Exploring Anime Trends: A Data-Driven Exploration

Github Link: https://github.com/SatvikaReddy/dataVis_project

Name	Umass Email	Spire ID
Megha Chakravorty	mchakravorty@umass.edu	34009311
Satvika Reddy Gavireddy	sgavireddy@umass.edu	34012398
Janani Lakshmi Pasham	jpasham@umass.edu	34023678

Table of Contents

Background and Motivation

Overview

Anime has become an increasingly global phenomenon, with millions of users actively engaging in ratings, reviews, and discussions across various platforms. However, anime rankings and popularity scores can be influenced by numerous factors beyond just quality—such as genre biases, rewatch behavior, and shifting user preferences over time.

This project seeks to analyze a large dataset of anime-related information, focusing on trends in scoring behavior, ranking shifts, and user engagement patterns. Key areas of interest include:

- Identifying biases in anime ratings, such as whether certain genres consistently are watched by which gender or if the source type of the anime data.
- Heat Maps to evaluate genre and anime related information for a country
- Analyzing user demographics (age, location, gender) to determine how different groups perceive and rate anime, potentially revealing representation gaps in the industry.
- Examining anime-watching habits across various user groups, including completion and drop rates, seasonal viewing patterns.

By leveraging data visualization, we aim to present clear and insightful trends that offer a deeper understanding of anime consumption and how user-driven ratings shape perceptions of quality and popularity.

Motivation

The motivation for this project stems from both a personal interest in anime and the broader implications of media analytics. As streaming platforms and recommendation systems become increasingly influential in content discovery, understanding how user ratings and preferences evolve can help refine these systems.

Anime has evolved from a niche interest to a global entertainment powerhouse, with millions of viewers engaging in discussions, reviews, and ratings across various platforms. However, the way anime is perceived and ranked is often influenced by more than just its intrinsic quality. Factors such as genre biases, user demographics, and evolving trends can significantly impact anime ratings and popularity scores.

Understanding these underlying influences is crucial, as rankings and scores shape industry decisions, influence viewer recommendations, and even affect production trends. This project aims to uncover patterns in scoring behavior, user engagement, and ranking shifts by analyzing a large dataset of anime-related information.

A key motivation behind this research is to highlight biases in anime ratings—whether certain genres consistently appeal to specific demographics or if the source material influences reception. Additionally, by examining user demographics, we can identify representation gaps in the industry and understand how different age groups, genders, and locations shape anime trends.

Furthermore, analyzing anime-watching habits, such as completion and drop rates or seasonal preferences, provides insight into viewer engagement. By leveraging data visualization techniques, we aim to present clear, data-driven insights that enhance our understanding of anime consumption and the evolving landscape of audience-driven perceptions.

Project Objectives

- Major Research question - Analyze customer engagement and consumption with respect to anime attributes
- Identify trends in scoring behavior - do users rate certain genres higher? Do users who rewatch anime rate them higher? This can help us remove bias and normalize ratings in future
- Understand the change in anime rankings over time. This can help us understand if user preferences change over time. Producers can use this knowledge to invest and promote more popular genres and types of anime with respect to a particular time
- Check if scores are the best indicator of anime popularity. This evaluation can help improve marketing strategy for different animes.
- Analyze the most-watched, highest-rated. This data can be useful to anime platforms trying to acquire the rights to present these on their space. This will increase their user base.
- Analyze anime popularity based on different demographics like age, location and gender. If a certain demographic consistently scores a genre lower, it could indicate a lack of representation or appeal for that group. Consequently, this can highlight potential biases in anime production and marketing strategies.
- Identify which type of anime (TV shows, movies, OVA) users watch most. This can help us identify which form of entertainment is becoming more popular with changing times. Producers can take this into account for upcoming productions.
- Analyze anime-watching trends over the years and find patterns to and find patterns to understand audience preferences, predict emerging genres, and identify shifts in popularity and engagement.

Data

Dataset link: https://www.kaggle.com/datasets/dbdmobile/myanimelist-dataset/data

Data Source

Anime is a popular form of Japanese entertainment, comprising TV series, movies, and OVAs from diverse genres and audiences, which everyone loves. Known for its distinctive art style and compelling storytelling, anime covers genres like action, romance, fantasy, and sci-fi, each offering unique narratives. Analyzing anime data helps uncover user preferences, improve recommendation systems, and track popularity trends. It provides insights into genre preferences, demographic influences, and engagement patterns, enhancing user experience and retention. Additionally, it aids studios and streaming platforms in tailoring content and marketing strategies based on audience behaviour

- The datasets were collected from the **MyAnimeList** platform, a popular online community and database for anime and manga enthusiasts. The platform provides valuable information about anime shows, user profiles, and user scores for different anime.
- We will be using the **"Anime Dataset 2023"** dataset, which is publicly available on Kaggle.

Data Collection

Data was collected using Jikan API calls and web scraping. The Jikan API was used to retrieve anime details, including name, English name, synopsis, score, genres, and more. For the User Details Dataset, user profile information and anime statistics were obtained through API calls. Additionally, web scraping was performed to collect user scores for different anime.

Columns/Attributes

Columns in **final_animedataset.csv**:

1. *username* - The username of the user who rated the anime.
2. *anime_id* - The unique ID of the anime.
3. *my_score* - The rating/score given by the user to the anime.
4. *user_id* - The ID of the user.
5. *gender* - The gender of the user.
6. *title* - The title of the anime.
7. *type* - The type of the anime (e.g., TV series, movie, OVA, etc.).
8. *source* - The source material of the anime (e.g., manga, light novel, original, etc.).
9. *score* - The overall score of the anime.
10. *scored_by* - The number of users who have rated the anime.
11. *rank* - The ranking of the anime.
12. *popularity* - The popularity rank of the anime.
13. *genre* - The genre(s) of the anime.

Columns in **user-details-2023.csv**:

1. *Mal ID* - Unique ID for each user
2. *Username* - The username of the user.
3. *Gender* - The gender of the user
4. *Birthday* - The birthday of the user (in ISO format).
5. *Location* - The location or country of the user.
6. *Joined* - The date when the user joined the platform (in ISO format).
7. *Days Watched* - The total number of days the user has spent watching anime.
8. *Mean Score* - The average score given by the user to the anime they have watched.
9. *Watching* - The number of anime currently being watched by the user.
10. *Completed* - The number of anime completed by the user.
11. *On Hold* - The number of anime on hold by the user.
12. *Dropped* - The number of anime dropped by the user.
13. *Plan to Watch* - The number of anime the user plans to watch in the future.
14. *Total Entries* - The total number of anime entries in the user's list.
15. *Rewatched* - The number of anime rewatched by the user.
16. *Episodes Watched* - The total number of episodes watched by the user.

Visualization Designs

We follow the FDS methodology for sheets:

Next 5 sheets

Sheet 1: Brainstorm

Sheets 2-4: Initial (Alternative) Designs

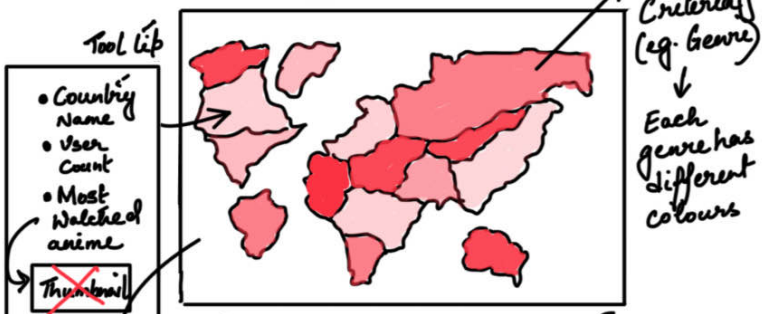
Sheets 5: Realization (Final Design)

IDEAS

Goal: Analyze Customer Behaviour, Engagement and Consumption with respect to Anime Attributes

Anime Attributes	User Attributes
Genre	Age
Type (TV, movies, etc.)	Gender
Source	Location
Producer	Completed
Score	On-hold
Rank	Watching
Popularity	Dropped
	Planned to Watch
	Rewatched
	Episodes Watched

VIS 1

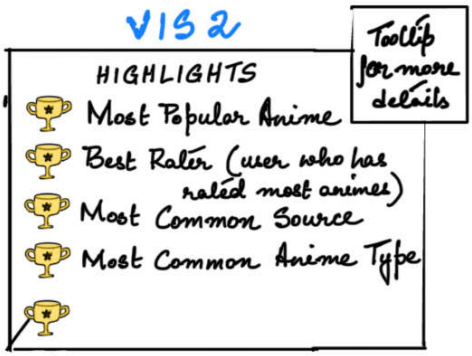


World Heat map — Identify locations with most anime viewers

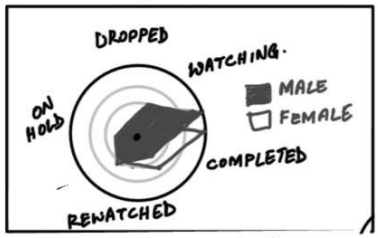
Only Countries (OPTION 1)

Countries but shows specific location on zooming (OPTION 2)

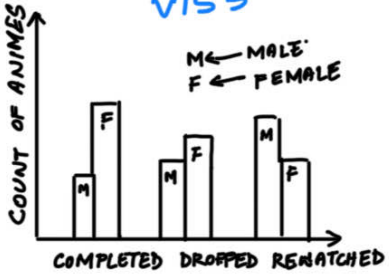
VIS 2



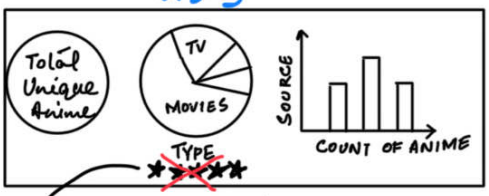
VIS 4



VIS 5

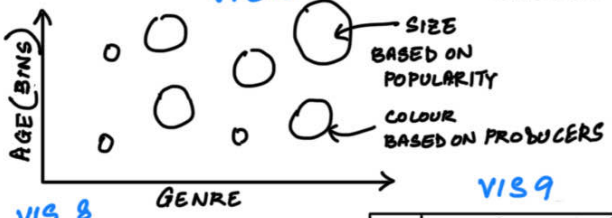


VIS 3

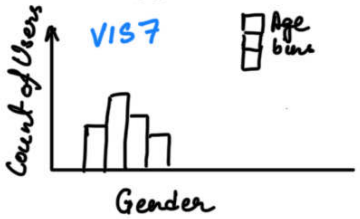


Most Popular Anime
→ No. of filled stars indicate %

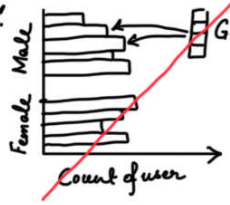
VIS 6



VIS 7



VIS 8



VIS 10

VIS 9

Tabular data

Source or Genre	Most Watched	Most Common Source	Most Common Type	Duration	...

Filter: 1, 2, 3, 4, 5, 6, 7, 9

Categorize:

(i) Since the map already filters by genre, change x-axis value in Vis 6 to "Source"

(ii) Remove thumbnail from Vis 1 to make sure that the tooltip info is not very crowded

(iii) Keep 1 Statistics related visualization (2 or 9), (iv) Keep either (4 or 5) which have same attributes

Combine and Refine:

(i) Combine 1, 2, 4, 6 to present information about multiple attributes

(ii) 4 over 5 → More complex and aesthetic.

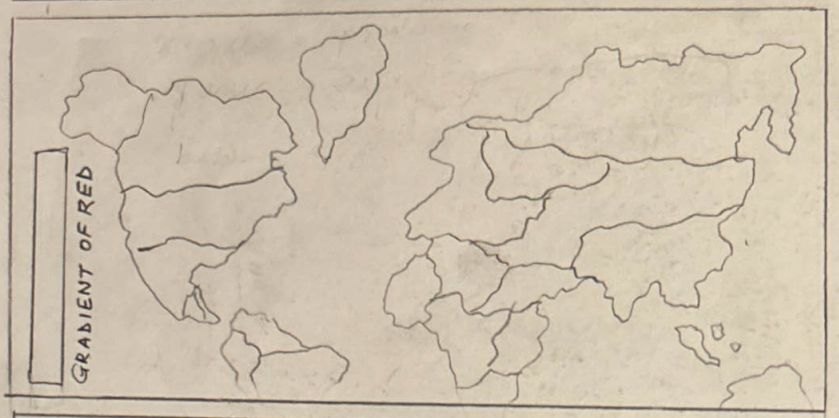
(iii) Heat map → Remove zoom feature because most location values are country-wise

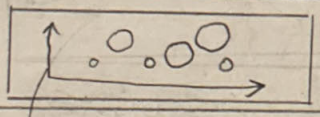
(iv) Dont have more detailed locations

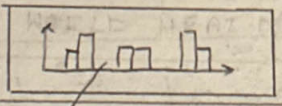
(v) Use "Bronze, Silver, Gold trophies in highlight for more details

Anime Posters for TOP 3 most popular animes across all geographical locations (changes in 10-15 secs)

• • •

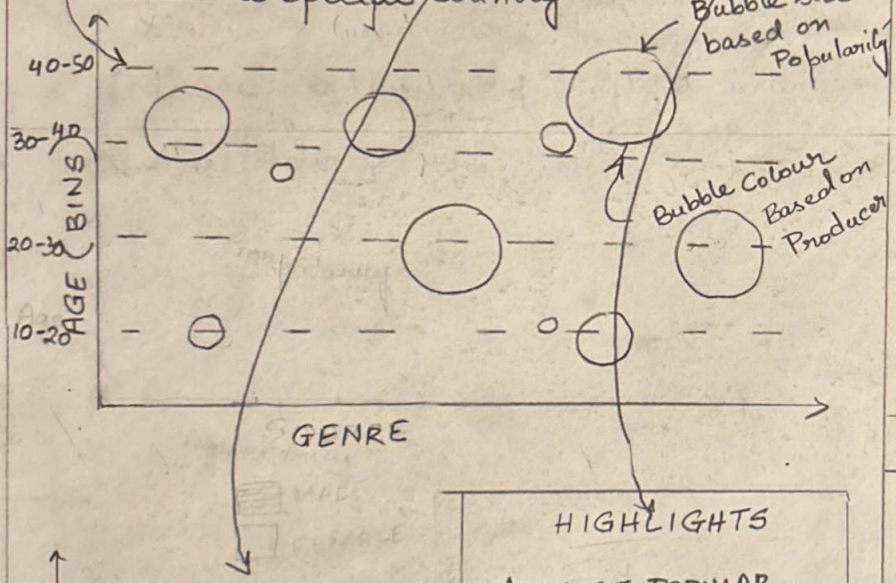


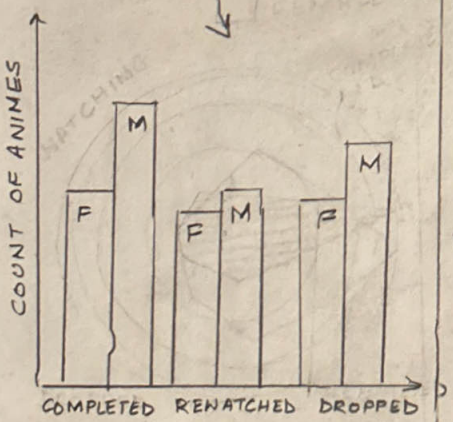




HIGHLIGHTS

Select a specific country





F ← Female
M ← Male
(denoted by colours)

HIGHLIGHTS

★ MOST POPULAR ANIME
★ BEST RATER
★ MOST COMMON SOURCE
★ MOST COMMON TYPE

FOCUS

METADATA

TITLE: LOCATION-WISE CUSTOMER ENGAGEMENT ANALYSIS

GROUP MEMBERS NAMES: MEGHA CHAKRAVORTY, SATVIKA REDDY GAVIREDDY, JANANI LAKSHMI PASHAM

DATE: 02/27/2025

TASKS: SHOWS LOCATION SPECIFIC DATA FOR CUSTOMER ENGAGEMENT AND BEHAVIOR

SHEET: 2

OPERATIONS

↳ User can hover over each location which tells them count of user from that location, average watched animes, average watched episodes

↳ In the highlights, if user hover over a winner in a category, it gives more details like (name of anime, japanese name, producer, type, source) or (age, no. of animes rated, location, most watched animes)

↳ User will see the visualizations in "layout" first. When they choose a location map will zoom in on the location and the 3 visualion in focus will load based only on data from that location

↳ Enable storytelling features to ensure users understand they have to select loc to proceed

DISCUSSION/DETAIL

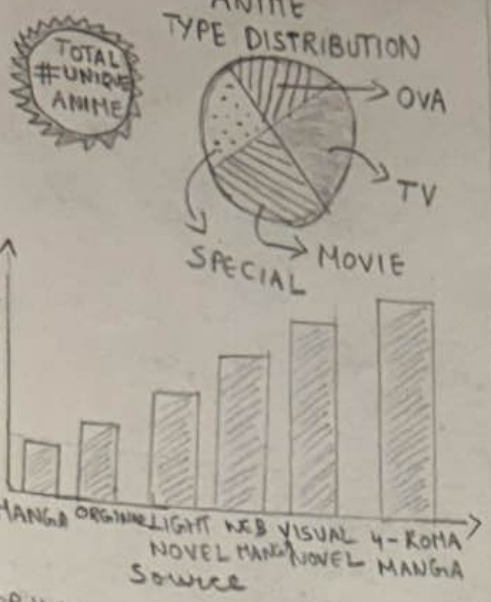
↳ Should world map be shown when nothing is selected?

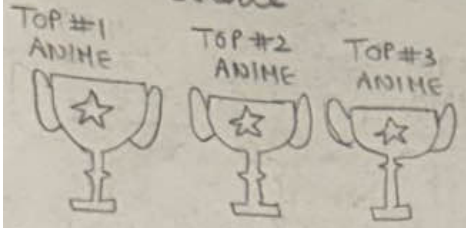
↳ Allow only country-wise location selection

↳ Should locations (say USA) show points for more specific locations (like NYC) where count of viewers is significant?

↳ Is drop-down better than tooltip for highlights section?

↳ Essential to clean location values during preprocessing





BASED ON POPULARITY AND RANK

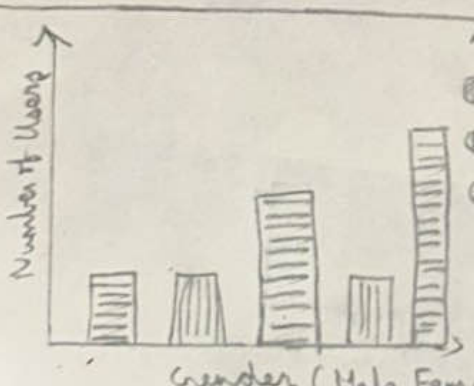
TITLE: SCORING TRENDS (USER PREFERENCES Vs. AVERAGE ANIME SCORES)

(THIS LAYOUT EMPHASIZES ON IDENTIFYING OUTLIERS AND PATTERNS IN USER ENGAGEMENT AND SCORING BEHAVIOR)

LAYOUT

FOCUS

→ Focuses on Drawing Attention to the top-rated and most popular anime.



The focus is on the distribution of users across different genders while simultaneously highlighting the age group segmentation. (Hover voting is used for age groups)

GEOGRAPHICAL HEATMAP VISUALIZING ANIME POPULARITY BY GENRE ACROSS COUNTRIES

ACTION	MYSTERY	FANTASY	THRILLER

FILTER BY GENRE

DIFFERENT COUNTRIES ARE COLOURED FROM LIGHT TO DARK BASED ON GENRE

(WE CAN KNOW WHICH COUNTRY HAS THE MOST VIEWS IN A PARTICULAR GENRE)

Focus is on analyzing the relationship b/w user-given scores and average anime scores. Color coding represents anime popularity, while point size indicates the volume of user ratings.

META DATA

TITLE: ANIME ENGAGEMENT INSIGHTS ACROSS COUNTRIES

GROUP MEMBER NAMES: JANANI LAKSHMI PASHAM, SATVIKA REDDY GAVIREDDY, MEGHA CHAKRAVORTY

SHEET NO: 3

DATE: 02/27/2025

TASKS: STORYTELLING VISUALIZATION TO ANALYZE ANIME ENGAGEMENT TRENDS, INCLUDING TYPE DISTRIBUTION, GENRE POPULARITY BY GEOGRAPHY, AND USER SCORING BEHAVIOR.

OPERATIONS

- Users can switch between different genres on the map from the options available to see how anime preferences vary across countries.
- User can also sort the source type in ascend. / descend. order.
- User can hover over elements on the scatter plot and geographic heatmap which reveals detailed information such as average scores, genre distribution and user engagement metrics
- The geographic heatmap allows users to click on specific country to update all other charts with region-specific data
- Users can click on trophy icons for the top-ranked anime which will display rank, popularity & aggregated user ratings

DISCUSSIONS

- Scatter plots can become cluttered due to large dataset, can we implement zoom or filter functionalities here?
- The heatmap might oversimplify data as it has only genre based selection
- Right now all charts are changed based on country, can we have more specific locations?

META DATA

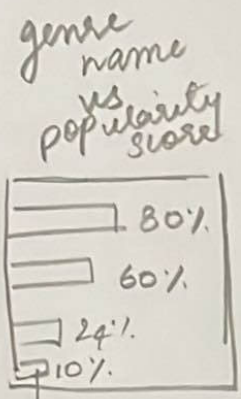
TITLE: ANIME VISUALISATONS - USA

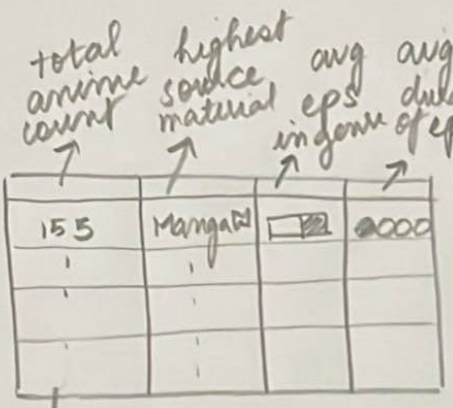
GROUP MEMBER NAMES: SATVIKA REDDY GAVIREDDY
MEGHA CHAKRAVORTY
JANANI LAKSHMI PASHAM

DATE: 02/27/2025

TASKS: shows the genre specific info about USA

SHEET No: 4





genre specific statistics overview

- USER can hover over the horizontal bar and it will give best anime in that genre
- NEXT USER can click on the right box on any genre

LAYOUT

FOCUS

→ The box gives basic statistics only, but the next focus on the specific genre

→ 3 visualisation specific do genre will pop up under the page

→ GRADIENTS/COLORS used to highlight genders, interests in heat map

OPERATIONS

DISCUSSION | DETAIL

→ It should take 8-10 weeks to implement.

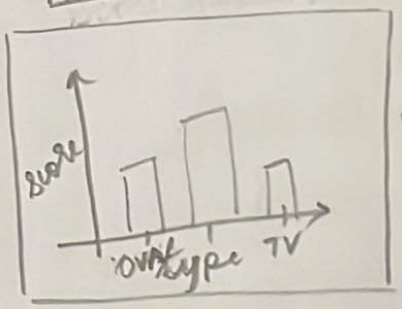
→ No specific hardware requirements other than laptops

→ DATASETS acquired from kaggle, need preprocessing

→ HTML, CSS, JAVASCRIPT as main stack.

→ RADAR CHART, bar graphs, heat maps, table with all information

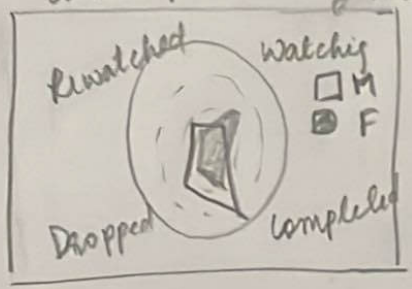
BEST ANIME



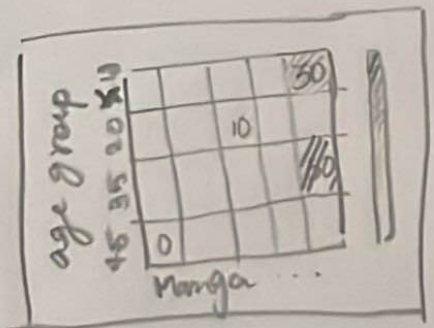
source distribution by type

BAR CHART

exploring anime consumption pattern by gender



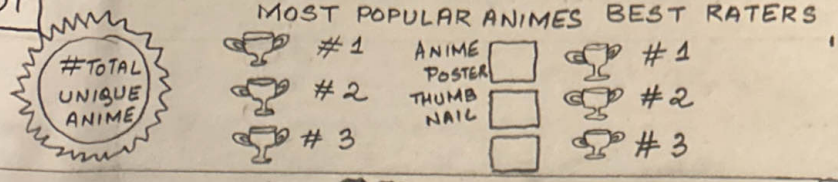
RADAR CHART

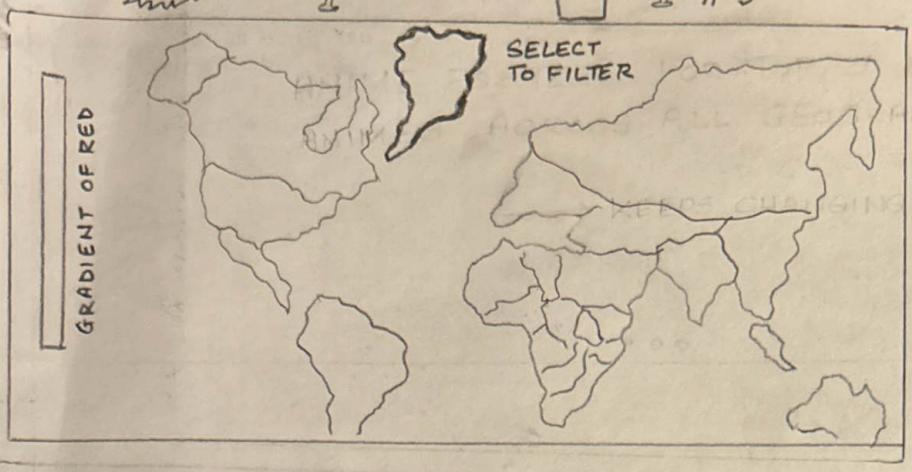


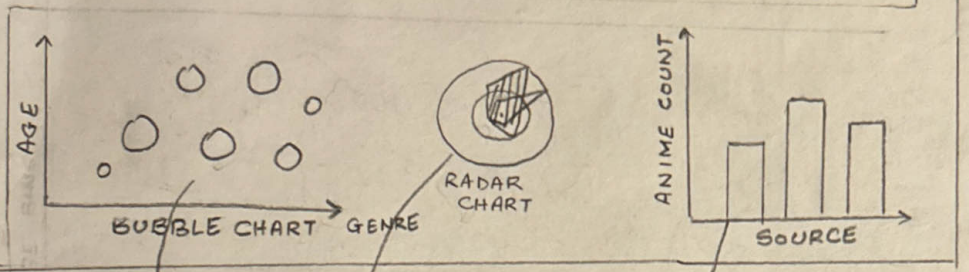
HEAT MAP

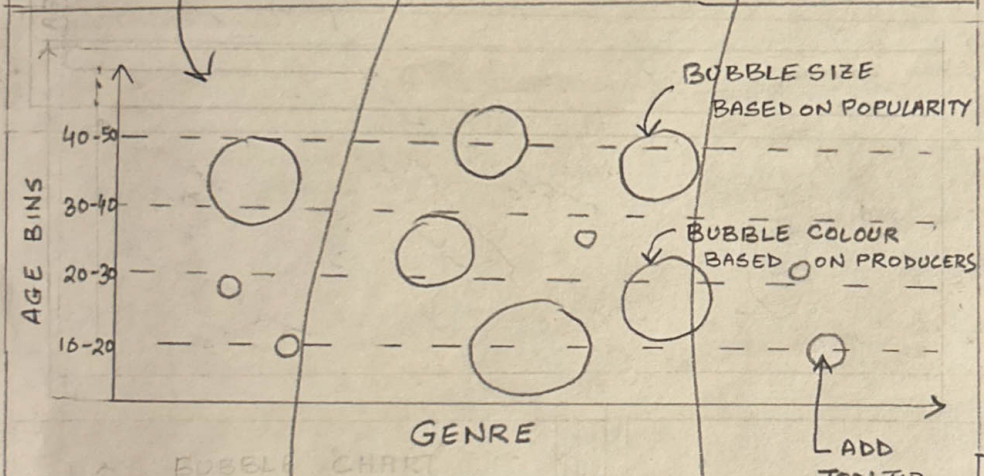
source pref by age group

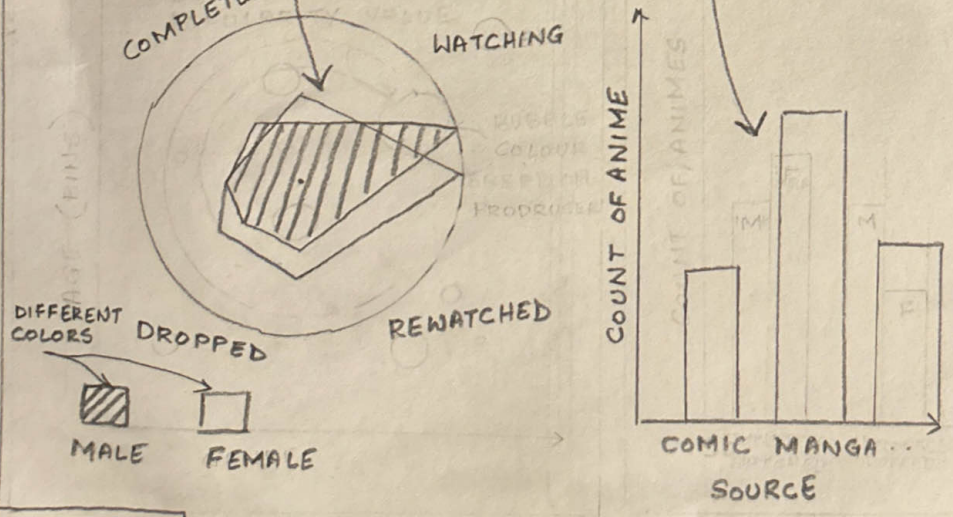
LAYOUT











FOCUS

METADATA

TITLE: LOCATION-WISE ANALYSIS OF CUSTOMER BEHAVIOR AND ITS CORRELATION TO ANIME ATTRIBUTES

GROUP MEMBER NAMES:
MEGHA CHAKRAVORTY
JANANI PASHAM
SATVIKA REDDY GAVIREDDY

DATE: 02/27/2025

TASKS: Shows customer behavior based on different anime attributes

SHEET: 5

OPERATIONS

↳ Heat map allows users to select countries. Once a country is selected following visualizations will be filtered based on that country

↳ Map allows tooltip interactions to display count of users, average watched animes, average watched episodes, etc.

↳ Enable story-telling style to make sure users select a specific country for better visualizations.

↳ Bar chart offers annotations with tooltips

↳ Display thumbnails of anime poster beside "most popular" anime names

DETAILS

↳ Preprocess locations data to filter out invalid locations

↳ Ensure valid locations are mapped to countries

↳ Total unique animes, most popular animes and best raters are with respect to all valid locs. and dont change with country-wise filtering.

↳ Add narrations between visualizations so that they are easy to follow as a story.

Data Pre-processing/ Exploratory Data Analysis

1. 1. Join final_animedataset.csv and user-details-2023.csv based by mapping user_id from the first and Mal ID from the later
2. Drop columns that we are not interested in for the project like username from the user-details-2023.csv
3. Handle missing or null values by either removing corresponding rows or imputing them with appropriate default values.
4. Convert location names into latitude and longitude coordinates for mapping them on a world map.
5. Convert the Birthday and Joined column to a datetime format.
6. Identify and remove duplicate entries in the dataset.
7. Verify the data types of columns and if needed convert them to correct datatypes.
8. Create composite columns to derive meaningful insights -
    a. Calculate age of user from Birthday column
    b. The number of days/years the users have been on the platform from the Joined column.

Design Evolution

Detailed steps for Data Preprocessing:

1. Since the file sizes are very large, we loaded the “final_animedataset.csv” to anime_df (a pandas dataframe) and “users-details-2023.csv” to user_df (a pandas dataframe) from the "dbdmobile/myanimelist-dataset" from Kaggle Hub directly using python.
2. Understood data quality by looking at the columns in the data using df.columns, datatype for each column and number of null values in each column using df.info for both the datasets
3. Merged anime_df and user_df by the “user_id” column in anime_df and “Mal ID” column in user_df to form a new pandas dataframe, df.
4. Dropped “Mal ID” to avoid duplicate columns
5. After analysing the merged data, null values were found in the following columns: Rank, genre, Gender, Birthday, Location. Since we have 16,963,197 data points after merging, we chose to drop the rows with any null values.
6. Age calculation - Converted the Birthday column values to datetime (datetime.fromisoformat(str(date_str))) and extracted the year from it (dt.year). Then, we calculated Age by subtracting the birthday year value from 2023, since all of this data was collected for 2023 (age = 2023 - dt.year). We are interested in the user's age at the time when the data was calculated.
7. We further reduced data size by dropping the Birthday and “Joined” columns, since we don’t plan to use them for our visualizations.
8. The values in the “Location” column needed to be cleaned since they contain different types of values from imaginary location, other planets to detailed locations. For this we used Python’s geopy library. In order to understand whether the location is valid, the code tries to convert the raw locations in the data to a Geopy Location object using geolocator.geocode function. If the conversion is successful, the location is valid, if it raises any error, the location is invalid.
9. After doing this validity check, we dropped all rows with invalid locations and saved and intermediate output in chunks.
10. We retrieved latitude, longitude positions of each location by converting the raw locations to Geopy location objects and extracting its latitude longitude attributes (location.latitude, location.longitude). With the latitude, longitude positions being known, we use the reverse_geocode function to extract the city, state, and country attributes that are mapped to these latitude, longitude positions. The chunks are updated with city, state, country locations after this.
11. We further dropped rows where any of the column values were missing (Dropping the rows makes sense because we have a lot of data)
12. Finally we filtered the data by country and chose all the data points where the country is USA.
13. Join final_animedataset.csv and user-details-2023.csv based by mapping user_id from the first and Mal ID from the later
14. We obtained the top 15 states by frequency. Using value_counts() function, we further filtered the data to keep only rows where state is one of the top 15

15. We further filtered this data to retain only the top 100,000 data points for easier data handling during data visualization.

Advantages and Disadvantages of Visualization Ideas:

We chose Visualization 4 over Visualization 5 - While both the plots encode gender and viewing pattern we chose the radar chart because it adds variety and aesthetic to the website overall.

1. **Visualization 1: Heat/Choropleth Map**
   Advantage:
   a. Clearly visualizes regional anime popularity and highlights potential genre, source preferences or rating biases across countries.
   b. Visually appealing and easy to convey which regions have the most anime enthusiasts
   c. Makes data filtering based on geographical locations easier
   Disadvantage:
   a. Requires in-depth data preprocessing
   b. Need to take a decision on whether it should be limited to the world map or the USA map
   c. May not be possible to compare the data from different states at the same time

Interesting visualization that allows interactivity and ties together other visualizations - **included in optional features**

2. Visualization 2: Anime viewing statistics
Advantages
a. Provides quick, at-a-glance insights (e.g., most popular anime, best rater, common source/type)
b. Simple and easy to understand
c. Highlights key insights across all regions
d. Allows interactive exploration supplemented by tooltips.

Disadvantages
a. Lacks depth or additional context behind the “most popular” or “best” labels.
b. May oversimplify trends, as it doesn’t show comparisons or distributions.

Approved this idea, since it is insightful, simple and easy-to-implement.

3. Visualization 3: Combination of 3 charts
Advantages
a. Displays multiple metrics (source types, total unique anime, popularity) in one compact layout.
b. Uses both pie and bar charts to provide a quick overview of different data dimensions.

Disadvantages
a. Combining multiple chart types in a small space can lead to visual clutter and confusion.

b. The pie chart may conceal minor segments if categories are not clearly defined or if their values are too small.
c. Star-based popularity ratings can be ambiguous without sufficient context (e.g., total votes, rating scale).

This idea was **rejected** due to dominating disadvantages.

4. Visualization 4: Radar chart of viewing engagement for different genders

Advantage:

a. Clearly displays viewing pattern across the demographic, more specifically with an embedded gender split.
b. Visually appealing, intuitive and different from other common plots, allowing viewers to immediately grasp the proportions of each category.
c. Conserves space in dashboards while still offering layered insights through nteractivity.

Disadvantage:

a. When certain categories have very low proportions, it becomes challenging to compare them effectively.

5. Visualization 5: Bar chart of Viewing engagement vs. Gender

Advantages

a. Clearly displays viewing pattern across the demographic, more specifically with an embedded gender split.
b. Standard bar charts are familiar and quickly convey relative magnitudes

Disadvantages

a. Adding too many categories can clutter the visualization, making it harder to read.
b. Emphasizes only total counts for each status without offering additional context (e.g., time trends).
c. Horizontal or vertical space can become an issue, given the number of visualizations we want to add

Between Visualization 4 and 5, **5 is chosen** because it is more aesthetically pleasing and we dont want to add multiple bar charts on our website.

6. Visualization 6: Bubble Chart of (Popularity, Genre, Age bins, Producers)

Advantages

a. Simultaneously displays genre (x-axis), age bins (y-axis), popularity (bubble size), and producer (color coding) - visually stimulating
b. Quickly highlights outliers where certain genres and age groups intersect with high or low popularity.
c. Consolidates several data attributes into one chart, saving space on the dashboard.

Disadvantages

a. Too many data points can overlap, obscuring individual bubbles and complicating comparisons.

b. Distinguishing between small differences in bubble sizes or subtle color variations may be challenging.
c. Plotting “Genre” on an x-axis can become messy if there are numerous or overlapping categories.

Accepted - Right amount of complexity, encodes various attributes adds interactivity

7. Visualization 7: Bar chart of Count of Users x Gender x Age bins

Advantages

a. Easily compares the number of users for each gender in a straightforward bar format
b. Subdividing bars by age bins reveals how different age groups distribute within each gender.
c. Familiar chart type that most audiences understand at a glance.

Disadvantages

a. Stacking or grouping multiple age bins can reduce readability.
b. Overly simple form of visualization

Rejected due to oversimplistic nature

8. Visualization 8: Horizontal Bar Chart (Gender, Count of Users, Genres)

Advantages

a. Side-by-side or stacked bars make it easy to compare male and female counts across genres.
b. Placing gender categories on the y-axis allows longer labels (if needed) without text overlap.

Disadvantages

a. Adding many genres or subcategories can create busy visuals with multiple bars stacked or grouped.
b. Stacked bars may make it harder to compare each genre segment precisely across genders.

Rejected due to a simplistic plot that can easily become crowded.

9. Visualization 9: Tabular Format of Key Metrics

Advantages

a. Easily displays exact values (e.g., most-watched, most common source, most common type) without ambiguity.
b. Users are accustomed to reading tables, making it simple to find specific data points.

Disadvantages

a. Does not immediately highlight trends or patterns; users must interpret data row by row.
b. Not visually appealing

Rejected due to lack of visual appeal

10. Visualization 10 - Bar Chart (Type vs. Score)

Advantages

a. Easily shows how different anime types stack up in terms of scores.

b. Bar heights provide an instant sense of which type outperforms or underperforms others.
c. Works well with filters and can be combined with other charts in a dashboard.

Disadvantages

a. Only displays a single metric (score) by type.

Rejected for being overly simplistic and encoding very few attributes.

Change in Must-have features

One of the visualizations we proposed was to create a heatmap in which the user can select the state they want to focus on. The other proposed visualizations - bar chart, radar chart and bubble chart will be filtered based on the selected state. This visualization will be complex with multiple intermediate components -

1. Thorough data preprocessing involved
    a. Identification of valid locations in the data
    b. Filtering the data only for the USA
    c. Extracting latitude, longitude positions and state from the data
2. Use a D3 projection to convert CSV coordinates to map coordinates.
3. Compute heat intensities from the point data and overlay a color gradient accordingly.
4. Implement event listeners on the plotted points for state-based filtering
5. Chart Synchronization: Filter and update the linked bar, bubble, and radar charts based on the selected area or cluster.

While we have completed the data preprocessing for this, given the complexity of this visualization, we will now include this in optional features and not must-haves as suggested in proposal feedback.

Implementation Details

The project leverages a modern web development stack to visualize and interact with anime consumption data:

- **Frontend Framework**: React.js – Used for building modular, reusable components and handling UI state.
- **Data Visualization Library**: Recharts – A composable charting library built on D3.js, optimized for React. Ideal for declarative chart definitions.
- **CSV Parsing**: Papaparse – A fast and powerful CSV parser for JavaScript used to fetch and process raw data directly from public directories.
- **Data Format**: Comma-Separated Values (CSV) files, served from the project's `/public` directory.
- **Styling**: Inline React styles with additional scoped layout styling for consistency and responsiveness.

Advantages of the Setup:

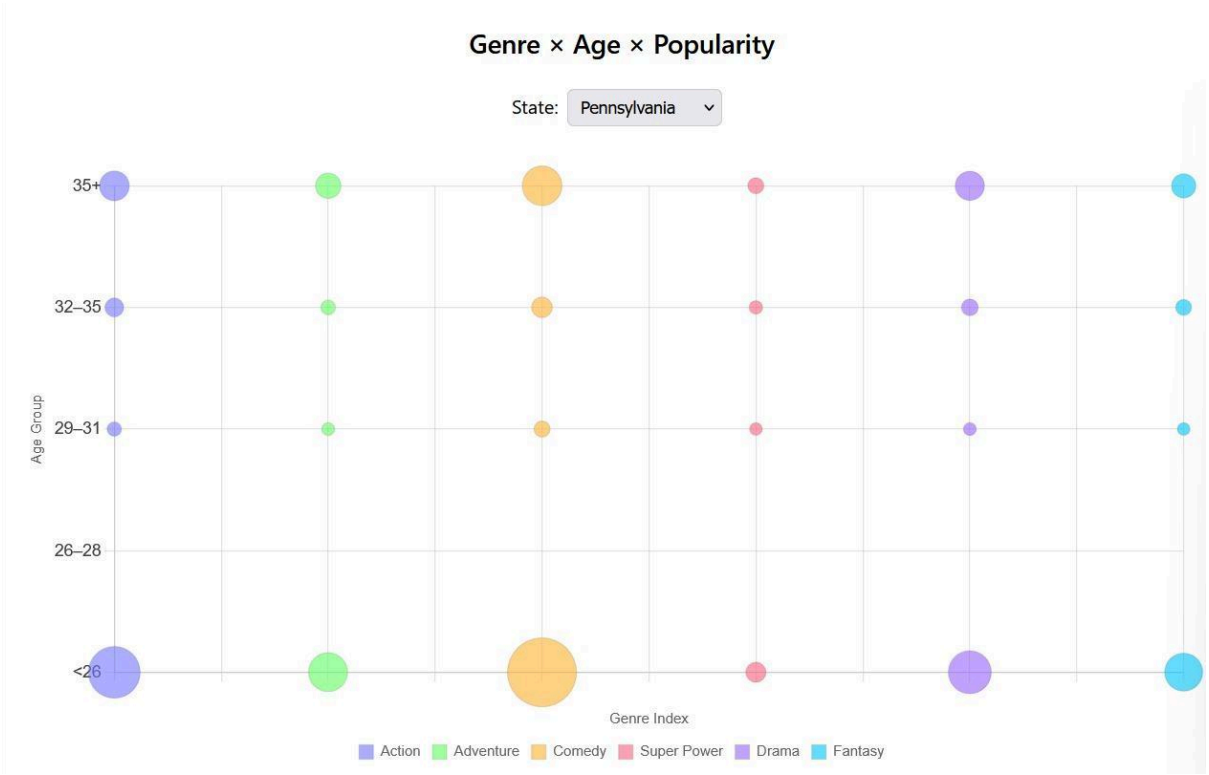
- **No Backend Required**: The app remains fully static and serverless while allowing dynamic behavior through CSV-driven data.
- **High Performance**: Local CSV parsing avoids the need for external APIs or slow server responses.
- **Extensible Design**: The system supports future data expansion (e.g., adding more countries, genres, or metrics) with minimal code changes.

Integration Flow:

Each chart component loads structured CSV data from the project's public directory using Papaparse. This data is parsed on the client side within `useEffect` hooks, ensuring real-time rendering without the need for a backend. The raw data is pre-grouped based on relevant dimensions such as gender, age group, country, or producer, then transformed into a format compatible with Recharts components.

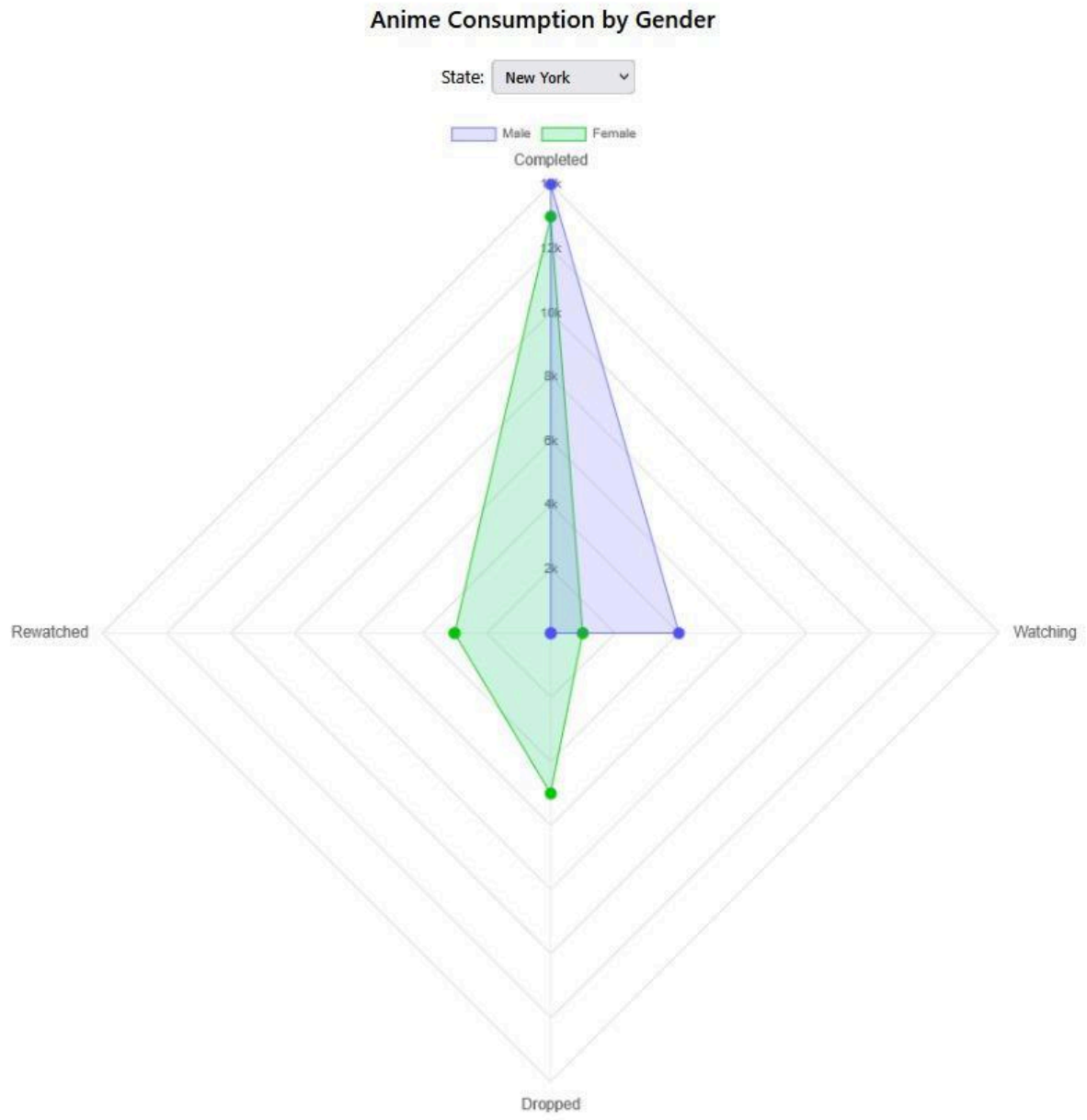
Once parsed, the data is passed into corresponding Recharts visualizations—Radar, Scatter (Bubble), and Bar charts—wrapped in `ResponsiveContainer` for adaptive design. Additional logic maps categorical data (e.g., genre, age) to numerical axes and assigns consistent colors to producers. Each component includes interactivity through dropdown filters, tooltips, and legends, making the charts dynamic and user-driven while maintaining a lightweight and scalable front-end structure.

Below are the initial visualizations of our website, we have a bubble chart, radar chart and a bar chart which dynamically show information regarding the different states of the United States.

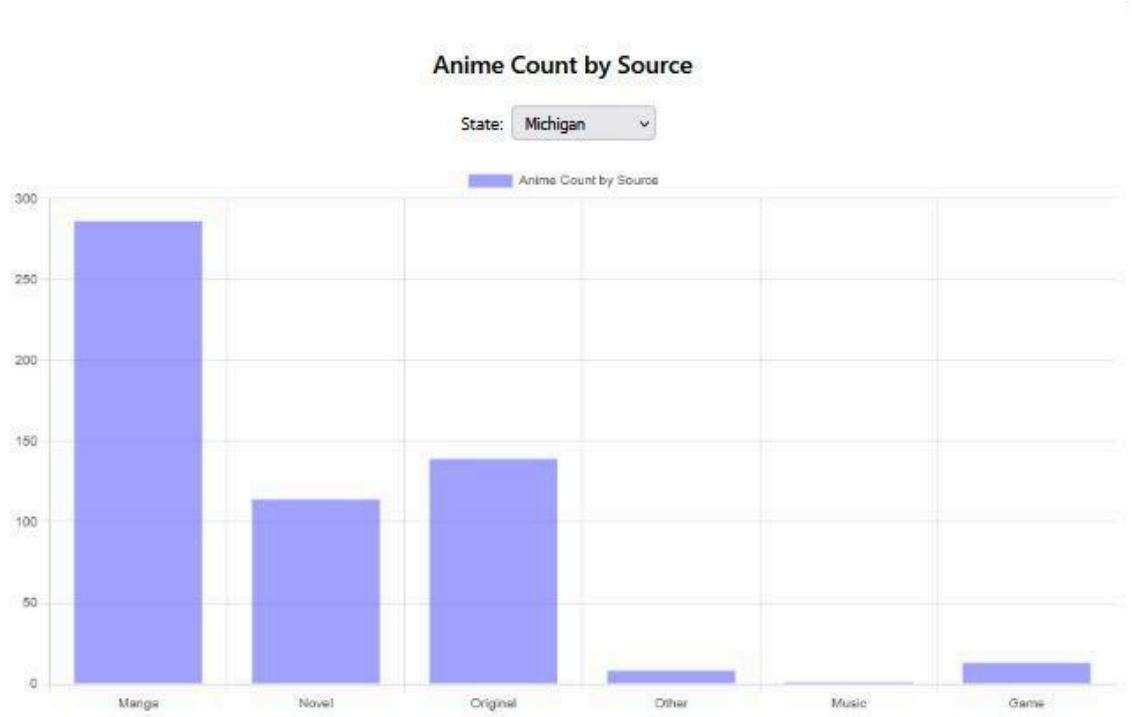


To analyze genre preferences across age groups, we implemented a bubble chart visualization titled “Genre × Age × Popularity”. This plot dynamically reflects user preferences in the states of the United States(the above doing it for Pennsylvania). Each bubble represents a specific genre-age group pairing, with its size indicating the relative popularity of that genre within the group. Colors distinguish different genres, allowing easy comparison across the grid. This visualization aids in identifying dominant genres among different age demographics, such as the strong preference for *Comedy* and *Action* among users under 26 for this particular state.

- The X-axis maps to anime genres (e.g., Comedy, Drama).
- The Y-axis corresponds to age groups (e.g., 13–17, 18–25).
- Bubble size (Z-axis) indicates popularity count. Multiple producers (e.g., Manga, Novel, Game) are visualized in color-coded layers for comparative clarity.



To examine anime consumption trends across genders, we implemented a radar chart titled “Anime Consumption by Gender.” This visualization maps consumption behaviors such as *Completed*, *Watching*, *Dropped*, and *Rewatched* across male and female users. Each axis represents a category of engagement, while the area covered by each color-coded polygon (blue for male, green for female) denotes volume. This format allows for intuitive comparison of patterns—highlighting, for instance, that while both genders show a strong tendency to complete anime series, females exhibit slightly more diversity across categories like rewatching and dropping.



The “Anime Consumption by Gender” radar chart visualizes gender-based patterns in anime engagement across four categories: *Completed*, *Watching*, *Dropped*, and *Rewatched*. This comparative analysis uses overlapping radar plots—blue for male and green for female viewers—to show the relative volume of engagement in each category. The chart highlights that both genders exhibit a dominant preference for completing anime series, with males slightly ahead. However, females tend to have higher activity in dropping and rewatching shows, suggesting broader interaction beyond simple completion. This form of visualization makes it easy to capture nuanced behavioral differences across user demographics.

Evaluation

Through our visualizations, we gained valuable insights into the world of anime consumption, particularly how user ratings and preferences are influenced by factors like genre, age, gender, location, and source material. We found that certain genres are consistently favored across age groups, while others are more niche and dependent on specific demographics. For example, action and fantasy genres tend to receive higher scores across most groups, while genres like slice-of-life or romance show more variance based on gender and age.

To answer our research questions—such as whether certain types of anime are more favored by specific demographics, how user engagement (rewatches, drops) affects rating patterns, and how anime rankings shift over time—we used a combination of heatmaps, radar plots, bubble charts, and bar graphs. These visualizations made abstract patterns more tangible, such as identifying representation gaps and potential biases in user behavior and anime production.

The visualizations worked well overall in conveying the story behind the data. For example:

- Radar plots helped compare engagement types across gender.
- Bubble charts allowed us to see the intersection of age, anime source, and viewing preferences.
- Bar charts effectively communicated categorical breakdowns.

However, there were some **limitations**:

- **Data completeness:** Not all users had complete demographic data (e.g., missing age, gender, or location), which restricted the depth of some analyses.
- **Visualization overload:** Some sheets became visually dense with too many variables, which might overwhelm a general audience.
- **Static charts:** Without interactivity, it was harder for users to explore trends on their own (e.g., filtering by genre or demographic).
- **Bias in ratings:** The dataset was skewed toward active users on MyAnimeList, which may not represent the broader anime-watching population.
- **Display bug:** The bar chart for Michigan appears empty by default. It only renders correctly when a user switches to another state and then returns to Michigan. This appears to be a rendering issue that needs debugging to improve user experience.

To further improve the project:

- We could incorporate interactivity using tools like Tableau Public dashboards or D3.js for drill-down analysis.

- Cleaning and enriching the demographic data further would allow for deeper segmentation.

- Using animations (like the poster reveal with trophies) could improve the narrative experience.

- More context/annotations on charts would help viewers unfamiliar with the data draw conclusions quickly.

Must Have Features

1. A key aspect of understanding customer behaviour is to analyze gender bias amongst anime viewers. It can be very well captured with a radar map, with even further segmentations of watched, rewatched, dropped, watching etc.
2. We also want to analyze how preferences towards anime, manga, comics etc. changes with age. We will segment it further by producers to understand if certain producers are popular amongst specific age groups. This can be best done with a bubble chart, with sources on the x-axis, age bins on the y-axis, the size of the bubble representing.
3. We also want to include the number badge for displaying the total number of unique anime across the globe
4. We also want to include the trophy icon for the top 3 most popular anime across the globe which is based on the popularity and rank metric.
5. We can use the bar chart visualization for **source vs. anime count to effectively** display the distribution of anime based on their origin, such as manga, light novels, and original works. This allows users to quickly identify the most common sources of anime adaptations.

Optional Features

- Heatmap highlighting which locations have the greatest number of anime lovers. This information is crucial because it tells us which countries have low customer engagement and can thus help us advertise it better in those countries
- We can have a scatterplot for effectively correlating the user-given scores with average anime scores, using color to represent anime popularity and size to indicate the volume of user ratings. This allows for a quick identification of outliers and emerging trends in user preferences.
- We can put anime posters in animation based on trophy colour i.e the 1st one will come on the screen first and with an interval of 5 seconds the second poster will come, post this the third poster with a bronze trophy beside it, for better aesthetics
- We can use the bar chart visualization
- We can also use scores given by users against the different types of the anime (e.g., TV series, movie, OVA, etc.) This gives us which type is most preferred by the users.
- We can also do another heat map between age group bins and the source of the anime. This tells us a lot of information about which source is most preferred and least preferred amongst each age group bin we have.
- We can include the bar chart visualization for gender, number of users, and age group effectively representing user demographics by showing the distribution of users across different gender categories while simultaneously highlighting age group segmentation using color coding.
- We can include the pie chart for visualizing the distribution of anime types (TV, Movie, OVA, Special) to show their relative proportions.

Project Schedule

Announce Your Project (**13th Feb**)

- Finalize the dataset to be used for analysis.
- Define project goals and key research questions.
- Assign roles and responsibilities to team members.
- Set up the GitHub repository and project documentation.

Project Proposal (**28th Feb**)

- Outline the methodology for data collection, preprocessing, and visualization.
- Identify potential biases and key factors influencing anime ratings.
- Plan visualization techniques (heatmaps, demographic trends, rating shifts).
- Submit a formal project proposal for feedback.

Project Review (**March 10**)

- Complete dataset preprocessing (handle missing values, standardize formats).
- Conduct exploratory data analysis (EDA) to identify trends and biases.
- Implement basic visualizations to understand rating distributions.
- Get initial feedback from peers/instructors to refine approach.

Project Milestone (**April 3**)

- Develop and refine core visualizations (e.g., genre trends, user behavior insights).
- Analyze biases in anime ratings based on demographics (age, gender, location).
- Investigate anime-watching patterns (completion/drop rates, seasonal trends).
- Begin integrating findings into a structured narrative.

Peer Feedback (**April 16**)

- Conduct internal peer reviews to evaluate clarity and accuracy.
- Test visualization usability and refine based on feedback.
- Ensure that insights are clearly communicated through data representation.

Project Screencast (**May 2**)

- Record a walkthrough video explaining the analysis, visualizations, and insights.
- Highlight key findings and showcase the impact of user-driven ratings.
- Ensure the screencast is polished and accessible for viewers.

Final Project Submission + Group Feedback (**May 6**)

- Integrate visualisations and include them into the final report.
- Ensure clarity, consistency, and proper documentation of methods.
- Submit the final project along with a GitHub repository update.
- Deliver a presentation summarizing key findings and takeaways.