IMDb Top 1000 Project - Key Questions & Analysis Report

# 1. What factors most influence IMDb ratings?

- Number of Votes was the strongest predictor of IMDb rating, indicating popularity and visibility play a major role.  
- Meta Score (critic reviews) closely aligns with IMDb scores, showing that critical reception impacts audience ratings.  
- Gross Revenue had a moderate to strong influence, with financially successful films often receiving higher ratings.  
- Director identity was a key factor, with well-known directors consistently delivering highly rated content.  
- Released Year had some influence, hinting at evolving rating behaviors over time.

# 2. Are certain genres or directors more successful?

- Director proved to be a top influencer, with names like Christopher Nolan and Quentin Tarantino frequently appearing in higher-rated entries.  
- Genre showed some effect, though genre combinations limited interpretability. Further decomposition (splitting genre tags) could improve insights.  
- Success appears to align more with individual talent (director) than broad genre classification.

# 3. Can we predict IMDb rating from available features?

- A Random Forest model achieved a Root Mean Squared Error (RMSE) of ~0.29, showing high prediction accuracy.  
- This confirms that IMDb rating is predictable from features like votes, reviews, and production metadata.  
- The model can be used for rating estimation or identifying potential high-rated content based on early metadata.