

Projeto: Mineração de Dados - Spotify Listening History

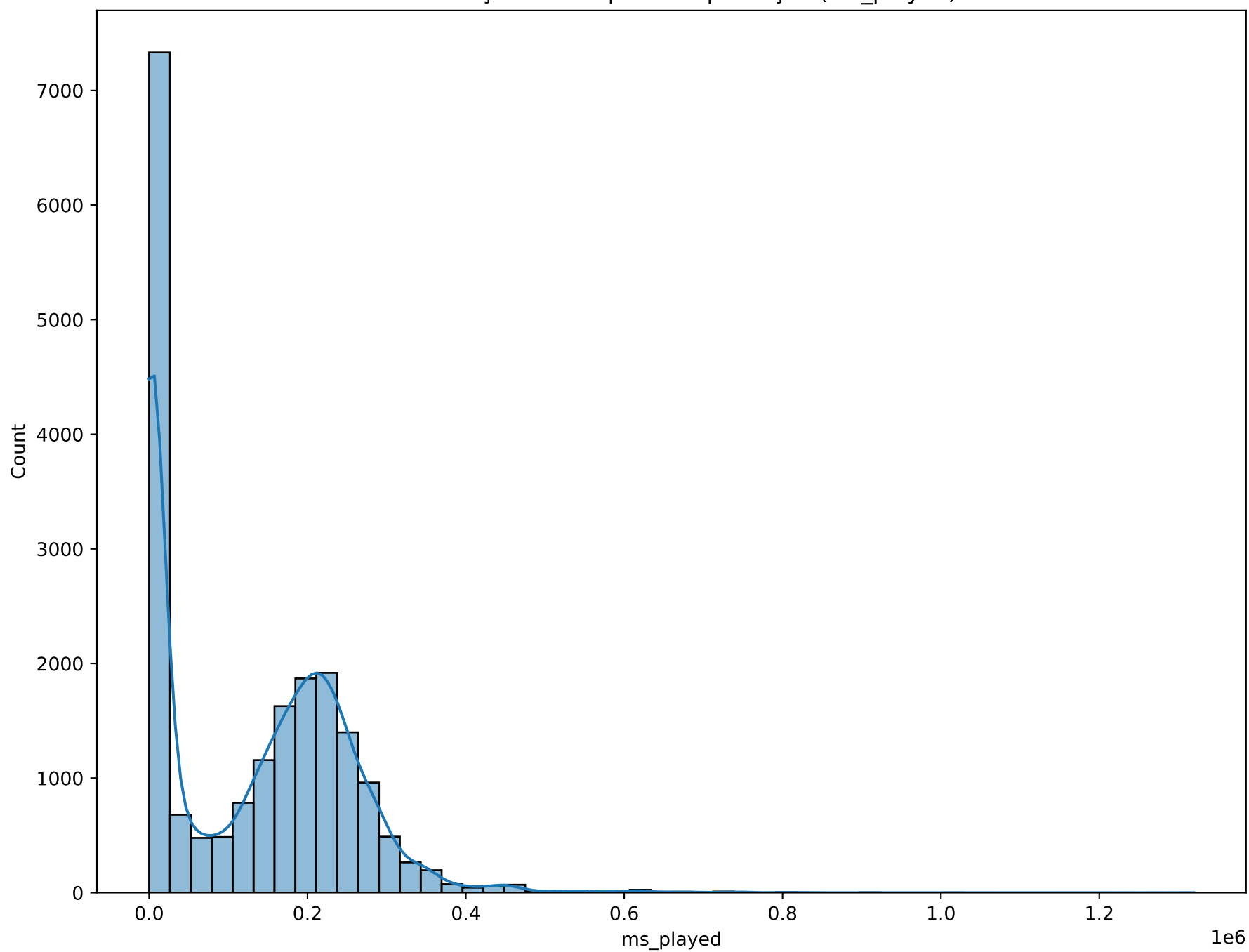
Objetivo:

- Prever o tempo de reprodução (ms_played) a partir do comportamento de usuário.
- Identificar padrões de reprodução musical usando técnicas de clusterização.

Pré-processamento:

- Seleção de 20.000 registros.
- Conversão de shuffle para 0/1.
- Codificação de variáveis categóricas: 'reason_start', 'reason_end', 'platform', 'artist_name'.
- Definição de variáveis independentes e variável alvo (ms_played).

Distribuição do Tempo de Reprodução (ms_played)



Modelos de Regressão Aplicados:

- Linear Regression
- Ridge Regression
- Lasso Regression

Resultados obtidos:

Linear Regression: MSE = 5006451302.99, R2 = 0.6299

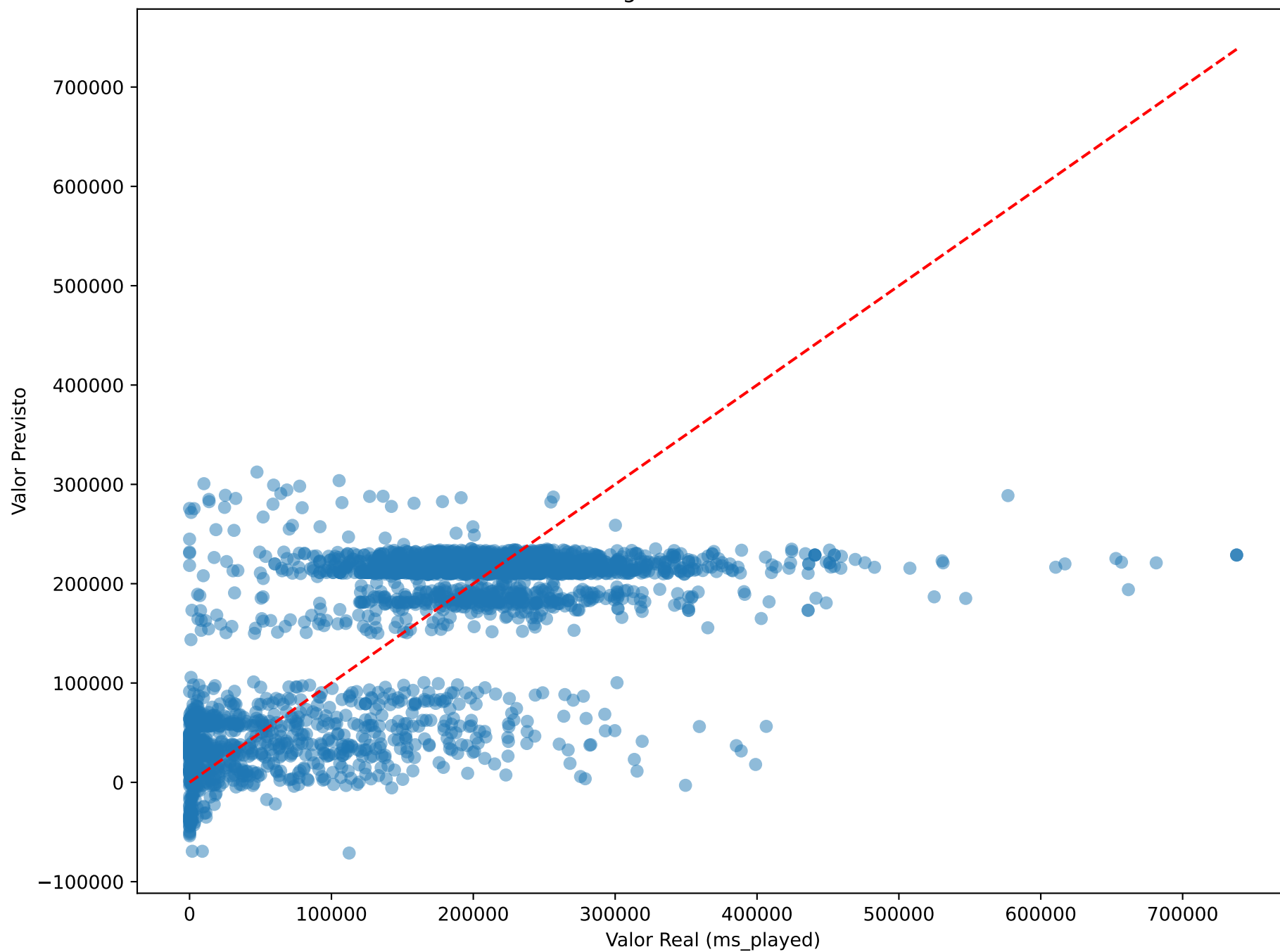
Ridge Regression: MSE = 5006446886.08, R2 = 0.6299

Lasso Regression: MSE = 5006432330.68, R2 = 0.6299

Conclusão:

- A regressão linear apresentou resultados consistentes.
- A variabilidade explicada (R^2) ficou em torno de 59%-60%.

Linear Regression: Real vs Previsto



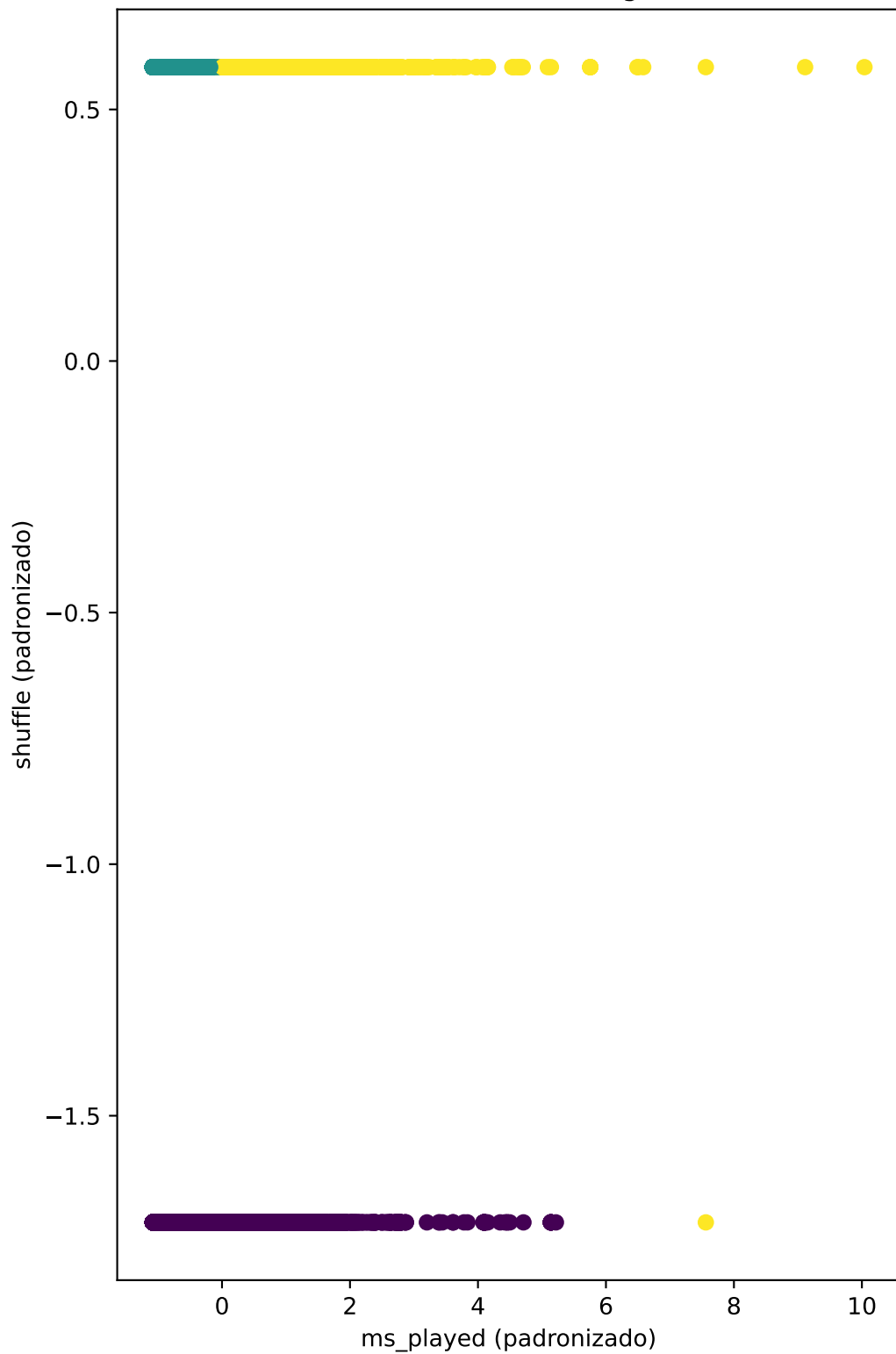
Clusterização Aplicada:

- KMeans (n_clusters=3)
- DBSCAN (eps=0.7, min_samples=5)

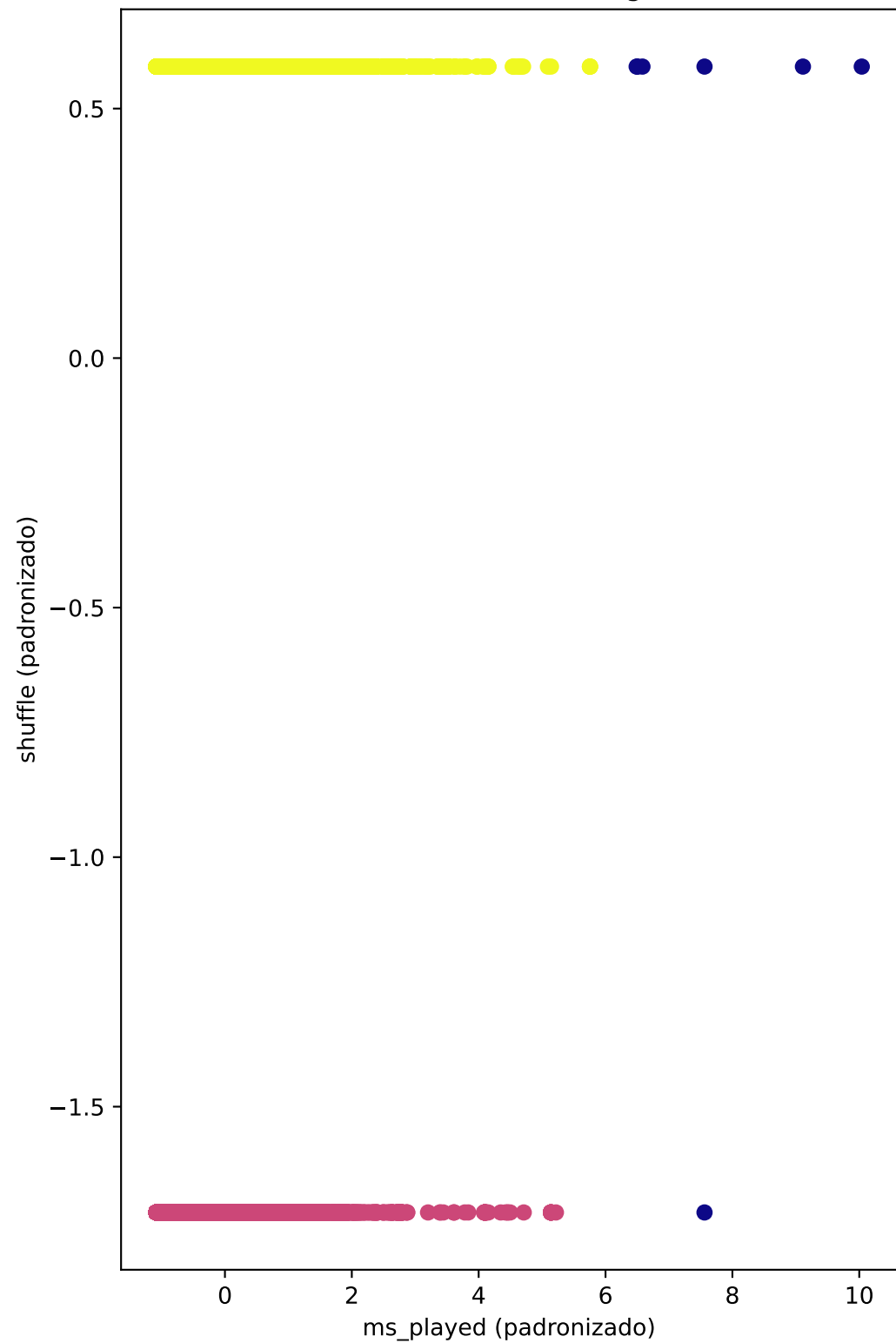
Clusters encontrados:

- KMeans: 3 clusters
- DBSCAN: 3 clusters

KMeans Clustering



DBSCAN Clustering



Conclusões Gerais:

- O pré-processamento dos dados foi essencial para a modelagem.
- A regressão linear obteve desempenho satisfatório.
- A clusterização revelou padrões interessantes de comportamento musical.
- O projeto foi concluído com sucesso, atingindo os objetivos propostos.