

# Does the inclusion of code in a Stack Overflow question increase the solvability of a post?

Sarah Telford  
Science, Engineering and  
Technology (SET)  
The University of Adelaide  
Adelaide SA Australia  
[a1810750@student.adelaide.edu.au](mailto:a1810750@student.adelaide.edu.au)

## CCS CONCEPTS

- **Stack Overflow** → Format
- **Stack Overflow** → Solution
- **Stack Overflow** → Questions

## KEYWORDS

Stack Overflow, format, user, solution, question

### ACM Reference format:

Sarah Telford. 2022. Does the inclusion of code in a Stack Overflow question increase the solvability of a post?. In *Proceedings of ACM Woodstock conference (WOODSTOCK'18)*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/1234567890>

## 1 Introduction

Stack Overflow is “a public platform building the definitive collection of coding questions & answers”<sup>1</sup>. According to Stack Overflows insights, “about 100+ million people visit” the site each month most of whom are professional or recreational programmers<sup>2</sup>. Many programmers and developers use Stack Overflow to ask questions and find solutions to issues and errors they encounter while developing code. Currently, there are “6,826,036 questions with no upvoted or accepted answers”<sup>1</sup> posted to Stack Overflow, with the site reporting a new question being asked every 13.6 seconds<sup>1</sup>. With such a saturated platform, users need to ensure that their posts are concise and easily understood. By looking at the impact including code segments in question has on the successful answer rates, users will be able to discern whether to include code in their own questions.

## 2 Previous research

Stack Overflow provides users with some “basic tips” to “improve your chances of getting an answer”<sup>3</sup>. Included in these tips are suggestions for introducing the issue before posting any code, as well as suggested formats for posting code. This suggests that the inclusion of code segments in a question improves the likeliness of getting an answer, however, it does provide any data about the success rates.

Research suggests that the usefulness of posted code segments gravely impacts the success of questions. A 2021 study by [Saikat Mondal](#), Mohammad Masudur Rahman, et al. from the University of Cornell found that “it is not always possible to reproduce the issues

from the code segments that may impede questions from receiving prompt and appropriate solutions”<sup>4</sup>. This study found that “a reproducible question has at least two times higher chance of receiving an accepted answer than an irreproducible question”<sup>4</sup>. Therefore, it is vital that included code segments are easily reproducible for questions to be successfully answered.

Further research suggests that there lies an issue with the explanations given by users providing solutions. According to Rodrigo F. Silva, M. Masudur Rahman, et al., solutions “containing both code examples and their explanations”<sup>5</sup> are harder to come by. Even solutions that contain both explanations and example code are sometimes unclear and hard to follow.

## 3 Data Collection

In order to answer if the inclusion of code in a Stack Overflow question increases the solvability of a post, data will need to be collected from Stack Overflow. Data regarding whether a submitted question contains code segments will need to be gathered from both unanswered, answered but unaccepted, and answered and accepted posts. Due to the large number of posts made to Stack Overflow, the sample pool for data collection will be limited to posts regarding the Python programming language. Python has been selected as the focus language because it is one of the most popular programming languages used by Stack Overflow developers. According to a 2020 survey conducted by Stack Overflow, 44.1% of users program using Python<sup>2</sup>.

Stack Overflow insights provide detailed data about users, posts, languages and formats used within the site. These insights will be used to gather data about Python posts and the rate of success in receiving an acceptable answer on questions containing and not containing code segments. Data will also be gathered about reasons questions have been left unanswered and what makes a success question.

## 4 Data Analysis

The data will be collected and used in a range of different statistical comparisons. Firstly, a t-test will be created comparing answered and unanswered questions to see if there is a significant difference in the means between the two groups. This will help to determine the accuracy and applicability of the results.

Percentages will be generated for questions

- answered and containing code segments

- unanswered and containing code segments
- answered and not containing code segments
- unanswered and not containing code segments

These will be compared to determine if there is a causal relationship between questions being answered and questions containing code segments. These percentages will be presented through bar graphs.

The confidence interval for these percentages will also be calculated. This interval will show the probability that a question will be answered based on whether a code segment had been included in the question. This interval helps statistically prove the relationship.

## 5 Course Relevance

In conducting this study, I will be utilising a selection of methodologies and techniques discussed throughout the term. Mainly, the focus of this paper will implement material covering data analysis and methodologies. This is because the focus of this study is to analyse data provided by Stack Overflow to determine if there is a casual relationship between unanswered questions and code segment inclusion.

Furthermore, this paper is relevant to my current degree, Bachelor of Software Engineering, because I often use Stack Overflow when trying to debug code or run into errors. By gaining insight on how to ask effective and efficient Stack Overflow questions, I will be able to better utilise this platform.

## REFERENCES

- [1] (2008 Accessed: 2020-9-08). Stack overflow.
- [2] Stack Overflow Developer Survey 2020. (2022). Retrieved 5 April 2022, from <https://insights.stackoverflow.com/survey/2020>
- [3] How do I ask a good question? - Help Center. (2022). Retrieved 5 April 2022, from <https://stackoverflow.com/help/how-to-ask>
- [4] Mondal, S., Rahman, M., Roy, C., & Schneider, K.. (2021). The Reproducibility of Programming-Related Issues in Stack Overflow Questions.
- [5] Silva, R., Rahman, M., Dantas, C., Roy, C., Khomh, F., & Maia, M.. (2021). Improved Retrieval of Programming Solutions With Code Examples Using a Multi-featured Score.