



**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)**

ANÁLISE SOBRE OS TOP 1000 CANAIS DA TWITCH - T2

**FLORIANÓPOLIS
2025**

1 INTRODUÇÃO E OBJETIVOS

Nos últimos anos, a Twitch consolidou-se como uma das principais plataformas de streaming ao vivo no mundo, atraindo milhões de usuários e criadores de conteúdo. A análise do comportamento e desempenho dos streamers tornou-se um tema de interesse, pois, a partir dela, é possível determinar engajamento da audiência, tempo de transmissão e estratégias de crescimento de canais entre outros fatores relevantes para criadores de conteúdo da plataforma. Com base nesse cenário, o presente trabalho tem como objetivo aprofundar a análise estatística dos 1000 canais mais assistidos na Twitch durante o ano de 2020, explorando variáveis como tempo de transmissão (Stream time), pico de espectadores (Peak Viewers), média de espectadores (Average Viewers) e número de seguidores (Followers), entre outras.

Neste segundo estudo, faremos o uso de estatísticas descritivas, estimação de parâmetros e testes de hipóteses para investigar padrões e associações presentes nos dados. Além disso, serão levantadas hipóteses relacionadas a médias, proporções, correlações e aos parâmetros de um modelo de regressão linear. Por meio dessas análises, buscamos entender mais profundamente sobre o funcionamento do ecossistema da plataforma Twitch.

2 MATERIAIS E MÉTODOS

Para desenvolvimento das hipóteses, nos apoiamos em algumas variáveis já discutidas no trabalho anterior. Em primeiro lugar, podemos citar as variáveis qualitativas trabalhadas, como Partnered e Language. A primeira é um valor booleano que faz referência ao programa de parceiros da plataforma, indicando que o streamer tem uma série de vantagens de engajamento, mas também tem alguns deveres à cumprir com a plataforma. Já a segunda, mostra qual é a linguagem predominante nas transmissões de um canal.

Além dessas, temos também as variáveis quantitativas: Followers, que representa o número absoluto de seguidores de um canal; Followers Gained, que representam a quantidade de seguidores que o streamer ganhou no período analisado; Average Viewers, indicando a média de público que uma transmissão teve enquanto estava no ar; Watchtime, que indica a quantidade de tempo que um criador de conteúdo ficou ao vivo na plataforma; e por fim, Peak Viewers, que indica o máximo de espectadores simultâneos na transmissão durante o período analisado.

3 RESULTADOS E DISCUSSÕES

3.1 ESTATÍSTICAS DESCRITIVAS

3.1.1 Programa de Parceria

Patterned	Proporção (%)	IC inferior (95%)	IC superior (95%)
True	93.33	84.41	100
False	6.67	0	15.59

Tabela 1: Estatísticas descritivas de Partnered em uma amostra de 30 streamers

A tabela apresenta a distribuição da variável “Partnered” em uma amostra de 30 streamers do top 1000 da plataforma Twitch. Os resultados indicam que a grande maioria dos streamers analisados possui um perfil “Patterned” (93.33%), com um intervalo de confiança de 95% variando entre 84.31% e 100%, Por outro lado, apenas 6.67% dos streamers não se enquadram nessa categoria, com um intervalo de confiança de 95% entre 0% e 15.59%.

3.1.2 Número de Seguidores

Estatística	Valor
Média amostral	373804.63
Desvio padrão amostral	324198.73
IC inferior (95%)	252746.73
IC superior (95%)	494862.43

Tabela 2: Estatísticas descritivas de Followers em uma amostra de 30 streamers

A tabela apresenta estatísticas descritivas de uma amostra composta por 30 streamers dentre os 1000 maiores da Twitch em número de seguidores. A média amostral encontrada foi de 373.804,63 seguidores, indicando que, em média, esses streamers possuem aproximadamente 374 mil seguidores. O desvio padrão amostral é de 324.198,73, o que evidencia uma grande variação no número de seguidores entre os streamers da amostra, sugerindo que alguns possuem audiências significativamente maiores do que outros. O intervalo de confiança de 95% para a média populacional foi calculado entre 252.746,73 e 494.862,43 seguidores, o que significa que podemos afirmar, com 95% de confiança, que a verdadeira média de seguidores dos 1000 maiores streamers da Twitch está dentro desse intervalo. Esses dados reforçam a ideia de que há uma grande

desigualdade na distribuição de seguidores dentro da plataforma, característica comum em redes sociais.

3.1.3 Espectadores Médios

Estatística	Valores
Média amostral	4201.00
Desvio padrão amostral	5078.24
IC inferior (95%)	2304.75
IC superior (95%)	6097.25

Tabela 3: Estatísticas descritivas de Average Followers em uma amostra de 30 streamers

A tabela apresenta estatísticas descritivas de uma amostra composta por 30 streamers da Twitch, focando na média de espectadores simultâneos (average viewers). A média amostral observada foi de 4.201 espectadores, indicando que, em média, esses streamers atraem cerca de 4.200 pessoas assistindo ao vivo. O desvio padrão amostral foi de 5.078,24, o que revela uma grande variabilidade no número de espectadores entre os streamers analisados. O intervalo de confiança de 95% para a média populacional foi calculado entre 2.304,75 e 6.097,25 espectadores, o que nos permite afirmar, com 95% de confiança, que a verdadeira média de espectadores dos maiores streamers da plataforma está dentro desse intervalo. Esses dados reforçam que, assim como no número de seguidores, há uma distribuição bastante desigual também no número de espectadores simultâneos, característica típica das grandes plataformas de streaming.

3.1.4 Número de Seguidores Ganhos

Estatística	Valores
Média amostral	110471.87
Desvio padrão amostral	105009.29
IC inferior (95%)	71260.76
IC superior (95%)	149682.98

Tabela 4: Estatísticas descritivas de Followers Gained em uma amostra de 30 streamers

A tabela apresenta estatísticas descritivas de uma amostra composta por 30 streamers da Twitch, followers Gained (seguidores conquistados), a média amostral foi de 110.471,87 seguidores, indicando que, em média, cada streamer da amostra ganhou

aproximadamente 110 mil seguidores no período analisado. O desvio padrão foi de 1.050.009,29, evidenciando uma variação extremamente alta na quantidade de seguidores ganhos, o que sugere que alguns streamers obtiveram picos de crescimento muito superiores à média. O intervalo de confiança de 95% varia entre 71.260,76 e 149.682,98 seguidores, o que significa que, com 95% de confiança, a média real de seguidores ganhos pela população está dentro desse intervalo.

3.1.5 Tempo em Transmissão

Estatística	Valores
Média amostral	95639.00
Desvio padrão amostral	44470.76
IC inferior (95%)	79033.34
IC superior (95%)	112244.66

Tabela 5: Estatísticas descritivas de Watchtime em uma amostra de 30 streamers

A tabela apresenta estatísticas descritivas de uma amostra composta por 30 streamers da Twitch, Stream Time (tempo total de transmissão), a média amostral foi de 95.639,00 minutos, indicando que, em média, os streamers transmitiram cerca de 95 mil minutos no período analisado. O desvio padrão foi de 44.470,76 minutos, mostrando uma variação considerável entre os streamers em relação ao tempo dedicado às transmissões. O intervalo de confiança de 95% para essa métrica foi calculado entre 79.033,34 e 112.244,66 minutos, o que nos permite afirmar, com 95% de confiança, que a média real de tempo de transmissão da população se encontra nesse intervalo. Ambos os resultados reforçam que há grande desigualdade no crescimento e no tempo de atividade entre os streamers, um padrão comum em plataformas de grande escala.

3.1.6 Linguagem

Linguagem	Proporção (%)	IC 95% inferior	IC 95% Superior
English	53.33	35.48	71.19
Russian	16.67	3.33	30
French	6.67	0	15.59
German	3.33	0	9.76

Chinese	3.33	0	9.76
Greek	3.33	0	9.76
Polish	3.33	0	9.76
Korean	3.33	0	9.76

Tabela 6: Estatística descritiva de Language em uma amostra de 30 streamers

A tabela apresenta estatísticas descritivas sobre o idioma utilizado por uma amostra de 30 streamers da Twitch, selecionados entre os 1000 maiores da plataforma. O idioma mais frequente foi o inglês, utilizado por 53,33% dos streamers da amostra, com um intervalo de confiança de 95% entre 35,48% e 71,19%, evidenciando a predominância do inglês na Twitch. O segundo idioma mais frequente foi o russo, com 16,67%, seguido por francês (6,67%), alemão, chinês, grego, polonês e coreano, todos com proporção de 3,33%. Os intervalos de confiança dos idiomas menos frequentes mostram sobreposição e ampla incerteza, o que sugere que, fora o inglês e o russo, a presença dos outros idiomas entre os maiores streamers é bem menos consistente e pode variar bastante. Esses resultados reforçam o domínio do inglês como principal idioma na plataforma, mas também indicam alguma diversidade linguística entre os principais criadores de conteúdo.

3.2 LEVANTAMENTO DE HIPÓTESES

3.2.1 Hipótese envolvendo média

Uma das variáveis que chama atenção no conjunto de dados, é a *partnered*, isso, pois grande parte dos streamers são parceiros da plataforma. Então, pode surgir a dúvida: “Será que criadores de conteúdo associados a plataforma tem mais seguidores dos que aqueles que não são associados?”. Então, trazemos como hipótese nula (H_0) a igualdade entre a média de seguidores dos dois grupos ($\mu_1 = \mu_2$) e como hipótese alternativa (H_1) que a média dos seguidores parceiros é maior ($\mu_1 > \mu_2$).

Para averiguar a hipótese, usaremos o Teste de hipótese para média de duas populações unilateral à direita pois temos ($\mu_1 > \mu_2$). Da análise imediata dos dados pudemos constatar que para:

Tipo de “Partnered”	Número de streamers	Média de seguidores	Desvio padrão
True (parceiro)	978	575.410	808.935
False (não parceiro)	22	331.943	527.037

Tabela 7: Média e Desvio para parceiros e não parceiros da plataforma

Temos uma estatística $t = 2.112$ e um p-valor 0.02283. Como o p-valor obtido é menor que 0.05 (referente à significância), podemos rejeitar a hipótese nula ao nível de 5% de significância. Isso indica que há evidências estatísticas para afirmar que, em média, streamers parceiros têm mais seguidores do que aqueles que não tem parceria com a plataforma.

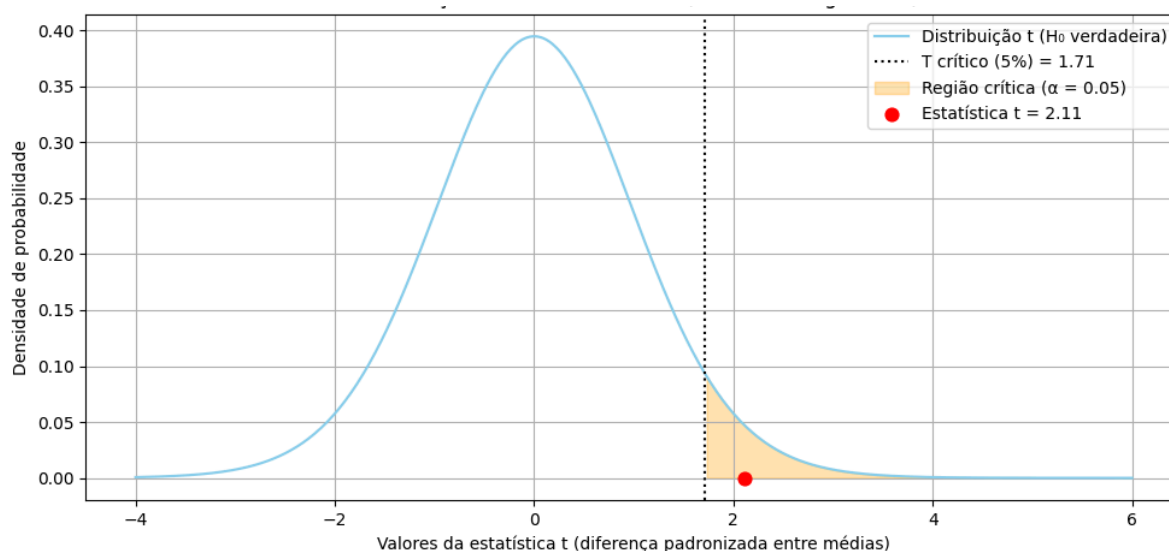


Figura 1: Visualização do teste T Unilateral (Parceria e Seguidores)

No entanto, vale destacar que o desvio padrão de ambos os grupos é muito elevado, o que é esperado em um conjunto de dados com muitos outliers (criadores que acumulam grandes audiências enquanto a maioria se concentra em patamares mais modestos). Além disso, consideramos que apesar de termos essa diferença estatisticamente significativa, ela não é “enorme” em termos absolutos, pois novamente, temos de considerar a distribuição assimétrica dos dados.

Concluindo, vemos que a partir do teste estatístico realizado, streamers parceiros da Twitch apresentam em média, um número maior de seguidores. Essa conclusão segue a lógica da plataforma, que oferece parceria aos canais com maior visibilidade, e ao mesmo tempo, oferecem benefícios que potencializam o crescimento do canal.

3.2.2 Hipótese envolvendo correlação

A primeira entrega do presente trabalho apresenta uma possível relação entre os seguidores e espectadores médios. Então, na presente seção, iremos verificar se essas duas variáveis estão de fato correlacionadas a partir de um teste de hipótese.

Iniciaremos definindo nossa hipótese nula (H_0), $\rho = 0$ (não há correlação linear entre seguidores e espectadores) e nossa hipótese alternativa (H_1), $\rho \neq 0$ (há correlação linear

entre seguidores e espectadores). Então, será feito um teste de hipótese para o coeficiente de correlação de Pearson, bilateral, testando se a correlação é diferente de 0 positivamente ou negativamente.

Realizando os cálculos com os dados disponíveis, temos que o valor da correlação de Pearson é 0.428, o que indica uma correlação positiva. Além disso, temos um p-valor muito próximo à 0 o que é uma evidência muito forte para reprovar H_0 (que diz que não há correlação) com um alto nível de confiança ($\alpha = 0,05$).

Estatística	Valor
Coeficiente de correlação de Pearson	0.428
p-valor	7.074502598363817e-46

Tabela 8: Coeficiente de correlação e p-valor retirados do código produzido para o trabalho

Esse resultado mostra que quanto mais seguidores um streamer tem, maior tende a ser a média de audiência. Abaixo, segue um gráfico de dispersão, que mostra a correlação entre as variáveis.

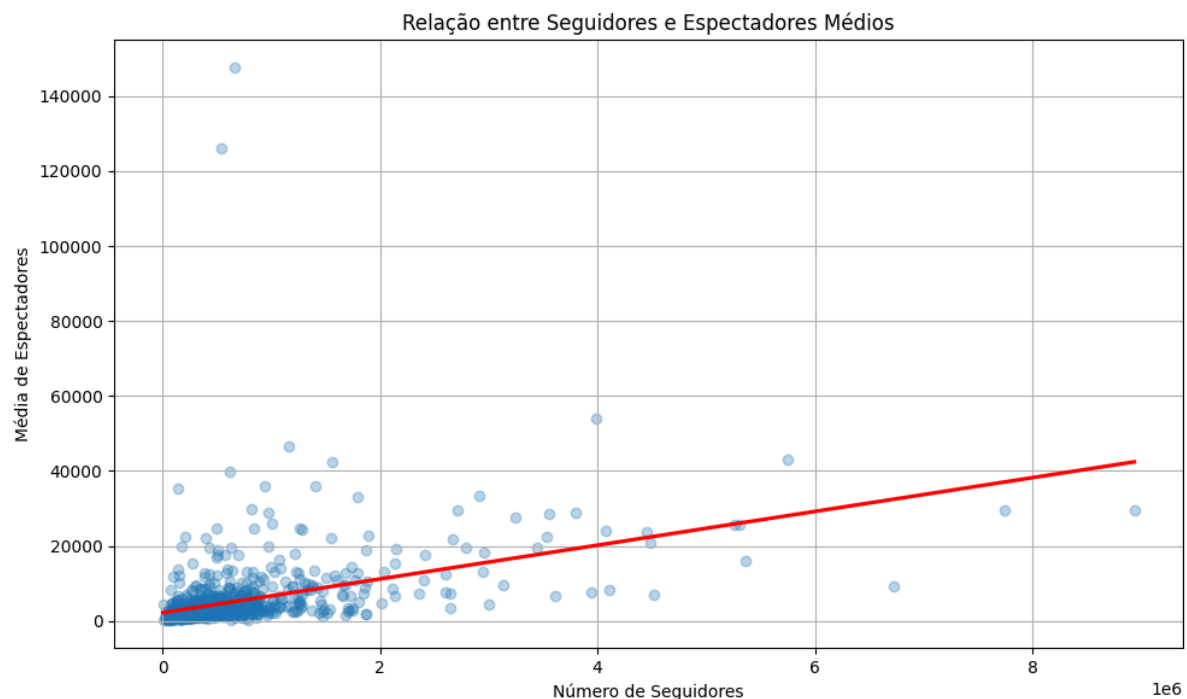


Figura 2: Gráfico que mostra a correlação entre número de seguidores e média de espectadores

Sendo assim, podemos notar que existem evidências estatísticas fortes que mostram que existe uma correlação positiva entre o número de seguidores e a média de

espectadores. Streamers com mais seguidores tendem a atrair mais audiência ao vivo, que também está de acordo com o comportamento esperado de plataformas de lives, nas quais o engajamento e base de seguidores se reforçam mutuamente.

3.2.3 Hipótese envolvendo proporção

Nesta etapa do trabalho, investigamos se há evidências estatísticas de que a proporção de transmissões com idioma principal em português é maior do que a proporção de transmissões em espanhol. Para isso, formulamos a seguinte hipótese nula (H_0): a proporção de transmissões em português é menor ou igual à proporção em espanhol ($p_1 \leq p_2$). A hipótese alternativa (H_1) é que a proporção de transmissões em português é maior do que a de transmissões em espanhol ($p_1 > p_2$). Utilizamos um teste de proporção para duas populações, com nível de significância $\alpha = 0,05$, e assumimos distribuição normal aproximada devido ao tamanho amostral relativamente grande.

Idioma	Contagem	Proporção (estimada)
Português	61	6,1%
Espanhol	68	6,8%

Tabela 9: Resultados do teste de hipótese sobre proporção entre idiomas

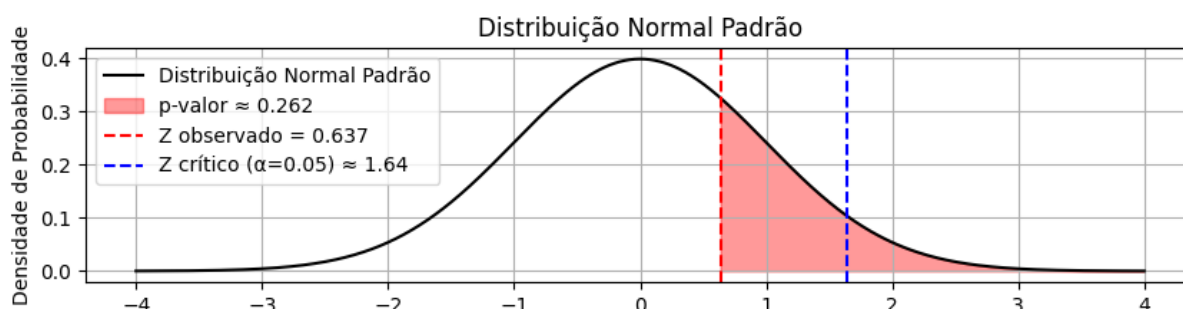


Figura 3: Resultados do teste de hipótese sobre proporção entre idiomas representados graficamente.

Os testes resultaram no valor para Z sendo 0,6372 e o p-valor sendo 0,2620. Com um nível de significância de 5%, observamos que p-valor é maior do que a significância, portanto não há evidência estatística suficiente para afirmar que a proporção de canais que transmitem em português é maior do que a de transmissões em espanhol.

3.2.4 Hipótese envolvendo parâmetros do modelo de regressão

Para essa etapa, primeiramente, escolhemos criar um modelo de regressão o qual modela o número máximo de espectadores que o canal obteve durante o período com base no número de seguidores que o canal tem. Segue representação gráfica do modelo na Figura 4.

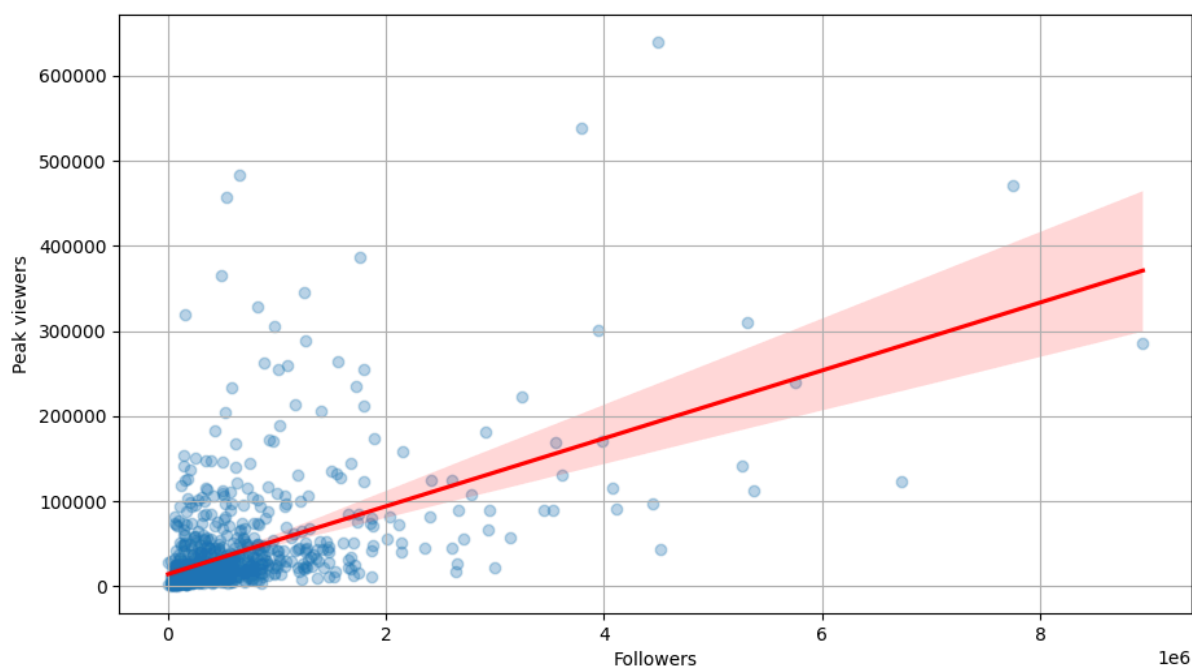


Figura 4: Modelo de regressão linear relacionando o Pico de Espectadores(eixo y) com Número de Seguidores(eixo x).

Com base no modelo apresentado, decidimos verificar se o número de seguidores, parâmetro β_1 , é realmente relevante para estimar a audiência mais alta já alcançada. Para isso, formulamos a seguinte hipótese nula (H_0): número de seguidores não influencia a audiência mais alta já alcançada ($\beta_1 = 0$). A hipótese alternativa escolhida (H_1) é que o número de seguidores influencia a audiência mais alta já alcançada ($\beta_1 \neq 0$).

Para testar se β_1 é relevante ou não para o modelo, utilizamos a análise de variância sobre o modelo para verificar sua utilidade. Já que Y é estimado apenas com uma variável, o teste verifica se ela é relevante ou não. Durante o teste, foi observado o valor 395,1 para a estatística F e o valor 2.52e-74 para p-valor. Com um nível de significância de 5%, observamos que p-valor é menor do que a significância, portanto podemos rejeitar H_0 , ou seja, verificamos que o número de seguidores influencia o pico de audiência mais alto já alcançado pelo criador.

4 CONSIDERAÇÕES FINAIS

O presente trabalho realizou uma análise estatística detalhada sobre os top 1000 streamers da Twitch, revelando um cenário de grande concentração de popularidade e desigualdade na distribuição de seguidores, espectadores e tempo de transmissão. As discrepâncias observadas entre a amostra de 30 streamers e os dados populacionais foram significativas, principalmente nos seguidores ganhos e no tempo total de stream, sugerindo que a amostra pode não ter capturado os streamers com maior influência na plataforma. Ao longo do levantamento das hipóteses, verificamos que streamers parceiros possuem, em média, mais seguidores do que aqueles que não são parceiros, resultado confirmado com evidências estatísticas sólidas. Também foi possível identificar uma correlação positiva entre o número de seguidores e a média de espectadores simultâneos, o que reforça a lógica de crescimento orgânico dentro da plataforma, onde uma base de seguidores maior tende a gerar uma audiência média mais elevada. Por outro lado, não encontramos evidências suficientes para afirmar que há maior proporção de transmissões em português em comparação com o espanhol, demonstrando que, neste aspecto, as línguas analisadas possuem presença semelhante na Twitch. Por fim, os testes de regressão mostraram que o número de seguidores é um fator relevante para prever o pico de audiência alcançado por um canal, confirmando a influência direta da base de seguidores no alcance máximo das transmissões. De forma geral, os resultados obtidos alinham-se com o comportamento esperado em grandes plataformas digitais, onde a popularidade tende a se concentrar e a gerar ciclos de crescimento contínuo para os streamers mais influentes.

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>