

Análise de dados: Crime & War Movies

O objetivo deste relatório é a exploração de dados sobre filmes de Crime & War, afim de identificar padrões financeiros, de popularidade de avaliações, e gerar insights estratégicos sobre suas rentabilidades. As análises buscaram responder questões como **quais combinações de subgêneros são mais lucrativas, como o orçamento impacta o retorno de resultado do filme, qual o ciclo adequado para franquias antes de uma eventual saturação do público**. As conclusões miram em auxiliar um executivo da industria cinematográfica a tomar decisões estratégicas de sua próxima produção, **transformando dados em recomendações acionáveis**.

Índice

1. Introdução:

Histórico do pipeline e análise de consistência estatística

2. Análises:

2.1 *Finanças*

2.2 *Avaliações e Popularidade*

2.3 *Tendências Temporais e Espaciais*

2.4 *Franquias x Filmes Únicos*

3. Encerramento:

Conclusões e Recomendações acionáveis

Pipeline dos Dados

RAW ZONE

- **Característica:** armazenar dados em estado bruto
- **Formato de objetos e fontes:** arquivos .csv e .json, e dados armazenados separados por fonte
- **Quantitativos:**
 - 'csv' IMDB = 244.543 registros únicos
 - 'json' TMDb = 44.641 registros únicos
- **Observações:**
 - 'csv' fornecido com todos os gêneros de filmes disponíveis
 - 'json' extraído (via API) já com filtro para manter apenas filmes do gênero Crime e/ou Guerra

TRUSTED ZONE

- **Característica:** armazenar tratados, desde a RawZone
- **Formato de objetos e fontes:** arquivos 'parquet' e dados armazenados ainda separados por fonte
- **Quantitativos:**
 - 'parquet' IMDB = 29.128 registros únicos
 - 'parquet' TMDb = 44.641 registros
- **Observações:**
 - dados do IMDB foram filtrados para gêneros - manter apenas registros que contém Crime e/ou Guerra
 - dados TMDb tratados sem mais filtros

REFINED ZONE

- **Característica:** armazenar dados tratados e modelados (multi-dimensionalmente), desde a *TrustedZone*
- **Formato de objetos e fontes:** arquivos 'parquet' e fontes combinadas, dados do 'JOIN' realizado
- **Quantitativos:**
 - 'parquet' IMDB e TMDB (combinados) = **1.528 registros únicos**
- **Observações:**
 - IMDB têm os dados base, e dados TMDB cumprem papel de enriquecimento
 - filtros de '*NOT NULL*' em campos primordiais às análises

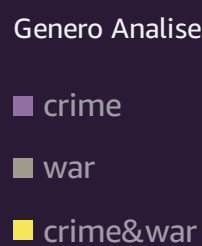
*Fontes dos dados: 1) IMDB - Internet Movie Database (<https://www.imdb.com>) | 2) TMDB - The Movie Database (<https://www.themoviedb.org/>)

Verificação de consistência da amostra (estatística descritiva básica)

'Crime&War' são filmes que contém os dois gêneros, simultaneamente. A amostra contém apenas 2 filmes nessa condição - e será descartada.

'Crime&War' são filmes que contém os dois gêneros, simultaneamente. A amostra contém apenas 2 filmes nessa condição - e será **descartada**.

A amostra contém apenas 2 filmes nessa condição - e será descartada.



A amostra é desbalanceada entre Crime e War, requerendo uso de medianas para análises equiparadas.

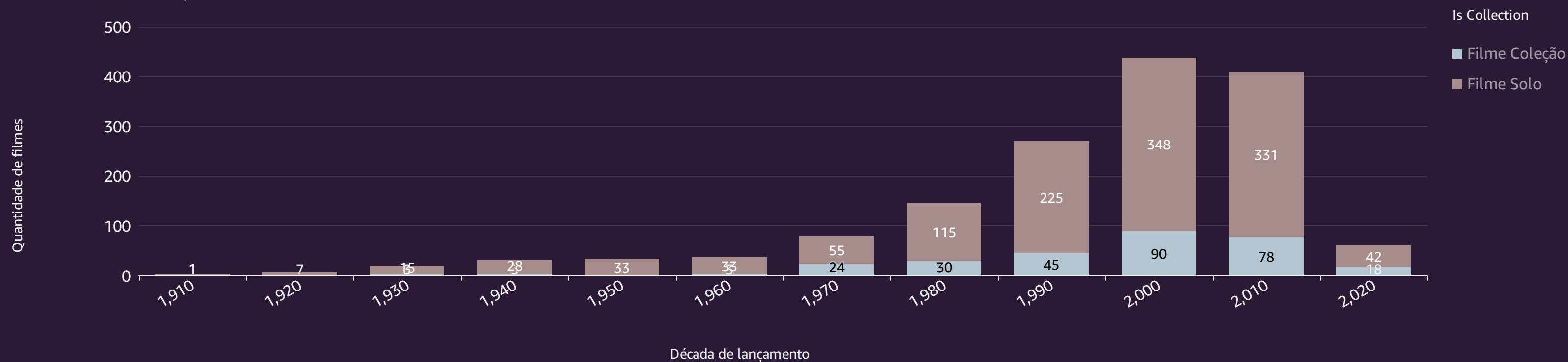
Os gêneros associados apresentam forte relação com Crime em categorias como Action, Drama e Thriller. War possui maior presença em History e War x Sci-F

Os gêneros associados apresentam forte relação com Crime em categorias como Action, Drama e Thriller. War possui maior presença em History e War x Sci-F



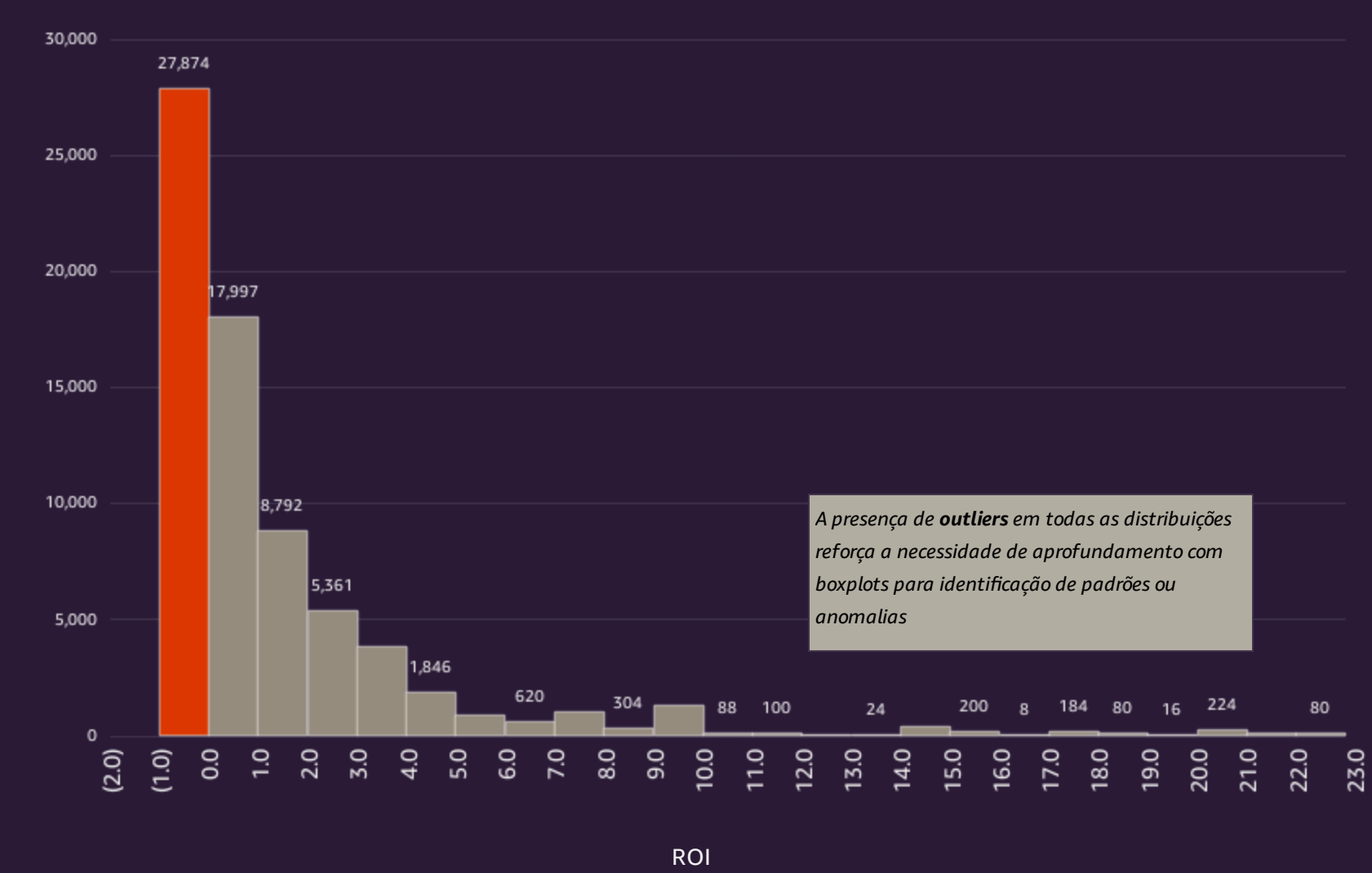
Entre 1980 e 2010, a proporção de filmes franquias e solos mantém consistência (cerca de 16%). O declínio após 2010 reflete dados limitados até 2022 + Pandemia (Covid-19).

Entre 1980 e 2010, a proporção de filmes franquias e solos mantém consistência (cerca de 16%). O declínio após 2010 reflete dados limitados até 2022 + Pandemia (Covid-19).



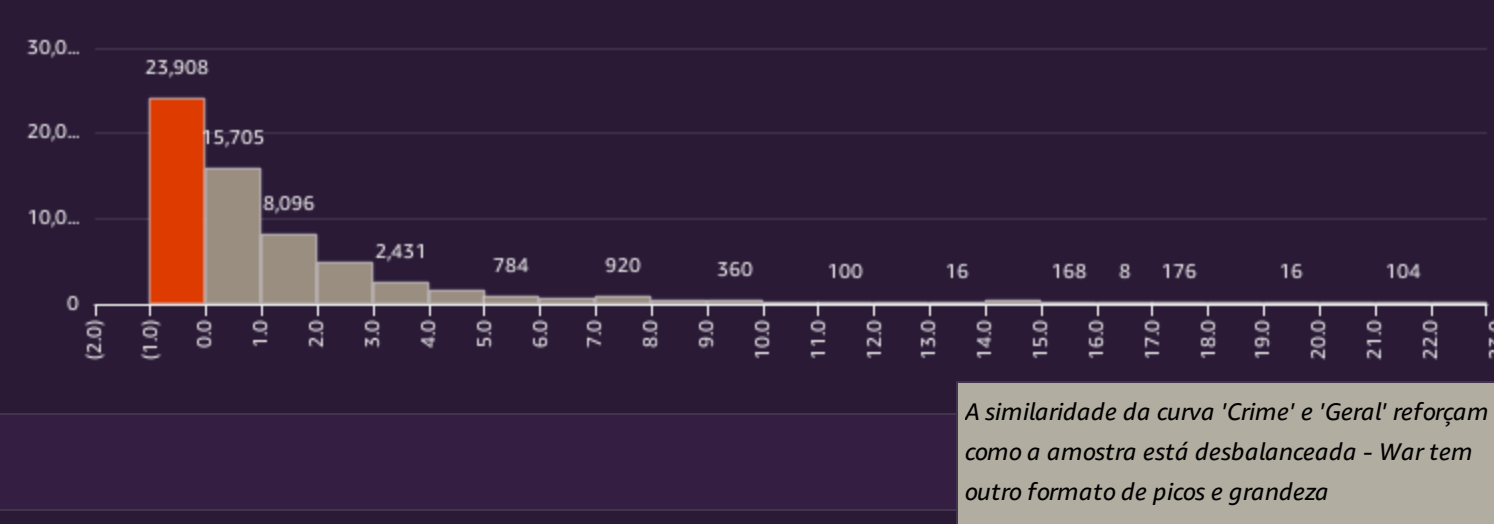
A maior concentração de ROI está entre -2 e 2, com picos de exceções na faixa negativa (-2) e positiva (1). Essa distribuição reflete um cenário balanceado, porém com outliers expressivos em valores extremos.

A maior concentração de ROI está entre -2 e 2, com picos de exceções na faixa negativa (-2) e positiva (1). Essa distribuição reflete um cenário balanceado, porém com outliers expressivos em valores extremos.



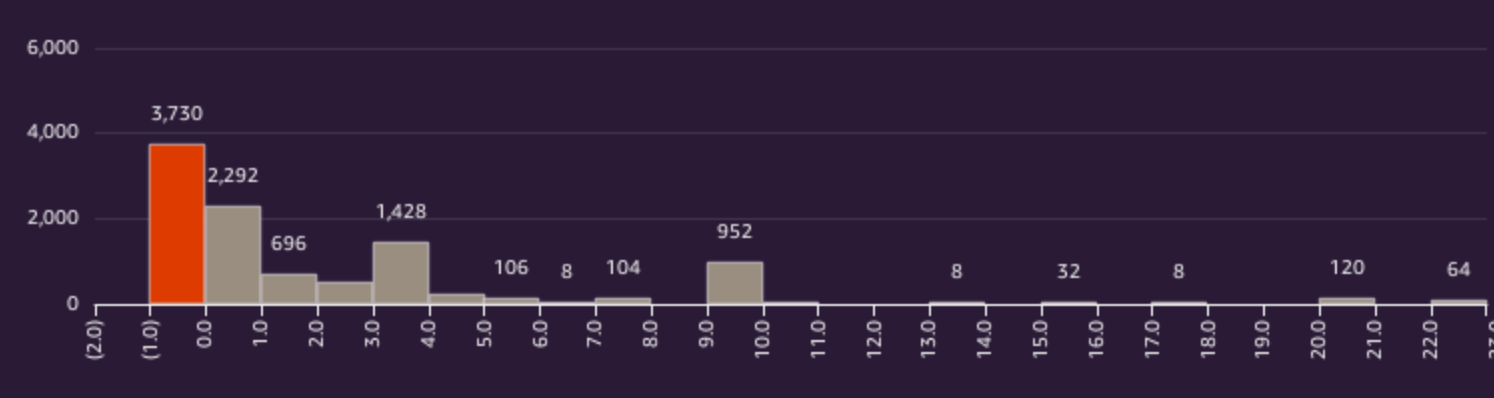
O ROI do gênero Crime segue o padrão geral, com maior concentração entre -2 e 2 e picos destacados em -2 e 1, indicando maior variabilidade

O ROI do gênero Crime segue o padrão geral, com maior concentração entre -2 e 2 e picos destacados em -2 e 1, indicando maior variabilidade



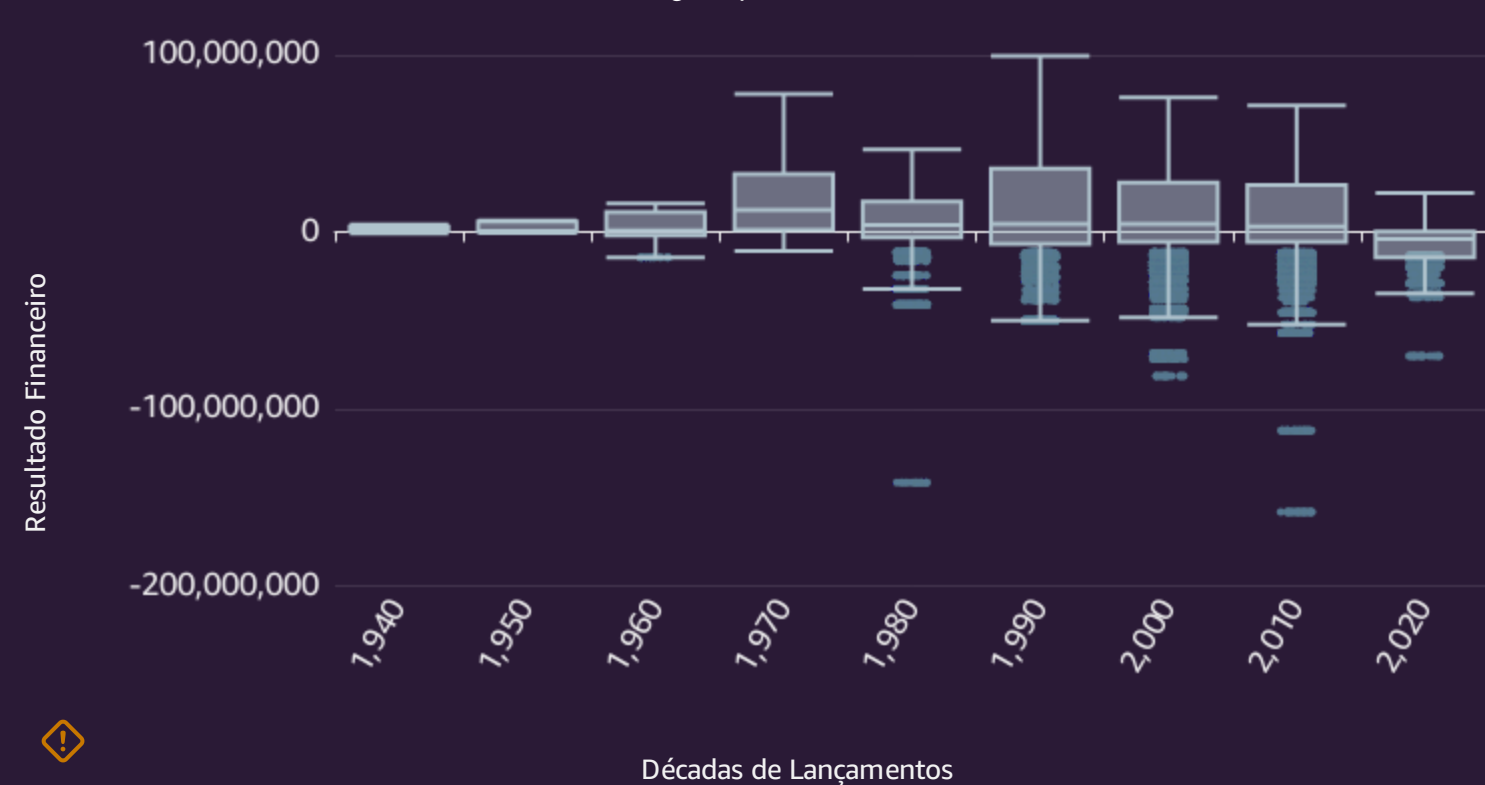
A distribuição do gênero *War* concentra-se entre -1 e 1, com picos isolados que sugerem necessidade de análise mais detalhada

A distribuição do gênero *War* concentra-se entre -1 e 1, com picos isolados que sugerem necessidade de análise mais detalhada



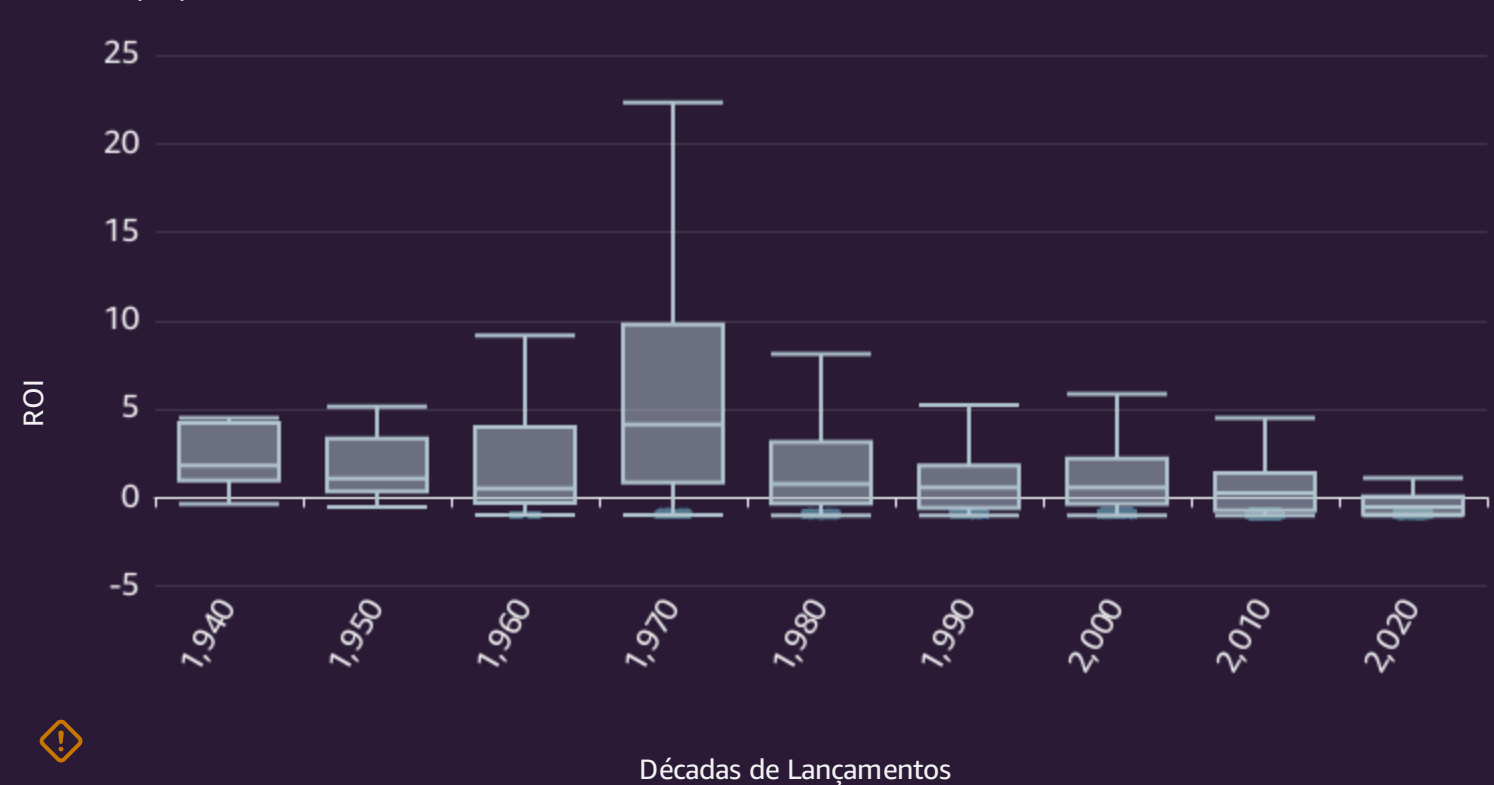
O resultado financeiro mostra consistência geral ao longo das décadas, com outliers inferiores significativos surgindo desde os anos 1970, destacando variabilidade em alguns períodos

O resultado financeiro mostra consistência geral ao longo das décadas, com outliers inferiores significativos surgindo desde os anos 1970, destacando variabilidade em alguns períodos



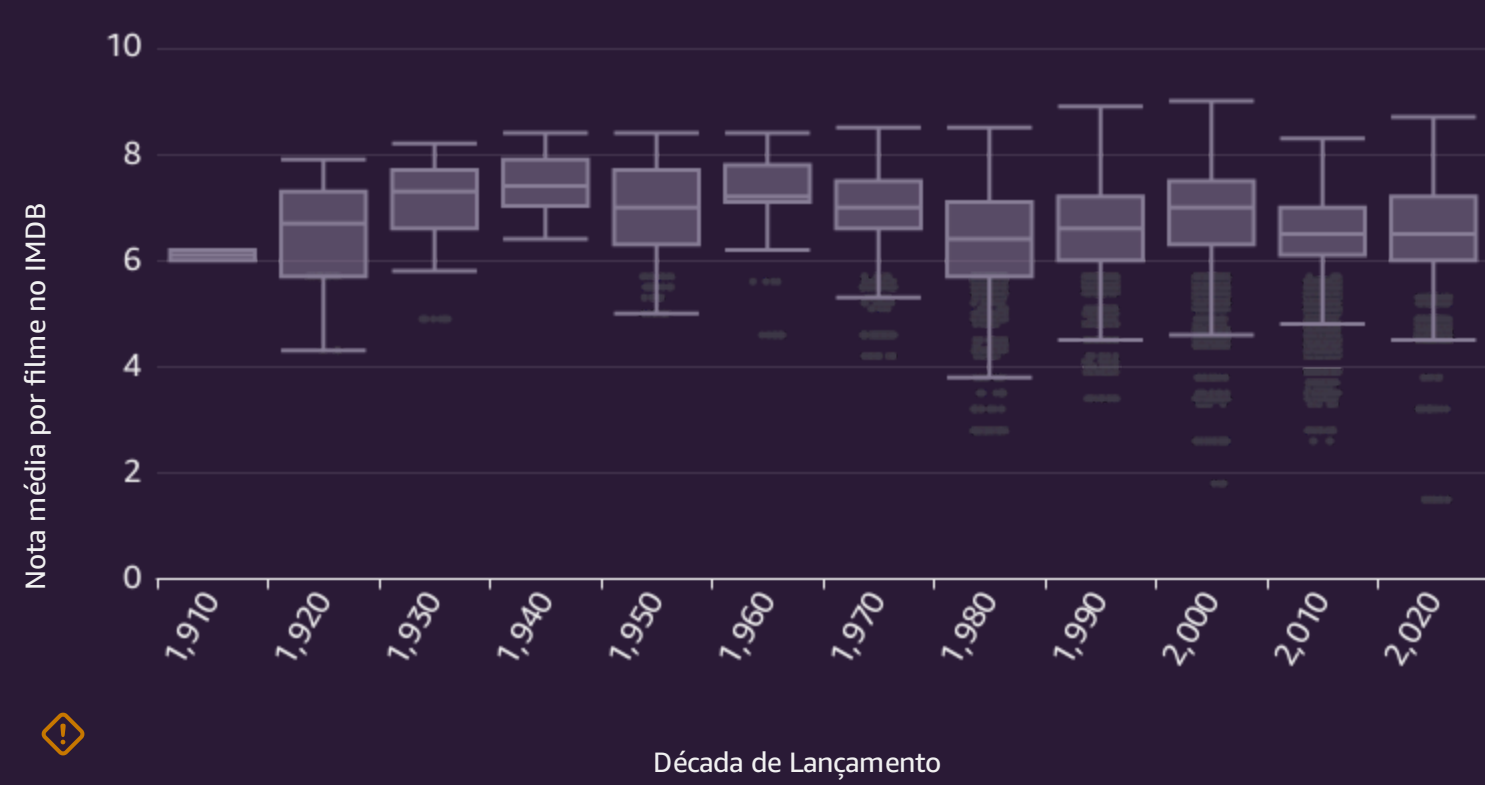
Apesar da estabilidade geral no ROI, a década de 1970 apresenta um fenômeno isolado que eleva o valor máximo de forma desproporcional

Apesar da estabilidade geral no ROI, a década de 1970 apresenta um fenômeno isolado que eleva o valor máximo de forma desproporcional



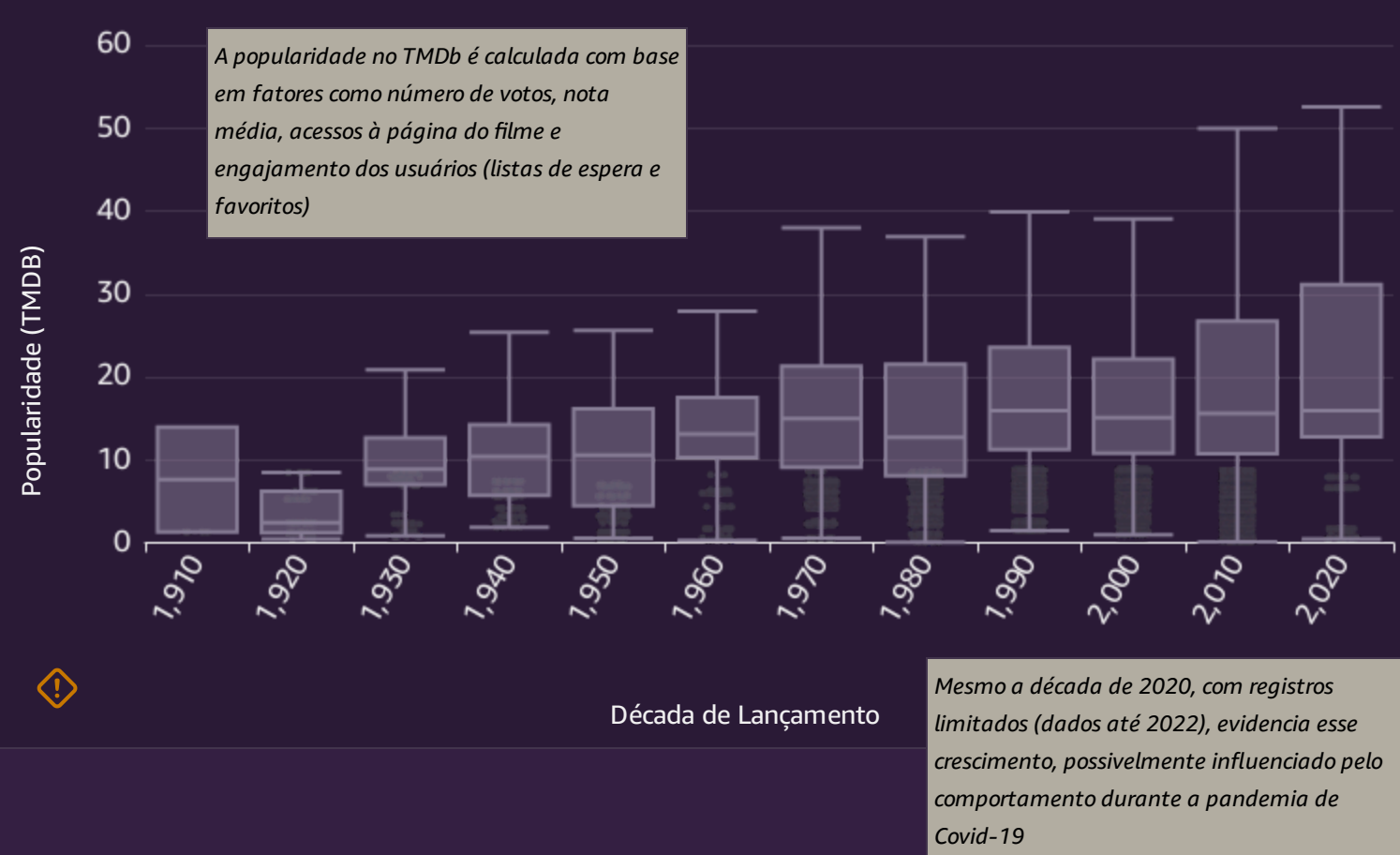
As notas médias dos filmes no IMDB apresentam estabilidade ao longo das décadas, com aumento de outliers inferiores a partir dos anos 1980

As notas médias dos filmes no IMDB apresentam estabilidade ao longo das décadas, com aumento de outliers inferiores a partir dos anos 1980



A popularidade dos filmes mostra crescimento consistente, com ampliação na dispersão e outliers inferiores em décadas recentes

A popularidade dos filmes mostra crescimento consistente, com ampliação na dispersão e outliers inferiores em décadas recentes



Conclusões sobre a Consistência Estatística + Tratamentos Complementares Aplicados

- Os dados da *RefinedZone*, como estão, podem gerar interpretações enviesadas e conclusões imprecisas. Há uma necessidade evidente de tratamentos adicionais para assegurar maior robustez analítica.

- Os dados da *RefinedZone*, como estão, podem gerar interpretações enviesadas e conclusões imprecisas. Há uma necessidade evidente de tratamentos adicionais para assegurar maior robustez analítica.
- Massas de outliers foram identificadas, e sua análise separada, por meio de clustering ou outros métodos, poderia fornecer insights úteis. Entretanto, essas métricas devem ser excluídas das análises centrais para evitar distorções.
- Para cumprir nosso objetivo de gerar análises históricas consistentes que forneçam padrões e insights estratégicos, optaremos por descartar outliers, reduzindo os riscos associados às métricas financeiras e de avaliação do público.

- **Filtro Temporal:** focaremos em produções lançadas entre 1980 e 2020, evitando picos ou vales (extremos) de dados, e que não reflitam o contexto cultural e histórico atual.

- **Aplicação de Filtros Estatísticos:** adotaremos apenas os dados entre os quartis Q1 e Q3 dos boxplots das métricas de Média de Notas, ROI e Resultado Financeiro, assegurando uma base mais estável para análise, desprezando outliers (especialmente os inferiores que não interessam como objetivo da análise).

Nota: As análises a seguir consideram os dados após a aplicação do tratamento complementar à *Refined Zone*.

Análise: Finanças

- Quais combinações de gênero, geram mais receita?
- Como o orçamento impacta o resultado financeiro?

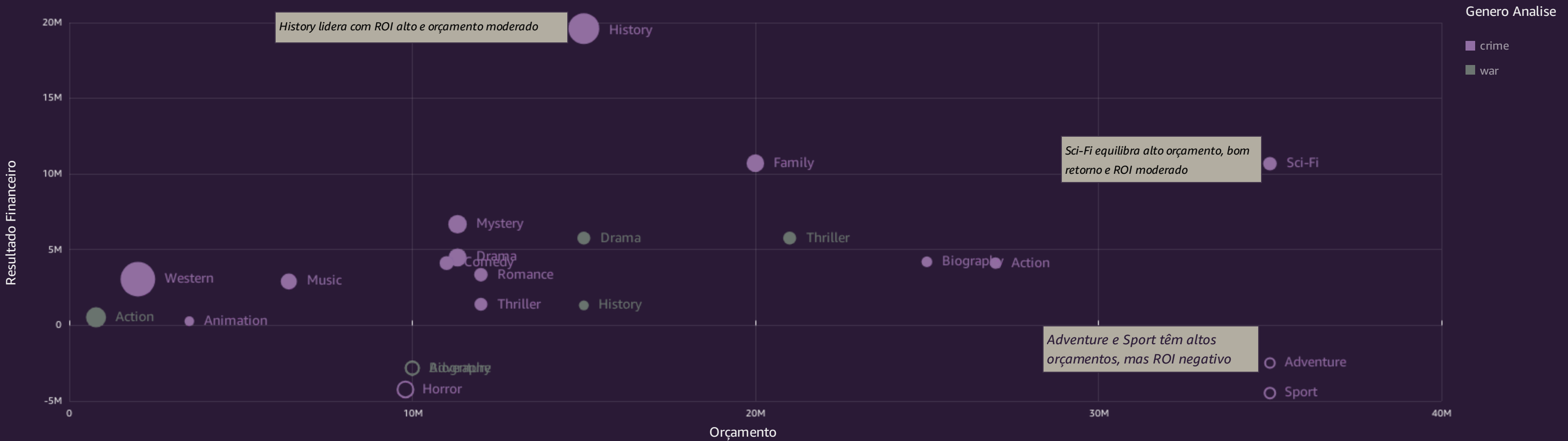
Receitas medianas: gêneros associados a Crime e Guerra

Gêneros associados como Crime + Sci-Fi lideram arrecadação, enquanto War + Thriller apresentam relevância moderada.



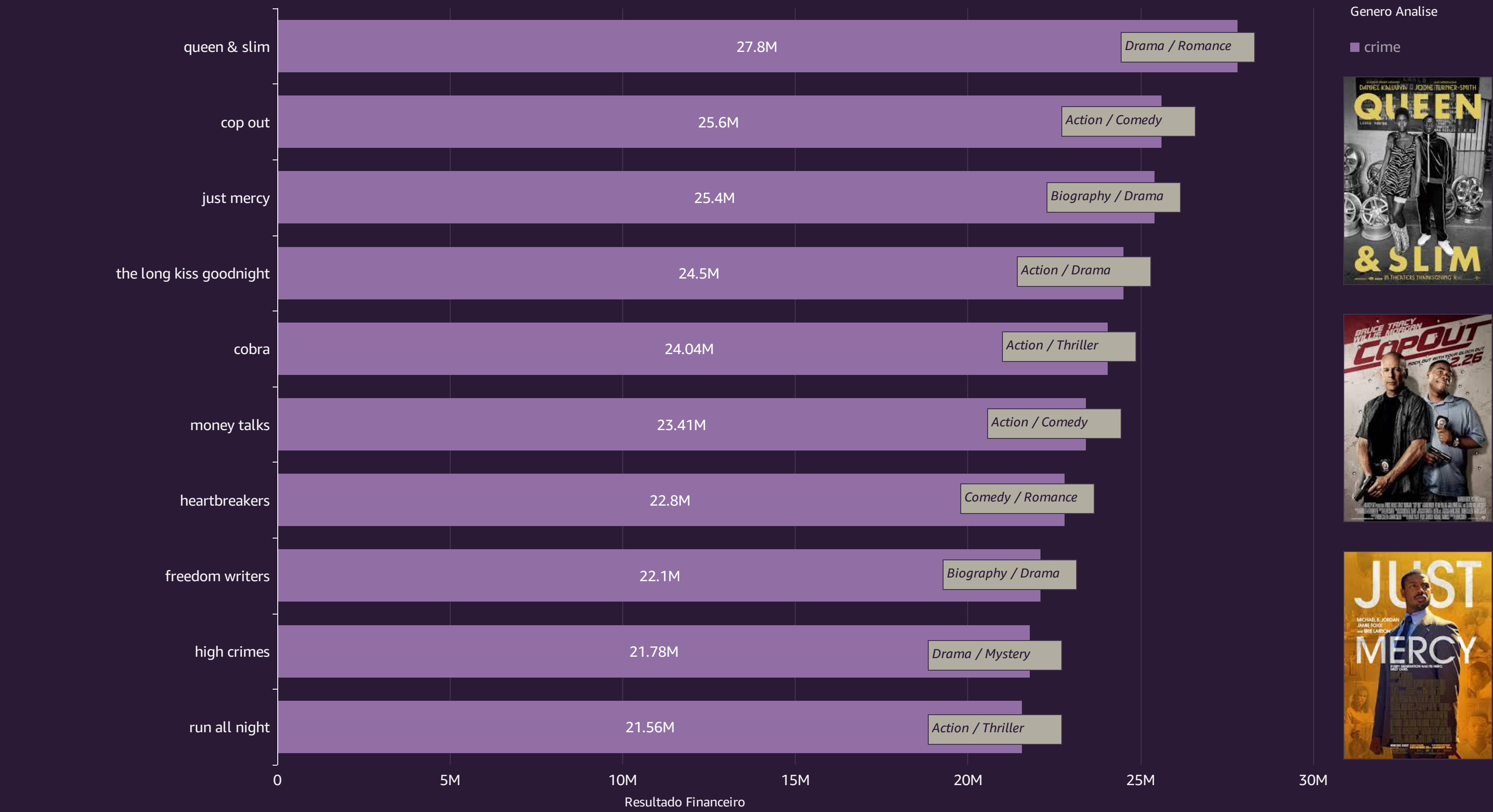
Dispersão entre Orçamento e Resultado Financeiro: padrões e desvios de ROI

Hipótese: filmes que investem mais no orçamento retornam mais lucro? Nem sempre, como mostram History, Sci-Fi, Adventure e Sport.



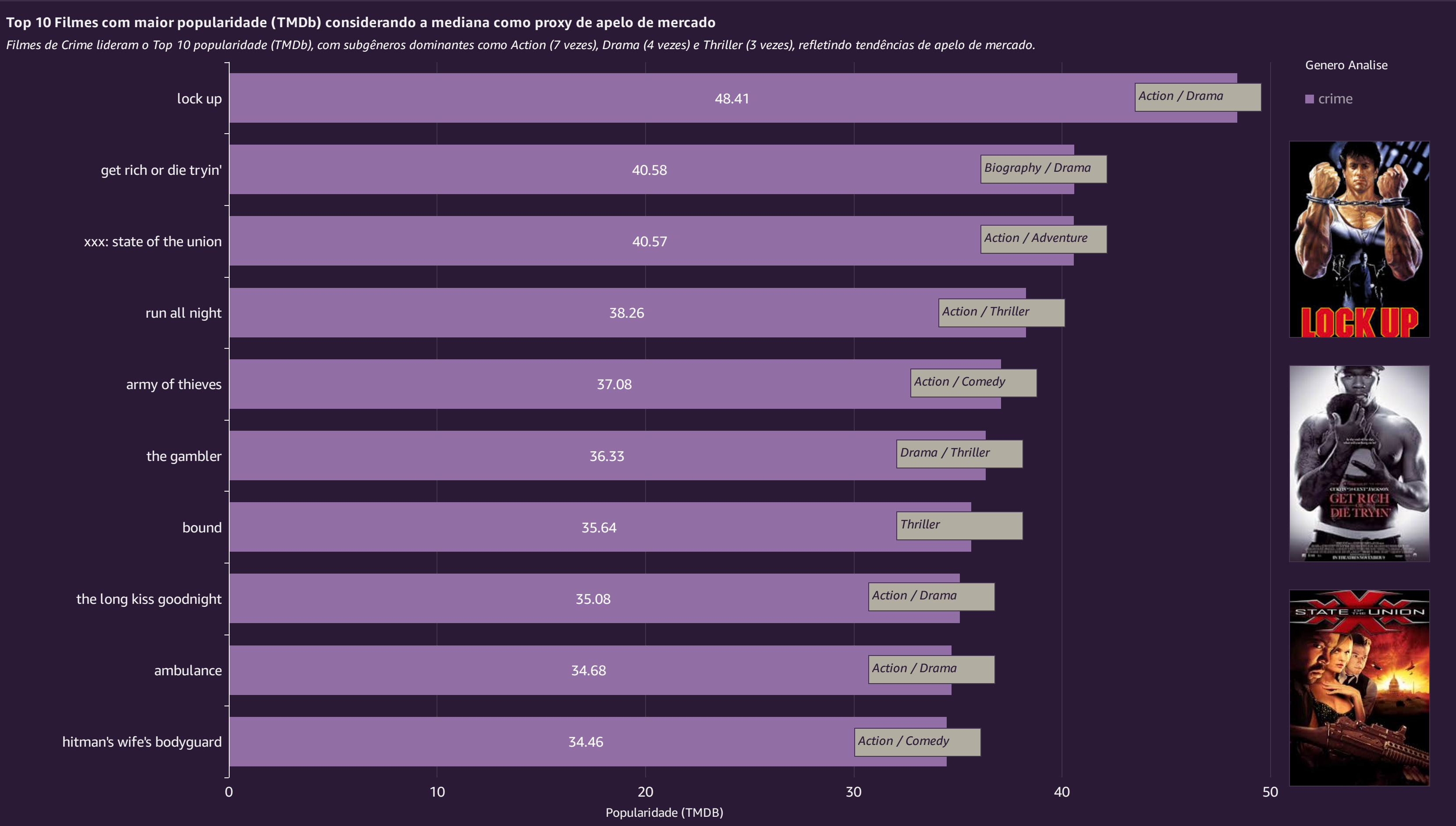
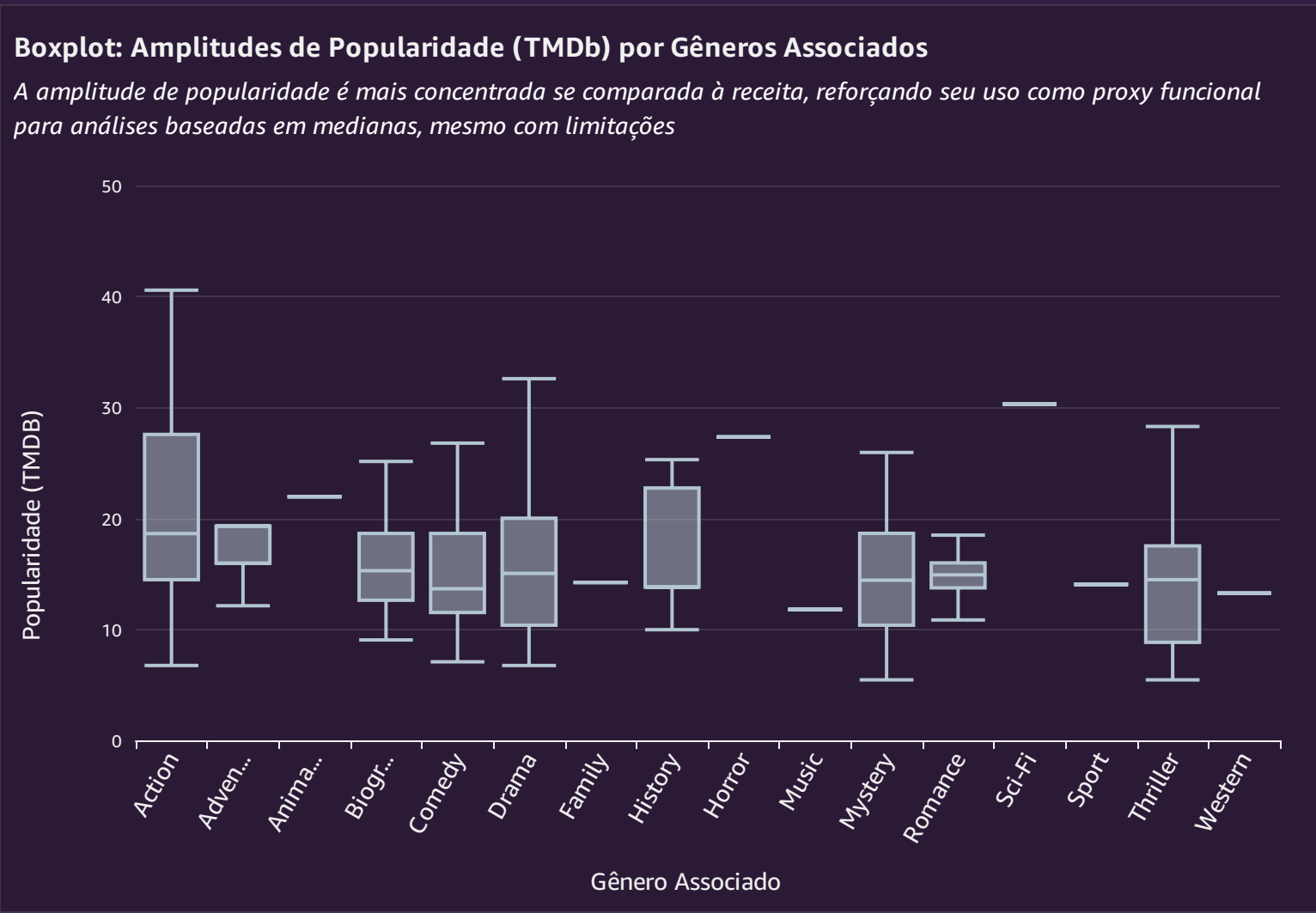
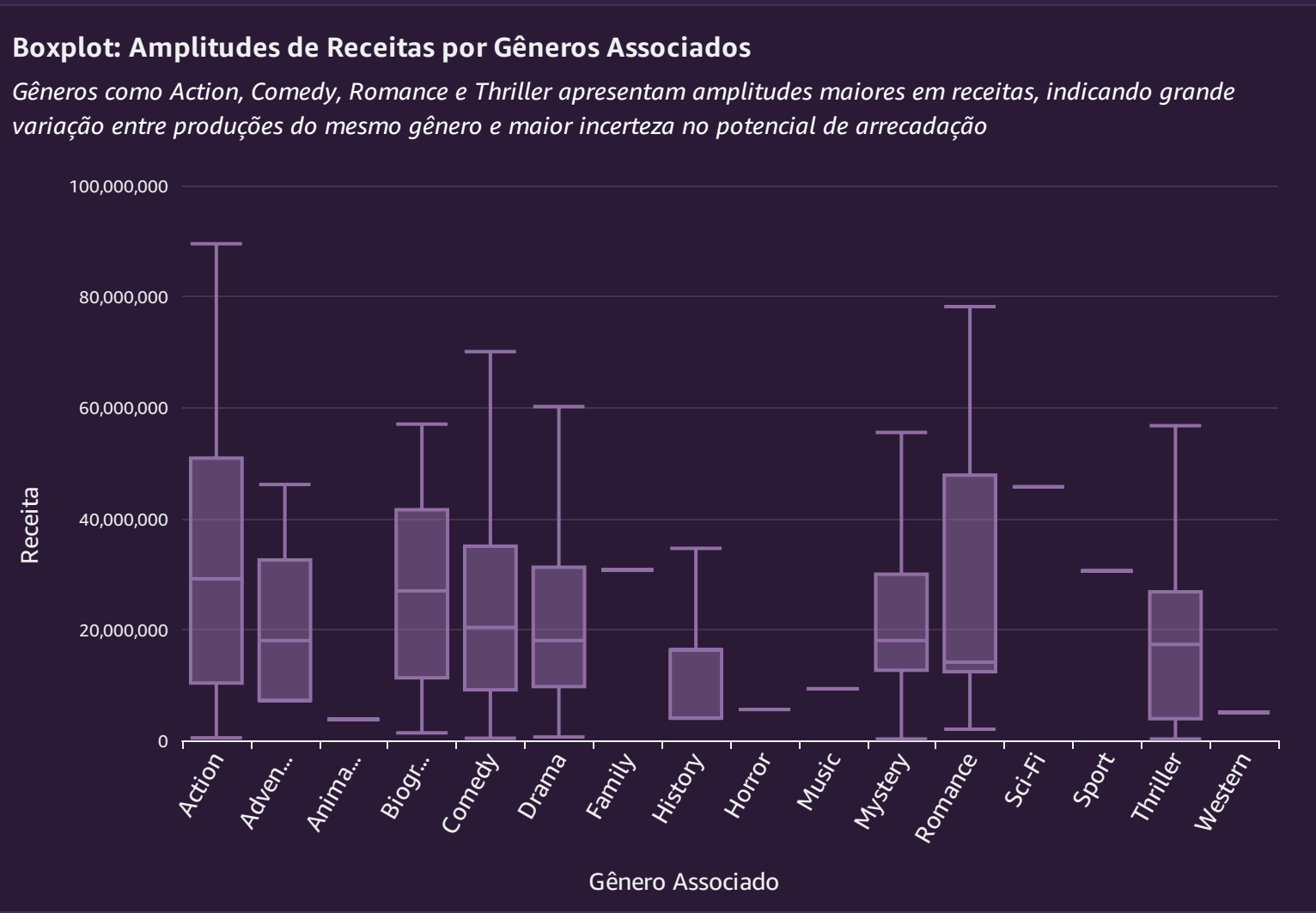
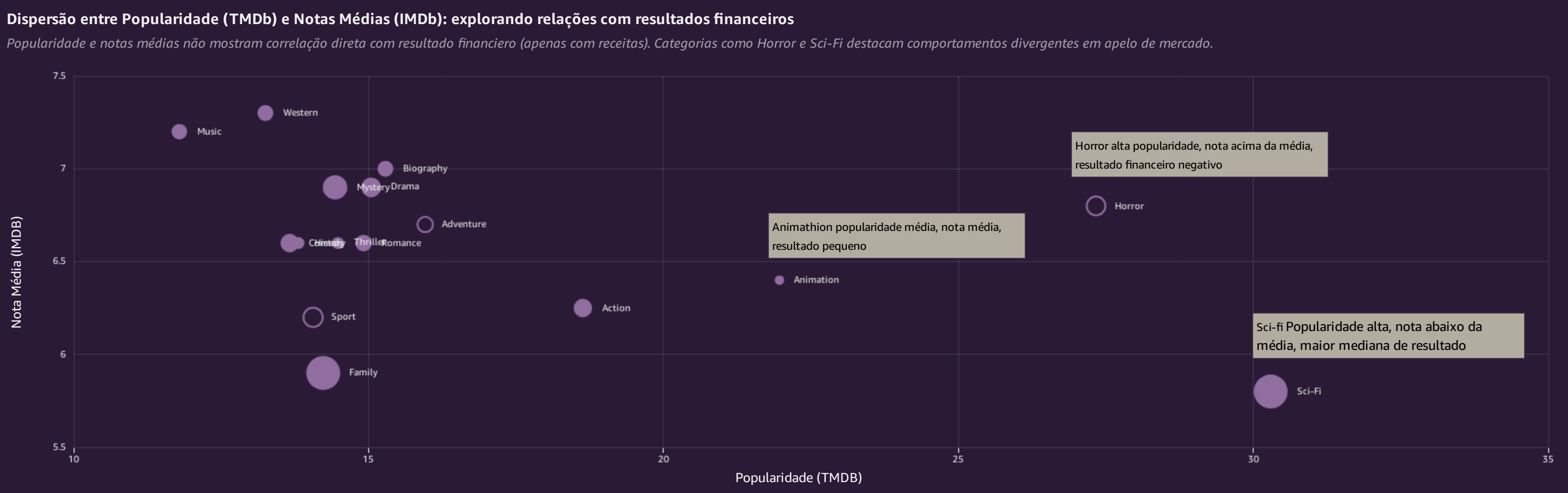
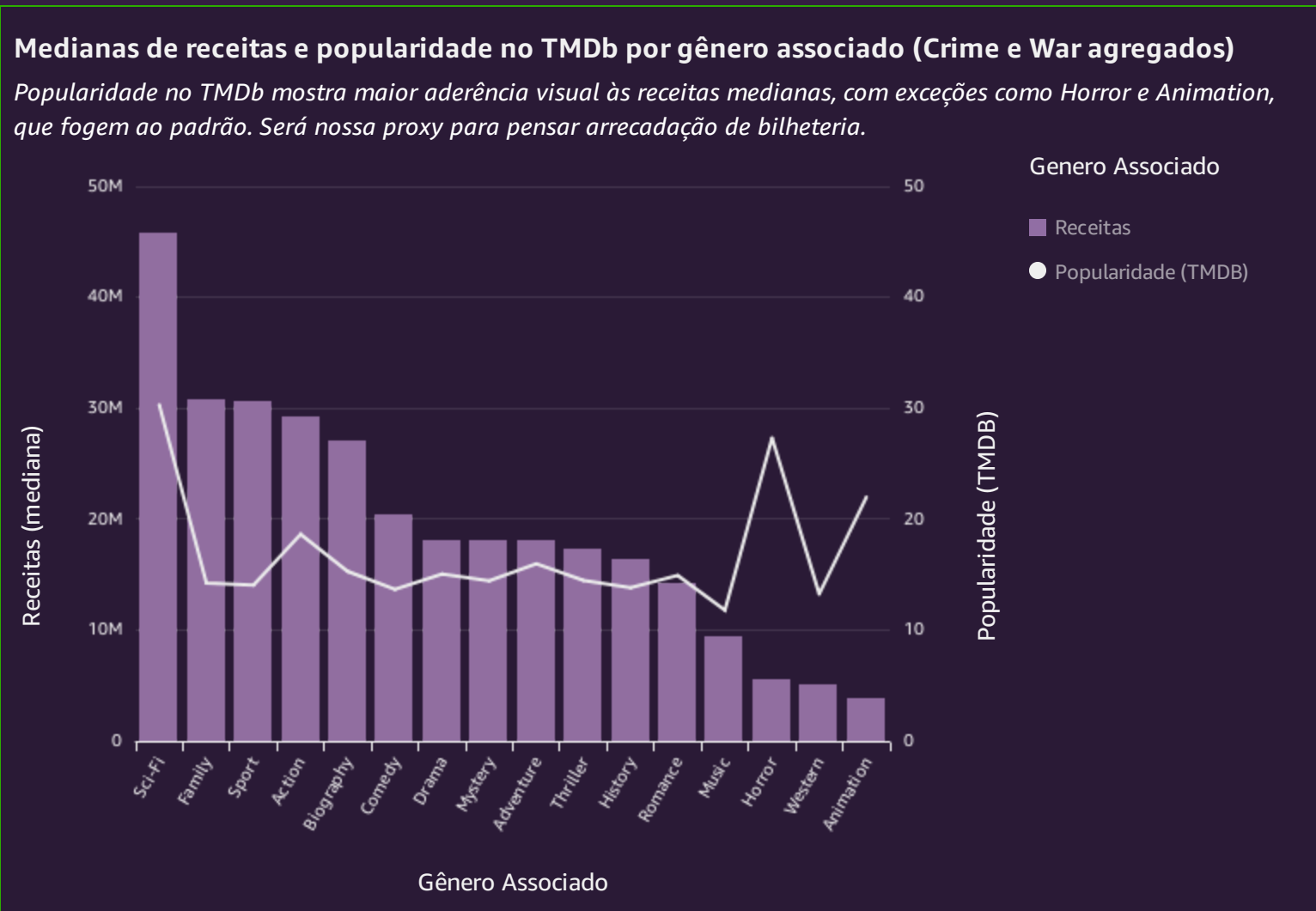
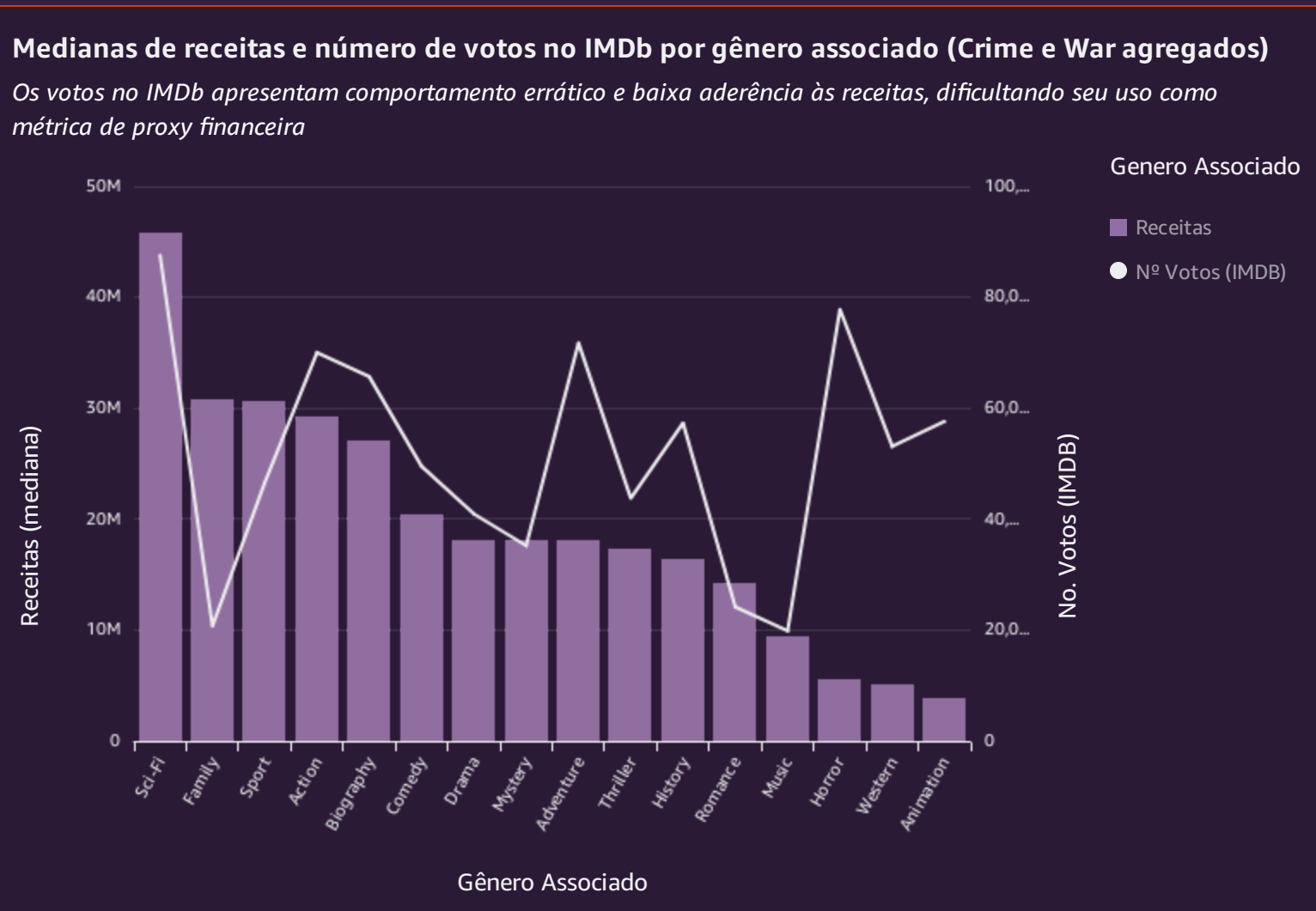
Top 10 Filmes com maiores resultados financeiros (gênero Crime domina a amostra)

Filmes de Crime lideram o Top 10 resultados financeiros, com subgêneros populares como Action (5 vezes), Drama (4 vezes) e Comedy (3 vezes).



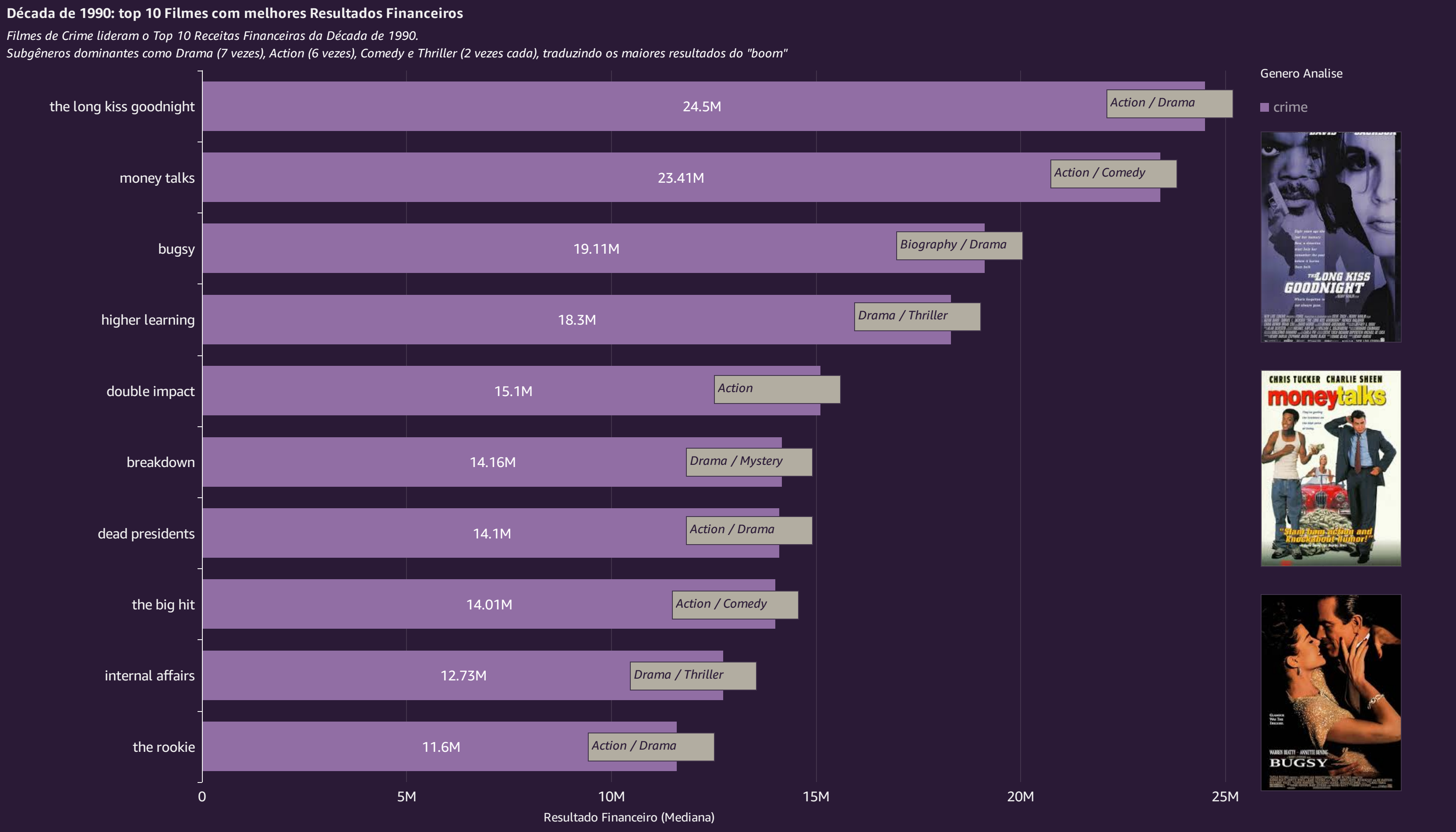
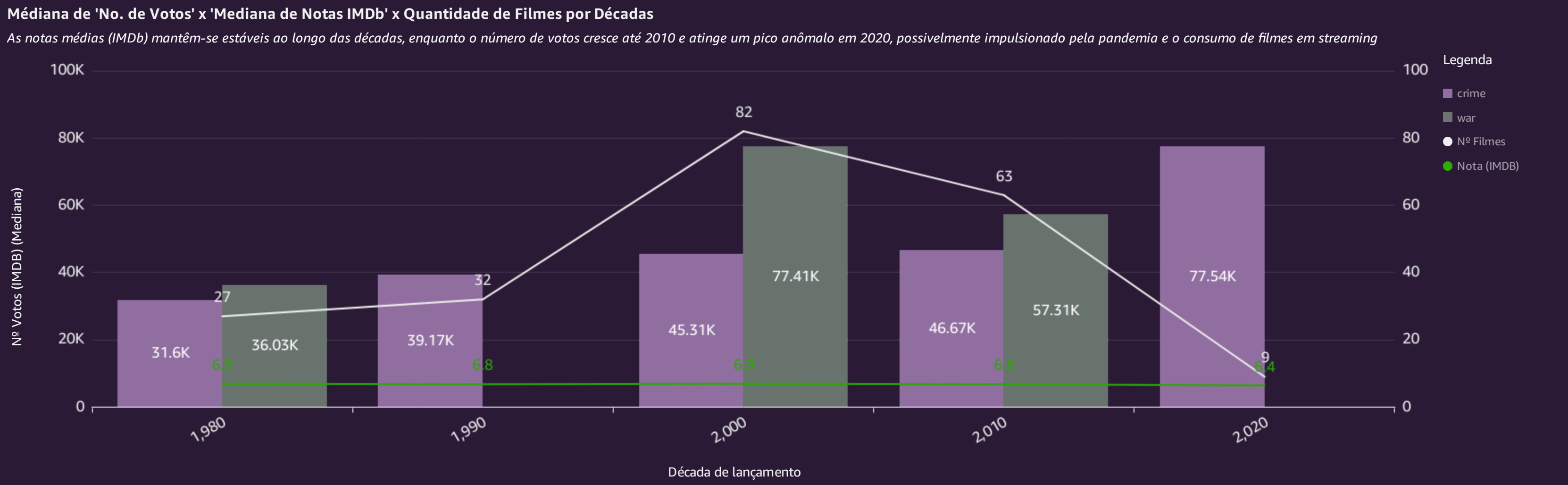
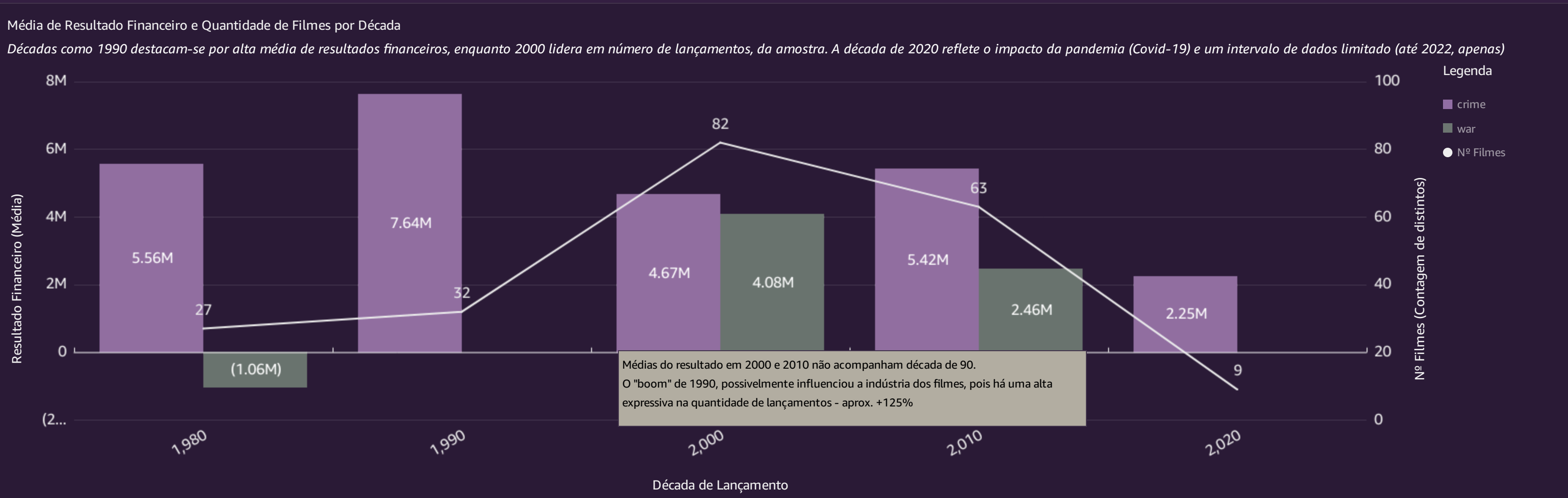
Análise: Avaliações e Popularidade

- A popularidade ou quantidade de votos de um filme se expressam na receita?
- As notas médias dos filmes têm correlação com as receitas?
- Quais gêneros associados costumam ter maior popularidade média? Há relação com resultado financeiro?



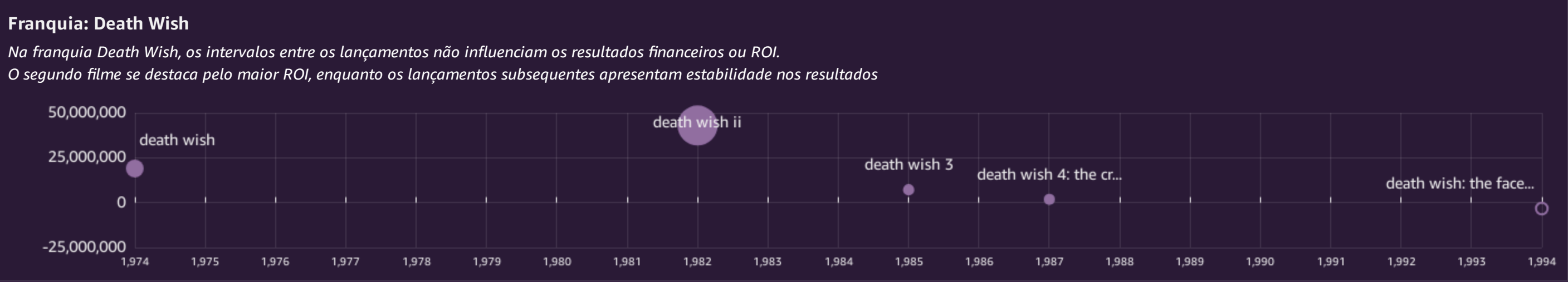
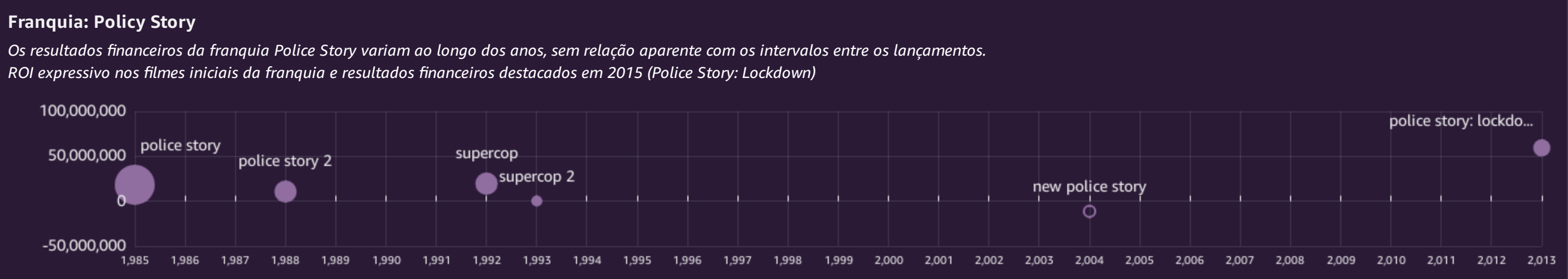
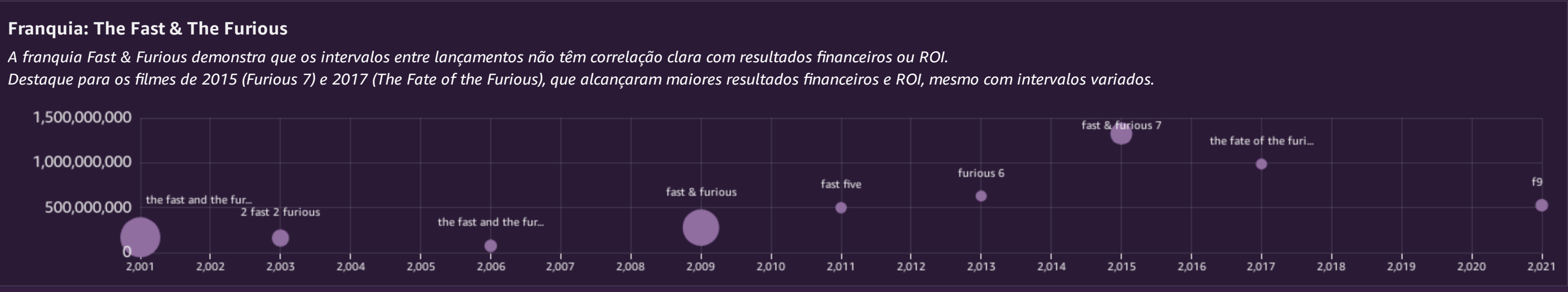
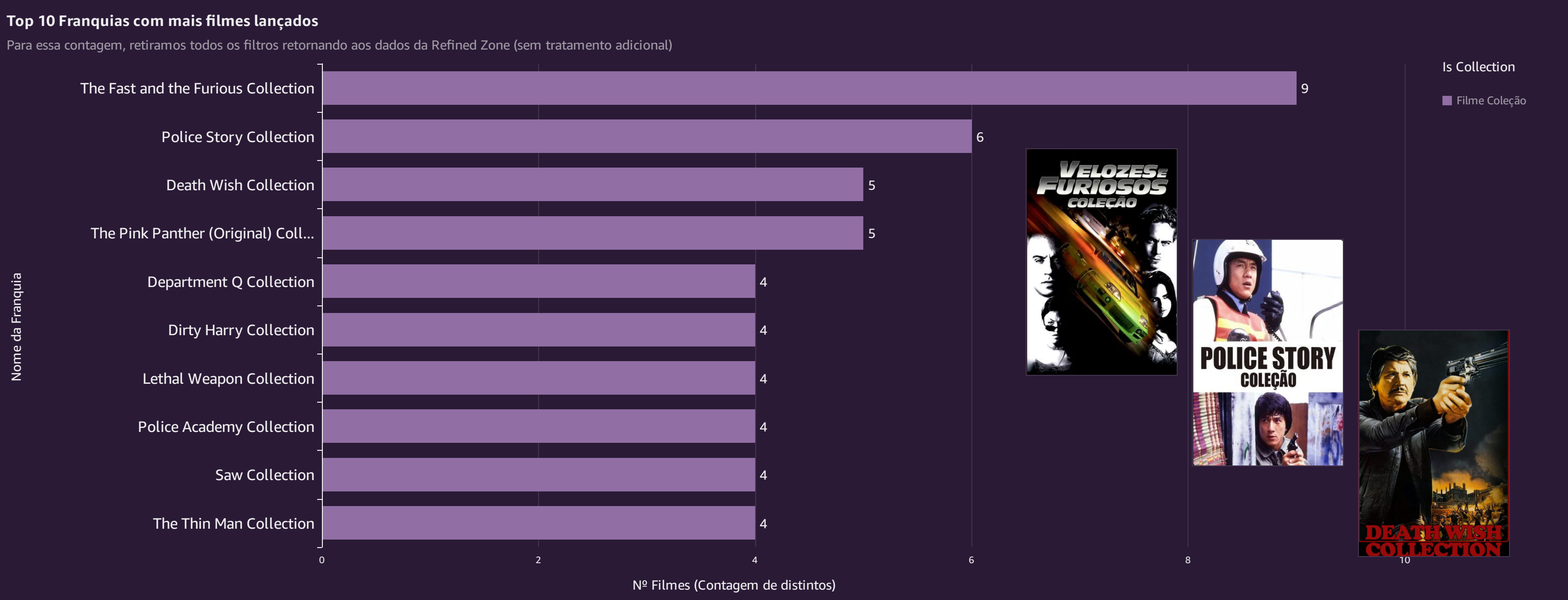
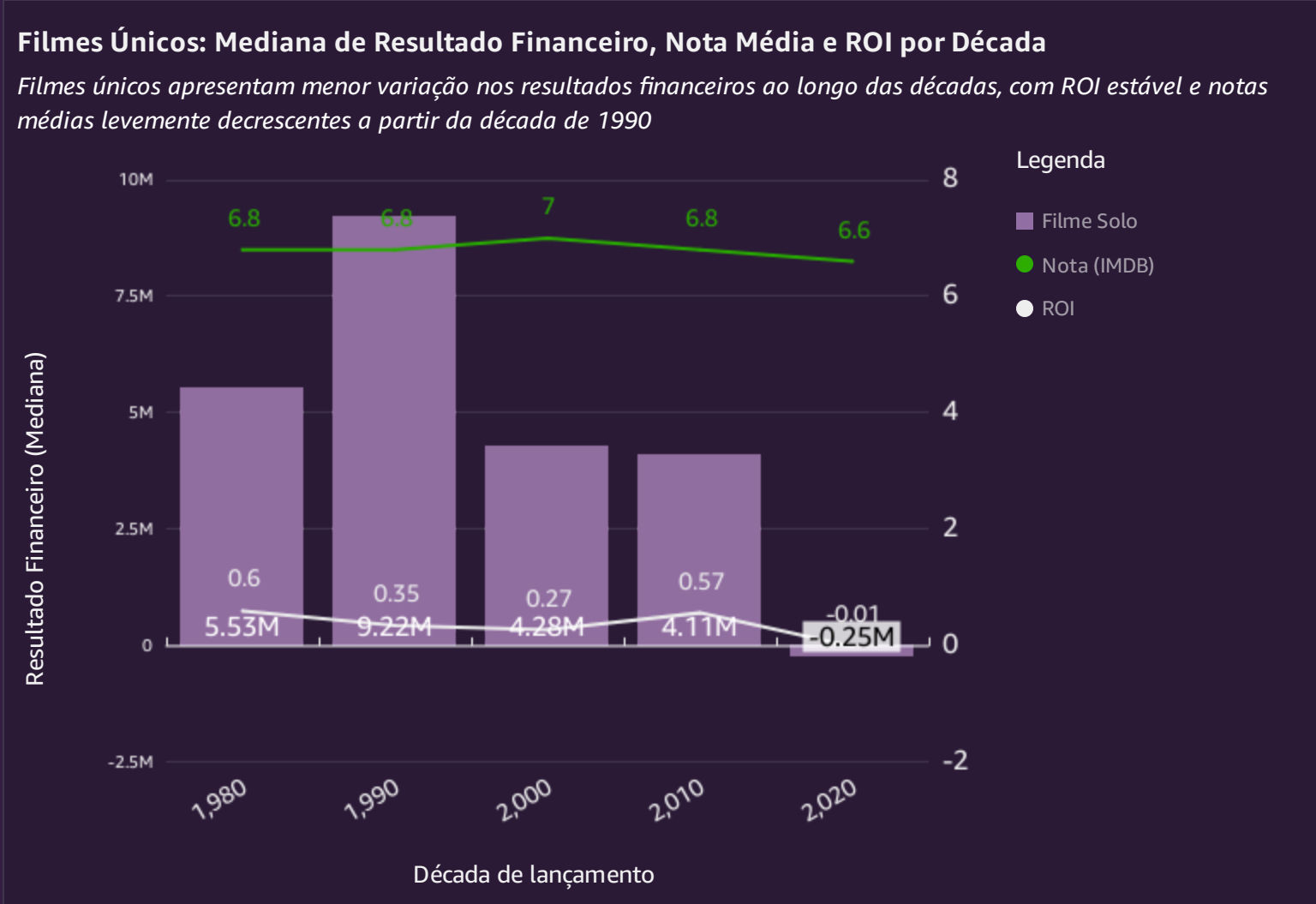
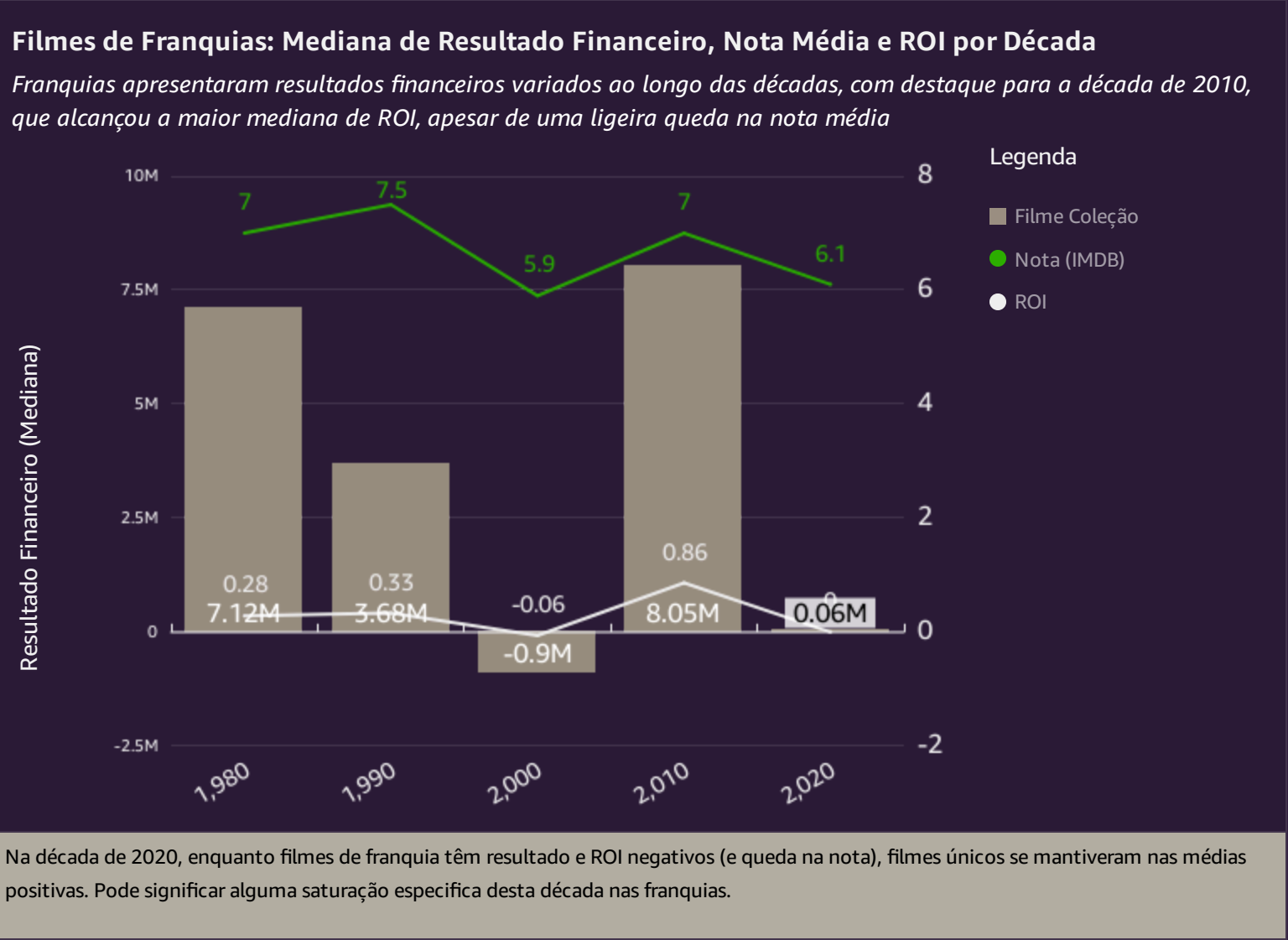
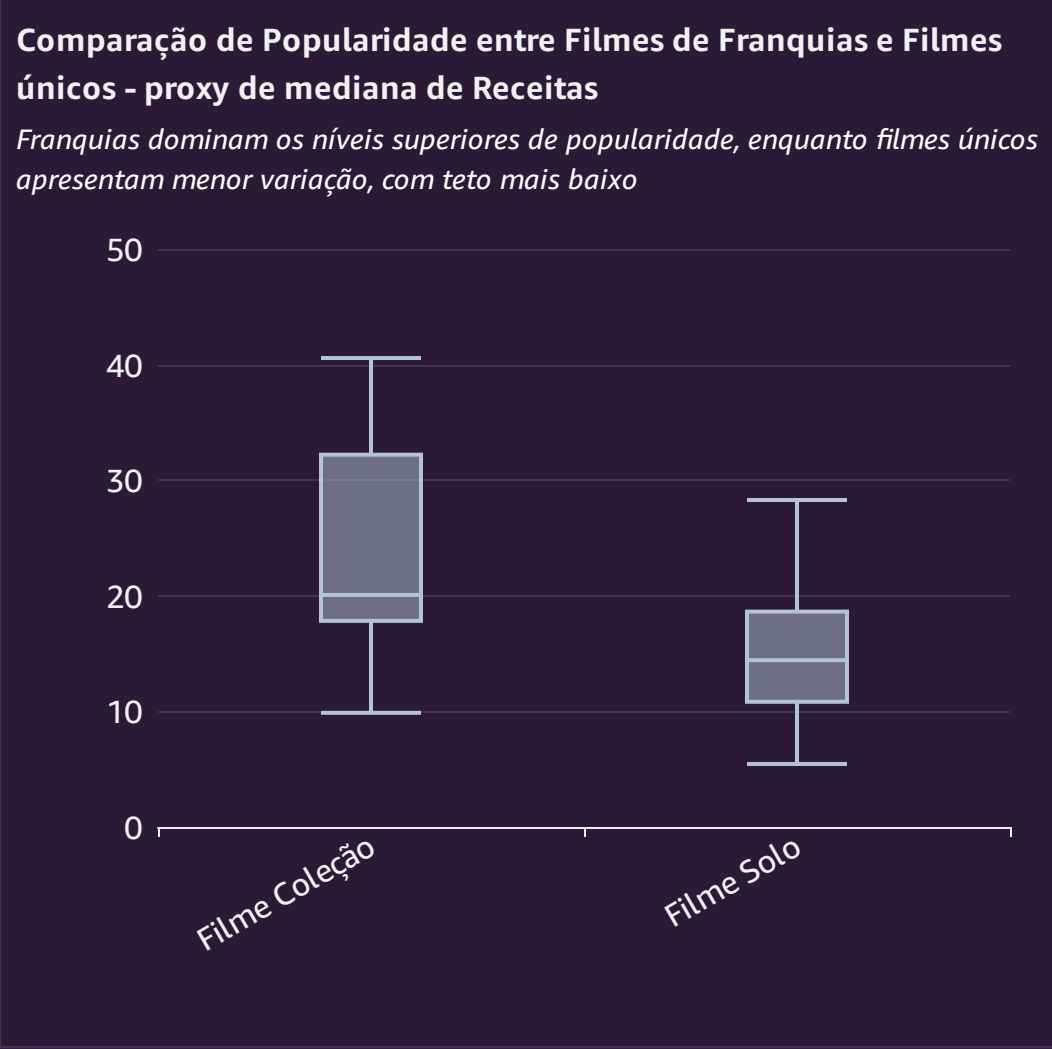
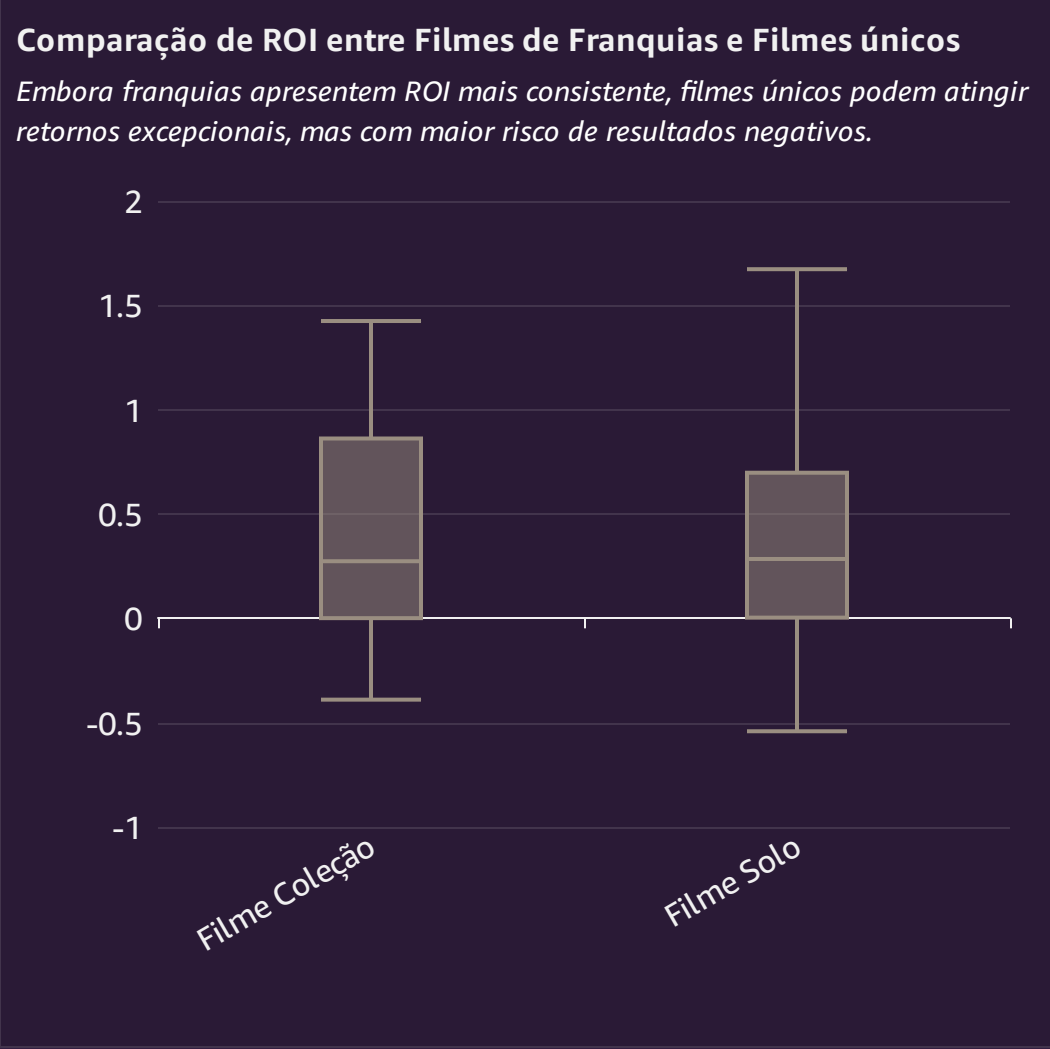
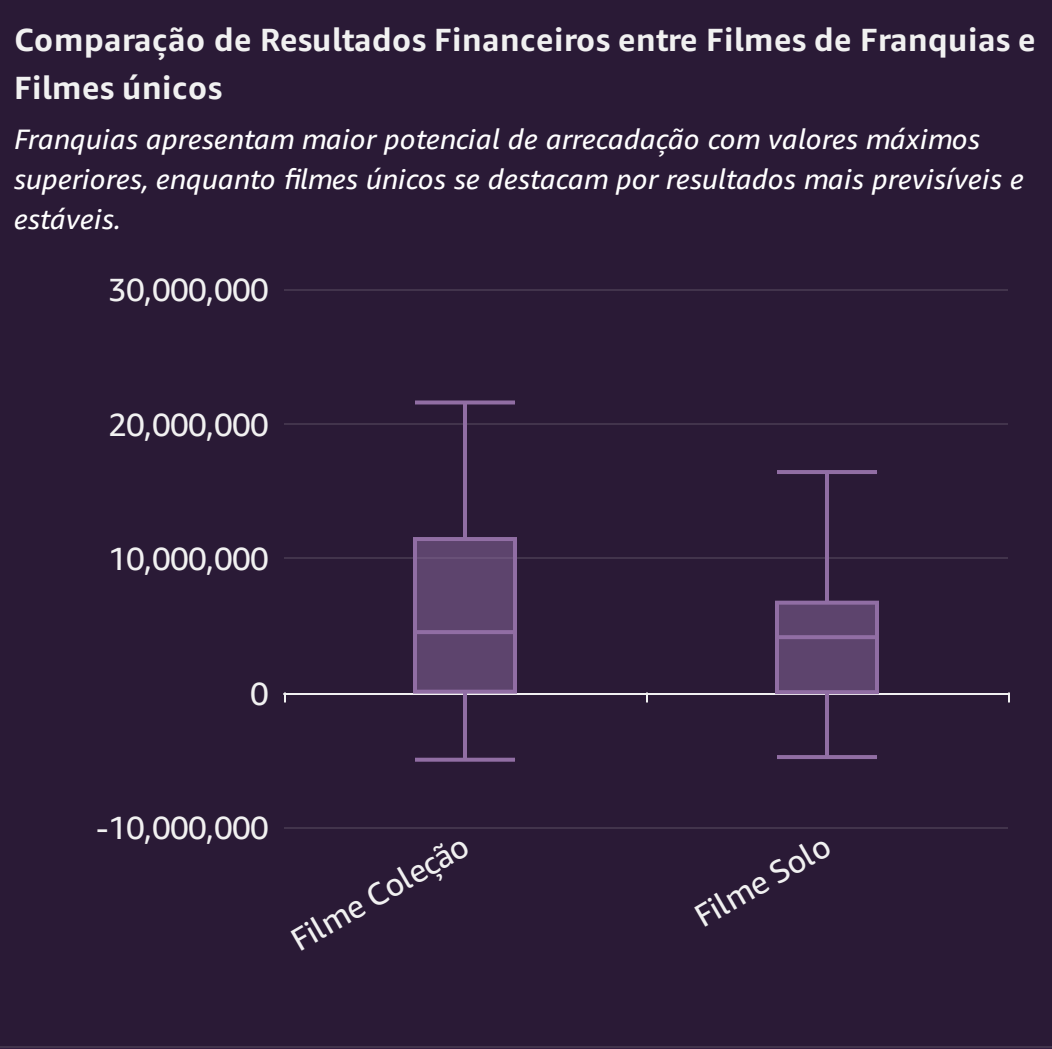
Análise: Tendências Temporais e Espaciais

- Quais décadas foram mais lucrativas para filmes de Crime e Guerra?
- Existe período de pico de lançamentos e avaliações?
- Quais países se destacam em produção (quantidade) e receita média?



Análise: Franquias x Filmes Únicos

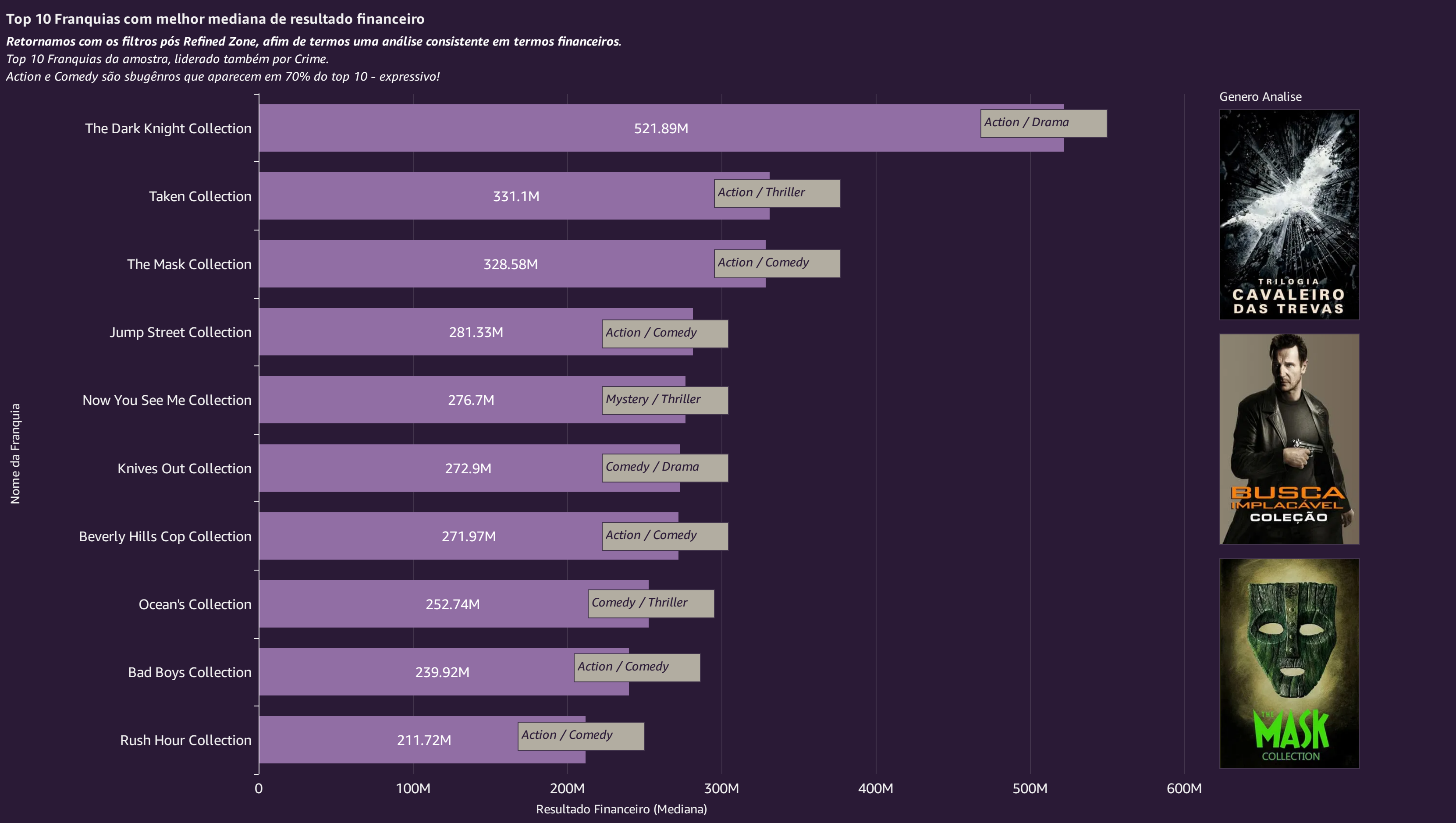
- Franquias 'Crime/ War' são mais ou menos lucrativas que filmes únicos?
- As sequências de filmes de uma franquia, mantém os resultados ou ROI? O tempo entre lançamentos influencia algo?
- Avaliações são impactadas por filmes de franquia x solo?



Baseando-se nas três dispersões ao lado, observa-se um padrão de saturação na segunda metade ou no último terço das franquias. Esse efeito se manifesta na queda progressiva do ROI, muitas vezes levando a resultados negativos nos lançamentos finais. A perda de apelo do público e o aumento dos custos de produção podem ser fatores determinantes nesse declínio.

Por outro lado, Police Story: Lockdown (2013) exemplifica como reboots e revitalizações podem reverter essa tendência. Diferente da trajetória decrescente vista em outras franquias, este lançamento conseguiu alcançar o melhor resultado financeiro da série, mostrando que a renovação pode trazer novo fôlego ao mercado.

Esses padrões sugerem a necessidade de um aprofundamento maior no segmento para identificar oportunidades estratégicas em reboots e entender melhor os fatores que contribuem para a saturação e o declínio financeiro das franquias ao longo do tempo.



Conclusões e Recomendações Acionáveis

Investimento em Franquias com Planejamento de Ciclo de Vida

- Franquias apresentam potencial de arrecadação elevado, mas tendem à saturação na segunda metade do seu ciclo.
- Franquias que contemplem, desde o início, uma estratégia de reboot ou renovação ao final do ciclo, podem ter um aproveitamento superior.

Otimização de Combinações de Gêneros

- A maioria dos filmes da amostra (desbalanceada) foram Crime, e por isso apareceram como destaque, de forma geral - mesmo nas comparações em medianas, com War.
- Subgêneros (de Crime), como History, Sci-Fi, Family mostraram bons retornos financeiros e ROIs consistentes, em suas medianas.
- Observando os top 10 resultados, subgêneros de Action, Drama, Thriller podem elevar consideravelmente os resultados - ainda que quando observados nas medianas são menos expressivos.

Estratégia de Orçamento Baseada em Subgêneros

- Nem sempre grandes orçamentos geram ROI proporcional.
- Ajustar investimentos de acordo com o subgênero, favorecendo History (bom ROI com recursos moderados) e equilibrando aportes em Sci-Fi (que registrou uma boa relação custo-benefício).
- Orçamentos de investimento médio têm maior probabilidade de retornar bom resultado financeiro.

Expansão Geográfica Estratégica

- Mercados fora do eixo tradicional (EUA, Reino Unido, Alemanha, França) mostraram potencial não totalmente explorado.
- Investir em War no Leste Europeu, por exemplo, ou Crime na Oceania, além de fortalecer coproduções internacionais para mitigar riscos e ampliar alcance de público, são oportunidades.

Timing de Lançamentos Otimizado

- Filmes de franquias "recentes" (2010+) mantêm médias de receita positivas com melhores resultados financeiros e ROI, comparado aos filmes únicos.
- A década de 1990 apresentou números financeiros expressivos, sugerindo modelos de sucesso que podem ser revisados e adaptados.

Estratégia de Popularidade como Proxy de Receita

- Popularidade no TMDb apresenta forte relação com receitas de bilheteria, salvo exceções pontuais (Horror e Animation).
- Adotar métricas de engajamento online como indicador precoce de potencial de vendas, fortalecendo campanhas de marketing digital e ampliando a visibilidade antes do lançamento.
- Um bom engajamento online, pode refletir diretamente na arrecadação do filme. Sci-fi + Crime é um ótimo exemplo disso.