INVESTASI DI INDUSTRI FILM
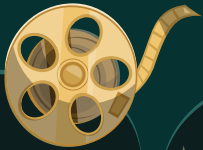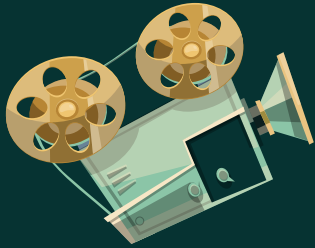
Benedicta Winni L
Des 2022

TABLE OF CONTENTS

Background

VS

Worldwide Box
Office Performance:
$2,797,732,053

Worldwide Box
Office Performance:
$218,838,386

Reference: Endgame & Little Women

Main Objectives

Menganalisis faktor yang dapat mempengaruhi profit dari suatu film.

Menganalisis karakteristik film yang dapat memberikan keuntungan dalam berinvestasi.

Data

- Terdiri dari **7668** judul film dari tahun 1980 hingga 2020
- Terdapat **15 variabel**:

- Budget
- Company
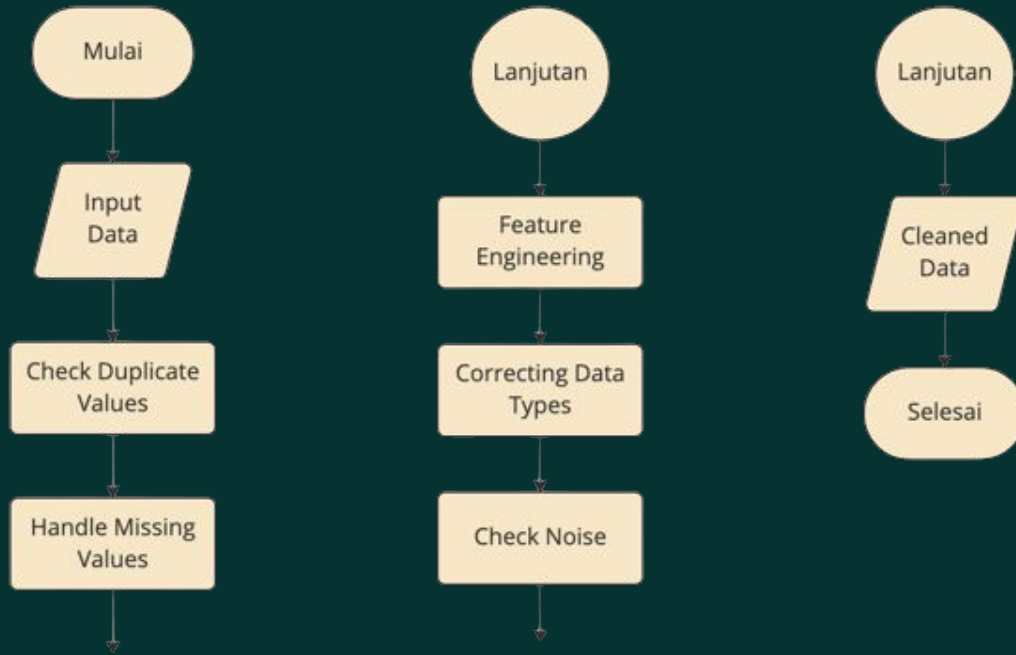- Country
- Director
- Genre

- Gross
- Name
- Rating
- **Released**
- Runtime

- Score
- Votes
- Star
- Writer
- Year

Reference: Movie Industry Data

Data Preprocessing

Mulai

↓

Input Data

↓

Check Duplicate Values

↓

Handle Missing Values

↓

Lanjutan

↓

Feature Engineering

↓

Correcting Data Types

↓

Check Noise

↓

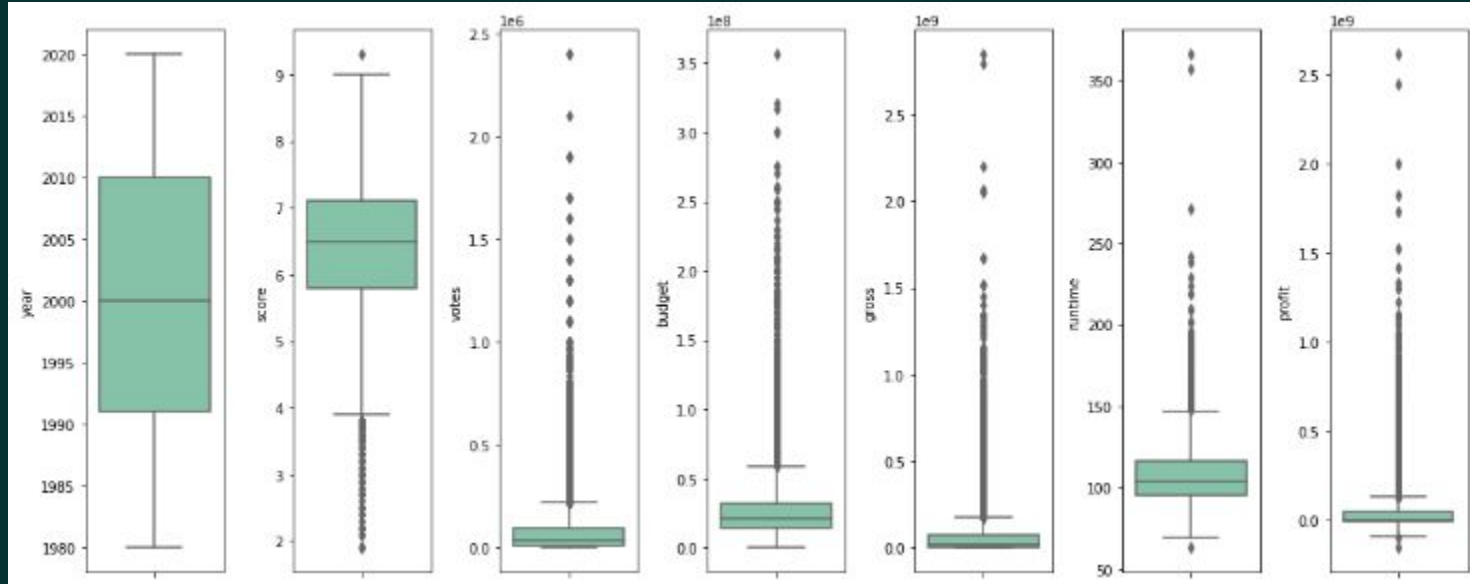Lanjutan

↓

Cleaned Data

↓

Selesai

Handling Missing Values

Feature Engineering

- Menambahkan kolom baru **profit**.
- Melakukan **regroup** pada variabel **rating**.

Not Rated –> NR
Unrated –> NR
TV–MA –> NC–17
TV–PG –> PG
X –> NC–17
TV–14 –> PG–13

Outliers Analysis

Perkembangan Jumlah Film (1980–2020)
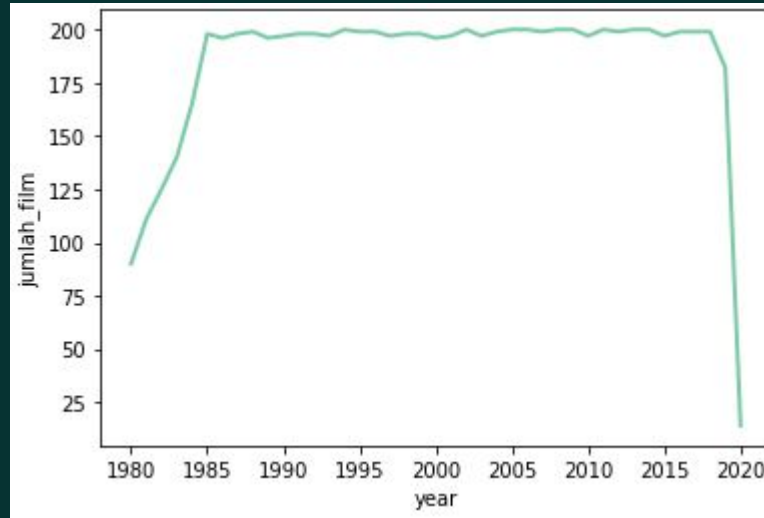
Genre dan Rating dengan Profit Terbesar (2010-2020)

Rata-Rata Profit berdasarkan Genre Film (2010-2020)

Rata-Rata Profit berdasarkan Rating Film (2010-2020)
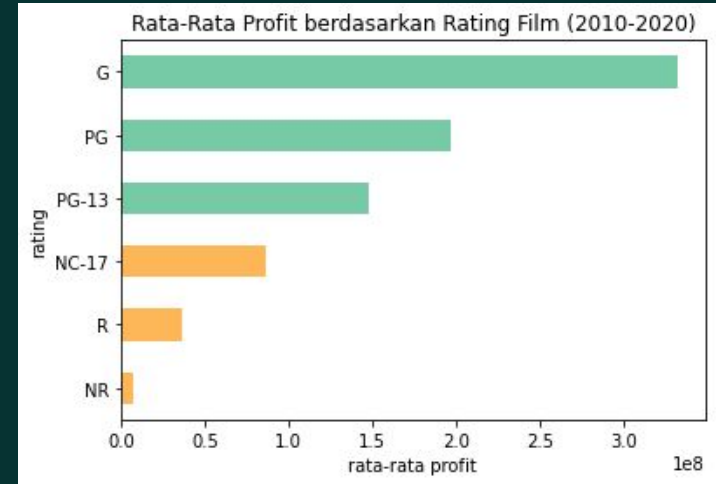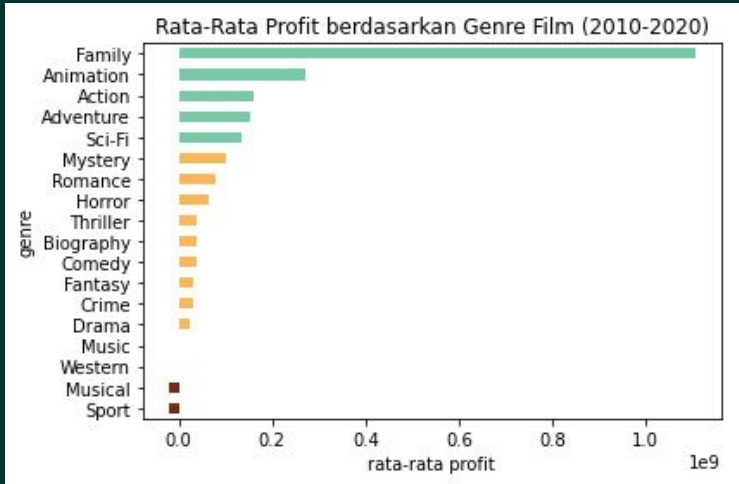
Perkembangan Profit Film yang Termasuk Top 5 Genre
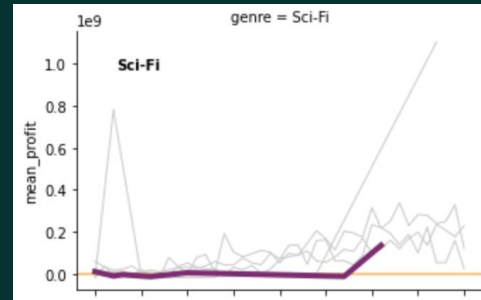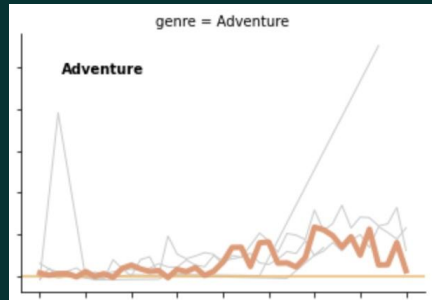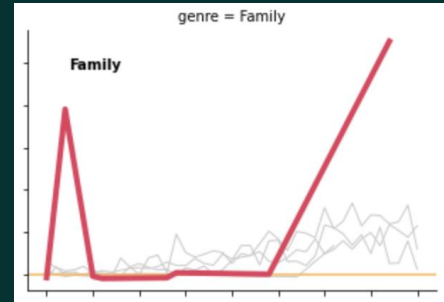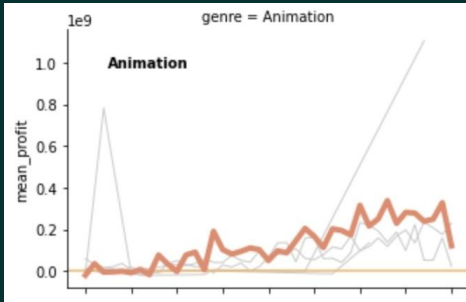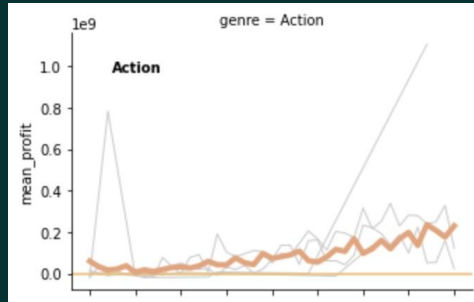
Pairwise Correlation
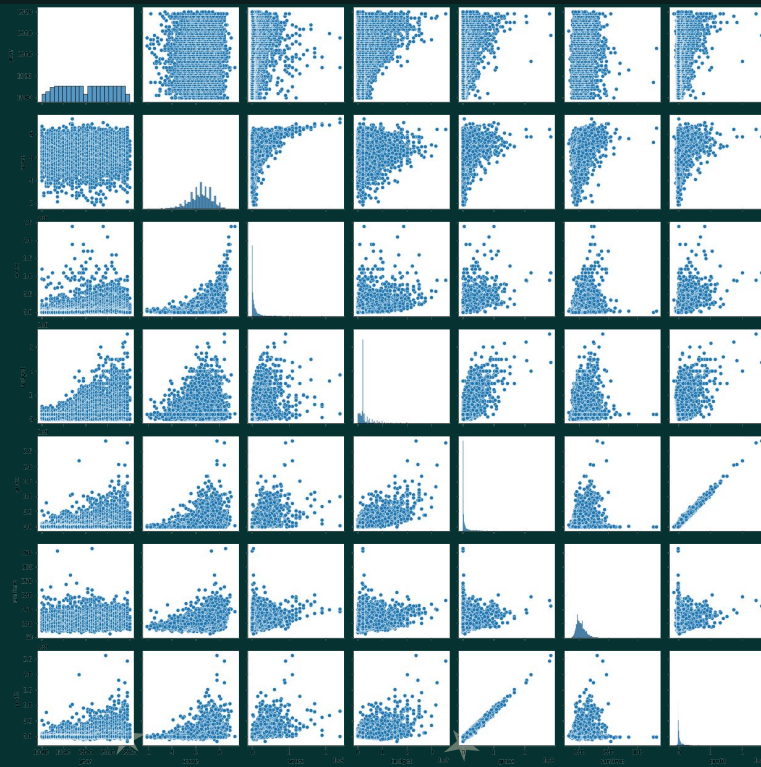
Modeling

Variabel	Koefisien
Year	2.572e+06
Score	2.405e+06
Votes	6.249e+07
Budget	5.773e+07
Runtime	-4.717e+04
Genre	-1.328e+18
Rating	4.223e+19

Model Evaluation

Model	MAE (Train)	RMSE (Train)	MAE (Test)	RMSE (Test)
Multiple Linear Regression (Asumsi heterogeneity dan normalitas tidak terpenuhi)	49.302.066	98.178.785	1.731.851.949.402.676	2.549.898.100.765.335
Ridge Regression	49.301.522	98.178.871	1.731.961.797.309.738	2.550.059.920.592.202
Lasso Regression	49.302.066	98.178.785	1731851962087328	2549898119445452

Hyperparameter Tuning (Ridge)

Alpha	RMSE (Ridge Regression)	RMSE (Lasso Regression)
0,01	101.972.729	101.970.416
0,1	101.827.435	101.970.414
1	100.845.656	101.970.393
10	99.348.595	101.970.185

Model	MAE (Train)	RMSE (Train)	MAE (Test)	RMSE (Test)
Ridge Regression (alpha = 10)	48.914.129	98.049.388	44.624.793	75.278.893

Model Comparison

Model	MAE (Train)	RMSE (Train)	MAE (Test)	RMSE (Test)
Ridge Regression (alpha = 10)	48.914.129	98.049.388	44.624.793	75.278.893
Ridge Regression	49.301.522	98.178.871	1.731.961.797.309.738	2.550.059.920.592.202
Lasso Regression	49.302.066	98.178.785	1731851962087328	2549898119445452

Actionable Insights

- Faktor yang perlu diperhatikan: .**votes** dan **budget**
- Karakteristik yang dapat memberikan: genre film **Family, Action, Animation, Adventure**; rating **G, PG,atau PG-13**; **score film yang tinggi**; **budget film tinggi** (diekspektasi memberikan profit tinggi)
- Melakukan analisis lebih lanjut dengan mempertimbangkan engagement actor/actress yang menjadi pemeran utama, produser, director, serta perusahaan produksi film

Thank You