

Projeto: Mineração de Dados - Spotify Listening History

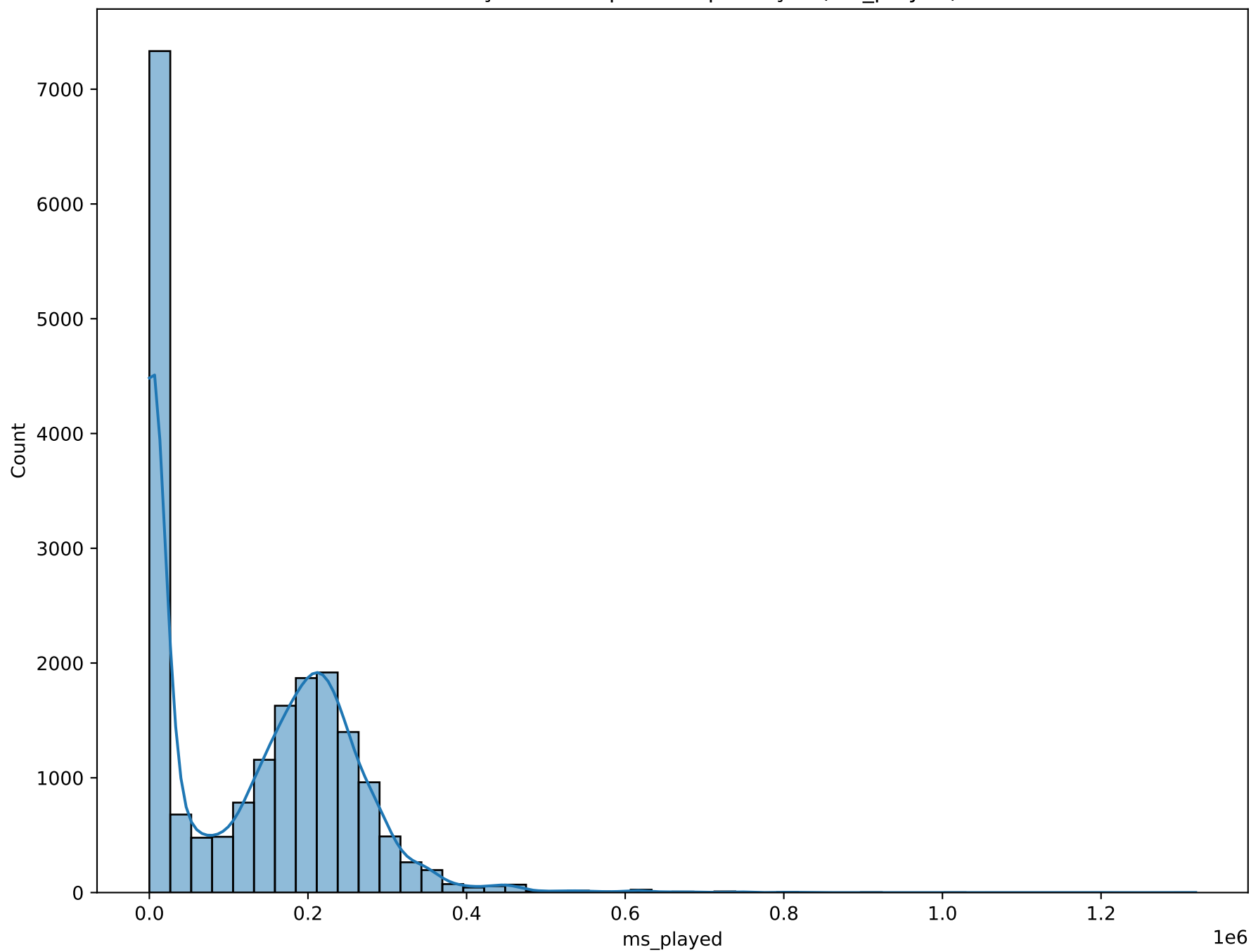
Objetivo:

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

Pré-processamento dos Dados:

- Seleção de 20.000 registros aleatórios.
- Conversão da variável 'shuffle' para valores 0 (False) e 1 (True).
- Codificação numérica de variáveis categóricas ('reason_start', 'reason_end', 'platform', 'artist_name').
- Verificação de dados faltantes:
 - Total de valores nulos encontrados: 0
- Nenhuma linha precisou ser removida.
- Dados prontos para a divisão entre treino e teste.

Distribuição do Tempo de Reprodução (ms_played)



Modelos de Regressão Aplicados:

- Linear Regression: método clássico para prever uma variável contínua.
- Ridge Regression: regressão linear com regularização para evitar overfitting.
- Lasso Regression: regressão linear que pode zerar coeficientes irrelevantes.
- ElasticNet: combinação de Ridge e Lasso.
- Bayesian Ridge: incorpora probabilidade na previsão dos coeficientes.

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

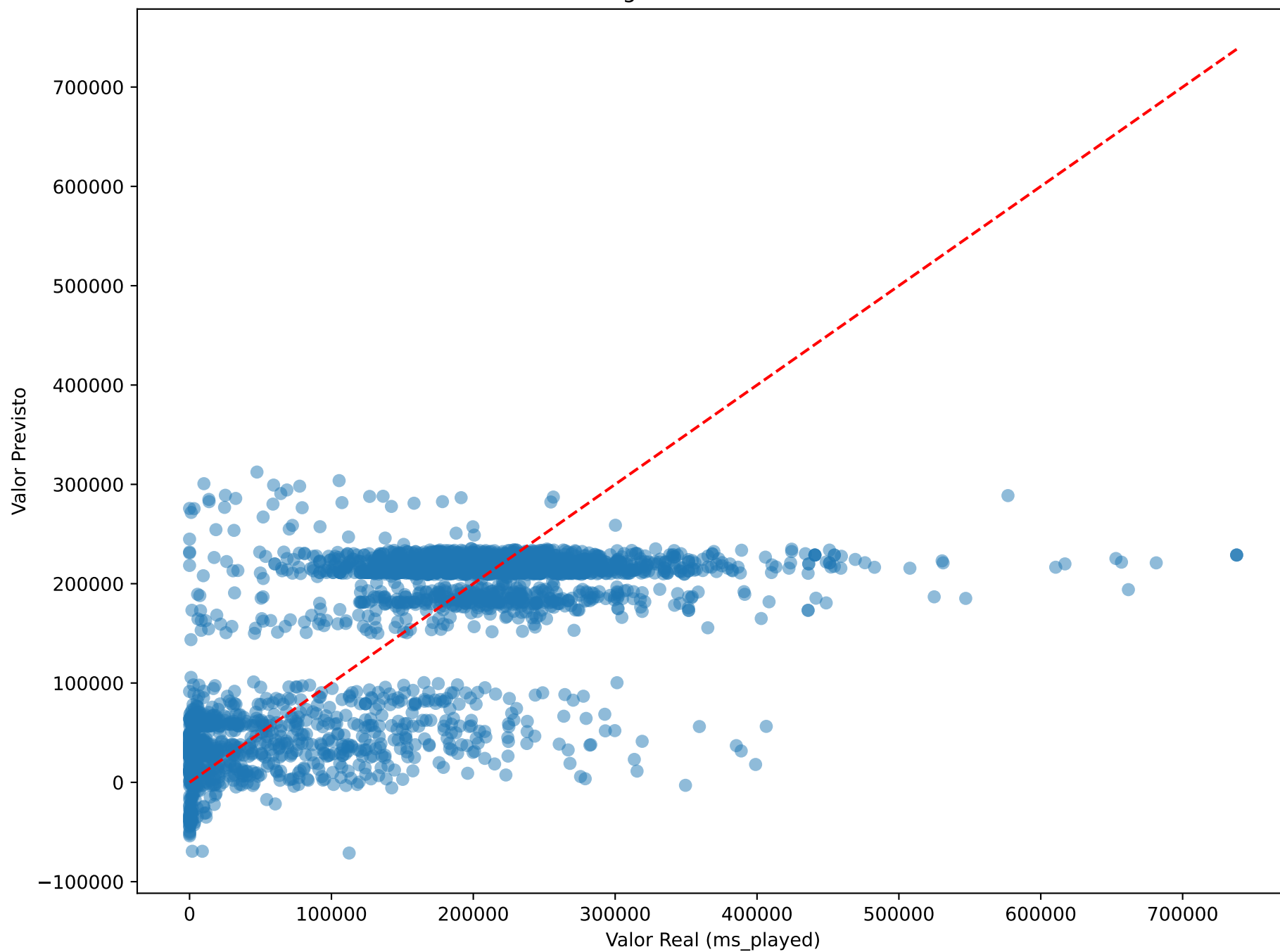
ElasticNet Regression: MSE = 5023760820.39, R2 = 0.6287

Bayesian Ridge Regression: MSE = 5006242108.59, R2 = 0.6299

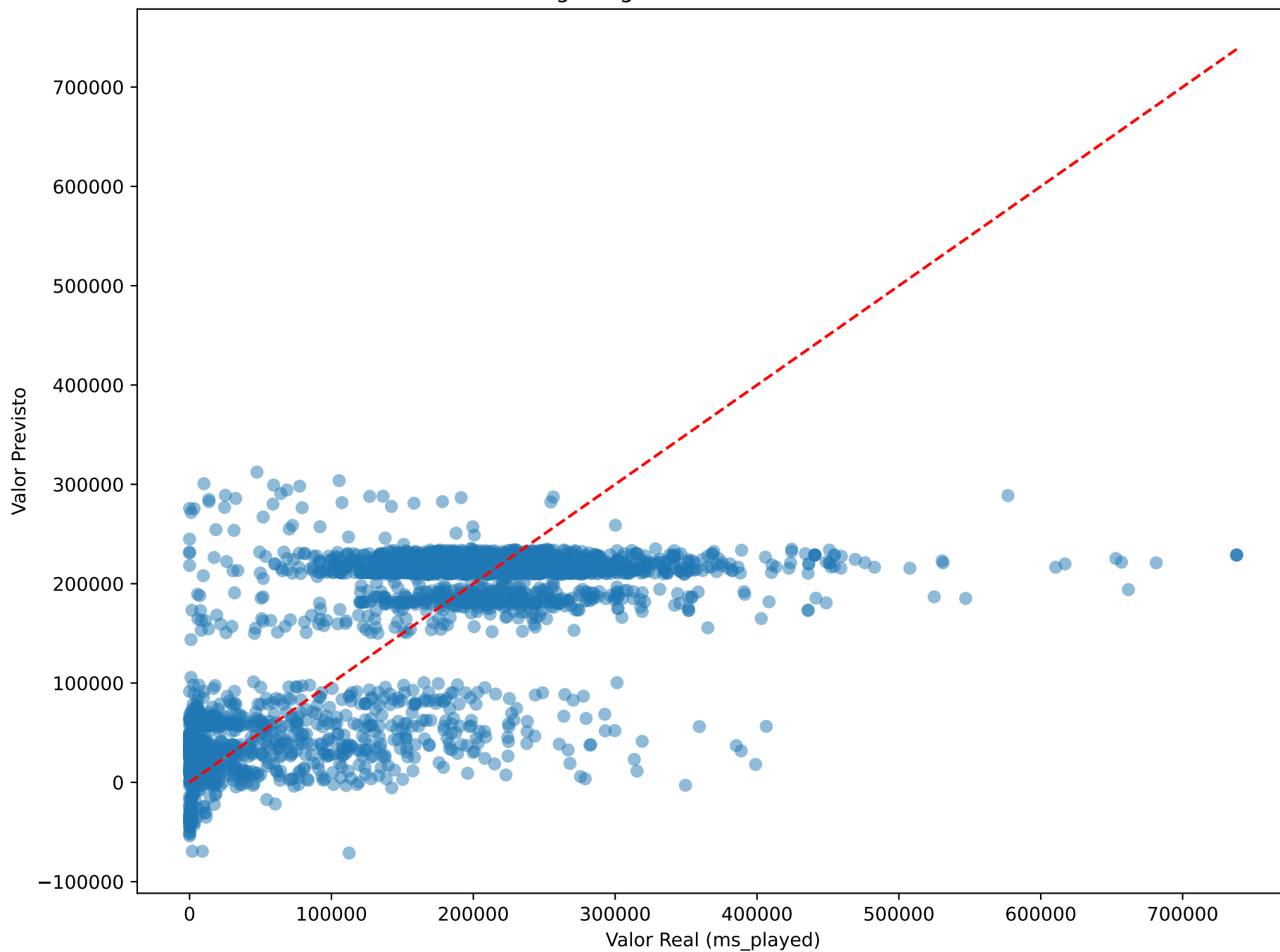
Conclusão:

- Modelos lineares e regularizados apresentaram desempenhos muito próximos.

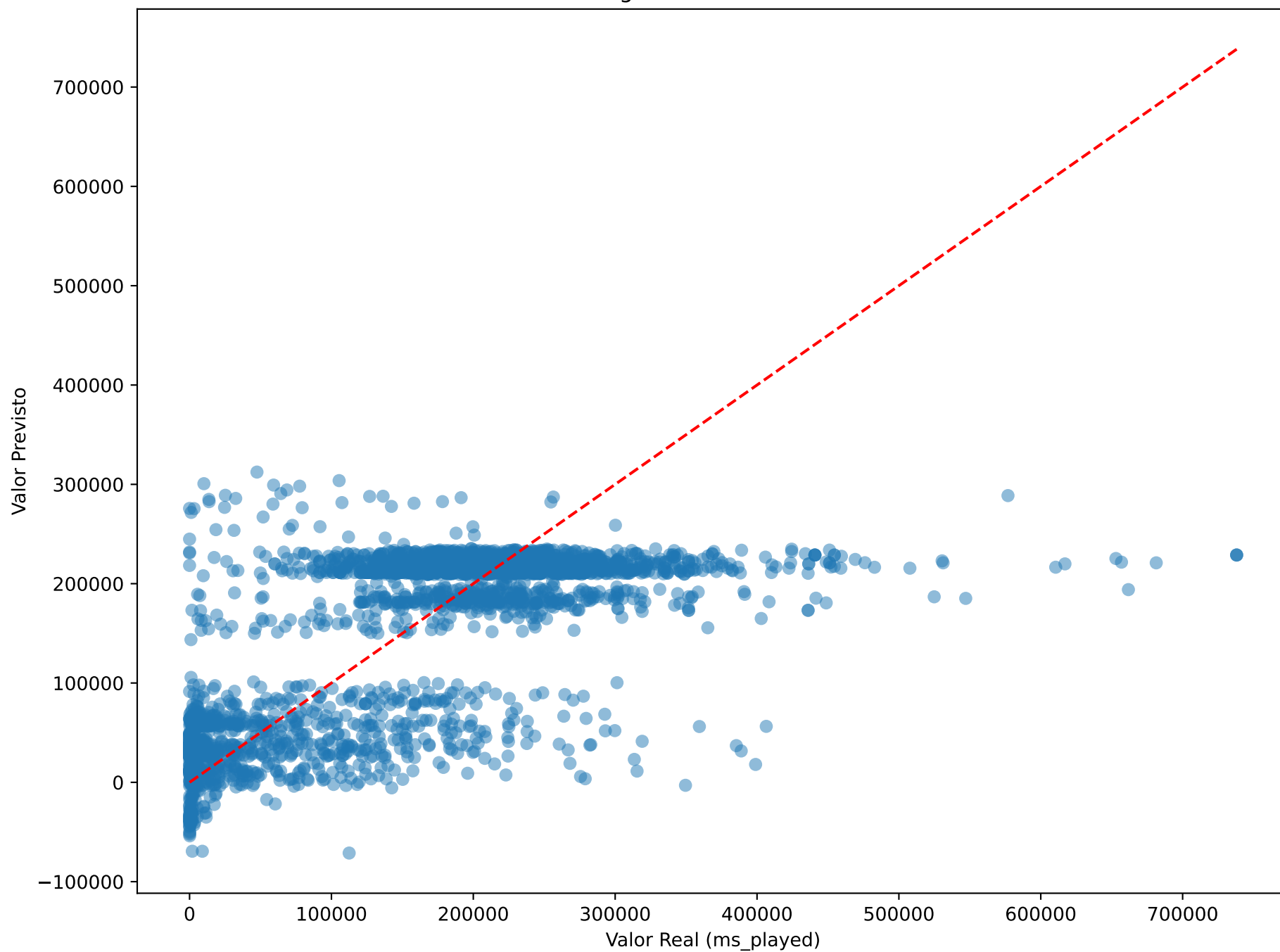
Linear Regression: Real vs Previsto



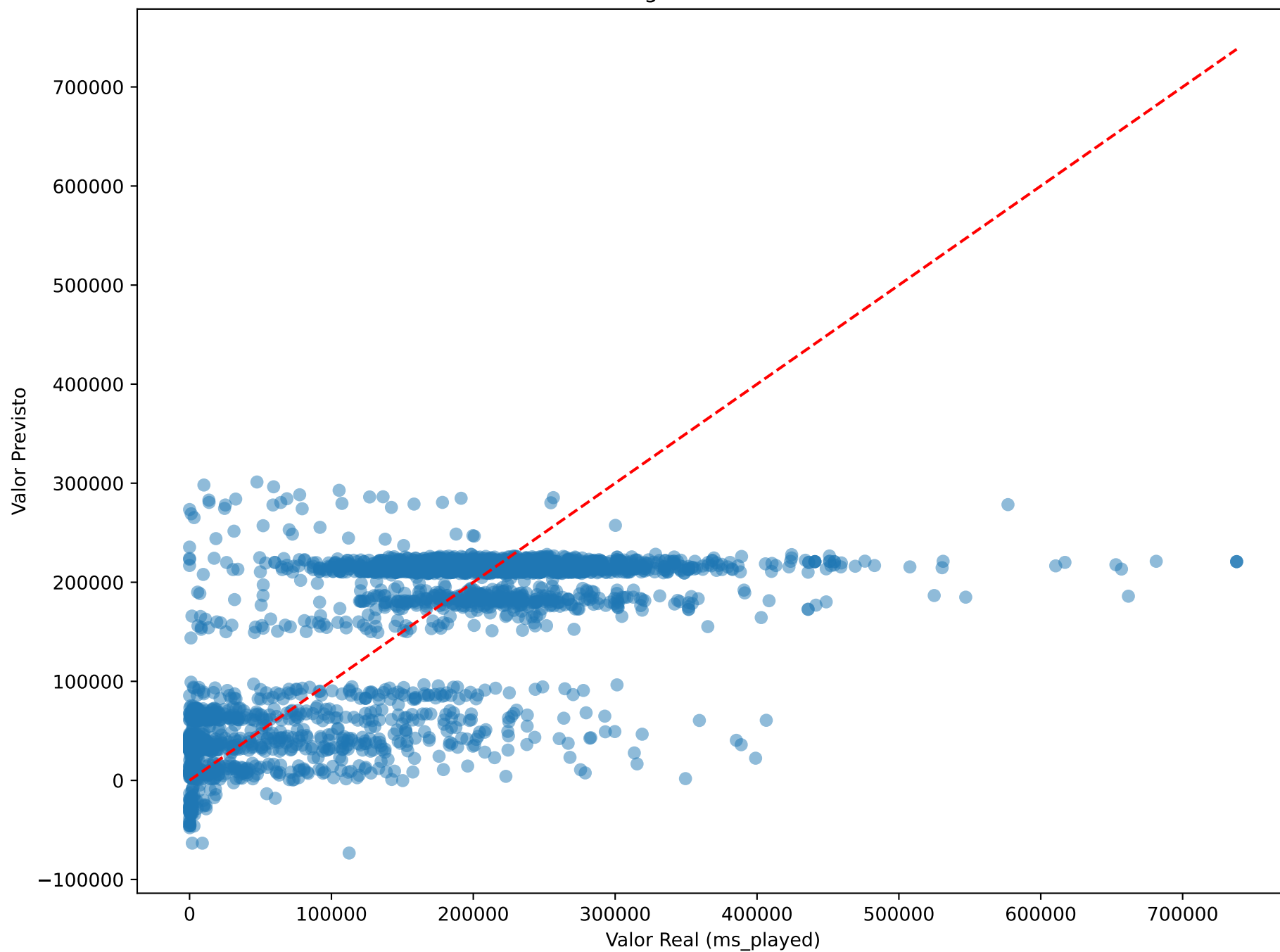
Ridge Regression: Real vs Previsto



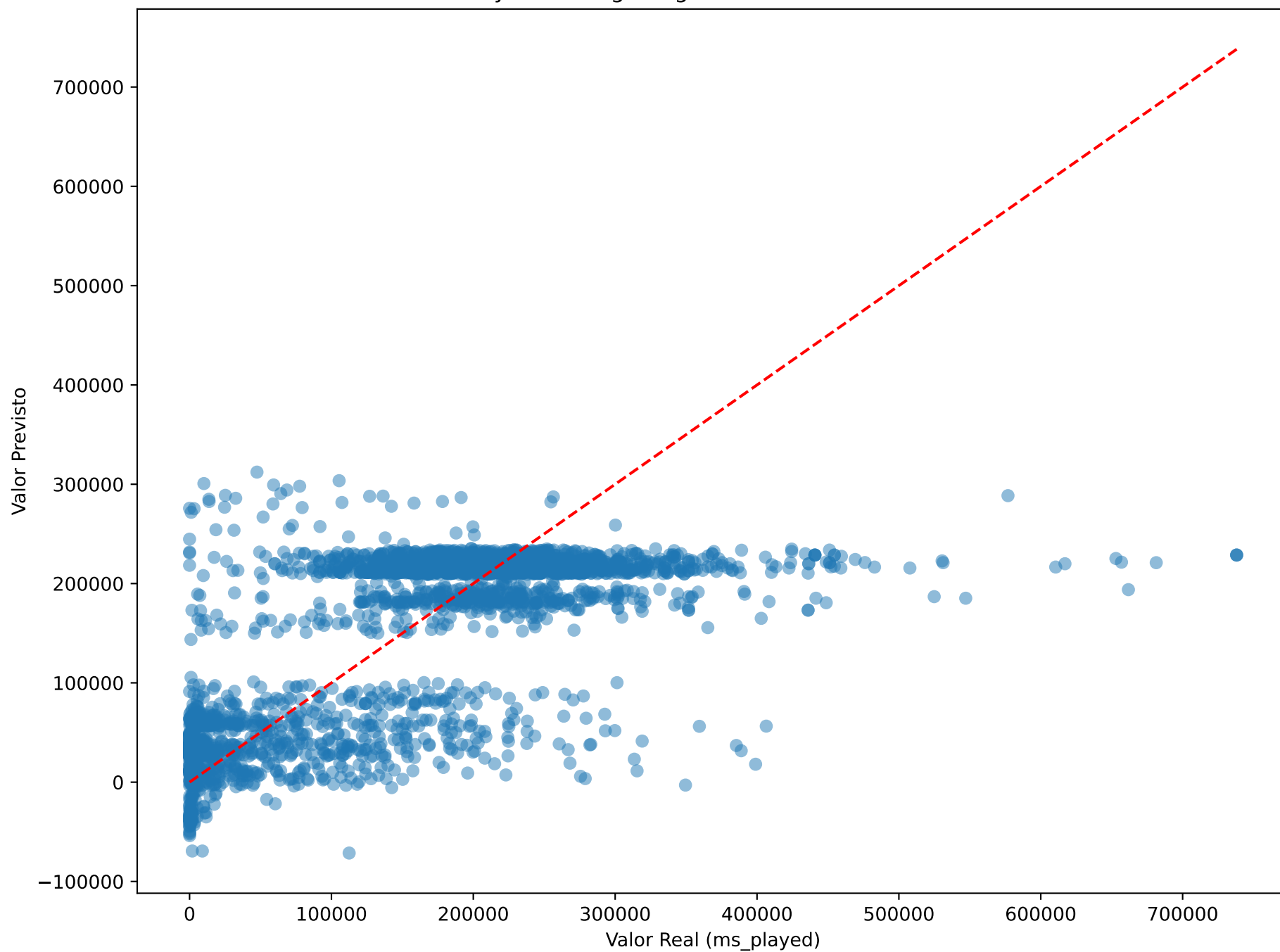
Lasso Regression: Real vs Previsto



ElasticNet Regression: Real vs Previsto



Bayesian Ridge Regression: Real vs Previsto



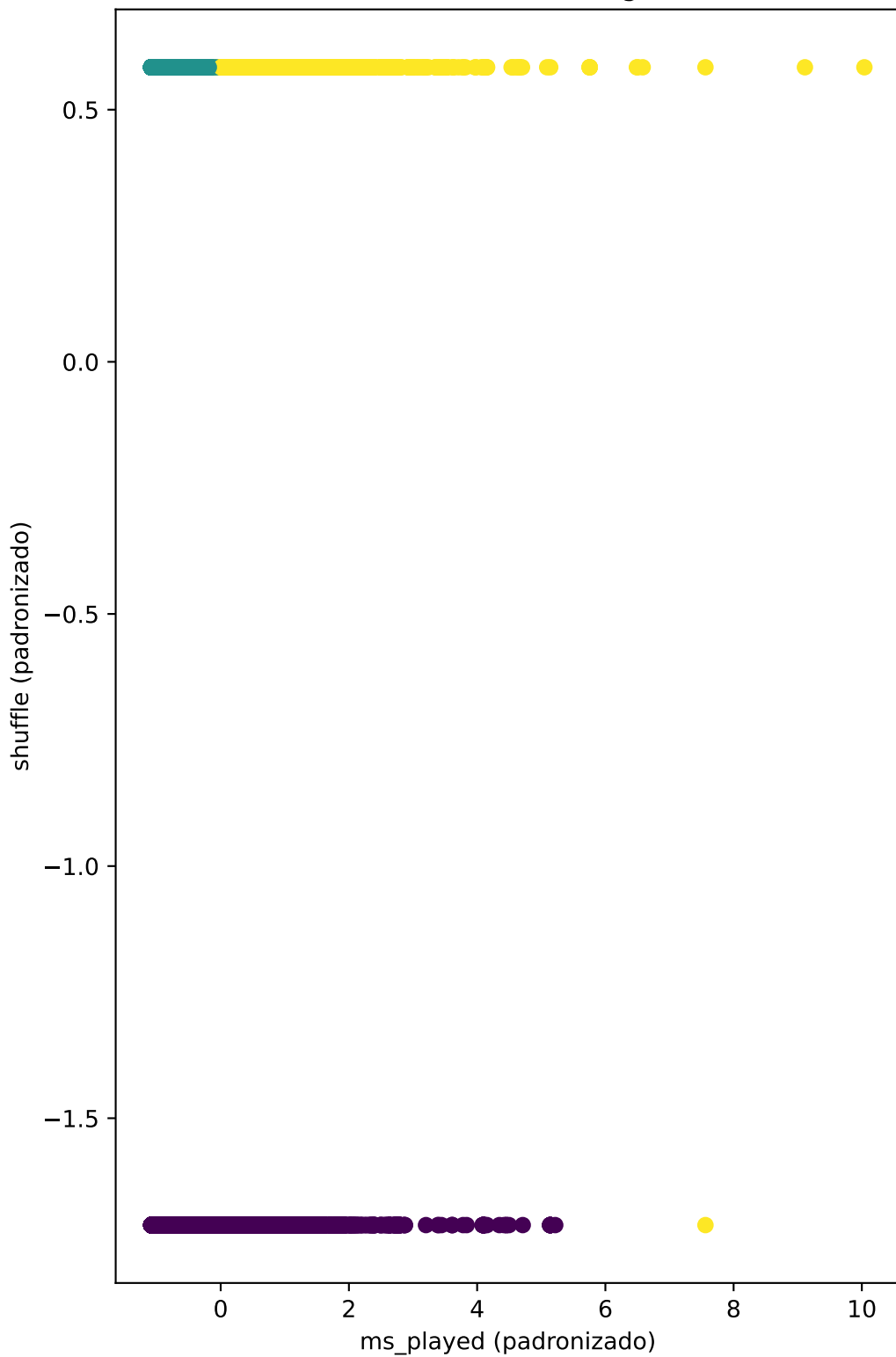
Clusterização Aplicada:

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

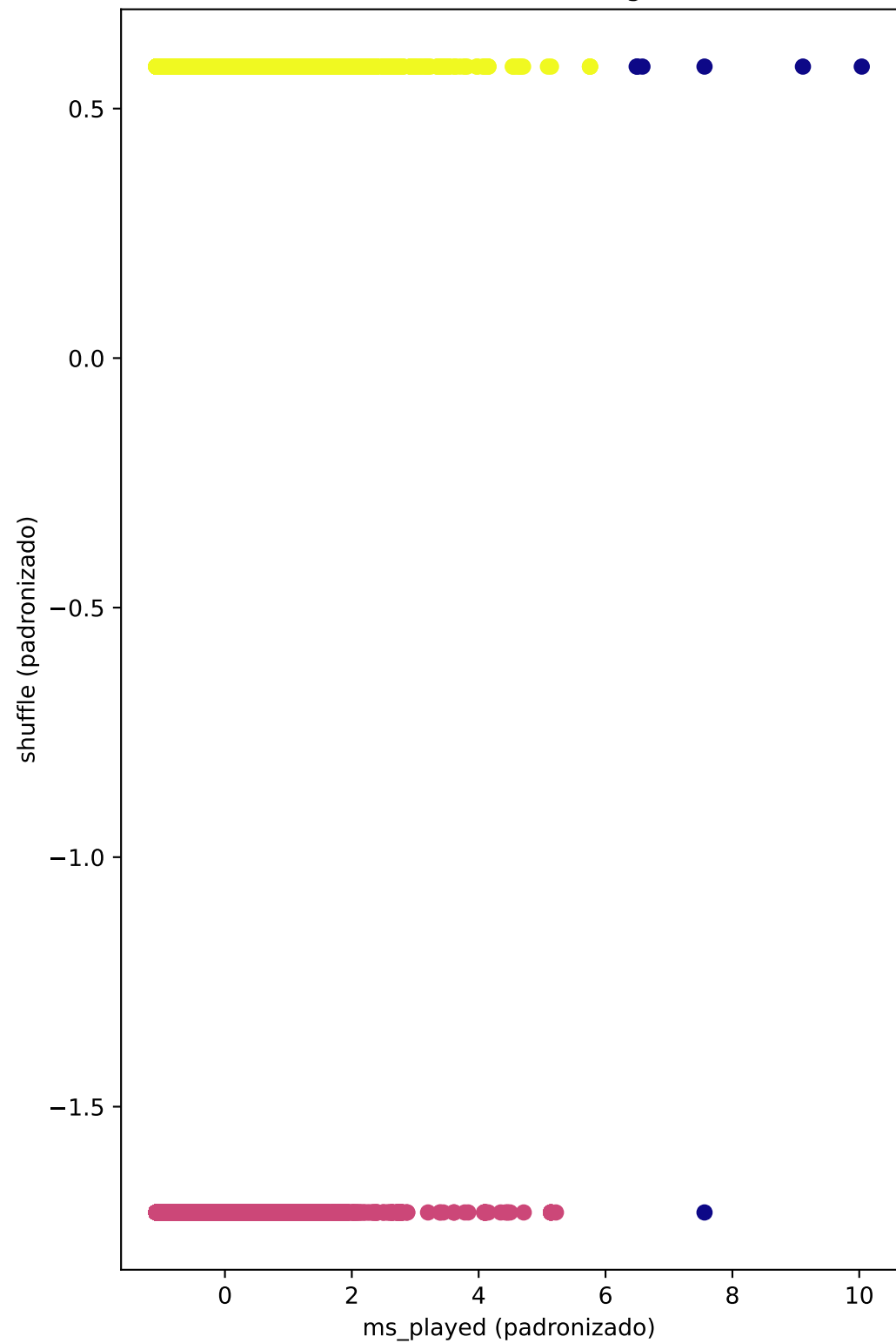
Clusters encontrados:

- KMeans: 3 clusters
- DBSCAN: 3 clusters

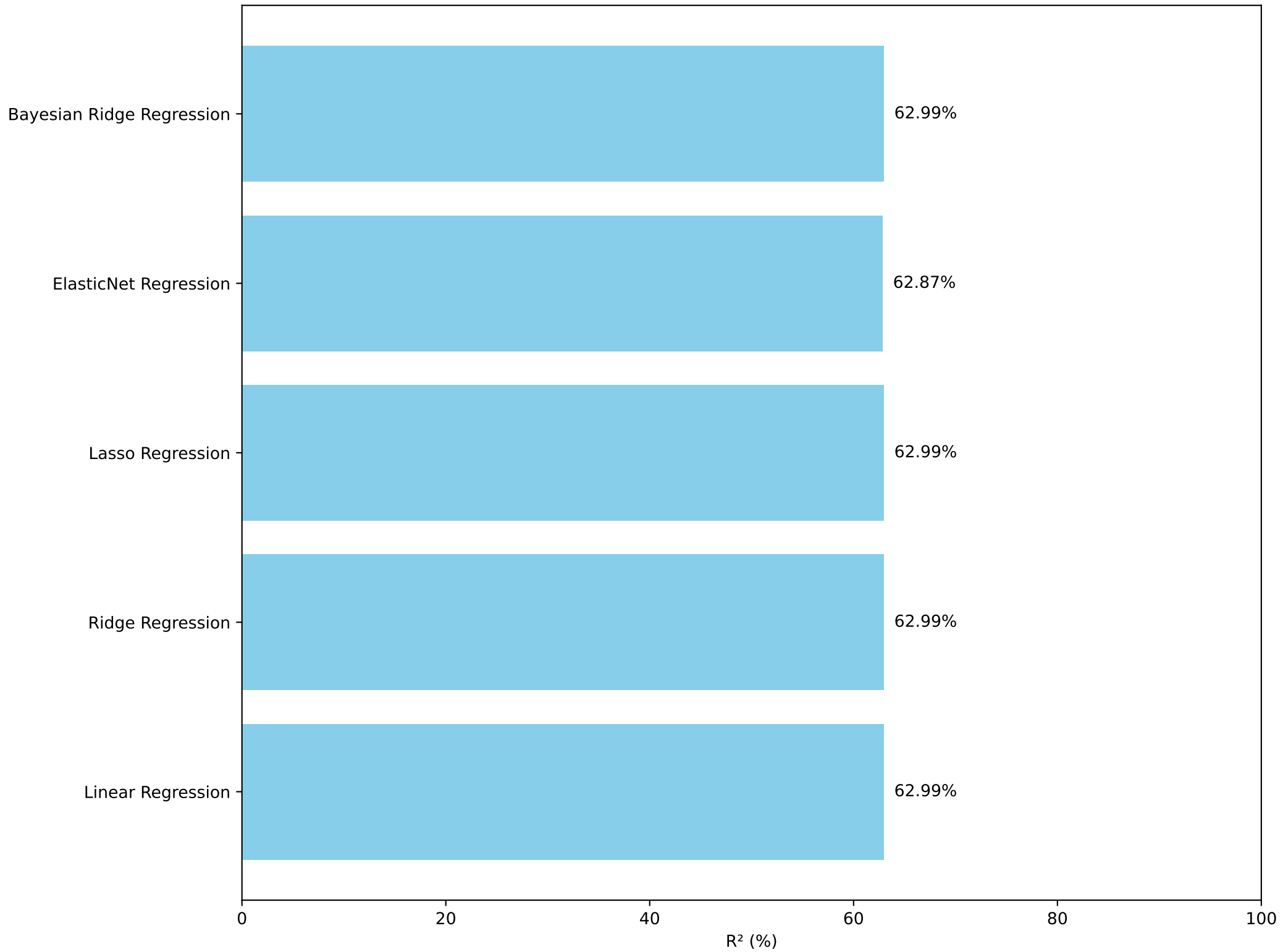
KMeans Clustering



DBSCAN Clustering



Comparação de Desempenho dos Modelos de Regressão



Conclusões Gerais:

- O pré-processamento dos dados garantiu qualidade na modelagem.
- Modelos lineares e regularizados tiveram desempenhos semelhantes.
- A clusterização identificou padrões de comportamento musical.
- O projeto foi concluído com sucesso.