

Análise de Padrões em Posts de Ensino de Engenharia de Software no Reddit

Amanda Moura de Souza¹, Carlos Emanuel Silva e Melo Oliveira¹,
Diego Machado Cordeiro¹, Giovanni Bogliolo Sirihal Duarte¹,
Gustavo Andrade Alves¹, Luiz Gustavo Mendes Santos¹, Pedro Ramos Vidigal¹

¹Pontifícia Universidade Católica de Minas Gerais (PUC Minas)
Caixa Postal 1.686 – 30535.901 – Belo Horizonte – MG – Brazil

Abstract. *This study explores the characteristics of Reddit posts that contribute to knowledge sharing and exchange in the field of software engineering. Despite Reddit's popularity for discussions on various topics, little is known about the specific elements of posts that encourage meaningful interactions and collaborative learning in software engineering. The research examines aspects such as post format, content type, multimedia usage, and language, aiming to identify which factors increase engagement and discussion quality. This understanding is essential for developers seeking to filter useful information, community moderators, and content creators, as well as to inform the development of more effective recommendation systems. The findings may contribute to enhancing the online learning experience and promoting more productive interactions on social platforms.*

Resumo. *Este estudo explora as características de postagens no Reddit que contribuem para o compartilhamento e a troca de conhecimento na área de engenharia de software. Apesar da popularidade do Reddit para discussões de temas variados, pouco se sabe sobre os elementos específicos das postagens que incentivam interações significativas e aprendizado colaborativo na área de engenharia de software. A pesquisa examina aspectos como o formato da postagem, o tipo de conteúdo, o uso de multimídia e a linguagem utilizada, com o objetivo de identificar quais fatores aumentam o engajamento e a qualidade das discussões. Esse entendimento é essencial para desenvolvedores que buscam filtrar informações úteis, moderadores de comunidades e criadores de conteúdo, além de informar o desenvolvimento de sistemas de recomendação mais eficazes. Os resultados podem contribuir para melhorar a experiência de aprendizado online e incentivar interações mais produtivas nas plataformas sociais.*

1. Introdução

A área de engenharia de software vem se expandindo consideravelmente nos últimos anos, inclusive em fóruns online, onde usuários compartilham conhecimento. O Reddit, uma das maiores plataformas da internet, abriga várias comunidades dedicadas exclusivamente à engenharia de software, servindo como pontos de encontro para estudantes e entusiastas discutirem conceitos da área.

1.1. Problema

O Reddit é amplamente utilizado como uma plataforma para discussões técnicas e compartilhamento de conhecimento na área de engenharia de software. No entanto, pouco se sabe sobre quais elementos específicos das postagens são mais eficazes para fomentar interações significativas e promover o aprendizado entre os entusiastas e profissionais da área.

Essa lacuna de conhecimento limita a capacidade de otimizar o uso do Reddit como uma ferramenta para o desenvolvimento colaborativo de conhecimento em engenharia de software. Características como formato da postagem, tipo de conteúdo, uso de multimídia, escolha de palavras-chave e profundidade dos comentários podem influenciar significativamente o engajamento da comunidade e a qualidade das discussões técnicas.

Compreender essas dinâmicas pode fornecer insights valiosos para criadores de conteúdo, moderadores de comunidades técnicas e para a estruturação de sistemas de recomendação em plataformas sociais voltadas ao aprendizado colaborativo. Portanto, há uma necessidade urgente de investigações empíricas que analisem quais elementos de postagens são mais relevantes para estimular o engajamento e facilitar a troca de conhecimento especializado em engenharia de software.

1.2. Motivação

No cenário atual da engenharia de software, desenvolvedores enfrentam um fluxo constante e massivo de informações, muitas vezes oriundas de plataformas como o Reddit, onde ocorrem discussões técnicas, tutoriais e debates sobre práticas de programação. Essa vasta quantidade de conteúdo cria uma oportunidade e um desafio: oferece um espaço rico para aprendizado e troca de experiências, mas também dificulta a tarefa de discernir quais postagens agregam valor ao desenvolvimento profissional e técnico.

Dado o volume de informações, é fundamental que os desenvolvedores sejam capazes de identificar e priorizar as postagens que apresentam insights práticos, conceitos inovadores ou soluções comprovadas aplicáveis ao seu trabalho. A capacidade de distinguir conteúdos valiosos é essencial para que o tempo investido em redes sociais resulte em um ganho real de conhecimento.

Além disso, a sobrecarga de informações não filtradas pode reduzir a eficácia do aprendizado contínuo, habilidade indispensável para acompanhar as rápidas mudanças tecnológicas e metodológicas na programação. Compreender quais características de postagens são mais eficazes para promover a aquisição de conhecimento relevante no Reddit é uma questão relevante tanto para desenvolvedores quanto para pesquisas em sistemas de recomendação e análise de conteúdo em plataformas sociais.

Explorar essas questões permitirá otimizar o fluxo de informação para desenvolvedores e contribuir para o desenvolvimento de mecanismos que melhorem a experiência de aprendizado online e incentivem a troca de conhecimento de forma direcionada e eficaz.

1.3. Justificativa

O uso crescente de plataformas sociais como Reddit para discussões técnicas destaca a necessidade de entender melhor os mecanismos que facilitam a disseminação de conhecimento nesse contexto. O Reddit, com milhões de usuários, representa uma oportunidade

significativa para novos insights. Identificar os elementos de postagens mais eficazes para engajamento e aprendizado pode informar o desenvolvimento de melhores sistemas de recomendação, auxiliando desenvolvedores a navegarem no vasto conteúdo disponível e se concentrarem no que agrega valor. O estudo possui implicações práticas para moderadores e criadores de conteúdo, permitindo-lhes adaptar estratégias para promover discussões produtivas e incentivar uma cultura de aprendizado contínuo na comunidade. A pesquisa proposta tem potencial para contribuir com a literatura acadêmica, oferecendo um modelo analítico aplicável a outros contextos online, e fornecer diretrizes para melhorar a eficácia das interações educacionais em redes sociais. Copy

1.4. Objetivo

O objetivo principal deste trabalho é analisar as características de postagens populares sobre engenharia de software no Reddit, com a finalidade de identificar padrões que contribuam para um ensino eficaz e uma troca significativa de conhecimento. Esta análise se concentra em dois aspectos fundamentais: a qualidade do conteúdo educacional dos vídeos compartilhados e a geração de discussões e interações educativas do ponto de vista de estudantes e entusiastas da área.

Para alcançar esse objetivo, o estudo buscará:

1. **Identificar Elementos-Chave:** Examinar quais características específicas das postagens — como formato, tipo de conteúdo, uso de multimídia, linguagem e comentários — estão associadas a níveis mais altos de engajamento e apreciação na comunidade de engenharia de software no Reddit.
2. **Promover Interações Educativas:** Investigar como as postagens fomentam discussões produtivas e interações educativas, avaliando a natureza das conversas geradas, a diversidade de perspectivas e o potencial para aprofundamento técnico e teórico.
3. **Desenvolver Diretrizes Práticas:** Com base nos padrões identificados, elaborar diretrizes práticas para criadores de conteúdo e moderadores que desejam otimizar suas postagens para maximizar o impacto educacional e o engajamento da comunidade.
4. **Contribuir para Sistemas de Recomendação:** Fornecer insights valiosos para o desenvolvimento de sistemas de recomendação mais eficazes, capazes de destacar conteúdos que realmente agreguem valor ao aprendizado dos usuários, facilitando a navegação e seleção de informações relevantes.

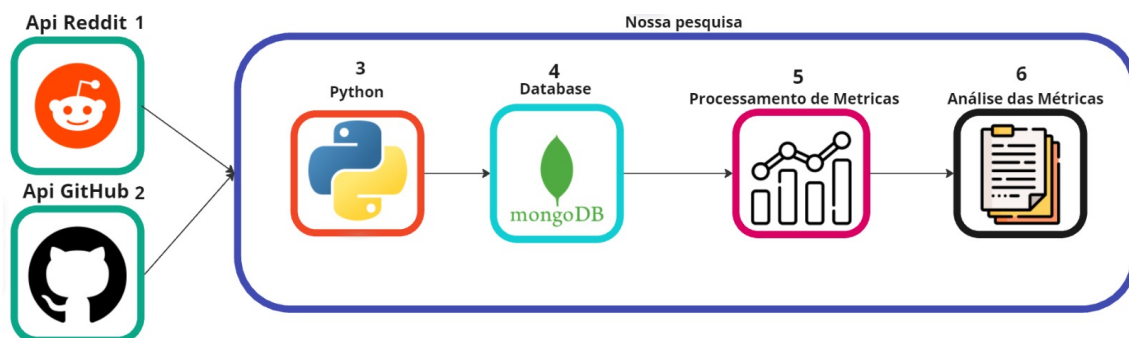
Ao atingir esses objetivos, este trabalho pretende não apenas enriquecer a compreensão acadêmica sobre a dinâmica de compartilhamento de conhecimento em plataformas sociais, mas também oferecer soluções práticas que aprimorem a experiência de aprendizado online para estudantes e entusiastas de engenharia de software.

2. Questões de pesquisa

1. **RQ 1 - Posts de quais linguagens têm maior engajamento?** : Para esta questão serão usadas as métricas de engajamento (UpVote e Comentários) , e a citação da linguagem no título do post. A hipótese estabelecida é que Python, JavaScript e HTML/CSS serão as linguagens que mais geram engajamento.

2. **RQ 2 - Quais tipos de mídia geram mais engajamento?:** Para esta questão serão usadas as métricas de engajamento (UpVote e Comentários) e o formato em que o post foi feito (Link, Video, Texto, Enquete). A hipótese inicial é que os posts em formato de discussão e texto são mais populares que enquetes e links.
3. **RQ 3 - O uso de tags influencia no engajamento?:** Para esta questão serão usadas as métricas de engajamento (UpVote e Comentários), e a citação da tag no post. A hipótese estabelecida é que posts com tags apesar de estar em menor quantidade terão maior engajamento

3. Metodologia



A Metodologia segue um total de 6 passos, sendo eles:

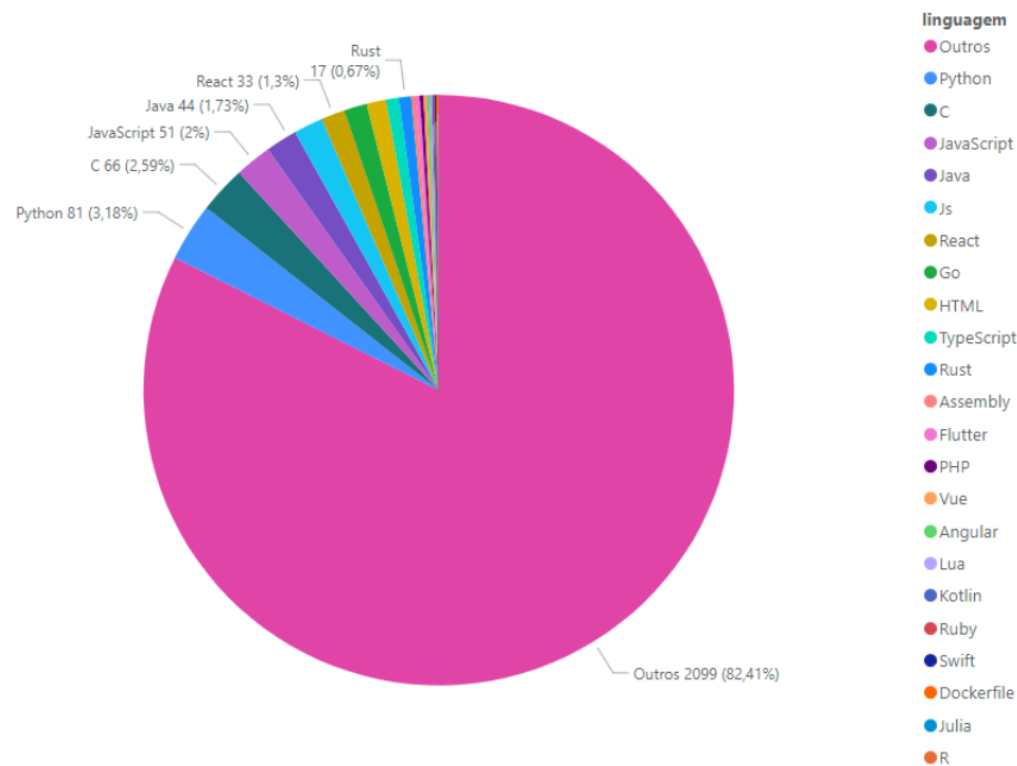
1. **API Reddit:** A API gratuita do Reddit será utilizada para coletar uma amostra de dados dos posts nos subreddits r/learnprogramming, r/webdev e r/programming na plataforma. Os dados coletados são: 'tags', 'descrição do post', 'título do post', 'upvotes', 'comentários', 'compartilhamentos'.
2. **API Github:** Será usada a API do GitHub para coletar dados sobre as linguagens de programação e, posteriormente, montar uma base para análise dos posts das comunidades do Reddit.
3. **Código Python:** Será criado um script em Python que realiza requisições à API para extrair informações relacionadas aos posts nos subreddits selecionados. O script coletará os dados citados no item 1 desses vídeos e vai armazenar em um banco de dados MongoDB.
4. **MongoDB:** Os dados serão armazenados em um banco de dados MongoDB, que é um sistema de banco de dados não relacional orientado a documentos.
5. **Processamento de Métricas:** Processar os dados usando Power BI. Algumas métricas, como upvotes e Comentários serão usado para calcular engajamento, em seguida, realizando o cruzamento do engajamento com outras métricas estabelecidas anteriormente.
6. **Análise das métricas:** Interpretação das métricas geradas a partir do script em Python visando responder às 3 questões inicialmente levantadas no modelo GQM, visando entender as características dos posts mais bem-sucedidos.. Uma análise comparativa será realizada entre as hipóteses iniciais levantadas (sobre padrões de sucesso e ensino no Reddit) e os resultados efetivamente obtidos com os dados coletados, buscando validar ou refutar as premissas da pesquisa.

4. Resultados

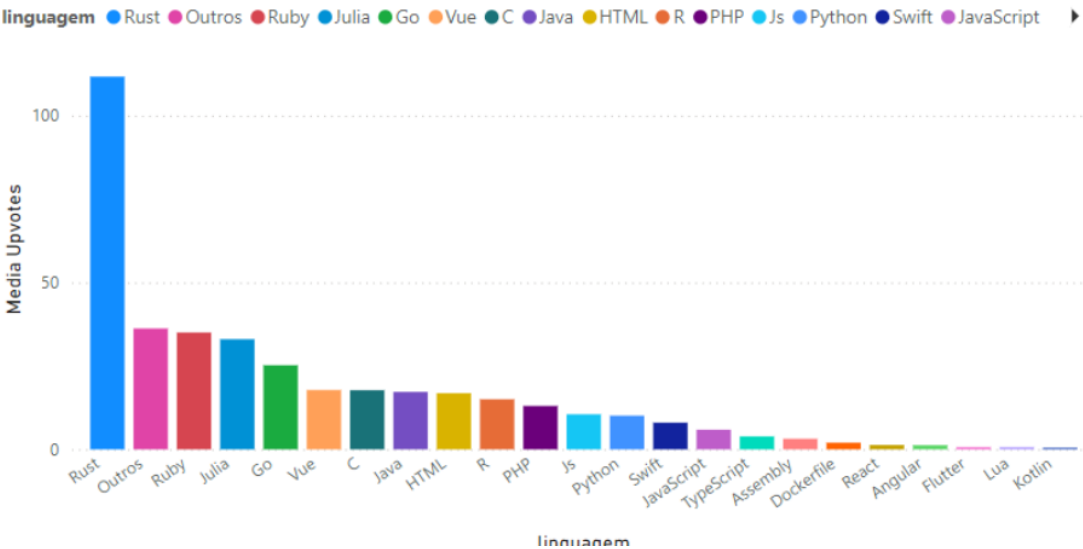
Essa área é dedicada a mostrar os resultados obtidos a partir da ferramenta power BI

4.1. RQ1 - Engajamento por linguagens

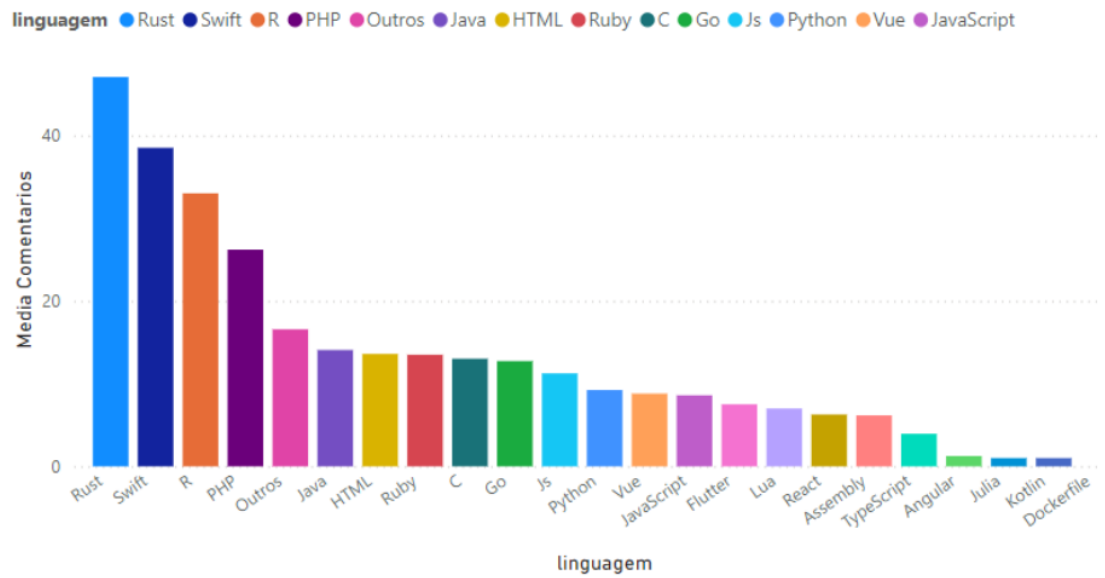
Medida por linguagem



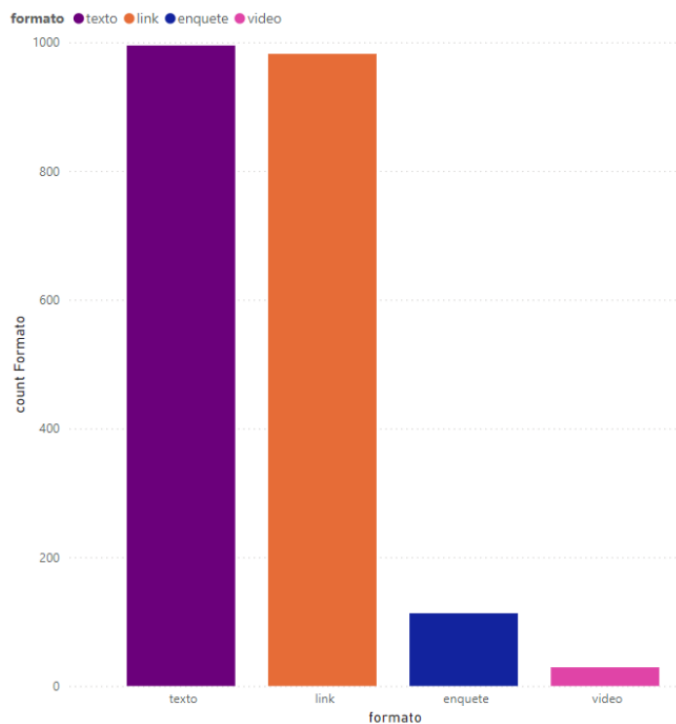
Media Upvotes por linguagem e linguagem



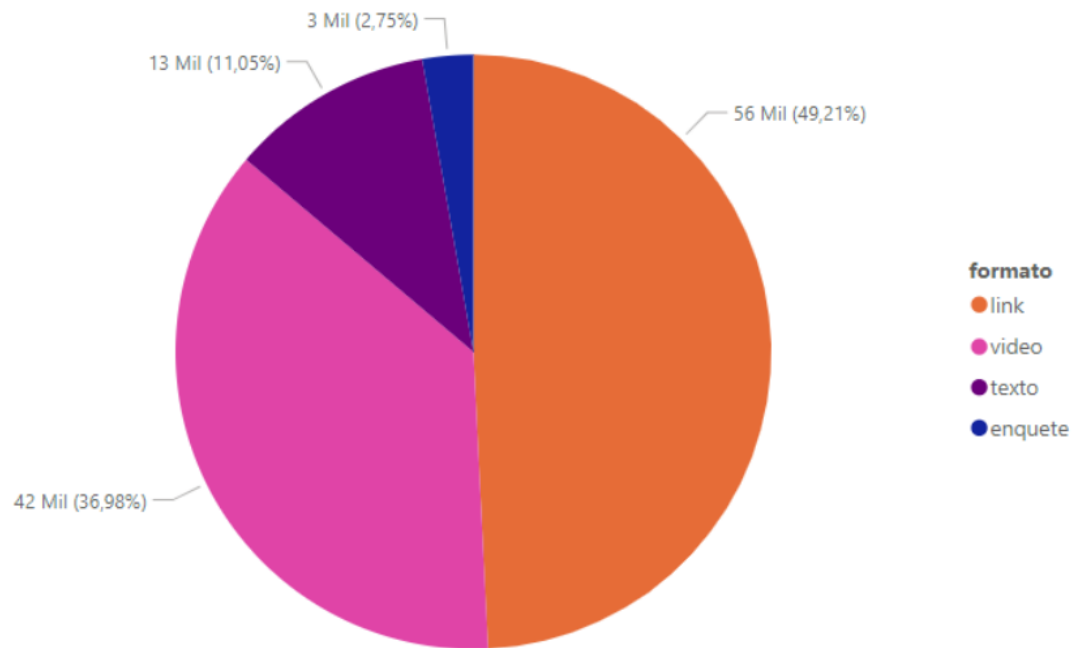
Media Comentarios por linguagem e linguagem



4.2. RQ2 - Engajamento por formato de post



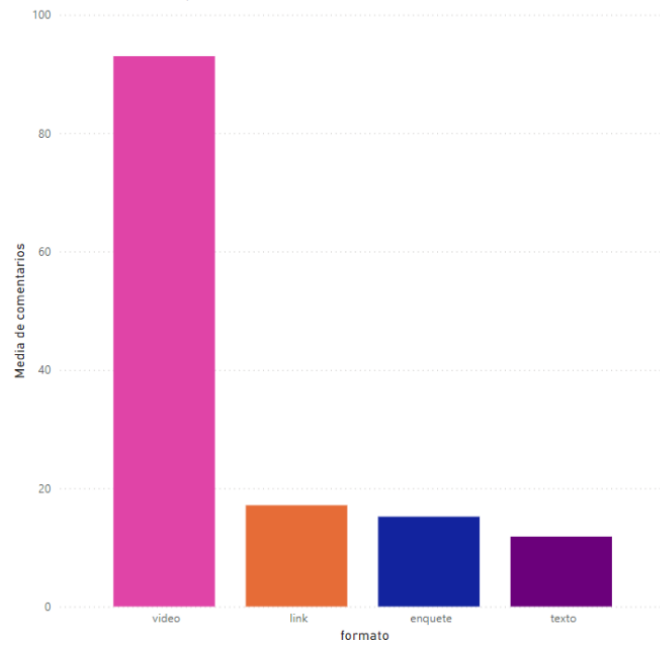
Soma de upvotes por formato



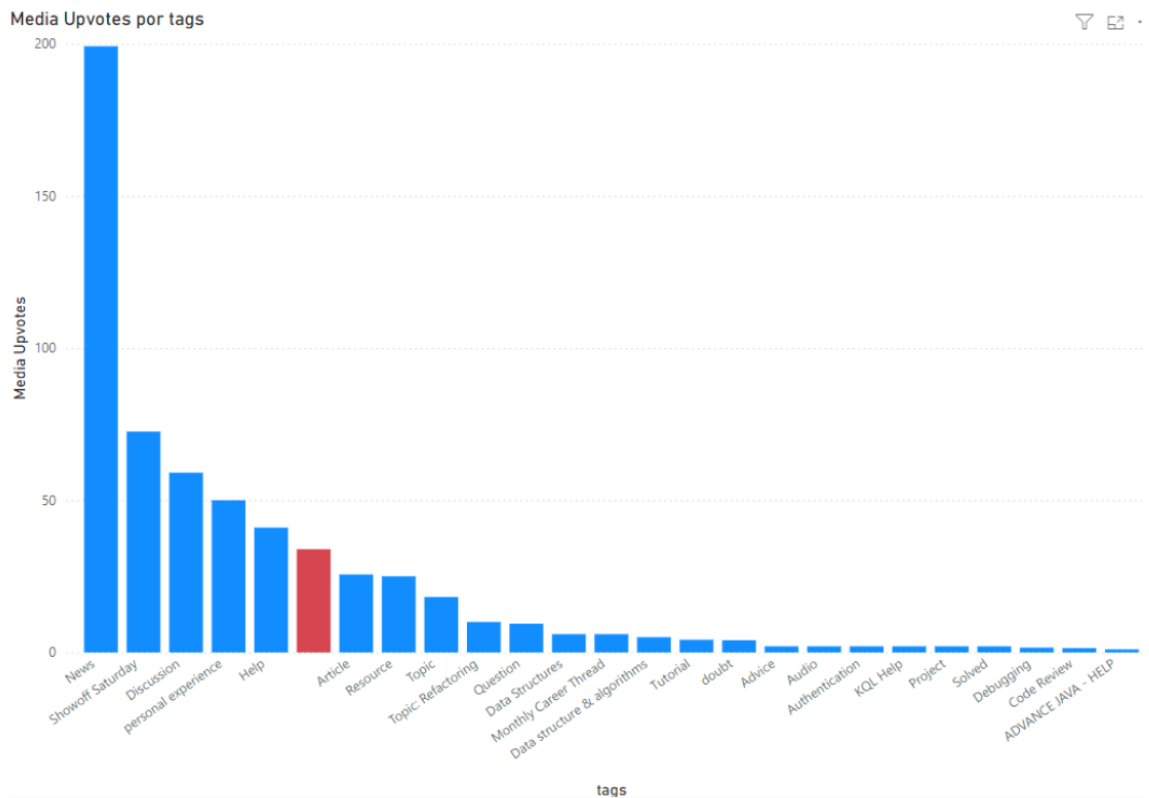
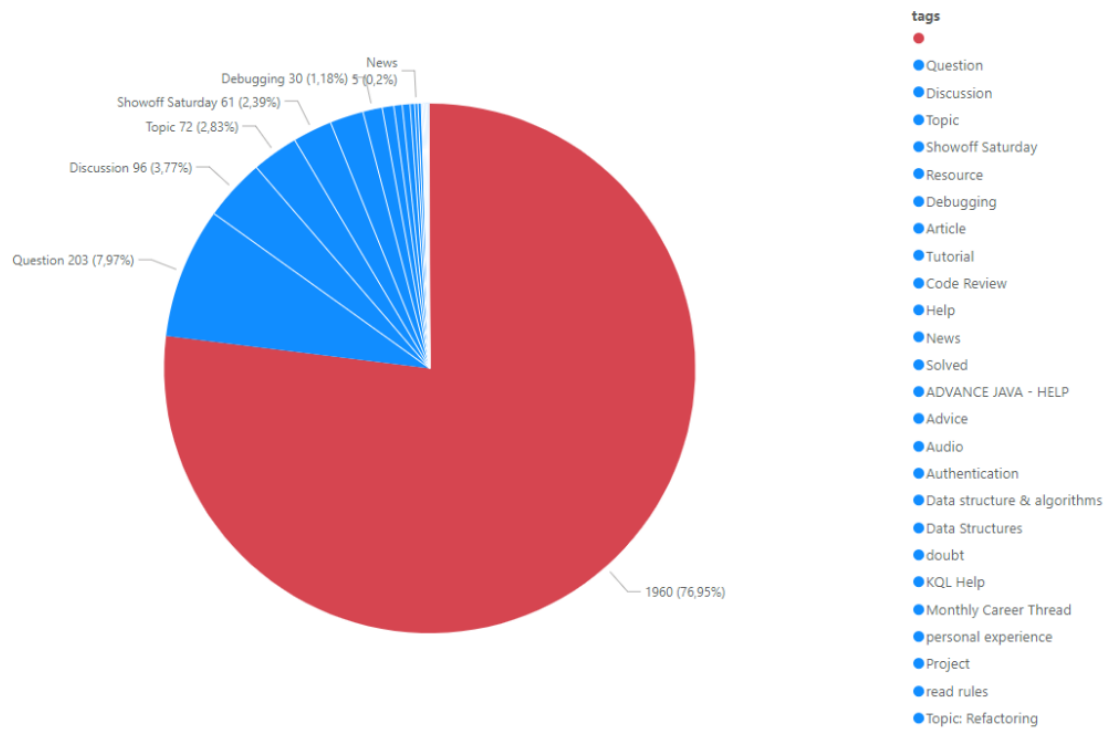
Media de comentarios por formato e formato



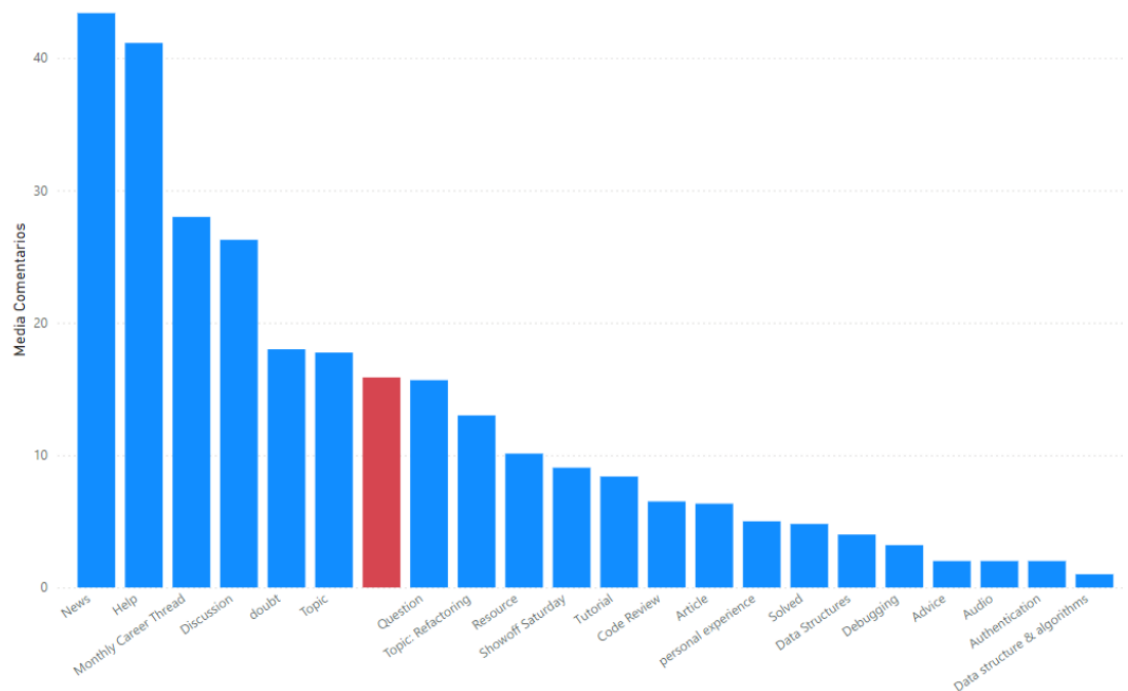
formato video link encuesta texto



4.3. RQ3 - Engajamento por tags



Media Comentarios por tags



5. Análise dos resultados

Essa sessão é dedicada a mostrar a análise dos resultados para cada uma das questões estabelecidas, comparando o resultado com as hipóteses levantadas inicialmente

5.1. RQ 1 - Posts de quais linguagens têm maior engajamento?

A hipótese estabelecida inicialmente foi que Python, JavaScript e HTML/CSS seriam as linguagens que mais geram engajamento. Esta hipótese foi negada, quando olhamos apenas número de posts que mencionam essas linguagens elas mostram grande volume porém, ao calcularmos o engajamento, podemos perceber que linguagens como Rust, Swift e Julia, receberam mais atenção e engajamento.

5.2. RQ 2 - Quais tipos de mídia geram mais engajamento?

A hipótese estabelecida inicialmente foi que os posts em formato de discussão e texto são mais populares que enquetes e links. Essa hipótese se provou errada, ao analisarmos o engajamento dos posts, podemos perceber que posts que são links para sites externos, como sites de notícias ou artigos são os posts com maior engajamento na plataforma

5.3. RQ 3 - O uso de tags influencia no engajamento?

A hipótese estabelecida inicialmente foi que posts com tags apesar de estar em menor quantidade terão maior engajamento. Essa hipótese foi confirmada, apesar de representar um número muito menor, os posts que contêm tags, principalmente as tags de New e Showoff Saturday, foram os posts que fizeram mais sucesso com maior engajamento

6. References

Referências usadas como base para criação do artigo [Fang et al. 2023]. [Liu and Anwar 2022]. [Whitfield et al. 2021].

The references must be listed using 12 point font size, with 6 points of space before each reference. The first line of each reference should not be indented, while the subsequent should be indented by 0.5 cm.

References

- Fang, H., Vasilescu, B., and Herbsleb, J. (2023). Understanding information diffusion about open-source projects on twitter, hackernews, and reddit. In *2023 IEEE/ACM 16th International Conference on Cooperative and Human Aspects of Software Engineering (CHASE)*, pages 56–67.
- Liu, Y. and Anwar, M. (2022). Learning programming in social media: An nlp-powered reddit study. In *2022 Fourth International Conference on Transdisciplinary AI (TransAI)*, pages 55–58.
- Whitfield, C., Liu, Y., and Anwar, M. (2021). Surveillance of covid-19 pandemic using social media: a reddit study in north carolina. In *Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB '21*, New York, NY, USA. Association for Computing Machinery.