

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS  
Instituto de Ciências Exatas e Informática (ICEI)  
Sistemas de Informação - Processos e Qualidade de Software

Relatório de coleta e análise de dados do GitHub - Trabalho 2  
Leonardo Ávila e Marlon Dias

Link do repositório: <https://github.com/LeonardoWillian/PQSAAnalise>

Objetivo: Identificar o nível de variação entre a qualidade dos programadores que estão envolvidos no VSCode a partir de seus repositórios e participações no VSCode.

Questão: Qual a taxa de variação de qualidade, relacionando atributos de repositórios pessoais, dentre os programadores do repositório VSCode?

Métricas:

- 1 - Soma da quantidade de forks de seus repositórios;
- 2 - Soma da quantidade de stars de seus repositórios;
- 3 - Quantidade de issues que programador é responsável do VSCode ;
- 4 - Quantidade de pull requests feitos pelo programador que foram aceitos no VSCode;

Modelos de Qualidade:

Utilizando dos forks, stars e pull requests aceitos dos repositórios, iremos medir a qualidade do programador, entendendo que ele tem uma ideia interessante no repositório(star) e que seu código é bem escrito o suficiente para engajar novos trabalhos(forks).

Utilizando da quantidade de respostas nos issues, entende-se que o programador tem interesse e comprometimento pelas discussões que norteiam o VSCode, portanto, isso corrobora para seu destaque naquele meio.

Como calcular a variação?

Pretendemos utilizar desvio padrão e/ou coeficiente de variação nas somatórias das métricas. Qual fórmula utilizar ainda não está totalmente claro para nós e estamos pesquisando um pouco nas anotações da disciplina de Estatística.

Explicando melhor, quando falamos sobre a qualidade do programador no repositório do VSCode, queremos dizer principalmente do quanto ele commita no repositório. Quando um pull request é aceito, se torna um commit em nome do programador que solicitou o pull request. Nesse trabalho, coletamos a somatória dos forks e stars dos repositórios como uma métrica mais individual, e os issues e commits no VSCode como uma métrica que coloca cada um deles lado a lado.

Segue abaixo o resultado da coleta de cada uma dessas métricas:

Programador	Forks	Stars	Issues	Pull requests aceitos
Tyriar	113	623	225	2286
isidorn	3	1	101	3661
rebornix	83	224	290	750
sandy081	63	162	217	2493
bpasero	17	51	281	5694
aeschli	7	6	201	2167
jrieken	61	156	262	4559
ramya-rao-a	26	38	5	755
joaomoreno	125	571	187	5100
egamma	7	11	18	237
<b>Média</b>	<b>50,5</b>	<b>184,3</b>	<b>178,7</b>	<b>2770,2</b>

A primeira percepção que temos com os números acima, é que a quantidade de forks stars, em repositórios pessoais, individualmente falando, não tem influência direta na quantidade de pull requests que um programador faz em determinado repositório. A nível de exemplo, bpasero e isidorn possuem pouquíssimos stars e forks em repositórios pessoais, mas tiveram uma grande quantidade de pull requests aceitos no VSCode. A quantidade de issues pelas quais o programador está responsável no repositório segue uma melhor uniformidade se comparada com a quantidade de pull requests no repositório. Isso possivelmente tem relação com o nível de experiência ou confiança que o programador representa para aquele repositórios. Como exceções, temos egamma e ramya-rao-a.

Programador	forks	stars	issues	pull requests aceitos
Tyriar	3906,25	192457,69	2143,69	234449,64
isidorn	2256,25	33598,89	6037,29	793524,64
rebornix	1056,25	1576,09	12387,69	4081208,04
sandy081	156,25	497,29	1466,89	76839,84
bpasero	1122,25	17768,89	10465,29	8548606,44
aeschli	1892,25	31790,89	497,29	363850,24
jrieken	110,25	800,89	6938,89	3199805,44
ramya-rao-a	600,25	21403,69	30171,69	4061031,04
joaomoreno	5550,25	149536,89	68,89	5427968,04
egamma	1892,25	30032,89	25824,49	6417102,24

<b>Desvio Padrão</b>	<b>43,06</b>	<b>218,97</b>	<b>97,98</b>	<b>1822,21</b>
----------------------	--------------	---------------	--------------	----------------

A tabela acima demonstra os desvios e o desvio padrão para cada métrica.

	Forks	Stars	Issues	Pull requests aceitos
Coeficiente de variação	0,85269	1,1880992	0,548297	0,657788989

A tabela acima mostra o coeficiente de variação de cada uma das métricas. Como já comentado anteriormente, os programadores têm quantidades de issues e pull requests mais uniformes entre si, porém, existem várias discrepâncias quando verificamos a variação de forks e stars.

Dados os valores acima, podemos perceber que os dados de repositórios pessoais de um programador não são conclusivos para trabalhar uma comparação de qualidade(considerando como principal fator a quantidade de pull requests aceitos), e que uma melhor abordagem seria trabalhar em cima de dados de um repositório onde todos tenham participação(como pull requests e issues).

Respondendo à pergunta, a variação de qualidade dos programadores do VSCode está entre 0,5 e 0,7, considerando os coeficientes de variação de issues e pull requests. Considerando que quanto menor o coeficiente, mais uniforme é a amostra, esses valores demonstram uma certa oscilação na qualidade dos programadores, o que é validado pelas quantidades de issues e pull requests individuais, que variam de 5 a 290 e 237 a 5100, respectivamente.

Quanto ao objetivo, concluímos que dados de repositórios pessoais não tem relação com a qualidade de um programador, e muito menos fazem sentido para uma comparação entre vários programadores.