Spotify Prediction Tool

AnalySis

AGENDA

01 BACKGROUND

PROBLEM, SOLUTION, & DATA

02 METHODOLOGY & FINDINGS

DATASET
EXPLORATORY ANALYSIS
LINEAR REGRESSION
DEMO

03 CONCLUSION

HUMAN-CENTERED APPROACH
INSIGHTS & FUTURE RESEARCH
SOURCES
QUESTIONS

01

Background

The problem?

Considering 10% of artists account for 98% of streams between 2014 and 2020, what tools can artists use to protect themselves among such tough competition?

Our Solution

A Linear Regression model that can predict the number of streams/success of a song.

The Dataset

The dataset is derived from the Spotify API, processed through Gigasheet, to analyze music trends and user preferences.

02

Methodology & Findings

Dataset

- Spotify data was collected from Gigasheet under a restricted license.
- Processing was minimal. Only rows with missing values were dropped.

```
# Data after dropping null values
>Spotify_data_df.head()

      Stream  Danceability  Energy  Loudness
Speechiness  Acousticness
0  1040234854         0.818   0.705    -6.679
0.1770       0.008360
1   310083733         0.676   0.703    -5.815
0.0302       0.086900
2    63063467         0.695   0.923    -3.930
0.0522       0.042500
3   434663559         0.689   0.739    -5.810
0.0260       0.000015
4   617259738         0.663   0.694    -8.627
0.1710       0.025300
```

Exploratory Analysis

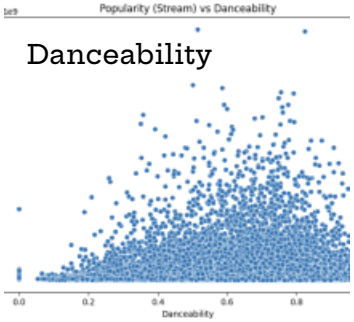
1) Assess the correlation of attributes compared to number of streams.
    - Visualize as Scatter Plots

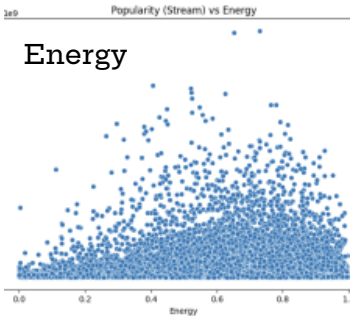
1) Identify the most positively correlated attribute.
    - Correlation Matrix

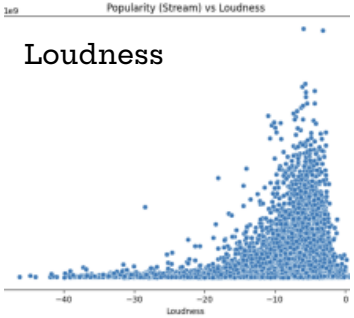
```
# Compute the correlation matrix for the selected attributes
correlation_matrix = df.corr()

# Identify the attribute with the highest absolute correlation to popularity
most_correlated_attribute = popularity_corr.abs().idxmax()
highest_correlation = popularity_corr[most_correlated_attribute]
```

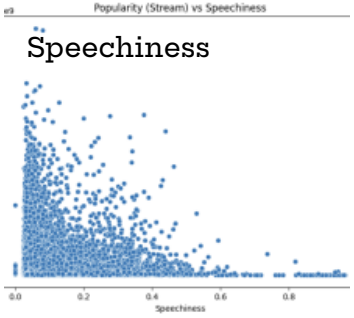
Scatter Plots of Correlations

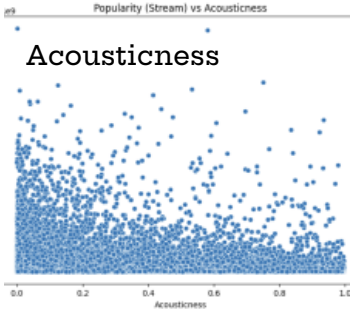
- “Loudness” has the highest positive correlation.
- “Danceability” and “Energy” follow “loudness.”
- All other attributes are negatively correlated or not correlated to streams.

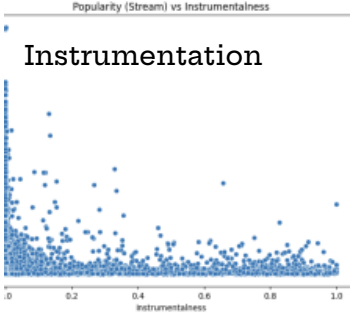


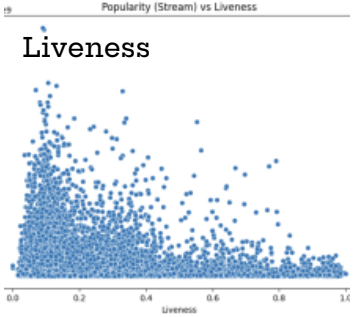


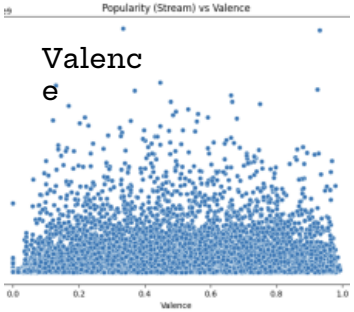


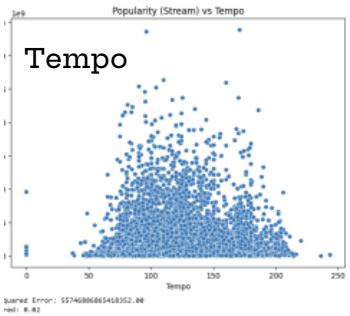






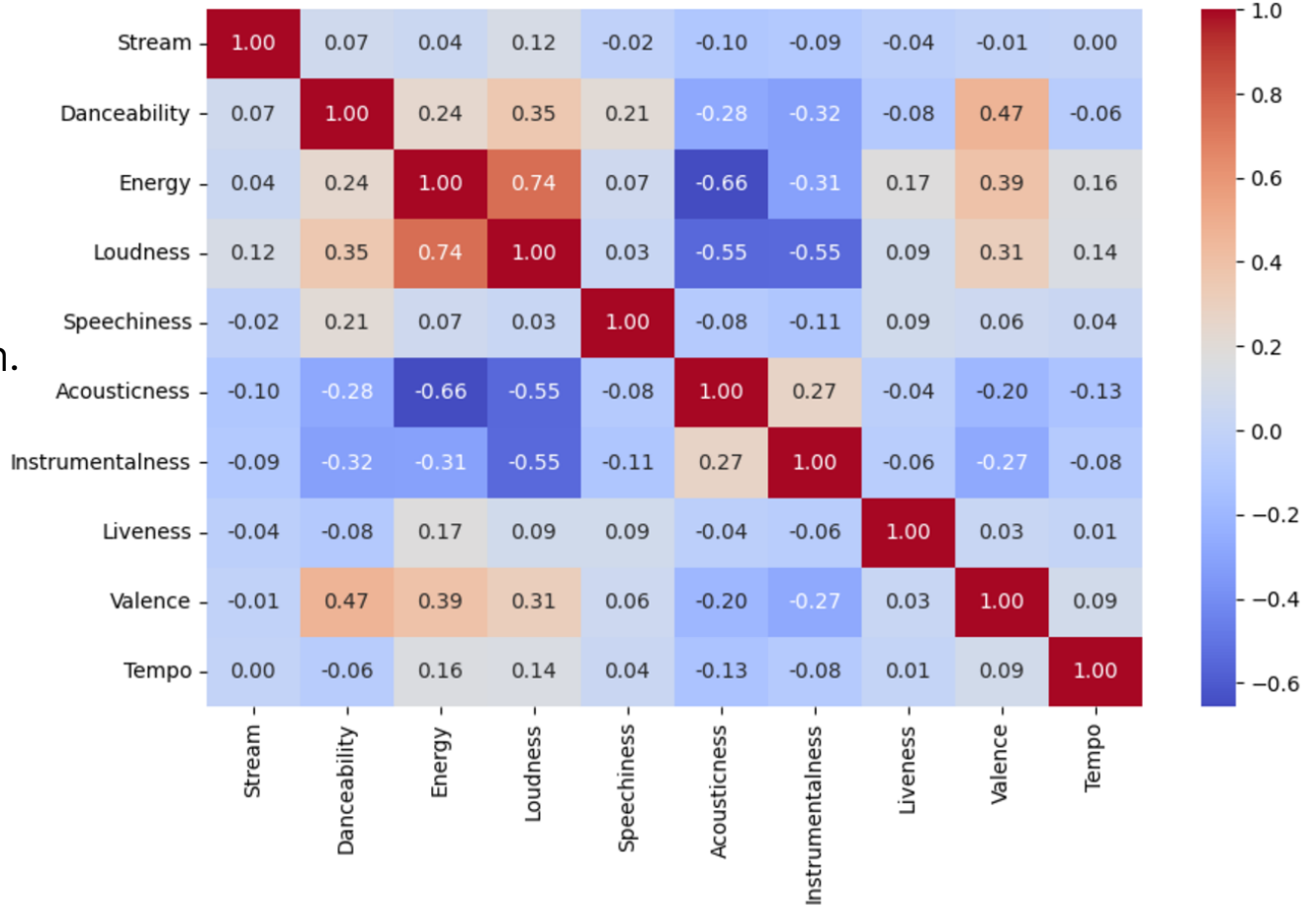






Correlation Matrix

- “Loudness” is the attribute that most correlates with stream.
- It has a correlation coefficient of 0.12.

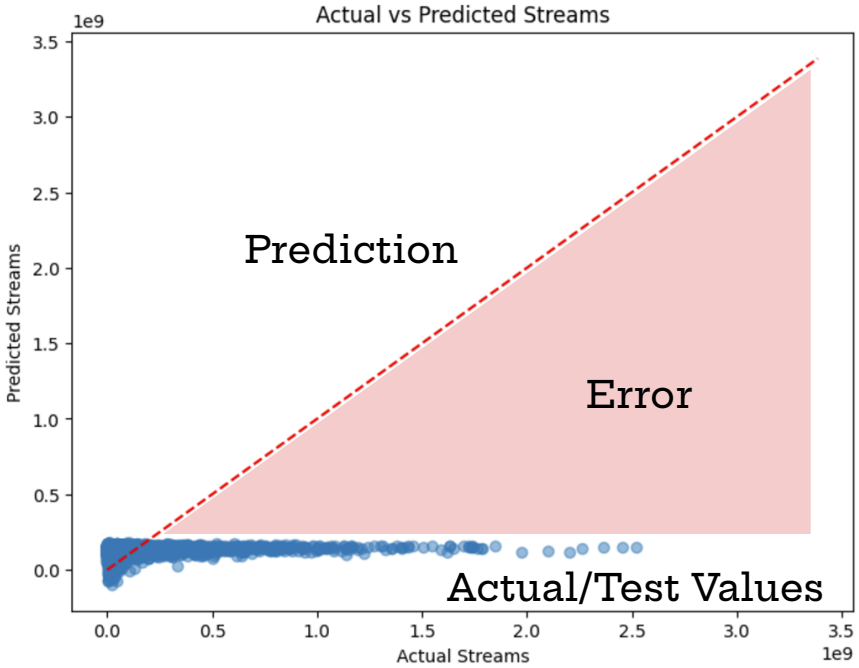


Linear Regression

3) Use most positively correlated attribute, “loudness,” to predict song streams.
    - Linear Regression Model

4) Assess bias of model.
    - Residuals
    - Means squared Error
    - R-squared

```
# Initialize and train the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions on the test set
y_pred = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
```



Our model grossly underestimates actual stream values.

Residuals = .9381 (1010) = Error = (predicted - test)**

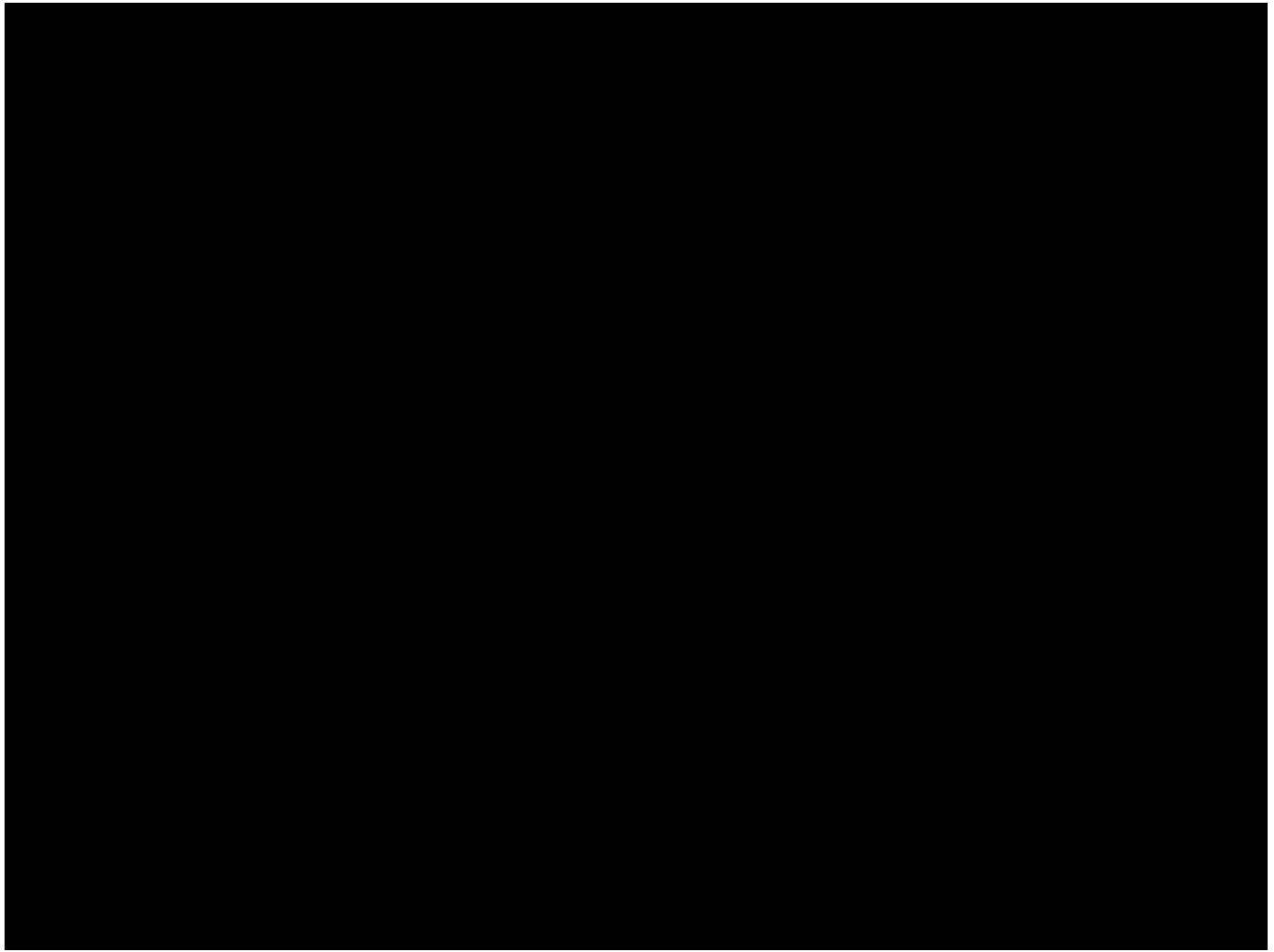
Our model's prediction is very distant from actual stream values.

Mean Squared Error = 5.575 (1016)**

Our model explains 20% of the variation in streams based on "loudness."

R-squared = 0.02

DEMO



03

Conclusion

Human Centered Approach

- The purpose of our project is to create a tool for musicians by analyzing music success
- Analyzing Bias
  - Our model has limited bias because it is simple.
- The licensing and usage of our data is ethical and socially responsible

LIMITATIONS & FUTURE RESEARCH

- Major limitation: low accuracy
  - High error + grossly underestimating test values
- In Future: develop a model with more features like “danceability” and “energy.”
  - Improve predictions
- Explore K-Nearest Neighbor (KNN) as an alternative model.
  - Compare results to linear regression model
- Perform a LIME analysis on new models
  - Understand the individual influence of specific features

SOURCES

Pohl, N. (2022, August 16). Analysis of the competition and Markets Authority Report on music streaming. The Musicians' Union. https://musiciansunion.org.uk/news/analysis-of-the-competition-and-markets-authority-report-on-music-streaming#:~:text=It%20goes%20on%20to%20say,%5D%20cannot%20sustain%20a%20living.%22

Spotify Terms of Use: https://www.spotify.com/us/legal/end-user-agreement/; License: https://developer.spotify.com/terms#section-iii-licenses-and-permissions

Thank You

Questions?