**FUNDAÇÃO GETÚLIO VARGAS**

Campus sede – Rio de Janeiro

GRADUAÇÃO EM CIÊNCIA DE DADOS E INTELIGÊNCIA ARTIFICIAL

**ALEX JÚNIO MAIA DE OLIVEIRA**

**BRUNO FERREIRA SALVI**

**JOÃO PEDRO JERÔNIMO DE OLIVEIRA**

**MATHEUS VILARINO DE SOUZA PINTO**

**THALIS AMBROSIM FALQUETO**

**APRIMORANDO VISUALIZAÇÕES DE DADOS: OBJETIVOS, ANÁLISES E FUNDAMENTOS**

Trabalho de A2 apresentado à Graduação em Ciência de Dados e Inteligência Artificial da Fundação Getúlio Vargas (FGV) como requisito parcial para o curso de Análise Exploratória de Dados e Visualização, orientado pelo professor Walter Wagner Carvalho Sande.

Rio de Janeiro – Rio de Janeiro

2024

SUMÁRIO

**1 INTRODUÇÃO............................................................................................................ 3**

**2 DESENVOLVIMENTO................................................................................................. 3**

**3 CONCLUSÕES......................................................................................................... 14**

**REFERÊNCIAS 15**

# **1 INTRODUÇÃO**

Na avaliação proposta da primeira metade do curso de “Linguagens de Programação”, não fizemos uma quantidade satisfatória de gráficos. Por essa razão, decidimos aprimorar os gráficos de um projeto realizado na avaliação da segunda metade da disciplina de “Introdução a Ciência de Dados”, durante o primeiro semestre da graduação. Com tais melhorias, temos o objetivo de tornar as representações visuais dos dados mais impactantes e esclarecedoras.

Nessa disciplina, os objetivos do trabalho envolveram realizar uma raspagem de dados, integrar os dados com a “API” do “ChatGPT” e implementar algum tipo de análise e visualização de dados. No nosso caso, optamos por fazer a raspagem sobre dois sites de jogos, “IMDB” e “STEAM”, com o intuito de comparar alguns dos principais jogos e empresas que utilizam tais plataformas digitais.

Vale lembrar que a base de dados aborda uma parte dos jogos de cada empresa em cada site analisado, isso pois o foco central do projeto era a integração e a análise com o GPT, e não uma extensa base de dados. Para verificar o resultado do trabalho original, basta clicar no link do [repositório do GitHub](https://github.com/jaopredo/gpt-analysis).

# **2 DESENVOLVIMENTO**

Contudo, desta vez, ao invés de confeccionarmos outro dashboard, iremos contar uma história sobre os dados coletados.

Pensamos em dar ênfase em alguns dados específicos, nesse caso, a empresa ‘Valve’, que, na maioria dos gráficos teve ou seus jogos, ou a própria empresa como elemento de destaque por estar sempre nas melhores posições.

Desta forma, iremos revisar detalhadamente cada gráfico gerado e discutir possíveis melhorias. O trabalho – visualmente falando – foi dividido em três partes: as primeira e terceira telas de dashboards, que relacionam as empresas, seus respectivos jogos e as pontuações de alguns comentários dos jogos, contêm as visualizações que serão revistas.

Sabendo disso, vamos iniciar a transformação do primeiro dashboard:

A screenshot of a computer screen

Description automatically generated

Figura 1 – Vista geral do primeiro dashboard

Olhando mais de perto os dois primeiros gráficos:



A graph with a red line

Description automatically generated

Figuras 2 e 3 – Gráficos que mostram os 5 melhores e os 5 piores jogos ranqueados.

Percebemos que as escolhas tidas para os gráficos são ilógicas, uma vez que não há sentido aparente possuir uma linha que liga as avaliações dos jogos ranqueados, isso ocorre pois não há uma associação direta entre eles. Além disso, também é perceptível a inutilidade da linha do eixo abscisso separando em um gráfico e no outro não. Além disso, sente-se falta de uma legenda apropriada.

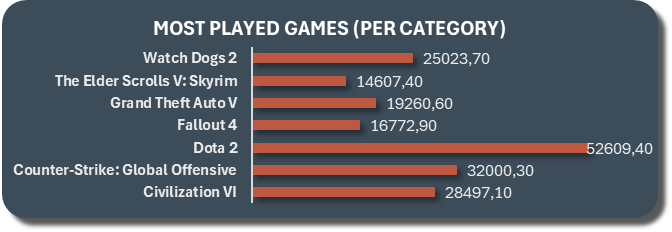
Portanto, mesclamos os dois gráficos, atribuindo a cor azul aos maiores “rates”, a cor laranja aos menores rates e incluímos a média de todos os jogos “entre” as duas tabelas. Com isso, pode-se ver que dois dos cinco jogos melhores ranqueados são da empresa Valve (Half-Life: Alyx e Portal 2), o que poderia incrementar mais na história.

A graph of a number of people

Description automatically generated with medium confidence

Figura 4 – Gráfico reestruturado correspondente aos gráficos das figuras 2 e 3.

Partindo para o terceiro gráfico do dashboard:



**LOWEST RANKED GAMES**

Figura 5 – Gráfico que mostra os jogos mais jogados por categoria (seletor de dados).

Neste gráfico, percebe-se algumas sutis melhorias: retirar o sombreamento, adicionar a unidade de medida utilizada para realizar a contagem, arredondar e ordenar os valores e destacar o maior valor, por ser destoante. Fazendo as alterações descritas, ficamos com gráfico da figura 6, que destaca, mais uma vez, um jogo da Valve (Dota 2, em azul):

A graph of a number of games

Description automatically generated with medium confidence

Figura 5 – Gráfico reestruturado correspondente ao gráfico da figura 4 (ainda com seletor de dados).

Observando agora os gráficos quatro e cinco do dashboard:

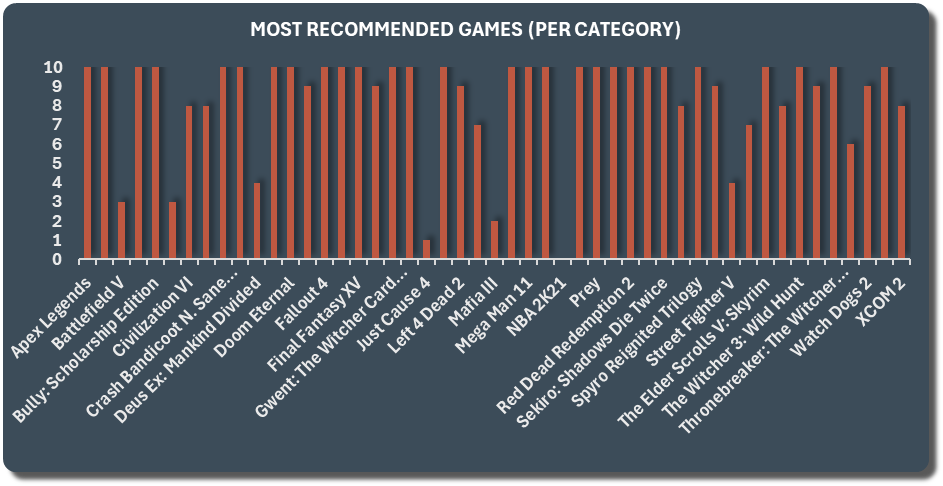


Figura 6 – Gráfico que representa os jogos mais recomendados por categoria.

**A colorful circle with numbers and text

Description automatically generated**

Figura 7 – Gráfico que representa as categorias com as melhores avaliações.

Nestes dois gráficos, pensamos em retirar essa quantidade enorme de colunas do gráfico da figura 6, que não foi confeccionado corretamente, visto que existem mais colunas do que legendas, e rearranjar o gráfico da figura 7 os mesclando, melhorando assim a legenda.

Dessa maneira, transformamos cada barra em uma categoria e o seu valor associado às médias das avaliações, além da ênfase nos dois maiores principais valores e a explicação do seu significado como legenda.

A graph of a graph with text

Description automatically generated with medium confidence

Figura 8 – Gráfico reestruturado correspondente aos gráficos das figuras 6 e 7.

Partindo para o último gráfico do dashboard:

**A red and white graph

Description automatically generated with medium confidence**

Figura 9 – Gráfico que representa o quão proveitoso é um jogo por categoria (seletor de dados).

Este gráfico é similar aos dois iniciais, a representação de área é ilógica tendo em vista que a ligação entre cada jogo não faz sentido, por mais que eles sejam da mesma categoria.

Além disso, pensamos em ordenar, tirar a sombra e o contorno. Para isso, julgamos necessário alterar para um gráfico de barras horizontal, selecionando agora os 10 melhores “rates” por categoria, evitando a quantidade desnecessária de colunas. Com tais mudanças, pode-se novamente se atentar nos dois jogos no topo: ambos são, mais uma vez, da empresa Valve.

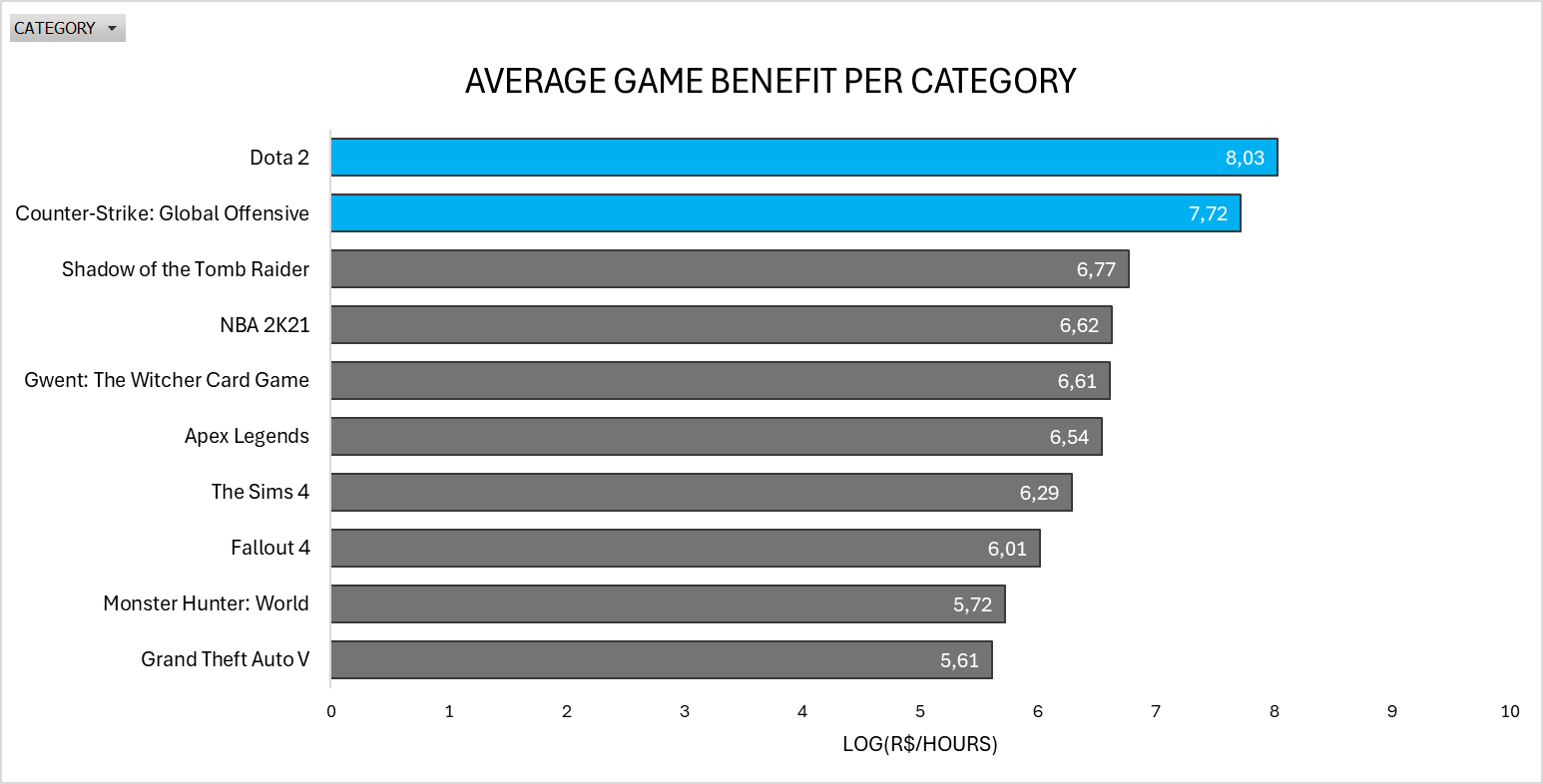
****

Figura 10 – Gráfico reestruturado correspondente ao gráfico da figura 9.

Além disso, o “benefit/beneficial” (proveitoso) foi especificado na lengenda, e é normalizado numa escala de 0 a 10 pela fórmula:  
“=LOG10(([@[HOURSSPENT]]/([@PRICE]+1)\*([@RATE]+1)/11)+0,0001)\*10/8)+5”.

Neste momento, iremos para a outra metade da visualização gráfica do dashboard:

A screenshot of a graph

Description automatically generated

Figura 11 – Vista geral do terceiro dashboard

Olhando mais de perto o primeiro gráfico:

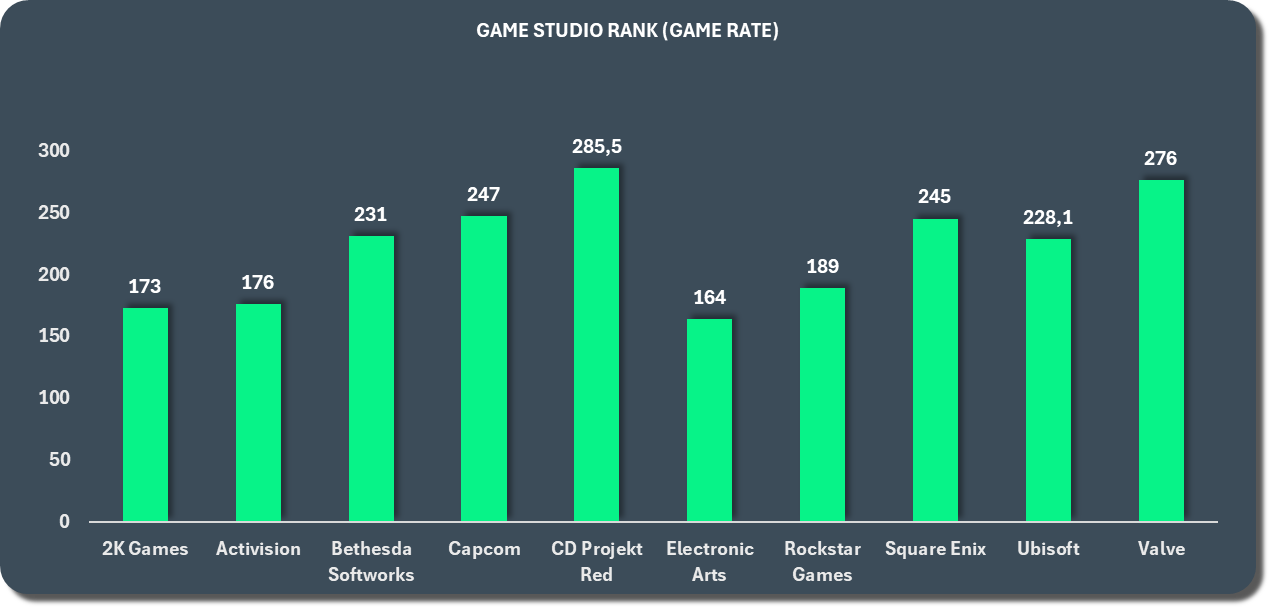


Figura 12 – Gráfico que representa a soma das avaliações de cada empresa de jogos.

Neste gráfico, vemos algumas inconsistências, tais como a falta de organização do eixo abscisso, colunas muito estreitas e falta de informações relacionadas ao eixo ordenado. Além do errado ordenamento das barras do sombreamento.

Fazendo as melhorias e modificando o parâmetro do gráfico, em que anteriormente era a soma total dos ‘rates’ e, agora, optamos pela média das avaliações, além de recolocar a legenda.

Analogamente aos outros gráficos, a empresa Valve é destaque novamente. Ao aplicar o storytelling e destacando a cor da coluna da Valve por estar novamente em primeiro, conseguiríamos convencer o público de que a empresa é a melhor da base de dados.

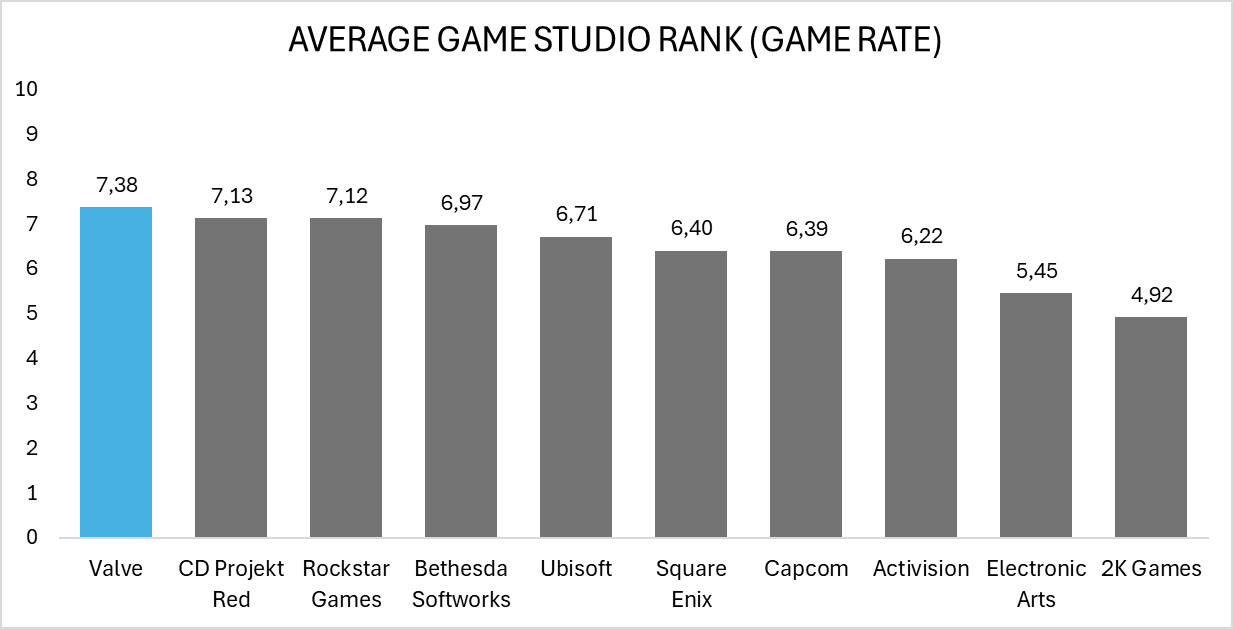


Figura 13 – Gráfico reestruturado correspondente ao gráfico da figura 12.

Avançando para o segundo gráfico do dashboard:

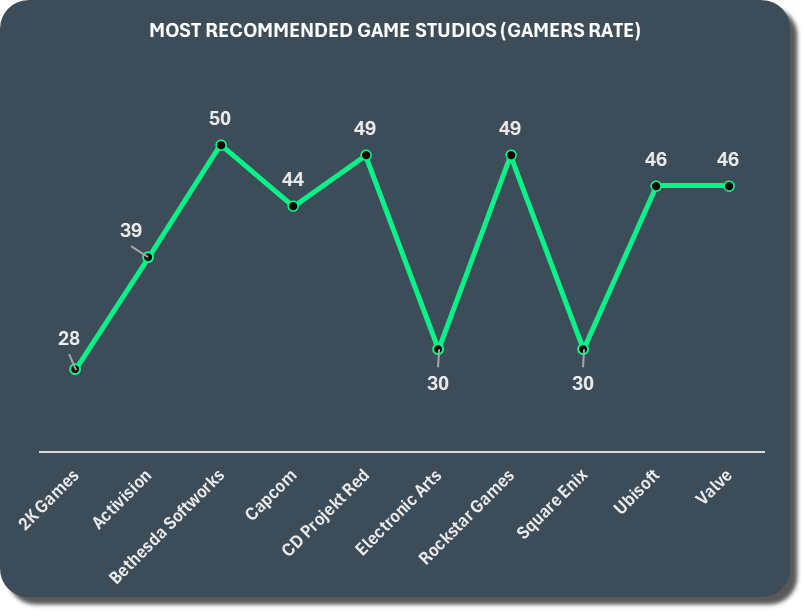


Figura 14 – Gráfico que representa as empresas mais recomendadas.

Nesse gráfico, nós nem entendemos ou sequer lembramos de como foi realizado. Por esse motivo, refizemos com outra forma.

O objetivo do novo gráfico é, baseando-se nas médias de cada jogo, mostrar as notas intermediárias de cada empresa dentro de cada categoria, visando ressaltar quais empresas têm maior aclamação do público dentro de cada setor de jogo.

A graph of different colored lines

Description automatically generated

RATE: Chat GPT analysis from the comments of the users about the game.

Figura 15 – Gráfico reestruturado correspondente ao gráfico da figura 14.

Conseguimos, com esse novo gráfico, destacar como o foco em uma categoria de jogo pode ser benéfico para o estúdio, destacando as empresas Valve e Bathesda.

Partindo agora para o penúltimo gráfico:

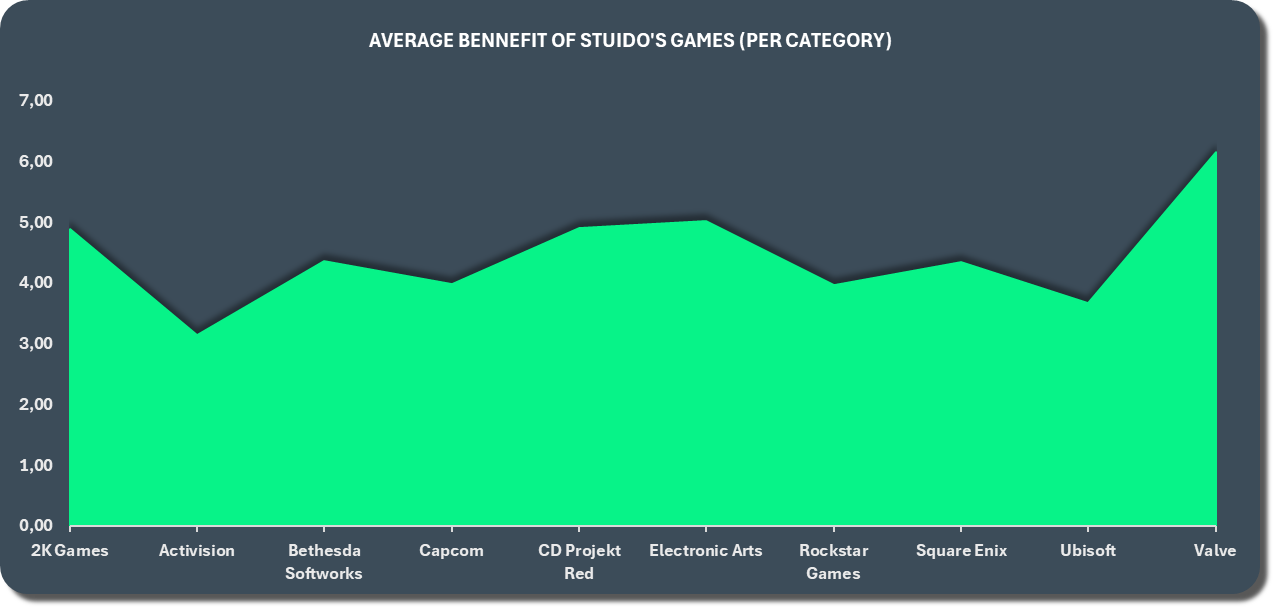
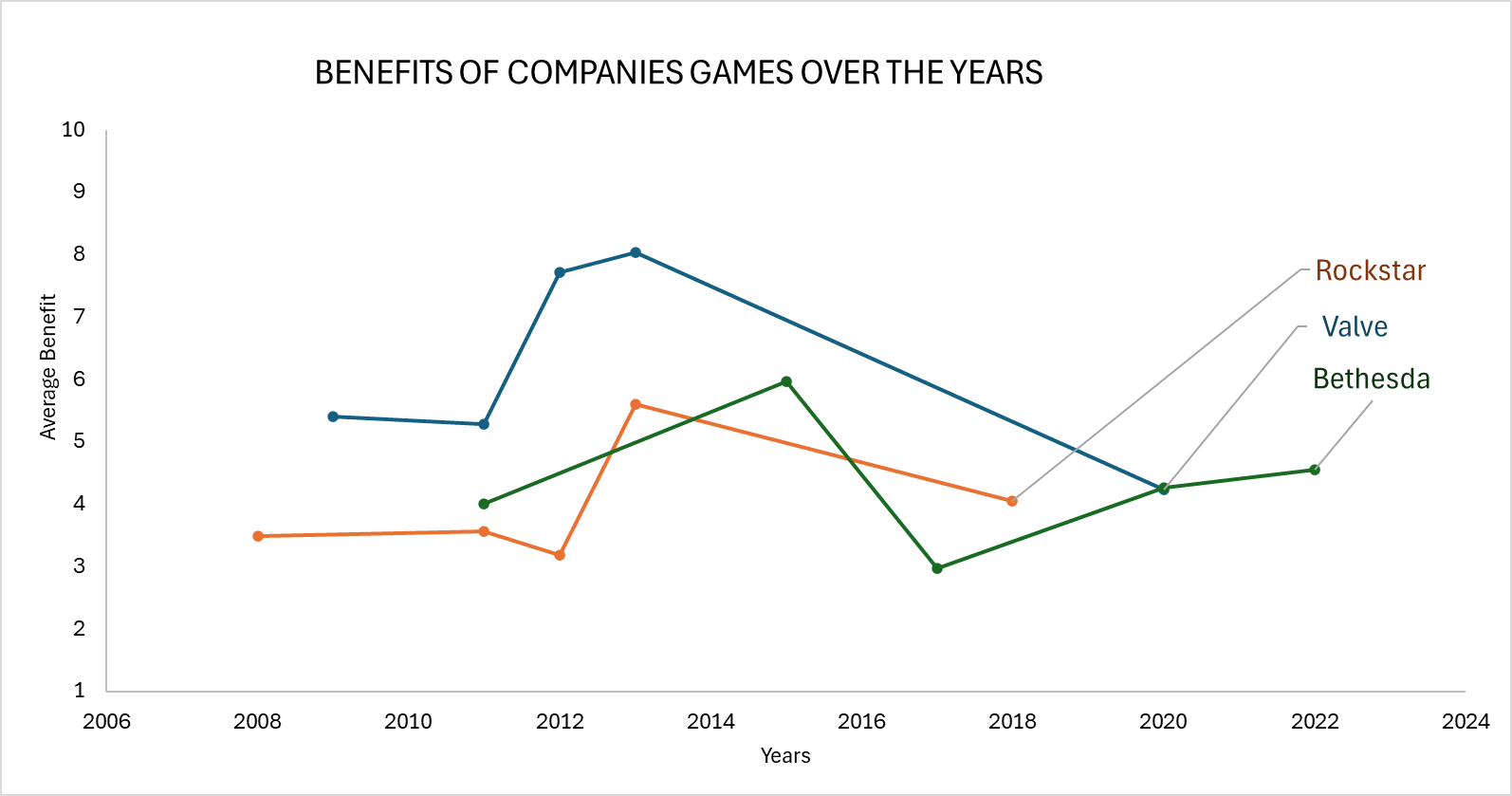


Figura 16 – Gráfico que representa o quão proveitosa é uma empresa de jogos (por categoria).

Neste gráfico, percebemos que é mais um dos gráficos em que o tipo de visualização não possui o menor sentido aparente, uma vez que os estúdios de cada jogo não têm ligação direta.

Por isso, e pensando em não fazer mais uma ilustração com barras, pensamos em transformar em um gráfico que compara o “benefit” de alguns jogos de cada estúdio pelo ano de lançamento (não todos, pois alguns estúdios não tinham uma quantidade mínima de jogos aceitável para ser comparada), priorizando uma interpretação limpa.

Com isso, a narrativa analisada sinaliza que, novamente, a empresa Valve continua sendo relevante, destacando-a em mais um gráfico.



BENEFIT: 0-10 scale measuring how beneficial a game is.

Figura 17 – Gráfico reestruturado correspondente ao gráfico da figura 16.

Por fim, iremos alterar o último gráfico da análise e visualização feita no projeto da segunda metade do curso de Introdução a Ciência de Dados:

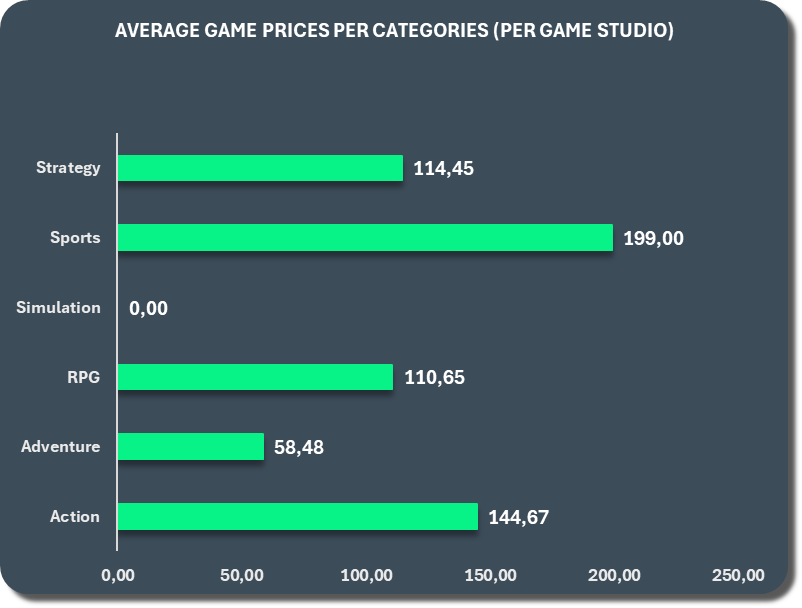


Figura 18 – Gráfico que representa a média de preço por jogo por categorias por empresa (seletor de dados.

Neste gráfico, temos todos os problemas já vistos anteriormente. São eles: ordenamento e estreitamento das colunas, sombreamento, entre outros.

Além de resolver tais problemas, enfatizamos a coluna com maior preço, retiramos o eixo, que não é necessário para compreender o gráfico, e colocamos a legenda do eixo x para identificar a moeda.

Especificamente nesse gráfico, seria difícil dar ênfase em uma empresa ou jogo em específico, por isso apenas realizamos as mudanças necessárias e citadas anteriormente.

A graph of numbers and a bar

Description automatically generated with medium confidence

Figura 19 – Gráfico reestruturado correspondente ao gráfico da figura 18.

**3 CONCLUSÕES**

Dessa forma, observa-se que a melhoria proposta nos gráficos originais permitiu uma evolução da plena compreensão da base de dados originária, além de destacar padrões e tendências que antes não eram evidentes.

Portanto, a partir de tais visualizações reestruturadas e dos dados coletados, podemos inferir que a empresa “Valve” e seus jogos relacionados merecem destaque, pois aparecem, na maior parte dos gráficos, em posições elevadas.

Ademais, a categoria de jogos de ação detém a maior concorrência dentre todas as categorias, refletindo seu apelo universal e ampla base de jogadores, o que sugere oportunidades estratégicas para exploração de mercado e novos lançamentos nessa área.

**REFERÊNCIAS**

KNAFLIC, Cole Nussbaumer. *Storytelling with data: a data visualization guide for business professionals*. Hoboken: Wiley, 2015.