Musical Themes Across Decades

Team 26 - Mitali Bidkar, Sumana Chilakamarri, Rhea Chitanand, Sarah Friedrichs, Kavya Golamaru

Motivation/Introduction

- Use machine learning to classify songs by decade based on lyrical themes
- Track how musical and cultural themes evolve over time
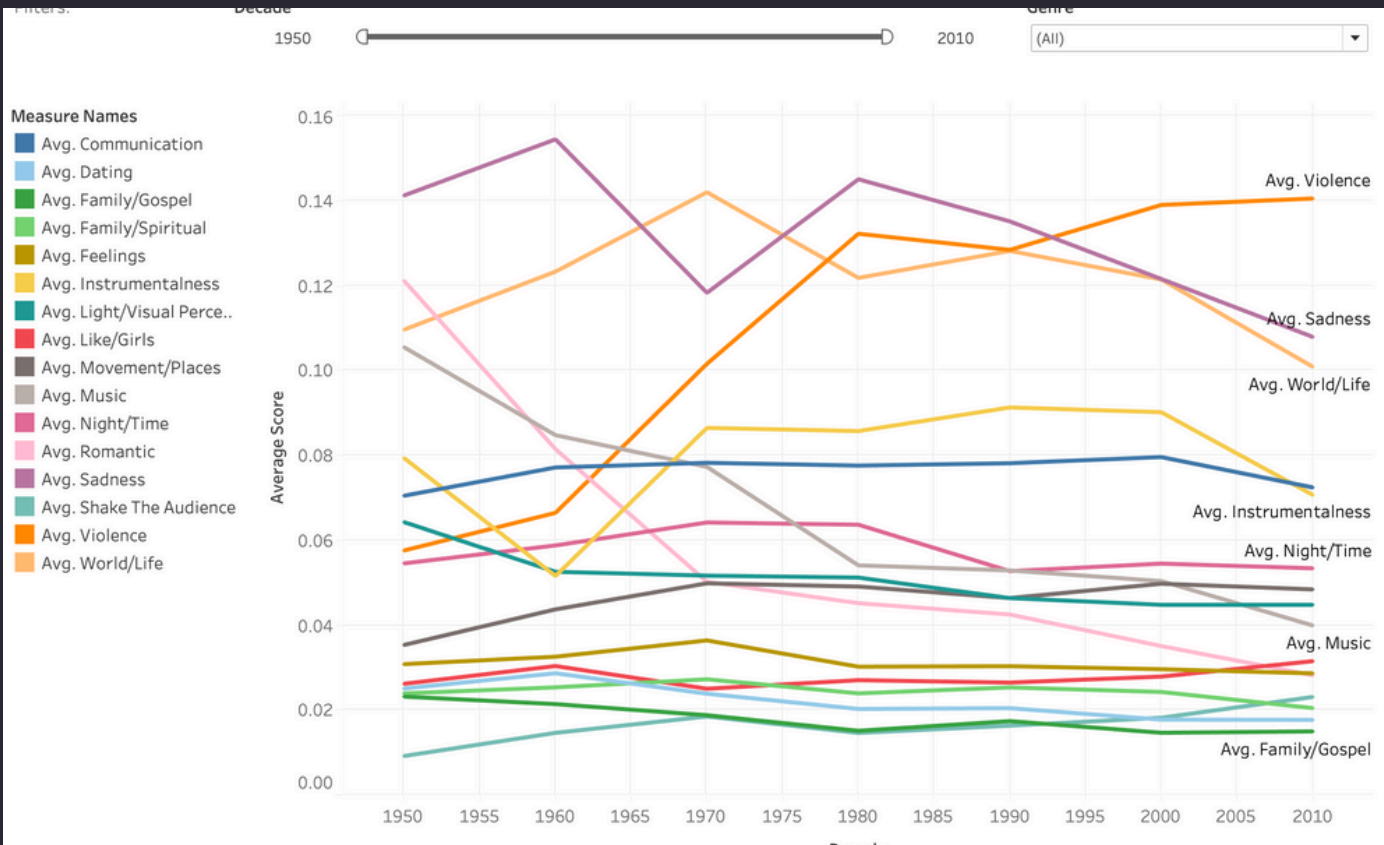- Reveal data-driven insights on how music reflects societal change

About the Data

- Kaggle "Music Dataset: 1950–2019" with lyrics and metadata
- 28K+ songs, 31 features (e.g., danceability, loudness, acousticness)
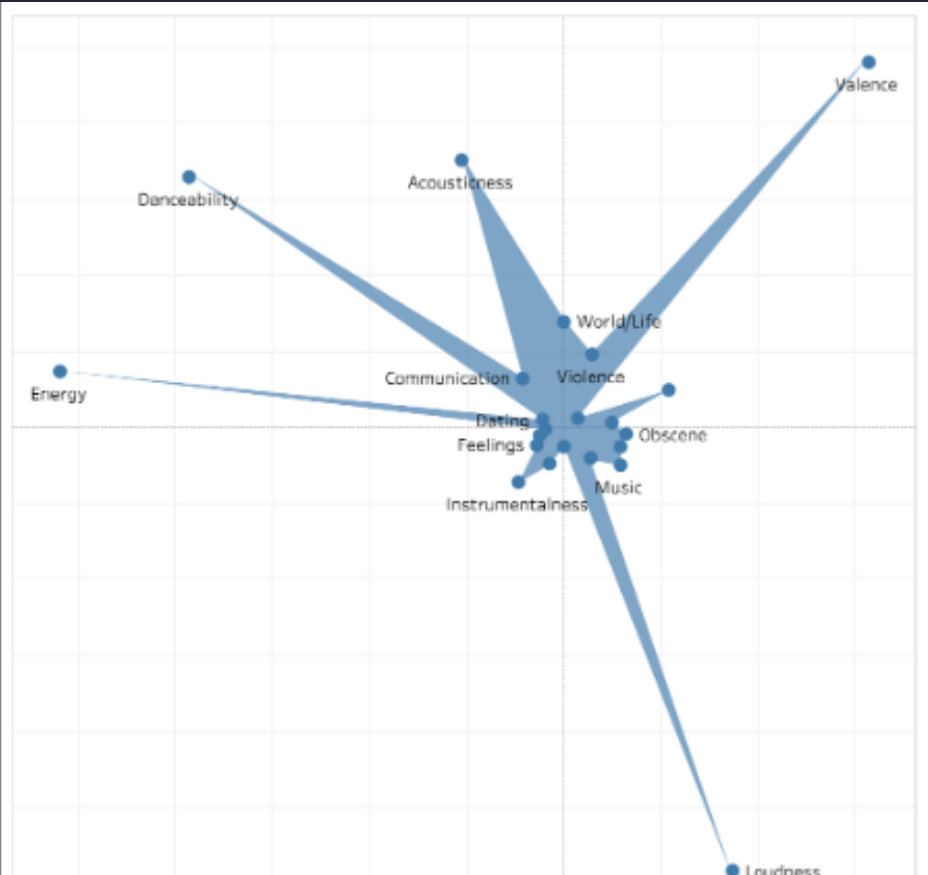- Temporal data, 27.7 MB

Our Approaches

Trend Visualizations

Line Graph of Themes
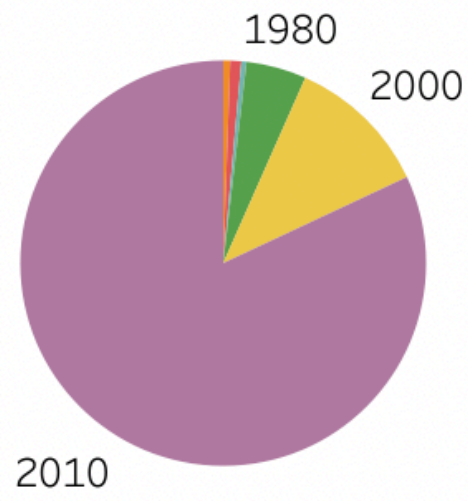

Radar Chart of Theme Prevalence

Heat Map of Theme Correlations by Genre

	communi..	dating	family/g..	family/s..	feelings	light/vis..	like/girls	moveme..	music	night/ti..
communica..		-0.0483	-0.0587	-0.0790	-0.0157	-0.1503	-0.0658	-0.1502	-0.0646	-0.0456
dating	-0.0483		-0.0018	-0.0592	0.0172	-0.0754	0.0081	-0.0740	-0.0103	0.0439
family/gosp..	-0.0587	-0.0018		0.0086	-0.0106	-0.0625	-0.0206	-0.0212	0.0052	-0.0198
family/spiri..	-0.0790	-0.0592	0.0086		-0.0432	-0.0296	-0.0663	-0.0446	0.0303	-0.0540
feelings	-0.0157	0.0172	-0.0106	-0.0432		-0.0589	0.0111	-0.0587	-0.0406	0.0098
light/visual ..	-0.1503	-0.0754	-0.0625	-0.0296	-0.0589		-0.0744	-0.0746	0.0232	-0.0408
like/girls	-0.0658	0.0081	-0.0206	-0.0663	0.0111	-0.0744		-0.0551	-0.0139	-0.0156
movement/..	-0.1502	-0.0740	-0.0212	-0.0446	-0.0587	-0.0746	-0.0551		-0.0238	-0.0483

Experiments

Participants: 12 users with varied familiarity with music & Machine Learning

Key Questions Explored (5 point scale):
- Are the visuals clear and insightful?
- Does the design help understand model outputs?
- What pain points are encountered?

Findings:
- High clarity ratings (averaging 4/5 to 5/5)
- Most users found meaningful insights
- Understanding of ML varied between users (scores 3/5-5/5)

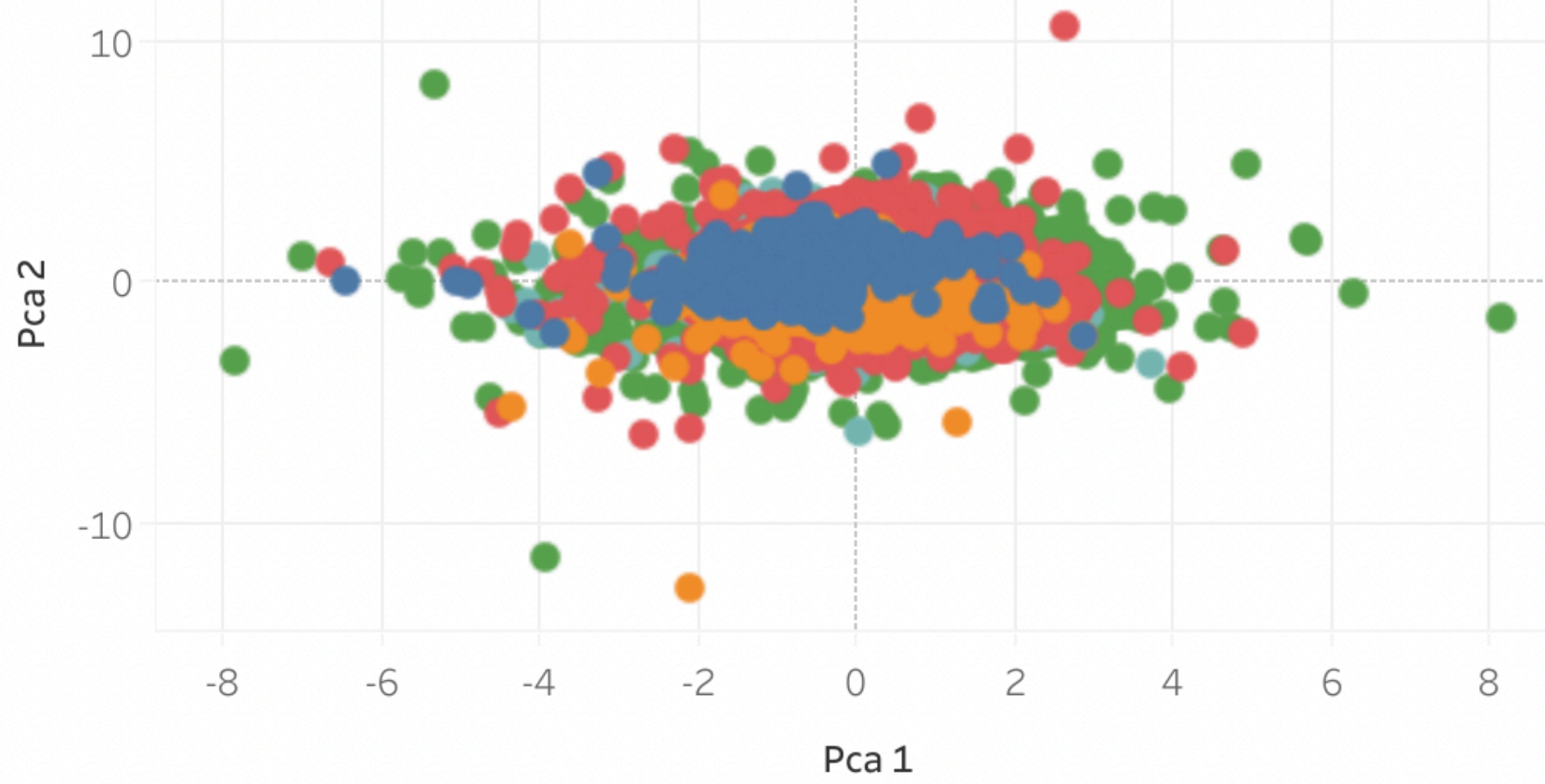**Validation**: Historical parallels validate thematic scoring, confirming the visualizations reflect real lyrical trends

Model Implementation

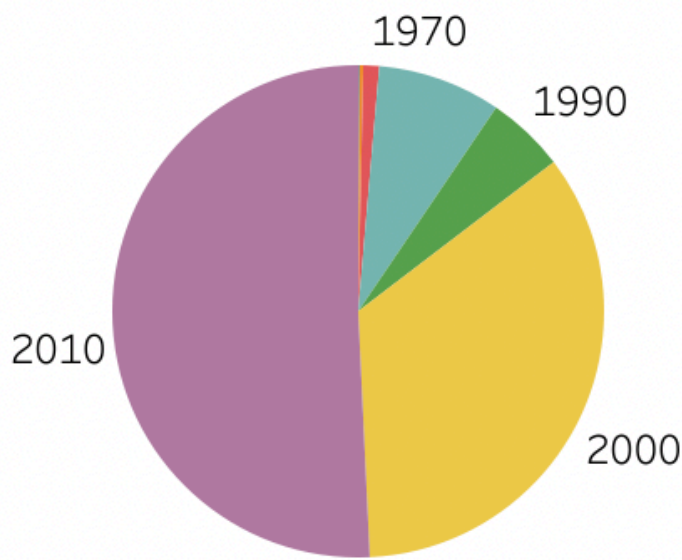- **Random Forest + XGBoost:** 22-dimensions; downsampling; GridSearchCV; best F1-score ~38%; PCA
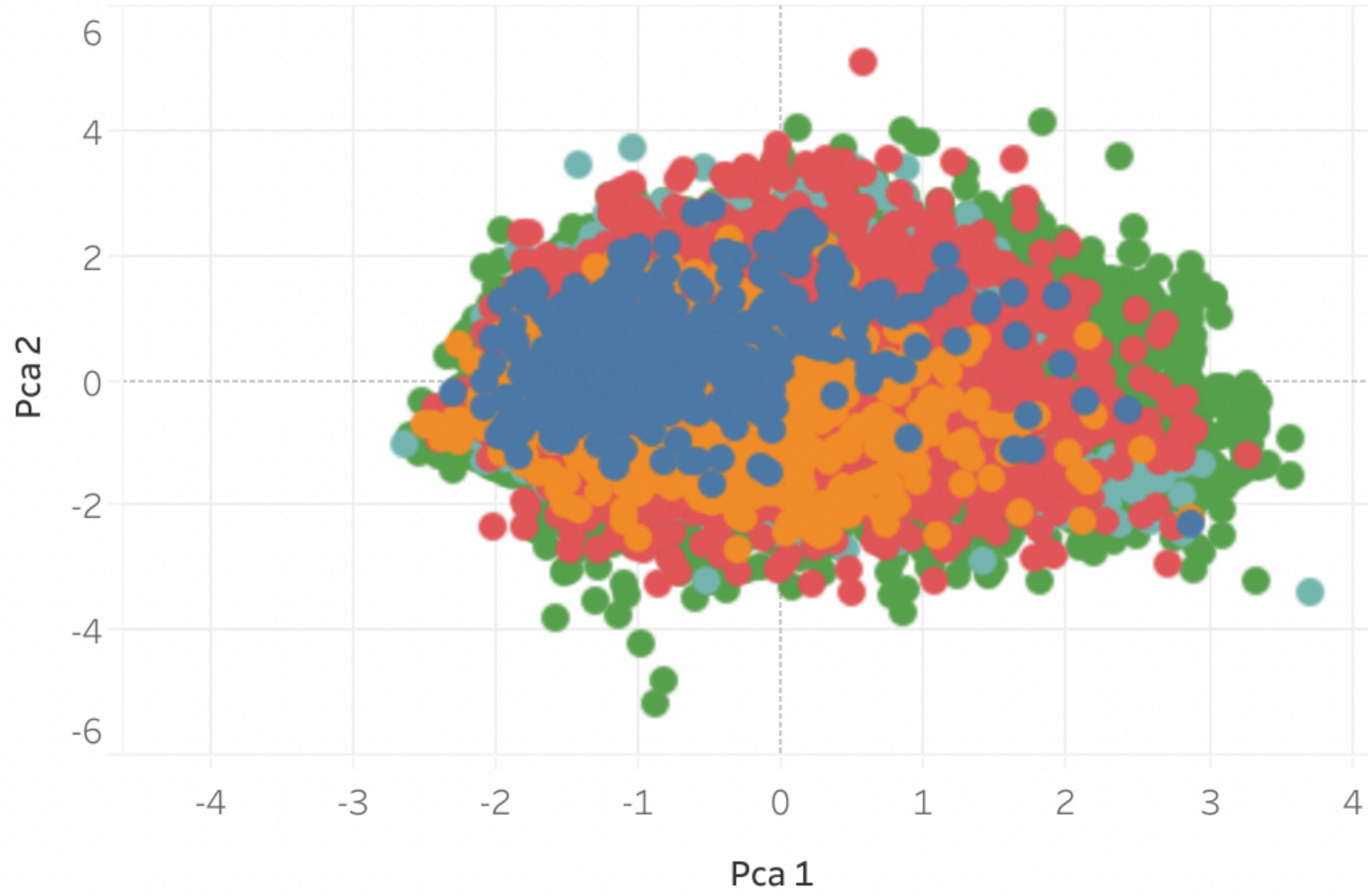

Results + Evaluation

Link to **Final Tableau Dashboard**

- **XGBoost performed slightly better overall**
  - **F1 Score: 38% Accuracy: 39.4%**

- Introduces temporal analysis to lyrical theme research.
- Proposes a novel decade classification method
- Moves beyond prior focus on genre classification and sentiment analysis of audio features.
- Offers interpretable insights into how music reflects cultural shifts.