



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

JUAN VICTOR LYRA BARROS E BARROS

**ANDROMEDEV:
UMA ANÁLISE SOBRE A EXPERIÊNCIA EM
DESENVOLVIMENTO DE CÓDIGO ABERTO NA UFCG**

CAMPINA GRANDE - PB

2021

JUAN VÍCTOR LYRA BARROS E BARROS

ANDROMEDEV:

**UMA ANÁLISE SOBRE A EXPERIÊNCIA EM
DESENVOLVIMENTO DE CÓDIGO ABERTO NA UFCG**

**Trabalho de Conclusão de Curso
apresentado ao Curso Bacharelado em
Ciência da Computação do Centro de
Engenharia Elétrica e Informática da
Universidade Federal de Campina
Grande, como requisito parcial para
obtenção do título de Bacharel em
Ciência da Computação.**

Orientadora: Lívia Maria Rodrigues Sampaio Campos.

CAMPINA GRANDE - PB

2021



B277a Barros, Juan Victor Lyra Barros e.
Andromedev: uma análise sobre a experiência em desenvolvimento de código aberto na UFCEG . / Juan Victor Lyra Barros e Barros. - 2021.

14 f.

Orientadora: Profa. Dra. Livia Maria Rodrigues Sampaio.

Trabalho de Conclusão de Curso - Artigo (Curso de Bacharelado em Ciência da Computação) - Universidade Federal de Campina Grande; Centro de Engenharia Elétrica e Informática.

1. Código aberto. 2. Mentoria. 3. Andromedev. 4. Desenvolvimento de código aberto. 5. OpenDevUFCEG. 6. Cultura de código aberto. 6. Programas de mentoria em desenvolvimento de código aberto. I. Sampaio, Livia Maria Rodrigues. II. Título.

CDU:004.4'415(045)

Elaboração da Ficha Catalográfica:

Johnny Rodrigues Barbosa
Bibliotecário-Documentalista
CRB-15/626

JUAN VICTOR LYRA BARROS E BARROS

ANDROMEDEV:

**UMA ANÁLISE SOBRE A EXPERIÊNCIA EM
DESENVOLVIMENTO DE CÓDIGO ABERTO NA UFCG**

**Trabalho de Conclusão de Curso
apresentado ao Curso Bacharelado em
Ciência da Computação do Centro de
Engenharia Elétrica e Informática da
Universidade Federal de Campina
Grande, como requisito parcial para
obtenção do título de Bacharel em
Ciência da Computação.**

BANCA EXAMINADORA:

**Professor Dr.(a.) Livia Maria Rodrigues Sampaio Campos
Orientador – UASC/CEEI/UFCG**

**Professora Dr.(a.) Everton Leandro Galdino Alves
Examinador – UASC/CEEI/UFCG**

**Professor Tiago Lima Massoni
Professor da Disciplina TCC – UASC/CEEI/UFCG**

Trabalho aprovado em: 25 de Maio de 2021.

CAMPINA GRANDE - PB

ABSTRACT¹

Andromedev was a mentoring event in open source development that had its first edition based on consolidated events of the same nature. To evaluate it, forms were used to capture data about the development process and information about the projects. However, the volume of data collected had a negative impact on its interpretability. In this context, the present work develops a descriptive analysis of the deliveries generated in Andromedev, in order to identify the target audience of the event; examine your evaluating artifacts; verify possible characteristics that led to the success of projects; and to look for factors that engaged students to continue contributing after the end of the event. A good adherence to the expected target audience was noted, however some aspects need to be improved for upcoming editions, such as weekly deliveries, the period of the event and the division of mentor time between various projects, this being a characteristic that impacted in the conclusion of projects and in the desire to continue contributing for the apprentices.

¹ Caso seu artigo esteja em inglês, coloque aqui o resumo em português; caso esteja o artigo em português, coloque aqui o resumo em inglês.

Andromedev: uma análise sobre a experiência em desenvolvimento de código aberto na UFCG

Juan Victor Lyra Barros e Barros*
juan.barros@ccc.ufcg.edu.br
Universidade Federal de Campina Grande
Campina Grande, Paraíba

Orientadora: Lívia Maria Rodrigues Sampaio
Campos*
livia@computacao.ufcg.edu.br
Universidade Federal de Campina Grande
Campina Grande, Paraíba

RESUMO

O Andromedev foi um evento de mentoria em desenvolvimento de código aberto que teve sua primeira edição baseada em eventos consolidados de mesma natureza. Para avaliá-lo, foram utilizados formulários para a captura de dados sobre o processo de desenvolvimento e informações sobre os projetos. Contudo, o volume de dados coletados impactou negativamente em sua interpretabilidade. Nesse contexto, o presente trabalho desenvolve uma análise descritiva sobre as entregas geradas no Andromedev, a fim de identificar o público-alvo do evento; examinar seus artefatos avaliadores; verificar possíveis características que levaram ao sucesso de projetos; e buscar fatores que engajaram estudantes a continuar contribuindo após o fim do evento. Notou-se uma boa aderência ao público-alvo esperado, contudo alguns aspectos precisam ser melhorados para próximas edições, como as entregas semanais, o período da realização do evento e a divisão de tempo de mentor entre vários projetos, sendo esta uma característica que impactou na conclusão de projetos e no desejo de continuar contribuindo pelos aprendizes.

PALAVRAS-CHAVE

Código Aberto; Mentoria; Evento; Andromedev.

1 INTRODUÇÃO

A OpenDevUFCG [5] é uma organização estudantil local da cidade de Campina Grande que tem como principal objetivo a inclusão de estudantes do curso de Ciência da Computação da Universidade Federal de Campina Grande (UFCG) na cultura do desenvolvimento de código aberto. A equipe da OpenDevUFCG coordena projetos que auxiliam os estudantes da UFCG em problemas recorrentes vivenciados por eles em sua realidade local, bem como motiva esses estudantes a contribuírem com projetos de código aberto. Além disso, realiza apresentações, *workshops* e eventos com o intuito de promover a cultura do código aberto e atrair novos contribuidores.

Dentre as iniciativas já realizadas pela OpenDevUFCG, destaca-se o Andromedev [7], evento criado em 2020 que teve seu funcionamento inspirado em programas de mentoria consolidados, como Google Summer of Code (GSoC) [3] e Outreachy [6]. Programas de mentoria em desenvolvimento de código aberto buscam criar um ambiente em que o mentor auxilia o aprendiz no uso das

tecnologias e códigos envolvidos no projeto, bem como introduz o aprendiz à comunidade. Em um estudo realizado com estudantes que participaram do GSoC [11], foi sugerido que a principal motivação para a participação desses estudantes no evento foi a possibilidade de alavancarem a carreira, além de ter sido mostrada uma forte intenção dos estudantes de continuar a contribuir com os projetos. Além disso, evidências apontam que o GSoC criou fortes laços entre mentores e aprendizes, tendo uma conversão de até 18% de aprendizes em mentores em edições futuras [13]. Tendo como base o sucesso e os benefícios desses programas, a equipe da OpenDevUFCG adaptou o evento para a sua realidade local.

O Andromedev teve o objetivo de estimular a participação de estudantes em projetos de *software* e criar naturalmente novas comunidades que tenham a cultura de código aberto. O evento conseguiu, em sua primeira edição, a participação de 21 organizações, com 50 projetos de código aberto. Através do sistema de mentoria, o Andromedev movimentou diversas organizações no desenvolvimento de seus projetos, tanto na comunidade acadêmica da UFCG (com a participação de professores e grupos estudantis), quanto nas organizações externas a ela.

Todo o processo de desenvolvimento foi acompanhado pela OpenDevUFCG através do preenchimento de formulários realizado pelos mentores e aprendizes participantes do evento. A coleta desses dados teve o intuito de identificar os resultados obtidos nos projetos, bem como levantar fatores que impactaram o desenvolvimento deles no evento. Aliado a isso, os dados poderiam, também, ser utilizados para a validação e adequação das regras e normas estabelecidas no regulamento. Contudo, transformá-los em informação útil à organização não se mostrou uma tarefa trivial, uma vez que o volume de dados gerados foi muito alto e, eventualmente, conteve registros acometidos por erro humano.

Ao final da primeira edição do Andromedev, surgiram várias questões sobre a realização do evento, tais como: “Quais fatores estimulam estudantes a continuar contribuindo para os projetos, mesmo ao fim do evento?”, “Qual é o perfil dos inscritos e aprendizes alcançados no evento?”, “Existe alguma característica nos projetos que aponte para a não conclusão deles no período do evento?”. Para responder a essas perguntas, este trabalho realizou, com o auxílio de um ferramental de *software*, uma análise quantitativa dos dados gerados na primeira edição do evento, visando entender aspectos relacionados com a organização do evento e o sistema de mentoria adotado. Os resultados dessa análise poderão gerar novos parâmetros que auxiliem na preparação das próximas edições, como também verificar o impacto da cultura de desenvolvimento de código aberto sobre os participantes do evento.

*Os autores retêm os direitos, ao abrigo de uma licença Creative Commons Atribuição CC BY, sobre todo o conteúdo deste artigo (incluindo todos os elementos que possam estar contidos, tais como figuras, desenhos, tabelas), bem como sobre todos os materiais produzidos pelos autores que estejam relacionados ao trabalho relatado e que estejam referenciados no artigo (tais como códigos-fonte e bases de dados). Essa licença permite que outros distribuam, adaptem e evoluam seu trabalho, mesmo comercialmente, desde que os autores sejam creditados pela criação original.

Ao longo deste trabalho, trazemos uma visão geral sobre o Andromedev na seção 2, explanando sobre os principais papéis envolvidos e como ocorreu o desenvolvimento do evento. A seguir, trazemos na seção 3 a descrição do processo metodológico utilizado. A seção 4 traz o conjunto dos resultados obtidos e, posteriormente, reunimos as discussões que eles geram na seção 5. Por fim, reunimos na seção 6 as limitações encontradas no desenvolvimento deste trabalho, bem como elencamos algumas sugestões de pesquisas futuras.

2 CONHECENDO O EVENTO

Esta seção traz uma descrição detalhada sobre o Andromedev e contém informações obtidas através do regulamento do evento [1], da sua página principal [7] e de informações fornecidas pela OpenDevUFMG.

Em sua primeira edição, o Andromedev funcionou de forma totalmente remota, desde o processo seletivo até a entrega dos resultados finais. Para entender melhor a estrutura do evento, serão explanados a seguir os principais papéis envolvidos (mentor e aprendiz), bem como a definição de um projeto no contexto do evento.

Dentre as principais definições de um projeto, destaca-se a característica de ser algum produto de *software* que tem um escopo pré-estabelecido, que realiza contribuição não trivial em um repositório e adota alguma licença de código aberto reconhecida pela Open Source Initiative [4]. O papel de mentoria é desempenhado pelos responsáveis por acompanhar o projeto do ponto de vista gerencial. Cabe à pessoa que exerceu esse papel definir quais rumos o projeto iria tomar e qual seria o produto final. As principais atividades de mentoria objetivavam: acompanhar um aprendiz no desenvolvimento do projeto; desenvolver reuniões de acompanhamento para esclarecimento de dúvidas; e auxiliar o aprendiz a se familiarizar com a comunidade do projeto. Cada projeto no Andromedev foi acompanhado por pelo menos dois mentores, um principal e um auxiliar. Não existiam limitações quanto à atuação de um mentor em apenas um projeto, então era possível que um mentor participasse de múltiplos projetos.

O papel de aprendiz foi desempenhado por estudantes de Ciência da Computação da UFMG que tinham interesse em participar de projetos de código aberto. Os aprendizes deveriam realizar o desenvolvimento das atividades relacionadas aos projetos. Buscou-se criar um evento que tivesse, como público-alvo, estudantes iniciantes no curso ou que participassem de grupos sub-representados, para proporcioná-los a oportunidade de participar de projetos de desenvolvimento de *software*. Contudo, não foram estabelecidas regras restritivas para alunos de outros grupos, sendo apenas aconselhada às organizações participantes a preferência por aprendizes do público-alvo esperado.

Devido às restrições impostas pela pandemia do SARS-CoV-2, a UFMG acabou adotando uma série de medidas para a realização do ensino remoto, conhecido como Regime Acadêmico Extraordinário (RAE). Dessa forma, o evento acabou sendo realizado em paralelo com as aulas em RAE e os estudantes tiveram que conciliar as horas das disciplinas com uma carga horária pré-estabelecida para o projeto no evento.

Durante a execução do Andromedev, foram realizados três tipos de entregas: a entrega (ou *feedback*) semanal, a entrega parcial e a

entrega final. As três entregas foram feitas via formulário e contavam com regimes de preenchimento específicos. A entrega semanal consistia em respostas a uma lista de perguntas que objetivavam identificar como foram as interações entre mentores e aprendizes naquele escopo de tempo. Já a entrega parcial tinha o objetivo de avaliar a execução do projeto e seus principais produtos. Esta foi realizada duas vezes e em semanas diferentes, para verificar a evolução durante o evento. Por fim, a entrega final tinha o objetivo de sintetizar as outras duas entregas discutidas acima, validar a execução do projeto e trazer uma projeção sobre o futuro dele após o término do Andromedev.

3 METODOLOGIA

Este trabalho segue uma abordagem quantitativa com objetivo descritivo, usando dados coletados durante o evento Andromedev em 2020. Os dados, cedidos pela OpenDevUFMG, foram analisados com o auxílio de ferramentas de *software* para responder às questões de pesquisa. Os artefatos gerados para as análises realizadas neste trabalho podem ser encontradas em um repositório no GitHub [8]. Nesta seção, serão detalhados os dados, as questões de pesquisa e os procedimentos utilizados neste trabalho.

3.1 Objetivos e Questões de Pesquisa

Este trabalho tem como objetivo avaliar a primeira edição do *Andromedev* através de uma análise quantitativa dos dados gerados durante o evento, considerando aspectos relacionados com a sua organização e o sistema de mentoria adotado. Buscando atingir esse objetivo, foram especificadas as seguintes questões de pesquisa:

- Q1: O público-alvo do Andromedev foi atingido como esperado?
- Q2: Quais artifícios de avaliação utilizados no evento tiveram resultados mais significativos?
- Q3: Quais fatores levaram à conclusão dos projetos dentro do prazo do evento?
- Q4: Quais fatores levaram estudantes de computação da UFMG a participarem e continuarem como contribuidores após a participação no evento?

3.2 Dados

Ao total, foram utilizados dados recolhidos de dez formulários, disponibilizados no formato CSV pela OpenDevUFMG. Os formulários reúnem dados sobre três visões do evento: as entregas obrigatórias, preenchida por mentores e aprendizes; o processo seletivo, que incluem as inscrições de mentores, aprendizes e organizações; e a pesquisa de interesse no evento, desenvolvida para o público que seguia as redes sociais da organização. Todos esses formulários foram realizados via Google Forms ¹ pela OpenDevUFMG durante a execução da primeira edição do Andromedev, e podem ser acessados publicamente [2]. Além disso, a organização também disponibilizou um arquivo CSV extra (chamado “*Overview*”) que contém o mapeamento dos aprendizes e mentores em seus respectivos projetos. A listagem dos arquivos disponibilizados, origens e quantidade de registros pode ser encontrada na Tabela 1.

¹<https://www.google.com/intl/pt-BR/forms/about/>

Tabela 1: Relação de arquivos cedidos para análise

Nome do Arquivo	Origem dos dados	Total de Registros	Total de Colunas
EF_Aprendiz	Entrega final dos aprendizes	41	23
EF_Mentoria	Entrega final dos mentores	34	19
EP1	Primeira entrega parcial	67	9
EP2	Segunda entrega parcial	54	10
FS_Aprendiz	<i>Feedbacks</i> semanais de aprendizes	218	10
FS_Mentoria	<i>Feedbacks</i> semanais de mentores	196	10
Inscritos	Inscrição de aprendizes	145	16
Interessados Previos	Pesquisa de opinião prévia ao Andromedev	107	17
Mentoria	Inscrição de mentores e mentores auxiliares	57	5
Organizacoes	Inscrição de organizações	23	8
Overview	Gerado pela OpenDevUF CG	50	5

O conjunto dos dados utilizados para essa análise não pode ser disponibilizado publicamente por conter informações pessoais e sensíveis dos participantes; todavia, todas as manipulações realizadas via codificação podem ser reproduzidas com a utilização dos artifícios gerados neste trabalho [8].

3.3 Procedimentos

Para a realização das análises objetivadas neste trabalho, foi necessária a adoção de um processo envolvendo três etapas: seleção de dados relevantes, pré-processamento e processamento.

3.3.1 Seleção de dados relevantes. Dada a grande quantidade de dados disponibilizados, foi necessário realizar um levantamento sobre as perguntas presentes em cada formulário, bem como as possíveis respostas para cada uma delas. Para isso, criou-se um dicionário (disponível no repositório com o nome “Dicionario”) que ajudou, posteriormente, a filtrar quais campos são essenciais para responder às questões de pesquisa. Dentre esses campos, fez-se necessário filtrar aqueles que identificavam os registros, a exemplo do *e-mail* na inscrição de aprendizes. Além disso, foram utilizadas diversas colunas da entrega final devido a possibilidade de levantar características relevantes sobre os projetos.

3.3.2 Pré-processamento. Em seguida, foi utilizada uma fase de tratamento dos dados, visando eliminar ruídos, padronizar entradas e preencher dados faltantes respeitando a semântica dos dados existentes. Nessa fase, os dados foram manipulados utilizando a linguagem Python² com o auxílio de bibliotecas de análise de dados

²<https://www.python.org/>

³. A escolha desse ambiente foi pautada na robustez e popularidade que acompanham essas ferramentas, principalmente no contexto científico e de análise de dados [10]. O *script* que realizou esse tratamento foi disponibilizado em um relatório na ferramenta Google Colab⁴, que permite a execução de um ambiente colaborativo *online* Python. Também foi mantida uma cópia desse relatório no repositório git desta pesquisa [8].

Os tratamentos aplicados nesse *script* ajudaram a:

- Gerar os registros faltantes para os casos em que mentores ou aprendizes não preencheram as entregas;
- Remover respostas duplicadas;
- Remover inconsistências devido ao mau preenchimento dos formulários (exemplo: aprendiz marcar projeto incorreto);
- Gerar conjunto de dados auxiliares obtido através da manipulação das tabelas já existentes, visando minimizar a realização de agrupamentos e outras filtragens na etapa posterior.

3.3.3 Processamento. Nessa fase, os dados pré-processados foram utilizados em análise descritiva, visando encontrar as principais métricas e estatísticas deles. Foram utilizadas métricas conhecidas da estatística, tais como frequências (absolutas e relativas), médias, medianas e desvios padrões. Além das ferramentas já utilizadas na fase de pré-processamento, também foi utilizada a biblioteca *matplotlib*⁵ para a construção dos gráficos.

Além disso, para responder às questões de pesquisa, fez-se necessário verificar quais fatores podem ter impactado na conclusão das metas estabelecidas para os projetos, e quais podem estimular os aprendizes a continuarem contribuindo após o término do evento. Para isso, foi utilizado o teste exato de Fisher [9], que visa a avaliação de associação entre classes de variáveis categóricas independentes [14].

O teste pode ser utilizado como uma alternativa ao teste Qui Quadrado por não apresentar limitações no uso em amostras pequenas de dados. Para utilizá-lo, firma-se duas hipóteses: a hipótese nula (H0), representando a inexistência de associação entre as variáveis; e a hipótese alternativa (H1), representando a existência de associação. Em seguida, é estabelecido um nível de significância α , o qual representa a probabilidade de rejeitar a hipótese nula, dada que ela é verdadeira. Por fim, é construída uma tabela de contingência entre as variáveis utilizadas, aplicado os cálculos probabilísticos sobre os valores da tabela e gerado um p-valor. Este deve ser comparado com o valor α e se for identificado um valor inferior ou igual, rejeitamos a hipótese nula. No caso contrário, aceitamos a hipótese alternativa e assumimos que existe uma relação de associação entre as variáveis testadas. O algoritmo para o teste exato de Fisher pode ser encontrado na comunidade Python, mas optou-se por uma versão do código desenvolvida em pacotes da linguagem R devido à implementação de extensões do teste que simplificam a manipulação de tabelas de contingência com tamanhos maiores que 2x2.

4 RESULTADOS

Nas subseções a seguir, serão apresentados os resultados obtidos das análises realizadas com base nas questões de pesquisa levantadas.

³<https://pandas.pydata.org/>

⁴<https://colab.research.google.com/>

⁵<https://matplotlib.org/>

4.1 Público-Alvo

4.1.1 Motivação dos interessados a aprender. Na procura de verificar as motivações do público-alvo para participar do evento como aprendiz, utilizamos dados obtidos da pesquisa de intenção prévio ao evento. Os mesmos podem ser encontrados no arquivo “InteressadosPrevios” e reúnem opiniões de 107 interessados em participar do evento como aprendiz. A pesquisa foi feita em consulta pública pela OpenDevUFMG antes do início do evento e levantava um formulário com várias questões que auxiliavam no desenvolvimento do regulamento. Neste formulário, foram levantadas algumas perguntas obrigatórias visando entender as motivações dos interessados em se tornar aprendiz. Os questionamentos focaram nas seguintes motivações: brindes; certificação de horas complementares; melhorar o currículo; no futuro, integrar o time do OpenDevUFMG; no futuro, ter mais oportunidades em laboratórios; e no futuro, ter prioridade em eventos do OpenDevUFMG. Para cada motivação, o entrevistado respondia com apenas uma das três opções possíveis: “considero desejável”, “considero essencial” ou “sou indiferente”. A Figura 1 condensa as respostas para essa pergunta.

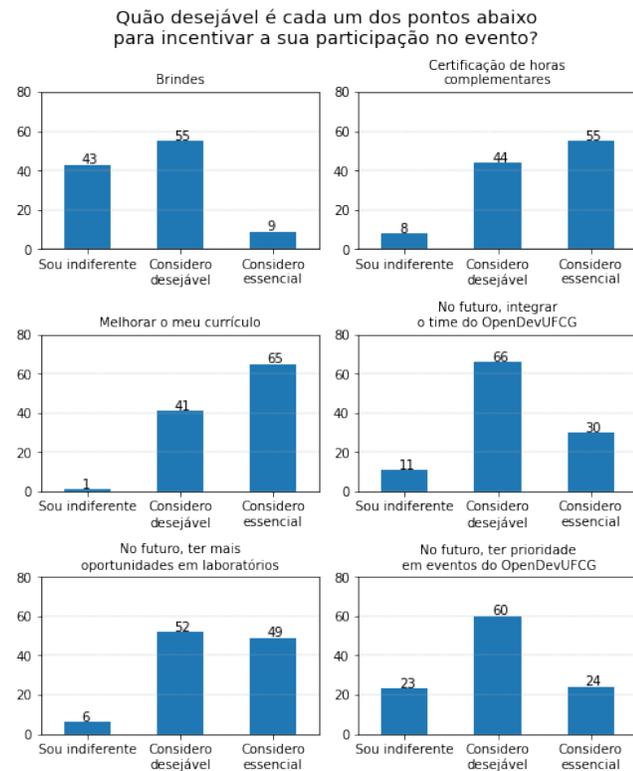


Figura 1: Frequência absoluta para motivação de participação

É possível verificar que “no futuro, ter mais oportunidades em laboratórios”, “certificação de horas complementares” e “melhorar o meu currículo” foram as maiores motivações dos possíveis inscritos a aprendiz no Andromedev. Em especial, a melhoria do currículo aparenta ser o fator mais importante dentre as opções listada por

apresentar o maior número absoluto para a opção “considero essencial”, com 65 ($\approx 61\%$) respostas, e ter o menor número absoluto para a resposta “sou indiferente”, com 1 ($\approx 1\%$) resposta. Em contrapartida, é possível verificar uma baixa motivação baseada em brindes, sendo registradas 9 ($\approx 8\%$) respostas para a categoria “considero essencial” e 43 ($\approx 40\%$) para “sou indiferente”.

4.1.2 Grupos sub-representados. Além da motivação, buscou-se analisar os 145 inscrições de aprendizes (formulário disponibilizado no arquivo “Inscritos”) para identificar qual foi o perfil dos inscritos, e, destes perfis, quais foram mais aprovados no Andromedev. Tendo como referência a intenção de trazer aprendizes que estão em grupos sub-representados ao evento, analisamos as inscrições e avaliamos as respostas para a pergunta mandatória: “Você se identifica com algum(ns) grupo(s) sub-representado(s) na sociedade e/ou na área de tecnologia?”. Além disso, filtramos as respostas dos inscritos aprovados para produzir um comparativo entre eles e o total de inscritos de cada grupo. A Figura 2 traz essa comparação, mostrando a frequência absoluta para as categorias levantadas nessa pergunta.

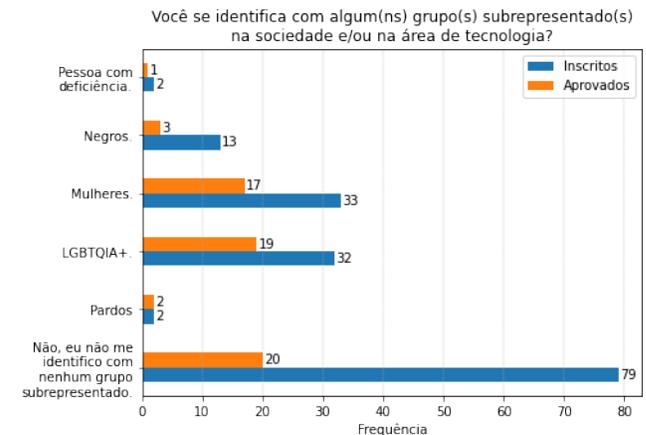


Figura 2: Frequência absoluta para pertencimento em grupos sub-representados

Entre os grupos sub-representados, temos o grupo de mulheres, com 33 inscritas, representado a classe mais frequente, seguida de LGBTQIA+ com 32 inscritos, negros com 13, pardos com 2 inscritos e pessoa com deficiência com 2 inscritos. Contudo, verificando a relação de pessoas que marcaram não se identificar com grupos sub-representados, nota-se uma predominância dessa classe, representando o valor máximo, entre as classes, de 79 inscritos. Levando em consideração que houve 150 inscritos no evento, podemos afirmar que os 66 inscritos (44%) restantes se identificam com pelo menos um grupo sub-representado. Analogamente, temos que dos 50 aprovados, 30 (60%) representam participantes de grupos sub-representados. De maneira geral, os valores relativos mostram um aumento no percentual de grupos sub-representados em comparativo com grupos não sub-representados.

Visando entender melhor esse aumento percentual de grupos sub-representados comentado acima, pudemos analisar essas categorias mais a fundo. Ao compararmos valores de inscritos e de

aprovados em cada uma das subcategorias, temos que nas classes de pessoas com deficiência, mulheres, LGBTQIA+ e pardos houve uma conversão de pelo menos $\approx 50\%$ de inscritos em aprendizes. Contudo, o grupo de pessoas que se identificam como negras, representado por 13 inscritos, apresentou um valor baixo de aprovação, tendo uma taxa de conversão de apenas $\approx 23\%$. Além disso, mesmo comparado com o grupo de pessoas não pertencentes a grupos sub-representados (que obteve $\approx 25\%$ de conversão), temos menos conversão de inscritos em aprendizes na classe de negros.

4.1.3 Do período de ingresso. Outra métrica a ser analisada é o perfil de ingresso na graduação em Ciência da Computação pela UFCG. Analogamente ao processo desenvolvido na seção 4.1.2, extraímos do formulário de inscrição de aprendizes (disponível no arquivo “Inscritos”), as respostas para a pergunta mandatória que registrava o período do inscrito. Em seguida, filtramos aqueles inscritos que foram aprovados no evento, para analisar a distribuição de inscritos e aprovados. A Figura 3 mostra a frequência das respostas em escala logarítmica para os períodos encontrados nas respostas.

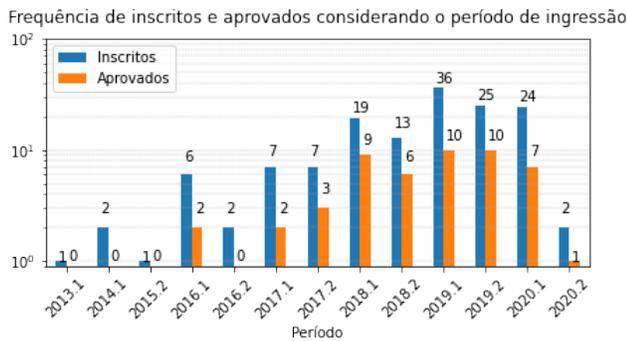


Figura 3: Frequência absoluta em escala logarítmica para período de ingresso em curso

Tomando como base o perfil esperado (estudantes do início do curso que ainda não tiveram oportunidade em projetos de desenvolvimento), é possível verificar uma predominância entre os períodos iniciais do curso. Ao analisarmos os valores para inscritos, encontramos que os três períodos com mais inscritos foram o “2019.1”, “2019.2” e o “2020.1”. É importante verificar que esses três períodos estão entre os quatro mais recentes até o momento do evento (Setembro de 2020), estando apenas o “2020.2” com valores baixos de frequência.

Para as estatísticas de inscritos temos média de ≈ 11 , desvio padrão de ≈ 11 e mediana de 7. A mediana acaba nos indicando que pelo menos 50% das frequências estão abaixo dela, mostrando certo desbalanceamento entre as frequências dos períodos. Contudo, como já discutido previamente, os períodos mais frequentes se concentram no início do curso, sendo justamente o público-alvo esperado a ser atingido. Nessa faixa de público, a maior parte dos estudantes estão iniciando o contato com o desenvolvimento de *software* em equipe e a experiência do evento pode ser um diferencial no currículo, alinhado com a motivação dos inscritos discutida na seção 4.1.1.

Para as estatísticas de aprendizes temos média de ≈ 4 , desvio padrão de ≈ 4 e mediana de 2. Nesse caso, a mediana também nos alerta para um possível desbalanceamento entre as classes, visto que pelo menos 50% dos períodos teve 2 ou menos aprovados. Ainda podemos observar que três períodos não apresentaram inscritos aprovados: “2013.1”, “2015.2” e “2016.2”. Apesar disso, o desvio padrão nos mostra uma maior uniformidade entre as classes se comparado ao valor encontrado para inscritos.

4.2 Artefatos

Visando examinar os artefatos de avaliação dos projetos utilizados no evento, reunimos os dados das entregas realizadas objetivando extrair métricas a partir delas. Os formulários que representam essas entregas foram disponibilizados publicamente [2] e a Tabela 1 traz a listagem dos nomes dos arquivos em que podem ser encontrados cada um deles.

Para compreendê-los, tornou-se necessário identificar a quantidade de projetos, aprendizes e mentores que participaram do evento, e assim descobrir a quantidade de registros esperados para cada entrega. Dessa forma, utilizamos o arquivo “Overview”, que traz a relação de projetos aprovados com seus respectivos mentores e aprendizes. Ao total, foram realizados 50 projetos e, com base nisso, construímos na Tabela 2 a relação das entregas obrigatórias com suas frequências esperadas e encontradas, de acordo com o regulamento do evento [1].

Tabela 2: Informações sobre as entregas obrigatórias do evento.

Formulário	Preenchido por	Freq. Esperada	Freq. Encontrada
Feedback Semanal	Aprendiz	400	196 (49%)
	Mentor	400	188 (47%)
Entrega Final	Aprendiz	50	41 (82%)
	Mentor	50	33 (66%)
Primeira E. Parcial	Aprendiz ou Mentor	50	45 (90%)
Segunda E. Parcial	Aprendiz ou Mentor	50	44 (88%)

De forma geral, observamos que as entregas finais e *feedbacks* semanais tiveram alguns valores preocupantes. Para os casos de entregas realizadas por mentores, o *feedback* semanal teve uma adesão de apenas 47% dos registros esperados, menos da metade do esperado para o evento. Para os aprendizes, o *feedback* semanal também obteve sua menor adesão dentre as entregas, apresentando 49% dos registros esperados. As entregas parciais, por sua vez, apresentam altas porcentagens de respostas, 90% e 88% respectivamente, demonstrando uma grande aceitação dessa entrega.

4.2.1 Entregas semanais. Visando entender os motivos para a baixa aceitação dos *feedbacks* (entregas) semanais, analisamos mais a fundo essa entrega. Estes acompanharam os projetos durante todo o evento e são neles que aprendizes e mentores registram semanalmente o andamento dos seus progressos em relação aos seus

respectivos projetos. O formulário de submissão se mantinha o mesmo para todas as semanas e existia um campo que marcava a semana ao qual era relativo aquele registro. Utilizando as respostas para esse formulário, conseguimos extrair a relação de aprendizes e mentores que realizaram o seu preenchimento em função das semanas de submissão. A Figura 4 mostra o resultado dessa análise para as oito semanas de evento.

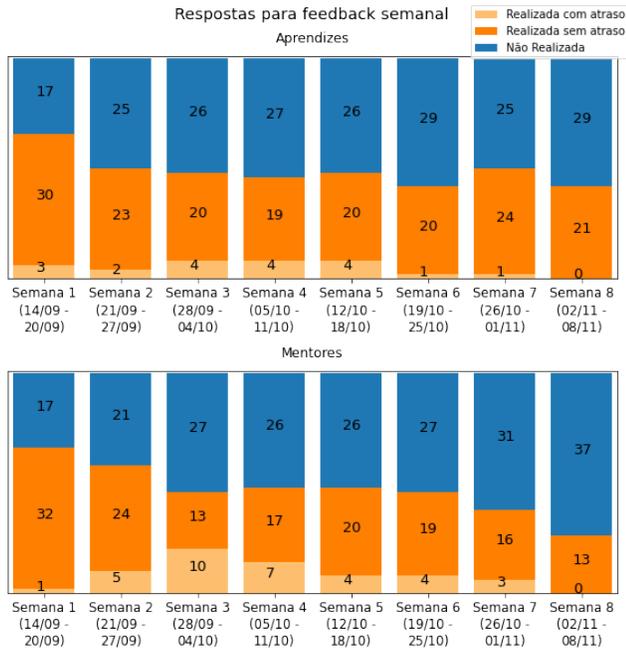


Figura 4: Frequência absoluta para respostas ao feedback semanal, subdivididas entre aprendizes e mentores

É possível identificar uma queda na quantidade de respostas à medida que o evento foi avançando. Passada a primeira semana, pelo menos 50% dos aprendizes não realizaram a submissão da resposta para o acompanhamento semanal, demonstrando uma baixa adesão em relação a esta entrega. Os registros para as semanas seguintes oscilam entre 21 a 24 respostas registradas. O mesmo comportamento de queda é notado no gráfico para mentores, contudo tendo sua queda a partir da segunda semana de submissão, apresentando valores que oscilaram de 13 a 24 registros.

Além do não preenchimento, ainda existe um percentual de respostas que foram preenchidas com atraso. Isso acabou comprometendo a interpretação das respostas para essa entrega, pois a submissão tardia acaba por trazer incerteza às respostas, uma vez que o aprendiz ou mentor pode não lembrar com clareza todos os aspectos levantados neste formulário. Quando observamos os atrasos de mentores, eles acabam se concentrando na semana 3 e 4, com 10 e 7 registros, respectivamente.

Tentando compreender o motivo dessas faltas para os mentores, podemos avaliar respostas encontradas na entrega final de mentores, formulário identificado no arquivo “EF_Final” [2]. Nela, destacamos a pergunta que tratava sobre possíveis melhorias que os mentores acreditam ser importantes para as próximas edições. O

campo era opcional e registrava respostas abertas para essas sugestões. Nele, encontramos um registro que aponta para dificuldades no preenchimento dos *feedbacks* semanais. Além disso, também foi realizada uma sugestão de melhoria que atuasse na notificação sobre o prazo dessas entregas. Contudo, estes foram registros isolados e necessitam de estudos qualitativos mais específicos para serem representativos.

Posto isso, para próximas edições, é necessário desenvolver estratégias que atuem nessas entregas e estimulem um maior número de respostas. Tais estratégias podem atuar na facilitação no preenchimento desses formulários e na utilização de notificação para que não haja atrasos nas submissões.

4.2.2 Entregas Parciais. As entregas parciais, disponíveis nos formulários “EP1” e “EP2” [2], representam o artefato mais significativo em porcentagem de respostas. Essas entregas correspondem aos formulários realizados durante a execução do evento sendo feita apenas uma submissão por projeto. Nas duas entregas parciais, foram registradas, respectivamente, 45 e 44 submissões, representando um total de 88% e 90% de participação. Não é possível analisar atrasos para essa entrega, uma vez que não foi permitido realizar nenhuma submissão após o prazo de submissão.

Segundo informações fornecidas pela OpenDevUFCG, diferentemente do *feedback* semanal, essas entregas eram acompanhadas de um *e-mail* enviado pela organização do evento, que levantava instruções de preenchimentos e prazos. Além disso, essas entregas foram implementadas em formulários separadamente, ou seja, cada uma das submissões era acompanhada por formulários diferentes. Isso pode ter motivado a alta porcentagem de respostas para essa entrega.

4.2.3 Das Entregas Finais. Por fim, foram coletadas informações sobre o projeto ao final do evento, como sendo uma última entrega. Essa entrega foi enviada por aprendizes e mentores separadamente e reuniu, respectivamente, 41 (82%) e 33 (66%) registros. Os formulários para as entregas finais de aprendizes e mentores estão disponibilizados, respectivamente, nos arquivos “EF_Aprendiz” e “EF_Mentoria”. Assim como nas entregas parciais, foram enviadas as instruções para preenchimento e prazos de envio, os quais deveriam ser respeitados, por não haver a possibilidade de envio atrasado.

Observando as informações obtidas nas três categorias de entregas, é possível extrair uma série de informações sobre o desenrolar do projeto que se encontram em respostas abertas em formato textual, sendo uma rica fonte de informação para entender melhor os motivos dos atrasos e faltas; contudo, foge ao escopo deste trabalho analisar esses dados.

4.3 Associação de características com conclusão de projeto

Esta subseção analisa os fatores que levaram à conclusão dos projetos dentro do prazo do evento. É importante ressaltar que, por conclusão, nós nos referimos ao cumprimento das atividades esperadas para os projetos durante o Andromedev, e não à participação do projeto até o fim do evento.

Nesse cenário, utilizou-se a entrega final de mentores e aprendizes, que traz em uma de suas perguntas obrigatórias a visão pessoal do aprendiz/mentor sobre a conclusão do projeto, se ele considera

o projeto concluído ou não. As respostas para essa pergunta foram cruzadas entre mentores e aprendizes para verificar divergências entre as duas partes. Foi notado que apenas dois projetos tiveram opiniões divergentes quanto a essa questão, demonstrando um bom alinhamento de boa parte dos projetos em relação às respostas de mentores e aprendizes. Dessa forma, foi possível filtrar nas respostas dos aprendizes aqueles que pertenciam a esse grupo convergente, ou aqueles em que não houve resposta do mentor sobre a questão. Ao total, foram 39 projetos filtrados, dos quais 28 projetos houve convergência de opinião, e 11 projetos em que apenas o aprendiz submeteu resposta. Casos em que só o mentor submeteu não foram encontrados. Ainda, 21 projetos não foram concluídos em sua totalidade, enquanto os demais 18 cumpriram as metas esperadas.

Em seguida, realizou-se o levantamento de características dos projetos para verificar a existência de associações entre elas e a não conclusão do projeto no período do evento. Para capturá-las, foram utilizados dados recolhidos da entrega final dos aprendizes, tendo em vista que ela elenca uma série de questionamentos que levantam fatores importantes dos projetos. Dentre os dados disponíveis, foram escolhidas cinco características e essa escolha teve como base a seleção de perguntas que fossem obrigatórias, tivessem em seu formulário um conjunto limitado de classes de respostas (até quatro classes diferentes) e que só fosse possível marcar uma resposta. Essas medidas foram tomadas para uniformizar os tipos de variáveis manipuladas e facilitar a aplicação dos testes de associação, tratados posteriormente. As características foram retratadas pelas respostas das seguintes perguntas:

- Quem realizou a maior parte do acompanhamento no projeto?
- Você participou de algum projeto de pesquisa ou desenvolvimento em paralelo no período do Andromedev?
- Em relação ao RAE, você acredita que...
- Em relação à dedicação esperada de minha mentora/ meu mentor...
- O seu mentor(a) dividiu o tempo de mentoria com outros aprendizes (interno ou externo ao evento)?

Para verificar a associação entre essas características dos projetos com o fato do projeto ter sido concluído com sucesso, foi utilizada a abordagem por testes de associação não paramétricos para variáveis categóricas. O objetivo desses testes é, através da análise sobre a frequência na qual são distribuídas as respostas para cada uma das variáveis, avaliar se existe associação entre as variáveis testadas. Para amostras pequenas de dados, utiliza-se frequentemente o teste exato de Fisher. Os passos para a utilização são comentados a seguir e podem ser encontrados no *script* de nome “Processamento”, disponível no repositório deste trabalho [8].

Para cada uma das cinco características levantadas, foram estabelecidas as seguintes hipóteses:

- H0: não há associação entre a característica e a conclusão do projeto;
- H1: há associação entre a característica e a conclusão do projeto.

Posteriormente, construíram-se as tabelas de contingência para cada uma das características analisadas, sempre com relação ao fato que indicava a conclusão do projeto. Com as tabelas prontas, aplicamos os cálculos para descobrir os p-valores de cada característica

e compará-los a um nível de significância (α) de 5% (valor padrão). Para os casos em que p-valor foi menor ou igual que o α , rejeitamos a H0 e aceitamos H1, caso contrário, aceitamos a H0 e rejeitamos H1. A Tabela 3 traz os p-valores e as hipóteses aceitas para cada uma das características.

Tabela 3: Resultados de testes obtidos para a relação entre as características e a conclusão do projeto dentro do evento.

Característica	p-valor	Hipótese Aceita ($\alpha = 5\%$)
C1	$\approx 0,02$	H1
C2	$\approx 0,70$	H0
C3	$\approx 0,01$	H1
C4	$\approx 0,75$	H0
C5	$\approx 0,05$	H1

A interpretação dos testes nos indica associação com três características: C1, que representa quem realizou o acompanhamento do projeto (mentor principal, mentor auxiliar, os dois juntos); C3, que mede os impactos do RAE sobre as entregas do aprendiz; e C5, que indica se o mentor auxiliou outros aprendizes em paralelo. As características que não apresentaram associação significativa foram a participação de aprendizes em projetos de desenvolvimento paralelo (C2) e a visão do aprendiz sobre a dedicação esperada da mentoria (C4). As subseções a seguir trazem as análises aprofundadas para os resultados do p-valor encontrados nos casos em que se notou associação entre as classes.

4.3.1 Acompanhamento do projeto (C1). Dada a presença de associação, analisamos a tabela de contingência, representada na Tabela 4. O acompanhamento realizado só pelo mentor auxiliar somou seis projetos e acabou sendo mais associado a projetos que não cumpriram o prazo adequadamente, uma vez que não houve registros para casos em que foi concluído. Além disso, o conjunto de projetos mentoreados apenas pelo mentor principal, somando 32 projetos, tendeu a ser ligeiramente maior para a conclusão no prazo do evento, com 17 ($\approx 53\%$) projetos. Para os projetos acompanhados por mentor principal e auxiliar tivemos o total de um caso e ele foi concluído no prazo do evento. Dessa forma, pode-se notar que projetos que concluíram no prazo estão mais associados a projetos mentoreados pelo mentor principal.

Tabela 4: Tabela de contingência para conclusão do projeto e quem acompanhou o projeto.

		Conclusão de Projeto		Total
		Não	Sim	
C1	Mentor Principal	15	17	32
	Mentor Auxiliar	6	0	6
	Mentor Principal e o Mentor Auxiliar	0	1	1
	Total	21	18	39

4.3.2 *Regime Acadêmico Extraordinário (C3)*. Podemos verificar a associação apresentada entre a percepção dos impactos no rendimento devido ao RAE e a conclusão dos projetos na tabela de contingência (Tabela 5). Verifica-se que aqueles aprendizes que sentiram pouca ou nenhuma influência do RAE, representados pelas classes “Afetou pouco meu rendimento no projeto” (9 respostas) e “Não afetou meu rendimento” (4 respostas), geralmente estavam associados a projetos que concluíram no período correto.

Tabela 5: Tabela de contingência para conclusão do projeto e impacto do RAE.

		Conclusão de projeto.		Total
		Não	Sim	
C3	Afetou bastante meu rendimento no projeto e não consegui conciliar os dois.	3	0	3
	Afetou meu rendimento no projeto, mas consegui conciliar os dois.	15	8	23
	Afetou pouco meu rendimento no projeto.	3	6	9
	Não afetou meu rendimento no projeto.	0	4	4
Total		21	18	39

Analogamente, os projetos em que o aprendiz sentiu mais influência do RAE, representados pelas respostas restantes, são aqueles que estão mais associados à não conclusão do projeto no tempo do evento. Para os casos em que o aprendiz acredita que não conseguiu conciliar os dois, todos os quatro projetos não foram concluídos. Já para os aprendizes que acreditam que o RAE afetou o rendimento mas conseguiram conciliar (somando 23 projetos), em 15 ($\approx 65\%$) deles o aprendiz não conseguiu concluir o projeto a tempo.

4.3.3 *Divisão de tempo de mentoria (C5)*. Alguns mentores também participaram de múltiplos projetos, desenvolvendo um acompanhamento dividido entre os seus aprendizes. Essa característica acabou apresentando uma associação com a conclusão do projeto no tempo adequado. Dessa forma, construímos a tabela de contingência (Tabela 6) para entender melhor esse resultado.

Tabela 6: Tabela de contingência para conclusão de projeto e divisão de tempo de mentoria

		Conclusão de projeto		Total
		Não	Sim	
C5	Não	7	12	19
	Sim	14	6	20
Total		21	18	39

A classe de projetos que tiveram o mesmo mentor em múltiplos projetos, representado por 20 projetos, consistiram em 14 (70%) não concluídos no evento. Já a classe dos projetos que não tiveram divisão de tempo na mentoria somaram 19 projetos ao total e 7 ($\approx 36\%$) não concluídos. À vista disso, temos que projetos que

tiveram mentores dividindo o tempo de mentoria com múltiplos aprendizes foram mais associados com a não finalização do projeto no período do evento.

4.4 Associação de características com participação pós-evento

Para analisar as características que influenciaram no desejo dos aprendizes de continuar a colaborar com a comunidade relativa ao projeto que participou no Andromedev, realizou-se uma abordagem semelhante àquela discutida na seção 4.3. Foram utilizadas as mesmas características utilizadas previamente, mas dessa vez buscou-se relacioná-las com a intenção do aprendiz de continuar a contribuir na comunidade. Para isso, extraímos também da entrega final do aprendiz, a seguinte pergunta mandatória: “Você irá continuar a contribuir com esse ou outro projeto dessa organização?”. A resposta para a pergunta variava em uma escala de 1 (“muito improvável”) a 5 (“muito provável”).

Para as possíveis respostas, apenas as classes “3”, “4” e “5” foram contempladas. Tendo em vista que as duas primeiras tiveram, respectivamente, 8 e 7 respostas, optou-se por unir estas classes para evitar problemas ao criar as tabelas de contingência com frequências muito pequenas, uma prática comum para esse tipo de cenário. Dessa vez, foram utilizados todos os registros dos aprendizes que submeteram a entrega final, contemplando 41 projetos divididos entre: 15 registros para a resposta “3 ou 4” e 26 registros para “5”. Mais uma vez, utilizamos o teste exato de Fisher para avaliar a associação, contudo dessa vez buscamos avaliar a associação entre as características dos projetos e a intenção do aprendiz de continuar a contribuir. Para as hipóteses do nosso teste, temos que:

- H0: não há associação entre a característica e a intenção do aprendiz de continuar a contribuir;
- H1: há associação entre a característica e a intenção do aprendiz de continuar a contribuir.

Realizamos os cálculos para o p-valor em cada uma das características e utilizamos o mesmo nível de significância $\alpha = 5\%$ utilizado previamente. Os resultados foram expressados na Tabela 7.

Tabela 7: Influência das características dos projetos e a continuidade na comunidade de código aberto.

Questão	p-valor	Hipótese Aceita ($\alpha = 5\%$)
C1	$\approx 0,07$	H0
C2	$\approx 0,44$	H0
C3	$\approx 0,27$	H0
C4	$\approx 0,56$	H0
C5	$\approx 0,05$	H1

Para C1, C2, C3 e C4 encontramos p-valores superiores ao nível α , nesses casos, aceitamos a hipótese nula e entendemos que não existe associação significativa entre essas características e a intenção do aprendiz de continuar contribuindo para a comunidade do projeto. As características representam, respectivamente: o tipo de mentor que realizou o acompanhamento, se principal ou auxiliar; a participação ou não do aprendiz em projetos de desenvolvimento paralelos ao Andromedev; a relação do aprendiz com o RAE; e a

dedicação do mentor no projeto. Para C5, temos a aceitação da hipótese alternativa, ou seja, houve associação entre a característica que indica a participação do mentor em múltiplos projetos e a intenção do aprendiz de continuar contribuindo.

Além disso, dada a proximidade do p-valor encontrada na característica 1 (p-valor $\approx 0,07$) com o nível de significância, também realizamos sua análise. As subseções a seguir trazem essas análises sobre as tabelas de contingência das duas características.

4.4.1 Acompanhamento do projeto (C1). A Tabela 8 traz a tabela de contingência que representa a relação das respostas entre o tipo de mentor que mais acompanhou o projeto e a intenção do aprendiz de continuar contribuindo. É possível observar que nos projetos mentoreados por mentores principais, 23 dos 33 aprendizes marcaram uma avaliação de “5” para a intenção de contribuir após o término do evento, o que representa $\approx 69\%$ dos projetos que pertencem a essa categoria. Já para os aprendizes que foram acompanhados pelo mentor auxiliar, 5 dos 7 marcaram a opção “3 ou 4”, representando $\approx 71\%$. A categoria de aprendizes que foram acompanhados por mentor principal e mentor auxiliar só obteve um aprendiz e ele registrou a resposta “5”.

Tabela 8: Tabela de contingência para intenção de contribuição e quem acompanhou o projeto.

		Desejo de contribuir		Total
		3 ou 4	5	
C1	Mentor Principal	10	23	33
	Mentor Auxiliar	5	2	7
	Mentor Principal e o Mentor Auxiliar	0	1	1
Total		15	26	41

Visto isso, temos que a presença do mentor principal no acompanhamento do projeto está mais associada com projetos em que o aprendiz tem mais interesse em continuar a desenvolver ao fim do evento. Já a presença de apenas um mentor auxiliar parece ter mais associação com aprendizes com respostas neutras-positivas a essa pergunta.

4.4.2 Divisão de tempo de mentoria (C5). Por fim, a tabela de contingência (Tabela 9) mostra a relação entre as respostas para a intenção de continuar a contribuir e a característica de ter um mentor auxiliando múltiplos aprendizes. Nela, é possível notar que os projetos que acompanharam apenas um aprendiz tiveram 16 dos 20 aprendizes marcando a resposta “5”, ou seja, 80% das respostas. Para projetos em que o mentor acompanhou mais de um aprendiz, temos uma distribuição mais homogênea: 11 ($\approx 52\%$) marcaram “3 ou 4” e 10 ($\approx 47\%$) marcaram “5”.

Portanto, a associação mostra que projetos acompanhados por mentores auxiliando múltiplos aprendizes tiveram reações mais neutras-positivas (3 ou 4) quanto ao desejo de continuar contribuindo para o projeto ou organização ao fim do Andromedev. Já em projetos em que o mentor atuou com apenas um aprendiz, houve uma maior quantidade de reações positivas (5).

Tabela 9: Tabela de contingência para intenção de contribuição e divisão de tempo de mentoria.

		Desejo de contribuir		Total
		3 ou 4	5	
C5	Não	4	16	20
	Sim	11	10	21
Total		15	26	41

5 DISCUSSÃO

Na intenção de avaliar se o público-alvo do Andromedev foi atingido como esperado, foram analisadas as motivações dos inscritos a participarem do evento como aprendiz e como se deu o perfil dos aprendizes que foram selecionados. Com isso, tem-se que o público-alvo do evento foi atendido parcialmente como esperado. A motivação dos aprendizes está alinhada com o desenvolvimento de seus currículos profissionais, mostrando alinhamento com o evento. Além disso, a busca por aprendizes iniciantes, representados por estudantes do início do curso, também parece ter sido atingida, uma vez que grande parte dos inscritos se concentram nos quatro primeiros períodos da graduação. Contudo, é preciso salientar também que os valores para as frequências encontradas em cada período podem ser afetados por fatores externos que não foram mensurados no formulário de inscrição, tais como a limpeza de currículo ou truncamento de curso. Dessa forma, a análise pode ser mais aprofundada se nas próximas edições forem coletadas perguntas que contemplem esses fatores. Ademais, apesar de ter apresentado uma alta taxa para a participação de aprendizes pertencentes a grupos sub-representados, o grupo de pessoas negras ainda teve resultados insatisfatórios.

Para avaliar os artefatos utilizados no evento, analisamos a frequência a qual cada uma das entregas obrigatórias foram preenchidas, além de verificar os atrasos no preenchimento delas. Dentre as entregas, os *feedbacks* semanais representam as piores porcentagens de preenchimento. Observamos que menos da metade dos registros esperados para essa entrega foram efetuados, mesmo considerando atrasos envolvidos. Em contrapartida, as entregas que obtiveram as maiores porcentagens de preenchimento foram as entregas parciais, em especial a primeira entrega, que obteve 90% das respostas esperadas.

Buscando entender os fatores para os aprendizes não terem conseguido concluir seus projetos em sua totalidade durante o período do evento, avaliamos características obtidas dos projetos para identificar associações. Encontramos que três fatores apresentam associação significativa com a conclusão dos projetos no Andromedev: o acompanhamento por mentores principais; o menor impacto sentido pelo aprendiz em relação ao RAE; e o acompanhamento por um mentor ligado a apenas um aprendiz. Contudo, é importante salientar que esses valores podem não refletir necessariamente em resultados insatisfatórios obtidos nos projetos, uma vez que a pergunta utilizada para avaliar esse aspecto não considerava gradações possíveis para a conclusão de um projeto. Dessa forma, podem existir aqueles que não concluíram todas as metas estipuladas, mas tenham cumprido entregas satisfatórias.

Além disso, avaliamos as intenções dos aprendizes de continuarem a contribuir para os projetos após o fim do evento. Mais uma

vez, o fato do mentor principal ter acompanhado o projeto esteve mais associado com intenções positivas de contribuir. Ademais, a presença de um mentor acompanhando apenas um aprendiz esteve mais associado a projetos em que o aprendiz teve uma reação positiva para continuar contribuindo após o fim do Andromedev.

Com esses resultados podemos indicar alguns aspectos que podem ser melhorados para as próximas edições do Andromedev. Visando melhorar a participação de grupos sub-representados, é necessário desenvolver estratégias no regulamento para o balanceamento desses grupos, garantindo a representação significativa de todos eles. Ademais, tem-se que o fato do evento ter sido realizado, majoritariamente, no mesmo período do RAE, afetou negativamente nos resultados do evento. Por isso, a OpenDevUFCEG pode avaliar alternativas para flexibilizar a carga horária obrigatória, possibilitando que os estudantes possam conciliar melhor os projetos com a graduação.

Outro aspecto que deve ser considerado é quanto ao acompanhamento de mentores em múltiplos projetos. Apesar desses mentores serem associados com resultados que não se alinham com os objetivos do evento, eles representam uma parcela significativa dos projetos da primeira edição. Foram registrados pelo menos 21 deles no formulário de entrega final, representando pelo menos 42% dos projetos participantes do Andromedev. Dessa forma, é necessário criar soluções que otimizem os resultados para esses mentores, e, para isso, é necessário entender os motivos para esses projetos terem essa associação.

Por fim, também precisamos ressaltar a importância de acompanhar os projetos após o final do evento. Em um estudo de interações capturadas ao longo de cinco anos do projeto de código aberto Hadoop ⁶, mostrou-se que menos de 20% dos recém-chegados tornaram-se contribuidores de longo prazo [12]. Por isso, é essencial ter um levantamento de como estão esses projetos após o fim do evento, uma vez que este trabalho focou apenas na análise sobre o desejo do aprendiz de continuar a contribuição e pode não refletir diretamente com o engajamento real.

6 LIMITAÇÕES E SUGESTÕES DE PESQUISA

Para consolidar as questões de pesquisa levantadas, é importante repetir as análises realizadas com bases de dados atualizadas. Uma vez que o trabalho foi realizado apenas com os dados da primeira edição, é natural que exista uma incerteza atrelada à quantidade de registros utilizados no trabalho. Estatisticamente, o número de projetos acaba sendo pequeno e isso traz dúvidas quanto aos aspectos que podem influenciar na conclusão do projeto, bem como o desejo de continuar a contribuir vindo do aprendiz. A utilização dos dados de edições futuras pode auxiliar a entender melhor os aspectos temporais e atemporais que esses projetos apresentam, e otimizar os resultados das próximas edições.

Também é possível utilizar os dados de campos abertos, considerados neste trabalho, como fonte mais robusta de informações, bem como os *links* para os repositórios dos projetos participantes do Andromedev para realizar raspagem de novas métricas e assim descobrir novos fatores que influenciam na participação dos aprendizes no evento. Esses repositórios trazem informações de

frequência de *commits*, requisições de alteração de código e quantidade de linhas de código alteradas, o que pode ajudar a entender melhor as atividades durante a execução do evento, mas também os resultados posteriores a ele.

7 AGRADECIMENTOS

Gostaria de agradecer à minha orientadora, Lívia Sampaio Campos, pela disponibilidade em acompanhar esta pesquisa, bem como ter me encorajado a escolher a temática de código aberto para este estudo. Também gostaria de agradecer à equipe da OpenDevUFCEG, que disponibilizou os dados para esta pesquisa e me fez acreditar que até as ideias mais impossíveis podem se tornar realidade. Por fim, agradeço àqueles que firmaram a base da minha educação: a minha família, em especial, a minha mãe, que sempre foi a minha maior impulsionadora.

REFERÊNCIAS

- [1] 2020. Regulamento do Andromedev. <https://drive.google.com/file/d/1ggJiprXrroiE8tRNwhMpXTKAOrHFVAKQ/view> [Online; acessado em 11 de Maio de 2021].
- [2] 2021. Formulários do Andromedev. <https://drive.google.com/drive/u/0/folders/1QVC1BCzvgSoUu3ue3Q41FTz1Z17ppY9a> [Online; acessado em 12 de Maio de 2021].
- [3] 2021. Google Summer of Code. <https://summerofcode.withgoogle.com/> [Online; acessado em 29 de Abril de 2021].
- [4] 2021. Open Source Initiative. <https://opensource.org/> [Online; acessado em 02 de Maio de 2021].
- [5] 2021. OpenDevUFCEG. <https://opendevufcg.org/> [Online; acessado em 29 de Abril de 2021].
- [6] 2021. Outreachy. <https://www.outreachy.org/> [Online; acessado em 29 de Abril de 2021].
- [7] 2021. Página Principal do Andromedev. <https://andromedev.opendevufcg.org/> [Online; acessado em 29 de Abril de 2021].
- [8] 2021. Repositório das análises no Github. https://github.com/JuanBarros2/tcc_andromedev/ [Online; acessado em 02 de Maio de 2021].
- [9] Ronald Aylmer Fisher. 1992. Statistical methods for research workers. In *Breakesthroughs in statistics*. Springer, 66–70.
- [10] Wes McKinney. 2019. *Python para análise de dados: Tratamento de dados com Pandas, NumPy e IPython*. Novatec Editora.
- [11] Jefferson O Silva, Igor Steinmacher, and Marco A Gerosa. 2016. Os Programas de Engajamento em Software Livre Atraem Estudantes à Colaboração Voluntária? Um Estudo Empírico do Google Summer of Code. In *Anais do XIII Simpósio Brasileiro de Sistemas Colaborativos*. SBC, 119–133.
- [12] Igor Steinmacher, Igor Wiese, Ana Paula Chaves, and Marco Aurélio Gerosa. 2013. Why do newcomers abandon open source software projects?. In *2013 6th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)*. 25–32. <https://doi.org/10.1109/CHASE.2013.6614728>
- [13] Erik H. Trainer, Chalalal Chaihirunkarn, Arun Kalyanasundaram, and James D. Herbsleb. 2014. Community Code Engagements: Summer of Code Hackathons for Community Building in Scientific Software (*GROUP '14*). Association for Computing Machinery, New York, NY, USA, 111–121. <https://doi.org/10.1145/2660398.2660420>
- [14] Jacques Wainer et al. 2007. Métodos de pesquisa quantitativa e qualitativa para a Ciência da Computação. *Atualização em informática* 1, 221-262 (2007), 32–33.

⁶<https://hadoop.apache.org/>