Engineering a Decision-Support System for Risk-Aware Movie Production Strategy

This project transforms raw, unreliable movie data into a decision-ready system designed to support financial risk management and capital allocation.

Role: Lead Data Analyst / Data Engineer (Solo Project) | **Tools:** Excel (Power Query), Python (Pandas), Relational Modeling | **Dataset:** TMDb 5000 Movie Dataset (Raw and Unstructured)

THE CHALLENGE

The Problem: Raw movie datasets are often unreliable due to missing financials and deeply nested structures. Decisions based on incomplete or corrupted data increase the risk of systematic overinvestment in high-budget failures.

The Goal: Transform 4,800+ rows of messy, JSON-encoded data into a clean relational database for a simulated production studio to identify high-ROI talent and budget "sweet spots."

THE EXECUTION

I built an end-to-end data pipeline to ensure data integrity and reduce analytical bias:

- **Data Recovery and Cleaning:** Resolved 3,400+ encoding errors (Mojibake) to prevent underrepresentation of international talent. Adopted a preservation-over-deletion strategy, manually researching missing financials for 100+ films to ensure historical coverage.
- **Architectural Engineering:** De-nested complex JSON structures into a six-table relational model using Power Query and Python. Standardized metrics into a Normalized ROI Index for fair comparisons across different eras.
- **Strategic Feature Engineering:** Developed a Consistency Score to evaluate directors based on sustained profitability rather than isolated high-performing projects.

THE INSIGHTS

The engineered dataset revealed business patterns previously hidden by data noise:

- **The Mid-Budget Opportunity:** Films with budgets between $20M–$50M showed a 35% higher median ROI than mega-budget productions, indicating a lower-variance investment profile with strong scalability.
- **The Safe Hands Effect:** Identified a distinct group of directors who maintained a 100% profitability rate across varying budget sizes, highlighting the value of consistency in risk-managed planning.

STRATEGIC RECOMMENDATIONS

- **Weighted Capital Allocation:** Prioritize mid-budget productions led by historically consistent directors to stabilize portfolio-level returns.
- **Risk Hedging:** Allocate limited capital to high-variance blockbusters as brand and visibility drivers rather than primary profit engines.

KEY SKILLS DEMONSTRATED

- **Technical:** Advanced Power Query for data modeling, Python-based data cleaning, Relational Database Design (Star Schema).
- **Analytical:** Financial modeling, Risk assessment, Feature engineering.
- **Business:** Strategy development, Data storytelling, Decision-support design.

Full page technical documentation, data validation scores, and transformation logic are available for review via my portfolio.