist D
Artist C
Artist B
Artist A
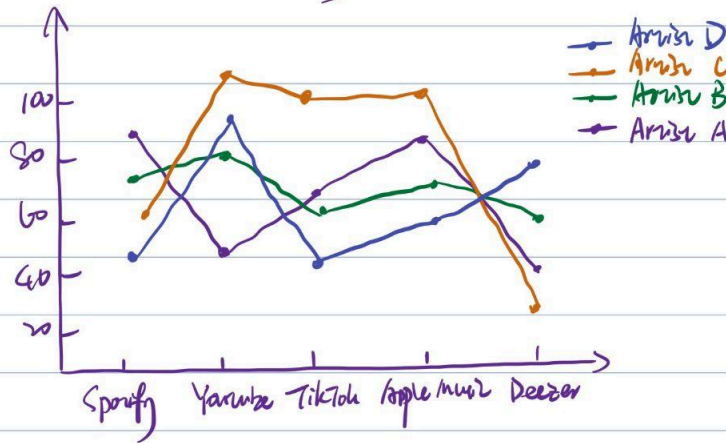
P10

Xinyue Li

Q8

Youtube Views v.s Spotify Playlist

P11

Q5 Artists Consistenly Across platforms

P12

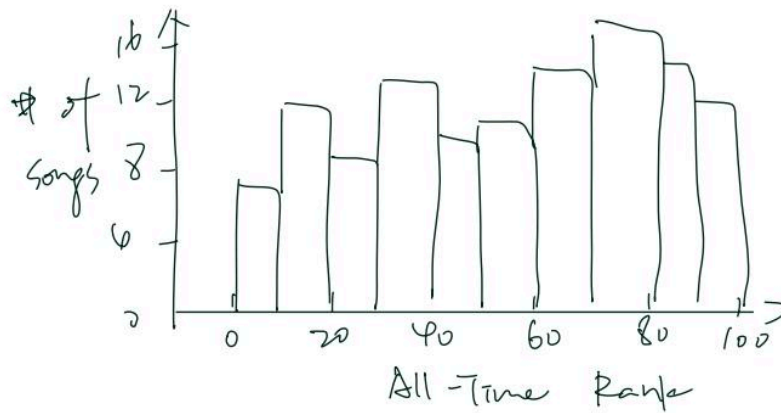|	Spotify	Youtube	TikTok	Apple Music	Pandora
Artist C	78	88	73	81	82
Artist B	83	46	50	57	64
Artist A	61	71	56	56	78

Q6. Top Song in Each Month
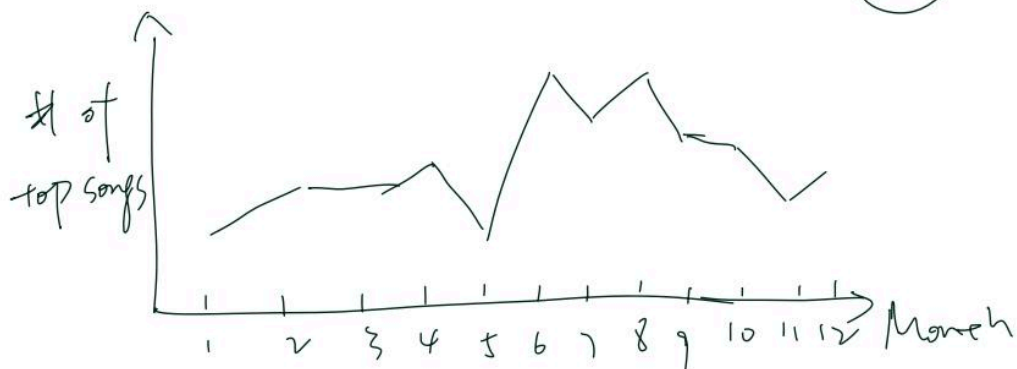
P13
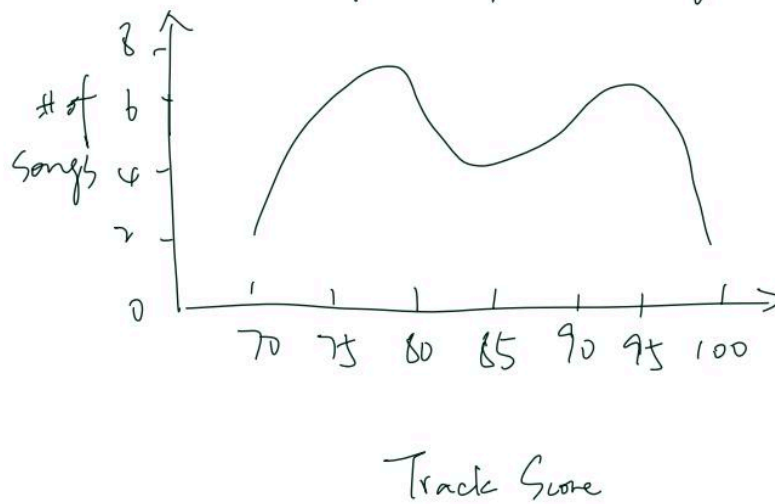
Q10   Distribution of All-Time Rank
            for Top 100 Songs

Q7   Distribution of Track Scores
          for Top 50 Songs

Final choice:

Q5   Artists Consistently Across platforms

	Spotify	YouTube	TikTok	Apple Music	Pandora
Artist C	70	88	73	81	82
Artist B	83	46	50	57	64
Artist A	61	71	56	56	78

(P12)

Spotify playlist counts vs. Spotify Streams Q4

Spotify Streams

Spotify playlist Counts

(P1)

Q2: TikTok Views vs. Spotify Streams.

Spotify Streams

8000
6000
4000
2000
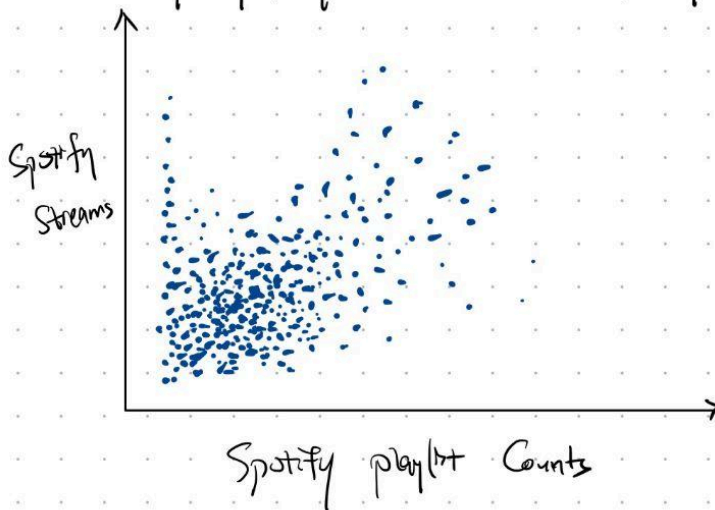
0   2000   4000   6000
TikTok Views
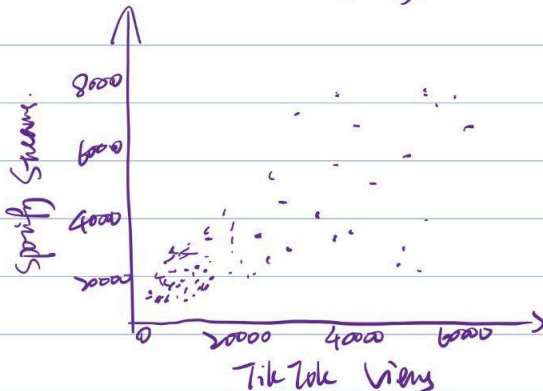
(P7)

Q6. Seasonal Song Popularity Trend
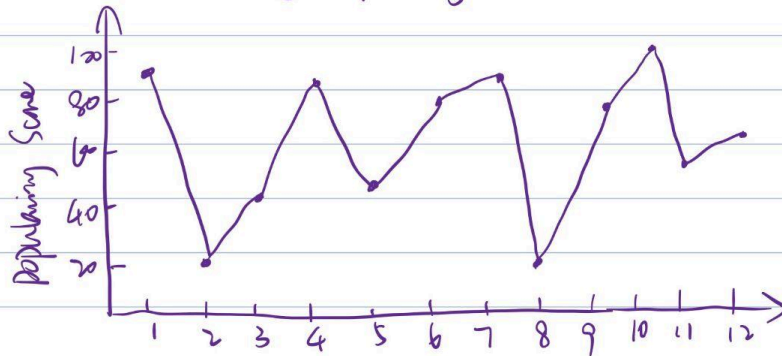
P9

Q7 Distribution of Track Scores for Top 50 Songs

P15

These plots were chosen as they could best represent significant findings from our dataset in a clear and easy-to-comprehend format for our audience. The Q5 (Artists Across Platforms) heatmap is an excellent visualization of artist popu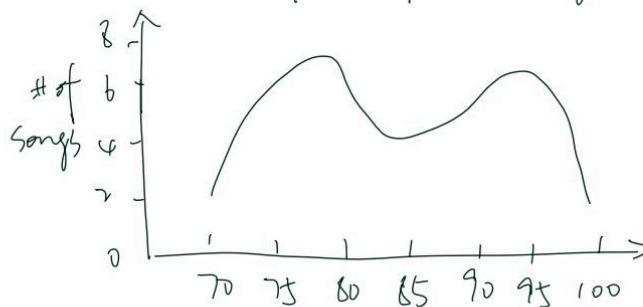larity differences between streaming platforms, and it is an insightful indicator of platform-specific dominance. Q2 (TikTok Views vs. Spotify Streams) and Q4 (Spotify Playlist Counts vs. Streams) scatter plots were selected to find correlations among engagement metrics, which can be helpful in platform-specific trend determination. The Q6 (Seasonal Song Popularity Trend) line chart would be chosen to analyze how song popularity changes over time, and the Q7 (Distribution of Track Scores for Top 50 Songs) histogram clearly indicates how track scores are distributed. All of these sketches are perfectly aligned with our analytical goals and facilitate storytelling using data visualization in a proper way.

Storyboarding

Main message
- a. Insights

    Xinyue Li:
    1. Consistency Across Platforms Leads to Long-Term Success
        - Artists who maintain high rankings across multiple platforms tend to sustain their popularity longer.
        - This suggests that a strong presence across streaming and social media platforms is key to long-term music industry success.
    2. Explicit Content Does Not Significantly Affect Streaming Performance
        - Contrary to common assumptions, explicit songs and non-explicit songs perform similarly on Spotify in terms of total streams.
        - This suggests that platform restrictions on explicit content may not heavily impact long-term popularity.

    Haochen Ding:
    1. Seasonal Trends in Music Popularity
        - Certain months, such as January, November, and July, see higher streaming numbers.
        - This could indicate strategic timing for music releases, where winter months might be associated with holiday season engagement, and summer might benefit from seasonal anthems.
    2. Cross-Platform Popularity Is Not Always Directly Linked
        - While TikTok virality can boost awareness, it does not guarantee high streaming numbers on Spotify.
        - YouTube engagement, however, shows a moderate positive relationship with Spotify playlist reach, suggesting that video content may help sustain long-term streaming success.

    Yinuo Yang:
    1. Playlist Inclusion Drives Long-Term Popularity
        - Songs with higher playlist reach on platforms like Spotify and Apple Music tend to sustain long-term streaming success.
        - Being featured in curated playlists (e.g., Spotify's Top 50, Today's Hits) significantly increases a song's exposure and maintains steady streaming growth over time.
    2. Emerging Artists Can Compete with Established Stars
        - Some lesser-known artists achieve high rankings across platforms, proving that newcomers can break into the mainstream if their songs gain traction.
        - This suggests that viral moments, playlist placements, and cross-platform presence can help newer artists compete with industry veterans.

b. Main Insight
   - The main insight for our audience (music industry professionals) is: "Cross-platform presence is the key to long-term music success—songs and artists that gain traction across multiple platforms (Spotify, YouTube, TikTok) sustain higher popularity and longevity in the industry."

   - "So What?": To maximize success, artists and music industry professionals should focus on strategic cross-platform promotion, ensuring their songs gain visibility not just on one streaming service but across multiple platforms to maintain long-term engagement and chart presence.
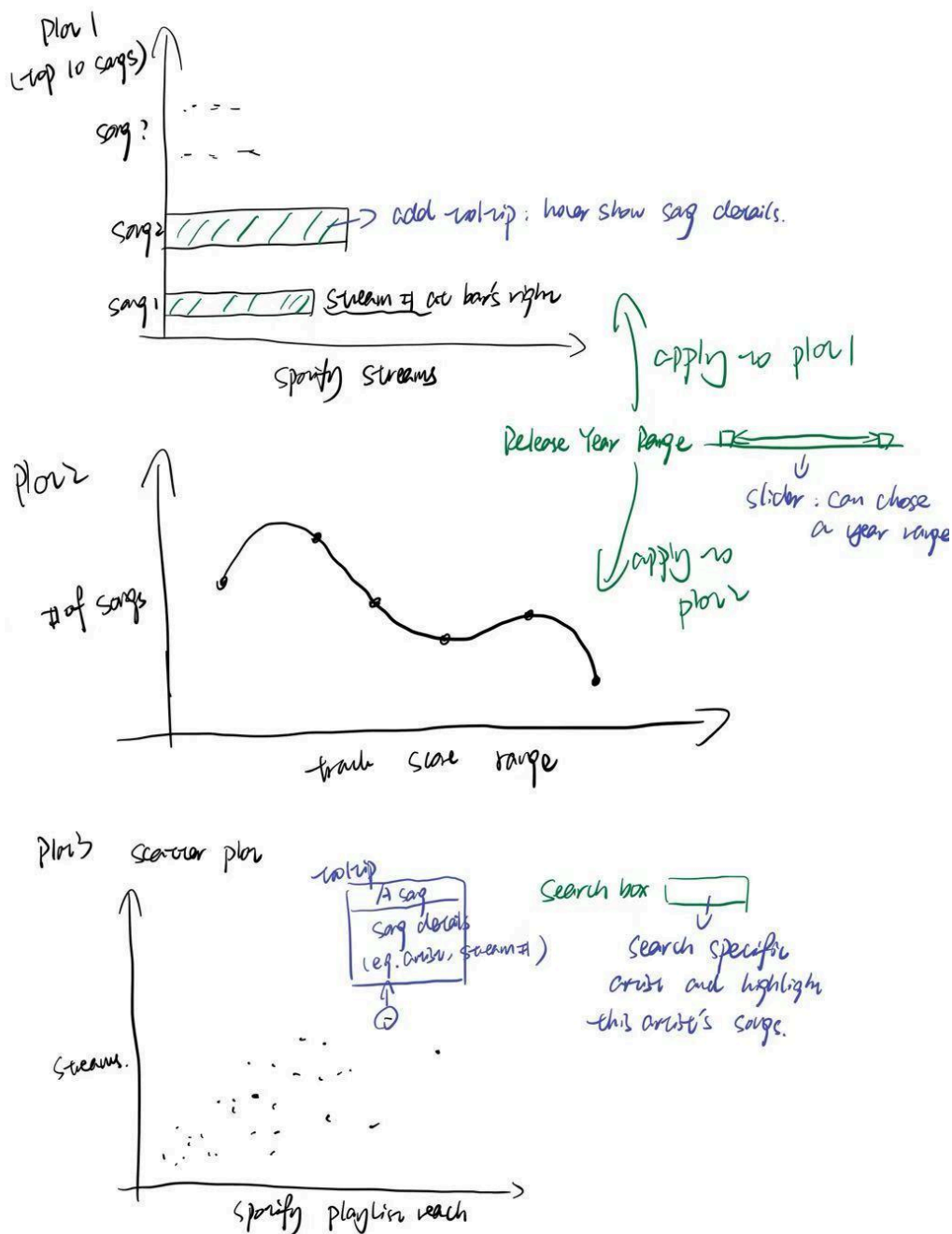
c. Reason for the choice
   - Our team chose the insight that cross-platform presence is key to long-term music success because the data shows that songs and artists sustain popularity longer when they gain traction across multiple platforms rather than relying on a single viral moment. Playlist inclusion on Spotify and Apple Music plays a crucial role in maintaining streaming numbers over time, while seasonal trends indicate that strategic release timing can further enhance performance. Additionally, YouTube engagement has a stronger connection to sustained success than TikTok virality, suggesting that multi-platform marketing is essential for long-term industry relevance. For music professionals, this means prioritizing playlist placements, cross-platform engagement, and strategic timing to maximize an artist's success beyond short-term trends.

Storyboard Flow:
   - Hook – Why do some songs sustain popularity while others fade?
   - Rising Insights – Key findings on cross-platform impact from our data.
   - Main Message – Cross-platform presence is crucial for music longevity.
   - Supporting Evidence – Playlist reach, emerging artists' breakthroughs, and multi-platform impact.
   - Solution – Strategic actions for artists and industry professionals to optimize engagement across multiple platforms.

Prototype v1

Plot 1 (top 10 songs)

song ?

song 2 → add tooltip: hover show song details.

song 1 — Stream # cot bars right

Spotify Streams

} apply to plot 1

Release Year Range ⟷ slider: can chose a year range

Plot 2

# of songs

track score range

} apply to plot 2

Plot 3 scatter plot

tooltip
A song
song details
(eg. Artist, Stream#)

Search box

Search specific artist and highlight this artist's songs.

Streams.

Spotify playlist reach

Haochen Ding:
I first structured the project, and set up version control using Git. I created the HTML framework to integrate visualizations smoothly.
I also created the Top 10 Most Streamed Songs plot. I build an interactive bar chart that dynamically updates based on a release year range slider. The chart sorts songs by streams, formats the x-axis in billions, and updates the title to reflect the selected time period, improving readability and user experience.

Yinuo Yang:

I built a scatter plot showing the correlation between Spotify Playlist Reach and Streams, where each song is represented as a fixed-size green point. The chart scales the x-axis for playlist reach and the y-axis for streams, ensuring clear data distribution, and includes interactive tooltips that display song name, artist, stream count, and playlist reach on hover. There is also a searchable artist filter that allows users to select an artist and highlight only their songs.

Xinyue Li:

I developed a line chart that visualizes the relationship between Track Score and the Number of Songs among the Top 100 most popular tracks. To enhance user interaction, I implemented a year range slider, allowing users to dynamically filter songs by release year. The chart bins track scores into 50-point intervals, ensuring a clear trend representation. Additionally, I integrated interactive tooltips that display the Track Score Range and Number of Songswhen hovering over a data point, improving readability and user engagement.