



**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS FLORIANÓPOLIS
INE-DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA
INE5405 - PROBABILIDADE E ESTATÍSTICA**

**ANTÔNIO ESCOBAR TORRES (22204119)
EDUARDO BISCHOFF GASEL (22200355)
VITOR PRAXEDES CALEGARI (22200379)**

Resumo Expandido - T1

**FLORIANÓPOLIS
2025**

1 INTRODUÇÃO E OBJETIVOS

Nos últimos anos, o crescimento da indústria de streaming tem chamado a atenção das grandes marcas. Com milhões de usuários consumindo e criando conteúdo diariamente, compreender os padrões e tendências desse ecossistema tornou-se fundamental para criadores, empresas e pesquisadores. Neste trabalho, usaremos métodos de probabilidade e estatística para analisar um conjunto de dados sobre streams da plataforma Twitch do ano de 2020, explorando variáveis como tempo assistido (Watchtime), tempo de transmissão (Stream time), pico de espectadores (Peak Viewers), média de espectadores (Average Viewers), número de seguidores (Followers), programa de parceria (Partnered) e idioma da transmissão (Language) de diversos criadores de conteúdo.

A partir da análise estatística dessas variáveis, buscamos responder a perguntas essenciais para o entendimento do comportamento da audiência na Twitch. Por exemplo, será que existe uma correlação entre o idioma da transmissão e o número médio de espectadores? Como o tempo de transmissão impacta o crescimento de seguidores? Ao aplicar técnicas como análise exploratória de dados e medidas de tendência central esperamos ter insights relevantes sobre o mercado de streaming.

2 MATERIAIS E MÉTODOS

Para realização das análises, usamos a base de dados *Top Streamers on Twitch*[1]. Trata-se de um compilado de informações que fazem referência aos 1000 canais mais assistidos da plataforma e dados referentes às suas transmissões (serão detalhados ainda nesta seção). Os dados são referentes ao ano de 2020, então vale considerar que o retrospecto do ano em questão é influenciado pelas políticas de lockdown por conta do coronavírus, que definitivamente aumentaram o consumo de diversas plataformas de streaming.

Outro fato importante, é que mesmo que a plataforma tivesse como ideia inicial ser uma rede de streams de games, hoje em dia há uma considerável parte dos criadores que produzem seus conteúdos com outros objetivos. Como exemplo, entre os tópicos mais populares de streams temos: “Just Chatting” e “IRL” nas quais os criadores conversam com seu público, sem estar necessariamente mostrando algo na tela, podcasts e também esportes tradicionais, como futebol e basquete.

A base de dados escolhida possui 11 colunas, sendo a primeira delas o nome do canal, e logo, os dados relacionados, como:

Watchtime: Trata-se de um índice quantitativo e representa o tempo em que o canal foi assistido durante o período (em minutos) para isso, é somado a quantidade de horas em que cada um dos seus espectadores permaneceu na transmissão.

Stream Time: Trata-se de um índice quantitativo e representa a quantidade de tempo que o canal ficou transmitindo conteúdo. Lembrando que diferentemente dos canais de televisão tradicionais, uma live pode ser encerrada, então normalmente, a transmissão só é difundida quando o criador está de fato em frente ao computador interagindo. Isso não exclui o fato de que podem existir canais que estão sempre transmitindo, alternando entre partes efetivamente ao vivo e reprises.

Peak Viewers: É o máximo de espectadores simultâneos que a stream teve durante o período. De acordo com a aplicação, são contabilizados tanto espectadores cadastrados e registrados, quanto usuários sem conta, assistindo como visitantes. É um índice quantitativo discreto, pois é representado por números inteiros, uma vez que o pico de visualizações não pode ser “três usuários e meio”.

Average Viewers: Representa a média de espectadores simultâneos ao longo de toda a transmissão. A média de espectadores de um canal na Twitch é obtida através da coleta periódica do número de visualizações simultâneas ao longo de uma transmissão. Esses dados são então utilizados para calcular a média aritmética, representando o número médio de viewers durante aquele período. Embora seja derivado de uma contagem, como a variável é uma média, pode assumir valores decimais, por isso, é classificada como quantitativa contínua.

Followers: Número total de seguidores do canal no momento da coleta dos dados. Esse valor representa a base acumulada de usuários que optaram por acompanhar o criador de conteúdo para serem notificados sobre novas transmissões. Assim como peak Viewers, é de caráter quantitativo e discreto.

Followers gained: Quantidade de novos seguidores adquiridos durante o período analisado. Esse dado ajuda a medir o crescimento do canal e pode indicar o impacto de determinadas transmissões ou eventos especiais. Também é uma variável do tipo quantitativo e discreto.

Partnered: Indica se o canal faz parte do programa de parceiros da Twitch. Canais parceiros recebem benefícios exclusivos, como monetização via anúncios e inscrições, além de maior personalização e suporte da plataforma. É um índice qualitativo e nominal, uma vez que ser ou não parceiro da plataforma, é uma opção, e não significa que um canal está “melhor avaliado” que o outro.

Language: Especifica o idioma principal da transmissão. Esse dado permite identificar a distribuição linguística do público da Twitch e analisar como a escolha do idioma pode impactar o engajamento e o crescimento do canal. É também uma variável do tipo qualitativa nominal.

As variáveis mencionadas foram escolhidas pois representam aspectos fundamentais da performance, alcance e características dos canais da Twitch. As

selecionamos guiados pelo objetivo de realizar uma análise estatística abrangente, que permita investigar correlações, tendências e padrões de comportamentos entre os principais streamers da plataforma ao longo do ano de 2020.

3 RESULTADOS E DISCUSSÕES

3.1 GRÁFICOS E TABELAS DE FREQUÊNCIA

Nesta seção, as variáveis escolhidas serão apresentadas por meio de tabelas de frequência e gráficos adequados. Em relação às variáveis qualitativas, foi utilizado o gráfico de barras verticais para facilitar a visualização entre as diferentes frequências associadas a cada rótulo do gráfico.

O gráfico de histograma foi escolhido para representar as variáveis quantitativas, pois dessa forma podemos agrupá-las em intervalos, mostrando de forma clara quantos casos caem em cada faixa. Os histogramas também permitem uma representação visual das assimetrias presentes no gráfico que podem ser facilmente identificadas ao observar a diferença de tamanho entre as barras.

3.1.1 Watchtime

Histograma do Tempo assistido

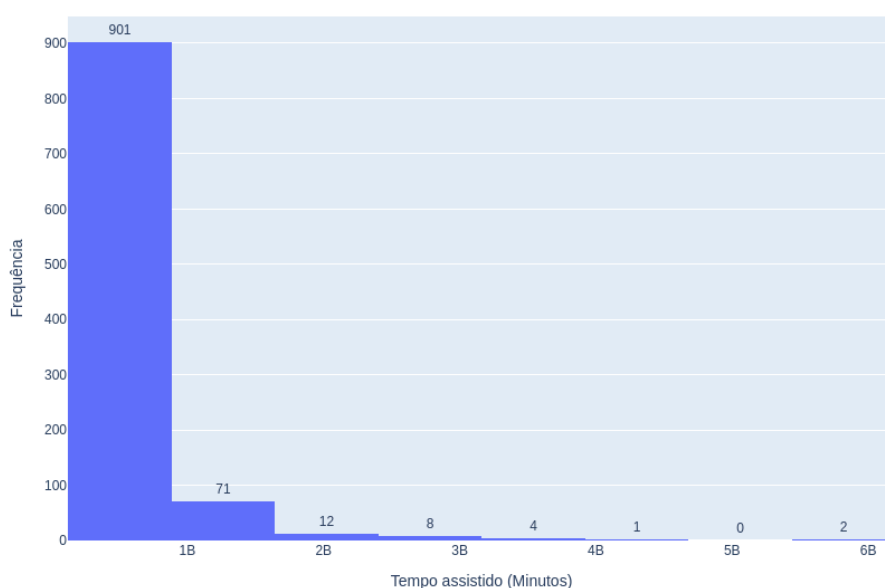


Tabela de Frequência: Watch time(Minutes)

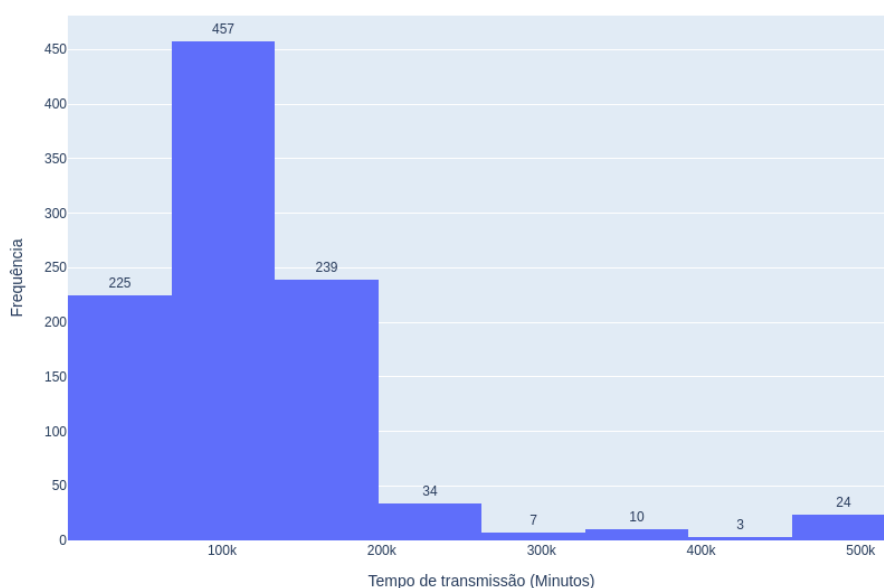
Faixa	Frequência	Porcentagem (%)
(122192849.999, 881438962.5]	901	90.1
(881438962.5, 1640685075.0]	71	7.1
(1640685075.0, 2399931187.5]	12	1.2
(2399931187.5, 3159177300.0]	8	0.8
(3159177300.0, 3918423412.5]	4	0.4
(3918423412.5, 4677669525.0]	1	0.1
(4677669525.0, 5436915637.5]	0	0.0
(5436915637.5, 6196161751.0]	3	0.3
Total	1000	100.00

A distribuição de watchtime entre os 1000 principais streamers, em um universo que já contava com cerca de 3 milhões de criadores ativos na Twitch em 2020[2], revela uma assimetria marcante, com 90,1 % dos canais concentrados no primeiro intervalo e frequências que caem rapidamente nos intervalos seguintes. Isso mostra que, mesmo ao limitar a análise aos mais famosos, há uma grande disparidade em que, poucos perfis concentram audiências extraordinariamente altas, enquanto a maioria atinge watchtimes consideravelmente mais modestos.

3.1.2 Stream time

Faixa	Frequência	Porcentagem (%)
(3464.999, 68212.5]	225	22.5
(68212.5, 132960.0]	457	45.7
(132960.0, 197707.5]	239	23.9
(197707.5, 262455.0]	34	3.4
(262455.0, 327202.5]	7	0.7
(327202.5, 391950.0]	10	1.0
(391950.0, 456697.5]	3	0.3
(456697.5, 521446.0]	25	2.5
Total	1000	100.00

Histograma do Tempo de transmissão



Os gráficos nesta seção destacam que parte dos principais streamers transmitem entre 1,9 e 4,3 horas por dia útil (22,5 % do grupo). Já 45,7 % ficam na faixa de 4,3 a 8,5 horas por dia útil — e mais 23,9 % entre aproximadamente 8,5 e 12,6 horas por dia útil. A partir deste ponto, as frequências caem rapidamente: apenas 7,9 % ultrapassam 12,6 horas por dia útil e menos de 5 % chegam a cerca de 20,9 horas ou mais por dia útil. Esses números mostram que, mesmo entre os mais famosos, a maioria se mantém em patamares moderados de transmissão por dias úteis, e apenas um pequeno grupo dedica jornadas

superiores a 12,6 horas diárias de streaming. Vale lembrar que é prática comum na plataforma Twitch realizar reprises de conteúdos já transmitidos, o que aumenta o tempo em transmissão sem necessariamente refletir o tempo real de transmissão ao vivo.

3.1.3 Average Viewers

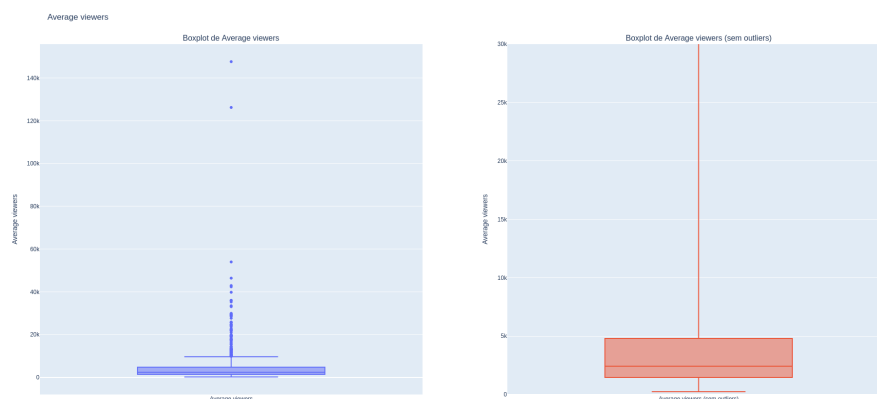


Tabela de Frequência: Average viewers		
Faixa	Frequência	Porcentagem (%)
(234.999, 18661.0]	955	95.5
(18661.0, 37087.0]	38	3.8
(37087.0, 55513.0]	5	0.5
(55513.0, 73939.0]	0	0.0
(73939.0, 92365.0]	0	0.0
(92365.0, 110791.0]	0	0.0
(110791.0, 129217.0]	1	0.1
(129217.0, 147644.0]	1	0.1
Total	1000	100.00

Ao analisar os dados de Average Viewers, pode-se notar que poucos canais concentram médias de espectadores muito elevadas, ultrapassando até mesmo a marca de cem mil espectadores, enquanto a grande maioria mantém-se abaixo dos vinte mil. Ao analisarmos as medidas de posição, observamos que o 1º quartil situa-se em torno de 1458 espectadores, a mediana em cerca de 2425 e o 3º quartil próximo de 4800, resultando num IQR (intervalo interquartil) de aproximadamente 3342, valor muito inferior à amplitude total dos dados. Além disso, a tabela de frequência mostra que 95,5 % dos streamers possuem média de espectadores inferior a 18661 espectadores, confirmando que a maior parte dos transmissores está concentrada em uma faixa bem abaixo dos outliers de maior audiência.

3.1.4 Followers

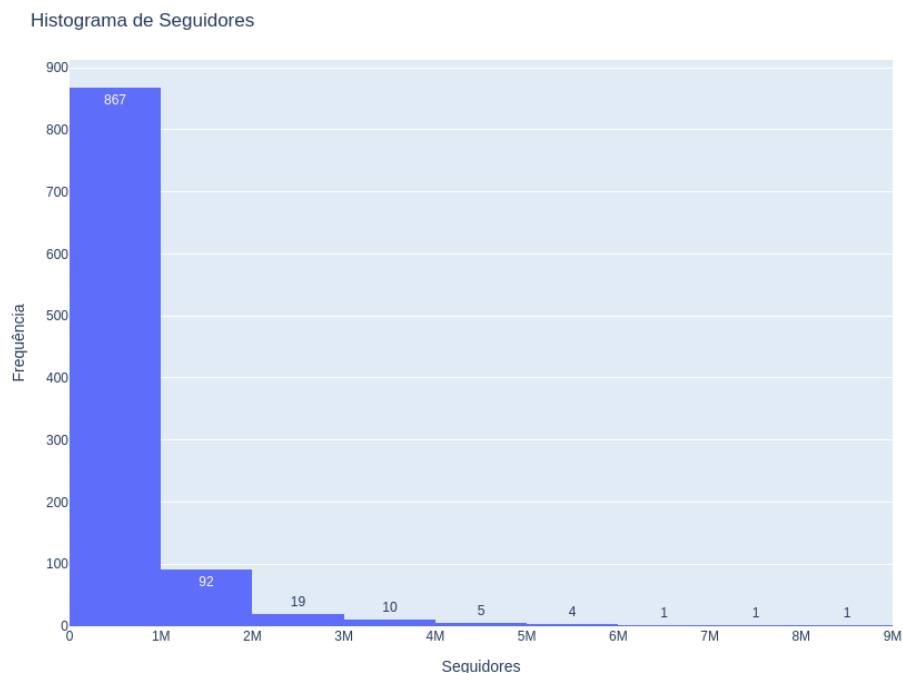


Tabela de Frequência: Followers		
Faixa	Frequência	Porcentagem (%)
(3659.999, 1120565.375]	884	88.4
(1120565.375, 2237470.75]	81	8.1
(2237470.75, 3354376.125]	16	1.6
(3354376.125, 4471281.5]	10	1.0
(4471281.5, 5588186.875]	5	0.5
(5588186.875, 6705092.25]	1	0.1
(6705092.25, 7821997.625]	2	0.2
(7821997.625, 8938904.0]	1	0.1
Total	1000	100.00

O gráfico de Histograma mostra que a distribuição dos seguidores tem uma grande concentração de streamers com números de seguidores entre 1 milhão e 2 milhões, enquanto apenas cerca de 41 realmente passam essa marca e apenas 3 passam dos 6 milhões. Podemos ver pela tabela de frequências que o maior número de seguidores foi de quase 9 milhões, o que compreende o princípio de Pareto, o qual diz que em plataformas digitais, um pequeno número de criadores gera um grande número de engajamento. Vale lembrar que esse é um top 1000 dos streamers, ou seja, em um âmbito geral essa comparação se torna ainda mais esparsa caso comparada a todos os canais da twitch.

3.1.5 Partnered

Distribuição de transmissores por parceria

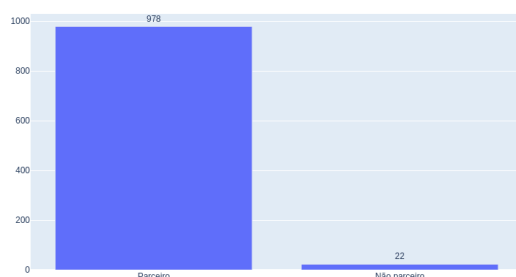
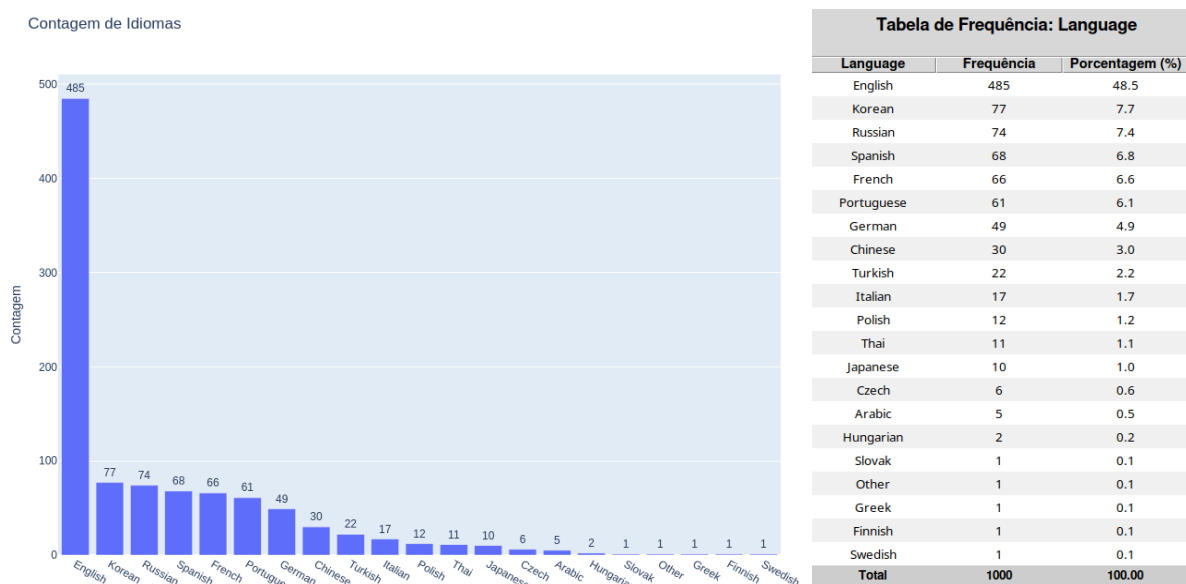


Tabela de Frequência: Partnered		
Partnered	Frequência	Porcentagem (%)
True	978	97.8
False	22	2.2
Total	1000	100.00

O programa de parceria da Twitch é destinado a streamers que comprovam consistência e relevância em seus canais. Para se qualificar, é preciso cumprir requisitos mínimos como transmitir um certo número de horas por mês e manter uma média de espectadores. Uma vez aprovado, o criador passa a receber benefícios financeiros diretos (receita de assinaturas e bits), maior customização para seu canal, acesso a ferramentas avançadas de moderação e prioridade de visibilidade nos diretórios da plataforma.

Entre os top 1000 streamers, 978 (97,8 %) são parceiros e apenas 22 (2,2 %) não possuem esse status, revelando que o selo de parceria praticamente coincide com os canais de maior audiência. Essa quase universalidade do programa de parceria indica uma forte concentração de recursos e visibilidade nas mãos desses streamers, sendo relativamente rara a presença de criadores em destaque sem o reconhecimento oficial da plataforma.

3.1.6 Language



Em relação aos dados obtidos ao verificar os diferentes idiomas dos maiores transmissores da Twitch, nota-se que há uma forte concentração na língua inglesa, com 48,5 % dos top 1000 canais transmitindo em inglês. Em seguida, porém com menor presença, o coreano (7,7 %), o russo (7,4 %), o espanhol (6,8 %), o francês (6,6 %) e o português (6,1 %). Ou seja, cerca da metade dos streamers opera em apenas um idioma, enquanto o restante se distribui em outras línguas com frequências menores. Essa assimetria indica que, embora existam criadores significativos em diversas línguas, o mercado global de streaming na Twitch é dominado por poucos idiomas majoritários, o que pode indicar prováveis barreiras de alcance das transmissões para públicos que falam línguas menos difundidas.

3.2 MEDIDAS DE RESUMO

Estatísticas Quantitativas								
Coluna	mean	std	min	25%	50%	75%	max	moda
Followers	570054.068	804413.367652	3660.0	170546.25	318063.0	624332.25	8938903.0	3660.0
WatchTime(Min	418427930.13	549635514.083	122192850.0	163189893.75	234990787.5	433739917.5	6196161750.0	122192850.0
StreamTime(mi	120515.16	85376.2013642	3465.0	73758.75	108240.0	141843.75	521445.0	51600.0
AverageViewer	4781.04	8453.68496466	235.0	1457.75	2425.0	4786.25	147643.0	1205.0

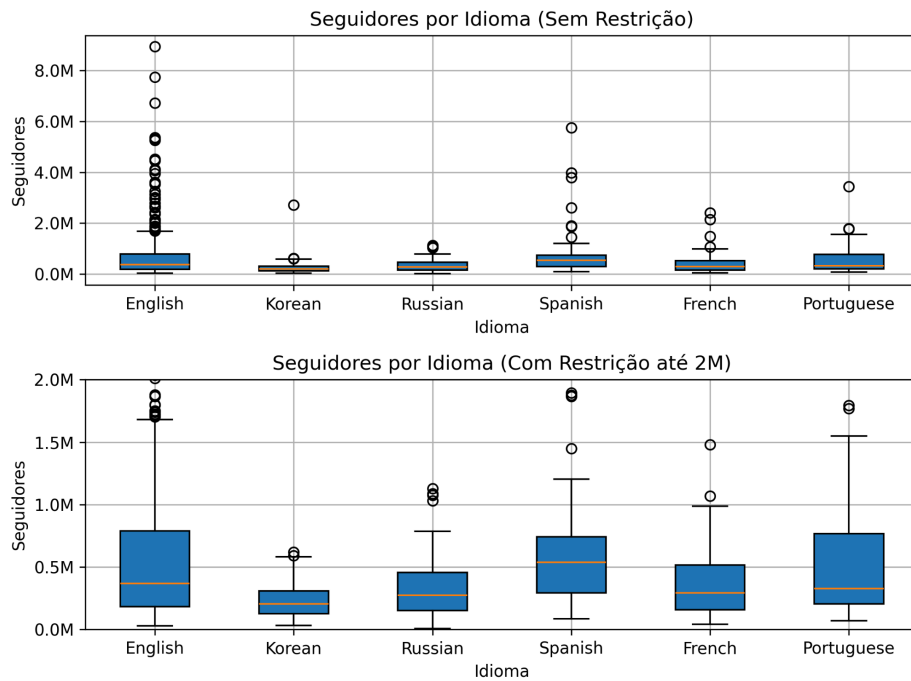
Existe um padrão claro de alta assimetria positiva nas variáveis quantitativas, principalmente Followers, WatchTime e StreamTime, a maioria dos canais tem um desempenho modesto quando comparado aos primeiros colocados, os quais acumulam números extremamente expressivos, o desvio padrão (std) elevadíssimo reforça essa concentração. Outro ponto a se observar é que a mediana (50%) é sempre menor que a média em todas as variáveis, isso sugere que os dados são “puxados” por “pontos fora da curva”. Isso reflete o *princípio de Pareto* que é muito comum em plataformas digitais: Poucos criadores geram a maior parte do engajamento. AverageViewers também mostra uma dispersão acentuada, mas com valores centrais menores.

Estatísticas Qualitativas				
Coluna	count	unique	top	freq
Partnered	1000	2	True	978
Mature	1000	2	False	770
Language	1000	21	English	485

A escolha aqui está dividida em (unique) que indica o total de valores diferentes, top (moda) e freq (frequência) tal qual aquele valor apresenta. Foi escolhido esse padrão pela descrição de frequência em cada variável já ter sido abordada nas tabelas de frequência, a tabela é bem auto-explicativa, indicando que parte esmagadora dos canais tem parceria com a plataforma, uma parte ainda significativa classifica suas transmissões como “livre para todos” e praticamente metade das transmissões são em inglês, provavelmente por acessibilidade e alcance internacional.

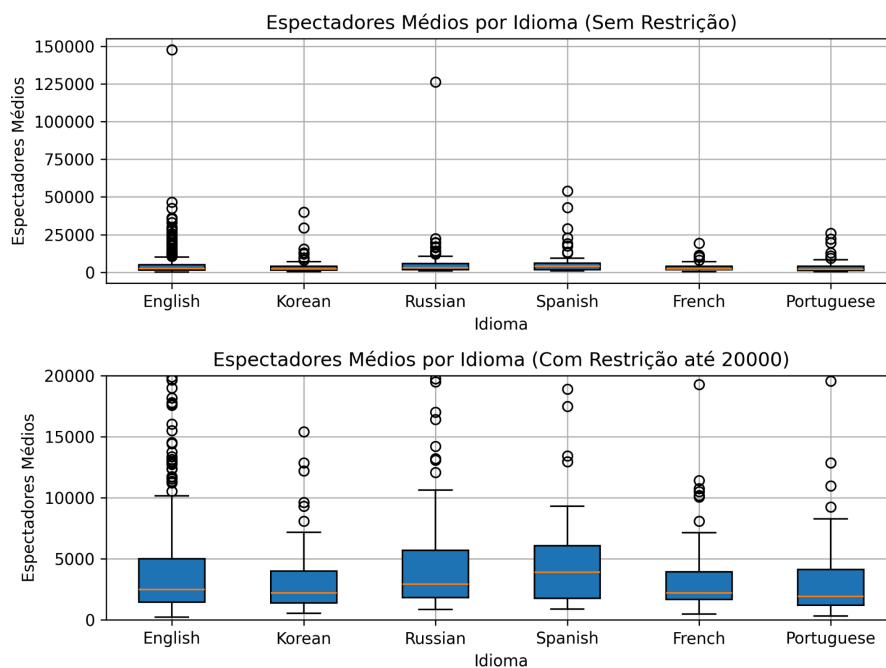
3.3 RELAÇÕES ENTRE VARIÁVEIS

3.3.1 Número de seguidores e idioma:



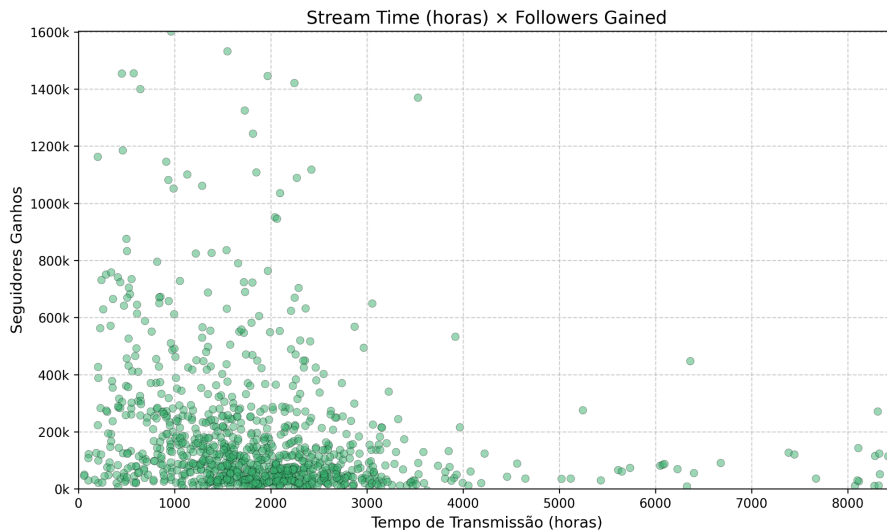
O Primeiro gráfico elaborado foi um boxplot, que representa a distribuição do número de seguidores dos canais em função do idioma. Esse gráfico foi escolhido, pois é o mais intuitivo para comparações entre categorias qualitativas e quantitativas. Ele permite observar a dispersão dos dados, mediana e presença de outliers dentro de cada um dos grupos linguísticos. Além disso, para esse par de variáveis, ele é interessante pois podemos ver que os canais que transmitem em inglês tem uma mediana de seguidores mais altas, no entanto, temos também uma maior variabilidade, indicando que há canais em inglês com números de seguidores muito alto, mas também alguns menores.

3.3.2 Média de espectadores e idioma



Pela mesma razão da relação detalhada acima, foi usado um gráfico do tipo boxplot, permitindo comparar a tendência central e a dispersão da audiência média. A comparação entre idiomas ajuda a compreender como o idioma pode influenciar o engajamento médio dos espectadores. Nele, podemos perceber no gráfico que novamente o idioma inglês é destacado, com valores medianos de audiência superiores e uma distância interquartílica longa, sugerindo menor variação na audiência.

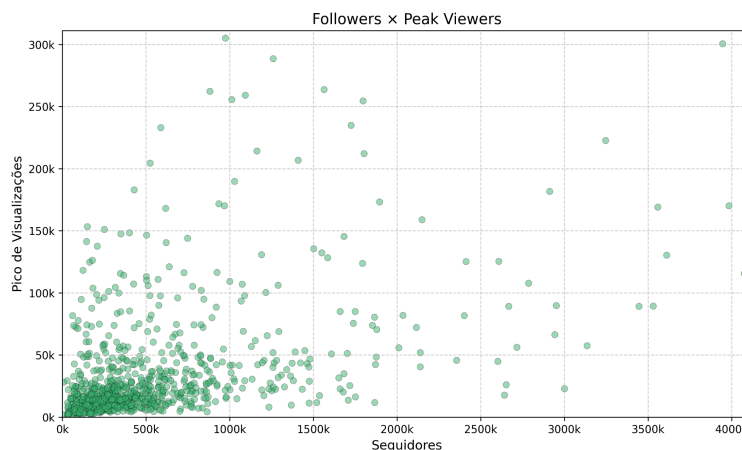
3.3.3 Tempo de transmissão e Seguidores ganhos:



Para analisarmos a relação entre o tempo de transmissão e o número de seguidores ganhos, foi utilizado um gráfico de dispersão. Acreditamos que esse gráfico é o mais apropriado pois estamos relacionando duas variáveis quantitativas, mostrando de forma direta a distribuição de pontos e presença (ou ausência de padrões).

Usando esse tipo de gráfico pode-se visualizar se há uma tendência clara - por exemplo - se canais que transmitem por mais tempo tendem a ganhar mais seguidores. Essa análise permite concluir que o tempo de transmissão é um fator relevante, mas não determinante para o crescimento do canal na plataforma.

3.3.4 Followers e Pico de espectadores:



A relação entre o número total de seguidores e o pico de visualizações foi analisada em um gráfico de dispersão, adequado por envolver duas variáveis quantitativas, que permite observar tendências, padrões e densidade de pontos em diferentes regiões do plano cartesiano. Podemos perceber tendências claras, canais com maior número de seguidores, tendem a alcançar picos mais altos de espectadores simultâneos, o que faz sentido, pois o canal tem maior base de alcance potencial.

Contudo, a dispersão dos pontos indica que a relação não é perfeitamente linear, existindo canais com muitos seguidores que não atingem picos tão altos e canais com poucos seguidores que em momentos específicos, conseguiram atingir um grande número de espectadores simultâneos.

4 CONSIDERAÇÕES FINAIS

A análise estatística dos principais canais da plataforma Twitch permitiu compreender melhor as dinâmicas e padrões do ecossistema de streaming. Compreende-se que poucos canais concentram grande parte da audiência, que o idioma é capaz de influenciar significativamente o desempenho da live e também a quase universalidade de criadores parceiros da plataforma entre os top 1000 maiores streamers. Com todos os dados e análises conseguimos entender que o sucesso de um conteúdo depende de uma combinação complexa entre visibilidade, consistência, conteúdo e engajamento.

Além disso, concluímos que a estatística é uma ferramenta poderosa para mapear essas relações e oferecer informações para que criadores entendam os desafios do mercado de streams.

5 REFERÊNCIAS

[1] Top Streamers on Twitch. Disponível em: <https://www.kaggle.com/datasets/aayushmishra1512/twitchdata?resource=download>.

Acesso em: 19 de abril de 2025.

[2] Twitch (service). Disponível em: [https://en.wikipedia.org/wiki/Twitch_\(service\)](https://en.wikipedia.org/wiki/Twitch_(service)) Acesso em: 01 de maio de 2025.

6 ANEXOS

Repositório no Github contendo os códigos utilizados durante a análise: <https://github.com/Vitor-Calegari/INE5405-ProbE>