

Taylor Swift's Eras Tour

A visual exploration of the biggest musical event of the year

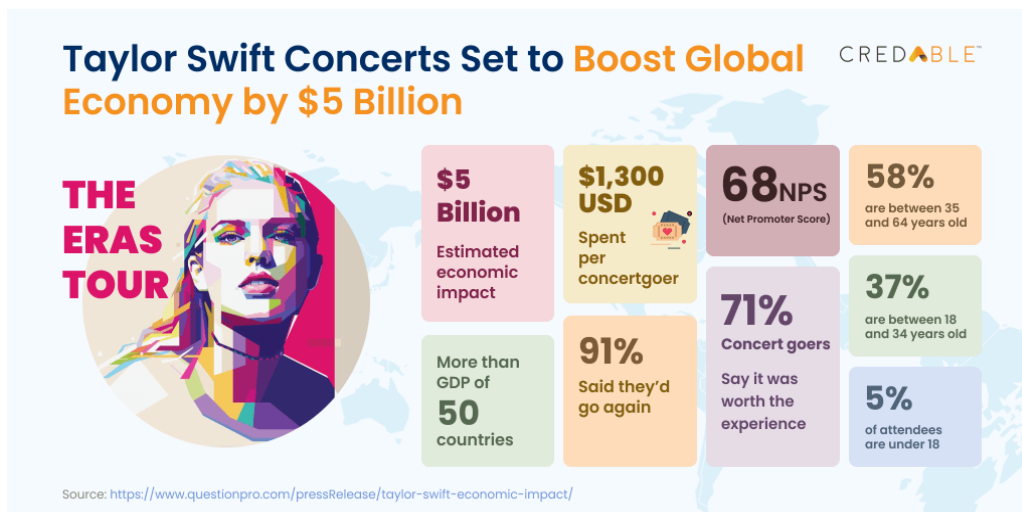
Team Members: Abhishek Dhar, Sanjana Jairam, Prithviraj Maurya, Snehal Patil

Motivation

Taylor Swift's Eras Tour has been one of the biggest events of 2023, so much so that she was awarded the Time Person of the Year. From drawing attention to Ticketmaster's non consumer friendly tactics, to boosting NFL viewership through dating rumors, Taylor Swift's impact was felt across various industries. Her tour also generated significant economic revenue within each venue's city, with an estimate stating that an average attendee spent about \$1,300 in tickets, hotel bookings, fares, merch etc. Another estimate states that the tour generated over \$5 Billion in economic impact, more than the GDP of 50 other countries. This exorbitant effect has leaders of lesser developed countries asking for tour dates in their region.

Due to the overwhelmingly large influence this show had on our culture, we chose to explore this summer phenomena in order to better understand the following objectives:

1. The intricacies of the show: How are the shows structured?
2. The show's global reach: Where have/will the performances taken/take place?
3. The revenue generated per show: How much money did the shows perform on average?
4. The streaming metrics of her songs and whether they influence the frequency of a song being played live.

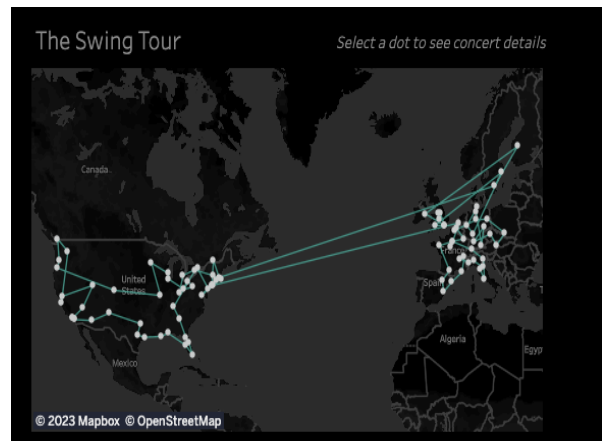


Prior Work

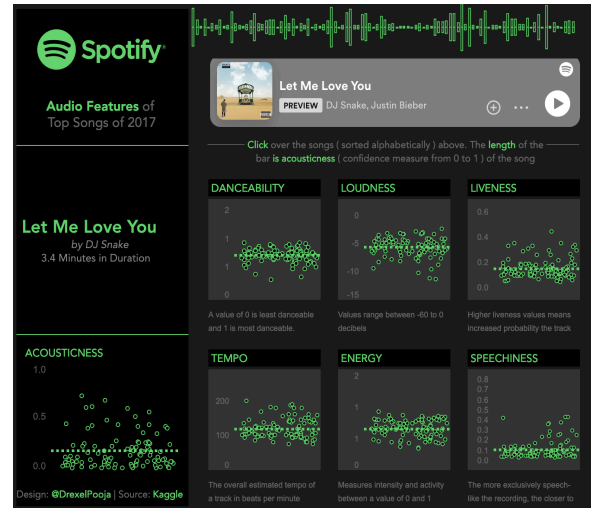
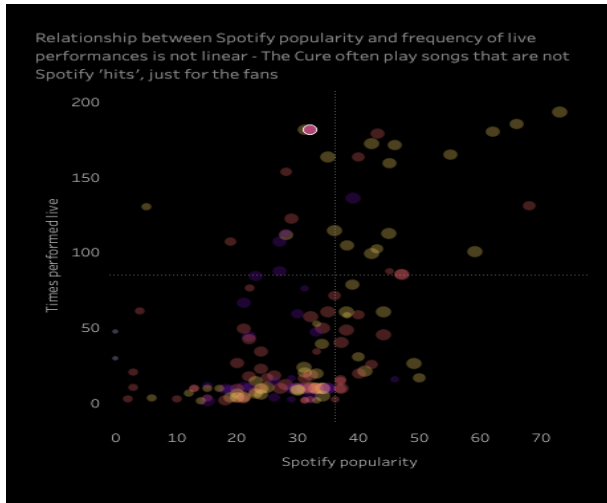
While there's been an abundance of articles discussing Taylor Swift's Eras Tour, many fall short in bringing the tour's impact to life through concrete visualizations. Various sources touch on aspects such as ticket sales, flight prices, and hotel expenses, yet few take the initiative to present this wealth of data in a visually compelling manner. For instance, [the NY Post's article](#), 'Rocket sleds, stunts, and a \$100M budget: Inside Taylor Swift's 'unbelievable' Eras Tour,' offers intriguing insights but lacks visual representations to truly capture the 'unbelievable' essence of the tour.

A separate [article](#) delves into Taylor Swift's discography, providing a comprehensive analysis of each album and dissecting her songs in intricate detail. The author successfully uncovers fascinating insights such as the most frequently used Major and Minor Keys, an extensive exploration of lyrical content, and collaborations with renowned songwriters. However, while she touches on the tour in relation to Taylor's albums, there's a missed opportunity to establish connections between the tour and the individual songs. Essential questions, such as the composition of Eras, the sequence of songs performed, and how metrics like danceability or loudness evolve throughout the tour, remain unexplored.

There have also been visualizations on various aspects around live performance for other artists that served as inspiration for some of our visualizations. For example, ['The Cure's 40-Year World Tours'](#) by Nina Lindell is an interesting visual depiction of all of The Cure's tours across their career. She uses a variety of visualization techniques such as a world map to show the many venues they performed at for a tour or a scatter plot to compare the number of times a song is performed live vs the song's Spotify Popularity. Both of these charts in particular are effective in demonstrating their stated objectives. The map uses the contrast in colors between the map and the (path between) venues to depict the enormity of a tour and the distances the band would have to travel for a given tour cycle. Meanwhile, the scatter plot is a good use of displaying the relationship between 2 dimensions of data (popularity and occurrence).



Another related visualization is titled ['The Secret Sauce in Spotify's Tracks'](#) by Pooja Gandhi which attempts to visualize various Spotify metrics for the top tracks as shown below. She constructs a small multiples visualization where each graph is a one dimensional horizontal strip plot for each metric. While interesting, the plot makes it difficult to understand the relationships between the metrics for a given song as it's impossible to identify the song from the plot. We attempt to improve upon this visualization by using a parallel coordinate plot to show this relationship between metrics for a given song, while also maintaining the overall relationship of the metric across all songs.



Contributions

As shown through the prior work above, we've demonstrated that there aren't enough visualizations surrounding the Eras Tour that provide readers with a clear understanding of the impact and details of the shows. Our hope with this project is to do just that by designing visualizations around every aspect of the tour: from the economic revenue generated, the locations of each show, and the music of the show itself. We also hope to introduce a more effective approach to the common task of visualizing various spotify metrics across songs.

Data Collection

Our project commenced with an exhaustive exploration of diverse data sources, meticulously consolidating them into a unified dataset tailored for our visualizations. This strategic amalgamation aligns seamlessly with the multifaceted nature of our project, driven by inquiries that intersect Taylor Swift's songs and her tours. The following constitutes the primary data sources we identified:

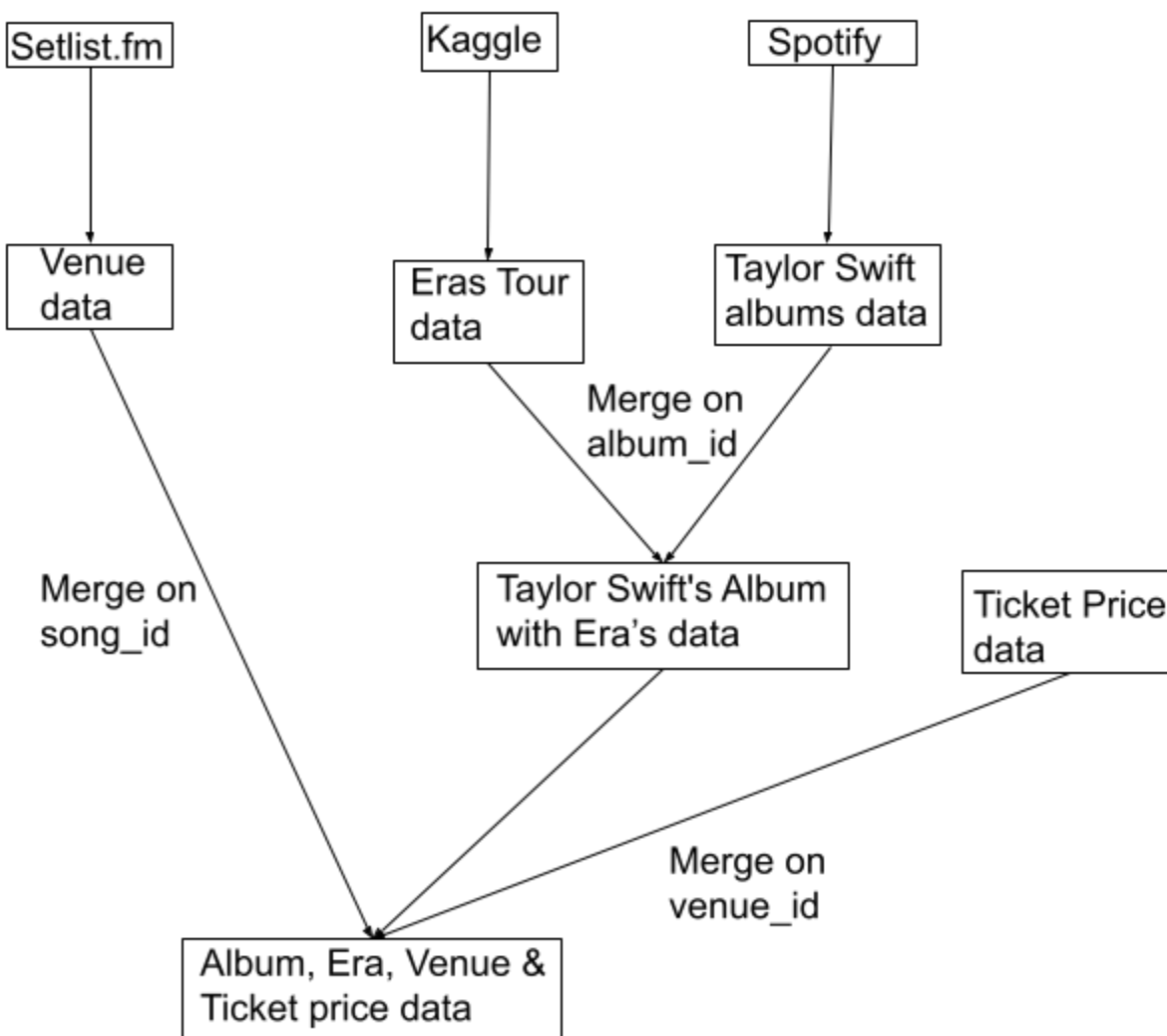
Setlist.fm API: This source furnishes data on setlists and show locations for a given artist. It serves as a vital repository, offering crucial insights into the locations of Taylor Swift's performances and the repertoire of songs featured at each show.

Spotify API: Leveraging data on artists and songs, including metrics like popularity, genre, and audio features, this API facilitates the mapping of each performed song to its corresponding Spotify metrics. This integration enriches our analysis by incorporating diverse dimensions of a song's characteristics.

Kaggle: A comprehensive dataset encompassing all of Taylor Swift's songs performed during each show in the Eras tour. This Kaggle dataset serves as a valuable foundation, providing a detailed record of song performances across various tour dates.

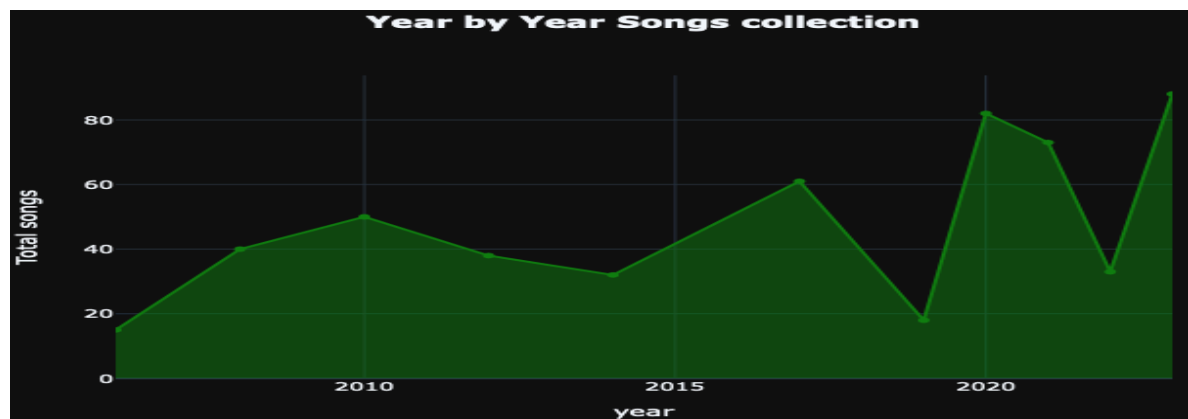
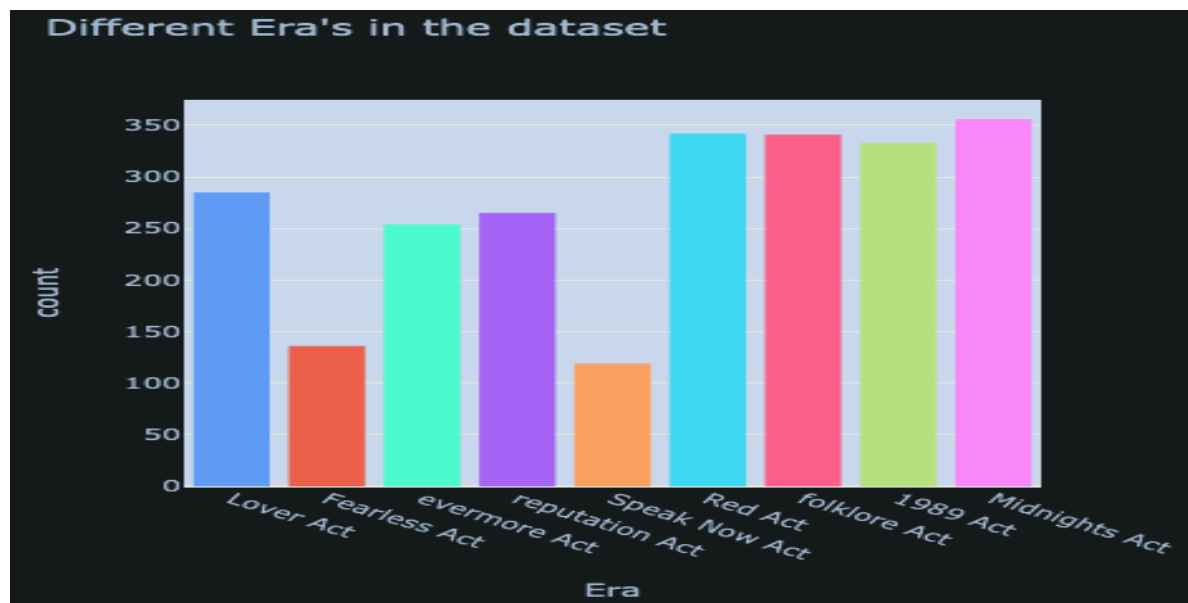
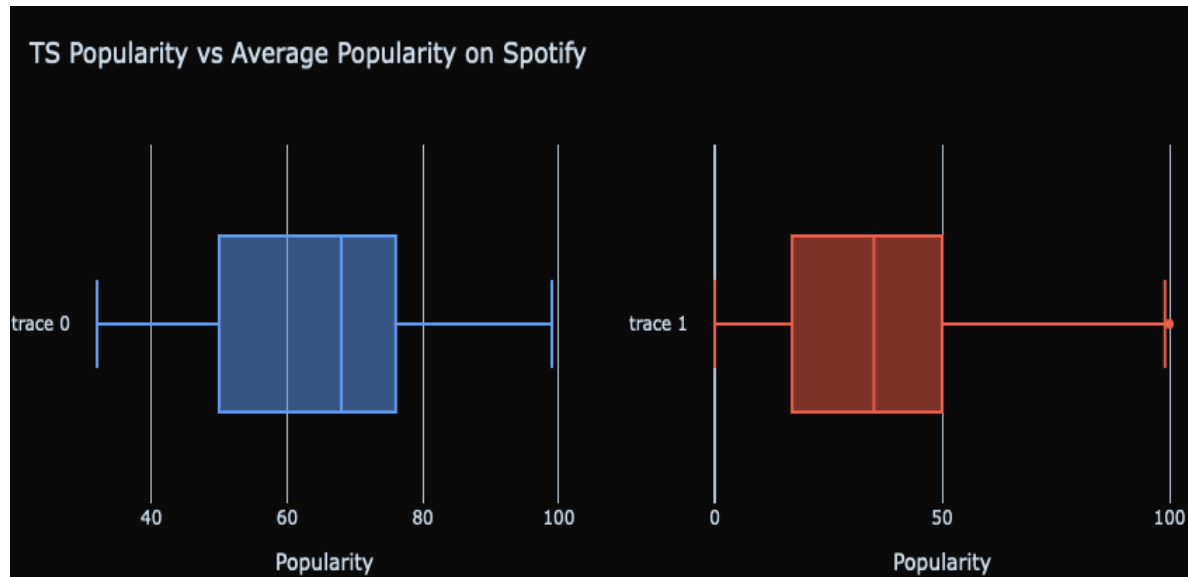
Survey of Ticket Prices: A 900+ person survey conducted by users of the Taylor Swift subreddit. This dataset provided information around the costs of a ticket according to the venue. This helped us better understand the economics behind the Eras Tour.

Data Processing Step:



Exploratory Data Analysis Step:

The subsequent phase involved visualizing the amassed data to comprehend its nuances. Employing Kaggle notebooks, we conducted Exploratory Data Analysis (EDA) on each dataset. This analytical step proved instrumental in extracting insights into various facets, including song attributes, lyrics, tour dynamics, and the overall composition of her performances. The ensuing visualizations encapsulate a glimpse of our exploratory journey.



Visualization methods:

Section 1: Understanding the Intricacies of the Eras Tour

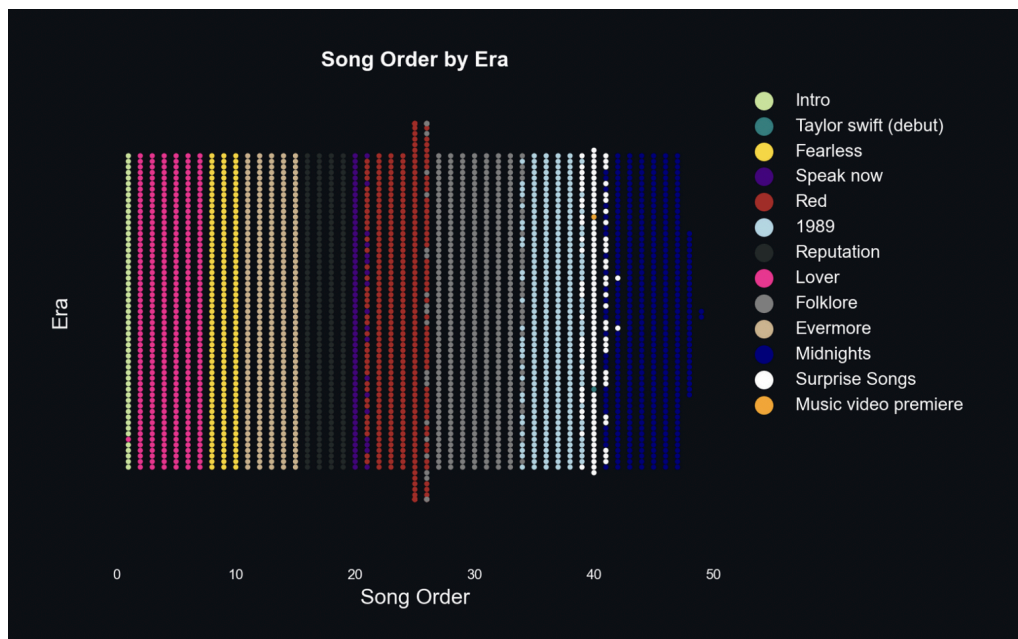
Visualization Type: Swarmplot, Bar Chart

Tech Stack: Pandas, Plotly, Streamlit (Python)

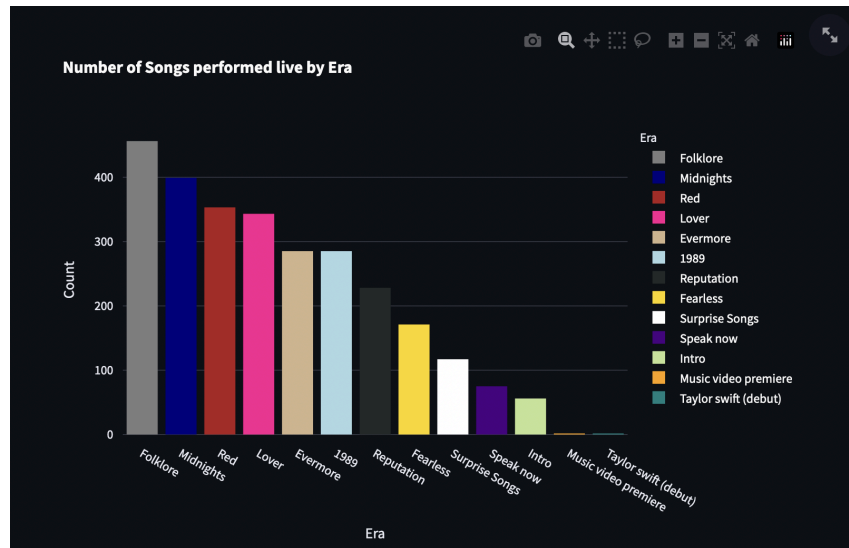
Data Source: Taylor Swift Eras Tour Data

The purpose of this section is to introduce the audience to The Eras Tour. Taylor Swift's Eras Tour is a 3+ hour long show that encompasses songs across all of her albums ("eras"). In order to best visualize this concept, we created a swarm plot (shown below) of all the songs performed on tour arranged in order of when they occurred in the show and colored by the era they belonged to. The symmetry of the data points show that the tour followed a strict structure, with each performance having a similar number of songs. The coloring of points best visualizes the concept of "Eras" as it shows that Taylor plays songs from the same album back to back in chronological order.

Visualization 1: Swarm Plot of Song Order by Era



Visualization 2: Bar Chart of Number of Songs Performed Live by Era



The next visualization is a bar chart showing the number of songs performed live by Era. We chose to go with a bar chart since we're trying to display the frequency count of categorical data and that length is the best way for the reader to perceive differences in magnitude. From this chart, we can see that "Folklore" had the most songs performed live. This can be because Folklore was the highest selling album in 2020 and responsible for her critical acclaim into the Alternative/Indie genre which brought in a diverse set of fans into her repertoire.

Section 2: Global Reach - Visualizing the expansive network of shows across the world and understanding its economic impact.

Visualization Type: Maps, Bar Charts, Radar plot

Tech Stack: D3.js, Plotly, Pandas

Data Source: Taylor Swift Eras Tour Data, 900+ Fans Surveyed about ticket prices

Observable notebook: <https://observablehq.com/@prithviraj/taylor-swift-the-eras-tour>

Introduction

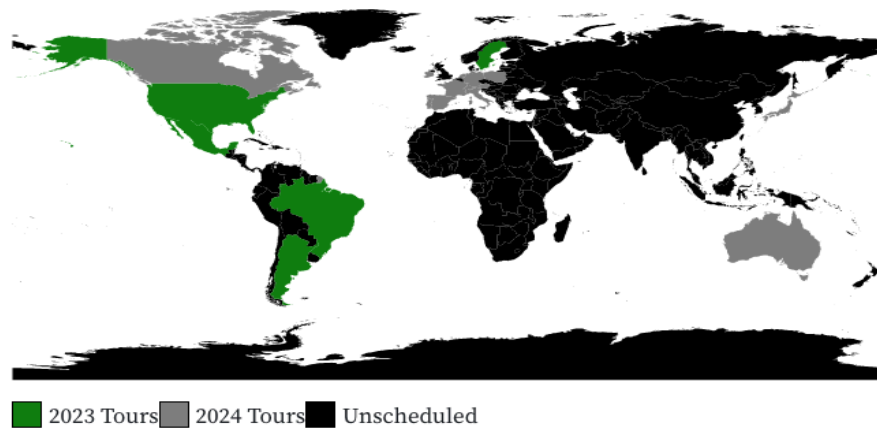
The Global Reach project delves into the dynamic landscape of Taylor Swift's Eras Tour, employing advanced data visualization techniques to bring to life the extensive network of shows worldwide. Leveraging technologies like D3.js, Plotly, and Pandas, the project extracts insights from the tour data, exploring not only the geographical expanse but also understanding its economic implications.

Data Collection

The data for the project was meticulously collected by scraping a blog post detailing the countries, cities, stadiums, dates, and the number of shows for Taylor Swift's Eras Tour in 2024. This information was compiled into a structured CSV file, serving as the foundation for developing insightful visualizations.

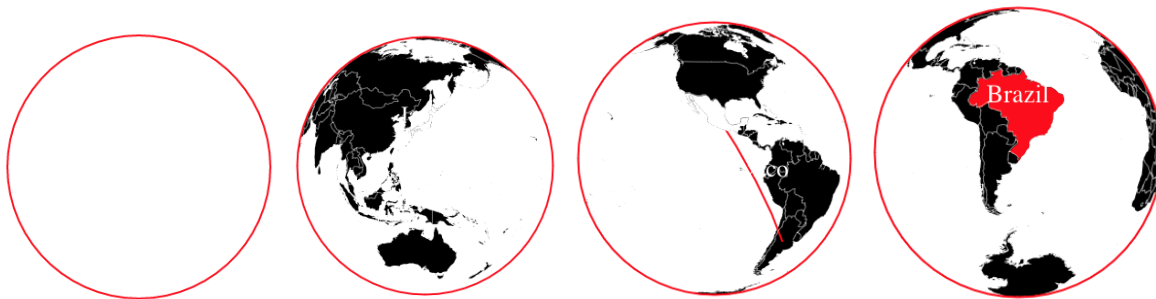
Visualization 1: Categorized Map of Touring Countries

The first visualization presents a comprehensive map of all countries included in the Eras Tour. Utilizing Plotly's geo capabilities, the coordinates of countries were projected onto the map. The equirectangular projection was employed for its ability to preserve relative sizes, ensuring an accurate depiction of the geographical distribution.



Visualization 2: Animated Tour Arcs on a Globe

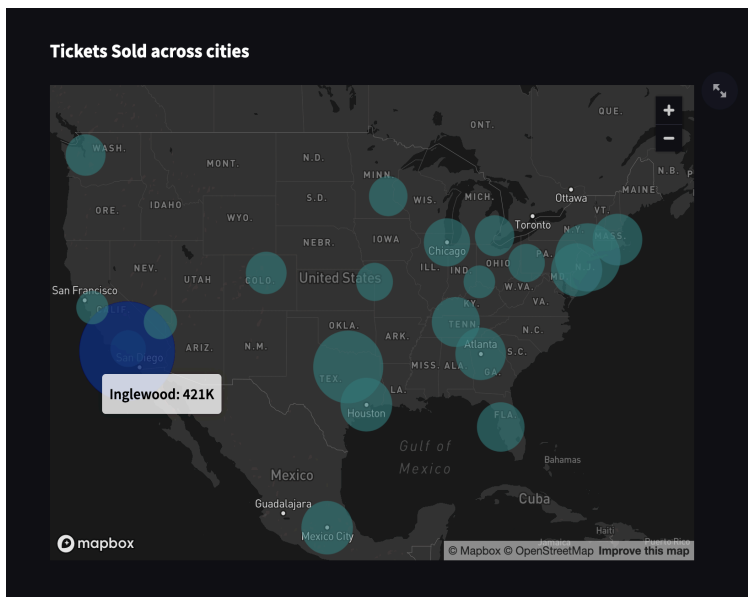
The second visualization introduces an animated map, showcasing arcs that connect countries based on the chronological order of tour dates. Drawing inspiration from the D3 [documentation](#), a canvas-based globe was constructed, dynamically moving to represent the changing coordinates of countries over time. We started by an empty globe on canvas then showing the country boundaries, then by putting arcs from one country to another. Coloring the current country and then adding the animation effect by rotating the globe.





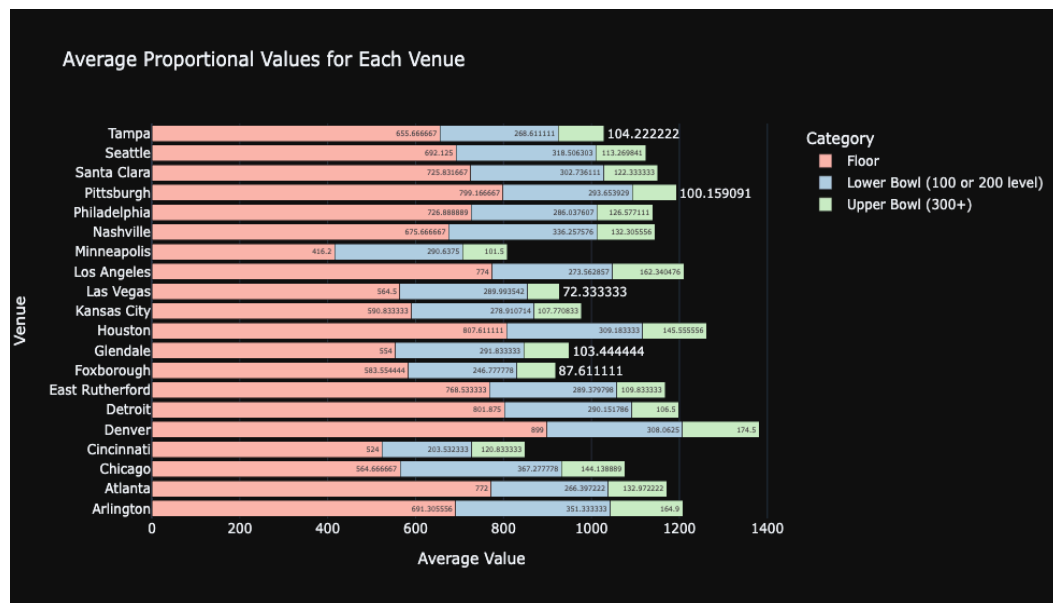
The map visualizations provide a rich understanding of the countries involved in the Eras Tour, offering insights into the global scale of Taylor Swift's performances. The animated arcs contribute a temporal dimension to the narrative, allowing viewers to follow the tour's journey over time, creating an engaging and dynamic visualization. While not explicitly visualized, the understanding of the tour's economic impact can be inferred from the extensive global reach and the number of shows in each country. The Global Reach project successfully captures the essence of Taylor Swift's Eras Tour through compelling visualizations. The combination of categorized maps and animated arcs provides a holistic perspective on the tour's global impact, making it a valuable tool for both fans and analysts.

Visualization 3: Map of tickets sold across the USA



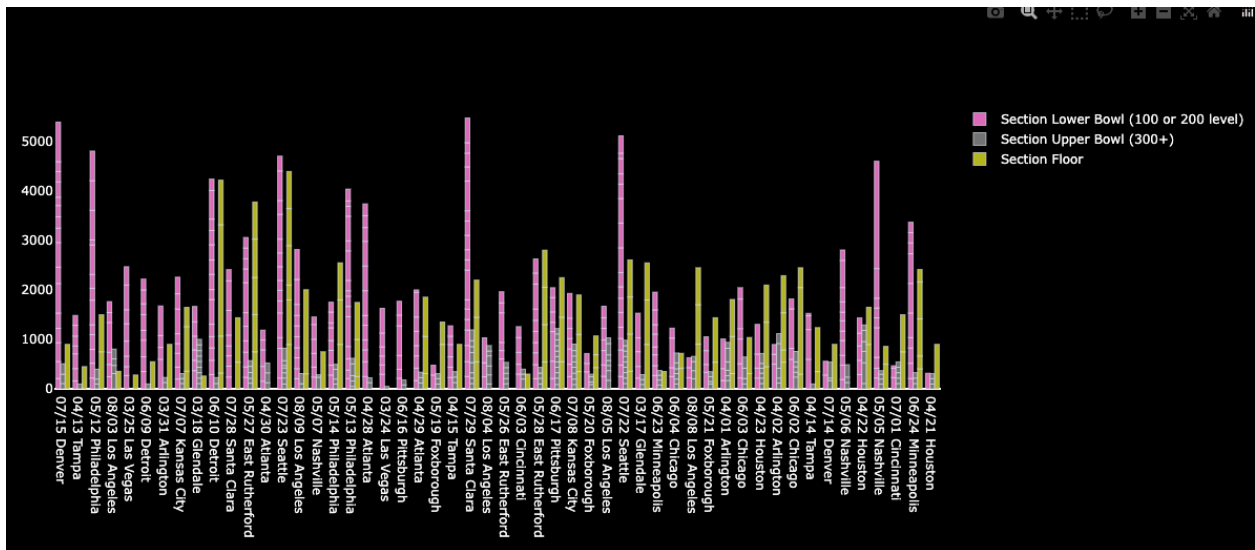
In the above visualization each uniform-colored point on the plot represents a city hosting one of Taylor's electrifying Eras Tour shows, with the area of the circle indicating the magnitude of tickets sold. We added the alpha channel to resolve the issue of circles getting overlapped. Hovering over any point reveals a pop-up window displaying the city's name and the exact number of tickets sold, providing an interactive experience. The scattered arrangement mirrors the geographical diversity of concert venues, allowing users to witness the tour's impact from bustling metropolises to more intimate settings. It's like a visual diary of the tour's success. We can see from the visualization that Inglewood sold the most tickets.

Visualization 4: Bar Chart of Average Proportional Values for Each Venue



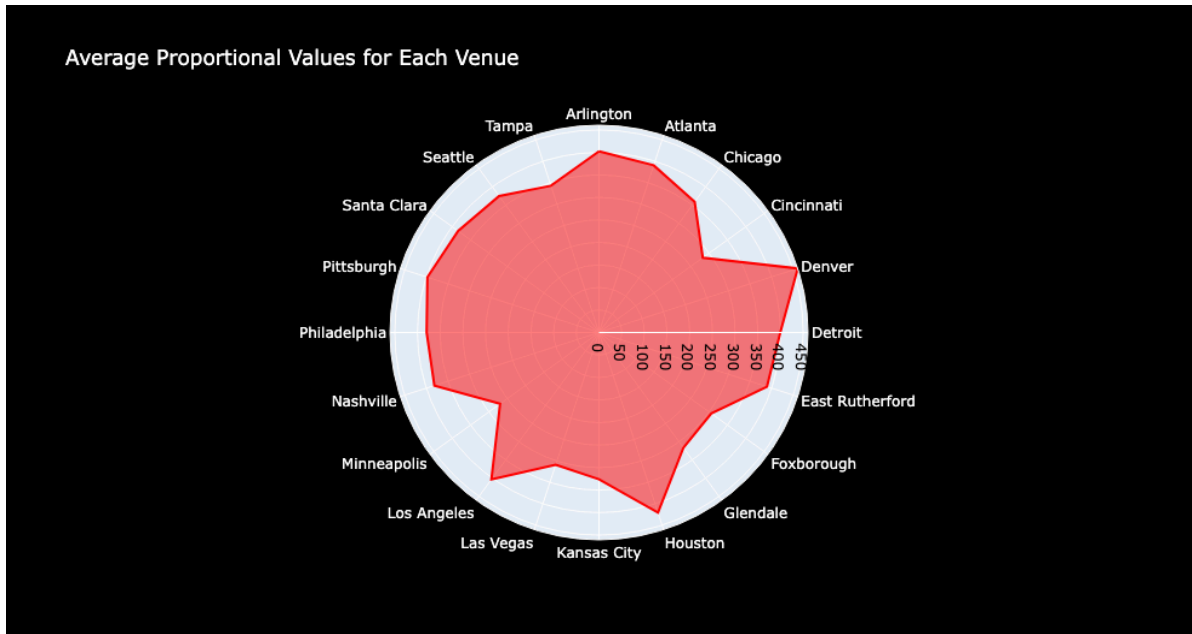
The above visualization shows us average ticket prices at different concert locations based on the seat category. From the visualization we get the insight that 'Floor' category seats have the highest price which makes sense as these seats are closest to the concert stage. The first(lower) level seats have average prices and lowest are upper level seats. Most expensive tickets are sold in Denver where average price can be somewhere around \$899 to lower level \$308 and upper level \$174.5.

Visualization 5: Bar Chart Breakdown of Tickets Sold Across Venues



The above visualization shows us revenue generated at different concert locations based on the seat category. From the visualization we get the insight that lower level category seats have been purchased the most then floor and at last upper level. This visualization is made from survey data of 900+ people who went for her concert. Even with this small set of data we see the majority of bars are above \$2500 and a good amount above \$3000; this shows how huge revenue Taylor swift's Eras tour concert generates.

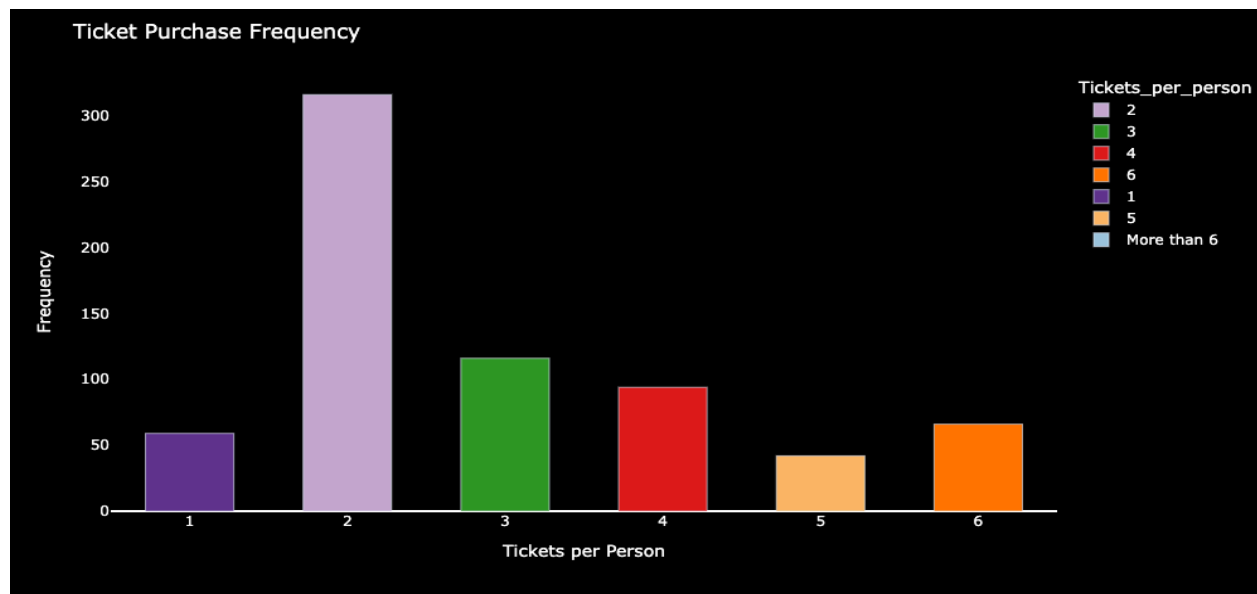
Visualization 6: Radar Chart of Average Proportional Values for Each Venue



The visualization above depicts average ticket prices across various locations in the US. For instance, if you plan to attend a concert in Detroit, expect ticket prices to hover around \$400, contingent upon your chosen seat category.

To generate this visualization, we initiated a feature engineering process, creating a new column to represent the average price for each location. Subsequently, we employed this average price as the parameter for the radial axis. The 'Venue' column serves as the theta axis, and no specific alpha channel was utilized in the creation of this visualization.

Visualization 7: Bar Chart of Ticket Purchase Frequency



The above visualization tells us in general if a person buys a ticket how many tickets they buy for Taylor Swift's concert. And through the above visualization on survey data we found people mostly purchase 2 tickets followed by 3 and 4. Out of 900+ entries 50+ times people bought 6 tickets.

Section 3: Encoding the Magic of Live Shows

Visualization Type: Small Multiples, Line Graph, Parallel Coordinate Plots

Tech Stack: Pandas, Plotly, Seaborn, Streamlit

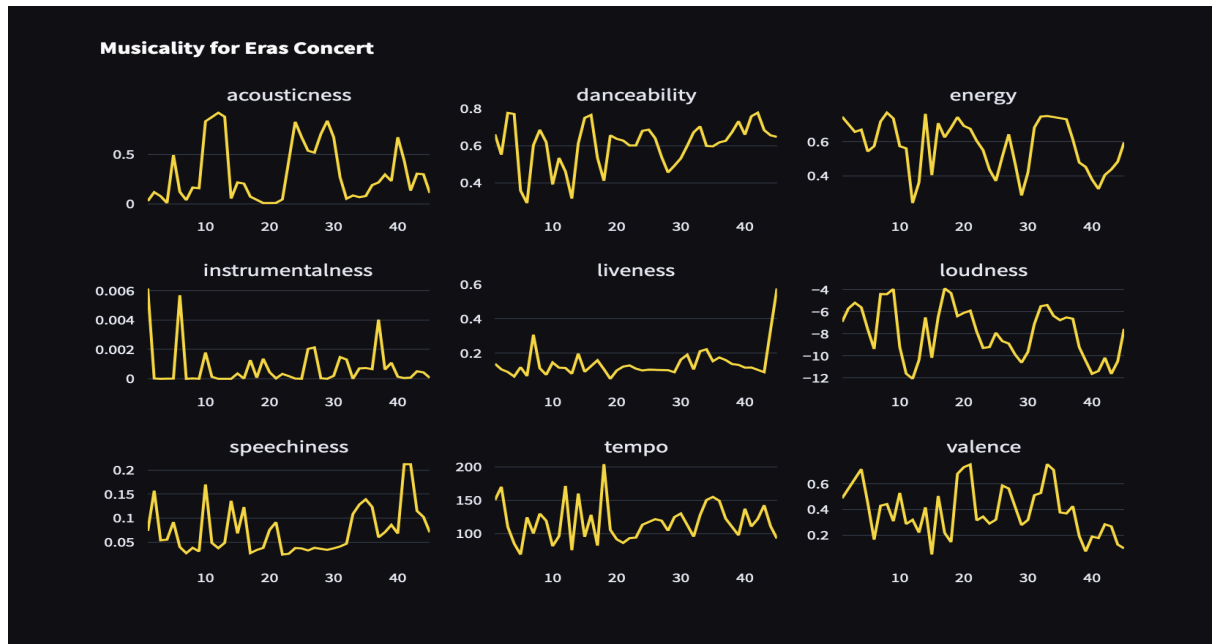
Data Source: Taylor Swift Eras Tour Data

Streamlit app: <https://erastour.streamlit.app/>

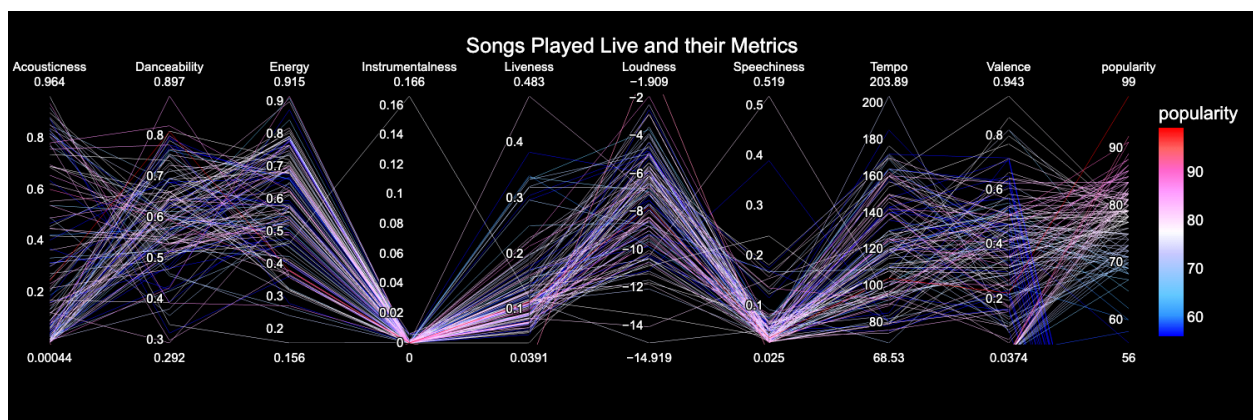
The purpose of this section is to quantify the characteristics of all the songs played live in order to get a better understanding into how a show performed by Taylor Swift might feel like. To do this, we referred to Spotify's metrics for each of the songs to conduct our analysis. In order to visualize these changes in metrics over time, we determined that a line plot for each metric would best represent the chronological nature of experiencing the different songs throughout the course of the show. To prevent messiness, we chose to display each plot on its own using a small

multiple approach so that the characteristics can be compared in a neat manner. Each plot shows the average value for that particular metric at a given point in the setlist. From this graph we can see that there are peaks in the “acousticness” around the 10 and 20-30 song mark in the show. The main reason the acoustic songs are placed this way can be to give the audience some time to rest. This is because in a long show you can often get tired after high energy songs and you might need some slow energy songs to re-energize you. Meanwhile, the energy is consistently high throughout the show as the danceability trends upward as the show progresses.

Visualization 1: Small Multiples of Musical Metrics across shows



Visualization 2: Parallel Coordinate Plot of Song’s Metrics



In order to display the metrics of each of the songs performed live, we chose to go with a parallel coordinate plot that's colored according to a song's popularity. As mentioned earlier in the prior works section, a small multiple doesn't help the reader gather any valuable insights from the graph. However, through a parallel coordinate plot, the reader can see trends across and within songs. For example, most of the songs played live are low in "Instrumentalness" and the most popular song also looks to be the loudest. Similarly, most of the songs are clustered together on the lower end of the "Liveness" and "Speechiness" scales. Overall, from this graph, we can see that the metrics across songs are quite varied barring for categories like Instrumentalness, Liveness, and Speechiness.

Section 4: Does Popularity Influence Song Choice?

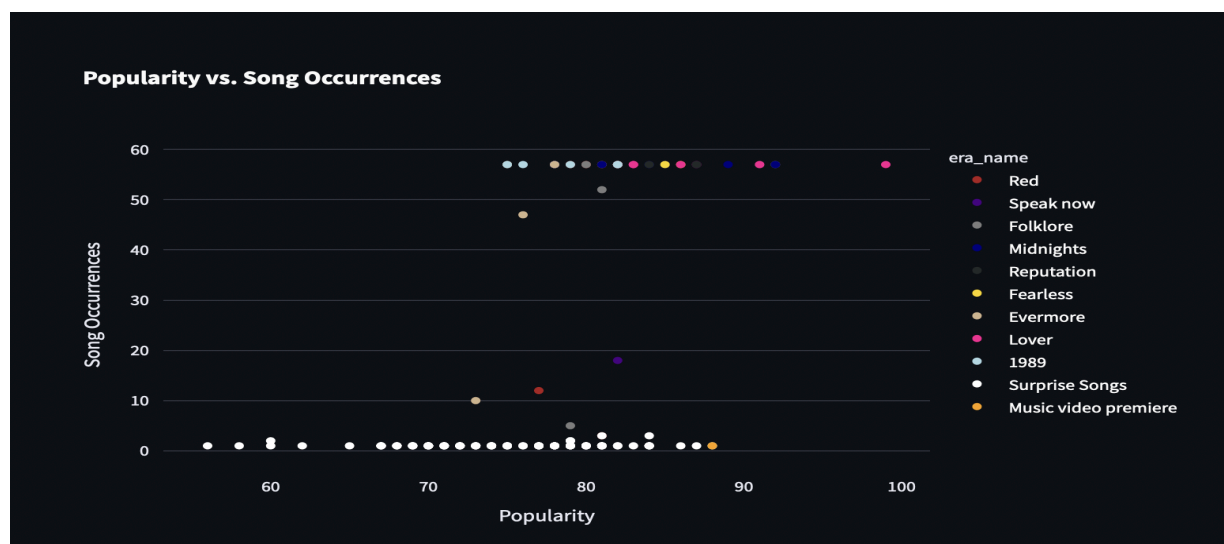
Visualization Type: Scatterplots, Swarmplots

Tech Stack: Pandas, Plotly, Seaborn, Streamlit

Data Source: Taylor Swift Eras Tour Data

The purpose of this section was to gather insights into whether the popularity of songs influences how frequently a song is played. In order to show this, we utilized two visualizations, the first one being a scatterplot that depicts the relationship between the number of times a song has been performed live and the popularity of the song according to Spotify - colored by era name.

Visualization 1: Scatterplot of Song's Popularity vs Frequency



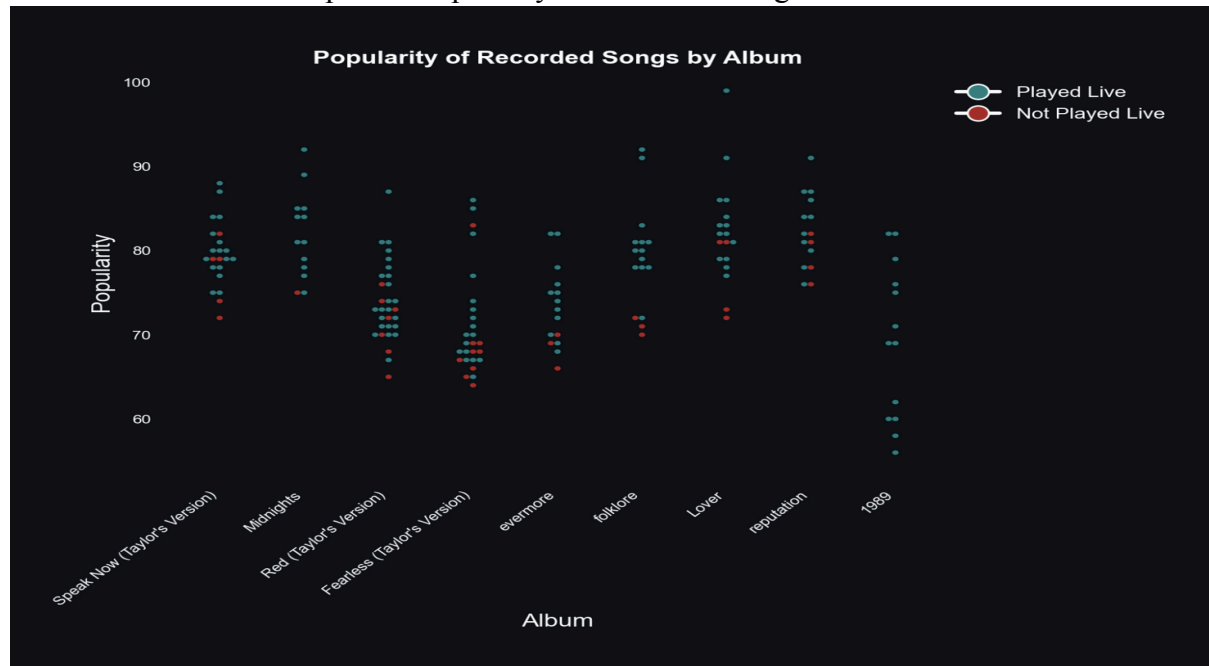
This graph provides us with some interesting observations:

1. The most frequently played songs are on the higher end of the popularity scale.
2. Taylor uses the 'Surprise Songs' section of the show to play songs that may not be all that popular compared to her other songs.

The second graph we opted to utilize was another swarm plot that showed the popularity of all recorded songs grouped by album. We then colored each point based on whether they were played live or not. This graph helps support the hypothesis that the most popular songs for a

given album are more likely to be played live. This is shown through the bottom songs for an album are often red (not played live) while the higher level songs in popularity are green (they've been played live).

Visualization 2: Swarm plot of Popularity of Recorded Songs



Conclusion:

- Through this exploration we've been able to answer all of our questions posed at the start of this report.
- Through a couple of visualizations, we were able to show that the shows were structured in a very planned manner.
- Each song was placed with intention in order to satisfy a larger purpose. For the most part, there was not much variation in songs across shows. However, there was consistently a "Surprise Section", which Swift uses as an opportunity to provide a unique experience for that particular show - sometimes even performing a less popular song.
- Through our map, we've been able to show that the tour is large in reach, spanning over four continents.
- We were also able to show that a higher popularity increases the likelihood that a song is performed at the show.
- Finally, through the small multiples and parallel coordinates plot, we were able to depict the change in song characteristics over the course of the show as well as how each song measures across each of the eight musical metrics.

Future Work:

1. Compare the gathered insights with her other tours over her career
2. Employ the same techniques on other artists' live shows to get a more holistic idea of live tours

References

- [1] [CBS News - Eras Tour Boosting Economy](#)
- [2] [Tableau Dashboard - Spotify Top Tracks](#)
- [3] [Tableau Dashboard - The Cure Band analysis](#)
- [4] [Tableau Dashboard - Heavy Metal Band analysis](#)
- [5] [Setlist API Documentation](#)
- [6] [Spotify API Documentation](#)
- [7] [Kaggle Dataset - Taylor Concert Tour Impact on Attendance](#)
- [8] [What are the Eras](#)
- [9] [Eras Tour Money & Economy](#)
- [10] [Fan Ticket Survey](#)
- [11] [Countries covered in the tour](#)