1. **Problem Definition**

Our problem focuses specifically on the answers that have been deemed to be high quality. Specifically, we will be examining answers above a certain, as-yet-undetermined score threshold and the set of accepted answers. These answers will be evaluated against a set of metrics to answer the question of "What makes for a good answer on Stack Overflow?". We'll be examining the presence of code snippets, the presence of links, the length of the answer (in words), the answerer's reputation, and the response time for each answer to attempt to answer this question. The accepted answers and the highly-voted answers will, nominally, comprise two different data sets, but both are being measured in the same way. Our primary goal is to find trends in these metrics across the data, so that we can identify aspects of the "good" answers and, perhaps, be able to apply these trends to the general problem of writing better documentation, both in crowdsourced form and otherwise.

1. **Input and Output**

The two source sets of data we will be using are the "accepted answers" and "Highly voted answer" against the metrics of "Response Time", "Number of tags", Length of the answer (in words)", "reputation of the answerer", "links provided as answers" and "code snippets". We would extract this information by running various queries on the Stack dump provided onto an excel sheet or any tabular format for that matter and plot them against the above two datasets of "highly voted answers" and "Accepted answers". The OUTPUT would be in the form of bar charts and graphs to conduct analysis and draw conclusions from the chart statistics. We can then put forward an analysis and conclusion stating how to analyse the quality of an answer to build a higher reputation and have your answer accepted and highly voted on Stack overflow.

1. **Sources of Information**

The information we use in our analysis is derived from two sources – highly voted answers (HVA) and accepted answers (AA).

1. **Highly Voted Answers (HVA)**

Highly voted answers are a subset of all submitted Stack Overflow answers which contains only the highest rated answers as voted by the community. To obtain this subset, we first look at the user-voted rating of each answer and determine a rating threshold derived from the average score. The set of highly voted answers is then defined as all answers that exceed the average score.

1. **Accepted Answers (AA)**

Accepted answers are answers to questions which the asker felt most adequately answered the prompt. Contrary to highly voted answers, accepted answers do not have to meet a certain score threshold and, in extraordinary cases, may even be the least preferred answer as rated by the community.