■ AI-Powered Data Report

Dataset: **Movie_collection_train.csv**

■ Dataset Overview

Rows	400
Columns	19
Numeric Columns	16
Categorical Columns	3
Missing Values	8

■ *AI Insight: The dataset has 400 rows and 19 columns. 16 numeric columns show varying levels of spread. 3 categorical columns capture distinct labels. There are 8 missing values across the dataset.*

■ Workflow Log

• File 'Movie_collection_train.csv' successfully loaded.
• DataFrame shape: (400, 19)
• Columns detected: ['Collection', 'Marketin_expense', 'Production_expense', 'Multiplex_coverage', 'Budget', 'Movie_length', 'Lead_ Actor_Rating', 'Lead_Actress_rating', 'Director_rating', 'Producer_rating', 'Critic_rating', 'Trailer_views', 'Time_taken', 'Twitter_hastags', 'Genre', 'Avg_age_actors', 'MPAA_film_rating', 'Num_multiplex', '3D_available']
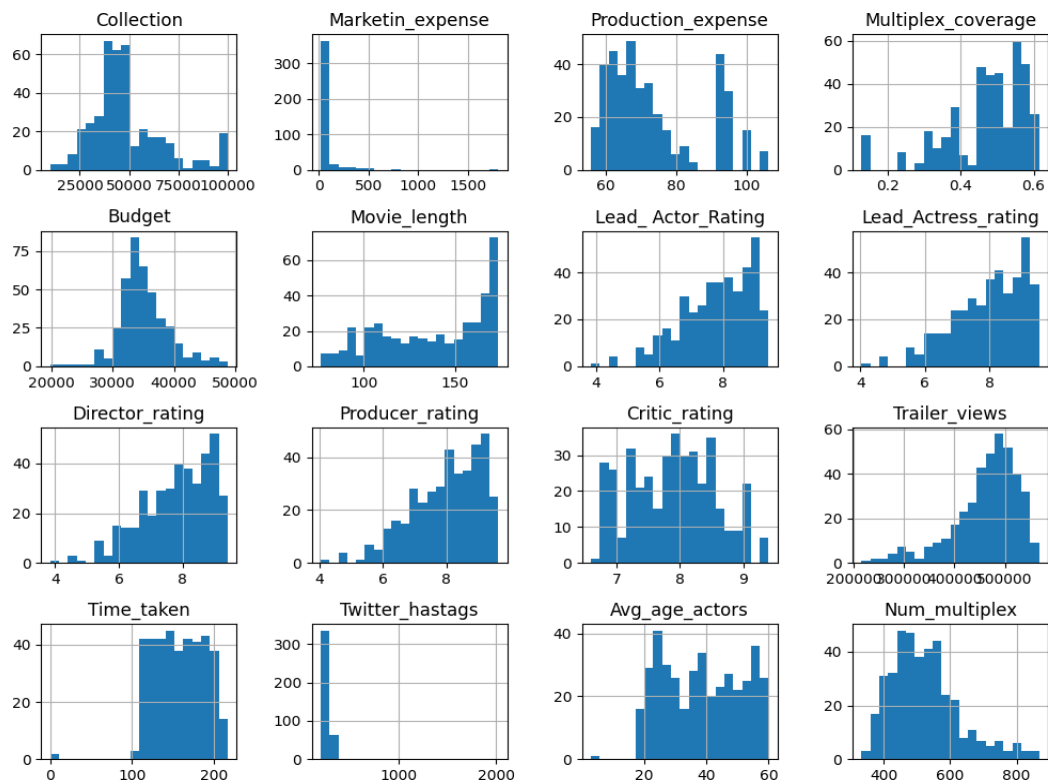
■ AI Narrative Insights

[TinyLlama error: [WinError 10061] No connection could be made because the target machine actively refused it]
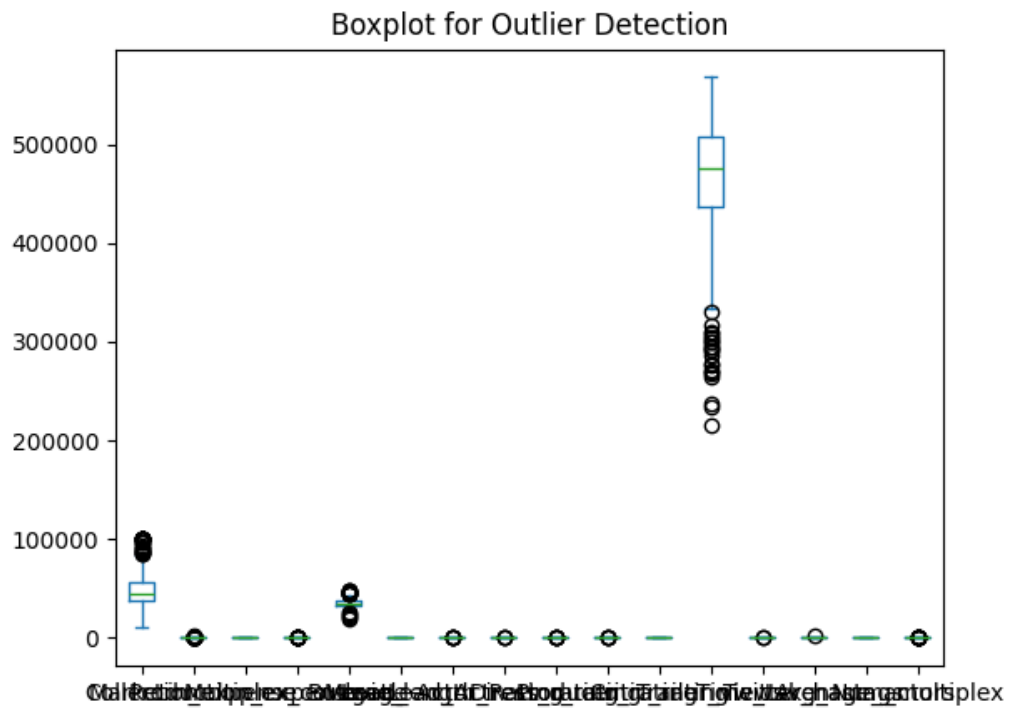
■ Weighted Statistics

Metric	Value
Collection	48646.50
Marketin_expense	55.02
Production_expense	73.83
Multiplex_coverage	0.47
Budget	35197.77
Movie_length	137.58
Lead_ Actor_Rating	7.81
Lead_Actress_rating	7.98
Director_rating	7.81
Producer_rating	7.99
Critic_rating	7.89
Trailer_views	463223.57
Time_taken	157.79
Twitter_hastags	258.08
Avg_age_actors	38.72
Num_multiplex	523.65

■ Visual Diagnostics

Distribution of Numeric Columns

Boxplot for Outlier Detection

Boxplot for Outlier Detection

Correlation Heatmap

Correlation Heatmap