Appendix to
Analyze This!
145 Questions for Data Scientists in Software Engineering

September 13, 2013
Technical Report
MSR-TR-2013-84

Microsoft Research
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052

Appendix to Analyze This! 145 Questions for Data Scientists in Software Engineering

Andrew Begel
Microsoft Research
Redmond, WA, USA
Andrew.Begel@microsoft.com

Thomas Zimmermann
Microsoft Research
Redmond, WA, USA
tzimmer@microsoft.com

ABSTRACT

In order to understand the questions that software engineers would like to ask data scientists about software, the software process, and about software engineering practices, we conducted two surveys: the first survey solicited questions and the second survey ranked a set of questions. Our analysis resulted in a catalog of 145 questions grouped into 12 categories as well as a ranking of the importance of each question. This technical report contains the survey text as well as the complete list of 145 questions with rankings.

1. INITIAL SURVEY

Survey on Software Development Analytics

We are researchers at Microsoft Research who are interested in understanding the questions that engineers would like to address with data. We would appreciate your feedback via this 10-minute survey. Your response will help us refine our hypotheses, develop better recommendations and tools, and focus our future research efforts around analytics for software development. For more information, please contact Thomas Zimmermann (TZIMMER) and Andrew Begel (ABEGEL).

Please do not take this survey more than once, even if you think your answers may not have been recorded the first time. It is difficult to filter duplicate submissions from the result set.

This survey is anonymous

Demographics

1. What best describes your primary work area?

- Development
- Test
- PM
- Other Engineer
- Other Non-Engineer

2. Which of the following best describes your role?
(Check at least 1 but no more than 6)

- Individual Contributor (IC)
- Lead
- Architect
- Manager
- Executive
- Other

3. How many years have you worked in the software industry?
(decimals OK)
(Min Number: 0 - Max Number: 99)

4. How many years have you worked at Microsoft? (decimals OK)
(Min Number: 0 - Max Number: 99)

Questions that you would like the answer to.

Suppose you could work with a team of data scientists and data analysts who specialize in studying how software is developed.

Please list up to five questions you would like them to answer. Why do you want to know? What would you do with the answers?

Here are some examples of questions that others have already asked.

- Do unit tests save more time in debugging than they take to write/run/keep updated?
- Do distributed version control systems offer any advantages over centralized version control systems?
- Is it really twice as hard to debug as it is to write the code in the first place?
- When does it make sense to reinvent the wheel vs. use an existing library?
- What impact does code quality have on our ability to monetize a software service?

5. Please enter your first question. Also, say why you want to know and what you would do with the answer.

(Max Characters: 2000)

6. Please enter your second question. Also, say why you want to know and what you would do with the answer.

(Max Characters: 2000)

7. Please enter your third question. Also, say why you want to know and what you would do with the answer.

(Max Characters: 2000)

8. Please enter your fourth question. Also, say why you want to know and what you would do with the answer.

(Max Characters: 2000)

9. Please enter your fifth question. Also, say why you want to know and what you would do with the answer.

(Max Characters: 2000)

Microsoft Research. Technical Report. MSR-TR-2013-84.

Follow up

11. Do you have any further comments on this topic or this survey?

(Max Characters: 2000)

12. If you would be willing to speak with us further about your responses, please enter your alias. (Or to keep your responses anonymous, click here to send us email before you submit).

(Max Characters: 256)

2. RATING SURVEY

This survey was a multi-page survey. It consisted of a start page and two pages of *randomly selected* questions to be rated by the respondent.

2.1 Start Page:

Microsoft Research Survey on Software Analytics

Software Engineering at Microsoft Research needs your help!

Please take a few minutes to fill out this two page survey about Software Development Practices. It is being conducted by Andrew Begel (abegel) and Tom Zimmermann (tzimmer) of Microsoft Research. We are interested in understanding the questions that engineers would like to address with data. Your answers will help us refine our hypotheses, develop better recommendations and tools, and focus our future research efforts around analytics for software development.

Your private and confidential response is important to us as you are one of a small number of randomly selected Microsoft employees and thus represent many of your peers. It will take about 10 minutes to complete the survey.

Thanks!

Andrew Begel (abegel) and Thomas Zimmermann (tzimmer)

2.2 Example of a Rating Page

Page 1 of 2:
Productivity / Reuse and Shared Components

Suppose you could work with a team of data scientists and data analysts who specialize in studying how software is developed.

In a previous survey, we asked Microsoft engineers to give us a list of questions they wanted the team to answer. Which ones do you think are the most important to work on first?

Please answer the following question for each item below: "In your opinion, how important is it to have a software data analytics team answer this question?"

	Essen-tial	Worth-while	Unim-portant	Unwise	*I don't un-derstand*
1. How can we measure the productivity of a Microsoft employee?	○	○	○	○	○
… (more questions)	○	○	○	○	○

3. THE 145 QUESTIONS

The 145 questions including ratings are included in the file

145Questions.xlsx

The categories and subcategories are listed in the file

Categories.xlsx