

TEMPLATE PARA ENTREGA DO PROJETO DA DISCIPLINA
Programação para Dados
Fase 2

Nome do estudante: Mariana Segovia Penha

[Link do repositório no GitHub.](#)

1) Quais são os dez jogos mais bem avaliados, de acordo com o Metacritic?

	Name	Metacritic score	Release date
45752	Disco Elysium - The Final Cut	97	2019-10-15
64935	Persona 5 Royal	97	2022-10-20
49557	Half-Life	96	1998-11-08
36775	Half-Life 2	96	2004-11-16
32406	BioShock	96	2007-08-21
57205	Grand Theft Auto V	96	2015-04-13
28170	Portal 2	95	2011-04-18
28484	Portal 2	95	2011-04-18
12746	Sid Meier's Civilization® IV	94	2006-10-25
55150	Sid Meier's Civilization® IV	94	2006-10-25

Os dez títulos mais bem avaliados foram recuperados através da ordenação do DataFrame pelo atributo Metacritic score em ordem decrescente. A lista é liderada por um empate entre Disco Elysium - The Final Cut (97) e Persona 5 Royal (97). O critério para desempate, conforme a solicitação, aplicou a ordenação secundária por "Release date" em ordem crescente (do mais antigo para o mais recente).

1. Disco Elysium - The Final Cut (97)
2. Persona 5 Royal (97)
3. Half-Life (96)
4. Half-Life 2 (96)
5. BioShock™ (96)
6. Grand Theft Auto V (96)
7. Portal 2 (95)
8. Sid Meier's Civilization® IV (94)

O jogo com a data de lançamento mais antiga (Disco Elysium, 2019-10-15) ocupa a primeira posição.

2) Para jogos de role-playing, qual o número médio e máximo de: DLCs, avaliações positivas, avaliações negativas e materiais de demonstração (número de capturas de tela e filmes, somados)?

```
... dlc_mean      0.952826
    dlc_max      2366.000000
    pos_mean     1516.528426
    pos_max      964983.000000
    neg_mean      247.193493
    neg_max      129925.000000
    demo_mean     0.000000
    demo_max     0.000000
    n_jogos      12295.000000
    dtype: float64
```

A análise do gênero Role-Playing (RPG), que totaliza 12.295 jogos na base de dados, demonstra os resultados da imagem acima, no código, as métricas de agregação foram calculadas utilizando os métodos `mean()` e `max()` do Pandas.

O desvio-padrão elevado e o contraste entre média e máximo são significativos. A média de Avaliações Positivas (1516.53) é inflada pela presença de outliers de alto impacto (até 964.983 reviews), característicos de títulos blockbuster do gênero. Da mesma forma, o número máximo de DLCs (2366) representa uma forte anomalia, indicando que a distribuição dos DLCs é extremamente assimétrica e dominada por um ou poucos jogos. A média zero em Materiais de Demonstração resulta da aplicação de `fillna(0)` no pré-processamento, indicando grande presença de dados nulos nessas colunas do dataset.

3) Quais são as cinco empresas que mais publicam jogos pagos na plataforma? Para tais empresas, qual o número médio e mediano de avaliações positivas de seus jogos pagos?

	jogos_pagos	pos_mean	pos_median
publisher_main			
Big Fish Games	443	7.363431	5.0
????	308	133.996753	9.5
8floor	239	4.460251	3.0
?????	209	161.406699	5.0
Strategy First	162	276.450617	23.0

A análise da mediana de avaliações positivas (que minimiza a distorção dos outliers) revela um contraste. Enquanto os líderes por volume (Big Fish Games, 8floor) têm medianas baixíssimas (5.0) e (3.0), respectivamente, indicando que a maioria de seus jogos recebe pouca atenção, a Strategy First (23.0) demonstra um engajamento inicial significativamente maior com seus títulos. Isso sugere que a Strategy First, apesar do menor volume, possui uma percepção de qualidade mais consistente na maioria de seus lançamentos pagos.

** A presença das entradas ????? no Top 5, que representam 517 títulos pagos combinados, indica um desafio de qualidade de dados. Geralmente resultantes de problemas com ausência de metadados (valores nulos ou strings não processáveis no campo Publisher original). Esses publishers anônimos detêm um volume considerável de lançamentos. **

4) O número de jogos que suportam o sistema operacional Linux cresceu entre 2018 e 2022?

	year	qtd_linux
0	2018	7854
1	2019	7615
2	2020	9474
3	2021	12308
4	2022	13961

Sim, o número de jogos lançados com suporte ao sistema operacional Linux cresceu significativamente no período de 2018 a 2022, como demonstrado acima. O suporte a Linux cresceu de \$7.854\$ jogos em 2018 para \$13.961\$ jogos em 2022. Excluindo uma ligeira queda entre 2018 e 2019, o crescimento foi constante e acelerado, resultando em um aumento de aproximadamente \$77.7\%\$ no total de títulos que oferecem suporte a Linux em apenas quatro anos.

5) Qual a evolução anual da média de Avaliações Positivas (a partir de 2010) entre jogos Pagos e Gratuitos (F2P)? Em particular, como essa tendência se manifesta no gênero de alto volume Indie? (demonstrado no gráfico 3)

--- Pergunta Adicional: Média de Avaliações Positivas (Pago vs Gratuito) por Gênero/Ano ---				
	year	genero_principal	is_paid	Positive
0	2010	action	False	12755.666667
1	2010	action	True	3014.901961
2	2010	action,adventure	False	5353.250000
3	2010	action,adventure	True	15607.363636
4	2010	action,adventure,casual	True	1042.500000
5	2010	action,adventure,casual,indie	True	39.000000
6	2010	action,adventure,casual,rpg	False	8091.000000
7	2010	action,adventure,indie	True	9555.000000
8	2010	action,adventure,indie,rpg	True	430.000000
9	2010	action,adventure,strategy	True	905.500000

A ideia é olhar para o engajamento médio (quantidade de reviews) e comparar como ele varia entre jogos gratuitos e pagos ao longo do tempo. O foco no gênero Indie foi escolhido porque é um dos que mais aparecem na base de dados, o que ajuda a entender melhor como funciona essa diferença entre F2P e Pago dentro de um nicho relevante como esse.

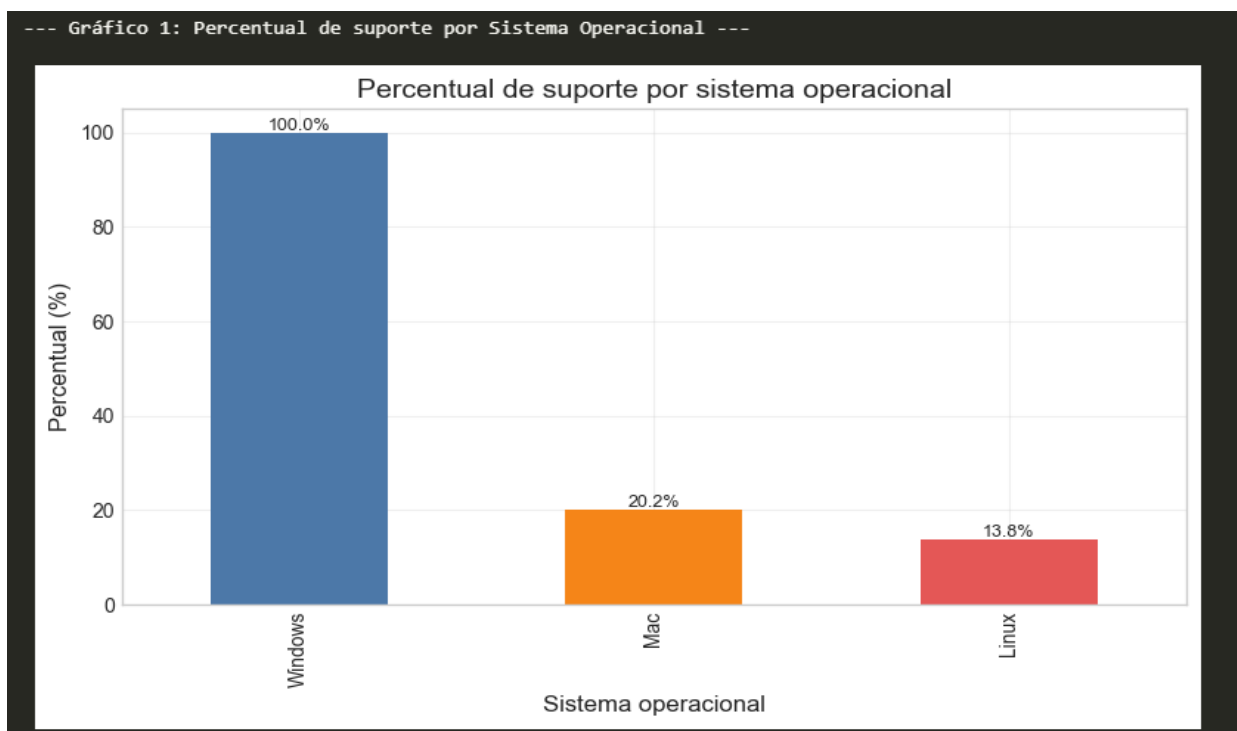
6) GRÁFICO 1: Percentual de jogos que possuem suporte para cada sistema operacional.

Foi gerado um gráfico de barras que apresenta o percentual de jogos compatíveis com Windows, Mac e Linux. A análise foi conduzida após o pré-processamento dos dados, onde os campos de suporte foram normalizados para interpretar corretamente os valores originais do CSV, que continham variações como "FALSO" e símbolos mascarados como #####. O gráfico revela que:

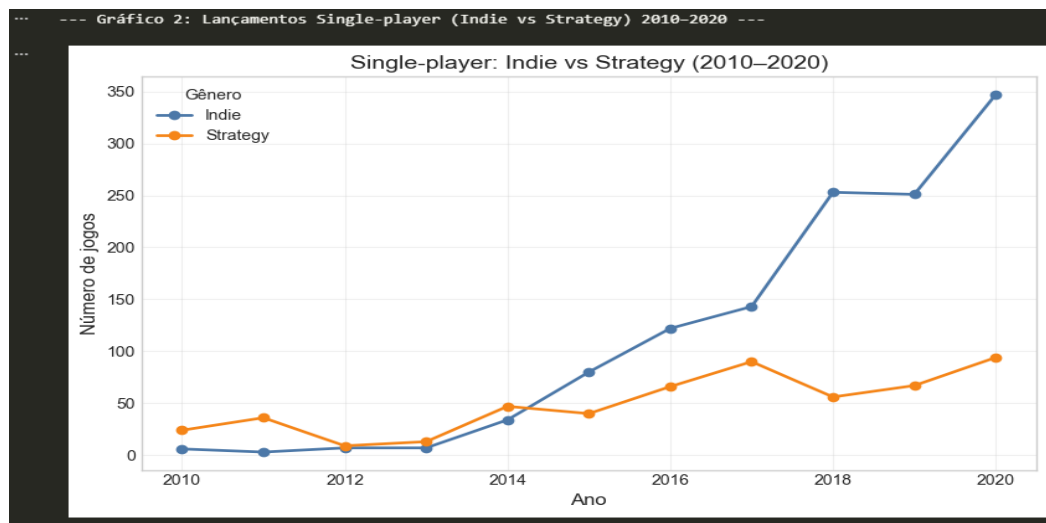
- Windows possui suporte em praticamente todos os jogos do conjunto, com 100% de compatibilidade.
- Mac aparece com suporte em aproximadamente 20,2% dos jogos.
- Linux tem suporte em cerca de 13,8% dos títulos.

Esses resultados indicam uma forte predominância do Windows como plataforma principal no mercado de jogos, enquanto Mac e Linux ainda representam uma fatia significativamente menor.

- O dataset foi carregado com pandas e tratado com funções específicas para limpeza e transformação.
- Os campos de suporte a sistemas operacionais foram convertidos para valores booleanos por meio da função `limpar_so`, que interpreta corretamente os valores textuais.
- O gráfico foi gerado com matplotlib, utilizando a função `grafico_percentual_por_sistema(df)`, que calcula o percentual de suporte com base no total de jogos.



7) **GRÁFICO 2:** Número total de jogos single-player do gênero Indie e estratégia lançados por ano entre 2010 e 2020.



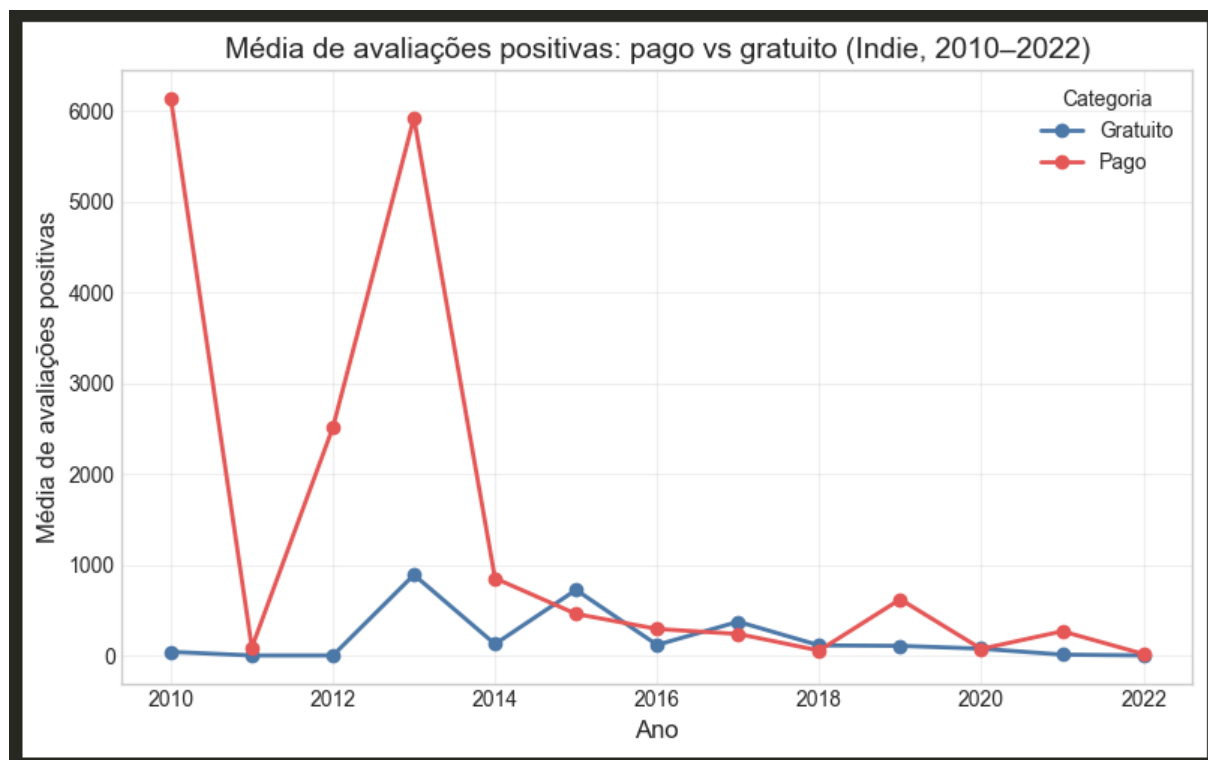
- **Indie:** Demonstra uma trajetória de crescimento exponencial acentuada após 2015, caracterizando o boom do desenvolvimento indie e a redução das barreiras de entrada na Steam.
- **Strategy:** Mantém uma curva de crescimento linear e estável ao longo da década. A menor inclinação sugere um nicho de desenvolvimento mais especializado e maduro, com menor saturação de lançamentos em comparação com o gênero Indie.

8) **GRÁFICO 3:**

Como solicitado, o gráfico abaixo seria um acréscimo à pergunta adicional criada de forma autoral, onde foi verificada a evolução anual da média de avaliações positivas (a partir de 2010) entre jogos Pagos e Gratuitos. A análise mostra, em particular, como essa tendência se manifesta no gênero Indie, que possui alto volume na plataforma.

Permanecendo no segmento citado, a performance máxima de engajamento pertenceu aos jogos pagos nos anos iniciais. No entanto, na média, a taxa de sucesso no engajamento entre modelos *Pago* e *Gratuito* se tornou equivalente no mercado pós-2015.

O gráfico, gerado pela função `grafico_paid_free_por_ano` (em `viz.py`), utiliza o método `unstack()` após o agrupamento para separar as linhas de dados entre jogos pagos (`is_paid = True`) e gratuitos (`is_paid = False`), permitindo a comparação de séries temporais.



Concluindo o projeto, deixo aqui novamente o [link](#) para o repositório dele, onde consta histórico de commits e READ.ME.