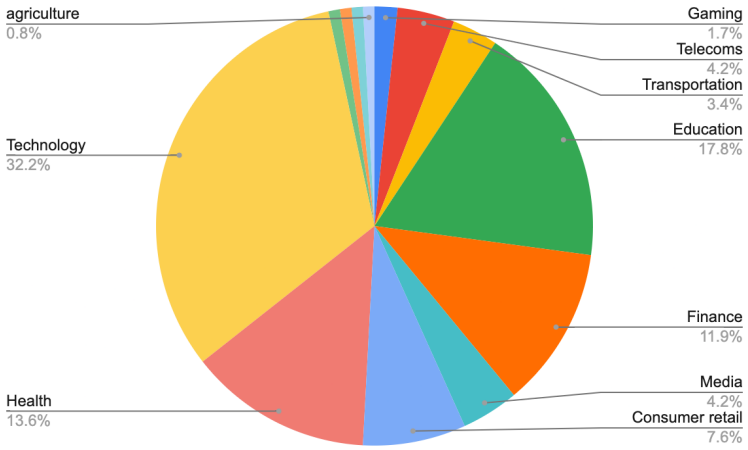
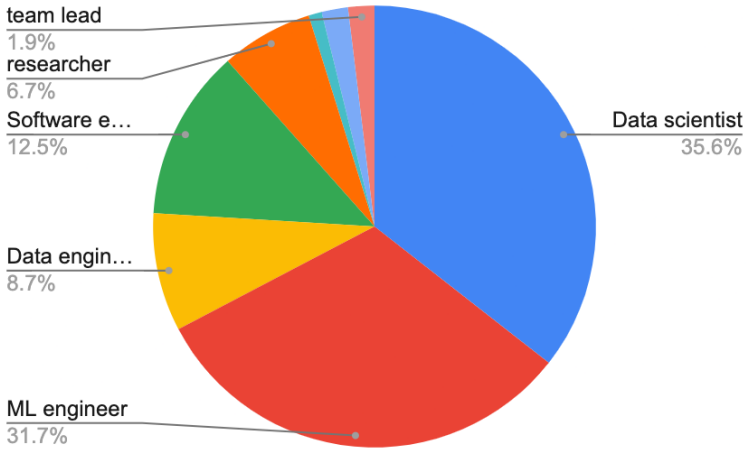
Practitioner Survey Study Protocol

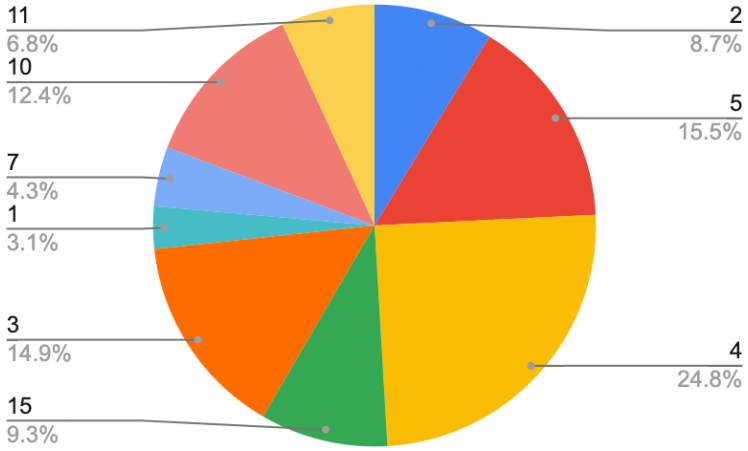
1. Protocol Title	**ML Experiment Management Tools: Practitioner Survey**
2. Background	An empirical study protocol for eliciting the opinions of practitioners on ML experiment management tools
3. Primary Objectives	Empirical findings for use as a guide and requirement for designing new and improved ML experiment management tools.
4. Secondary Objectives	Provide empirical insight on ● The kind of experiments conducted by practitioners and the specialized management tools they use. ● The challenges of experiment management with/without specialized tools, and possible adoption barriers. ● The benefits users derive from using the subject tools. ● Practitioners' preferences on tools features and paradigms.
5. Study Design	**The questionnaire instrument:** We created the questionnaire using Google Forms—a popular online survey tool. The Questionnaire consists of several sections (described below) with mandatory and non-mandatory questions. We utilized Likert-scale, open and closed-ended question forms. **Instrument Validation:** We validated our questionnaire instrument by performing 4 rounds of dry-runs, where 1 master student, 2 Ph.D. students, and 1 post-doc filled the question while providing feedback. We improved the questionnaire based on identified issues and feedback. **Participants:** A. Recruit practitioners who have working experience with ML experiment management tools. B. Upon practitioner's agreement to participate: a. Provide survey instrument (Google Form) via email. b. Participants learn about the survey objective and the background knowledge of the ML

	experiment tools. **The questionnaire content:** C. Upon the introduction of the survey instrument, we obtained permission from practitioners to elicit the survey responses. D. To ensure only relevant participants are considered, the survey identifies participants who perform ML experiments and ended the immediate survey for those who do not. E. The survey branches into two questionnaire sections with respective questions for a. Practitioners performing tools **with** experiment management tools b. Practitioners performing tools **without** experiment management tools. F. The survey questions are Likert-scale, open and closed-ended questions. G. The survey obtained demographic information of the survey participants **Analysis:** H. Systematically analyze the participants' answers using both qualitative and quantitative approaches. I. Report key findings.
6. Target Audience & sampling strategy	The primary target participants are practitioners with ML model experimentation/prototyping experience. Our sampling strategy: • Group A: Recruitment from a local ML conference • Group B: Recruitment via authors or ML experiment projects on GitHub • Group c: Recruitment via Freelance services targeting ML practitioners. We ensured the recruitment of relevant participants in Group A by verbal confirmation. For Group C, we checked the project history of the freelancer to confirm they have a relevant work history. For all groups, we ensured only those meeting this criterion are considered in the survey using Step **D** of our study design

7. Participant incentives	The incentives for group A and B are primarily based on common interests. Many of the participants enlisted to obtain the report from our study when finalized. For group C, we incentivized participation with 15 USD per participant.
8. Rationale for the number of participants	We strive for a larger number of participants, and these were the final counts obtained. Group A- 24, Group B - 25, Group C - 32.
9. Response rate	Response rate (i.e. the statistics on participants we obtained from invitations) are • Group A - About 50%, • Group B - approximately 1%, • Group C - We accepted participation from about 60% of practitioners who indicated an interest in participation.
10. Participants Demographics	**Domain/industry statistics:**  **Role stats:**

<table>
<tr>
<td></td>
<td>


Years of Experience:



</td>
</tr>
<tr>
<td>11. Data Collection and Analysis</td>
<td>The participant responses were collected via Google Forms. We downloaded the data as a CSV file.

The data represent responses to questions from Tables I, II, and III of the paper.

We imported the data into Excel for our analysis and performed both qualitative and quantitative analysis using Excel.

Qualitative Analysis: This was performed for the open-ended questions, and we employed thematic analysis.

Quantitative Analysis: The analysis was performed using excel to determine the statistics provided in the descriptive reporting. This applies to the closed-ended questions and the Likert scale. For example, we obtained the percentage of</td>
</tr>
</table>

	participants who perform ML experiment with/without ML experiment management tools.
12. Supplement files / Materials	**Recruitment cover letter:** These are the cover letter sent via email or posted as an advert to recruit survey participants • Note: We also sent out a reminder to Group A after 2 weeks of the first email. **Survey questionnaire instruments:** This is the PDF copy of the Google Form (**Survey Questions.pdf**) sent to all participants. This includes i) an Introduction to the subject and objective of the survey, ii) Questions/forms to elicit relevant information for this study iii) Questions/forms to elicit participants' details. **Datasets/responses:** The responses obtained from participants are provided in the first tab of the Excel file — **"Survey Responses.xlxs"**. The Batch column indicates the response group. **Group A** responses are indicated as Batch 1, **Group B** as Batch 2, and **Group C** as Batch 3. **Codebook:** For the open-ended questions, we provide the summary of the codebook for the relevant question where applicable. This information is provided in separate tabs of the Excel file - ***"Qualitative analysis.xlxs"***
13. Response management	In total, we obtained responses from 81 participants. Some of the items of the questionnaire are mandatory. E.g., questions used to determine section branching are compulsory. We only obtained responses from participants who completed and submitted the questionnaire form. Hence, we do not track drop-outs. For the descriptive reporting of our analysis, we provide analysis results as percentages based on the obtained responses. This account for missing

	responses in non-mandatory questions. Note: We decided on a mix of mandatory and non-mandatory questions to improve the completion rate.
14. Construct Validity	To ensure the correct interpretation of our survey questionnaire, we tried to make participants understand the basis of experiment management tools. First, we provide a summarized explanation of the tools in the invitation letters. Second, the questionnaire instrument starts with an introduction section to aid a better understanding of the survey questions. In addition, we used the validation phase of the survey instrument (dry-runs conducted before the actual survey) to ensure the correctness of our survey questions and terminology by confirming that our dry-run participants understood the questions as intended. This resulted in an iterative process where all authors reviewed and revised the survey instruments for improvement. Lastly, we mitigated construct validity by targeting practitioners with relevant experience, ensuring they have domain knowledge, which can aid in correctly interpreting our survey questions.