

近代の映画評価と企業応用の分析 ラチゼ

<https://github.com/ahoucvtw>



目次



テーマについて

- 目的
- データソース



分析

- 決定木
- 他の分析方法



データ基礎俯瞰

- 評価定義
- 評価分類
- 映画ジャンル



結論

- これから映画を見る人
- 企業に対して





目的

一般民衆向け：

近年、日々の生活リズムが早まり、余暇を取ることが難しくなっている。

故に、地雷映画を避けるために、最も良い映画の特徴を事前に知る必要があります。

企業向け：

大ヒット映画を作ることが優先事項なので、事前に市場調査を行って、大衆に合わない映画の制作は避けることができます。



データソース



<https://www.themoviedb.org/>

The Movie Database (TMDB) は映画やテレビ番組、俳優、テレビ局、映画や番組会社などのオンラインデータベースです。

現在は39言語に対応し、おおよそ60万以上の映画、10万以上のテレビ番組、そして200万以上の俳優などデータが登録されています。



評価定義①

収集したデータによると、この3つのデータ
（**人気さ**、**平均投票点数**、**投票人数**）は基準
として判断できます。

movie_name	popularity	vote_average	vote_count
The Godfather	77.000	8.7	15394
The Shawshank Redemption	80.831	8.7	20624
दिलवाले दुल्हनिया ले जायेंगे	32.282	8.7	3381
Schindler's List	75.526	8.6	12301
Gabriel's Inferno Part III	30.058	8.6	933
...
The Open House	13.707	3.7	1181
The Human Centipede 3 (Final Sequence)	25.900	3.7	560
Disaster Movie	17.561	3.3	794
Alone in the Dark	17.477	3.2	443
Dragonball Evolution	47.398	2.8	1601



評価定義②

データの正規化:

3つのデータを正規化します。
人気さ、平均投票点数、投票人数。

評価分数(Evaluation_Point)の定義:

$$0.2(\text{人気さ}_{\text{正規化}}) + 0.3(\text{投票人数}_{\text{正規化}}) + 0.5(\text{平均投票点数}_{\text{正規化}})$$



評価定義③

良い映画：

評価分数(Evaluation_Point) > 4.5



普通映画：

4.5 > 評価分数(Evaluation_Point) >= 3

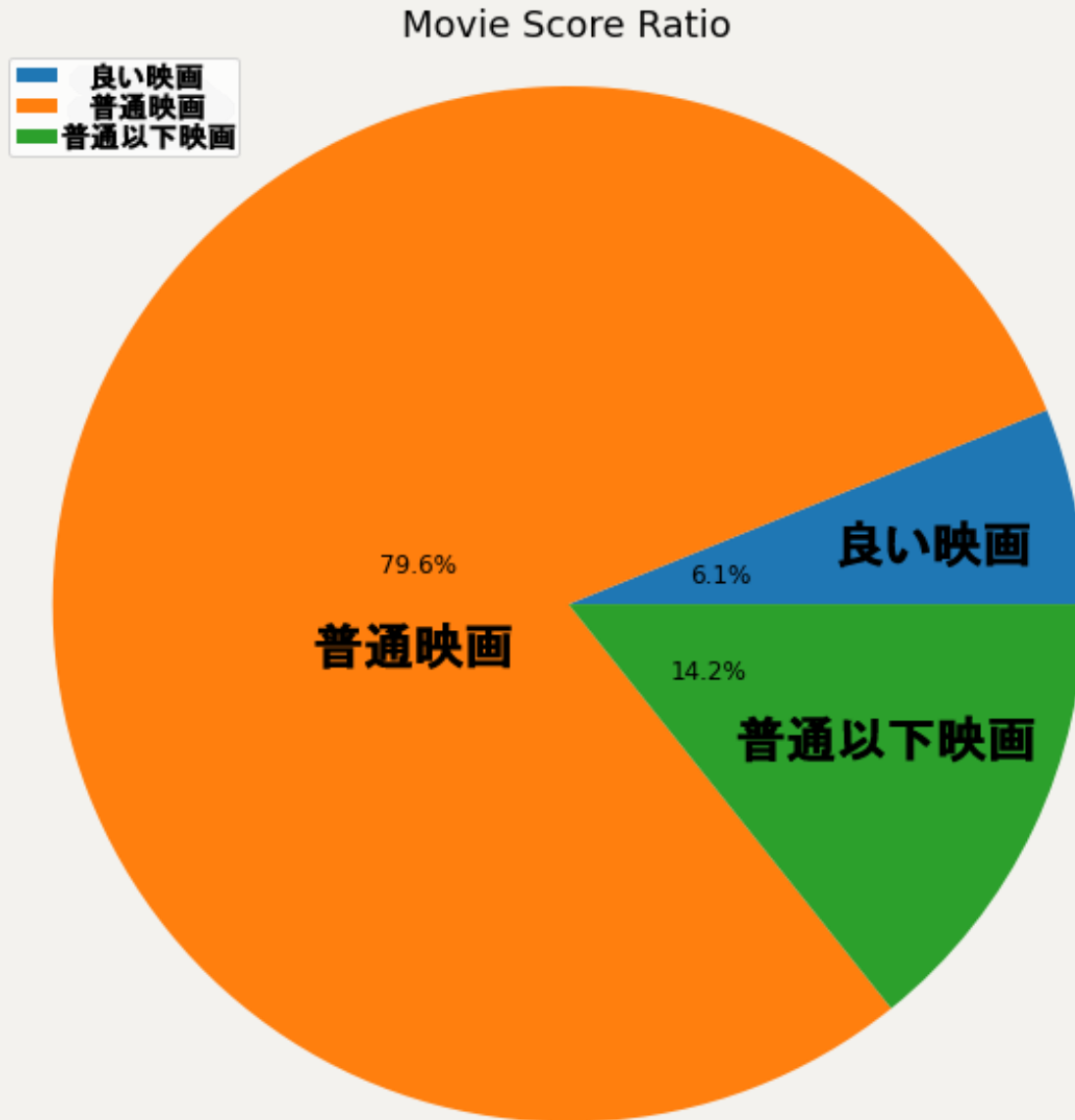


普通以下映画：

評価分数(Evaluation_Point) < 3



評価分類

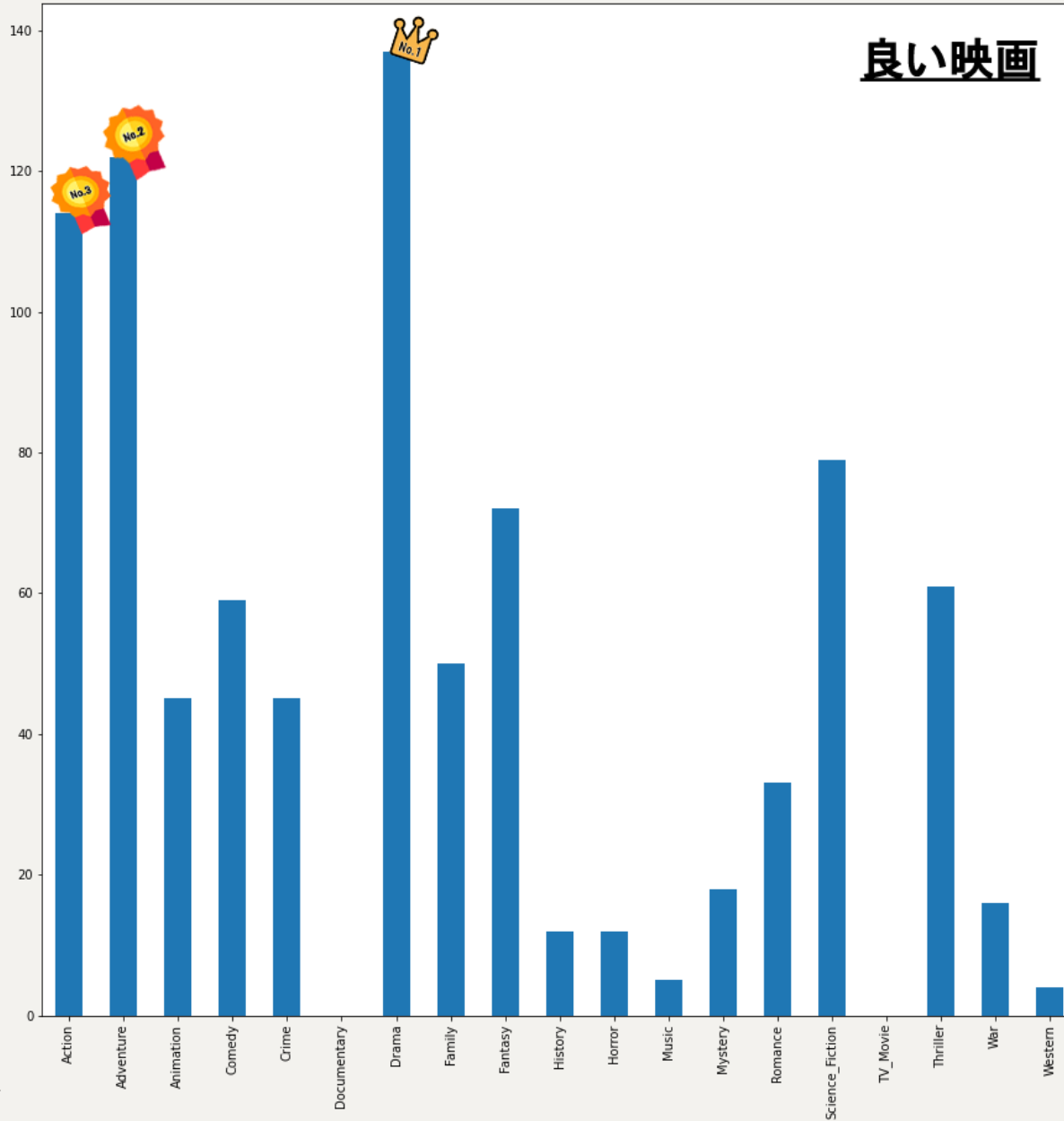


前の定義によると、今回のデータはおおよそ**80%**は**普通の映画**です。

そして**6%**は**良いの映画**、**14%**は**普通以下の映画**です。



映画ジャンル①

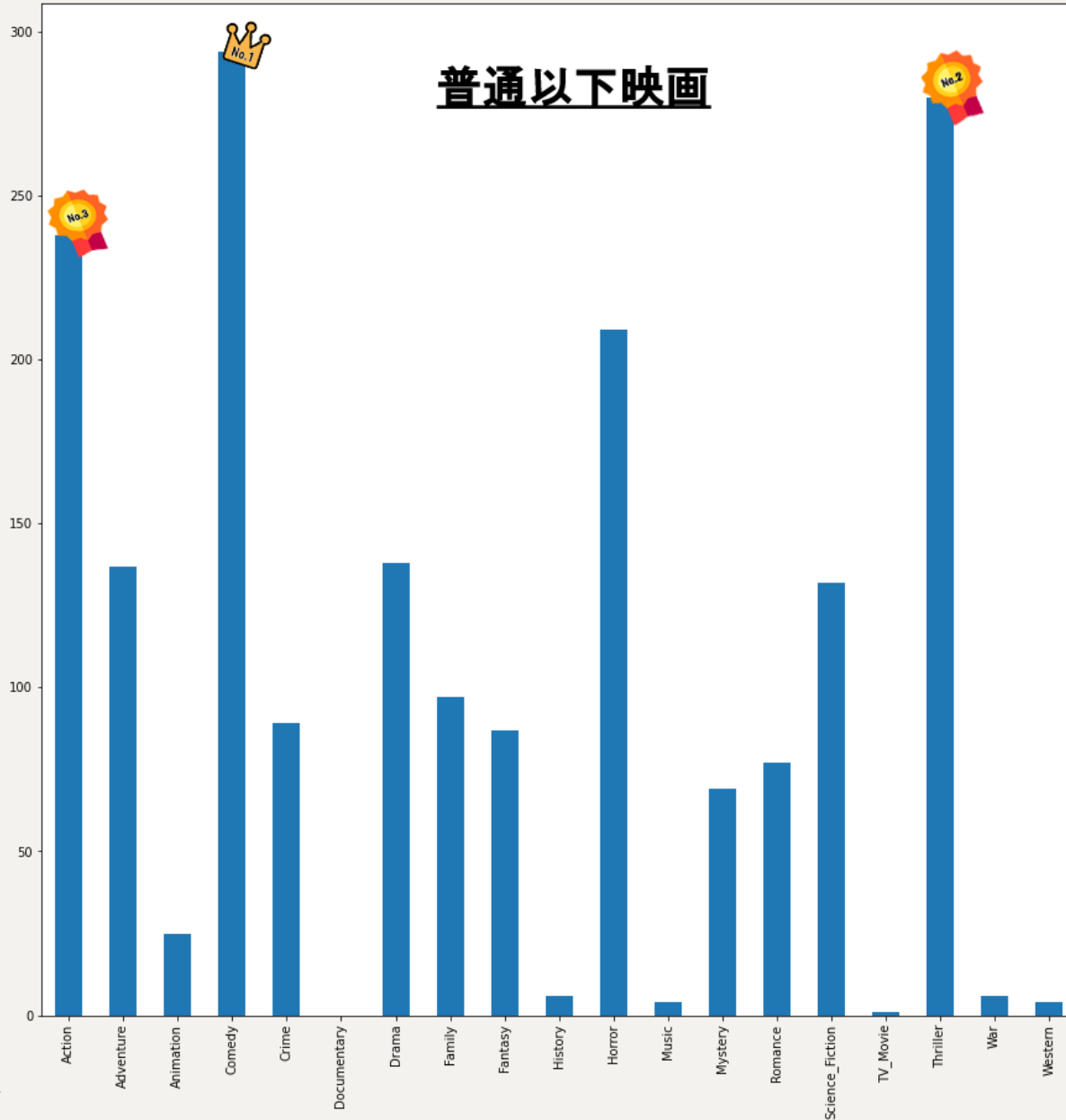


良い映画の統計で、多数人の映画見る条件は**ドラマ映画**です。

次は**冒険**や**アクション**です。

つまり、現在多数の人は**完全なストーリーを持つ映画**（架空か本物かを問わず）に興味を持つと私は考えています。

映画ジャンル②

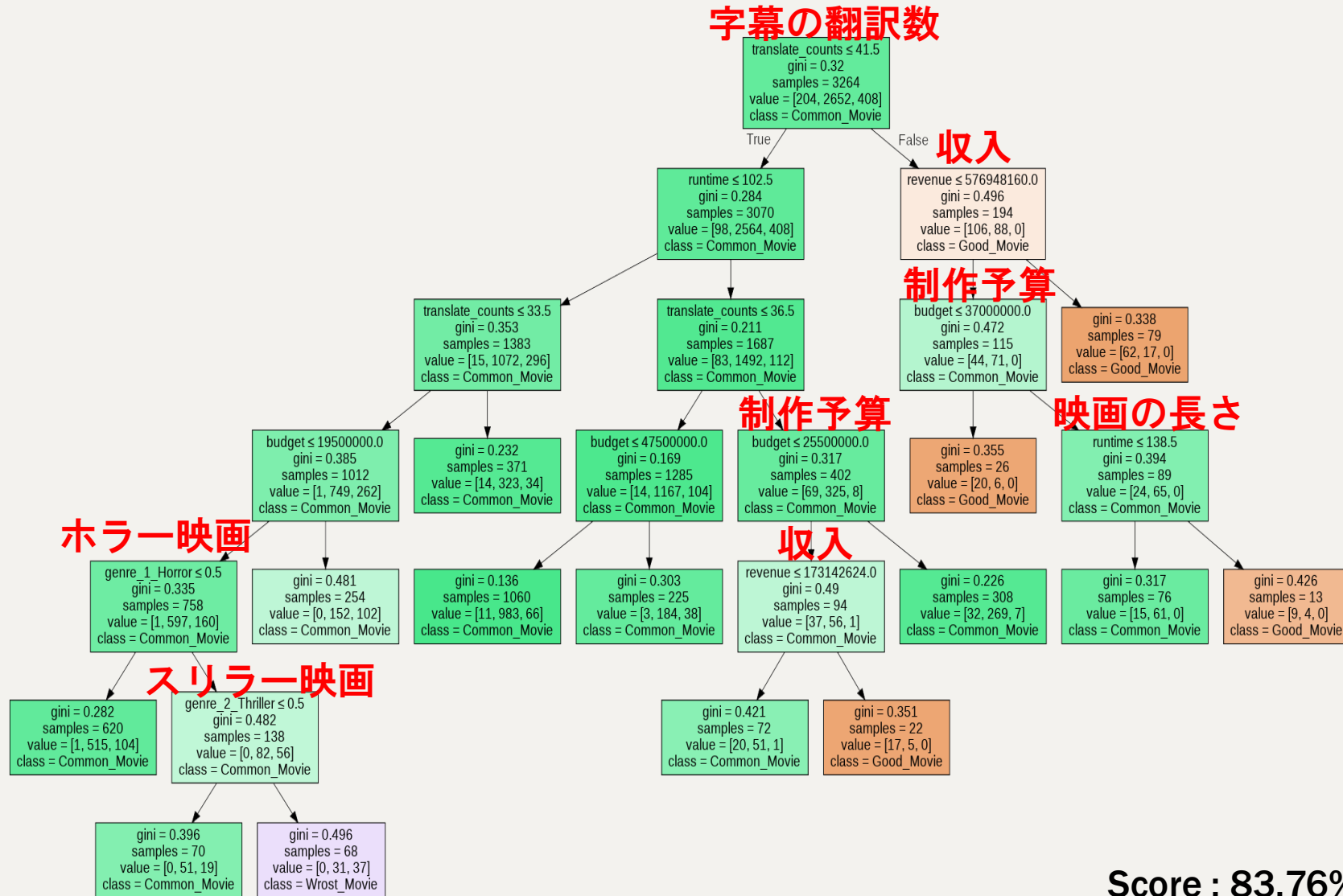


普通以下の映画は実際に人によって異なります。

左の統計を見ると、多数の人はコメディ映画とホラー映画が嫌いです。
なぜなら、コメディでは人によって笑うポイントが違い、スリラー映画では血や暴力的なシーンがよく出るからです。



決定木



Score : 83.76%

決定木分析によって大きく
良い映画になる条件：

1. 字幕翻訳 > 42言語
2. 収入 > 制作予算
3. 映画の長さ > 2時間

多数の普通以下映画になる
条件はスリラー映画と
ホラー映画



他の分析方法

ランダムフォレスト(Random forest) :

ランダムサンプリングされたトレーニングデータによって**学習した多数の決定木**を使用する。

Score : 81.25%

K近傍法(KNN) :

特徴空間における**最も近い訓練に基づいた分類の手法**です。

Score : 82.08%

KMeans :

クラスタリングのアルゴリズム。クラスタの平均を用い、与えられたクラスタ数**K個**に分類する。

Score : 81.32%





これから映画を見る人

映画を見るかどうかを選ぶときはいくつかの事前準備：

1. 映画の概要に興味があるかどうかを確認してください。
2. 映画の予告編が映画の概要と異なるかどうかを確認してください。
3. 対応する字幕翻訳があるかどうかを確認してください。
4. ホラー映画やスリラー映画は見ないほうがいい。もちろん、興味があれば見てください。



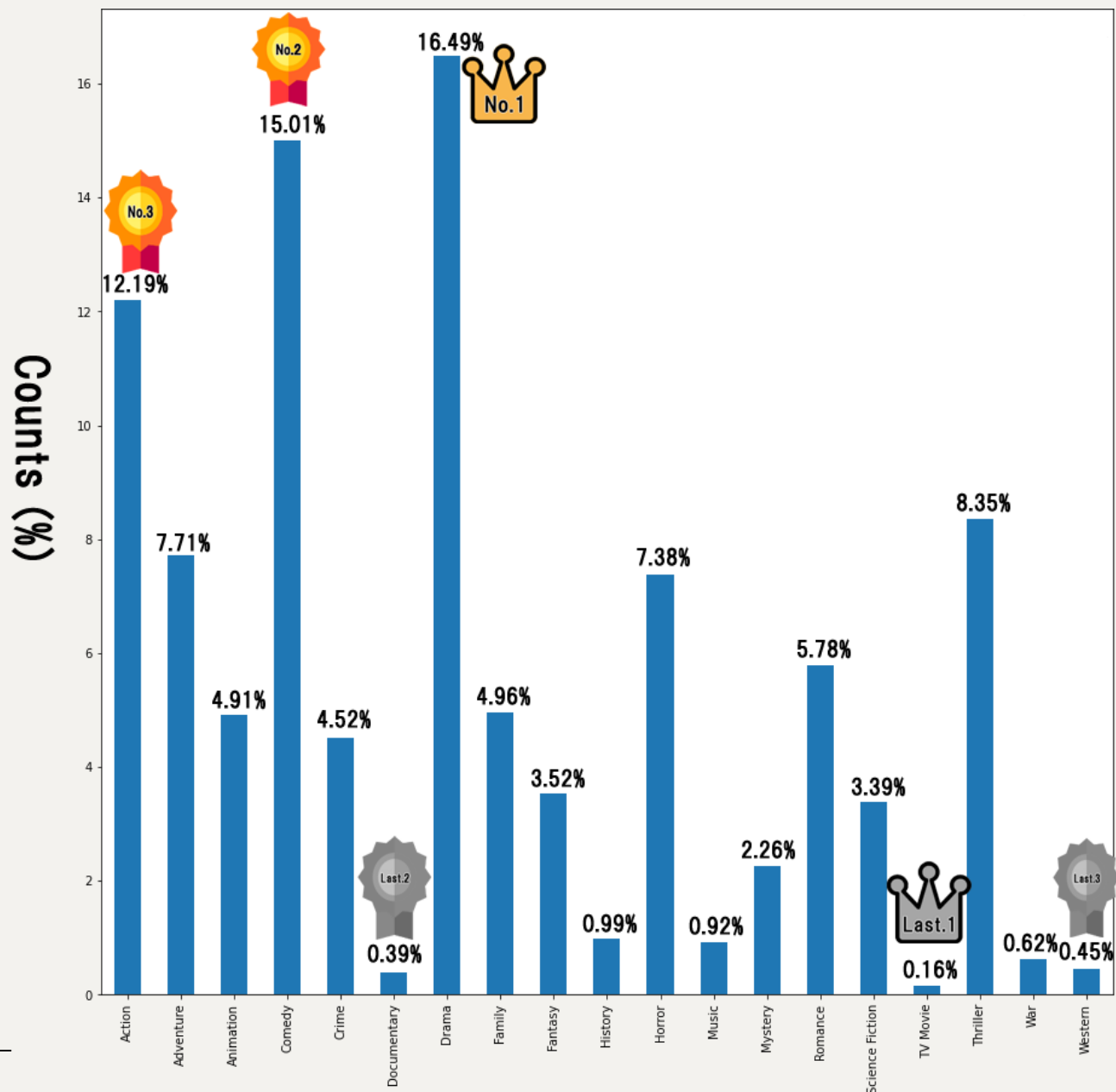
企業に対して①

映画が売れるかどうかは企業にとって大事なことです
から、収益性の高い映画を作るために、**映画のテーマ、
ジャンルの選択**は非常に重要です。

次は**投資利益率 (ROI) が高い**映画ジャンルの統計です。



企業に対して②



- 最も重要なことは**完全なストーリー**といくつか**コメディ**および**アクション**要素を持つ映画が売れるです。
- あまりに**ドキュメンタリー**や**西部劇**要素の制作は避けたほうがいいと思います。

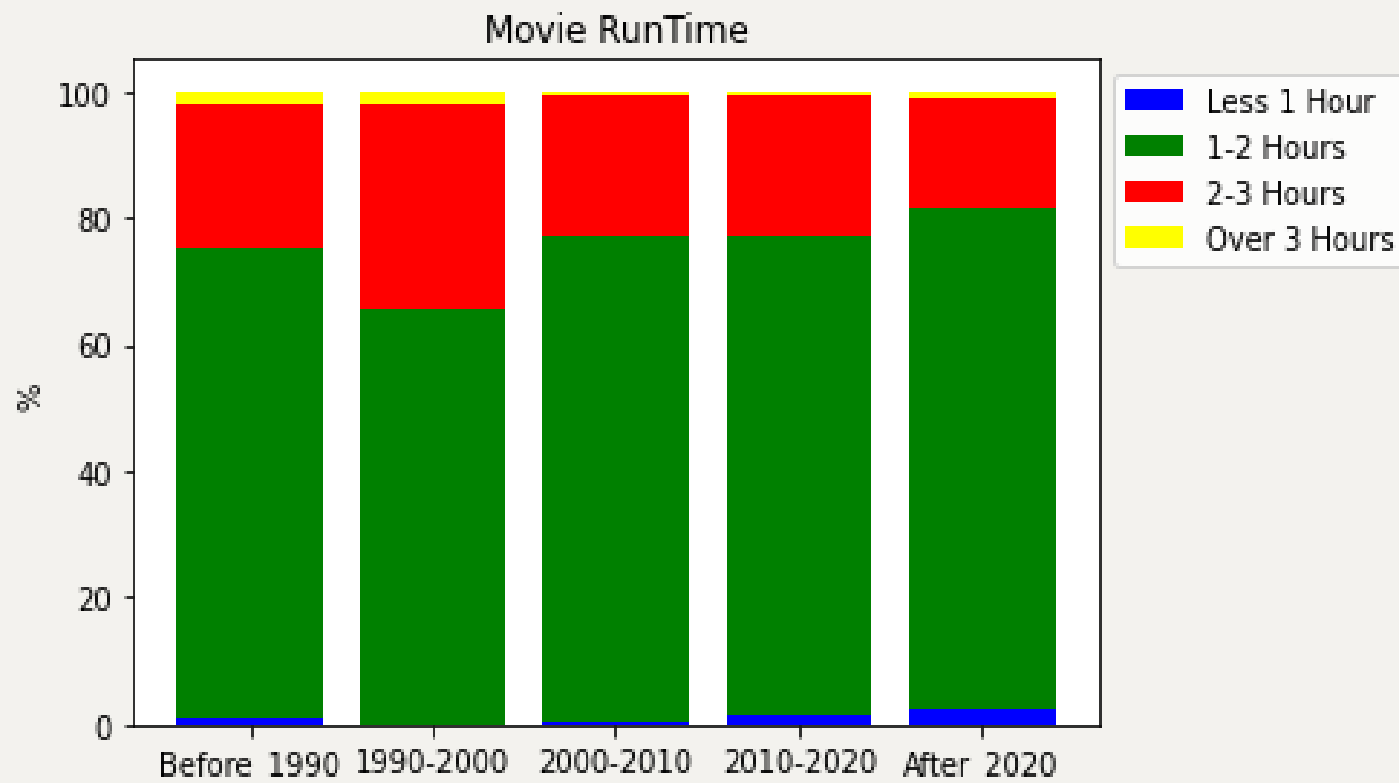
企業に対して③

また、人々の生活ペースが加速しているため、余暇をあまり取れず、**映画の長さ**も映画の視聴条件を選択する際に重要なポイントです。

次は近年、**投資利益率(ROI)が高い映画長さ**の統計です。



企業に対して④



- 映画の長さは**1～2時間**が一番、次は**2～3時間**です。
- **ほぼ3時間および3時間以上**の映画制作は避けた方がいいと思います。
- **1時間未満の映画**も増える傾向もあるので、映画制作の際にも考慮できると思います。

