

Uma primeira pergunta sobre o banco de dados é: qual é frequência com que os usuários avaliam os filmes? A figura 1 mostra quantos usuários deram a mesma quantidade de avaliações. Nessa figura não são apresentados todos os dados, apenas os mais significativos. Em uma rápida avaliação é possível dizer que a quantidade de avaliações mais frequentes é 175, com cerca de 4500 usuários com a essa quantidade de avaliações. Considerando que esses registros são de um período de 20 anos, podemos dizer que esses usuários avaliaram em média quase 9 filmes por ano. Outros casos de destaque são 30, 60 e 189 avaliações com 2300, 3000 e 2900 casos, respectivamente. A figura 2 mostra uma outra faixa de valores de quantidade de avaliações, percebe-se claramente a redução na escala de frequência e o maior espaçamento entre os picos de frequência. Por fim apresento a figura 3, que mostra uma faixa de valores de usuários com avaliações semanais em média. Baseando-se na média temporal de avaliações, pode-se concluir que o grupo de usuários é formado majoritariamente por pessoas que simplesmente gostam de cinema e querem compartilhar suas experiências.

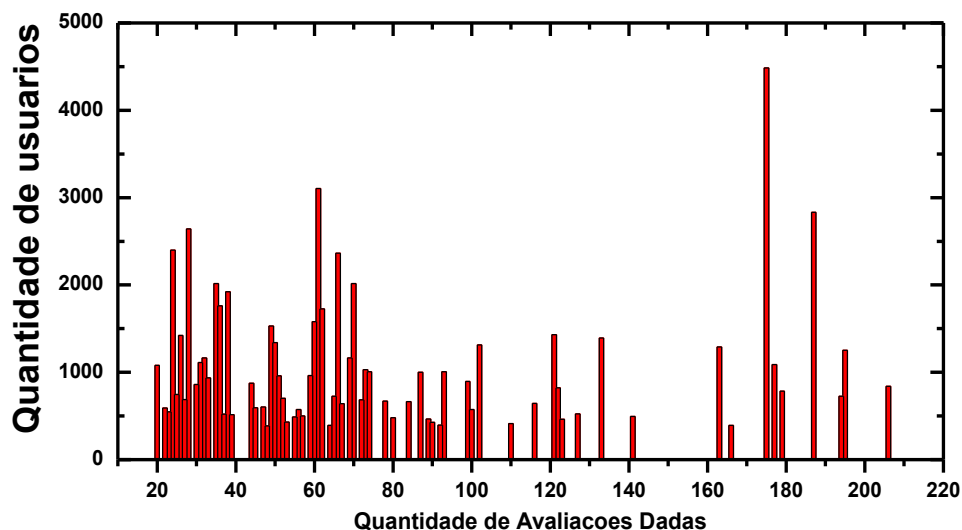


Figura 1: Quantidade de avaliações que cada usuário. Na figura não são apresentados todos os dados, para não prejudicar a visualização dos dados.

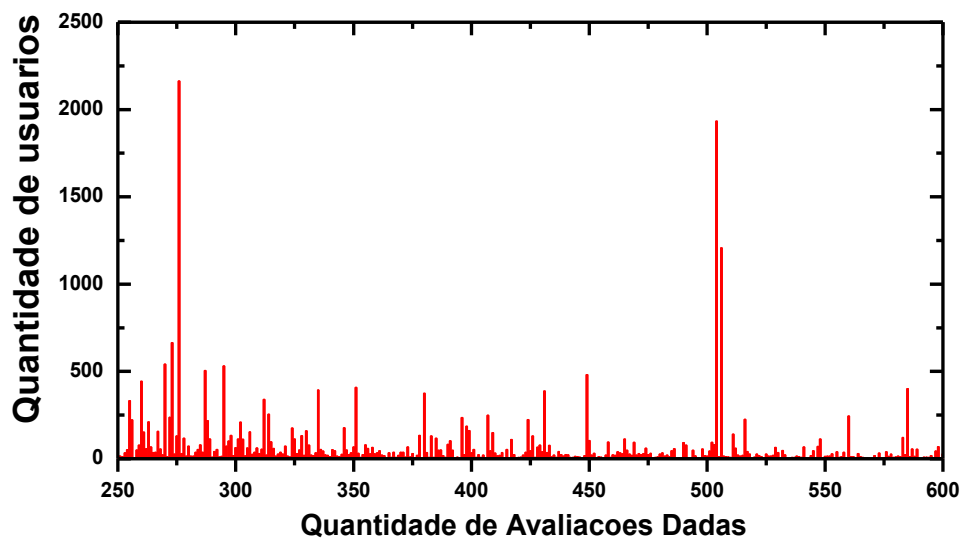


Figura 2: Quantidade de avaliações que cada usuário, considerando usuário que deram entre 250 e 600 avaliações.

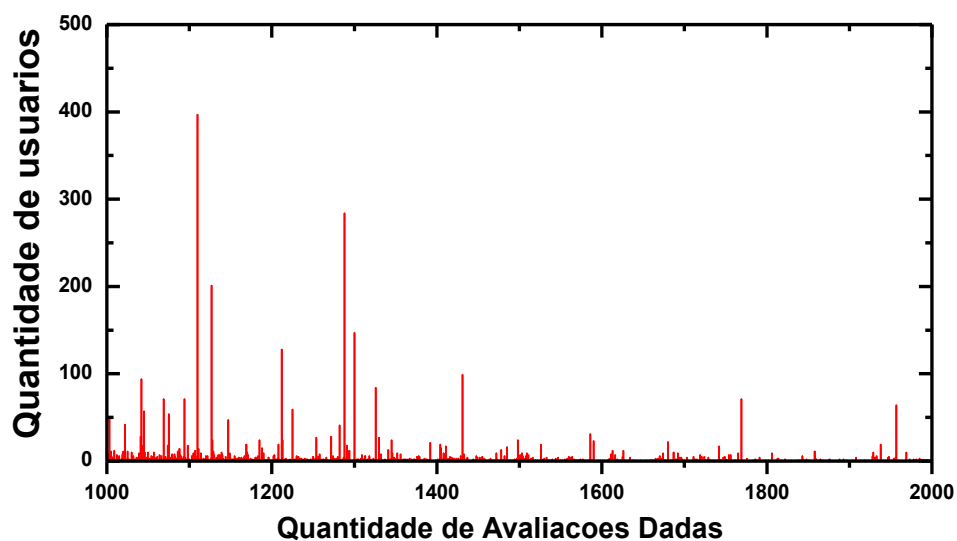


Figura 3: Quantidade de avaliações que cada usuário, considerando usuário que deram entre 1000 e 2000 avaliações.

A segunda etapa da exploração, segue na mesma linha, porém com a pergunta: quantas vezes cada filme foi avaliado? Analisando a figura 4, é possível verificar que 4000 filmes foram avaliados 50 mil vezes. Outros casos de destaque são filmes avaliados por 22, 13 e 2.5 mil usuários. Esses dados poderiam ser comparados com a bilheteria de lançamento dos filmes para tentar estabelecer uma relação entre avaliações e popularidade, para responder a seguinte questão: os filmes são mais avaliados por serem mais assistidos? Ou há outra explicação que envolva o gênero do filme, por exemplo?

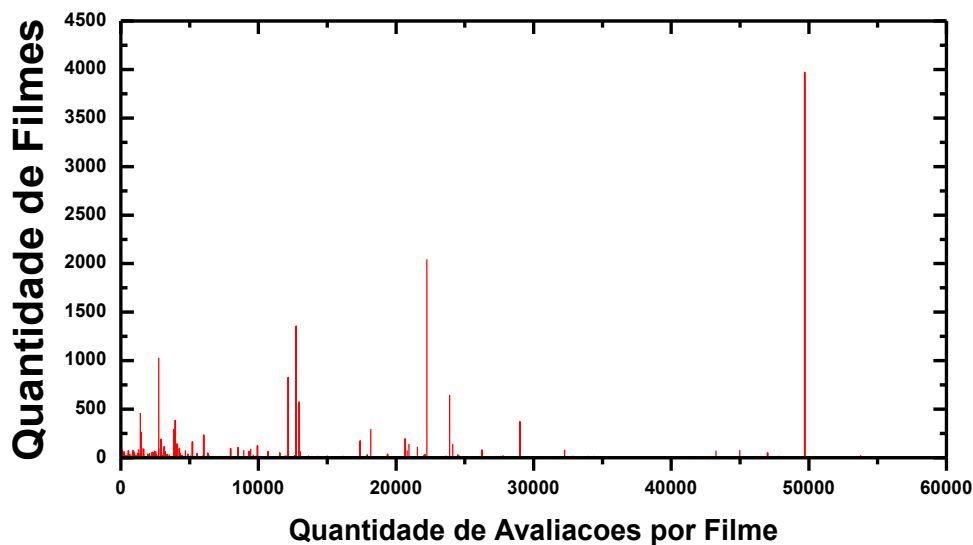


Figura 4: Quantidade de avaliações recebidas por cada filmes versus a quantidade de filmes que receberam a mesma quantidade de avaliações.

Para tentar responder à pergunta: porque alguns filmes são mais avaliados? Podemos explorar os dados disponíveis de diferentes maneiras. Uma primeira aproximação é estabelecer a quantidade de filmes avaliados por ano de lançamento. A figura 5 mostra esse resultado para o período de 1950 até 2015:

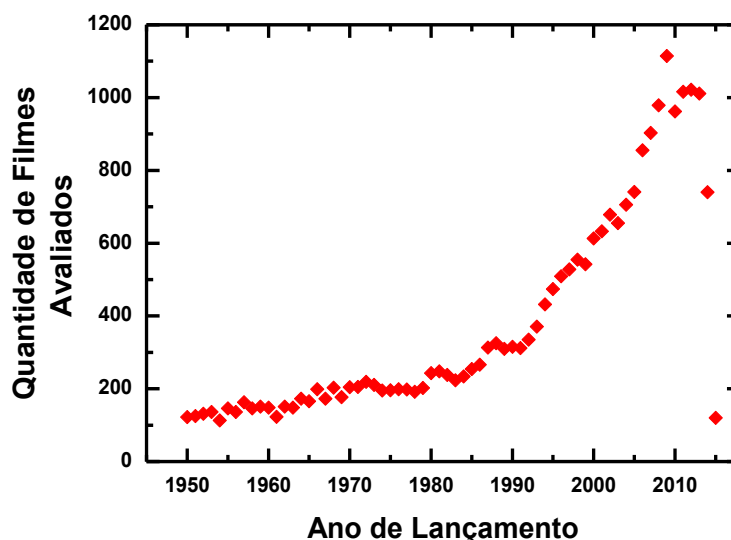


Figura 5: Quantidade de filmes avaliados por ano de lançamento. O dado de 2014 apresenta uma queda pois não houve tempo suficiente (entre o lançamento e o fechamento deste banco dados) para atingir um grupo maior de usuários. O ano de 2015 tem a menor quantidade pelo mesmo motivo, além disso o dado apresenta somente filmes de 2015 até março.

Para interpretar os dados apresentados na figura 5 é necessário consultar alguns dados sobre a indústria do cinema, serviços on-line de avaliação de filmes e serviços de *streaming*. O primeiro ponto a ser considerado é a quantidade de filmes lançados por ano, como por ser visto no site (Filmes por ano: <https://www.the-numbers.com/market/>) a quantidade total de filmes lançados por ano vem se mantendo constante sem grandes variações. Já podemos tirar uma primeira conclusão, a quantidade de filmes lançados não justifica por si só o aumento expressivo na quantidade de avaliações observado na figura 5.

No início da década de 1990 os lançamentos do Macintosh e Windows, permitiram que pessoas comuns acessassem a internet de casa; concomitantemente os primeiros sites focados na avaliação de filme surgiram como o IMDB (<https://en.wikipedia.org/wiki/IMDb>). Outros sites também foram lançados no final da mesma década como o Rotten Tomatoes (https://en.wikipedia.org/wiki/Rotten_Tomatoes). Isso pode explicar parte desse aumento a partir de 1990. O crescimento se dá até 2010, quando há um aparente platô (necessitamos de mais dados para termos certeza do comportamento). Nesse período de 20 anos (de 1990 a 2010), o surgimento de serviços de streaming como a Netflix (<https://pt.wikipedia.org/wiki/Netflix>) também pode ser apontado como uma possível explicação.

Para concluir, é crescente a facilidade de assistir qualquer filme (lançado em qualquer ano) em nossa casa. Porém essa facilidade não é convertida avaliações de filmes mais antigos, ou seja, os usuários tendem avaliar filmes lançados recentemente (considerando a data da avaliação). Isso nos leva a mais perguntas: os usuários estão assistindo filmes antigos? Se estiverem, esses usuários preferem não avaliar os filmes assistidos? Uma possível resposta é a existência de dois grupos entre os usuários, os que preferem ver apenas filmes recentes e aqueles que gostam de assistir desde filmes antigos aos mais recentes. Esse segundo grupo tem subdivisões, usuário que veem filmes lançados a partir de 1990, 1980, 1970 e assim por diante. Esse possível comportamento pode ser visto na figura 6, percebe-se que todo começo de década se inicia um ciclo aumento no número de filmes avaliados que termina no fim da década. Tal comportamento pode ser explicado por usuários que preferem filmes a partir de uma determinada década, tendo assim um efeito cumulativo (considerando que temos usuário que começam avaliar filme da década anterior, com os novos usuários).

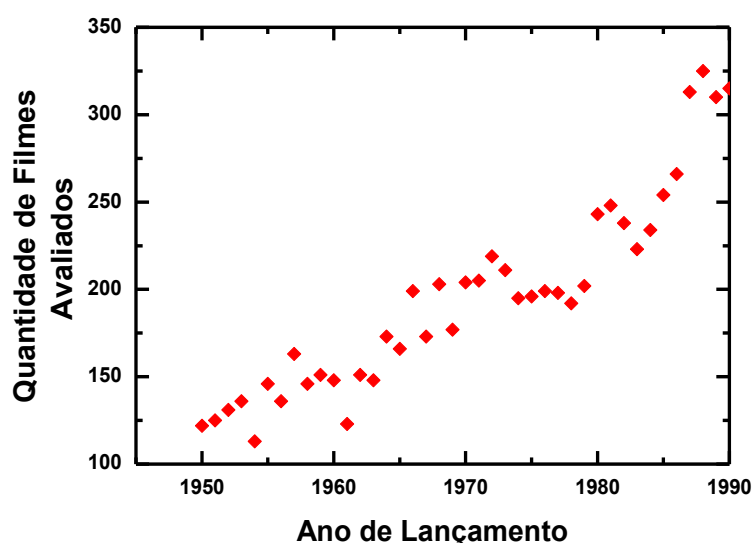


Figura 6: Quantidade de filmes avaliados por ano de lançamento, considerando o período entre 1950 e 1990

A figura 7 mostra a quantidade de filmes dos diferentes gêneros, nesses dados um filme pode ser contabilizado mais de uma vez, visto que alguns filmes têm mais de um gênero. Uma análise rápida nos mostra que filmes de drama têm mais avaliações, o segundo gênero com mais avaliações é a comédia. Para uma avaliação mais precisa, irei considerar filmes com apenas um gênero, assim os filmes poderão ser contabilizados apenas uma vez, a figura 8 mostra o resultado.

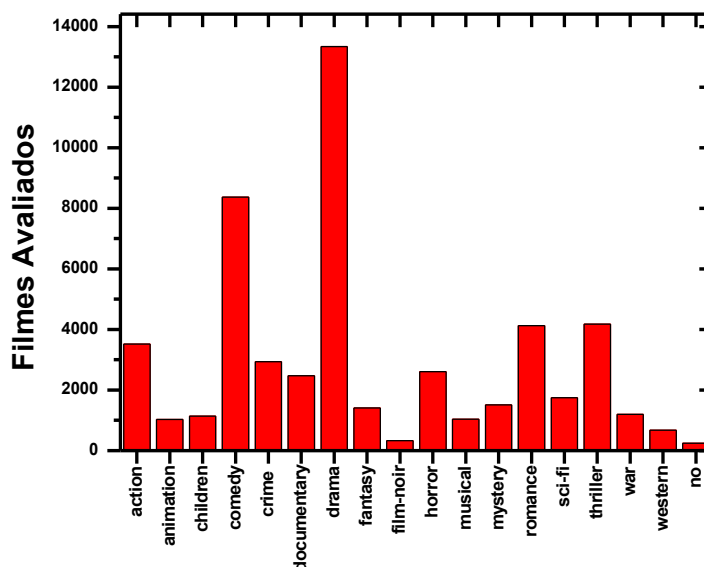


Figura 7: Quantidade de filmes, de cada gênero, avaliados

A figura 8 mostra claramente que drama e comédia ainda são os gêneros com a maior quantidade de avaliações. Gêneros que tinham certo destaque na figura 7, como romance, thriller e ação, apresentam uma redução considerável quando temos filmes com apenas um gênero, o que pode indicar uma não preferência dos usuários por filmes que sejam puramente desses respectivos gêneros. Podemos destacar o crescimento proporcional dos documentários, mas isso reflete mais o tipo de produção (um documentário geralmente não apresenta um segundo gênero) do que a preferência dos usuários.

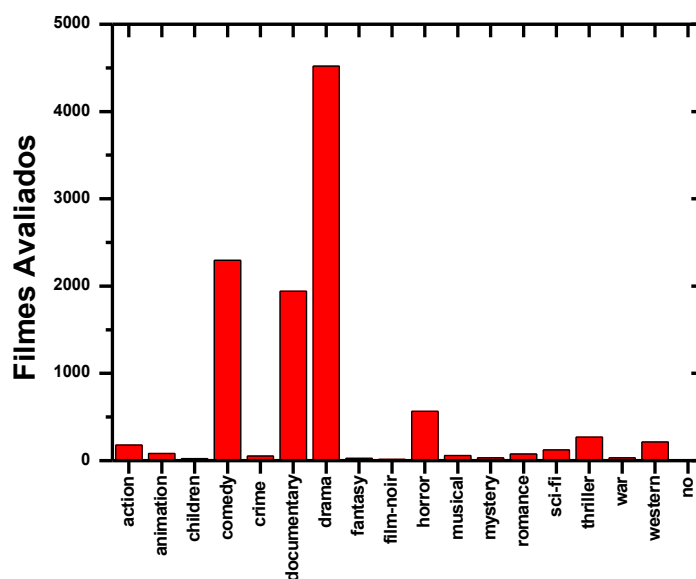


Figura 8: Quantidade de filmes (com apenas um gênero) avaliados

Uma pergunta que surge após a apresentação desses dados é: a quantidade de filmes lançados de cada gênero influencia de alguma maneira a quantidade de avaliações? Para responder essa questão é necessário conhecer os dados apresentados em: (<https://www.the-numbers.com/market/genres>) . É possível perceber que os gêneros mais produzidos de filmes são: drama e comédia. Isso pode explicar a maior quantidade de avaliações para esses gêneros. Um outro ponto que merece destaque é a bilheteria de cada gênero (relacionado diretamente com a audiência nos cinemas), os dois gêneros com maior bilheteria são aventura e ação (mesmo com uma quantidade menor de filmes produzidos). Uma outra conclusão parcial é: a maior parte dos usuários que avaliam os filmes não os assistiram nos cinemas; se bilheteria refletisse em mais avaliações, os líderes seriam aventura e ação, porém o que observamos é que usuários avaliam gêneros com mais opções disponíveis.

Dando prosseguimento, podemos estudar as notas dadas a cada gênero. Seguindo a lógica utilizada até agora, irei mostrar um dado mais geral para depois aprofundar a análise. Essa primeira abordagem irá mostrar apenas quantos filmes de cada gênero (o mesmo filme poderá ser contabilizado mais de uma vez, já que há filmes com múltiplos gêneros) recebeu uma nota alta (4, 4.5 ou 5) e quantos receberam uma nota baixa (0, 0.5, 1, 1.5, 2, 2.5 e 3). Isso nos dará uma pequena ideia da preferência dos usuários, a figura 9 mostra o resultado.

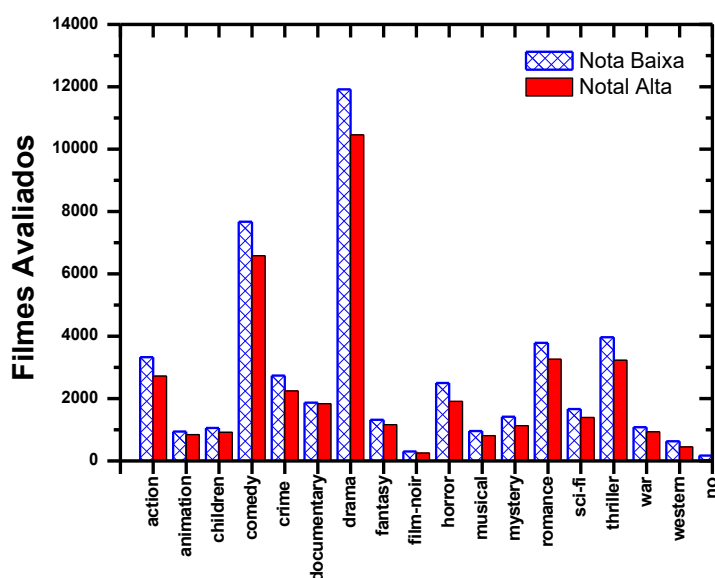


Figura 9: Quantidade de filmes (de cada gênero) avaliados que receberam pelo menos uma nota baixa (0, 0.5, 1, 1.5, 2, 2.5 e 3) – coluna azul, ou pelo menos uma nota alta (4, 4.5 ou 5) – coluna vermelha.

A figura 9 mostra que a quantidade de notas baixas é superior às notas altas. Antes de tirar qualquer conclusão sobre a grau de exigência dos usuários e/ou qualidade dos filmes, deve-se lembrar que há filmes com apenas uma avaliação enquanto outro com quase 50 mil avaliações, ou seja, há filmes que têm avaliações mais relevantes que outros do ponto de vista estatístico. Para ter um melhor entendimento, irei fazer dividir os filmes em dois grupos: filmes com até 500 avaliações, e filmes com mais de 500 avaliações, em seguida irei mostrar quantos filmes de cada conjunto tiveram nota média maior e menor que 3. As figuras 10 e 11 mostram os resultados.

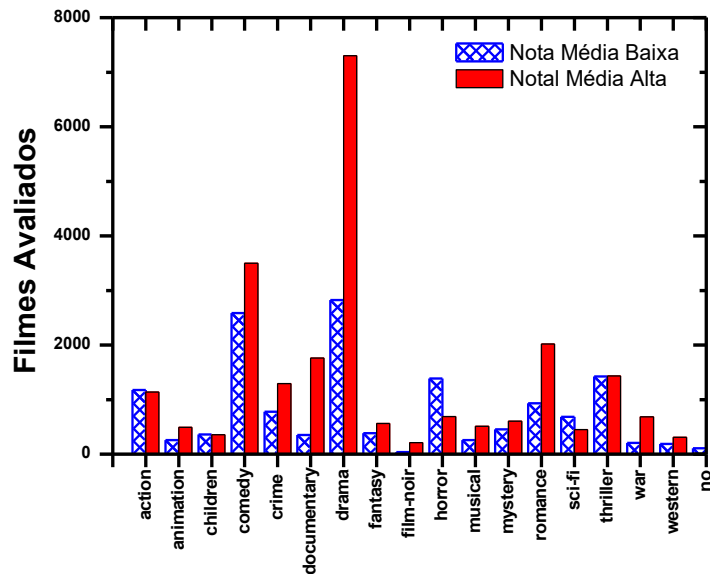


Figura 10: Quantidade de filmes (de cada gênero) avaliados que receberam nota média abaixo de 3 – coluna azul, ou nota média acima de 3– coluna vermelha. Esses filmes receberam no máximo 500 avaliações.

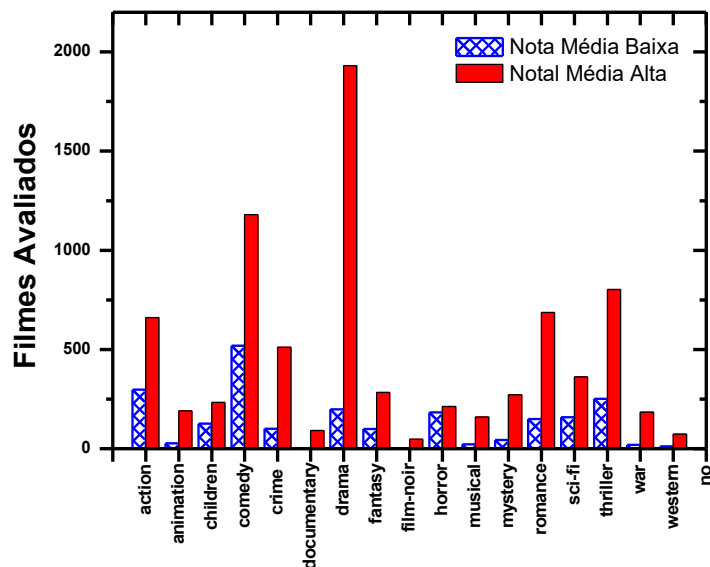


Figura 11: Quantidade de filmes (de cada gênero) avaliados que receberam nota média abaixo de 3 – coluna azul, ou nota média acima de 3– coluna vermelha. Esses filmes receberam no mínimo 500 avaliações.

Comparando as figuras 10 e 11, percebe-se que filmes com mais de 500 avaliações (figura 11) apresentam maior probabilidade de ter nota média maior que 3 (para tirar essa conclusão basta perceber que, proporcionalmente, os filmes da figura 11 tem mais avaliações maiores que 3, quando comparados com os filmes da figura 10). Quais conclusões pode-se tirar desse fato? Quanto mais avaliações um filme tiver maior é sua chance de ser bem avaliado, em futuras avaliações. Podemos conjecturar alguma explicação para esse fato. Dado que a maioria das avaliações foram dadas por usuários caseiros, que provavelmente usaram a internet para este fim, há grande chance desse usuário ter feito alguma pesquisa (usando serviços como IMDB ou apenas recomendações do NetFlix) para

escolher um filme. As pesquisas feitas vão mostrar filmes que têm maior aceitação entre os usuários, isso faz com que filmes com boas notas tenham maior probabilidade de serem sugeridos para usuários. Ou seja, quanto mais avaliações positivas mais chance de o filme ser indicado, maior a probabilidade de haver mais pessoas assistindo esse filme que consequentemente terá mais indicações.

Conclusões.

Mostrei inicialmente a quantidade de avaliações dadas por cada usuário, me permitindo concluir que as avaliações foram dadas por usuários que assistam filmes como passatempo (considerando média anual de avaliações). Em seguida mostrei a quantidade de vezes que cada filme foi avaliado, essa análise mostrou haver conjuntos bem distintos de filmes que foram avaliados 50, 22, 13 e 2.5 mil vezes. Para tentar responder à pergunta, porque há filmes com mais avaliações que outros, mostrei a quantidade de filmes avaliados por ano de lançamento, apresentei algumas referências externas sobre o assunto, e concluí que usuários tendem a avaliar filmes recentes.

Em seguida mostrei a quantidade de avaliações que cada gênero de filme recebeu, novamente com apoio de referências externas, cheguei à conclusão de que usuários que avaliam os filmes são predominantemente usuários caseiros (que assistem em lugares que não seja o cinema). Para finalizar estudei a nota dada pelos usuários a cada gênero, este estudo permitiu inferir que filmes que mais agradam os usuários terão mais chances de serem assistidos, e de receberem maiores notas.