What Makes a Song a Hit?

Predicting views and likes of a music track

About the dataset

- 26 variables
- 20717 different tracks
- Interesting variables -

Dataset

Interesting measures
- Acousticness
- Speechiness
- Instrumentalness
- Duration

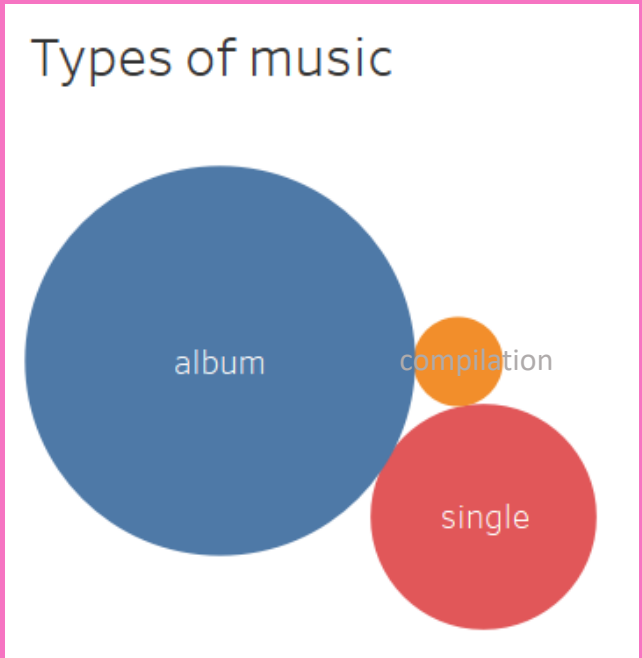
Unlocking Business Value

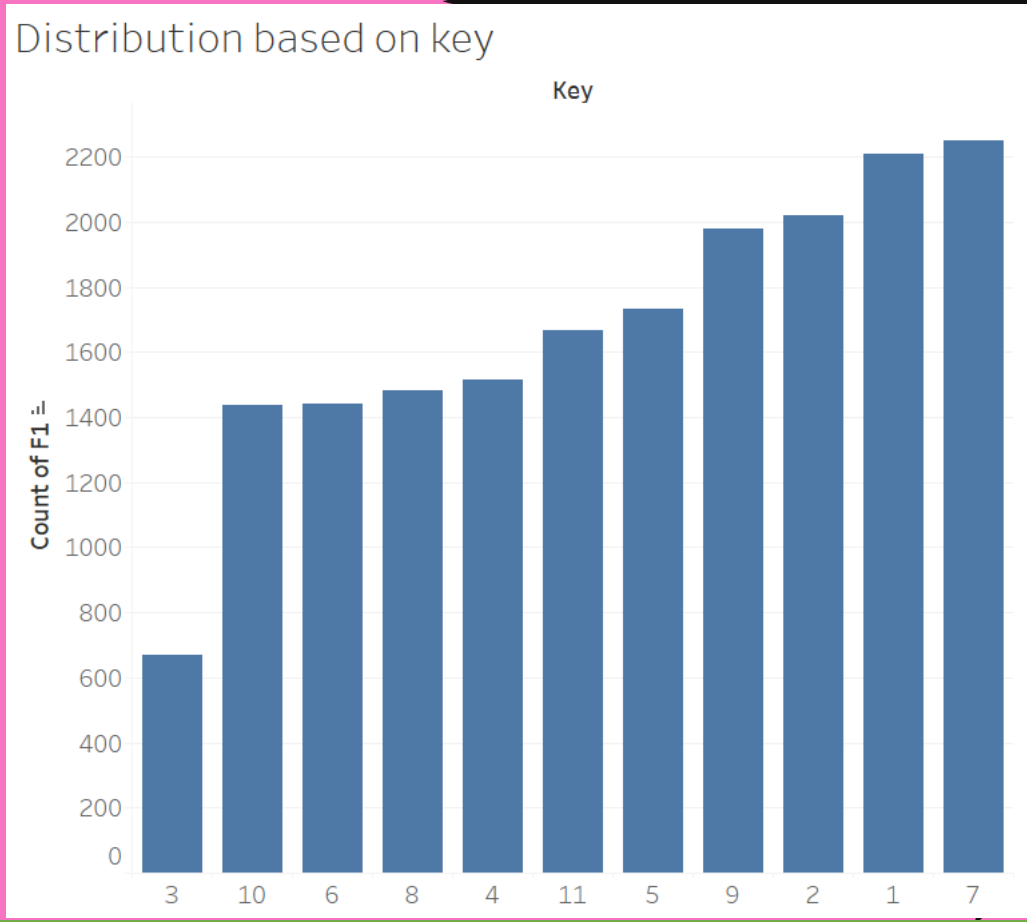
The dataset offers unparalleled business value by providing a comprehensive view of songs and artists through 26 key variables.

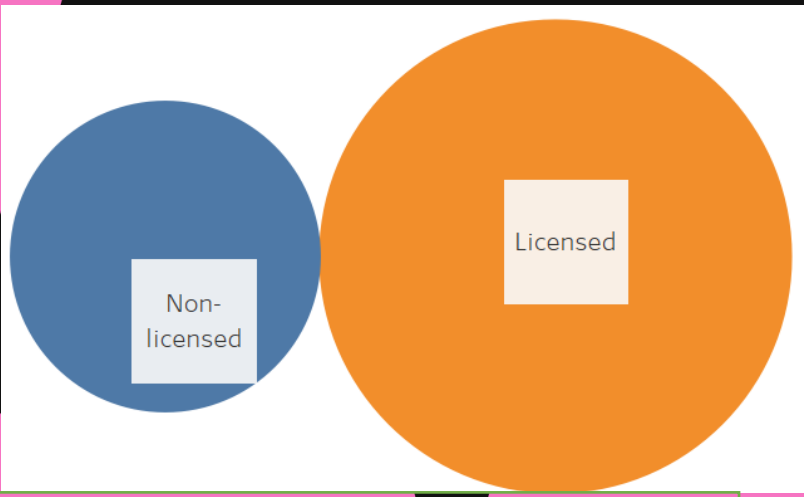
From Spotify streaming statistics to YouTube engagement metrics, it unveils trends in danceability, energy, and other musical features, guiding strategic decisions in marketing, playlist curation, and artist collaborations.

The dataset empowers us to understand audience preferences and identify number of views on videos using predictive analytics.

Overview

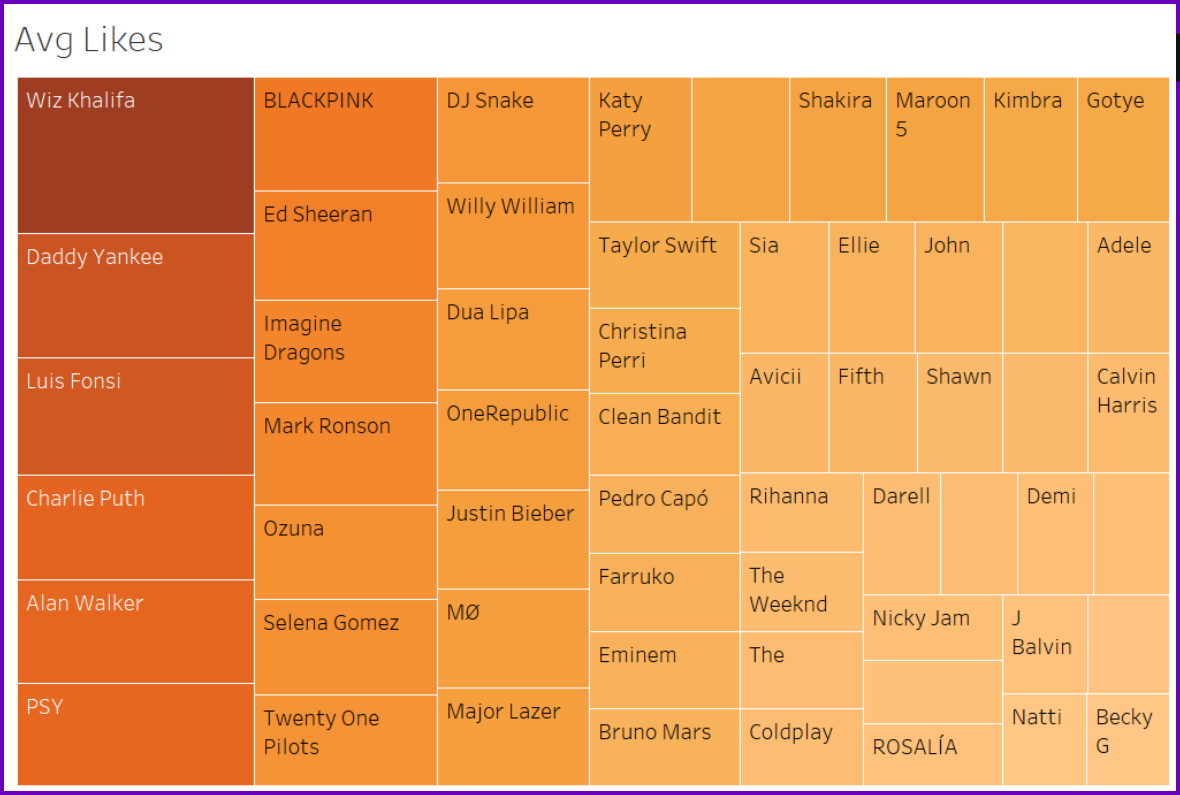


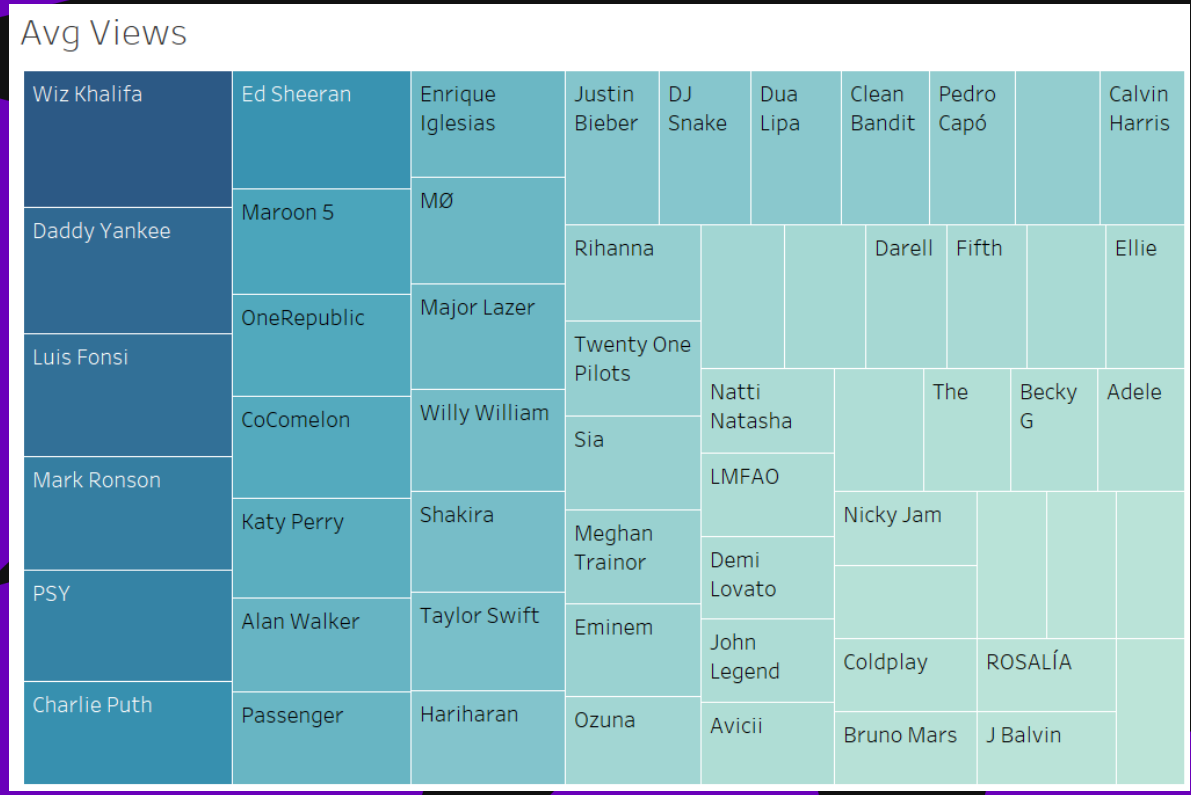




Insight: The charts show the structure of the data, like the number of singles, albums, and compilations. They also highlight whether the tracks are licensed or not and reveal the distribution of musical keys in the songs.

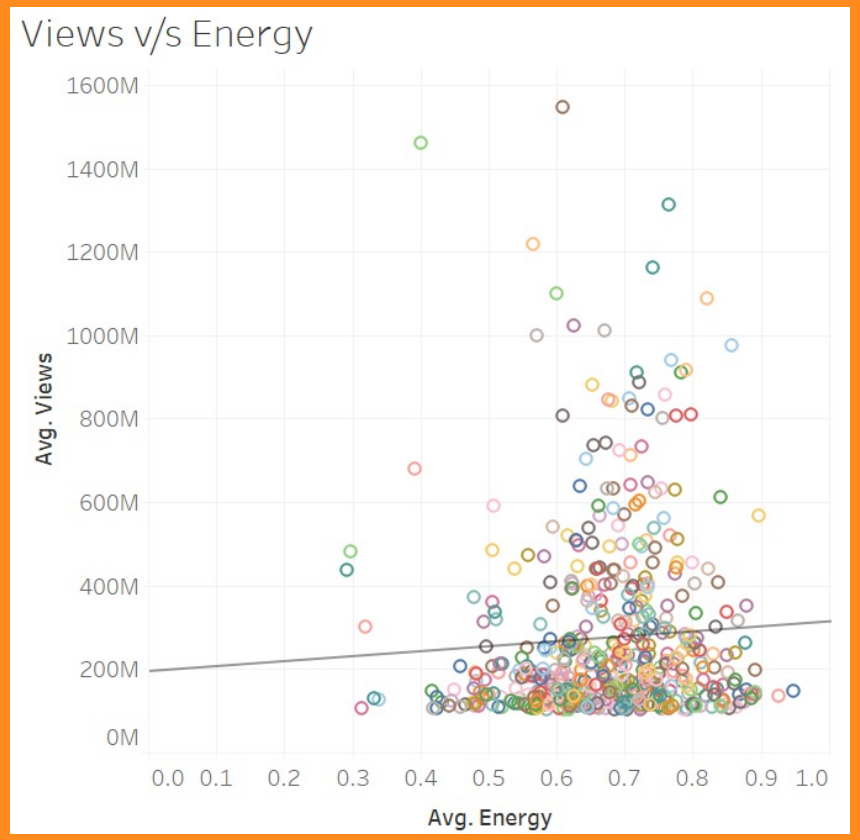
Top Artists

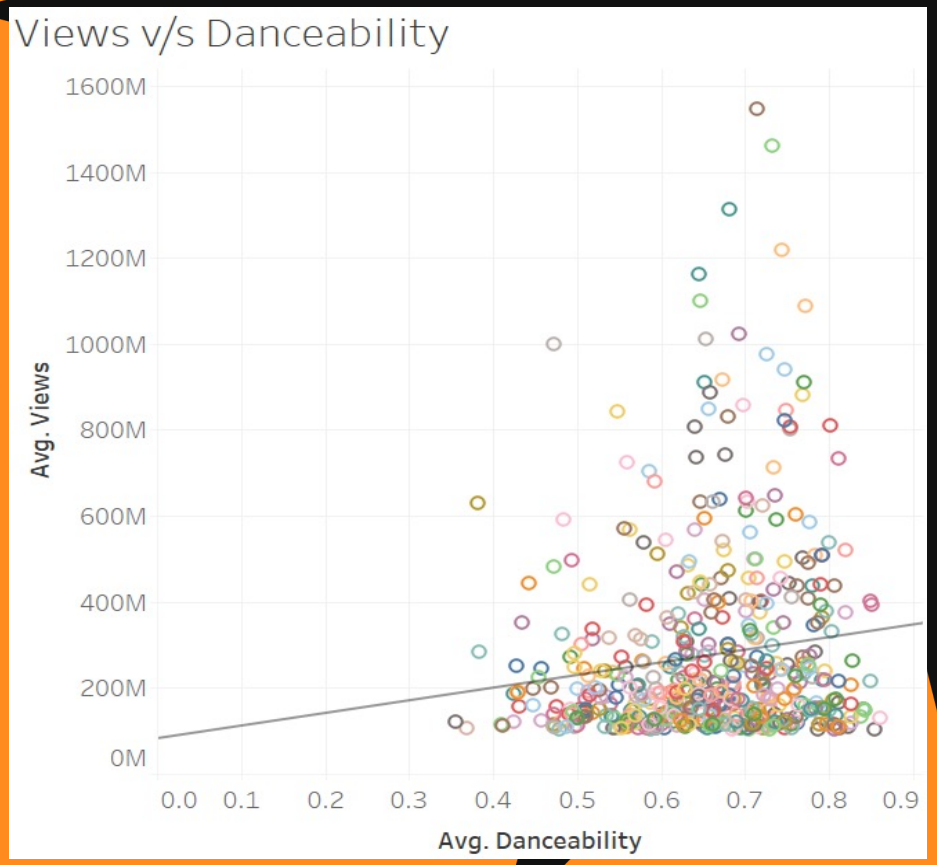




Insight: This insight suggests the potential to develop a dynamic dashboard capable of monitoring and analyzing the average likes and views garnered by individual artists over a specified timeframe.

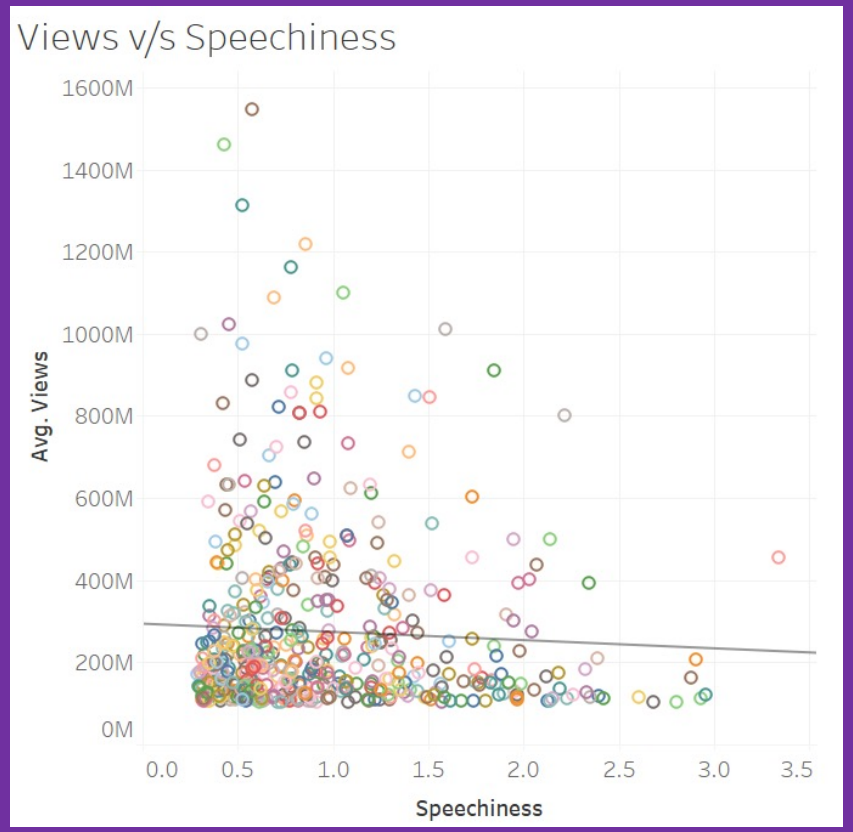
Correlations with Views

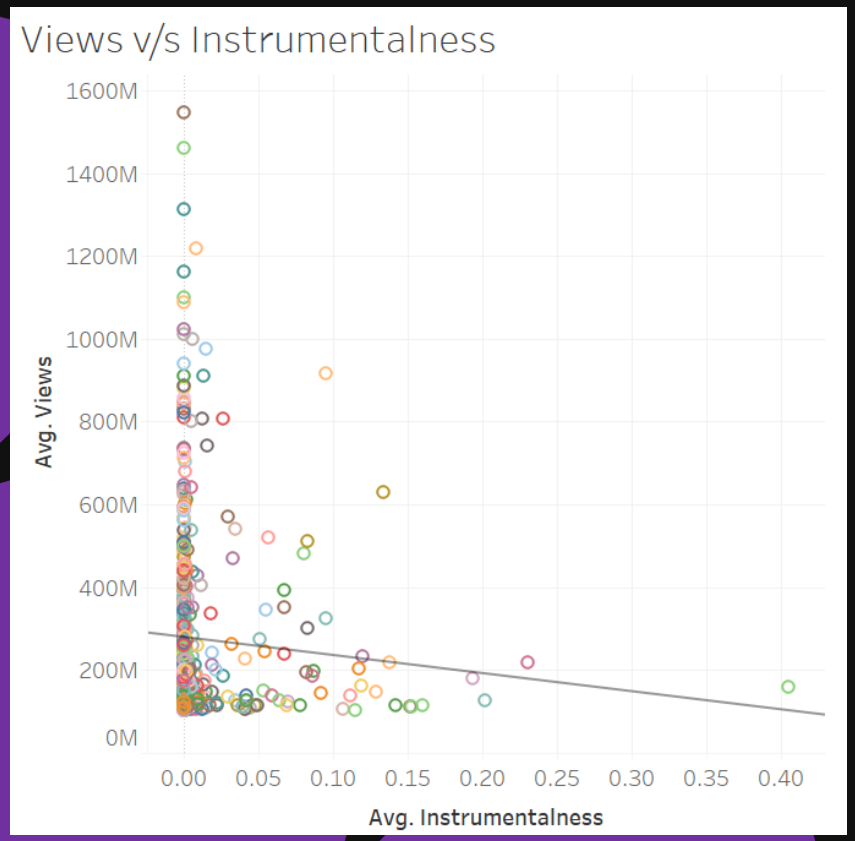




Insight: This insight suggests the potential to develop a dynamic dashboard capable of monitoring and analyzing the average likes and views garnered by individual artists over a specified timeframe.

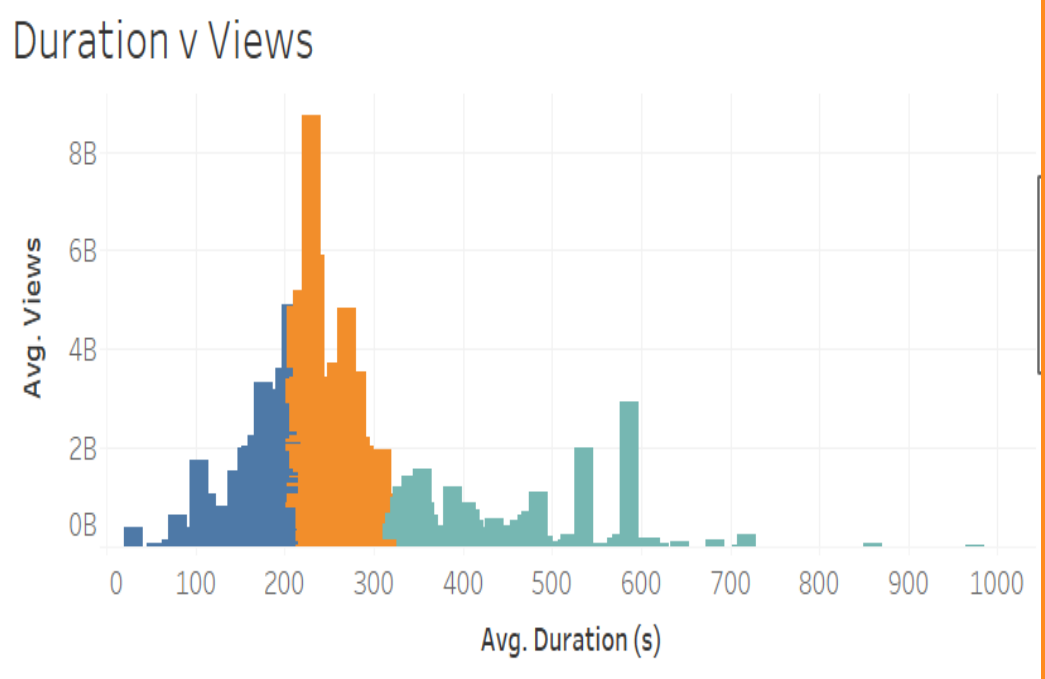
Correlations with Views

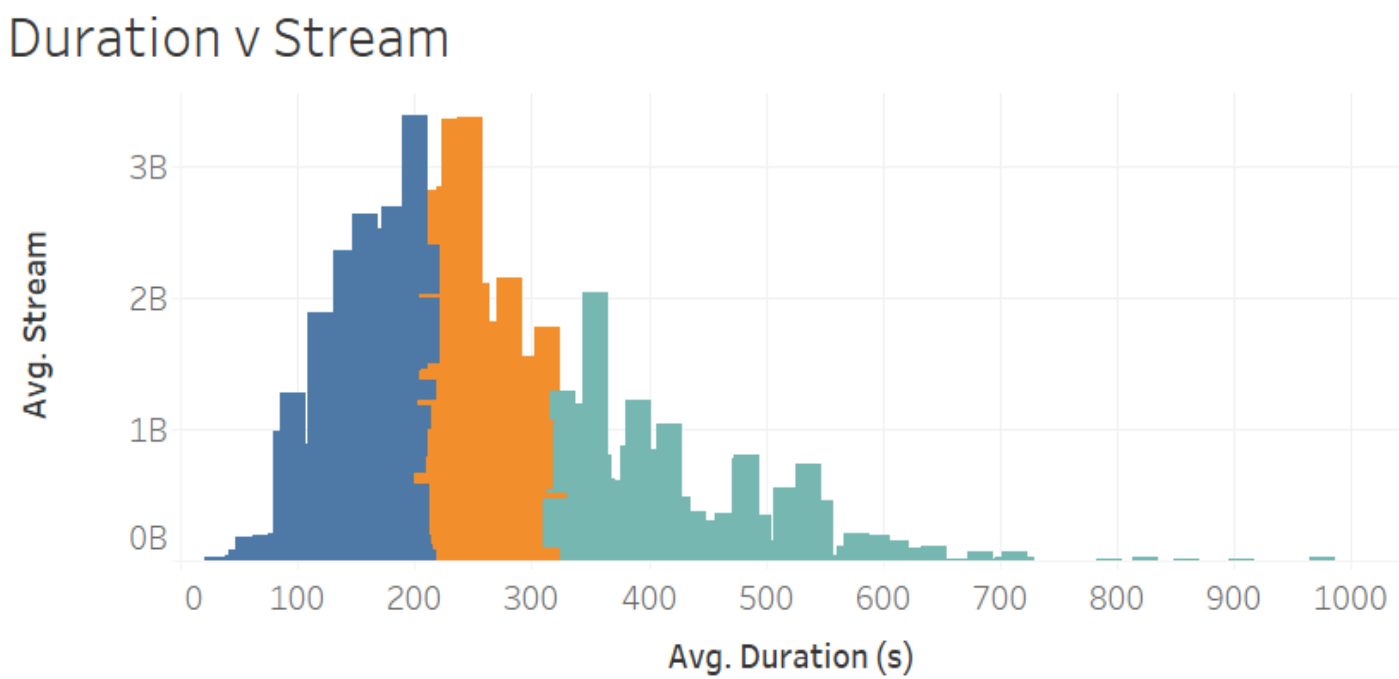




Insight: This insight suggests the potential to develop a dynamic dashboard capable of monitoring and analyzing the average likes and views garnered by individual artists over a specified timeframe.

Correlation of duration with Streams and Views

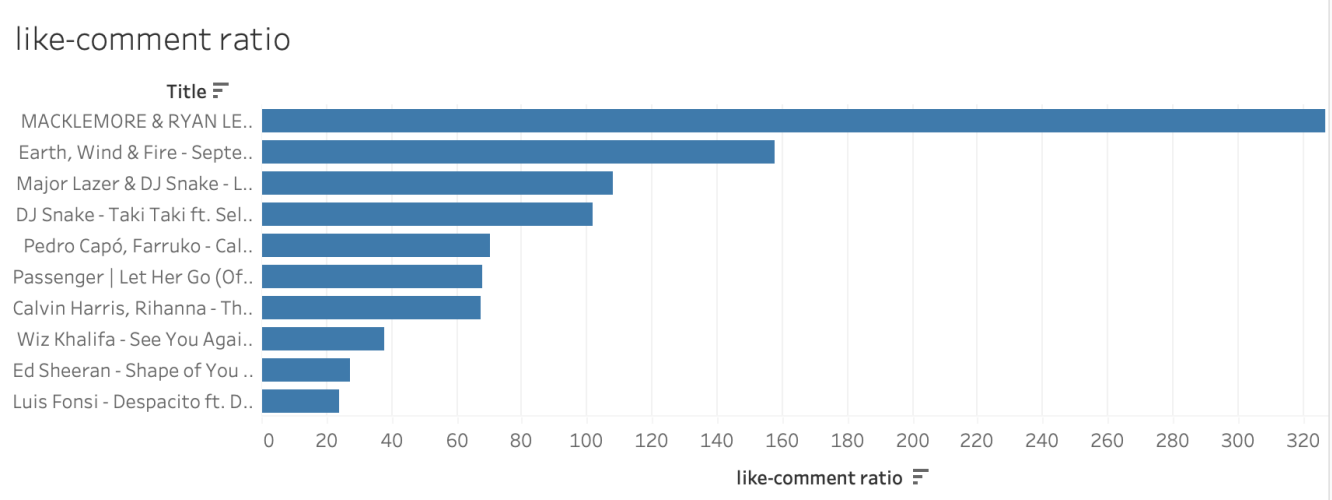




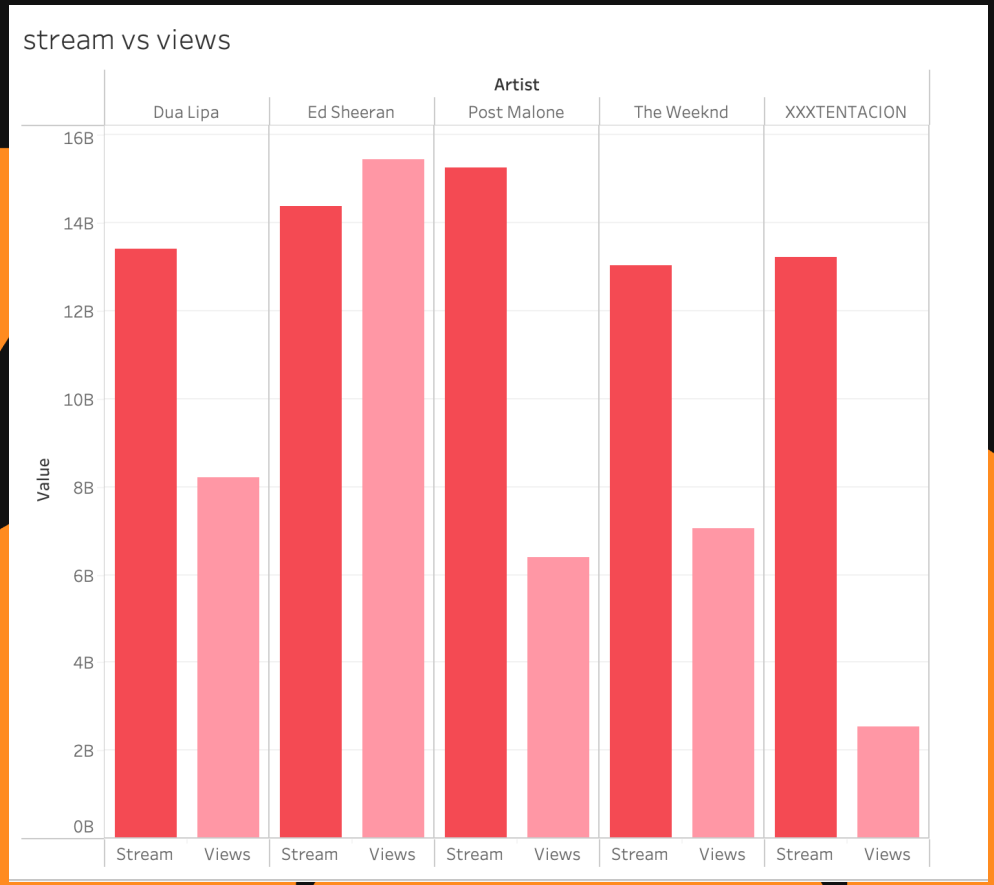
Insight: This insight suggests that there is a sweet spot of a song's duration which will get you higher views. Creators can utilize this to decide on their song's duration.

Top songs and artists

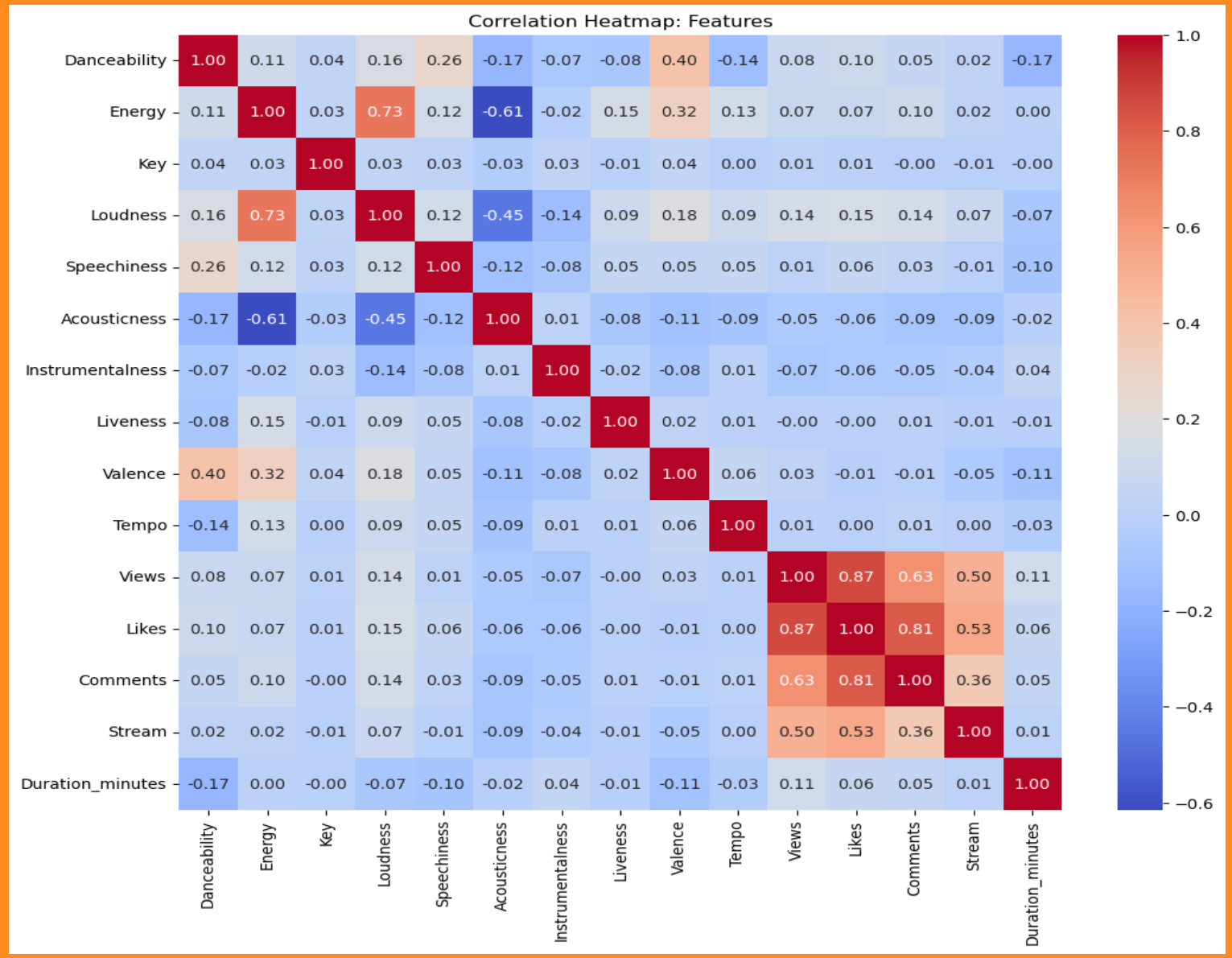
This chart is useful for visually comparing the popularity and reach of these artists across the two platforms. It clearly illustrates which artists have a stronger presence on Spotify versus YouTube..



This bar chart displays the top 10 videos based on the top views. It's a measure of audience engagement with the content on YouTube.



EXPLORATORY DATA ANALYSIS



Variable	VIF
Danceability	16.5931213
Energy	20.5116252
Valence	8.52089006
Loudness	8.83545171
Acousticness	2.79489575
Instrumentalness	1.08710019
Tempo	16.0642783
Duration_minutes	13.1819492

VARIABLE SELECTION TECHNIQUES USED

Using variable selection methods to streamline the model, making them more accurate and interpretable. Focusing on the most relevant variables for better predictability.

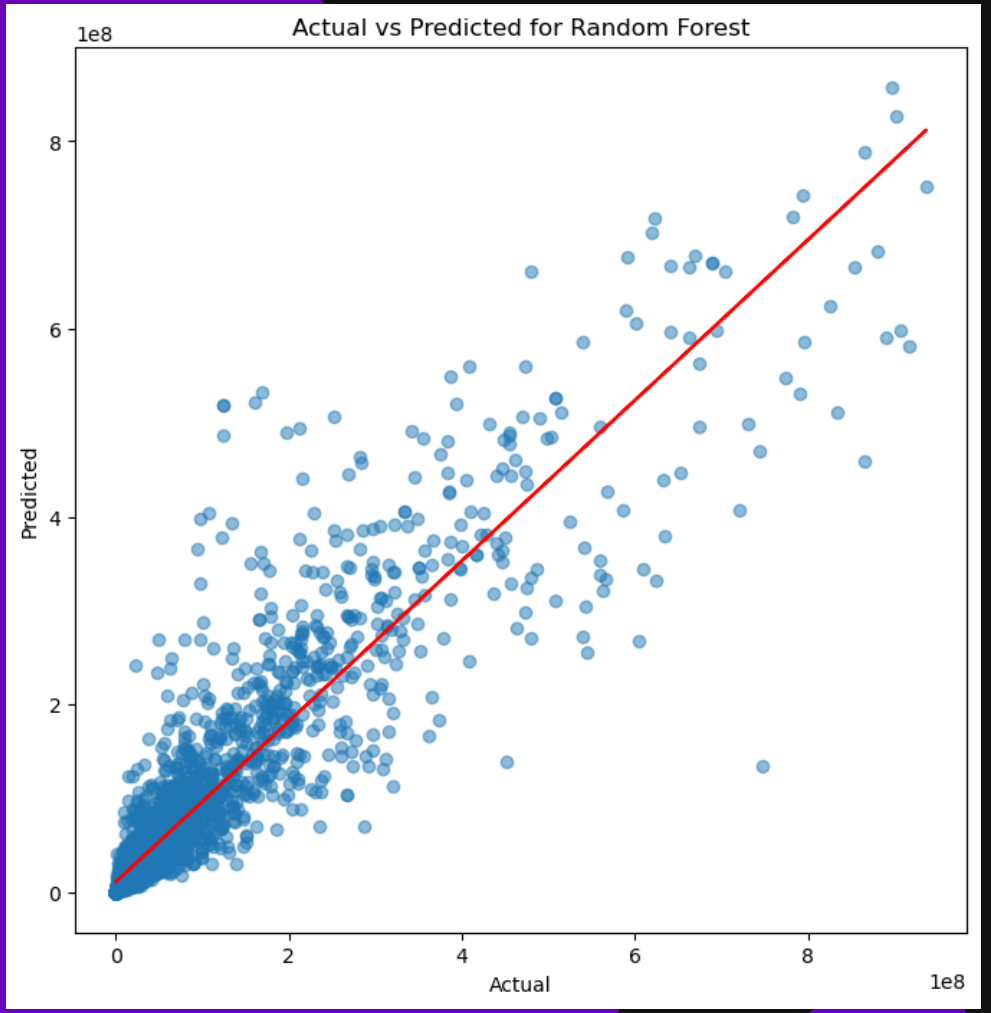
ALGORITHM	SELECTED VARIABLE
Lasso Variable Selection	['Key', 'Loudness', 'Speechiness', 'Acousticness', 'Instrumentalness', 'Liveness', 'Valence', 'Likes', 'Comments', 'Stream', 'Duration_minutes']
Forward Selection	'Valence', 'Likes', 'Comments', 'Duration_minutes'
Backward Selection	'Likes', 'Comments', 'Duration_minutes'
Stepwise Selection	'Speechiness', 'Valence', 'Likes', 'Comments', 'Duration_minutes'

MODEL EVALUATION

We selected Lasso regression for variable selection resulted in superior model accuracy compared to forward or stepwise selection methods.

The Random Forest model exhibited the highest accuracy with the lowest RMSE and the highest R-squared value of among the six utilized modelling techniques.

MODEL	RMSE	R2
Random Forest	48811929.79	0.838
Gradient Boosting	49873677.75	0.831
Linear Regression	58461739.01	0.768
Ridge Regression	58461729.15	0.768
Lasso Regression	58461738.99	0.768
Decision Tree	67075839.66	0.694



INSIGHTS WRAPPED

For producers

This wealth of information can potentially foster data-driven decision-making for record labels and studios. It can be a good way to identify artists, and new music, gems yet to be found.

For musicians

These insights can enhance targeting strategies of musicians, ultimately driving success in an ever-evolving and competitive music landscape.

For audience

Through these visualizations, we can discern that music exhibits diverse characteristics influencing its popularity. Given the myriad tastes people have in music, these visualizations enable individuals to explore and discover music that resonates with them on a profound level.

Thank You!