

Final Report

Thomas Gao, Robert Hosbach, Devesh Khandelwal, Joe Mirza

W241 | Summer 2021

Abstract

<>

Introduction

A wealth of existing research shows that using monetary incentives can increase survey response rates (see Armstrong, 1975; Kanuk & Berenson, 1975; and Duncan, 1979). James & Bolstein (1992) found that an incentive of only \$1 that was mailed with the survey significantly increased the response rate compared to people who received no incentive. More recently, Debell et al. (2020) found that including \$5 of visible money in a mailed survey increased response rates from roughly 43% to 47%. Such studies have also been conducted on response rates for surveys administered to college students. For instance, Szélenyi et al. (2005) found that a \$2 incentive led to increased response rates among college students compared to no incentive.

This body of research suggests that direct financial incentives can lead to higher survey response rates. But, while direct financial incentives have been studied in depth, the effect of indirect, philanthropic donations are comparatively less well understood. Gattellari & Ward (2001) found that offering donations to Australian surgeons’ alma maters actually led to lower response rates when compared to offering no incentive at all. Nesrallah et al. (2014) and Warwick et al. (2019) had similar findings in medical-related settings. Outside of the medical field, Pedersen & Nielsen (2016) also found that incentives promising donations to a good cause led to decreased response rates relative to no incentives in online surveys.

We conducted a field experiment to test the effect of two different incentives on the survey response rates from alumni of UC Berkeley’s Masters of Information and Data Science (MIDS) program. One incentive (hereafter referred to as the “direct incentive”) is a chance to be one of 10 winners of a \$25 Amazon gift card if the alumnus completes the survey. This direct financial incentive is akin to the direct financial incentives discussed above. The other incentive (hereafter referred to as the “philanthropic incentive”) is an opportunity to have us, the survey developers, donate funds to the Berkeley Student Food Collective (BSFC) if a sufficiently high survey response rate is achieved. While some studies found donation-based incentives counterproductive, our experiment takes a slightly different approach. Instead of promising a donation to BSFC for an individual’s response, we will test the effect of survey response rates when individuals are told that the donation will only occur if a certain threshold of group response rate is met.

In this experiment, we evaluated the effects of the direct and philanthropic incentives to cause higher survey response rates when compared to a control group that received no incentive. All three groups were given the same Qualtrics survey that we distributed via one of our personal email addresses; however, only alumni in the direct and philanthropic incentive groups received text in their emails regarding the incentive they were assigned. We hypothesized that both the direct and philanthropic incentives would cause higher survey response rates among the MIDS alumni. As mentioned, direct incentives have led to increased survey response rates in certain contexts, whereas donation-based incentives have shown far less promise. Nevertheless, we still expected the philanthropic incentive to garner higher survey response rates relative to the control group for the following reasons: 1) The treatment is a group incentive rather than an individual incentive, and being part of a group working towards a common goal can be inherently motivating; 2) The certainty of the donation to charity when the threshold is reached might be preferred over the uncertainty of receiving a monetary reward; and 3) Knowing you are helping to donate money to a good cause can evoke a sense of meaning when compared to receiving money individually.

In the following section we describe our methodology for conducting this experiment, including the tools we used, our data collection process, how we dealt with attrition and non-compliance, and other details. Next, we report our experimental results and provide a discussion of their meaning and generalizability. We also discuss potential follow-on opportunities from this experiment. Finally, we conclude our paper and provide a series of appendices containing the email text we used, Python and R code we used for data collection and analysis, the actual survey, and a summary of the survey responses we received.

Methodology

In this section we explain the key components of our experimental design, including our sampling frame, data collection and blocking, survey design, email sending strategy, and identification of compliers, non-compliers, and attriters.

Scraping alumni data

To obtain the official UC Berkeley School of Information (hereafter referred to as UCB I School) email addresses for all MIDS alumni, our team wrote a Python script to scrape selected pages from the UCB I School website’s people directory (“People,” 2021). The script relies on the Selenium package for Python, which is an open-source tool for browser automation (Muthukadan, 2021). The script is provided for reference in Appendix B.

Specifically, the team scraped information from all pages listing MIDS graduates, of which the website lists 1,105 as of the end of Spring 2021. These 1,105 alumni represent the sampling frame for this experiment. The web pages include not only the I School email address of each alumnus, but also the person’s name and graduation year.¹

Blocking

In order to improve the precision of our estimated causal effect, we randomly assigned the 1,105 alumni into the control and two treatment groups using blocks based on graduation year (2015-2021) and gender (male, female, or unknown). We expect blocking on graduation year to improve the precision of our estimated causal effect because it is plausible that earlier MIDS graduates might have lower survey response rates, on average, than alumni who more recently graduated from MIDS. Similarly, we blocked on gender because studies suggest that women may be more likely to respond to surveys than men in a variety of contexts (*e.g.*, Smith, 2008).

Alumni gender designations were not available to scrape from the UCB I School website’s people directory. Therefore, we inferred each alumnus’ gender based on the person’s name and photograph (from the people directory or LinkedIn, if available).² Out of the 1,105 MIDS alumni, we confidently classified 789 (71.4%) as male and 299 (27.1%) as female. We classified the remaining 17 (1.5%) alumni as “unknown” because we did not have enough information to confidently classify them as male or female.

We used the blockTools library in R to block-randomize the 1,105 alumni into the control and two treatment groups by blocking on graduation year and inferred gender (Moore & Schnakenberg, 2016). Figure 1 shows the number of alumni assigned to each of the three groups by graduation year and gender. Each cell in Figure 1 contains three numbers, which correspond to the number of alumni assigned to each of three groups (control and two treatment) for that block. For a given block, the same number of alumni are assigned to each group when possible. In cases where the number of alumni in the block do not evenly divide into three groups, the number of alumni assigned to each group for a given block never differs by more than one.

¹There were two missing email addresses and one missing graduation year in the scraped data. We identified the email addresses using the UCB I School Slack workspace and validated the missing graduation year using LinkedIn.

²We used a binary gender designation: male or female. Without more data, more specific gender designations were not possible. Additionally, blocking based on more disaggregated gender designations seems unlikely to improve the precision of our causal effect estimate relative to a binary gender designation.

Inferred Gender	Graduation Year						
	2015	2016	2017	2018	2019	2020	2021
Female	2	6	13	19	20	29	10
	3	6	13	19	20	29	10
	2	6	14	19	21	29	9
Male	11	32	36	46	60	60	19
	10	32	36	46	60	61	18
	10	32	36	46	59	61	18
Unknown	0	0	2	1	1	1	1
	0	0	2	1	1	1	0
	1	1	1	1	2	0	0

Figure 1: Number of MIDS alumni block-randomized into each of three groups (control and two treatment) based on inferred gender and graduation year

Survey design and delivery architecture

Our experiment followed a progression of randomization, treatment, and outcome measurement. Figure 2 depicts the ROXO representation of our experiment, where the same ROXO design was applied to each of the 21 blocks we randomized within (denoted by the blue arrows). In Figure 2, “R” represents the block-randomization step, “X₁” and “X₂” represent the different treatment emails sent to the direct and philanthropic incentive groups, and “O” represents measurement of survey completions (our outcome).

Inferred Gender	Graduation Year						
	2015	2016	2017	2018	2019	2020	2021
Female	R O						
	R X ₁ O						
	R X ₂ O						
Male							
Unknown							

Figure 2: ROXO construct applied to each of the 21 randomization blocks

We delivered the control and treatment messages to the three groups of MIDS alumni using emails sent from Devesh Khandelwal’s personal I-School email address. (We found from testing that using a less personal email address resulted in more of our emails landing in recipients’ “promotions” or “social” Gmail inboxes, rather than their “primary” inboxes.)

The subject lines of the emails to all three groups were identical for the initial emails as well as the reminder emails (see Table 1). Moreover, the body text of the emails were also identical for all three groups, except for the treatment text that was accentuated using bold typeface.

Table 1: Email subject lines for the initial and reminder emails

Email	Subject
Initial (July 2, 2021)	«Alumnus Name»-How Can MIDS Improve? Voice Your Opinion
July 9 and 16, 2021 Reminders	Reminder: «Alumnus Name»-How Can MIDS Improve? Voice Your Opinion
July 22, 2021 Reminder	ONE DAY LEFT : «Alumnus Name»-How Can MIDS Improve? Voice your opinion
July 23, 2021 Reminder	CLOSING TONIGHT: «Alumnus Name»-How Can MIDS Improve? Voice your opinion

For all three groups, the treatment text indicates that alumni who complete the survey will receive a summary of the survey results. In addition, the direct incentive group was told that alumni who complete the survey will be entered to win a \$25 Amazon gift card, and the philanthropic incentive group was told that if the overall survey response rate reaches 60%, we will donate \$250 to the Berkeley Student Food Collective. Table 2 shows the treatment text for all three groups, and Appendix A includes the original and reminder email templates sent to all three groups.

Table 2: Treatment email text for the control and two treatment groups

Group	Subject
Control	If you complete the survey, we will send you a summary of the results.
Direct incentive	If you complete the survey, we will send you a summary of the results and you will be entered to win an Amazon gift card for \$25. Ten respondents will be selected at random to receive a gift card.
Philanthropic incentive	If you complete the survey, we will send you a summary of the results. Additionally, if we achieve a 60% response rate, we will donate \$250 to the Berkeley Student Food Collective.

Our outcome was a measurement of survey completions. We designed a two-part survey using Qualtrics to ask about the perceived value of various aspects of the MIDS program (part one) as well as to understand the respondent’s rationale for enrolling in MIDS (part two). The first part asks the respondent to evaluate eight statements using a Likert scale. The second part asks the respondent to answer a single multiple choice question. We intentionally made the survey short—respondents should be able to complete the survey in 2-3 minutes—in order to maximize the likelihood of any given alumnus completing the survey. The survey questions and a high-level results summary are provided in Appendix C.

Sending emails and reminders

We sent initial emails to the control and two treatment groups on Friday, July 2, 2021 using Yet Another Mail Merge (YAMM). The YAMM service took approximately 12 minutes to send the emails to the 370 (approximately) alumni in each group. Table 3 indicates the start times for sending emails to each group.

Table 3: Start times for sending initial emails to the control and two treatment groups

Group	Email Batch Start Time on July 2, 2021 (PDT)
Direct incentive	9:26 am
Philanthropic incentive	10:14 am
Control	10:40 am

In addition to the initial email that we sent to all 1,105 alumni, we sent four follow-up reminder emails. We sent the first reminder email to all 1,105 alumni starting at 9:00 am (PDT) on July 9, 2021 (even to those who had already completed the survey). We sent the remaining three reminder emails starting at 9:15 am (PDT) on July 16, 8:00 am (PDT) on July 22, and 9:15 am (PDT) on July 23, 2021 only to alumni who had not already completed the survey by the evening prior to sending the emails. Figure 3 shows the cumulative number of completed surveys over time, with dashed lines signifying when we sent emails. The figure shows a clear pattern of a sharp increase in survey responses immediately after the emails are sent, with diminishing returns over time.

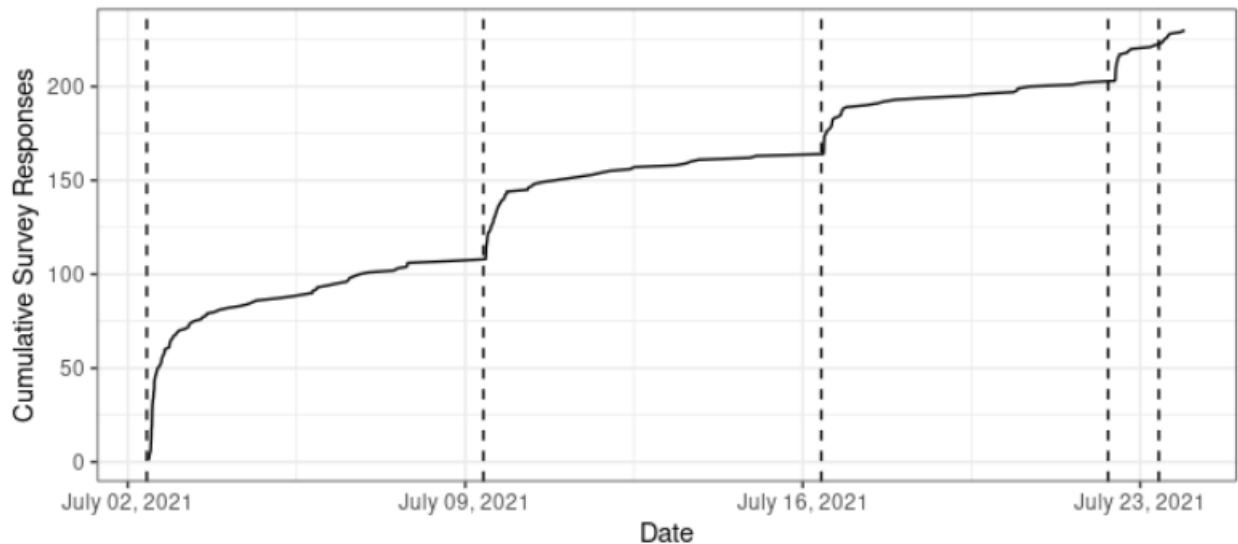


Figure 3: Cumulative survey completions over the experimental period with initial and reminder email send dates indicated by dashed vertical lines

Identifying compliers, non-compliers, and attriters

The YAMM service provides a near-real time merge status indicator for each email that is sent out. Possible status options are: sent, opened, clicked, responded, bounced, and unsubscribed. While YAMM does not guarantee the reported status to be 100% accurate, this is the only source of data that allows us to identify non-compliers in our study (“How Accurate Is Our Tracking Tool?” n.d.).

We placed the treatment text specific to the control, direct incentive, and philanthropic incentive groups in the body of the emails, rather than in the email subject lines. This allowed us to define a complier as someone who opened the email, given the assumption that if the person opened the email they read the treatment text. We therefore identified compliers as those alumni for whom YAMM reported that they opened, clicked, or responded to at least one of the emails we sent them. Conversely, non-compliers are defined as those who receive at least one of the emails, but never open any emails they receive. As a result, these alumni are never presented with a dosage of the treatment corresponding to the group to which we assigned them. We identified non-compliers in our study as those for whom YAMM reported that the email was sent, but not bounced (nor opened, clicked, or responded). Finally, attriters are defined as those alumni who never receive any of the emails we sent them. Any alumni for whom YAMM reported a bounced email from our initial email batch was deemed an attriter. Figure 4 is a flow diagram that shows the number of alumni in each group after accounting for attriters and non-compliers. In total, 229 MIDS alumni completed the survey, providing an overall response rate of approximately 21%.

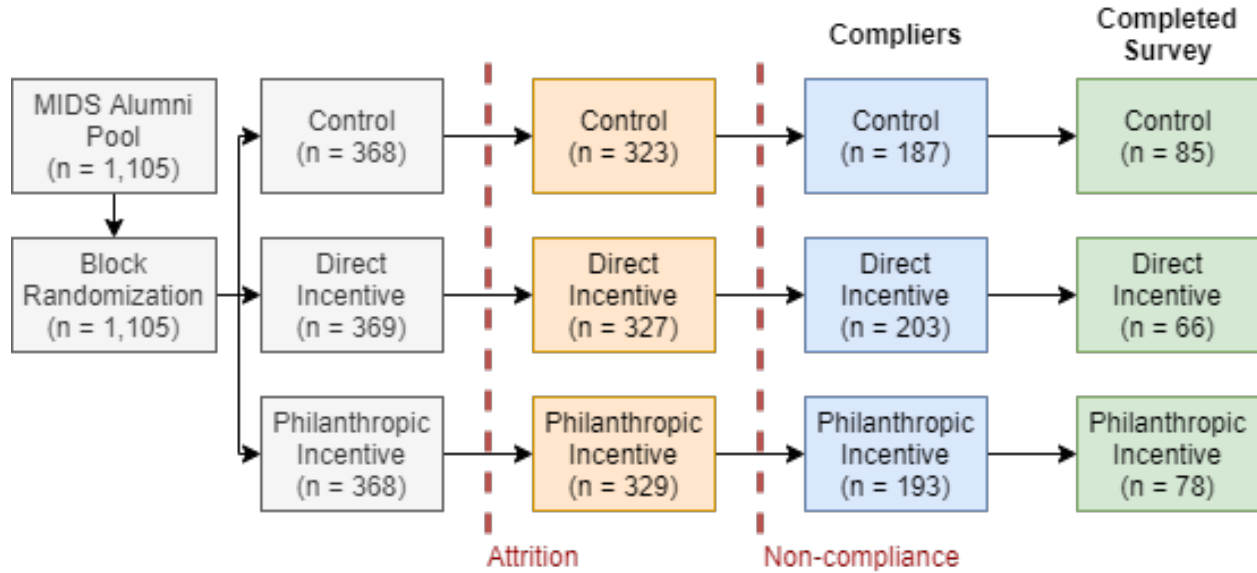


Figure 4: Flow diagram indicating the number of MIDS alumni by group after accounting for attrition and non-compliance

Calculating treatment effects

For this experiment, we calculated the following treatment effects: intent-to-treat (ITT) effect, complier-average causal effect (CACE), and heterogeneous treatment effects (HTEs) using gender and graduation year subgroups. When calculating these treatment effects, attriters were omitted from the analysis. We omitted attriters for two reasons: 1) We do not have a measured outcome for these alumni, and 2) Omitting attriters should not bias our estimated treatment effects because the group assignment has no influence on who attrits, which is determined by whether or not the person's email account is active.

The ITT effect calculation provides an estimated treatment effect based only on the group assignment. Because this calculation requires no discerning between compliers and non-compliers, the ITT effect is our primary result from this experiment. The CACE, in contrast, does require knowledge of the number of compliers and non-compliers in each group. Due to the aforementioned issues with YAMM, obtaining these estimates required us to make assumptions about who actually opened at least one of the emails they received. As a result, we do not have as much trust in our CACE estimates as we do for our ITT effect estimates. Finally, we analyzed HTEs for gender and graduation year subgroups based only on group assignment. This allowed us to avoid making compliance assumptions, and it means that the estimates we obtained should be interpreted as heterogeneous ITT effects.

Results

<>

Discussion

<>

Conclusion

<>

References

- Armstrong, J. S. (1975). Monetary Incentives in Mail Surveys. *The Public Opinion Quarterly*, 39(1), 111–116. Retrieved from <https://www.jstor.org/stable/2748074>
- Debell, M., Maisel, N., Edwards, B., Amsbary, M., & Meldener, V. (2020). Improving Survey Response Rates with Visible Money. *Journal of Survey Statistics and Methodology*, 8(5), 821–831. <https://doi.org/10.1093/jsam/smz038>
- Duncan, W. J. (1979). Mail Questionnaires in Survey Research: A Review of Response Inducement Techniques. *Journal of Management*, 5(1), 39–55. <https://doi.org/10.1177/014920637900500103>
- Gattellari, M., & Ward, J. E. (2001). Will donations to their learned college increase surgeons' participation in surveys?: A randomized trial. *Journal of Clinical Epidemiology*, 54(6), 645–649. [https://doi.org/10.1016/S0895-4356\(00\)00350-4](https://doi.org/10.1016/S0895-4356(00)00350-4)
- How accurate is our tracking tool? (n.d.). Retrieved from <https://support.yet-another-mail-merge.com/hc/en-us/articles/212022085-How-accurate-is-our-tracking-tool->
- James, J. M., & Bolstein, R. (1992). Large Monetary Incentives and Their Effect on Mail Survey Response Rates. *The Public Opinion Quarterly*, 56(4), 442–453. Retrieved from <https://www.jstor.org/stable/2749201>
- Kanuk, L., & Berenson, C. (1975). Mail Surveys and Response Rates: A Literature Review. *Journal of Marketing Research*, 12(4), 440–453. <https://doi.org/10.1177/002224377501200408>
- Moore, R. T., & Schnakenberg, K. (2016). *Package 'blockTools'*. CRAN. Retrieved from <https://cran.r-project.org/web/packages/blockTools/blockTools.pdf>
- Muthukadan, B. (2021). *Selenium with Python*. Retrieved from <https://selenium-python.readthedocs.io/>
- Nesrallah, G., Barnieh, L., Manns, B., Clase, C., Mendelssohn, D., & Guyatt, G. (2014). A charitable donation incentive did not increase physician survey response rates in a randomized trial. *Journal of Clinical Epidemiology*, 67(4), 482–483. <https://doi.org/10.1016/j.jclinepi.2013.11.005>
- Pedersen, M. J., & Nielsen, C. V. (2016). Improving Survey Response Rates in Online Panels: Effects of Low-Cost Incentives and Cost-Free Text Appeal Interventions. *Social Science Computer Review*, 34(2), 229–243. <https://doi.org/10.1177/0894439314563916>
- People. (2021). Retrieved from <https://www.ischool.berkeley.edu/people>
- Smith, W. (2008). Does Gender Influence Online Survey Participation? A Record-Linkage Analysis of University Faculty Online Survey Response Behavior. *Online Submission*.
- Szelényi, K., Bryant, A. N., & Lindholm, J. A. (2005). What Money Can Buy: Examining the effects of prepaid monetary incentives on survey response rates among college students. *Educational Research and Evaluation*, 11(4), 385–404. <https://doi.org/10.1080/13803610500110174>
- Warwick, H., Hutyra, C., Politzer, C., Francis, A., Risoli, T. J., Green, C., ... Mather, R. C. I. (2019). Small Social Incentives Did Not Improve the Survey Response Rate of Patients Who Underwent Orthopaedic Surgery: A Randomized Trial. *Clinical Orthopaedics and Related Research*, 477(7), 1648–1656. <https://doi.org/10.1097/CORR.0000000000000732>

Appendix A: Emails

This appendix contains the initial email templates that were sent out to the control and two treatment groups of MIDS alumni on July 2, 2021, as well as the reminder email templates that were sent in subsequent weeks. Areas in the templates demarcated by “«...»” indicate dynamic fields that are filled in by YAMM as the email gets sent.

Initial emails sent on July 2, 2021

Control group

Hi, «Alumnus Name»:

We are a group of MIDS students conducting research to gather insights on the experience of MIDS alumni. Please take a 2-3 minute survey to provide your opinion of the program. The insights from this survey will help us understand the value of the program as perceived by you, the alumni.

If you complete the survey, we will send you a summary of the results.

The deadline to submit your survey response is Friday, July 23, 2021 by midnight (PDT).

[Click here to take the survey](#) and ensure your voice is heard.

Thanks for your time, Devesh, Robert, Thomas, and Joe

Direct incentive group

Hi, «Alumnus Name»:

We are a group of MIDS students conducting research to gather insights on the experience of MIDS alumni. Please take a 2-3 minute survey to provide your opinion of the program. The insights from this survey will help us understand the value of the program as perceived by you, the alumni.

If you complete the survey, we will send you a summary of the results and you will be entered to win an Amazon gift card for \$25. Ten respondents will be selected at random to receive a gift card.

The deadline to submit your survey response is Friday, July 23, 2021, by midnight (PDT).

[Click here to take the survey](#) and ensure your voice is heard.

Thanks for your time, Devesh, Robert, Thomas, and Joe

Philanthropic incentive group

Hi, «Alumnus Name»:

We are a group of MIDS students conducting research to gather insights on the experience of MIDS alumni. Please take a 2-3 minute survey to provide your opinion of the program. The insights from this survey will help us understand the value of the program as perceived by you, the alumni.

If you complete the survey, we will send you a summary of the results. Additionally, if we achieve a 60% response rate, we will donate \$250 to the Berkeley Student Food Collective.

The deadline to submit your survey response is Friday, July 23, 2021 by midnight (PDT).

[Click here to take the survey](#) and ensure your voice is heard.

Thanks for your time, Devesh, Robert, Thomas, and Joe

Reminder emails sent on July 9 and July 16, 2021

Control group

Hi, «Alumnus Name»:

We are a group of MIDS students conducting research to gather insights on the experience of MIDS alumni. On July 2nd we sent you a link to a 2-3 minute survey that allows you to provide your opinion of the MIDS program.

If you have not yet completed the alumni survey, this is a reminder that you have «14 or 7» days left to respond.

If you complete the survey, we will send you a summary of the results.

[Click here to take the survey](#) and ensure your voice is heard.

Thanks for your time, Devesh, Robert, Thomas, and Joe

Direct incentive group

Hi, «Alumnus Name»:

We are a group of MIDS students conducting research to gather insights on the experience of MIDS alumni. On July 2nd we sent you a link to a 2-3 minute survey that allows you to provide your opinion of the MIDS program.

If you have not yet completed the alumni survey, this is a reminder that you have «14 or 7» days left to respond.

If you complete the survey, we will send you a summary of the results and you will be entered to win an Amazon gift card for \$25. Ten respondents will be selected at random to receive a gift card.

[Click here to take the survey](#) and ensure your voice is heard.

Thanks for your time, Devesh, Robert, Thomas, and Joe

Philanthropic incentive group

Hi, «Alumnus Name»:

We are a group of MIDS students conducting research to gather insights on the experience of MIDS alumni. On July 2nd we sent you a link to a 2-3 minute survey that allows you to provide your opinion of the MIDS program.

If you have not yet completed the alumni survey, this is a reminder that you have «14 or 7» days left to respond.

If you complete the survey, we will send you a summary of the results. Additionally, if we achieve a 60% response rate, we will donate \$250 to the Berkeley Student Food Collective.

[Click here to take the survey](#) and ensure your voice is heard.

Thanks for your time, Devesh, Robert, Thomas, and Joe

Reminder emails sent on July 22, 2021

Control group

Hi, «Alumnus Name»:

We are a group of MIDS students conducting research to gather insights on the experience of MIDS alumni. On July 2nd we sent you a link to a 2-3 minute survey that allows you to provide your opinion of the MIDS program.

If you have not yet completed the alumni survey, this is a reminder that you have 1 more day left to respond.

If you complete the survey, we will send you a summary of the results.

[Click here to take the survey](#) and ensure your voice is heard.

Thanks for your time, Devesh, Robert, Thomas, and Joe

Direct incentive group

Hi, «Alumnus Name»:

We are a group of MIDS students conducting research to gather insights on the experience of MIDS alumni. On July 2nd we sent you a link to a 2-3 minute survey that allows you to provide your opinion of the MIDS program.

If you have not yet completed the alumni survey, this is a reminder that you have 1 more day left to respond.

If you complete the survey, we will send you a summary of the results and you will be entered to win an Amazon gift card for \$25. Ten respondents will be selected at random to receive a gift card.

[Click here to take the survey](#) and ensure your voice is heard.

Thanks for your time, Devesh, Robert, Thomas, and Joe

Philanthropic incentive group

Hi, «Alumnus Name»:

We are a group of MIDS students conducting research to gather insights on the experience of MIDS alumni. On July 2nd we sent you a link to a 2-3 minute survey that allows you to provide your opinion of the MIDS program.

If you have not yet completed the alumni survey, this is a reminder that you have 1 more day left to respond.

If you complete the survey, we will send you a summary of the results. Additionally, if we achieve a 60% response rate, we will donate \$250 to the Berkeley Student Food Collective.

[Click here to take the survey](#) and ensure your voice is heard.

Thanks for your time, Devesh, Robert, Thomas, and Joe

Final reminder emails sent on July 23, 2021

Control group

Hi, «Alumnus Name»:

We are a group of MIDS students conducting research to gather insights on the experience of MIDS alumni. On July 2nd we sent you a link to a 2-3 minute survey that allows you to provide your opinion of the MIDS program.

If you have not yet completed the alumni survey, this is a reminder that you have a few hours left to respond.

If you complete the survey, we will send you a summary of the results.

[Click here to take the survey](#) and ensure your voice is heard.

Thanks for your time, Devesh, Robert, Thomas, and Joe

Direct incentive group

Hi, «Alumnus Name»:

We are a group of MIDS students conducting research to gather insights on the experience of MIDS alumni. On July 2nd we sent you a link to a 2-3 minute survey that allows you to provide your opinion of the MIDS program.

If you have not yet completed the alumni survey, this is a reminder that you have a few hours left to respond.

If you complete the survey, we will send you a summary of the results and you will be entered to win an Amazon gift card for \$25. Ten respondents will be selected at random to receive a gift card.

[Click here to take the survey](#) and ensure your voice is heard.

Thanks for your time, Devesh, Robert, Thomas, and Joe

Philanthropic incentive group

Hi, «Alumnus Name»:

We are a group of MIDS students conducting research to gather insights on the experience of MIDS alumni. On July 2nd we sent you a link to a 2-3 minute survey that allows you to provide your opinion of the MIDS program.

If you have not yet completed the alumni survey, this is a reminder that you have a few hours left to respond.

If you complete the survey, we will send you a summary of the results. Additionally, if we achieve a 60% response rate, we will donate \$250 to the Berkeley Student Food Collective.

[Click here to take the survey](#) and ensure your voice is heard.

Thanks for your time, Devesh, Robert, Thomas, and Joe

Appendix B: Scraping MIDS Alumni Data

This appendix contains the Python script we used to scrape the name, email, and graduation date for each MIDS alumnus from the UC Berkeley School of Information website's people directory.

```
from collections import namedtuple # Facilitate collect/convert into Pandas
import time # To pause between page loads

from bs4 import BeautifulSoup # To parse the HTML pages
import pandas as pd
# Selenium allows us to open a web browser and act on web pages programmatically
from selenium import webdriver
from selenium.webdriver.common.keys import Keys

# Open a Chrome browser
driver = webdriver.Chrome()

# This file contains a person's username and password
# (each on its own line) for logging into I-School
with open("creds.txt", "r") as file:
    creds = file.readlines()

# Navigate to login page and authenticate
driver.get("https://www.ischool.berkeley.edu/user/login?destination=home")
username = driver.find_element_by_id("edit-name") # Find the username entry
password = driver.find_element_by_id("edit-pass") # Find the password entry
username.send_keys(creds[0].strip()) # Fill in the username
password.send_keys(creds[1].strip()) # Fill in the password
driver.find_element_by_id("edit-submit").click() # Click the login button

# Navigate to first page of alumni, then pause for 5 seconds for loading
start_url = (
    "https://www.ischool.berkeley.edu/people?name=&role=126&degr=MIDS" +
    "&year%5Bvalue%5D%5Byear%5D=&spec=All&emp=&faculty_type=All"
)
driver.get(start_url)
time.sleep(5)

# Initialize namedtuple for data collection
Alumni = namedtuple('Alumni', ["name", "cohort", "email"])
# This list will store `Alumni` namedtuples for each person
all_data = []

# Scrape all data we can for each page
flag = True
page = 1
while flag:
    print(f"Starting page {page}.")
    # Get the HTML from the current browser page
    soup = BeautifulSoup(driver.page_source, 'html5lib')

    # This is a way to get a list of "chunks" of the HTML that
    # correspond to a given alumnus
```

```

data = [x.parent.text.strip() for x in soup.find_all(
    "div",
    class_='views-field views-field-field-profile-fullname')
]

for person in data:
    name = str(person.split("\n")[0].strip()) # 1st item is always the name
    try:
        # If an entry contains "20," that's the cohort
        cohort = str([x.strip() for x in person.split("\n") if "20" in x][0])
    except IndexError:
        cohort = ""
    try:
        # If an entry contains "@," that's the email
        email = str([x.strip() for x in person.split("\n") if "@" in x][0])
    except IndexError:
        email = ""
    # Append an `Alumni` namedtuple for the person to `all_data`
    all_data.append(Alumni(name, cohort, email))

try:
    # If there is another page of alumni, click the "next" button
    driver.find_element_by_class_name("pager__item--next").click()
    page += 1
    time.sleep(5) # Pause for not overloading server
except:
    print("Finished with the last page!")
    flag = False

results_df = pd.DataFrame(all_data)

```

Appendix C: Qualtrics Survey and Results Summary

This appendix contains the actual Qualtrics survey that was sent to each of the MIDS alumni, as well as a graphical summary of the completed survey responses. Figure 5 is the screenshot of the Qualtrics survey that was sent to each of the MIDS alumni. Figure 6 and Figure 7 are summary visualizations of the 229 responses we received from the MIDS alumni.

Berkeley
UNIVERSITY OF CALIFORNIA

Please indicate whether you agree with the following statements. If you are unable to answer the question, please choose "Not applicable".

	Strongly disagree	Somewhat disagree	Neutral	Somewhat Agree	Strongly Agree	Not Applicable
I would recommend MIDS to others interested in pursuing a career in data science.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
MIDS offers good value-for-money relative to alternatives, including self-education, other programs, etc.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
MIDS has helped me achieve my career objectives.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
MIDS provided unique benefits relative to other data science degrees.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have found MIDS for Life to be valuable post-graduation.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The MIDS network of students, alumni and instructors have helped me achieve my career objectives post-graduation.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The instruction for the program as a whole was valuable, including live sessions, assigned readings, problem sets, projects, Slack, etc.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
MIDS career services has helped me achieve my career objectives.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

What was your rationale for enrolling in MIDS?

☐ To build on an existing data science career you were already employed in.

☐ To enable a career change into data science.

☐ To add data science to your resume or skill set, for work in a non-data science role.

☐ Other

Figure 5: Screenshot of Qualtrics survey

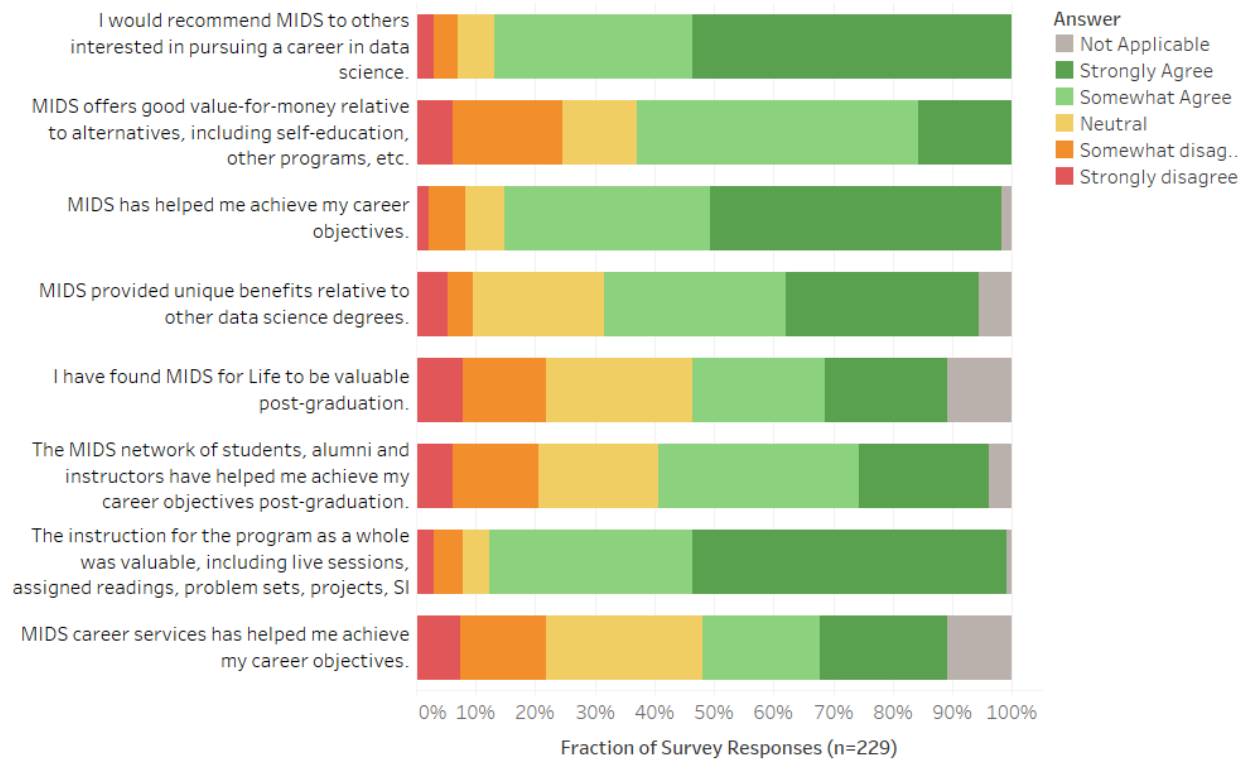


Figure 6: Summary of responses to the first set of statements from the Qualtrics survey

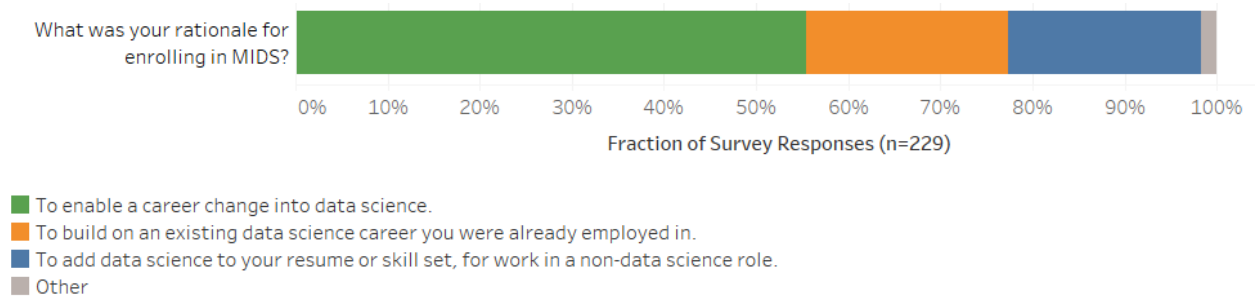


Figure 7: Summary of responses to the last question from the Qualtrics survey