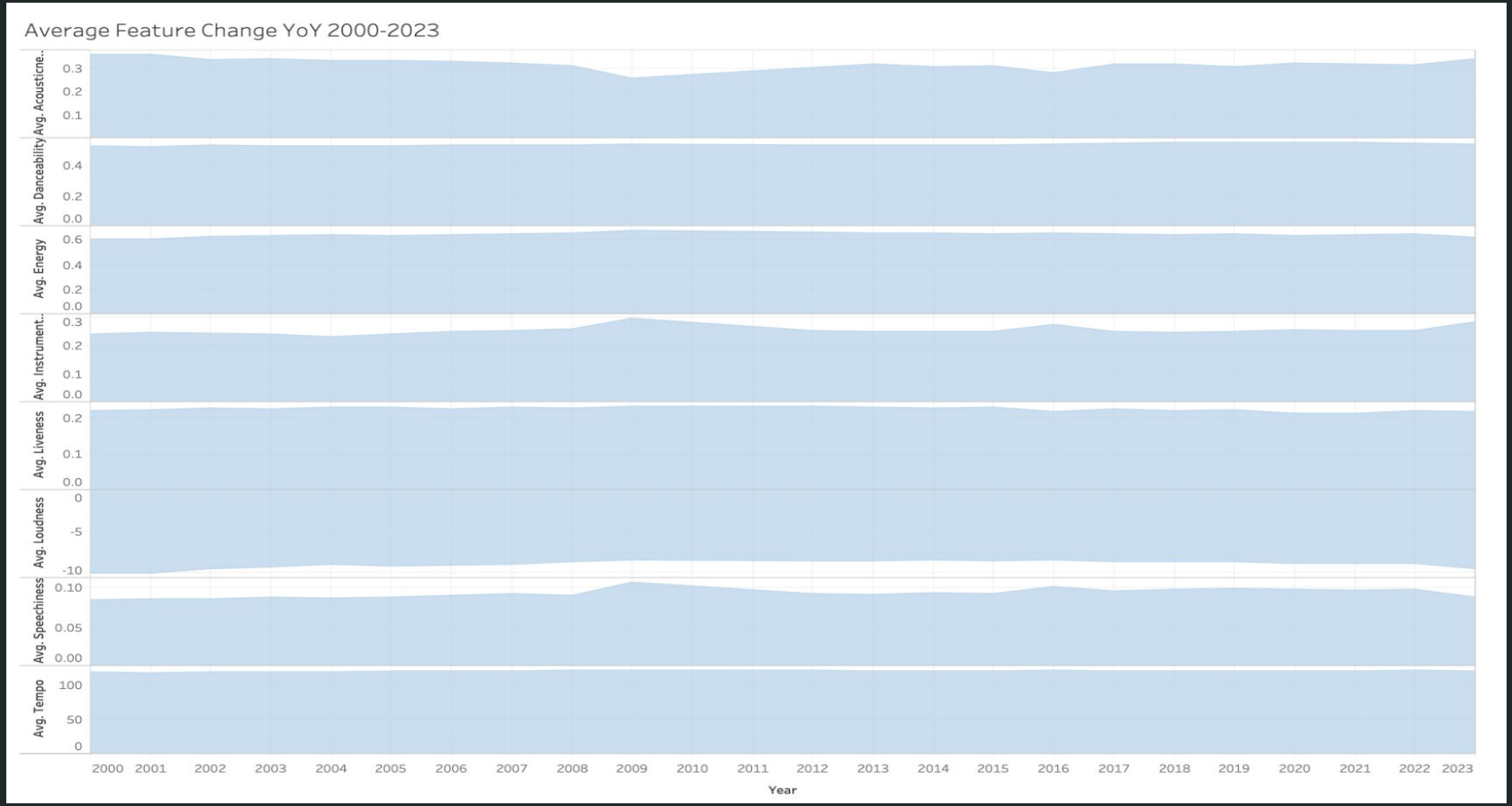
Spotify Music 1921-2023 - How Has Music Changed Over Time?

The Unknown

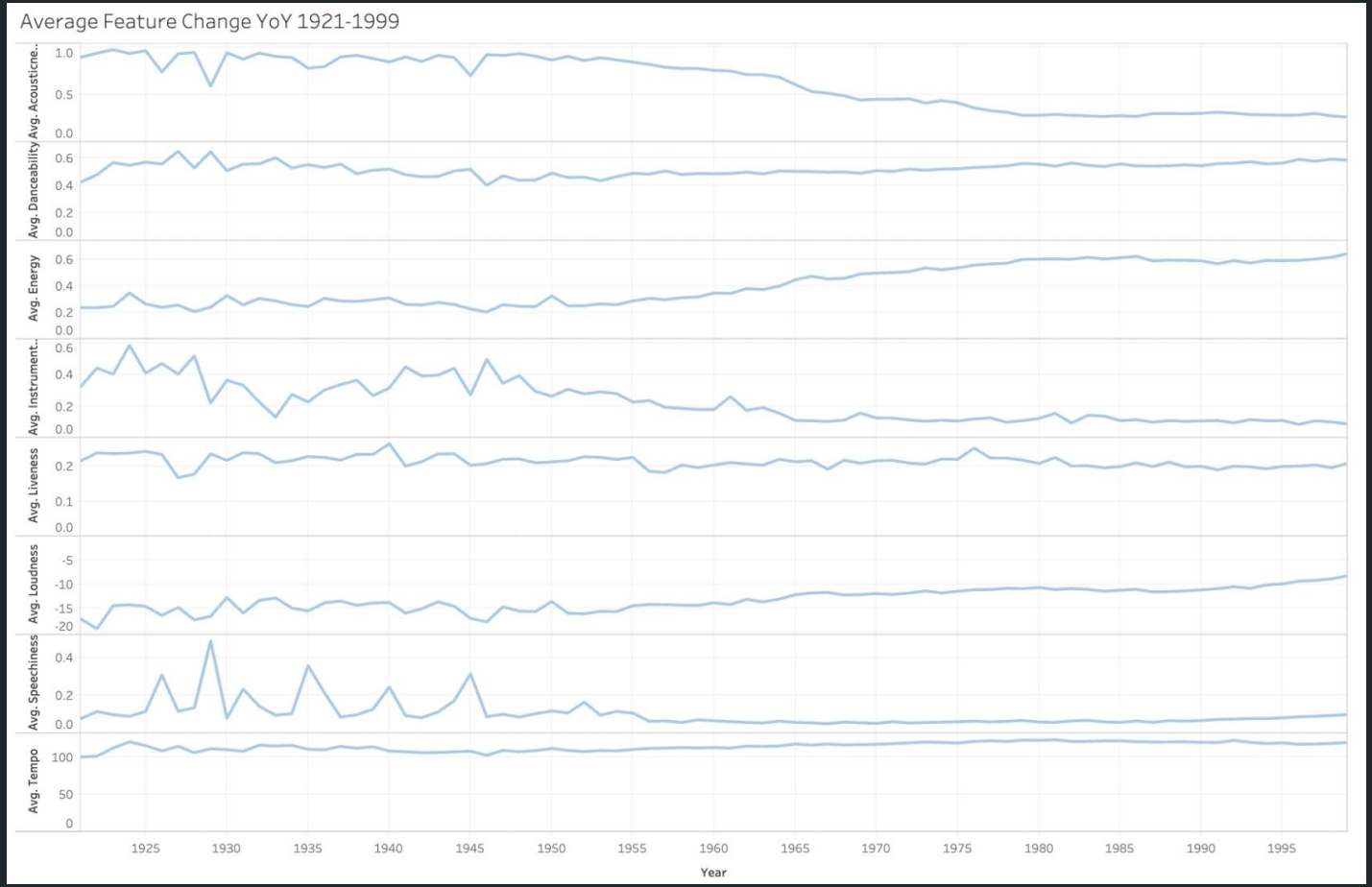
As society has evolved over time, so has its music preferences. But this is known.

How has music evolved over time?

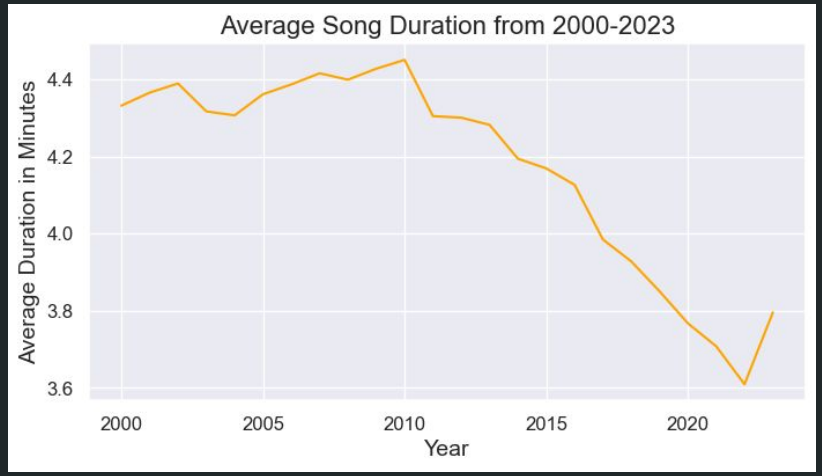
Tale of Music Evolution

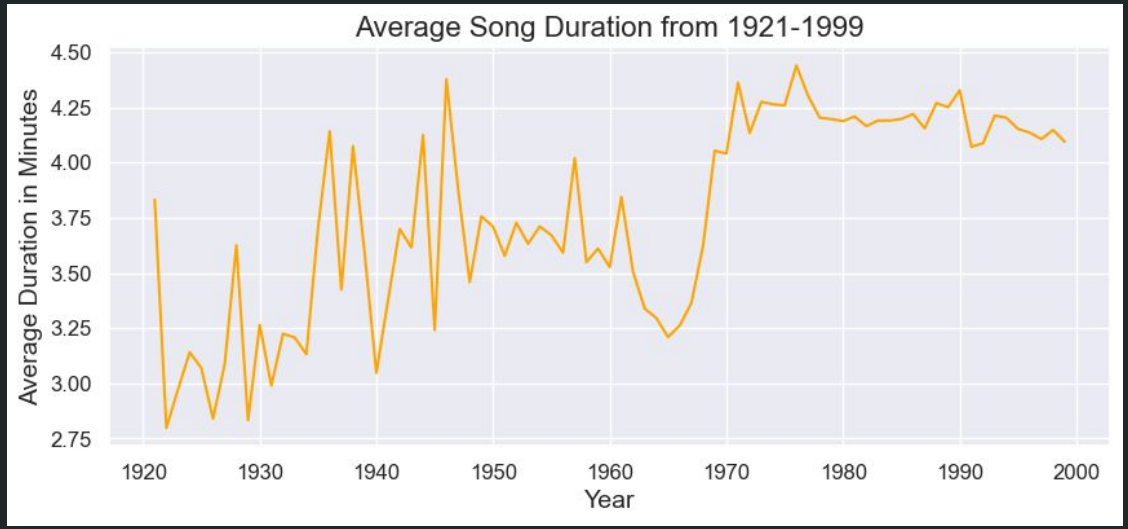


Tale of Music Evolution - Continued



Tale of Music Evolution - Continued





Who Might Care?

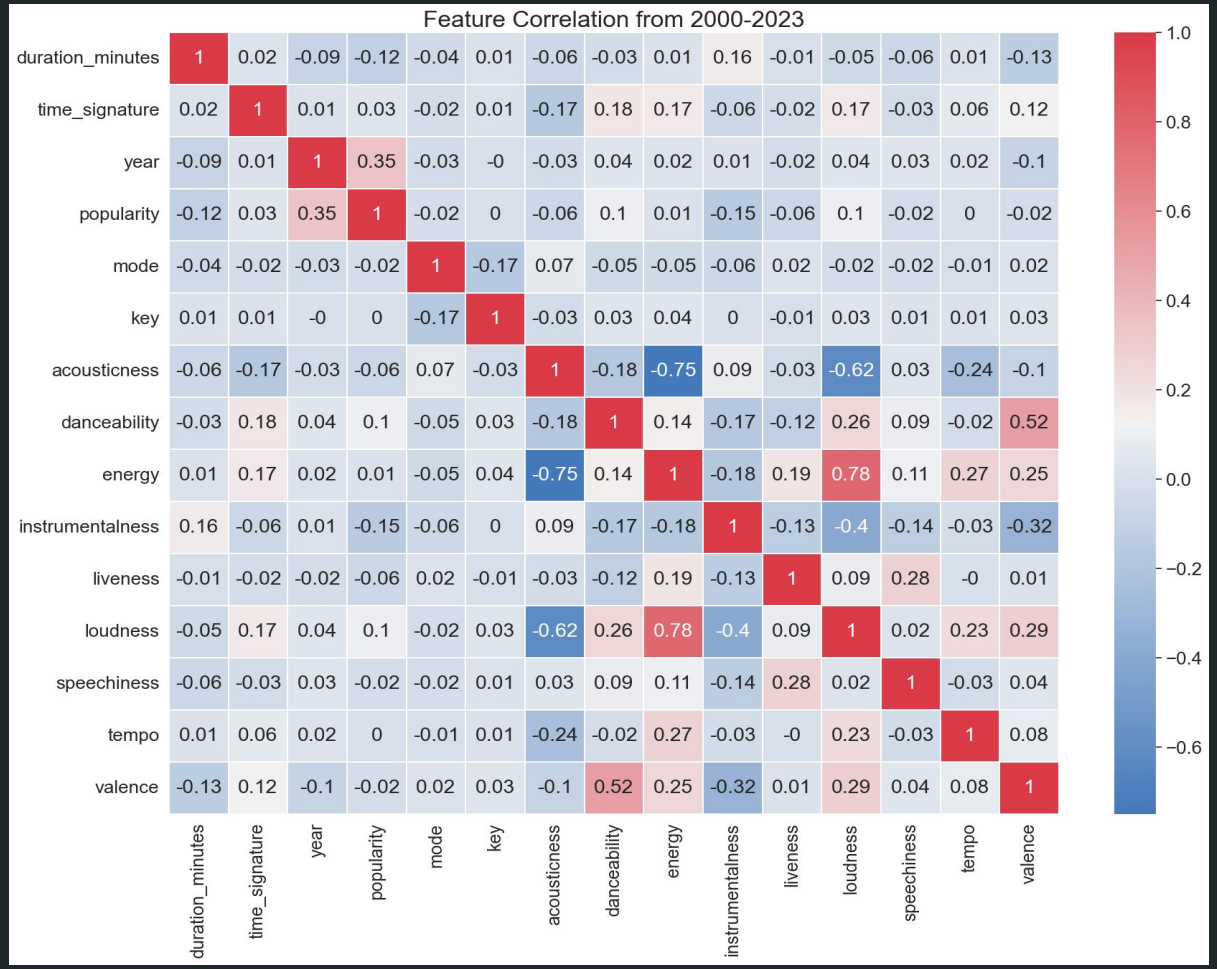
- Concert venues
- Artists
- Producers
- Merchandisers
- Fans
- Music connoisseurs

Getting to Know the Data

- Spans 1921-2023
- via Spotify’s API
- 1.2 Millions rows between two datasets
  - Dataset 1 spans 2000-2023
  - Dataset 2 spans 1921-1999
- 19 fields/features
- csv format

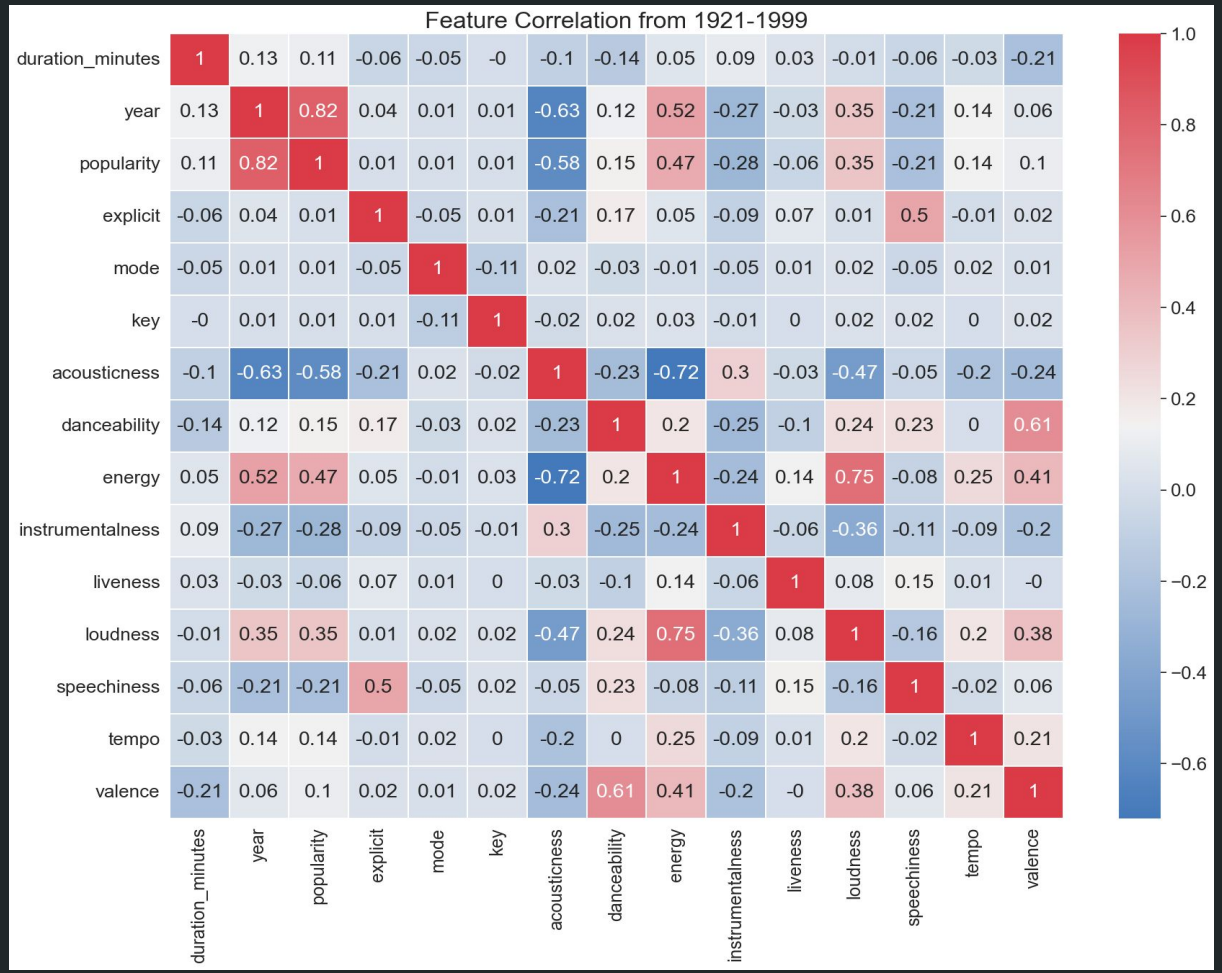
Analyzing Feature Correlations - Dataset 1

What stands out?



Analyzing Feature Correlations - Dataset 2

What stands out?



Regression Modeling - Supervised Machine Learning

- Of the 19 features, two were dropped
- 80% train/20% test split was applied to each model
- Target feature = popularity
- Tool = Python scikit learn

Model Metrics Leveraged

- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error (MAPE)
- R Squared

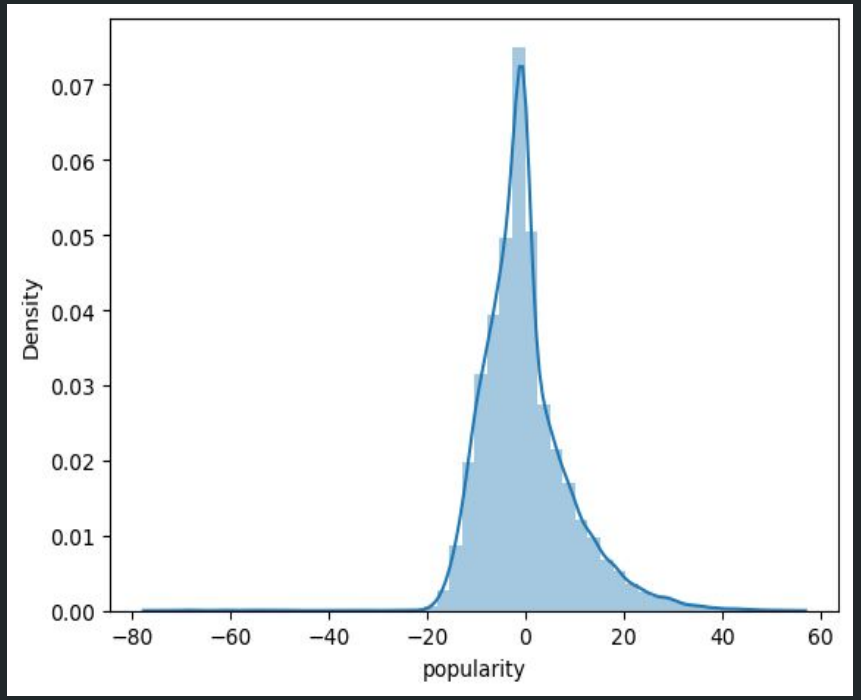
Model Comparisons - Dataset 1

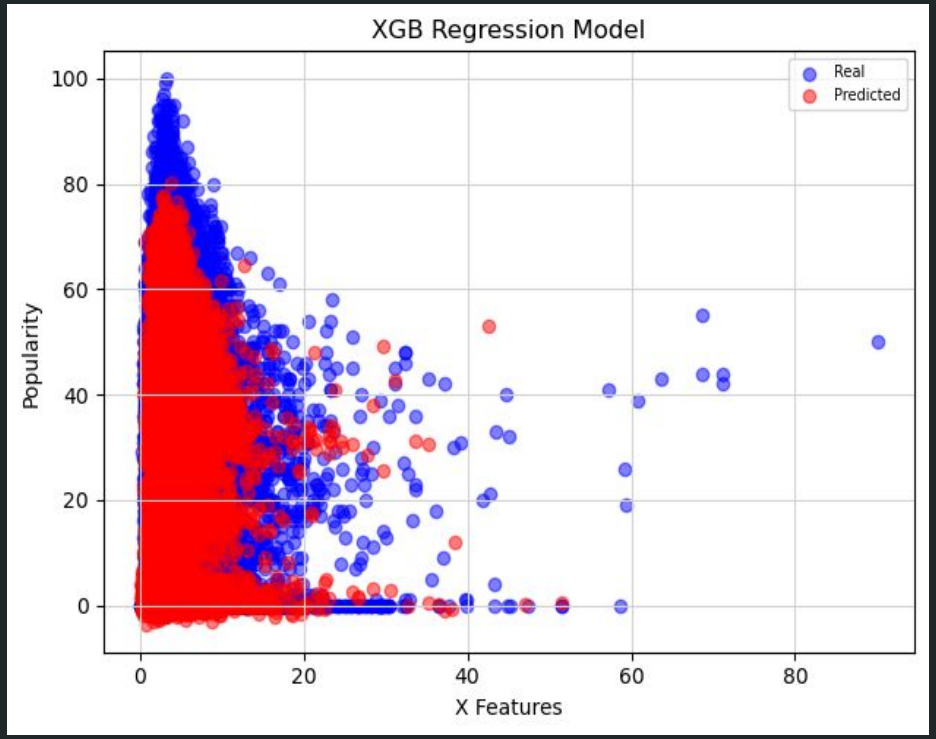
Model	RMSE	MAPE	R Squared
Random Forest	N/A	N/A	N/A
kNN	14.26	8958822157789828	19.37%
Decision Tree	N/A	N/A	N/A
Linear	14.45	9375453561387142	17.15%
Lasso	14.45	9375453561387940	17.15%
XGB	13.15	7131983477509248	31.38%

Model Comparisons - Dataset 2

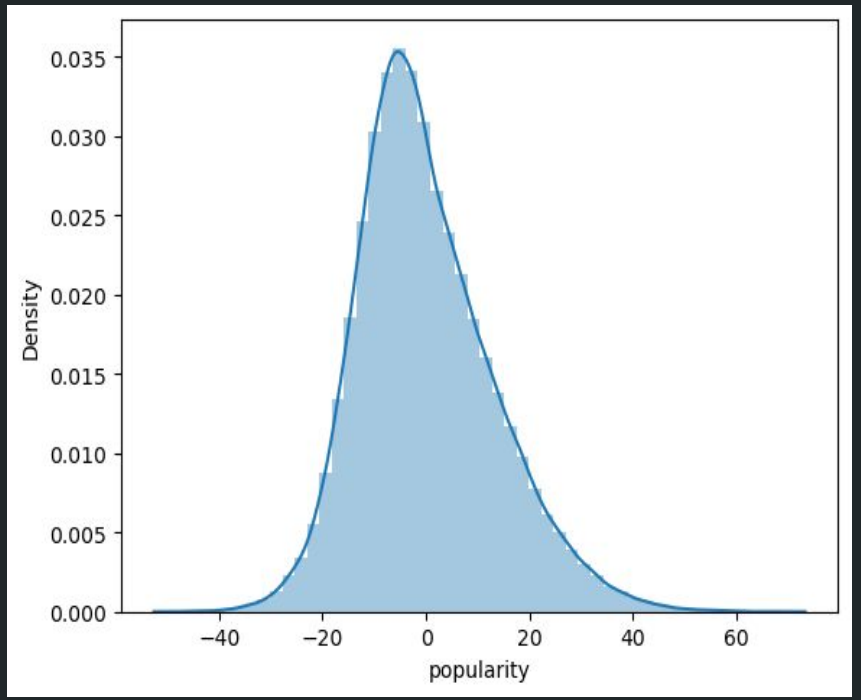
Model	RMSE	MAPE	R Squared
Random Forest	N/A	N/A	N/A
kNN	9.35	18775129153256 39.75	81.29%
Decision Tree	9.63	2025441916744209	80.15%
Linear	10.09	4887416424252879	78.20%
Lasso	10.09	48867280428335 69	78.20%
XGB	9.01	14442484095593 38.25	82.60%

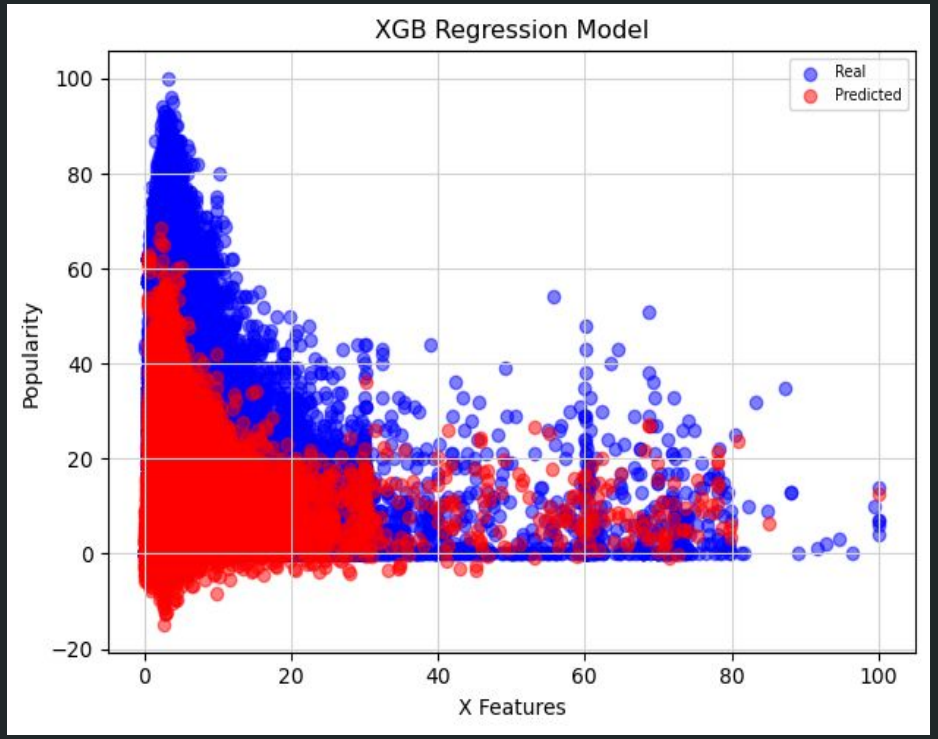
XGB Regression Model Performance - Dataset 1





XGB Regression Model Performance - Dataset 2





Model Improvements

- More computing/server power to tune applicable models
- If possible, gather more data tied to the years 1921-1999 (Dataset 2) so it better mirrors the data scale of Dataset 1

Takeaways

- Out of the six supervised regression models, the XGB regression models performed the best.
    - Its predictive power was quite strong
    - 80% train/20% test split was used
- Out of 19 features, 17 were leveraged for modeling