

Presentation of the data: Indicate how and where you found the data. Explain what you found interesting about the dataset. State in general terms what you identified as relevant questions to be answered with the help of your visualization.

- I found this data on Kaggle, something I found interesting about this data is that I am a big music listener and recently this year spotify released their wrapped and I got to see how I listened to music for this year. Something that I was interested in was how the rest of the world listened to music in 2023 and my project was supposed to be kind of similar to a spotify wrapped for the whole world in a way

Goals and objectives: Provide a list of the specific goals of your project. This should echo the list of action items that you wrote in your proposal.

- My main goal for this project was to analyze listening patterns between all the top songs for this year. In my proposal
 - The question I intend to answer is what constitutes a popular song in 2023? Was it the artist, was it the genre of music, was it the sound of the music (danceability , key, tempo, bpm etc..) and how did all of these come together.

Design: Discuss your main design decisions and explain how they are tied to your goals.

- Design-wise one of the first features I noticed about my dataset was a high number of dimensions and some of the dimensions were similar to each other while others didn't make much sense to compare. I realized that I wouldn't be able to represent all of this with just one single visualization, so I opted to make a dashboard of visualizations.
- Chosen Visualizations
 - **Scatter Plot:** One of the easiest ways to identify relationships between two dimensions is to use a scatter plot and allow filtering changing in size and color and other features increasing intractability. Since I was specifically interested in the most popular features I added the ability to filter by streams and artists

- **Radar Graph:** In this dataset we had 6 features of music each ranked by a percent. I thought the perfect visualization for this was a radar plot as it helped identify which musicality features were popular and which features were not as relevant. This was also interactable
 - **Correlation Heatmap:** This was very important in analyzing trends in the data finding out which features of the dataset had strong correlations or strong relationships between each other. This was also interactable able to filter by the top n% of songs to analyze the top songs within the top songs
 - **Artists Track Bar Chart:** One of the most overlooked ideas in music is that it is hard for smaller artists to break into the scene as many of the top artists dominate the space. This graph helped me visualize who were the top artists and the number of tracks they had in the top 2023 songs. This also allowed me to narrow down the top N artists and the what percent of songs they constituted for
- When doing an earlier project I came across documentation for plotly which is very similar to bokeh since it runs a javascript frontend while all of the backend coding is in python. Something I really liked about the aesthetics of the visualizations and how easy it was to integrate into a web app as the developers also made dash a front end library for web apps. Using dash I was able to add interactable buttons and other features. It was as simple as making my visualization then a new tab on dash and adding my buttons and using button callbacks to update my visualization. Was going to use bokeh but the dashboard web app with dash worked flawlessly with plotly.
 - My first idea was to run a PCA on the dataset to possibly eliminate dimensions, but I wasn't able to understand the library and could have maybe completed this with more time I was also trying to make a force directed graph of artist collaborations for 2023 which would have been really interesting, But I wasn't able to get the edges properly

- **Challenges:** Summarize the difficulties you encountered during the project and how they were addressed.
 - The main difficulty of this project was finalizing which visualizations I wanted to use. I wasted a lot of time making visualizations that never made it to the final cut of my project. Narrowing down my options was one of the hardest parts.
 - I already had a specific library in mind that I wanted to use for this project, but since we didn't use plotly for projects learning the documentation was another difficulty, but overall it was a lot simpler than MATPLOTLIB or Bokeh in my opinion
 - I started the project in Tableau, but realized it doesn't allow interactive visualizations or its support was poor
 - Analyzing the results took some time as some of the results I got contradicted each other and overall some of the results were ambiguous
- **Outcome:** Comment on your results and how they match your original objectives.
 - Summary of my findings
 - Listeners preferred songs with high danceability energy and valence, overall this result was unsurprising as these are all features that make a song "playable" on the radio
 - The top artists/artists with the most number of tracks tended to constitute most of the songs of the top 1,000 meaning that listeners were very loyal to these popular artists, but at the same time there were a lot of lesser known artists in this list
 - As far as correlation there were no strong correlations between musicality elements, BPM, Key vs Streams. What this indicated to me is that although the most popular songs within this top 1000 seemed to all have high energy. The songs in the top 1000 tended to have a lot of variety and be all over the place. Meaning there isn't a particular formula for popular music
 - Overall I'm pretty happy with the results as I was able to narrow down features of music that made it popular

- **Software:** Indicate any external dependencies of your software and give any necessary instructions to run it. If your data is large, explain where it can be downloaded from.
 - I used Dash, Panda and Plotly
 - Run the file **Python ShankeaFinalProject.py** and open the link to the local server the app is running on
- **Optional:** Provide a list of references you may have used to inform your work on this project.

IMAGES OF VISUALIZATIONS INCLUDED BELOW

Scatter Plot

Radar Chart

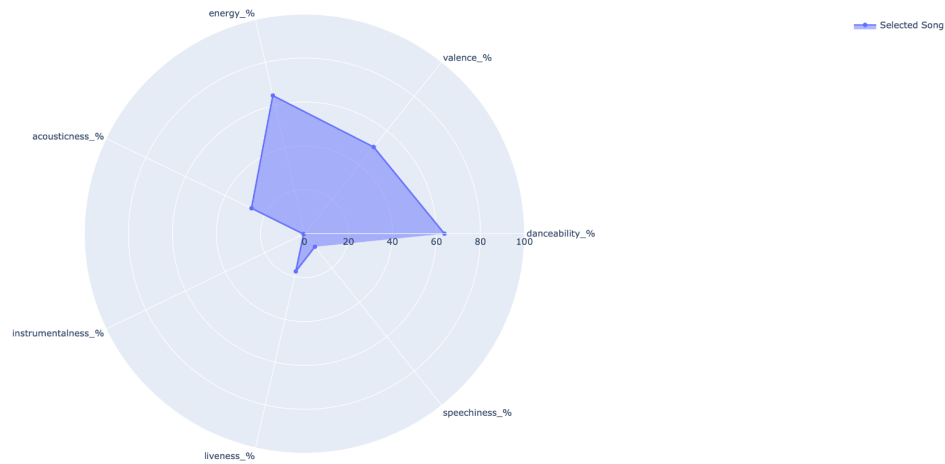
Correlation Heatmap

Top Artists Bar Chart

Average

Top N%10

Average Audio Features for Top 10% of Streams



Scatter Plot

Radar Chart

Correlation Heatmap

Top Artists Bar Chart

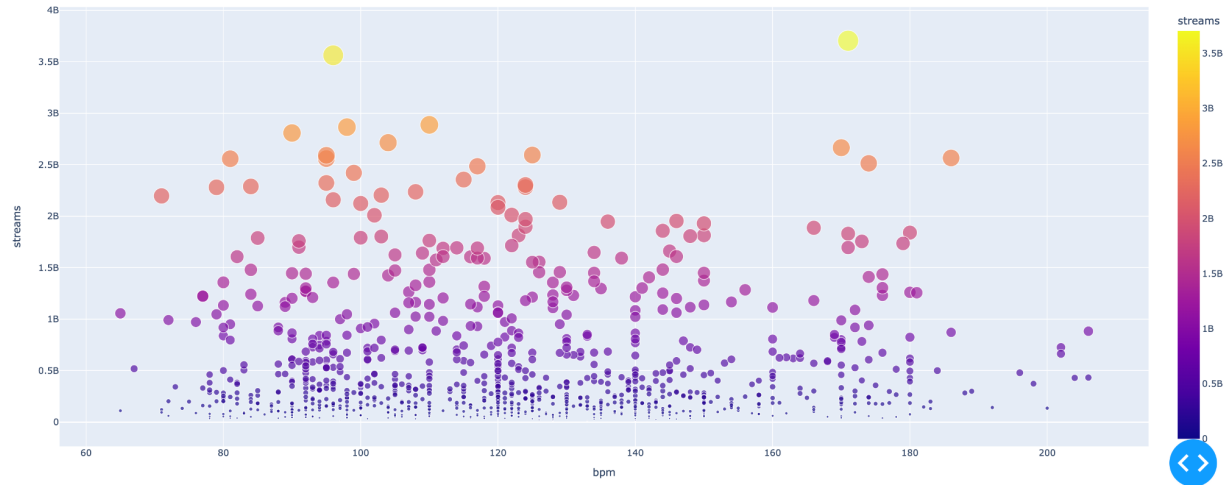
Select artist(s)

Stream Range:

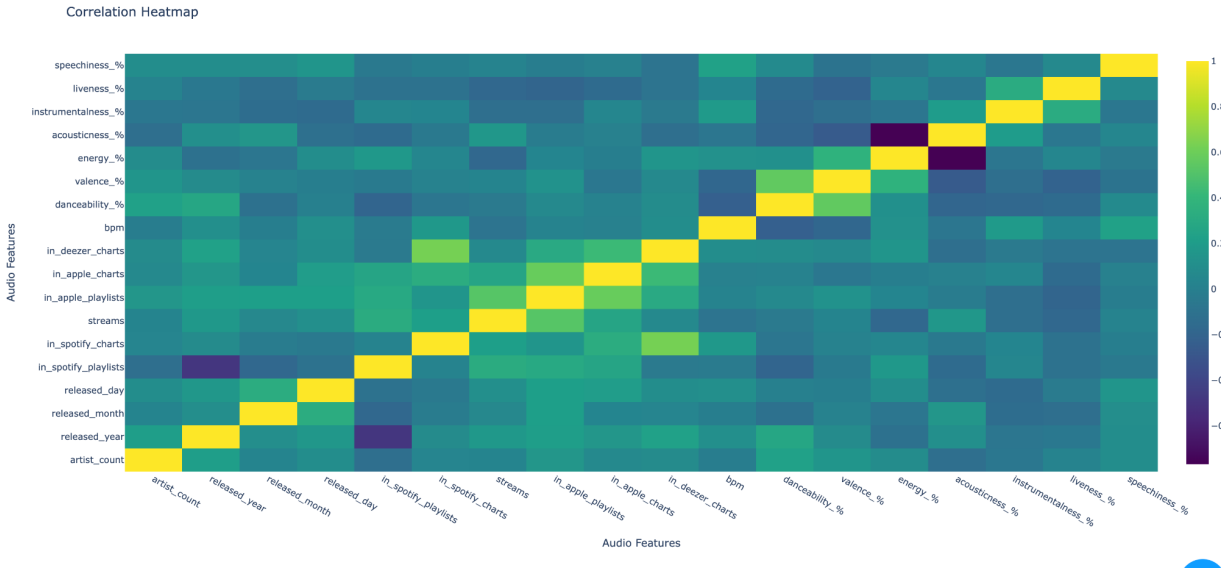
0M100M200M300M400M500M600M700M800M900M1000M1100M1200M1300M1400M1500M1600M1700M1800M1900M2000M2100M2200M2300M2400M2500M2600M2700M2800M2900M3000M3100M3200M3300M3400M3500M3600M3700M

Top N%Top N%

bpmtrack_name track_name track_name



Top N%:10



Top N Artists:30

