

# Ambiente Automatizado para Análise de Dados para o Parque da Ciência da UFMS

Daniel Yudi de Carvalho  
Luciana Montera (Orientadora)

Faculdade de Computação  
Universidade Federal de Mato Grosso do Sul (UFMS)  
Caixa Postal 549 – 79.070-900 – Campo Grande – MS  
daniel.yudi@ufms.br, luciana.montera@ufms.br

***Abstract.** This article presents the development of an automated environment for data analysis of feedback collected through an NPS (Net Promoter Score) system at the UFMS Science Park. Using Python and Dash, a data visualization platform was created that allows the management team to effectively analyze and interpret visitor feedback, aiming to improve the services offered and promote a better user experience.*

***Resumo.** Este artigo apresenta o desenvolvimento de um ambiente automatizado para análise de dados de feedback coletados por meio de um sistema NPS (Net Promoter Score) do Parque da Ciência da UFMS. Utilizando Python e Dash, foi criada uma plataforma de visualização de dados que permite à equipe gestora analisar e interpretar de forma eficaz os feedbacks dos visitantes, visando aprimorar os serviços oferecidos e promover uma melhor experiência aos usuários.*

## 1. Introdução

Os espaços não formais de aprendizagem, como museus e centros de ciência, desempenham um papel fundamental na complementação da educação formal, proporcionando experiências práticas e interativas que estimulam o interesse e a curiosidade dos participantes [1]. Segundo Bourdieu [2], a relação entre os museus e seu público é essencial para a construção de significados e experiências enriquecedoras. Na sociedade contemporânea, esses espaços ganham ainda mais relevância ao promoverem o diálogo entre ciência e sociedade [3].

O Parque da Ciência da UFMS é um exemplo de espaço não formal que busca aproximar a ciência da comunidade, oferecendo atividades educativas e lúdicas que contribuem para a divulgação científica e o enriquecimento cultural [4]. Criado com o objetivo de promover a interação entre a universidade e a sociedade, o Parque da Ciência disponibiliza monumentos de ciência interativos que estimulam e despertam a curiosidade dos visitantes.

Para assegurar que um espaço socioeducativo como o Parque da Ciência cumpra seu propósito de ser ao mesmo tempo atrativo e pedagógico, tornou-se essencial desenvolver um método eficiente para coleta e análise de feedbacks. Com essa necessidade em

mente, foi idealizado um aplicativo que promovesse uma maior interação entre os visitantes e os organizadores do parque. Além disso, visando otimizar os futuros planejamentos de melhorias, foi concebida uma segunda aplicação, projetada para oferecer aos usuários uma visão mais clara e detalhada dos feedbacks coletados pela primeira ferramenta.

Para avaliar a qualidade dos serviços prestados e identificar oportunidades de melhoria, foi implementado um aplicativo de avaliação baseado no sistema NPS (*Net Promoter Score*). O NPS é uma métrica amplamente utilizada para medir a satisfação e lealdade dos clientes [7], sendo adaptada neste contexto para avaliar a experiência dos visitantes ao Parque, após a realização de visita guiada por monitores do Parque da Ciência da UFMS. A compreensão da experiência do visitante é essencial para a melhoria contínua do espaço.

O volume e a complexidade dos dados coletados podem dificultar sua interpretação ou omitir informações relevantes para a gestão. Assim, a análise dos dados coletados pela aplicação NPS (também chamada de feedbacks) é essencial para compreender a satisfação dos visitantes, identificar pontos fortes e fracos dos serviços oferecidos e orientar estratégias de aprimoramento das visitas guiadas.

Este trabalho apresenta o desenvolvimento de um ambiente automatizado para análise de dados, utilizando Python e Dash, que permite a visualização e interpretação de forma simples e direta dos feedbacks coletados após as visitas. O restante deste trabalho está organizado da seguinte forma: a seção 2 detalha a metodologia empregada para o desenvolvimento do ambiente; na seção 3 é apresentado o desenvolvimento do ambiente automatizado; na seção 4, são discutidos os resultados deste trabalho e, por fim, na seção 5 é apresentada a conclusão deste trabalho.

## 2. Metodologia

Para desenvolver o ambiente automatizado de análise de dados, utilizou-se a linguagem de programação Python devido à sua versatilidade e ampla gama de bibliotecas para ciência de dados [11]. Em particular, o *framework* Dash foi escolhido para a criação de aplicações web interativas, permitindo a construção de *dashboards* dinâmicos [12].

Os dados coletados pelo aplicativo NPS utilizado para avaliação das visitas ao Parque da Ciência são armazenados em arquivos em formato CSV. A cada visita, são cadastradas informações como nome do responsável pela visita, nome da escola, e-mail do servidor responsável, número de estudantes, cidade e bairro da escola visitante, e faixa de idades.

A segunda etapa de coleta envolve a obtenção de feedbacks de três tipos distintos de atores: os alunos visitantes, o(s) professor(es) responsável(is) pelo grupo e o monitor ou servidor do parque que acompanha o grupo. Os alunos expressam suas opiniões sobre o passeio unicamente pela escolha de uma “carinha” que representa sua percepção da experiência. Os responsáveis pela visita, além de escolher uma “carinha”, também podem selecionar caixas de seleção que complementam/justificam o voto escolhido com opções pré-definidas, além de um campo de texto livre para inserir feedbacks personalizados.

Como a aplicação de coleta de feedbacks foi usada poucas vezes até o momento, o volume de dados disponíveis é pequeno. Por isso, optou-se por gerar dados sintéticos para que a aplicação de análise fosse desenvolvida e validada. Utilizou-se uma abordagem

baseada em IA Generativa para simular os feedbacks coletados. Esta técnica permitiu a geração de um conjunto de dados que imitava as características esperadas dos dados reais, incluindo variáveis como avaliações numéricas, respostas a questões objetivas e comentários textuais.

Foram gerados dados sintéticos correspondentes a 22 escolas, totalizando 3000 participantes, provenientes de 10 municípios diferentes. O arquivo gerado e utilizado para a criação e validação do ambiente de análise está disponível para consulta <sup>1</sup>. Os arquivos reais gerados pela aplicação ficarão armazenados em disco na própria aplicação.

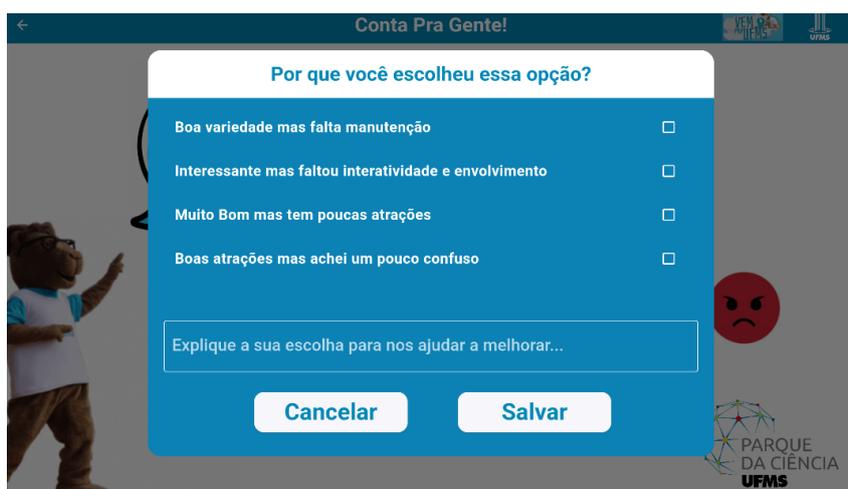
### 3. Desenvolvimento do Ambiente Automatizado

Antes de descrever o ambiente, faz-se necessário detalhar como os dados de uma aplicação NPS são analisados. As Figuras 1 e 2 mostram a tela principal de avaliação que é apresentada aos visitantes ao término da visita e as opções de justificativas, respectivamente. Nesta tela, o visitante irá escolher um ícone que representa a sua satisfação em relação à visita realizada. Quando o respondente é um professor ou responsável pela visita, a avaliação é complementada pela seleção de algumas justificativas, como “Variedade de Atrações”, “Boa estrutura, mas falta manutenção”, “Tem poucas atrações”.



Figura 1. Ícones para avaliação da visita

<sup>1</sup><https://drive.google.com/drive/folders/1JDyyZmgdKcAFxreQwQovAma-Gxdk1SXx?usp=sharing>



**Figura 2. Opções de feedback**

Dependendo da resposta do usuário, este é classificado como *promotor*, *passivo* ou *detrator* [7]. Cada avaliação contribui com um valor inteiro ((a) - 5 pontos; (b) - 4 pontos; (c) - 3 pontos; (d) - 2 pontos e (e) - 1 ponto) para o cálculo do NPS final de uma visita, conforme descrito na Equação 1.

$$NPS = \left( \frac{\text{Total de Promotores} - \text{Total de Detratores}}{\text{Total de Respondentes}} \right) \times 100 \quad (1)$$

Alguns usuários poderão ainda fornecer mais detalhes que justificam suas respostas. Estes poderão selecionar opções de feedback que reforçam o voto dado ou ainda oferecer um feedback customizado em texto livre. Assim, a cada visita é possível atribuir um nível de satisfação geral dos visitantes que varia entre 1 e 5.

Por meio de scripts Python, foram realizadas etapas de pré-processamento dos dados contidos no arquivo .csv resultante de uma visita, incluindo conversões de tipo e criação de colunas adicionais, como a classificação das avaliações em categorias de NPS (promotores, passivos e detratores). Os dados de múltiplos arquivos csv<sup>2</sup> foram agregados em um único *DataFrame*, permitindo uma análise consolidada das informações. Foram criados contadores para contabilizar o número de promotores, passivos e detratores, permitindo o cálculo do NPS para o período agregado. Além disso, uma função de agrupamento foi utilizada para agrupar os dados por escolas, o que permitiu a análise da distribuição das escolas que participaram da visita. Isso forneceu *insights* sobre quais escolas estavam mais engajadas com o Parque da Ciência.

Para a visualização dos dados, utilizou-se o Dash [14] a biblioteca *Plotly Express* [12]. Foram criados diversos gráficos interativos, como:

- Gráfico de barras apresentando a ocorrência dos feedbacks selecionados como “falta de funcionários”, “Boa estrutura”, “Bom atendimento”;
- Nuvem de palavras mostrando a ocorrência de palavras nos feedbacks em texto livre;

<sup>2</sup>resultante de um conjunto de visitas

- Gráfico de barras apresentando a distribuição de visitas por município;
- Gráficos de linhas para visualização da distribuição das visitas ao longo do tempo (por data e hora);
- Gráficos de pizza para a análise da distribuição de promotores, passivos e detratores;

A biblioteca Seaborn também foi utilizada para criar um *heatmap* que representa a correlação entre as opções selecionadas e as avaliações dos visitantes. Essa visualização é uma ferramenta poderosa para identificar padrões e relações significativas nos dados, como, por exemplo, comentários relacionados à organização associados a votos positivos ou observações sobre a estrutura vinculadas a votos negativos.

O dashboard foi desenvolvido seguindo princípios de usabilidade e design intuitivo. Utilizou-se o *Dash Bootstrap Components* para estilizar a aplicação, tornando-a responsiva e compatível com diferentes dispositivos, tais como tablet e celular. A interface permite que os usuários apliquem filtros interativos, como seleção de escolas, municípios, avaliações e intervalos de datas, atualizando os gráficos e métricas em tempo real.

Embora a abordagem utilizada tenha se concentrado em usabilidade e design intuitivo, é importante destacar que existem outras metodologias e princípios de design que poderiam ser aplicados, como o design universal, centrado no usuário, inclusivo ou emocional. Essas abordagens não foram implementadas nesta aplicação, mas poderiam ser consideradas em futuras evoluções para atender a públicos específicos ou ampliar o alcance e acessibilidade do sistema.

Implementou-se tratamento de erros e mensagens informativas para casos em que os filtros selecionados resultem em conjuntos de dados vazios, garantindo uma experiência consistente e evitando confusão por parte do usuário. Além disso, incluiu-se documentação e comentários no código para facilitar futuras manutenções e expansões.

O ambiente automatizado permite à equipe do Parque da Ciência acessar *insights* valiosos a partir dos dados coletados. As principais funcionalidades implementadas incluem:

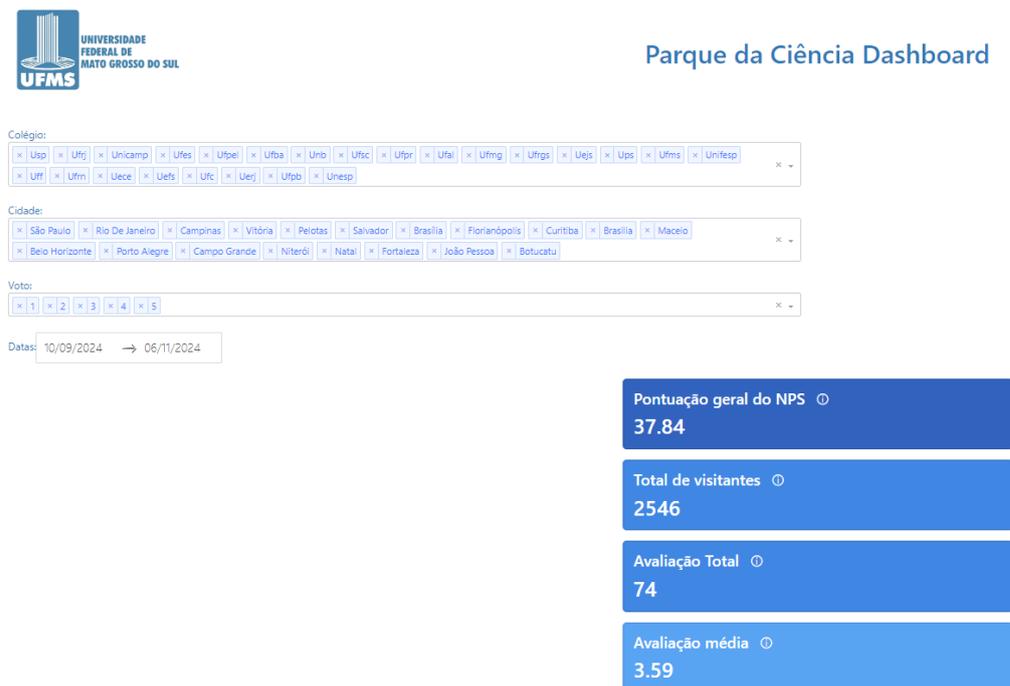
- Filtro de dados (região, data, escola) para *insights* personalizados
- Análise de dados temporais para estudo de tendências
- Análise de causas para avaliações negativas
- Identificação de oportunidades de melhoria
- Dashboard interativo para visualização intuitiva dos resultados
- Segmentação de públicos-alvo

#### 4. Resultados e Discussões

O desenvolvimento do ambiente automatizado para análise de dados de feedback do Parque da Ciência da UFMS apresenta-se como uma ferramenta promotora de percepção aos gestores quanto às visitas e trabalho prestado pelo Parque à comunidade interna e externa da UFMS e é capaz de auxiliar a tomada de decisões dos organizadores do parque para futuras melhorias. A utilização de Python e Dash permitiu a criação de uma plataforma interativa e acessível, transformando dados brutos em informações valiosas para a gestão. A plataforma está disponível no endereço <https://web-production-5848d.up.railway.app/>.

O dashboard proporciona uma maneira prática e visual de analisar a satisfação dos visitantes, facilitando a tomada de decisões e o monitoramento de atividades de forma eficiente.

A Figura 3 apresenta a porção inicial do dashboard, onde são apresentados os filtros disponíveis para a criação dos gráficos, como cidade, pontuação (NPS) e colégios que já visitaram o Parque. Com base nos filtros selecionados, são apresentados a pontuação geral do NPS das visitas, o total de visitantes e a pontuação média das avaliações.



**Figura 3. Visão inicial da aplicação Parque da Ciência da UFMS - Dashboard**

A pontuação média das avaliações é calculada conforme mostra a equação 2.

$$\text{Pontuação Média} = \frac{\text{Soma de todos os pontos}}{\text{Total de Visitantes}} \quad (2)$$

Nesta equação (2), o valor obtido representa a média das avaliações, refletindo o nível geral de satisfação dos visitantes.

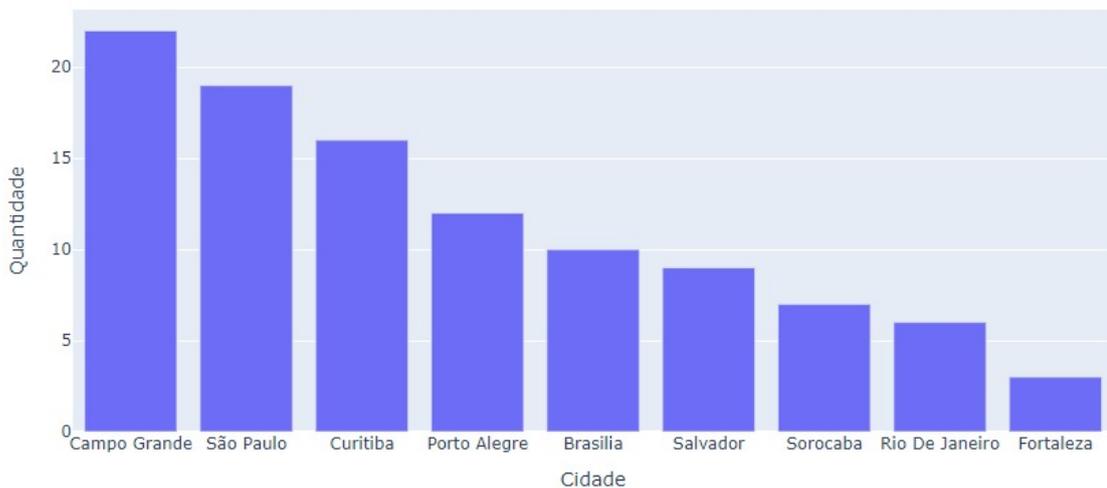
A seguir serão apresentados alguns dos gráficos gerados automaticamente pela aplicação, dada a seleção dos filtros disponíveis.

Na Figura 4 temos as ocorrências dos feedbacks selecionados após o monitor do parque votar. Essas opções servem para que o monitor possa reforçar o voto dado com uma justificativa pré-selecionada. Caso não haja, dentre as opções dadas, uma justificativa que endosse o voto do monitor, será possível oferecer o feedback em texto livre que também será coletado.

Na Figura 5 temos uma nuvem de palavras onde é possível identificar as frases e palavras mais utilizadas pelos usuários nas justificativas dadas à classificação assinalada.



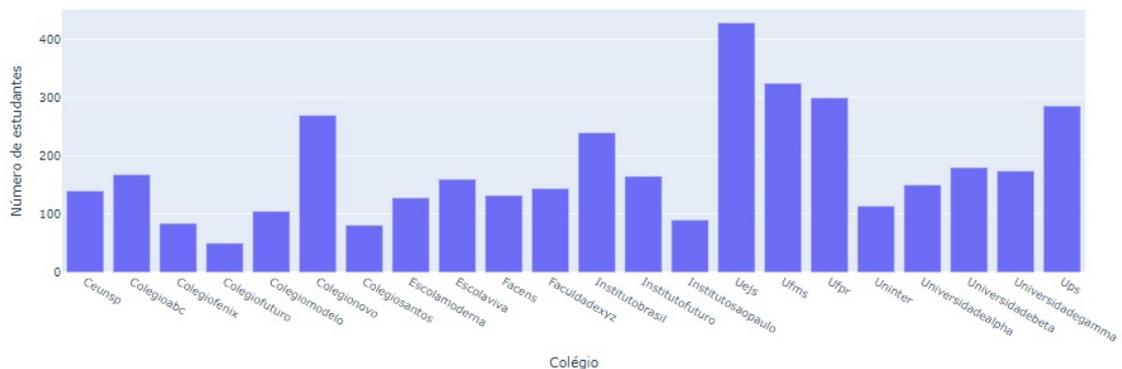
## Quantidade de Escolas por Cidades ⓘ



**Figura 6. Quantidade de visitas por cidade.**

Na Figura 7, é exibido um gráfico de barras que mostra o total de alunos por escola. Este gráfico permite visualizar quais escolas têm maior ou menor quantidade de alunos, oferecendo uma visão clara da distribuição do número de estudantes por instituição. Essas informações podem ser utilizadas para identificar possíveis oportunidades de engajamento com as escolas que têm uma maior ou menor participação e ainda apresentar potenciais escolas parceiras para projetos relacionados ao Parque da Ciência ou ainda projetos de interesse institucional.

## Total de alunos por escola ⓘ

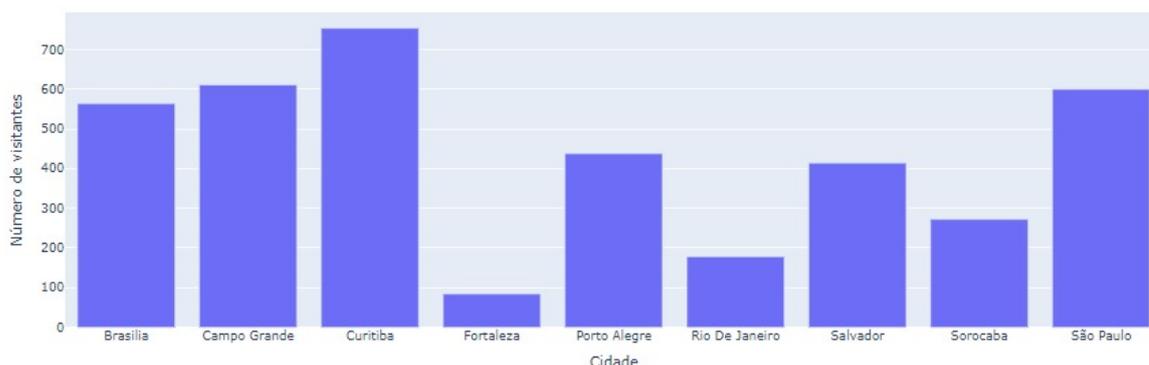


**Figura 7. Total de alunos por escola**

A Figura 8 mostra um gráfico de barras que apresenta o total de visitas por cidade. Este gráfico facilita a visualização de quais cidades apresentam o maior ou menor número

de visitas, permitindo uma análise mais aprofundada sobre a distribuição geográfica dos visitantes. Essas informações são valiosas para entender o alcance e a popularidade de determinados locais, apoiando a gestão em decisões estratégicas e ações de promoção direcionadas.

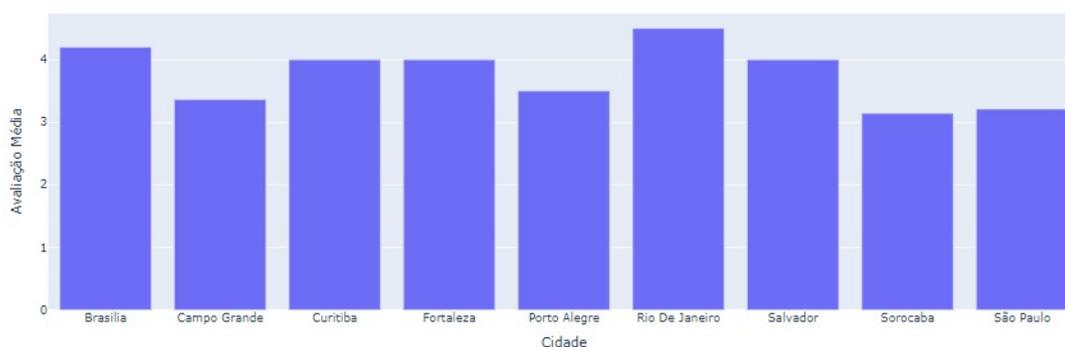
#### Total de visitas por cidade ⓘ



**Figura 8. Total de visitas por cidade**

A Avaliação Média por cidade é mostrada no gráfico da Figura 9. Este tipo de gráfico permite identificar diferenças na satisfação dos visitantes em diferentes localidades, ajudando a entender como a qualidade das experiências pode variar entre as cidades. Assim, a gestão pode utilizar essas informações para direcionar esforços de melhoria de maneira mais eficaz.

#### Avaliação média por cidade ⓘ



**Figura 9. Avaliação média por cidade**

A distribuição de visitantes por faixa etária é mostrada na Figura 10. Este gráfico facilita a compreensão de quais faixas etárias têm maior participação nas visitas. É de se esperar que a faixa etária predominante seja entre 8 e 16 anos, o que pode ajudar a equipe a direcionar esforços para atração de públicos de diferentes faixas etárias como,

por exemplo, pessoas idosas. O gráfico apresentado não reflete esta situação pois, como já informado, os dados utilizados neste trabalho são dados gerados aleatoriamente.

Distribuição de visitantes por faixa etária ⓘ

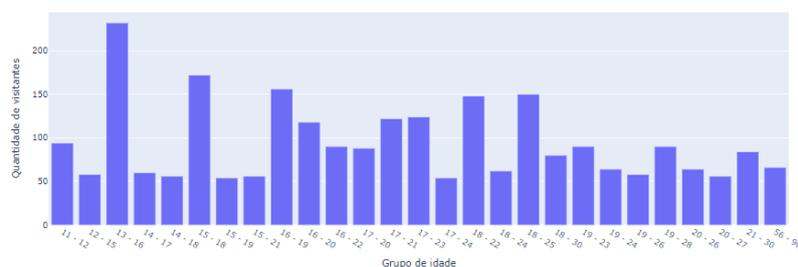


Figura 10. Distribuição de visitantes por faixa etária

NPS por cidade ⓘ

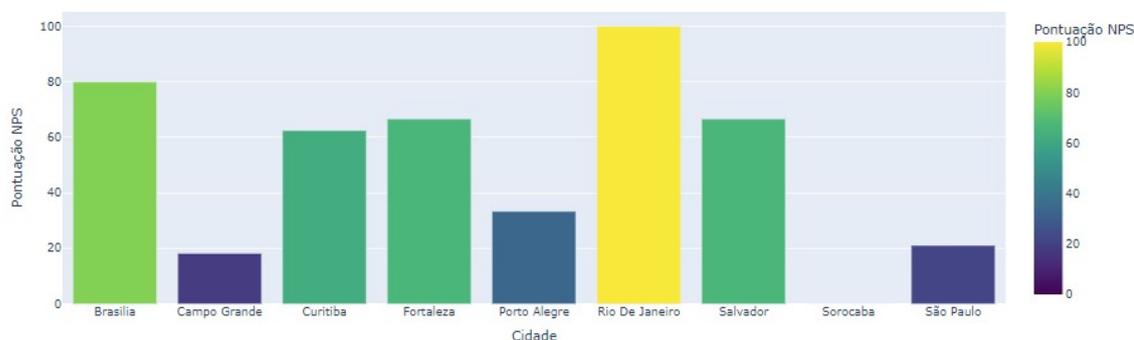


Figura 11. Gráfico indicadores de NPS

Na Figura 11, os indicadores de NPS são apresentados de acordo com a origem geográfica dos visitantes. Uma escala de cores é utilizada para indicar a pontuação NPS, que varia entre 100 (amarelo) e 0 (preto). Esta é uma métrica essencial pois resume a percepção dos visitantes.

Os gráficos das figuras 12 e 13, apresentam dados temporais que mostram o comportamento das visitas ao longo das horas do dia bem como dos dias da semana, respectivamente. Esses gráficos permitem identificar os períodos de maior e menor visita, possibilitando um planejamento mais eficiente de atividades e gerenciamento de recursos ao longo do ano e de políticas que melhorem a distribuição das visitas ao longo dos dias da semana e horários.

### Visitas por hora ①

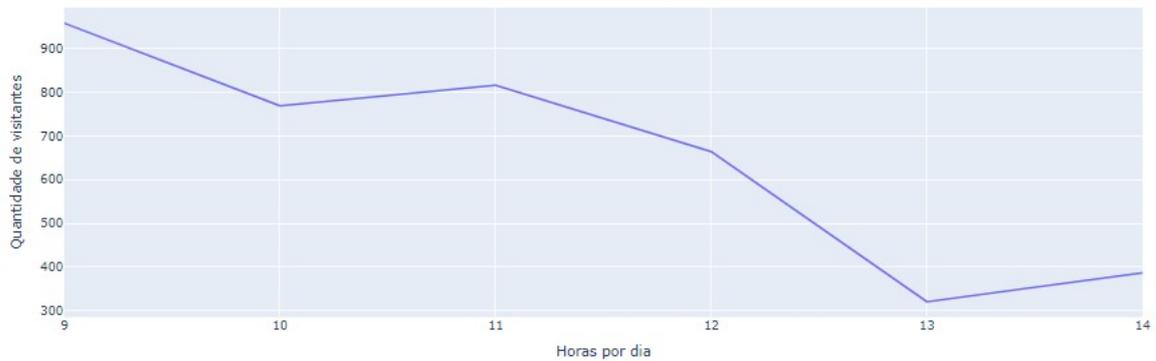


Figura 12. Dados temporais por hora

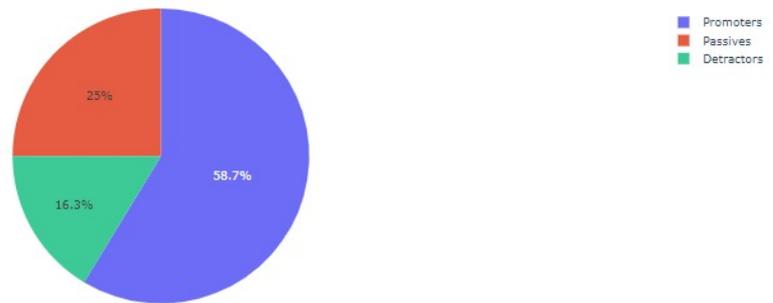
### Visitas por período ①



Figura 13. Dados temporais

Na Figura 14, é apresentada uma análise de NPS (*Net Promoter Score*), que avalia o nível de satisfação dos visitantes. O gráfico representa a distribuição dos visitantes nas categorias *promotores*, *passivos* e *detratores*, destacando suas respectivas porcentagens. Esses dados permitem entender melhor o perfil de satisfação dos visitantes, auxiliando em ações voltadas para a melhoria contínua do serviço, uma vez que deseja-se realizar visitas que maximizem o número de avaliadores *promotores*.

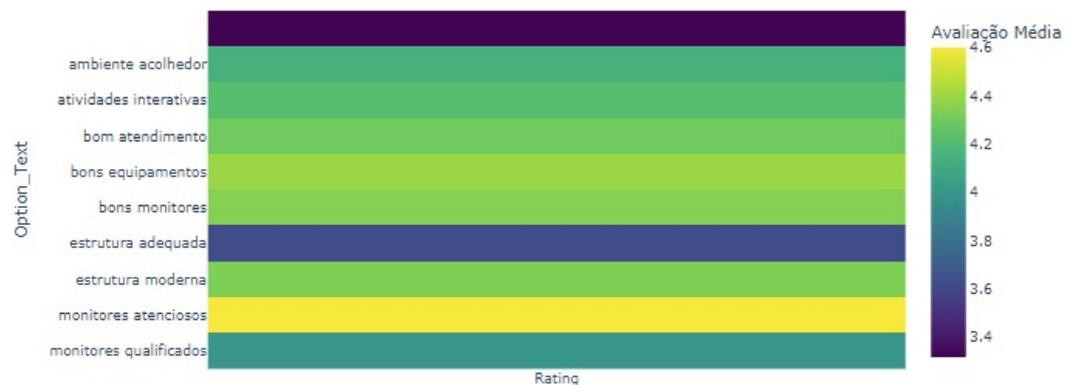
## Análise de Promotores, Passivos e Detratores ①



**Figura 14. Análise NPS**

Na Figura 15, é mostrado o gráfico de correlação entre as opções de feedback dos visitantes e suas respectivas avaliações, que permite analisar como cada tipo de feedback está correlacionado com a satisfação dos visitantes, fornecendo informações úteis para entender quais aspectos têm maior impacto, seja positivo ou negativo, nas avaliações realizadas pelos visitantes. A compreensão desta correlação é fundamental para a priorização de melhorias e a elaboração de estratégias que aumentem a satisfação dos visitantes.

## Correlação entre Opções e Votos ①



**Figura 15. Correlação entre opções e votos**

O ambiente desenvolvido facilita a análise de grandes volumes de dados, fornecendo *insights* sobre a satisfação dos visitantes, padrões de visitação e aspectos a serem aprimorados durante as visitas. Com isso, a equipe do Parque pode tomar decisões com base nas visitas já realizadas e, assim, elevar a qualidade dos serviços prestados.

O ambiente automatizado desenvolvido e disponibilizado oferece uma ferramenta valiosa para a equipe de gestão do Parque da Ciência da UFMS, pois contribui para a melhoria contínua dos serviços oferecidos e para o fortalecimento da relação entre a univer-

sidade e a comunidade. Dentre os direcionamentos que a análise apresentada pela ferramenta oferece, pode-se citar a identificação das potencialidades e fragilidades do serviço de visita guiada, bem como da infraestrutura geral oferecida pelo Parque, a adequação das políticas de divulgação/convite às visitas ao Parque que podem ser direcionadas a públicos específicos, cidades e horários.

Futuras expansões deste trabalho podem incluir:

- A integração de técnicas avançadas de processamento de linguagem natural para análise mais detalhada dos feedbacks textuais, como análise de sentimentos e identificação de tópicos.
- A incorporação de dados provenientes de outras fontes, como redes sociais, para ampliar a compreensão da percepção pública sobre o parque.
- A implementação de modelos preditivos para antecipar tendências de visita e satisfação.
- A realização de testes de usabilidade com a equipe gestora para aprimorar a interface e funcionalidade do dashboard.

## 5. Conclusão

A utilização de Python e Dash viabilizou a criação de uma plataforma interativa, transformando dados brutos em informações acionáveis. Com a análise dos feedbacks, a equipe gestora pode identificar pontos de melhoria e potencialidades, visando o aprimoramento contínuo dos serviços prestados. Futuramente, a integração de técnicas mais avançadas, como a análise de sentimentos e modelos preditivos, poderá enriquecer ainda mais a capacidade do sistema em oferecer *insights* valiosos, garantindo uma experiência cada vez mais positiva aos visitantes. Dessa forma, o Parque da Ciência da UFMS se consolida como um importante espaço de educação não formal e interação entre a comunidade e a universidade.

## Referências

- [1] Falk, J. H., Dierking, L. D. (2000). *Learning from museums: Visitor experiences and the making of meaning*. AltaMira Press.
- [2] Bourdieu, P., Darbel, A. (1969). *L'amour de l'art: les musées d'art européens et leur public*. Paris: Les Éditions de Minuit.
- [3] Giddens, A. (1991). *As consequências da modernidade*. São Paulo: Editora da UNESP.
- [4] Parque da Ciência da UFMS. (2018). *Sobre o Parque*. Disponível em: <http://www.parquedaciencia.ufms.br/sobre>. Acesso em: 28 out. 2023.
- [5] Pereira, L. F., Costa, J. M. (2019). Educação não formal e museus de ciência: potencialidades e desafios. *Cadernos de Educação*, 18(36), 123-139.
- [6] Beetlestone, J. G., Johnson, C., Quin, M., White, H. (1998). The science center movement: contexts, practice, next challenges. *Public Understanding of Science*, 7(1), 5-26.
- [7] Reichheld, F. F. (2003). The one number you need to grow. *Harvard Business Review*, 81(12), 46-55.

- [8] Davallon, J. (1992). Le musée est-il vraiment un média? *Publics et Musées*, 2, 99-124.
- [9] Kotler, P., Keller, K. L. (2017). *Administração de marketing*. 15ª ed. Pearson.
- [10] Godin, B., Gingras, Y. (2000). What is scientific and technological culture and how is it measured? *Public Understanding of Science*, 9(1), 43-58.
- [11] Van Rossum, G., Drake, F. L. (2007). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.
- [12] Plotly Technologies Inc. (2015). *Dash User Guide*. Disponível em: <https://dash.plotly.com/>. Acesso em: 28 out. 2023.
- [13] Dash Bootstrap Components (2018). *Dash User Guide*. Disponível em: <https://dash-bootstrap-components.opensource.faculty.ai/>. Acesso em: 28 out. 2023.
- [14] Plotly Technologies Inc. (2023). *Dash: Framework para construção de dashboards interativos*. Disponível em: <https://dash.plotly.com/>. Acesso em: 28 out. 2023.
- [15] McKinney, W. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, pp. 51-56.
- [16] Patki, N., Wedge, R., Veeramachaneni, K. (2016). The synthetic data vault. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 399-410.
- [17] Xu, L., Skoularidou, M., Cuesta-Infante, A., Veeramachaneni, K. (2019). Modeling tabular data using conditional GAN. In *Advances in Neural Information Processing Systems*, 32.