

CENTRO UNIVERSITÁRIO ESTÁCIO DE SERGIPE
TECNÓLOGO EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

VINÍCIUS ALMEIDA ANDRADE

ENTENDENDO AS RELAÇÕES DE DADOS ENTRE CONTEÚDO DA NETFLIX

ARACAJU

2025

Sumário

1. DIAGNÓSTICO E TEORIZAÇÃO	3
1.1 Identificação das partes interessadas e parceiros	3
1.2 Problemática e/ou problemas identificados	3
1.3 Justificativa	3
1.4 Objetivos/resultados/efeitos a serem alcançados (em relação ao problema identificado e sob a perspectiva dos públicos envolvidos).....	4
1.5 Referencial teórico (subsídio teórico para propositura de ações da extensão)	5
2. PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO	6
2.1 Plano de trabalho (usando ferramenta acordada com o docente)	6
2.2 Grupo de trabalho (descrição da responsabilidade de cada membro).....	6
2.3 Metas, critérios ou indicadores de avaliação do projeto	6
2.4 Recursos previstos.....	8
2.5 Detalhamento técnico do projeto	8
3. ENCERRAMENTO DO PROJETO	9
3.1 Relato de Experiência Individual (Pontuação específica para o relato individual).....	9
3.1.1 Contextualização.....	10
3.1.2 Resultados e discussão:	10
3.1.3 Reflexão aprofundada	20
3.1.4 Considerações finais	21
ANEXOS	22

1. DIAGNÓSTICO E TEORIZAÇÃO

1.1 Identificação das partes interessadas e parceiros

As partes interessadas neste projeto incluem a Netflix, seus assinantes e um público mais amplo que procura conteúdo, como filmes e séries. O projeto contém dados de 2008 a 2021 que fornece insights sobre tendências, preferências e padrões do público-alvo, resultando em um conteúdo cada vez mais adaptado, por meio de inclusão e da produção de conteúdos, ao que seus assinantes procuram na plataforma.

1.2 Problemática e/ou problemas identificados

A dificuldade em extrair valor e insights de um catálogo massivo de dados da Netflix é a problemática central que este projeto se propõe a abordar. O tema foi pensado para pessoas apaixonadas por filmes e séries, que muitas vezes se encontram perdidas devido à sobrecarga de informação causada pela vasta quantidade de conteúdo disponível, sem indicações claras ou dados organizados para suas escolhas. Essa complexidade também representa um grande desafio para a própria Netflix, que precisa transformar seus dados brutos de catálogo em informações úteis para decidir o que comprar/produzir, como divulgar e o que criar futuramente. Diante disso, o projeto foi elaborado para fazer o levantamento de dados e encontrar correlações e padrões, beneficiando a experiência do usuário e as estratégias da Netflix.

1.3 Justificativa

Este projeto se justifica por sua relevância acadêmica, aplicando e aprofundando conhecimentos do curso em Big Data, PySpark e análise de dados. A manipulação de um catálogo massivo como o da Netflix desenvolve competências essenciais em ciência de dados, cruciais no mercado atual. O código, ao buscar

padrões em volumes complexos, é um exemplo direto da aplicação de teorias e ferramentas.

Além disso, a motivação surgiu do interesse pessoal em streaming e em solucionar desafios digitais. A sobrecarga de informação na escolha de conteúdo é real para muitos, e a paixão por filmes/séries, junto ao desejo de aplicar conhecimentos para gerar valor e facilitar a vida do usuário, impulsionou o projeto. Transformar dados brutos em descobertas úteis para o usuário e a estratégia para a Netflix.

1.4 Objetivos/resultados/efeitos a serem alcançados (em relação ao problema identificado e sob a perspectiva dos públicos envolvidos)

- Analisar a composição geral do catálogo, apresentando os **números totais de filmes e séries**. Isso proporcionará ao usuário uma compreensão inicial da variedade disponível e auxiliará a Netflix a entender melhor seu inventário principal.
- Mapear a dinâmica de crescimento do catálogo através da **data de adição dos conteúdos (por mês, por ano e por mês/ano)**. Esse insight é crucial para a Netflix no planejamento de aquisições e lançamentos, enquanto permite ao usuário observar a frequência de renovação da plataforma.
- Identificar os padrões de gêneros e sua distribuição, explorando **a quantidade e quais os mais comuns de forma geral, os mais recorrentes por ano, por tipo (filme ou série) e em cada país**. Esses dados ajudarão o usuário a encontrar conteúdos de seu interesse e a Netflix a tomar decisões sobre novas produções e aquisições baseadas nas preferências do mercado.
- Determinar os **países com maior contribuição no catálogo**, oferecendo à Netflix um panorama geográfico da produção e ao usuário a possibilidade de explorar conteúdos de diversas origens.
- Explorar as **relações de colaboração entre diretores e atores**, quantificando quantas vezes trabalharam juntos. Essa análise fornece à Netflix insights sobre equipes criativas de sucesso e ao usuário uma forma de descobrir conteúdos por meio de parcerias conhecidas.

- Detalhar a participação de atores e diretores, apresentando **quantos filmes e quais filmes o ator participou, e quantos e quais filmes o diretor dirigiu**. Isso enriquece a capacidade do usuário de seguir o trabalho de seus profissionais favoritos e oferece à Netflix um panorama detalhado do portfólio de seus talentos.

1.5 Referencial teórico (subsídio teórico para propositura de ações da extensão)

Data Mining (Mineração de dados) é o processo de descobrir padrões úteis em grandes conjuntos de dados, como big data. Neste projeto, diversas técnicas de data mining foram empregadas para explorar e extrair informações relevantes do conjunto de dados da Netflix, visando uma melhor compreensão do conteúdo e de suas relações.

- Clustering (Agrupamento): Entender a distribuição do conteúdo, por exemplo agrupar filmes, series, gênero de cada um deles, atores...
- Análise associativa: Identifica regras que descrevem como itens estão relacionados para melhor compreensão dos dados, por exemplo: contar a frequência de colaborações entre atores e diretores.
- Análise de sequências: A análise da evolução da adição de conteúdo ao longo do tempo é uma análise de sequência, buscando padrões temporais.

Machine learning (Aprendizado de máquinas) é o processo utilizado para que sistemas aprendam com dados, identificando padrões e fazendo previsões e/ou decisões. No projeto foi utilizando a base desta etapa, fazendo a limpeza e organização dos dados (divisão, explosão, padronização), o que é crucial para melhor aprendizado.

Frameworks de Big Data são plataforma de software servem para guardar, processar e analisar uma quantidade enorme de dados. Elas oferecem ferramentas e jeitos de programar que ajudam a lidar com dados complexos, permitindo que quem trabalhar com dados e softwares desenvolva APIs. No projeto foi utilizado PySpark uma API do Apache Spark.

2. PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO

2.1 Plano de trabalho (usando ferramenta acordada com o docente)

Recursos Necessários:

- Acesso a computadores com software para análise estatística (Python, SQL, Spark, Google Colab).
- Acesso à internet.
- Acesso ao GitHub para coletar dados públicos.

2.2 Grupo de trabalho (descrição da responsabilidade de cada membro)

Vinícius (Desenvolvedor, Analista de dados e Relator)

- Coordenação geral do projeto acompanhando o progresso, definindo prazos e metas.
- Obtenção, limpeza, organização do conjunto de dados da Netflix.
- Desenvolvimento do código PySpark com SQL para realizar a análise dos dados.
- Implementação das análises específicas (atores, diretores, gêneros, tipo, países, datas de adição, combinações).
- Consolidação dos resultados em arquivo.
- Criação de gráficos e visualizações para representar os resultados com Power BI.
- Elaboração do texto escrito do trabalho.
- Preparação da apresentação oral do trabalho.
- Revisão final do trabalho escrito e da apresentação.

2.3 Metas, critérios ou indicadores de avaliação do projeto

- Meta 1: Analisar a distribuição e as tendências do conteúdo da Netflix, identificando os gêneros mais populares, os países com maior produção e a evolução do catálogo ao longo do tempo.
- Meta 2: Investigar as relações e colaborações na produção de conteúdo, como a frequência de trabalho entre atores e diretores.
- Meta 3: Organizar e apresentar os dados de forma clara e acessível, auxiliando os usuários na escolha de filmes e séries.
- Meta 4: Desenvolver um processo de análise de dados, utilizando ferramentas como PySpark.
- Meta 5: Consolidar e disponibilizar os resultados da análise em formatos adequados para uso em outras aplicações ou ferramentas, por exemplo Power BI.

Critérios de Avaliação:

- Relevância: A análise deve abordar questões importantes e fornecer insights úteis sobre o conteúdo da Netflix.
- Precisão: Os resultados da análise devem ser precisos.
- Completude: A análise deve demonstrar o que há de relevante nos dados, fornecendo assim uma visão ampla.
- Clareza e Organização: Os dados e os resultados devem ser apresentados de forma clara, organizada e fácil de entender.
- Impacto: O processo de análise deve gerar impacto na forma de escolha do público-alvo e no desenvolvimento do catálogo.

Indicadores de Avaliação:

- Quantidade de dados analisados: A pesquisa deve utilizar uma grande quantidade de dados abertos, com alta qualidade.
- Número de análises: Foram feitas análises dentro de cada um destes núcleos: gêneros, atores, diretores e países. Gerando resultados relevantes para escolhas dos assinantes e da Netflix.
- Qualidade dos arquivos gerados: Os arquivos foram gerados com formatação adequada e sem perda de dados.
- Número de perguntas de pesquisa respondidas: Todas as perguntas de pesquisa definidas foram respondidas de forma satisfatória.

2.4 Recursos previstos

- Acesso à internet
- Repositório do GitHub para obter dados abertos (https://raw.githubusercontent.com/jonates/opendata/refs/heads/master/netflix/netflix_titles.csv).
- Notebook no Google Colab para desenvolvimento e análise de dados.
- Google Drive para armazenamento do código desenvolvidos no Google Colab.
- Spark para processamento e manipulação de dados em escala.
- PySpark API do spark que facilita a interação com o Spark utilizando linguagens como Python ou SQL.

2.5 Detalhamento técnico do projeto

A solução foi desenvolvida através de transformações e agregações dos dados brutos que são lidos, gerando assim informações relevantes sobre o que foi analisado como números e tendências. Este projeto tem como intuito a limpeza e organização dos dados facilitando o entendimento das informações geradas e foi feito com os seguintes passos:

- Coleta de dados:
 - Obter dados sobre os títulos da Netflix baixados através do GitHub em formato CSV.
- Limpeza e organização dos dados:
 - Os dados brutos obtidos no GitHub podem ser confusos, por estar os dados desorganizados principalmente para fazer as análises pretendidas no projeto um exemplo disso é o elenco que estava descrito como country no dado bruto, o que foi necessário utilizar virgula como separador para obter os atores.
- Linguagens de programação:
 - Foi utilizado as linguagens Python e SQL que são umas das principais e são amplamente usadas para manipular grandes volumes de dados.
- Análise:

- Com os dados limpos e organizados foram feitas as análises com as linguagens como:
 - Contabilizar filmes e séries.
 - Analisar a frequência dos gêneros mais comuns.
 - Verificar quais países produzem mais conteúdos e identificar quais gêneros mais produzidos pelos mesmos.
 - Identificar atores e diretores que trabalham juntos com frequência.
 - Ver a variação ao longo do tempo na adição de conteúdos na plataforma.
- Visualização dos dados:
 - Para os resultados ficarem de fácil entendimento para tomar as decisões, o código salva os dados analisados em arquivos CSV e também os combina em um arquivo Excel, que pode ser aberto em programas como o Power BI para criar dashboards.
- Interpretação dos resultados:
 - A etapa final é entender através da análise o que os números e gráficos mostram.
- **Fluxo de Dados:**
 - O fluxo de dados começa com a coleta dos dados, logo após a leitura do arquivo, seguido por etapas de transformação, limpeza e organização com PySpark manipulando com as linguagens de programação para calcular números ex.: filmes/séries, extrair informações como datas, manipular dados ex.: gêneros e países, e relacionar colunas ex.: atores e diretores. Os resultados então são salvos em arquivos CSV e consolidados em um arquivo Excel.

3. ENCERRAMENTO DO PROJETO

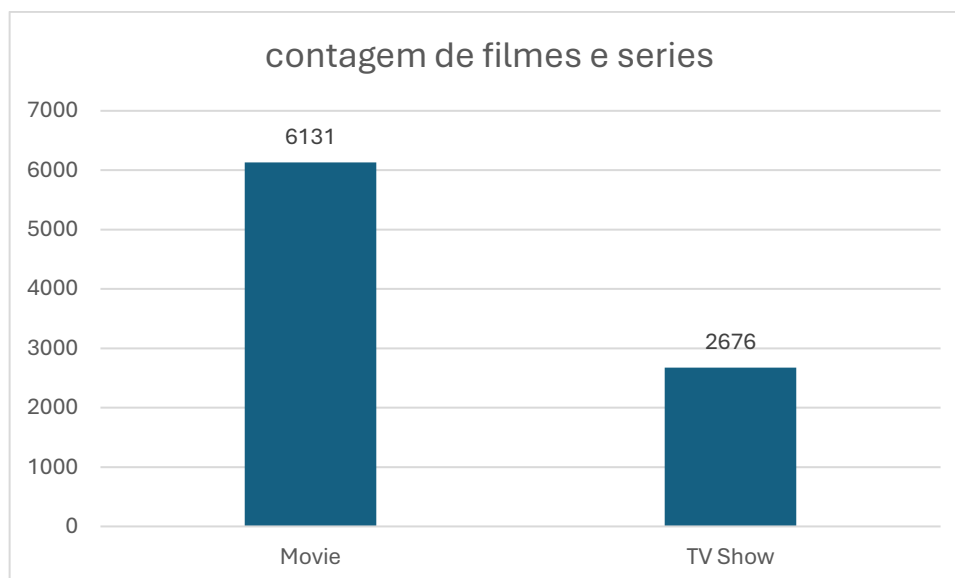
3.1 Relato de Experiência Individual (Pontuação específica para o relato individual)

3.1.1 Contextualização

O projeto consistia em analisar um conjunto de dados abertos, para retirar informações relevantes, escolhi trabalhar com dados públicos da Netflix, que é uma gigante e pioneira no mundo dos streamings, foi desafiador, pois estava lidando com grande parte de ferramentas e definições novas, me fez ter uma perspectiva diferente sobre análise de dados, onde tive uma apesar de desafiadora, uma ótima experiência. Como único membro do grupo, fui responsável por todas as etapas do projeto, desde a definição dos objetivos até a apresentação dos resultados.

3.1.2 Resultados e discussão:

Minha expectativa inicial foi de aprofundar os conhecimentos nas ferramentas específicas para de dados como Apache Spark e de linguagens como Python e SQL e basicamente esperava buscar padrões e como os dados se relacionavam, observei a importância da qualidade e confiabilidade dos dados logo ao procurar os dados abertos, após conseguir o conjunto de dados e finalmente começar a trabalhar com ele, vieram o que eu considero os principais desafios que foi o desenvolvimento, onde utilizei o Google Colab , tive uma experiência breve com o Spark, porém a utilização da API PySpark trabalhando com linguagens as quais já estava mais familiarizado , suavizou o impacto de primeiro momento e a limpeza dos dados que é o que mais se repetiu no desenvolvimento do código, onde se exige um esforço para lidar com inconsistências e valores ausentes. Mas a medida que fui superando as dificuldades, os estudos dos dados foram resultando em informações as quais eram respostas que eu procurava no início do projeto.



Títulos por filmes e series

Rótulos de Linha

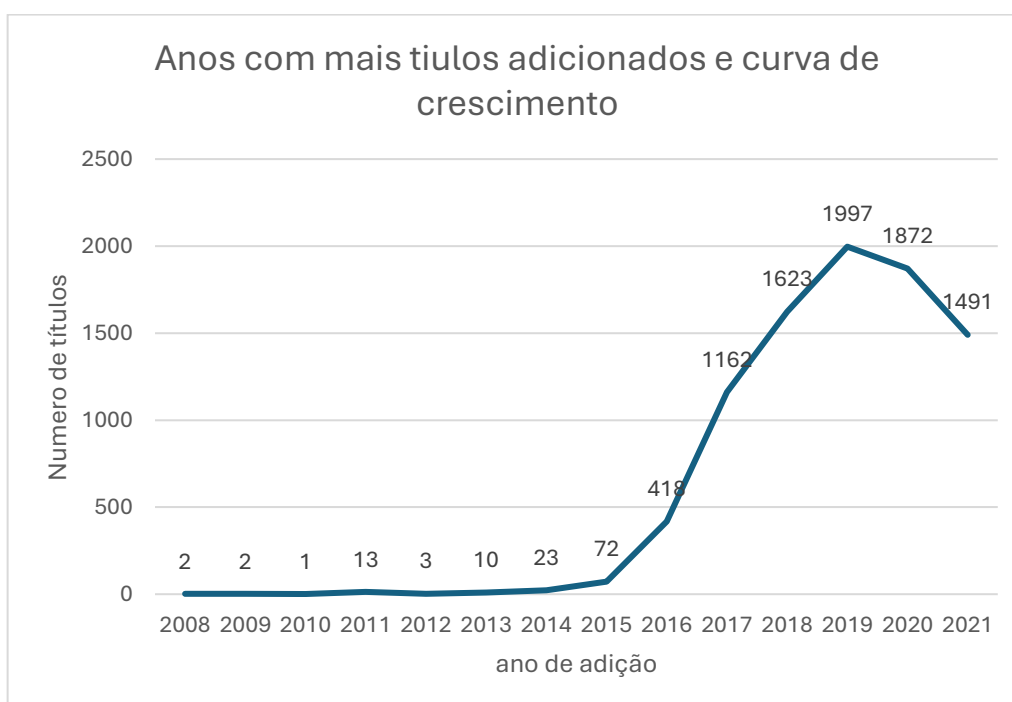
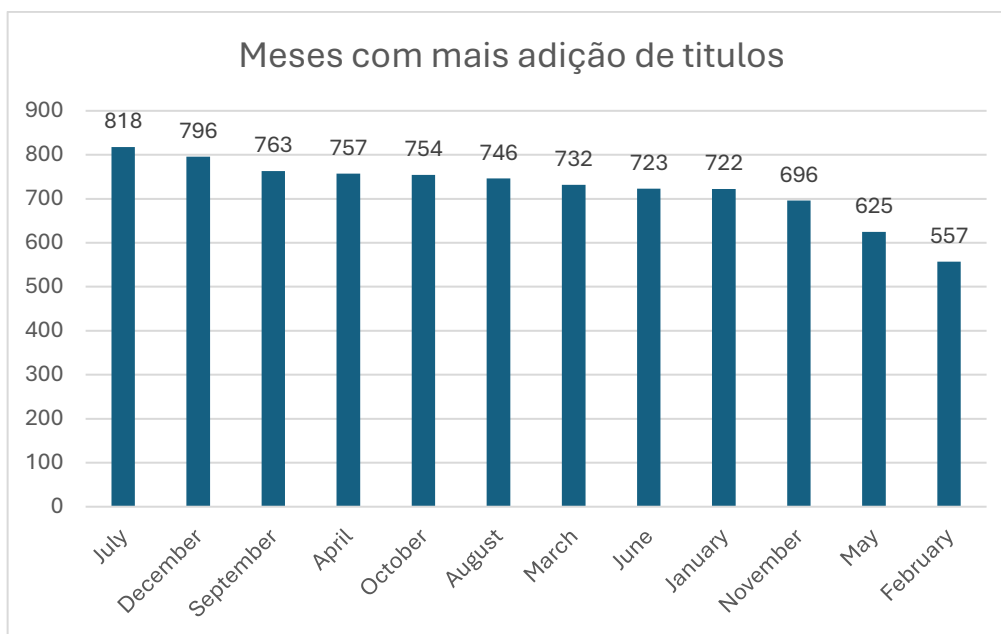
Movie

Swordfish, China Salesman, War Chhod Na Yaar, Magnus, Black or White, Isoken, The Blue Umbrella, Onc

TV Show

Green Frontier, Bad Education, Game Winning Hit, Stay Tuned!, The Naked Director, Jurassic World Camp





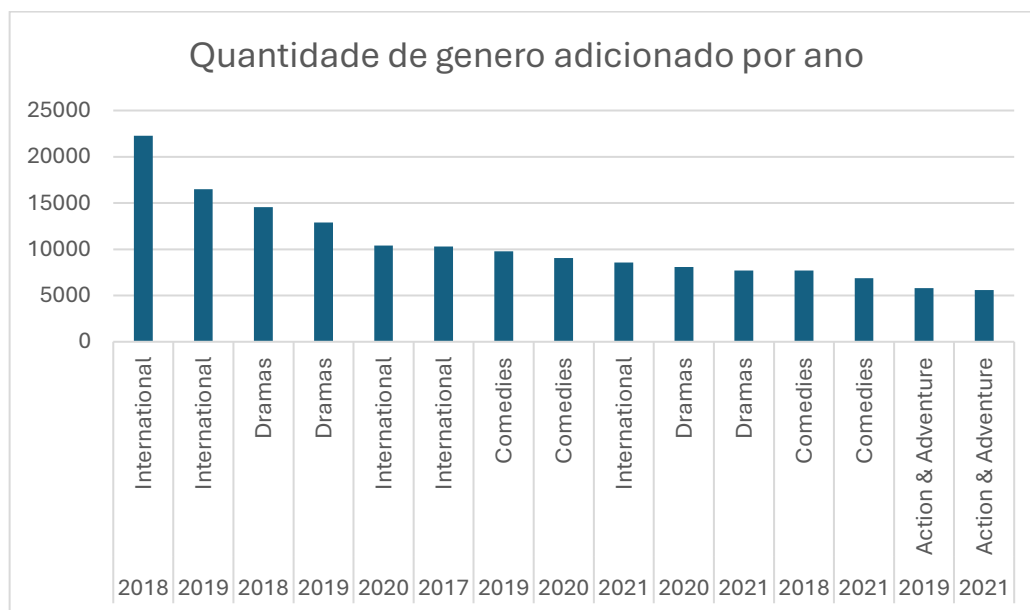
Títulos por ano com opção de filtragem por gênero

ano	titulos
2008	To and From New York, Dinner for Five
2009	Just Another Love Story, Splatter
2010	Mad Ron's Prevues from Hell
2011	The Ryan White Story, Adam: His Song Continues, Triumph of
2012	Casa de mi Padre, Being Elmo: A Puppeteer's Journey, Kung F
2013	The Short Game, The 4400, A Russell Peters Christmas, Aziz
2014	Chelsea Peretti: One of the Greats, The Battered Bastards of
2015	Jen Kirkman: I'm Gonna Die Alone (And I Feel Fine), Sammy &
2016	Color of Woman, Gentlemen and Gangsters, Skin Wars, The I
2017	Joaquín Reyes: Una y no más, Sohni Mahiwal, Magnus, Abby S
2018	Tamasha, TAU, Borderliner, A Billion Colour Story, China Sale
2019	Green Frontier, War Chhod Na Yaar, NOVA: Killer Floods, Iso
2020	The Magic School Bus Rides Again Kids In Space, The Outpost
2021	Oxygen, Canine Intervention, Beckett, Swordfish, Mommy Iss



Títulos com filtragem por gênero

genero
Rótulos de Linha
A Billion Colour Story, Opening Night, The Blue Umbrella, Bitch, Inxeba, The spiral, Listen, The Monster, The One An Easy Girl
An Imperfect Murder
Backfire
Bad Education, Secret of Success, My Dear Warrior, Game Winning Hit, Vincenzo, The Comedy Lineup, Stay Tuned
Bad Education, Still Game, Diva Brides, Horrid Henry, Bad Boy Billionaires: India, Derry Girls, White Gold, Edge of BASEketball, Sons of Ben, Accomplice, Radical: the Controversial Saga of Dada Figueiredo, The Game Changer
Beasts of No Nation
Beckett, Swordfish, Coffee & Kareem, China Salesman, Vettai, Beowulf, Triple Threat, The Time Machine, Office Bleach, Marvel Anime: Wolverine, CLANNAD, Darwin's Game, Knights of Sidonia, Love, Chunibyo & Other Delusional Borderliner, Oktoberfest: Beer & Blood, Blood & Water, Anne with an E, Spartacus, Warrior, Game Winning Hit, China Salesman, War Chhod Na Yaar, Magnus, Isoken, Inxeba, Pee Mak, Radical: the Controversial Saga of Dada Figueiredo



Títulos por ano com filtragem por gênero

genero	(Tudo)
Rótulos de Linha	
2008	
Dinner for Five	
To and From New York	
2009	
Just Another Love Story	
Splatter	
2010	
Mad Ron's Prevues from Hell	
2011	
Even the Rain	
Joseph: King of Dreams	
The Ryan White Story, Adam: His Song Continues, Triumph of the Heart, Joseph: King of Dreams, Victim of Beauty, Even the	
Triumph of the Heart, Quiet Victory: The Charlie Wedemeyer Story	
Victim of Beauty	
2012	
Being Elmo: A Puppeteer's Journey	
Casa de mi Padre, Kung Fu Panda: Holiday	
Kung Fu Panda: Holiday	
2013	
A Russell Peters Christmas, Aziz Ansari: Buried Alive, Iliza Shlesinger: War Paint, Russell Peters: Notorious	
Breaking Bad	
Breaking Bad, Jack Taylor	
Jack Taylor	

OS GÊNEROS MAIS POPULARES EM FILMES E SÉRIES

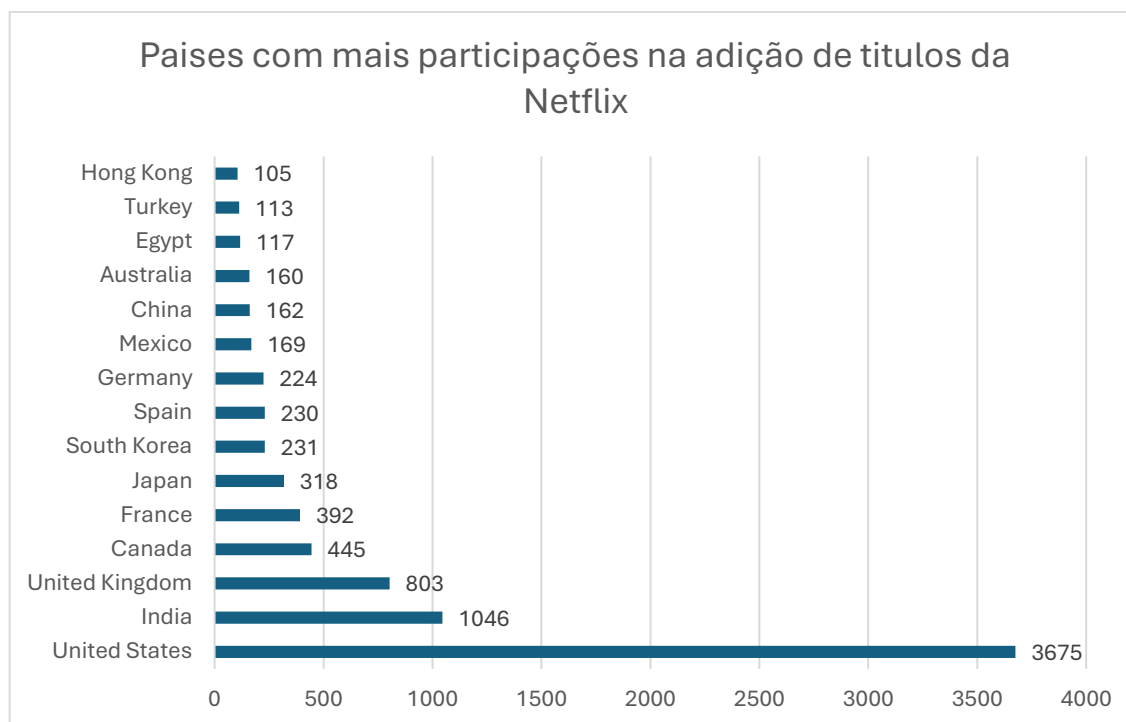
Movie		Movie		TV Show	TV Show		
		Independent ; 751	Children & Family; 641		Crime; 469		
					Kids' TV; 451		
		Romantic; 616			International; 1350	Docuseries; 395	
		Movie		TV Show		TV Show	Movie
International; 2748	Dramas; 2419	Documentaries; 866	Action & Adventure; 857	TV Dramas; 762	TV Comedies; 580	Thrillers; 573	
Comedies; 1670							

Títulos por filmes e series por gênero

Movie	International	Oxygen, Chatô: The King of Brazil, China Salesman, War Chhod Na Yaar, Sohni Mahiwal, Magnus, Isoken
Movie	Dramas	Oxygen, Chatô: The King of Brazil, War Chhod Na Yaar, Sohni Mahiwal, Black or White, The Blue Umbrella
Movie	Comedies	Too Handsome to Handle, Mommy Issues, War Chhod Na Yaar, Mama's Boy, Isoken, My Ex & Whys, The
TV Show	International	Green Frontier, Borderliner, Gentlemen and Gangsters, My Dear Warrior, Game Winning Hit, Life, Stay
Movie	Documentaries	Whispers, NOVA: Bird Brain, Amy Tan: Unintended Memoir, Born Strong, In Our Mothers' Gardens, For
Movie	Action & Adventure	Beckett, Swordfish, Coffee & Kareem, China Salesman, Vettai, Beowulf, Triple Threat, The Time Machin
TV Show	TV Dramas	Borderliner, Oktoberfest: Beer & Blood, Blood & Water, Anne with an E, Spartacus, Warrior, Game Win
Movie	Independent	A Billion Colour Story, Opening Night, The Blue Umbrella, Bitch, Inxeba, The spiral, Listen, The Monster,
Movie	Children & Family	True: Wuzzle Wegg Day, Magnus, The Christmas Chronicles, The Blue Umbrella, Double Dad, Incredible
Movie	Romantic	Tamasha, 10 Days in Sun City, Sohni Mahiwal, Isoken, Ready to Mingle, Last Summer, Mehendi Circus, I
TV Show	TV Comedies	Bad Education, Secret of Success, My Dear Warrior, Game Winning Hit, Vincenzo, The Comedy Lineup,
Movie	Thrillers	TAU, Oxygen, Forensic, Security, The Haunting of Molly Hartley, Berlin Syndrome, Mrs. Serial Killer, The
TV Show	Crime	Green Frontier, Borderliner, Gentlemen and Gangsters, A Perfect Day For Arsenide, Warrior, Somewhe
TV Show	Kids' TV	ChuChu TV Kids Songs, Learning Videos & Bedtime Stories, Power Rangers Zeo, Horrid Henry, Luna Pet
TV Show	Docuseries	The Staircase, Countdown: Inspiration4 Mission to Space, The Pyramid Code, A Year In Space, Bad Boy
Movie	Music & Musicals	Opening Night, This Is the Life, Vettai, Brad Paisley's Comedy Rodeo, Rush: Beyond the Lighted Stage, O
TV Show	Romantic	Color of Woman, Gentlemen and Gangsters, Bhaag Beanie Bhaag, Secret of Success, Club Friday To Be
Movie	Horror	Sinister 2, Stree, Pee Mak, The Haunting of Molly Hartley, Suzzanna: Buried Alive, The Monster, Case 39
Movie	Stand-Up Comedy	Joaquín Reyes: Una y no más, Gina Yashere: Skinny B*tch, ADAM SANDLER 100% FRESH, Amy Schume

CONTAGEM DE PAISES COM MAIS GENEROS ADICIONADOS

United States		United States		India	
Documentaries; 508		Action & Adventure; 402		Dramas; 829	
Children & Family; 390		Independent ; 385		Comedies; 677	
				International; 930	
				Dramas; 662	
				India	
				Comedies; 323	
				United Kingdom	
				International; 296	



Títulos por país

pais	títulos
United States	Oxygen, Swordfish, NOVA: Killer Floods, Black or White, Bitch, Once in a Lifetime Sessions with George Clooney
India	Tamasha, A Billion Colour Story, War Chhod Na Yaar, Vettai, Sohni Mahiwal, The Blue Umbrella, Abby Cadabby
United Kingdom	Bad Education, Beowulf, Arctic Dogs, White Gold, Eddie - Strongman, Sitting in Limbo, Learning Songs
Canada	Whispers, True: Wuzzle Wegg Day, The Pyramid Code, American Hangman, For the Love of Spock, Ann
France	Whispers, Oxygen, The Staircase, Power Rangers Zeo, The Chosen Ones, Is Love Enough? Sir, Inxeba, T
Japan	CLANNAD, Power Rangers Zeo, Naruto Shippūden the Movie: The Will of Fire, Terrace House: Tokyo 20
South Korea	Color of Woman, Animal Crackers, The Bros, Rookie Historian Goo Hae-Ryung, Love Rain, Tong: Mem
Spain	Animal Crackers, The Invisible Guest, Black Beach, Nevenka: Breaking the Silence, The Occupant, Pal
Germany	Borderliner, Devil's Bride, Oktoberfest: Beer & Blood, Rising High, Resident Evil: Afterlife, Inxeba, Mem
Mexico	La Viuda Negra, Sofía Niño de Rivera: Selección Natural, The Chosen Ones, Carlos Ballarta: El amor es
China	LuckyDays, Initial D, Beyond Skyline, Animal Crackers, The Boy, China Salesman, Set Off, Double Worl
Australia	Heartbreak High, The Bank Job, The Secret, Heidi, Ghost Rider, Glitch, The Adventures of Figaro Pho, E
Egypt	Monopoly (The Bank Of Luck), The Promise, Sameer Abu Alneel, Mama's Boy, Escaping Tel Aviv, Eugen
Turkey	Kids on the Block, Overnight Republic, Sen Kimsin?, Organize Isler, Fatma, Love 101, One-Way to Tomc
Hong Kong	Initial D, Doubles Cause Troubles, The Bund, Painted Faces, Revenge of the Green Dragons, SDU: Sex I
Nigeria	10 Days in Sun City, Ije: The Journey, Isoken, Lost in London, Crazy people, On the Real, Elevator Baby, (
Italy	Beckett, Summertime, Scandal in Sorrento, Suburra: Blood on Rome, The Ruthless, Security, Domino,
Brazil	Beckett, Alice Junior, Whindersson Nunes: Adult, Marco Luque: Tamo Junto, Chatô: The King of Brazil,

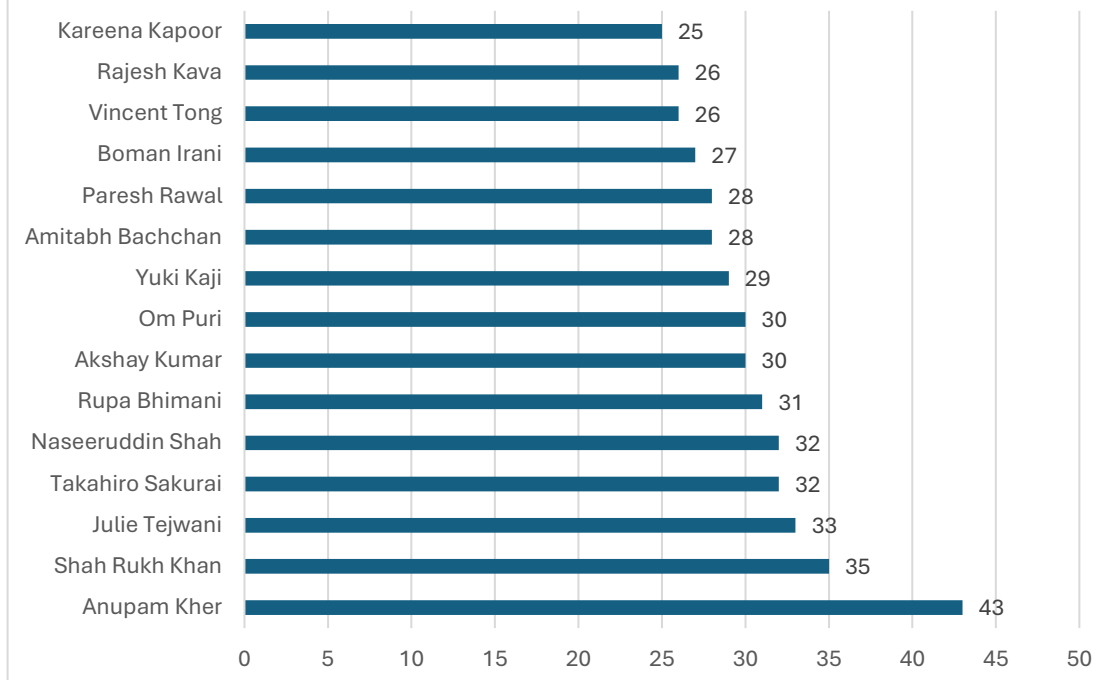
Quantidade de vezes que diretores e atores trabalharam juntos

diretor	atores	count
Rajiv Chilaka	Rajesh Kava	19
Rajiv Chilaka	Julie Tejwani	19
Rajiv Chilaka	Jigna Bhardwaj	18
Rajiv Chilaka	Rupa Bhimani	18
Rajiv Chilaka	Vatsal Dubey	16
Rajiv Chilaka	Mousam	13
Rajiv Chilaka	Swapnil	13
Suhas Kadav	Saurav Chakraborty	8
Toshiya Shinohara	Kappei Yamaguchi	7
Toshiya Shinohara	Houko Kuwashima	7
S.S. Rajamouli	Nassar	7
S.S. Rajamouli	Tamannaah Bhatia	7
Toshiya Shinohara	Kumiko Watanabe	7
S.S. Rajamouli	Ramya Krishnan	7
S.S. Rajamouli	Anushka Shetty	7

Títulos que cada ator e diretores trabalharam juntos

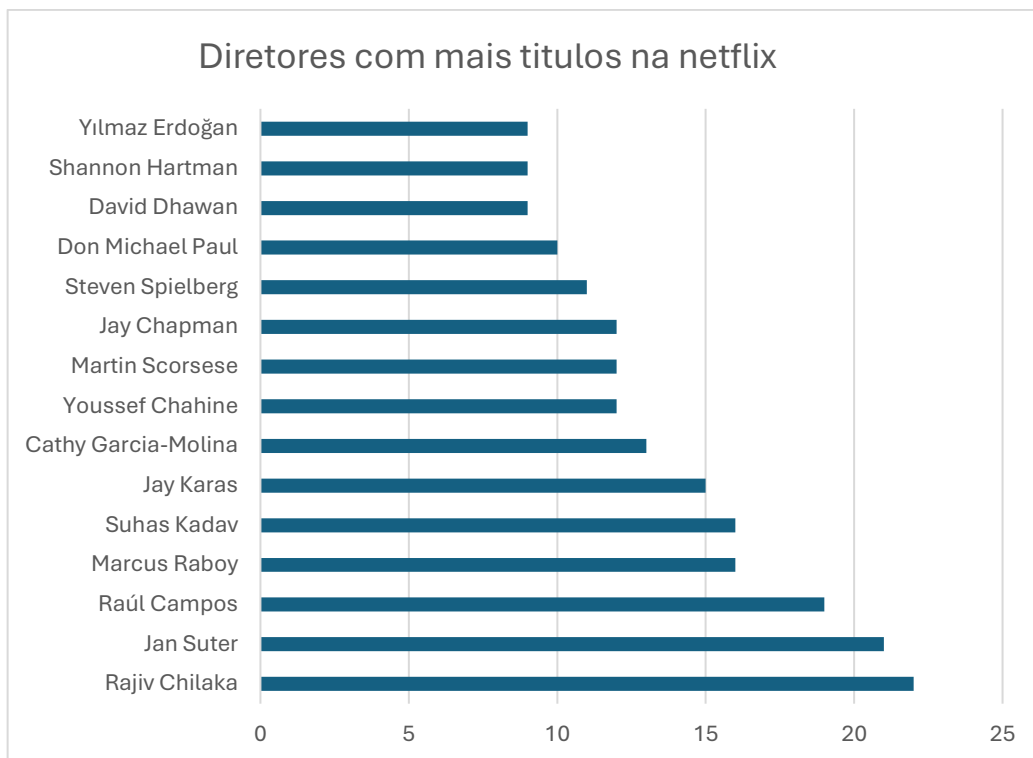
A	B	C
diretor	atores	titulos
Rajiv Chilaka	Rajesh Kava	Chhota Bheem and The ShiNobi Secret, Chhota Bheem Aur H
Rajiv Chilaka	Julie Tejwani	Chhota Bheem and The ShiNobi Secret, Chhota Bheem Aur H
Rajiv Chilaka	Jigna Bhardwaj	Chhota Bheem and The ShiNobi Secret, Chhota Bheem Aur H
Rajiv Chilaka	Rupa Bhimani	Chhota Bheem and The ShiNobi Secret, Chhota Bheem Aur H
Rajiv Chilaka	Vatsal Dubey	Chhota Bheem and The ShiNobi Secret, Chhota Bheem Aur H
Rajiv Chilaka	Mousam	Chhota Bheem And The Broken Amulet, Chhota Bheem and T
Rajiv Chilaka	Swapnil	Chhota Bheem And The Broken Amulet, Chhota Bheem and T
Suhas Kadav	Saurav Chakraborty	Motu Patlu Kung Fu Kings 4 The Challenge of Kung Fu Brother
Toshiya Shinohara	Kappei Yamaguchi	InuYasha the Movie 3: Swords of an Honorable Ruler, InuYas
Toshiya Shinohara	Houko Kuwashima	InuYasha the Movie 3: Swords of an Honorable Ruler, InuYas
S.S. Rajamouli	Nassar	Baahubali: The Beginning (Malayalam Version), Baahubali 2:
S.S. Rajamouli	Tamannaah Bhatia	Baahubali: The Beginning (Malayalam Version), Baahubali 2:
Toshiya Shinohara	Kumiko Watanabe	InuYasha the Movie 3: Swords of an Honorable Ruler, InuYas
S.S. Rajamouli	Ramya Krishnan	Baahubali: The Beginning (Malayalam Version), Baahubali 2:
S.S. Rajamouli	Anushka Shetty	Baahubali: The Beginning (Malayalam Version), Baahubali 2:

Atores com mais participação nos títulos da netflix



Títulos que os atores trabalharam

atores	titulos
Anupam Kher	Toilet: Ek Prem Katha, Prem Ratan Dhan Payo, Haseena Maan Jaayegi, Vivah,
Shah Rukh Khan	Anjaam, Billu, Pardes, Don, Swades, Raees, Kabhi Alvida Naa Kehna, Main Ho
Julie Tejjwani	Chhota Bheem Ka Troll Se Takkar, Chhota Bheem and The ShiNobi Secret, Chl
Takahiro Sakurai	K, Sirius the Jaeger, Mobile Suit Gundam: Iron-Blooded Orphans, Knights of Si
Naseeruddin Shah	Ram Prasad Ki Tehrvi, John Day, Mandi, Peepli Live, 7 Khoon Maaf, Bazaar, Ho
Rupa Bhimani	Chhota Bheem Ka Troll Se Takkar, Chhota Bheem and The ShiNobi Secret, Chl
Akshay Kumar	Toilet: Ek Prem Katha, Insan, Bhool Bhulaiyaa, Boss, Joker, The Shaukeens, Te
Om Puri	Project Marathwada, Disco Dancer, Bollywood Calling, Billu, Don, Kyun! Ho G
Yuki Kaji	K, NiNoKuni, GANTZ:O, B: The Beginning, Magi: The Labyrinth of Magic, Attack
Amitabh Bachchan	Amar Akbar Anthony, Wazir, Kyun! Ho Gaya Na, Lakshya, Hera Pheri, Kabhi Al
Paresh Rawal	Chup Chup Ke, Haseena Maan Jaayegi, Bhool Bhulaiyaa, Maya Memsaab, Phi
Boman Irani	Happy New Year, Ferrari Ki Sawaari, Don, Parmanu: The Story of Pokhran, Dri
Vincent Tong	Super Monsters Monster Pets, Super Monsters: The New Class, Ghost Patrol,
Rajesh Kava	Chhota Bheem and The ShiNobi Secret, Chhota Bheem Aur Hanuman, Mighty
Kareena Kapoor	Chameli, Uda Punjab, Heroine, Yuva, Don, Kabhi Khushi Kabhie Gham, Aitraa
Andrea Libman	Super Monsters: The New Class, Chip and Potato, Super Monsters: Santa's Su
Samuel L. Jackson	I Am Not Your Negro, Barely Lethal, Patriot Games, Pulp Fiction, Death to 202
John Cleese	Monty Python: Live at Aspen, George of the Jungle 2, Monty Python and the Hol
Tara Strong	Animal Crackers, Batman: The Killing Joke, Gabby's Dollhouse, The Rugrats M
Fred Tatasciore	Marvel's Hulk: Where Monsters Dwell, Marvel Super Hero Adventures: Frost F
Jigna Bhardwaj	Chhota Bheem and The ShiNobi Secret, Chhota Bheem Aur Hanuman, Little S
Daisuke Ono	K, Godzilla, GANTZ:O, Black Butler, Magi: Adventure of Sinbad, Yowamushi Pe
Kay Kay Menon	Black Friday, Ray, Life Ki Toh Lag Gayi, Penalty, Ek Khiladi Ek Haseena, Mumba
Nawazuddin Siddiqui	Peepli Live, Serious Men, Manorama Six Feet Under, Sacred Games, Petta, Ka
Junichi Suwabe	Fate/stay night: Unlimited Blade Works, Godzilla, Food Wars!: Shokugeki no S
Ashleigh Ball	The Hollow, My Little Pony Equestria Girls: Forgotten Friendship, My Little Pon
Ajay Devgn	Golmaal: Fun Unlimited, Kaal, Bhoot, Fitoor, Insan, Yuva, Company, Himmatu



Títulos que os diretores trabalharam com filtro para escolher diretor

diretor	(Tudo)
Rótulos de Linha	
A Babysitter's Guide to Monster Hunting	
A Beautiful Life, Revenge of the Green Dragons, Look for a Star	
A Billion Colour Story, Saeed Mirza: The Leftist Sufi	
A Bridge Too Far	
A California Christmas, Lady Driver, Wheels of Fortune, Roped	
A Champion Heart	
A Chaster Marriage	
A Choo	
A Christmas Prince: The Royal Baby, Judy Moody and the Not Bummer Summer, A Christmas Pr	
A Christmas Special: Miraculous: Tales of Ladybug & Cat Noir, Miraculous: Tales of Ladybug & C	
A Cinderella Story	
A Cinderella Story: Christmas Wish	
A Classic Horror Story	
A Dangerous Woman	
A Family	
A Family Affair	
A Family Man	
A Family Reunion Christmas	
A Faraway Land	
A Fortunate Man	
A Futile and Stupid Gesture	

3.1.3 Reflexão aprofundada

Com o projeto foi possível identificar padrões e tendências nos dados o que permite com que os resultados da pesquisa possam ser usados para implementar um

sistema que recomenda filmes ou gerar insights para tomadas de decisões da Netflix, ou de maneira mais casual em uma simples escolha do que assistir.

3.1.4 Considerações finais

A análise revelou padrões interessantes na distribuição de conteúdo da Netflix, podendo retirar diversas perspectivas sobre. Contudo o estudo foi limitado ao conjunto de dados disponível no GitHub, que não representa a totalidade do catálogo da Netflix. Além disso, a previsão de tendências futuras não foi abordada neste projeto.

Porém aprofundando os dados e após a experiência desse projeto, esse projeto poderia servir como balizador para uma implementação de um sistema ou site de recomendações de filmes por exemplo e embora utilizei o PySpark e tenha se mostrado uma ferramenta eficiente para o processamento de dados, outras ferramentas de Big Data, como Hadoop ou MongoDB, poderiam ser exploradas. A escolha do PySpark foi motivada pela sua facilidade de uso e integração com Python e SQL.

ANEXOS

<https://docs.google.com/document/d/1FSzBSyll3CeoAwzBW9GVAYULgQnDLZRwaUAkjkPNgfw/edit?usp=sharing>