

1. Qual a relação entre a popularidade dos repositórios e as suas características de qualidade?

- **Hipóteses:**

- Existe uma relação direta entre a popularidade dos repositórios e as suas métricas de qualidade. Acredito que a quantidade de contribuintes em projetos é diretamente proporcional à sua popularidade. Assim, mais pessoas contribuindo para um projeto faz com que exista um filtro mais rigoroso para boas práticas de codificação. Mais pessoas contribuindo significa um aumento na probabilidade de existirem contribuidores mais experientes e com um conhecimento mais maduro de boas práticas.

- **Metodologia:**

- Para responder a esta questão, foi escrito um script em Python para fazer as seguintes tarefas: foram buscados os 1000 repositórios Java mais populares do GitHub usando a API GraphQL. Para cada um desses repositórios, foram extraídas as **quantidades de estrelas** de cada um deles para se avaliar o **grau de popularidade**, bem como a url do repositório. Então, com a url de cada repositório, clonou-se cada um deles e executou-se sobre eles a ferramenta de análise estática de código CK. Ao se executar a ferramenta, filtrou-se as métricas de qualidade: i) CBO's das classes; ii) DIT's das classes e iii) LCOM's dos métodos das classes. Foi feito um tratamento dos dados obtidos para eliminar repositórios que apresentaram algum erro durante a execução do script ou que apresentaram dados corrompidos. Dos 1000 repositórios analisados, 40 tiveram de ser eliminados. Então, finalmente, foi feito o teste estatístico de Pearson, usando a biblioteca Pandas, para se verificar a relação de correlação entre a **popularidade** dos repositórios e as características de qualidade analisadas.

- **Resultados obtidos:**

- **Discussão sobre hipóteses vs resultados:**

2. Qual a relação entre a maturidade dos repositórios e as suas características de qualidade?

- **Hipóteses:**

- Existe uma relação direta entre a maturidade dos repositórios e as suas métricas de qualidade. Acredito que a quantidade de contribuintes em projetos é diretamente proporcional à sua maturidade. Assim, mais pessoas contribuindo para um projeto faz com que exista um filtro mais rigoroso para boas práticas de codificação. Mais pessoas contribuindo significa um aumento na probabilidade de existirem contribuidores mais experientes e com um conhecimento mais maduro de boas práticas.

- **Metodologia:**

- Para responder a esta questão, foi escrito um script em Python para fazer as seguintes tarefas: foram buscados os 1000 repositórios Java mais populares do GitHub usando a API GraphQL. Para cada um desses repositórios, foram extraídas as **idades em anos** de cada um deles para se avaliar a **maturidade**, bem como a url do repositório. Então, com a url de cada repositório, clonou-se cada um deles e executou-se sobre eles a ferramenta de análise estática de código CK. Ao se executar a ferramenta, filtrou-se as métricas de qualidade: i) CBO's das classes; ii) DIT's das classes e iii) LCOM's dos métodos das classes. Foi feito um tratamento dos dados obtidos para eliminar repositórios que apresentaram algum erro durante a execução do script ou que apresentaram dados corrompidos. Dos 1000 repositórios analisados, 40 tiveram de ser eliminados. Então, finalmente, foi feito o teste estatístico de Pearson, usando a biblioteca Pandas, para se verificar a relação de correlação entre a **maturidade** dos repositórios e as características de qualidade analisadas.

- **Resultados obtidos:**

- **Discussão sobre hipóteses vs resultados:**

3. Qual a relação entre a atividade dos repositórios e as suas características de qualidade?

- **Hipóteses:**

- Levantando-se uma hipótese pura sobre a relação entre a atividade de projetos e suas métricas de qualidade, diria que essa relação não existe. Entretanto, os dados analisados estão “contaminados”. Ou seja, os repositórios que estão sendo analisados já estão no TOP-1000 mais populares do GitHub e provavelmente essa característica por si só já exerce influência sobre suas características de qualidade. Dessa forma, levanto-se em conta que os repositórios analisados já estão entre os mais populares do GitHub, diria que, dentre eles, existe sim uma relação direta entre a atividade e as métricas de qualidade pelo mesmo motivo já citado anteriormente: mais atividade provavelmente significa mais contribuintes e isso levaria a uma participação mais provável de desenvolvedores mais experientes, que manteriam níveis elevados de qualidade.

- **Metodologia:**

- Para responder a esta questão, foi escrito um script em Python para fazer as seguintes tarefas: foram buscados os 1000 repositórios Java mais populares do GitHub usando a API GraphQL. Para cada um desses repositórios, foram extraídas as **quantidades de releases** de cada um deles para se avaliar a **atividade**, bem como a url do repositório. Então, com a url de cada repositório, clonou-se cada um deles e executou-se sobre eles a ferramenta de análise estática de código CK. Ao se executar a ferramenta, filtrou-se as métricas de qualidade: i) CBO's das classes; ii) DIT's das classes e iii) LCOM's dos métodos das classes. Foi feito um tratamento dos dados obtidos para eliminar repositórios que apresentaram algum erro durante a execução do script ou que apresentaram dados corrompidos. Dos 1000 repositórios analisados, 40 tiveram de ser eliminados. Depois, foi feita uma outra “limpeza” dos dados, removendo repositórios sem nenhuma release. Provavelmente, para estes repositórios, o lançamento de releases não é feito de forma sistemática e como etapa de lançamento de atualizações, o que faz com que a métrica “quantidade de releases” não sirva como ferramenta de medição de atividade para esses repositórios. Então, finalmente, foi feito o teste estatístico de Pearson, usando a biblioteca Pandas, para se verificar a relação de correlação entre a **atividade** dos repositórios e as características de qualidade analisadas.

- **Resultados obtidos:**

- **Discussão sobre hipóteses vs resultados:**

4. Qual a relação entre o tamanho dos repositórios e as suas características de qualidade?

- **Hipóteses:**

- Não existe nenhuma relação entre o tamanho de um projeto e suas características de qualidade. Essa ausência de relação já é bastante conhecida na Engenharia de Software pelos mais diversos motivos, sendo o principal deles o fato de que linhas de código não é uma boa métrica para tamanho ou complexidade de um software. Além disso, a complexidade de um sistema não determina de forma alguma sua qualidade. O que define a qualidade são diversos outros fatores, como: i) projeto bem feito e bem organizado; ii) experiência da equipe que trabalha no projeto; iii) condução do projeto; entre vários outros.

- **Metodologia:**

- Para responder a esta questão, foi escrito um script em Python para fazer as seguintes tarefas: foram buscados os 1000 repositórios Java mais populares do GitHub usando a API GraphQL. Para cada um desses repositórios, foram extraídas as url's do repositório. Então, com a url de cada repositório, clonou-se cada um deles e executou-se sobre eles a ferramenta de análise estática de código CK. Ao se executar a ferramenta, filtrou-se as métricas de qualidade: i) CBO's das classes; ii) DIT's das classes e iii) LCOM's dos métodos das classes, assim como as LOC's de cada classe. Foi feito um tratamento dos dados obtidos para eliminar repositórios que apresentaram algum erro durante a execução do script ou que apresentaram dados corrompidos. Dos 1000 repositórios analisados, 40 tiveram de ser eliminados. Somou-se, então, para cada repositório, as linhas da métrica "LOC" para se obter a **quantidade total de linhas de código** do projeto como métrica de **tamanho**. Então, finalmente, foi feito o teste estatístico de Pearson, usando a biblioteca Pandas, para se verificar a relação de correlação entre o **tamanho** dos repositórios e as características de qualidade analisadas.

- **Resultados obtidos:**

- **Discussão sobre hipóteses vs resultados:**