

# Question earnings prediction and analysis on the website Quora

1st, February 2022

## I. Introduction

In this report, I will describe my attempt at evaluating how much money can be earned from each posted question on the website Quora.

Quora is a social question-and-answer website based in California, United States. It is considered to be one of the biggest question-and-answer applications/websites in the world, along with others such as Answers, StackExchange and Quizlet.

To increase the traffic volume and search appearance on Google, Quora needs to host an immense number of different questions spanning all topics. The intuition is that any internet user will see their concerned questions posted on Quora, leading to their clicking on the link and visiting the website. In question-answer websites like Quora, users are noted to be more likely to answer existing questions rather than post a new question. To significantly increase the number of questions, Quora has devised a strategy called Partner Program in 2018, where users in this program are paid for regularly posting new questions, helping to increase the website's traffic volume.

Source of the data: The question data is directly based on my own posted questions on Quora, which I have requested Quora to send me a copy of the data. My Partner Program was in the Japanese language, so this report paper will only investigate the income scheme of the program in Japanese, which is probably inapplicable to programs in other languages. The income data in this report is strictly exclusive to me, but you can see my questions at the address:

<https://jp.quora.com/profile/Nguyen-Thanh-Luan>

Below this, I will discuss several machine learning methods for analyzing the correlations between the individual questions' income and their features. Finally, I will make predictions and provide a conclusion on the strategy to maximize the income from the Quora Partner Program.

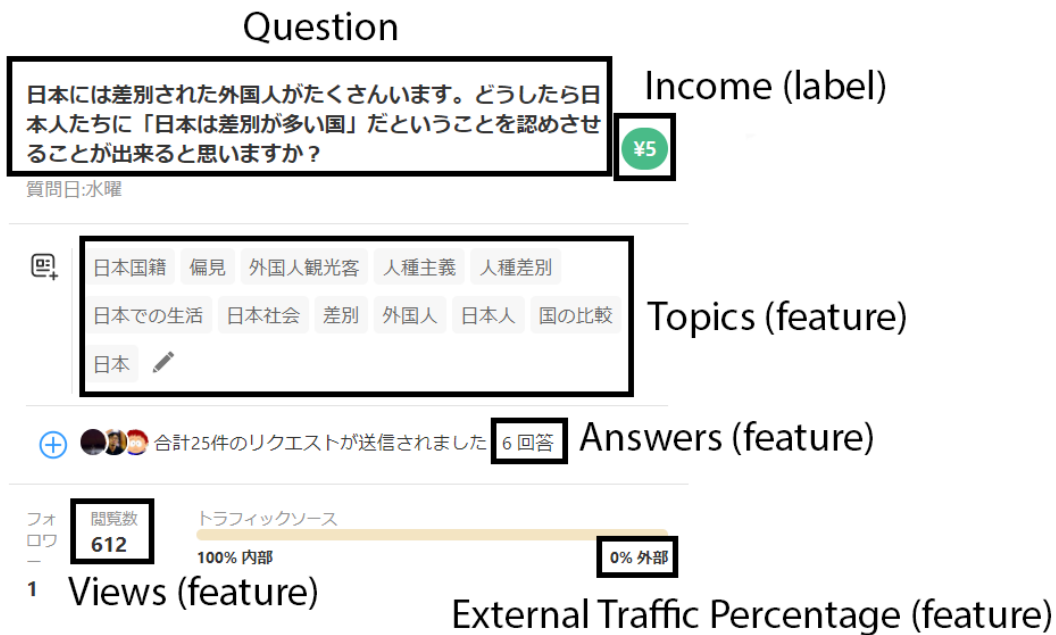
## II. Problem Formulation

The question that I am trying to answer is: given some certain characteristics of the posted questions, how much income the question is likely to earn?

According to official information, the question's income is based on how often external users click on the question, stay on the website for a certain duration of time and see the advertisements promoted by Quora. However, based on my experience, it is particularly difficult to predict whether the question will earn much money or not generate any income. This is the purpose of this paper: to systematically gather the necessary data and make a general analysis and prediction towards the income scheme of the Quora Partner Program.

For each question datapoint, various features can be recorded, but I opt to choose only four features because they are easy to be collected in a highly automated manner, namely: the number of answers, the

number of views, the external traffic percentage and the question topics. The illustration of the datapoint is clearly illustrated in Fig 1.



*Figure 1: A posted question on the Quora website with many different features*

### Summary of the machine learning problem

**Datapoint:** A single question with five properties: number of answers, number of views, external traffic percentage, question topics, and the generated income.

**Features:** The number of answers made to the question, the number of views when users click on the question, external traffic percentage and the topics of the question.

**Label:** Income of the question (in Japanese yen). The income information is not available when the question is posted. It is only known the day after the question posting => It is uncertain how much the question will earn at the time of posting.