# Does Joining Clubs Lead to Better Jobs?

by

David Wang

An honors thesis submitted in partial fulfillment

of the requirements for the degree of

Bachelor of Science

Undergraduate College

Leonard N. Stern School of Business

New York University

May 2017

| | |
|---|---|
| Professor Marti G. Subrahmanyam | Professor Vishal Singh |
| Faculty Adviser | Thesis Adviser |

# Does joining clubs lead to better jobs?

David H. Wang

May 2017

An honors thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science from the Undergraduate College of the Leonard N. Stern School of Business of New York University.

## Abstract

This analysis attempts to shed more light on the world of extracurricular activities at the undergraduate level, and how that might translate into future job prospects. I collected data from NYU Stern's Office of Student Engagement for club activity data, and built a web scraper to mine LinkedIn for graduate job placement data. While certain shortcomings prevented the analysis from being statistically conclusive, there were numerous other interesting findings regarding how graduates placed at clubs and how students engaged with clubs on campus. Key visualization tools were left out of this version of the thesis due to data sensitivity concerns.

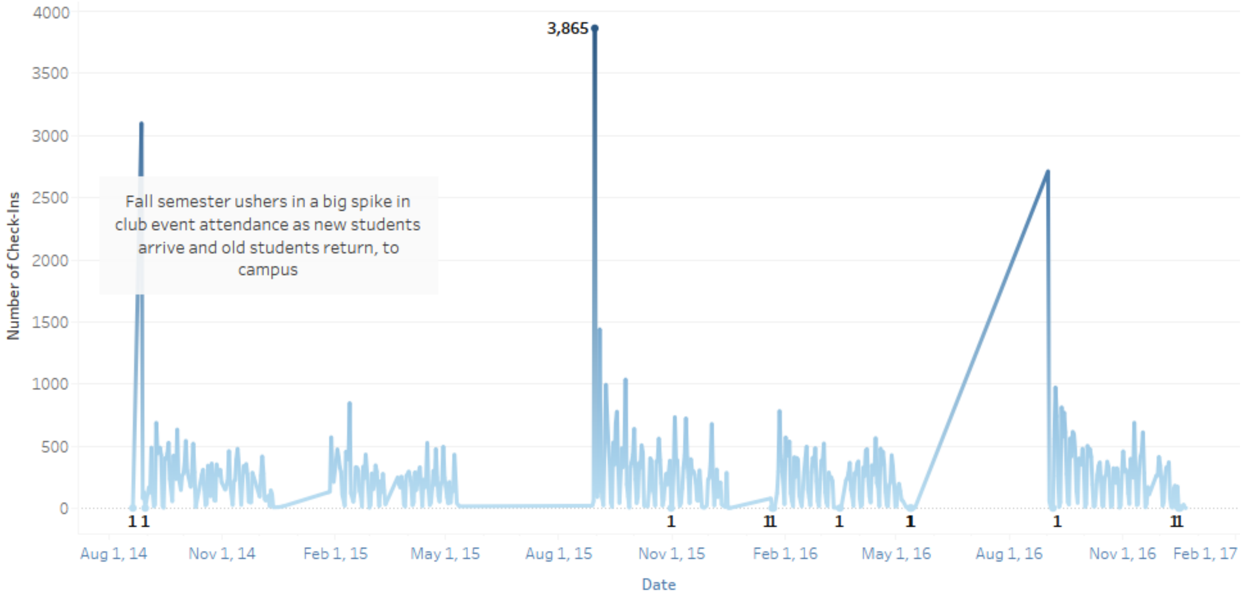
## Acknowledgements

- Vishal Singh (Thesis Adviser) - for supporting me these past few years both in SPUR and in the Honors Program as an insightful, sharp, and encouraging mentor of data, statistics, and life.
- Marti Subramanyam (Faculty Adviser) - for believing in me when I said that I wanted to do a non-traditional thesis that combined my passions in technology and business.
- Dean Ashish Bhatia & Mandy Lancour (OSE) - for supporting my data collection efforts and giving me great feedback as to the impact of my analysis on the Stern community.

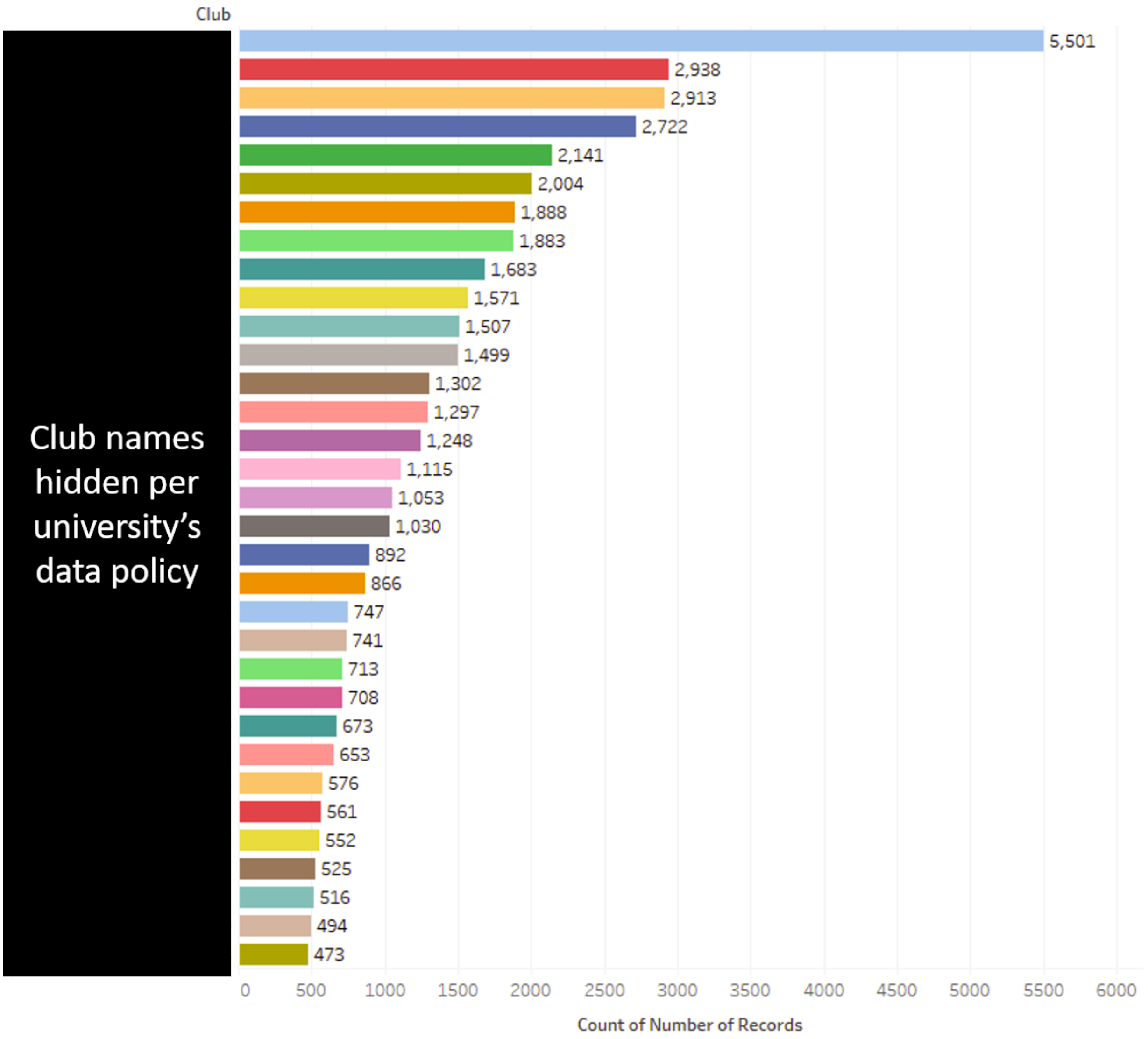
## Background

Every August, wide-eyed freshmen step foot into Stern and are inundated with amazing opportunities to join the vibrant extracurricular scene. From learning career skills, to meeting new friends, to networking into desirable jobs, clubs have so much to offer to Stern students. And indeed, Stern students take advantage of these offerings. Since the start of the 2014 Academic Year, 2,562 unique NYU students have recorded 51,134 check-ins to over 1,602 events, across 211 clubs who've held events at Stern.

Total Check-ins over Time



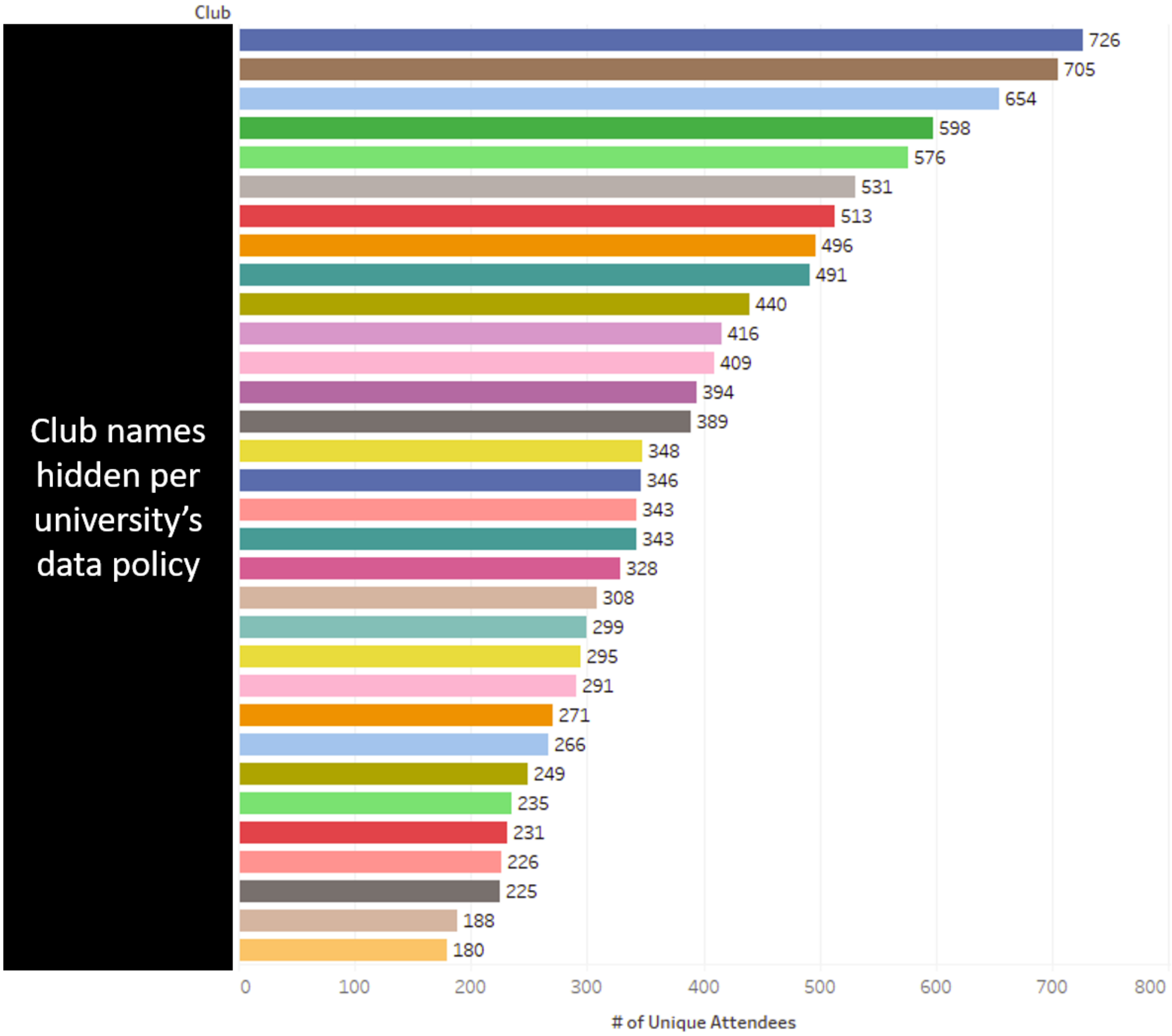
Top 30 Clubs by Total Check-Ins for 2016 Years


[redacted]

Trying to map all this analysis of club activity to a career path - in an ideal world (this is the type of data I would require [measure of intent, randomly assign clubs]), but since not an ideal world, what is the process to approach it to try to proxy it ??? say the problems up front, identify caveats, and say what you can learn from these data sets

In 2016, 179 clubs held events at Stern, most of them being Stern clubs. Of the top 10 clubs with the most check-ins, all of them are Stern clubs. However, this is aggregated across the whole year - so the same student could have checked in multiple times to various events. Therefore, a look into the total number of unique attendees for each club will reveal how big each club's audience really is.

Top 30 Clubs by Unique Attendees for 2016

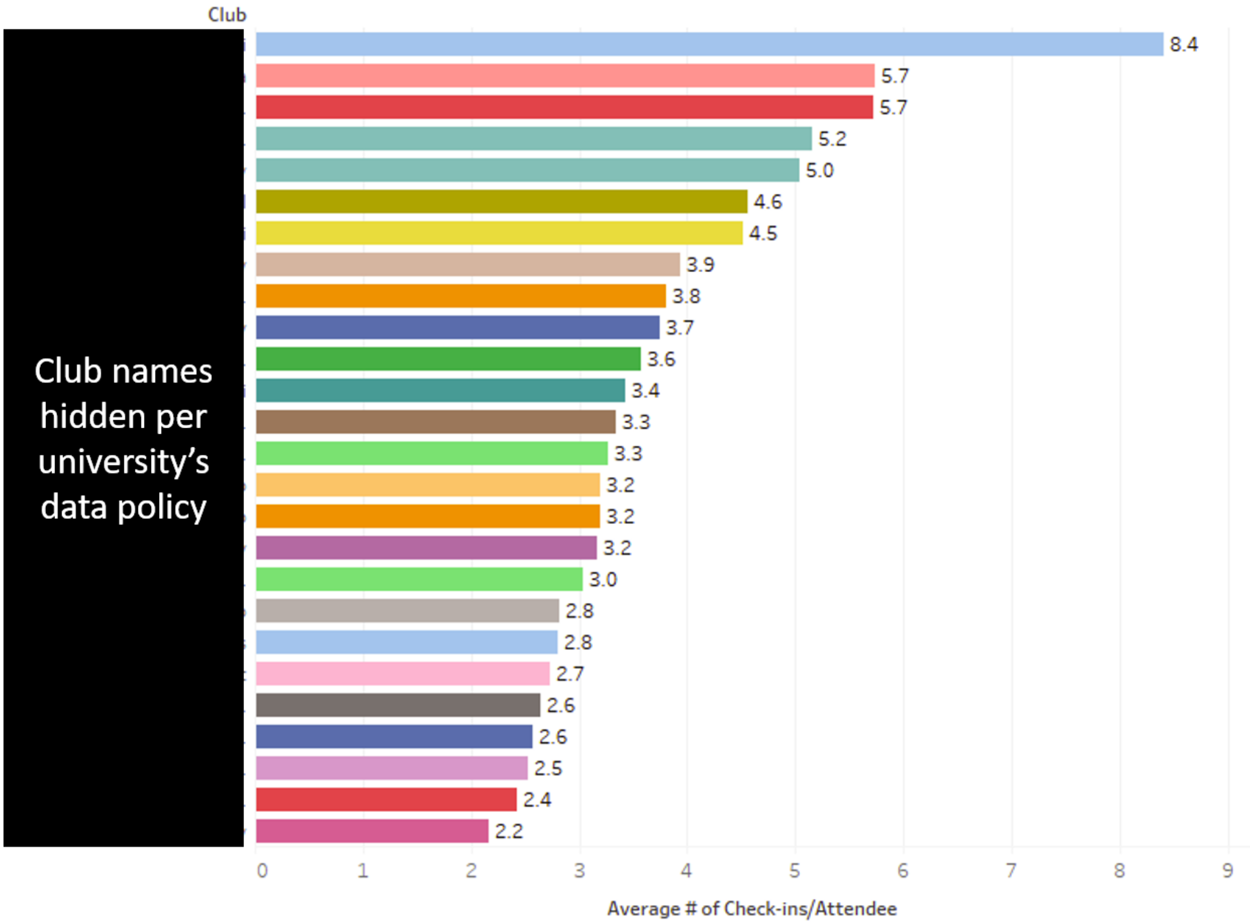

[redacted]

The top clubs by unique attendees differs slightly from that of check-ins. A side-by-side comparison shows that...

| Unique Attendees Ranking | Total Check-ins Ranking |
|---|---|
| 1. Stern Student Council | 1. Beta Alpha Psi |
| 2. Finance Society | 2. Stern Student Council |
| 3. Office of Student Engagement | 3. Investment Analysis Group |
| 4. Beta Alpha Psi | 4. Finance Society |
| 5. USWIB | 5. USWIB |
| 6. Management Consulting Club | 6. Management Consulting Club |
| 7. Business Analytics Club | 7. Inter-Club Council |
| 8. Investment Analysis Group | 8. Delta Sigma Pi |
| 9. Quantitative Finance Society | 9. Alpha Kappa Psi |
| 10. Alpha Kappa Psi | 10. Marketing Society |

Some of these differences are understandable - DSP and AKPsi are fraternities and thus have a capped number of attendees that can possibly attend an event (as some events are exclusive to fraternity members). These figures can be combined to find the Average Number of Check-Ins per Unique Attendee, for each club, to get a sense of the engagement level for each club's members.

Average # of Check-ins/Attendee for Clubs with more than 100 Unique Attendees



One club averaged an impressive 8.4 events per unique attendee. When asked to comment on these figures, an executive board member had this to say: "Our members are very committed. At the beginning of every semester, we tell all of our new candidates, that this organization is what you make of it. The more work you put into it, the more we can help you. At the end of the day, you can do the bare minimum of meeting enough requirements to cross as a member, but we always encourage our candidates to do more and see them go above and beyond!"

There's no doubt that undergraduates at Stern are very involved in their clubs. However, average Stern student only has so much time to dedicate to extracurricular activities. And frankly speaking, most Stern students join Stern clubs for one main reason: to better their career prospects, which can be achieved through either 1) learning valuable skills, or 2) networking with older students. Therefore, Stern students need clarity on this somewhat ubiquitous understanding of whether or not joining clubs helps in job placements, and furthermore, which clubs place better at which firms. It is very possible that the data reveals trends of how certain firms only hire from certain clubs - thus, revealing some internal loyalties between these clubs and firms. In the spirit of transparency and for the sake of helping freshmen and sophomores better allocate their time based on job interest, the following analysis will look into club participation as a measure of eventual post-graduation job placement

## Introduction

To keep the analysis concise and to avoid "noisy" data, the only career paths analyzed in this analysis are those pertaining to financial services, particularly investment banking. This is because investment banking is the most desired career path amongst Stern students. In addition, only pre-professional clubs will be analyzed. This is because it would be unfair to measure pre-professional clubs against clubs that are clearly social in nature, such as various fraternities or even the Stern Student Council. As long as the club has a significant component of its membership activities dedicated to educating students about a specific professional field, then it can be considered a pre-professional organization. Furthermore, my analysis will only focus on the top 11 investment banks from the 2017 Vault Banking 50 Guide (an industry-wide accepted standard for ranking investment banks by performance and prestige). Thus, the scope of the question at hand is condensed into: how does undergraduate pre-professional extracurricular involvement affect post-graduation job placement into the top 11 investment banks, according to Vault?

Note: I only included the Top 11 banks because LinkedIn seemed to have caught onto my scraper's activities while I was in the middle of scraping data for the 11th firm. This done introduce some sampling bias so just be sure to append all the results you see with "for the top 11 investment banks".

## Hypothesis

Our null hypothesis in this experiment says that "graduates going to 'better' firms for investment banking or similar jobs, do not statistically significantly differ in extracurricular participation." "Better" in this case, refers to the 2017 Vault IB Rankings.

## Data Collection Methodology

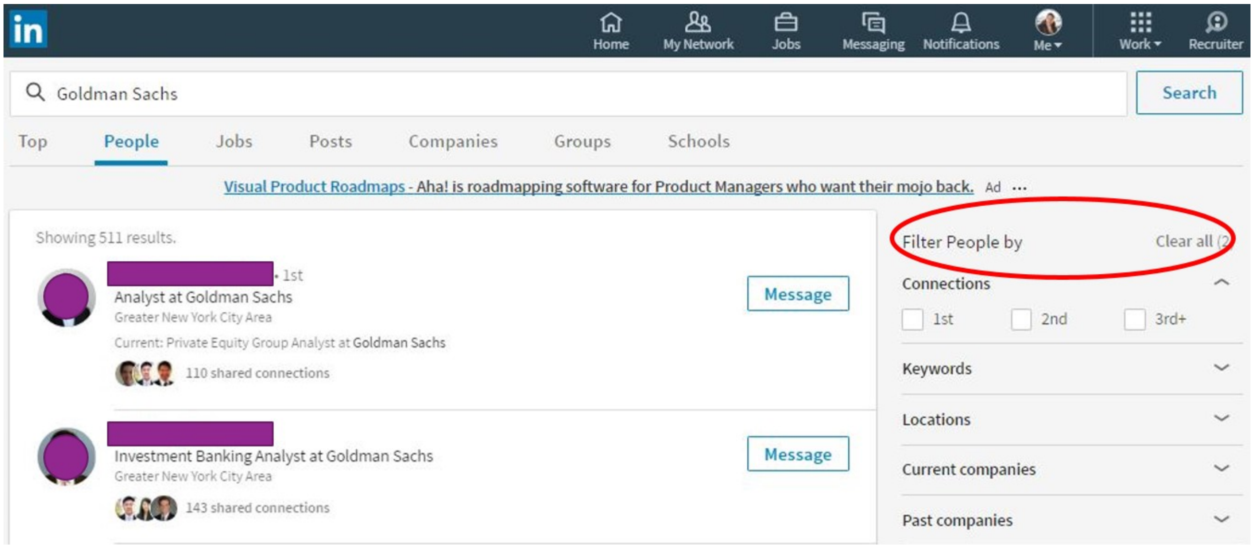
The data collected and cross-validated in two ways: LinkedIn web scrapers and primary research. LinkedIn web scraping will involve building a web scraper tool that scrapes LinkedIn public user profiles for information. Key information to gather will be: first job out of NYU Stern and club involvement (if any). This raw data set will be performed across all graduates of NYU Stern from 2013-2016. This date range was chosen since it is close to the years that I, myself, attended NYU Stern and thus I will have more domain expertise in being able to fill in missing information or detect data anomalies among my raw data set. This additional cross-validation step is very important as LinkedIn varies widely in terms of how people report where they work, what titles they have, and what school activities they were involved in. Primary research gathers the data from the Office of Student Engagement, current club leadership, and my own, manual verification via LinkedIn

Challenges to data collection will not be in the actual collection itself, but actually in understanding the data and making sure it is accurate. For example, on LinkedIn, anyone can claim they are part of a certain organization even when they are not. In addition, people may not list the organizations they are part of. Therefore, sample size is critical for this analysis in reducing outliers and identifying the true trends.

## Building the web scraper

Ironically, the most difficult and challenging part of this thesis is creating the scraper and it will receive almost none of the credit in terms of the end-deliverable. The technical challenge required to subvert LinkedIn's anti-bot detection security is tremendous. Thus, I had to be very careful in making sure that I gathered this information in a way that would avoid detection and not result in me getting my IP address banned. In addition, to protect anonymity of future students and to avoid condoning such behavior, my scraper's specific algorithmic tricks will not be disclosed.

The first step in building the scraper was to collect a list of names of every person that worked at each one of these top 11 firms. This can be done on LinkedIn search with the help of some filtering aspects on the side. For example, I filtered for people who worked at "JP Morgan" and were from "New York University - Leonard N. Stern School of Business". Then I copy and pasted this big list of names into excel, saved it as a CSV, and used some python code to strip out meaningless information until I was left with observations that contained: name, company, and current job title. The problem on the LinkedIn Search Results page was that each person's link did not have their profile URL. In addition, scraping the LSR page would be impossible due to the fact that these search results are only made possible after having created an account with LinkedIn. LinkedIn does not allow scraping period, so they certainly do not have permission fields that allow me to enter their site as a scraping bot despite me having an account with LinkedIn. Thus, this part of the process had to be manual. Now that I had the list of observations, I had to get the actual profile URL so I could visit the profile page and get the desired information.



Goldman Sachs

Visual Product Roadmaps - Aha! is roadmapping software for Product Managers who want their mojo back. Ad ···

Showing 511 results.

[redacted] • 1st
Analyst at Goldman Sachs
Greater New York City Area
Current: Private Equity Group Analyst at Goldman Sachs
110 shared connections

[redacted]
Investment Banking Analyst at Goldman Sachs
Greater New York City Area
143 shared connections

Filter People by Clear all (2

Connections
□ 1st □ 2nd □ 3rd+

Keywords

Locations

Current companies

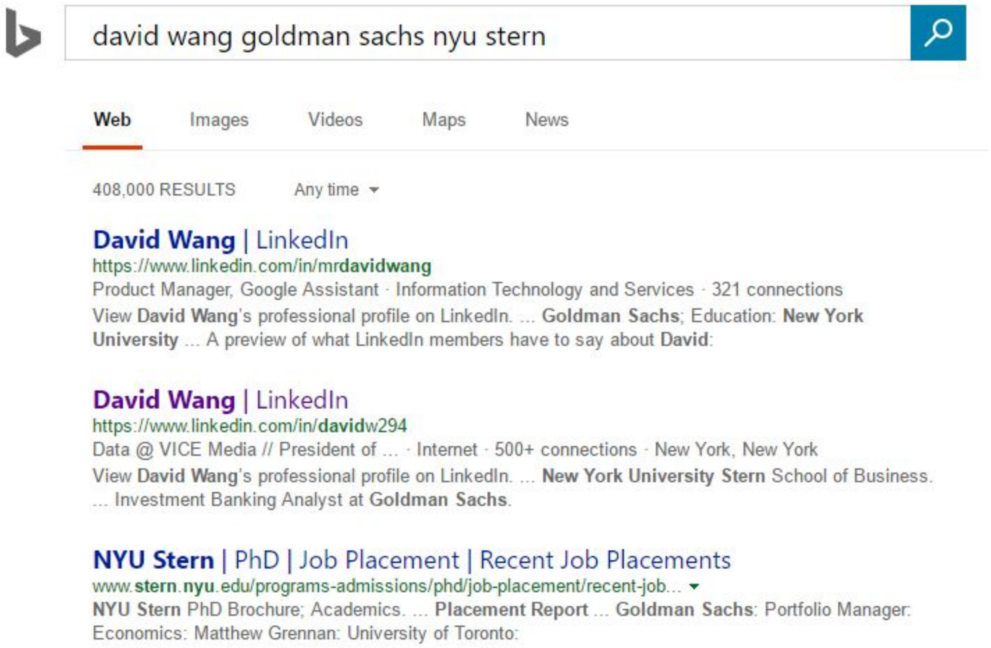
Past companies

The second step was to run each of these observations through a search engine so search results of the person would appear. For example, my script ran in a way that for every observation in the previous list, a search would be conducted for the name, firm, and current job title in the search string. This would hope to narrow down the person in the upcoming search results. When these search results appear, the script can break the page down into raw HTML, and save that information. This is important because each search result is hyperlinked to that person's LinkedIn profile URL. Thus, the true purpose of this step is to get the profile URLs. The search engine used was Bing, as Google is much more strict with automated searches on its platform. The output from this stage was, for each observation, a raw string of HTML on that first page.

```
# Snippet of code:
```

```
name = name.replace(" ", "+")
title = title.replace(" ", "+")
company = "Moelis"

url = "https://www.bing.com/search?q=" + name + "+" + title + "+" + company + "+nyu+stern" + "+linke
```



david wang goldman sachs nyu stern

408,000 RESULTS Any time

**David Wang** | LinkedIn
https://www.linkedin.com/in/mr**davidwang**
Product Manager, Google Assistant · Information Technology and Services · 321 connections
View **David Wang**'s professional profile on LinkedIn. ... **Goldman Sachs**; Education: **New York University** ... A preview of what LinkedIn members have to say about **David**:

**David Wang** | LinkedIn
https://www.linkedin.com/in/**david**w294
Data @ VICE Media // President of ... · Internet · 500+ connections · New York, New York
View **David Wang**'s professional profile on LinkedIn. ... **New York University Stern** School of Business. ... Investment Banking Analyst at **Goldman Sachs**.

**NYU Stern** | PhD | Job Placement | Recent Job Placements
www.**stern**.**nyu**.edu/programs-admissions/phd/job-placement/recent-job...
**NYU Stern** PhD Brochure; Academics. ... **Placement Report** ... **Goldman Sachs**: Portfolio Manager: Economics: Matthew Grennan: University of Toronto:

Note: I don't actually work at Goldman Sachs

The third step was to parse each HTML string to find the LinkedIn profile URL. Luckily, LinkedIn profiles have a very unique expression - they always follow a "https://www.linkedin.com/in/" pattern. Thus, searching for this specific string and stopping just after the username's last character will bring about that person's profile URL. Thus, the output from this stage is a list of profile URLs for all people at NYU Stern who worked at this firm, either in the past or currently.

```
# Snippet of code:

string_v = unicodedata.normalize('NFKD', response.text).encode('ascii','ignore')
m = re.search('https://www.linkedin.com/in/(.+?)"', string_v)
```

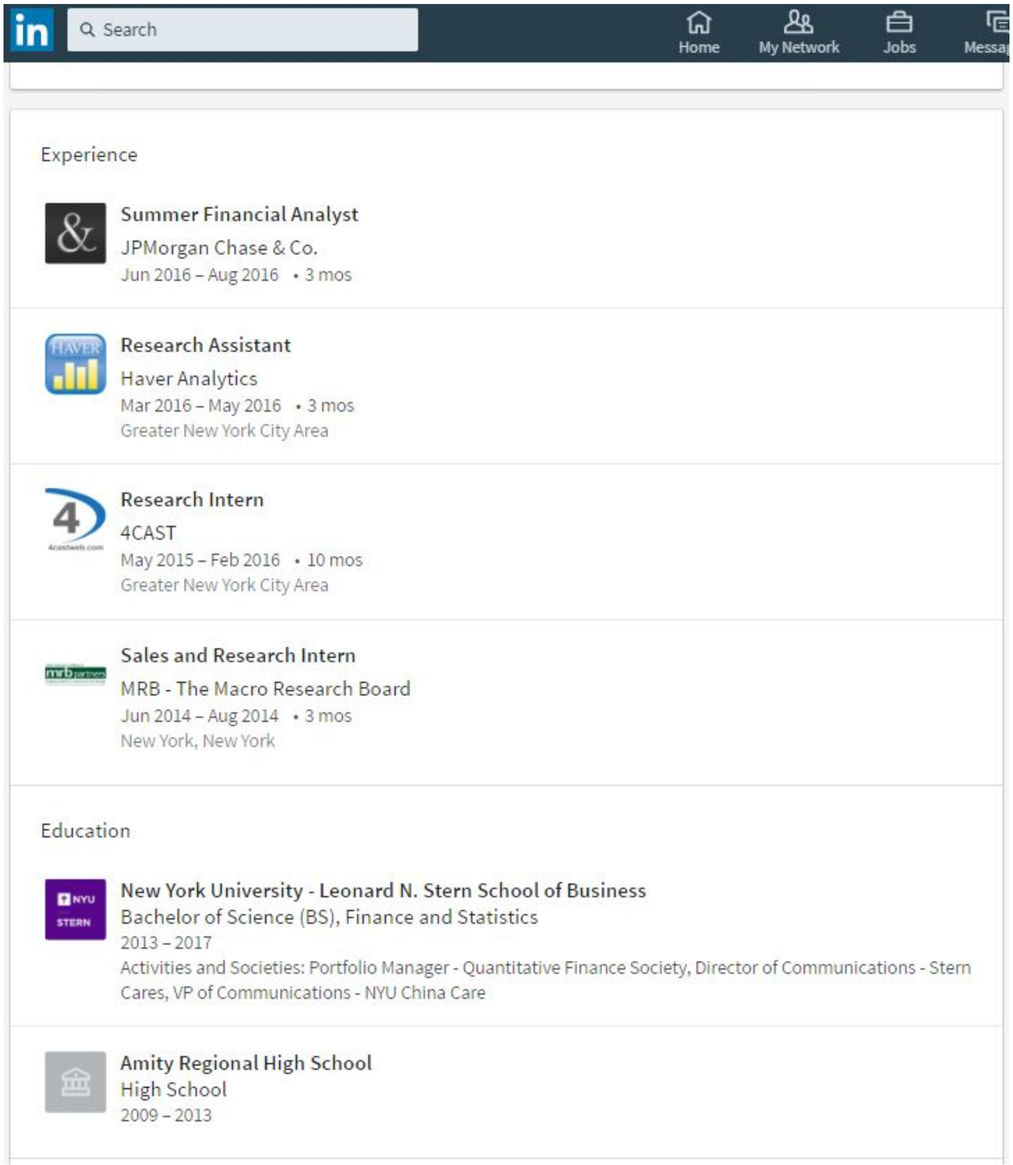
The fourth step was to visit each profile URL and collect the relevant information from our analysis. Since our analysis includes graduates from 2013, it is extremely plausible that they have moved on to another job by 2017. Thus, the scraper must devise an algorithm to find the job that the graduate did immediately upon graduation. The parameters for this search was essentially to search for jobs right after the person's graduation date, which is listed under their Education section. If that information is not present, then the scraper would search for the oldest job without the title "Summer" or "Intern" in the job title. Various trial-error methodologies led me to compile a list of algorithmic iterations that tried to capture every type of profile listing permutation. This step was arguably the most arduous because the data came in such varied formats. Thus, I opted for a "wide-net" approach in which I tried to gather as much as I could to avoid missing information. The output from this step was a raw list of observations that included the graduate's name, post-grad job title, firm, and graduation year.

```
# Snippet of code:

html = driver.page_source # load driver to html
  i = random.uniform(32, 395)
```

```
    time.sleep(i)
    driver.close()
    soup = BeautifulSoup(html) # use BS on html

    previous_raw = soup.findAll('li', { "data-section" : "pastPositionsDetails"})
```



Experience

Summer Financial Analyst
JPMorgan Chase & Co.
Jun 2016 – Aug 2016 • 3 mos

Research Assistant
Haver Analytics
Mar 2016 – May 2016 • 3 mos
Greater New York City Area

Research Intern
4CAST
May 2015 – Feb 2016 • 10 mos
Greater New York City Area

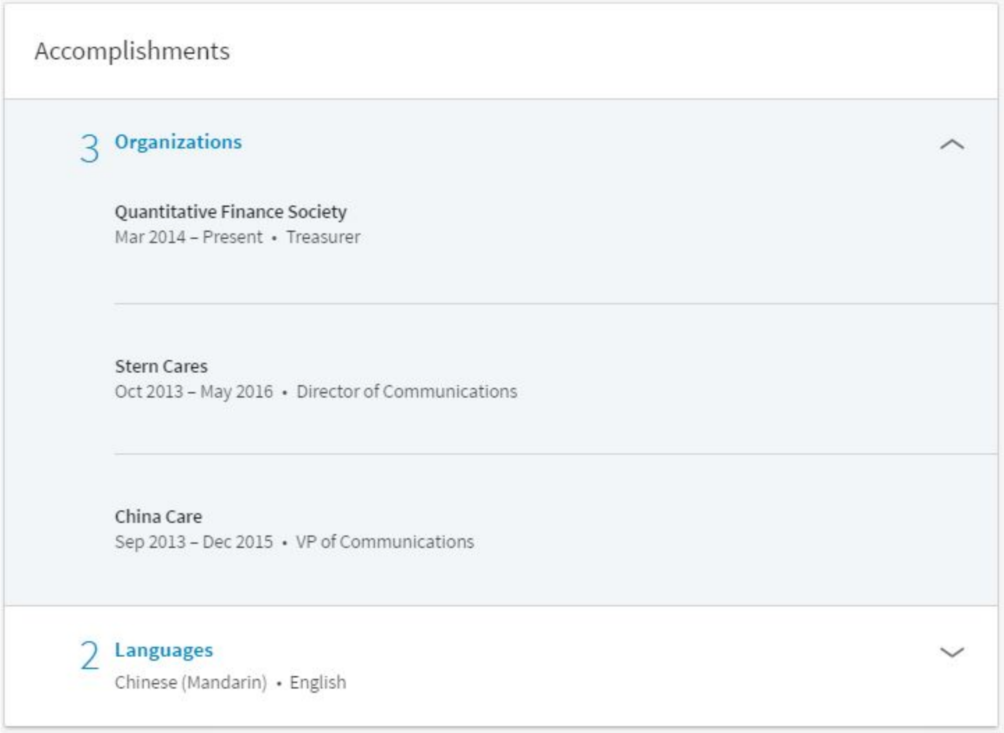
Sales and Research Intern
MRB - The Macro Research Board
Jun 2014 – Aug 2014 • 3 mos
New York, New York

Education

New York University - Leonard N. Stern School of Business
Bachelor of Science (BS), Finance and Statistics
2013 – 2017
Activities and Societies: Portfolio Manager - Quantitative Finance Society, Director of Communications - Stern Cares, VP of Communications - NYU China Care

Amity Regional High School
High School
2009 – 2013

The fifth step was to verify this data and trim it down to remove irrelevant observations. There was no intelligent way to do this, thus I manually went through more than a thousand LinkedIn profiles to verify, correct, and remove observations. In the end, I ended up with around 300 observations of graduates from 2013-2016 at the top 11 investment banks. Given the size of each Stern graduating class and the estimated percentages of how many of these graduates go into these jobs, 300 seemed to be a reasonable number.



Accomplishments

3 Organizations

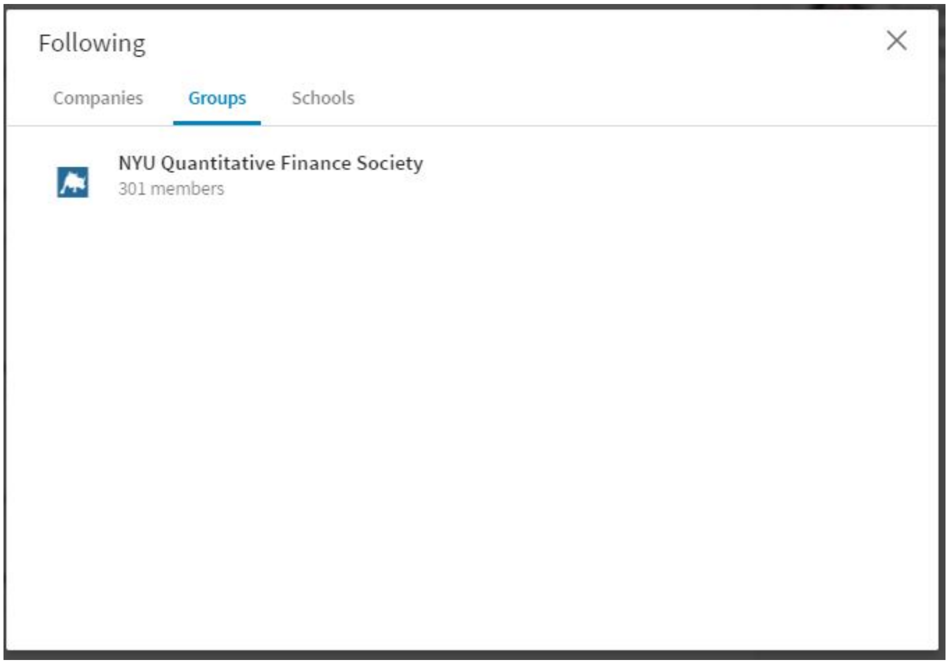
Quantitative Finance Society
Mar 2014 – Present • Treasurer

Stern Cares
Oct 2013 – May 2016 • Director of Communications

China Care
Sep 2013 – Dec 2015 • VP of Communications

2 Languages
Chinese (Mandarin) • English

Sometimes users listed their clubs under the “Organizations section”



NYU Quantitative Finance Society
301 members

…other times they put their clubs under “Groups”

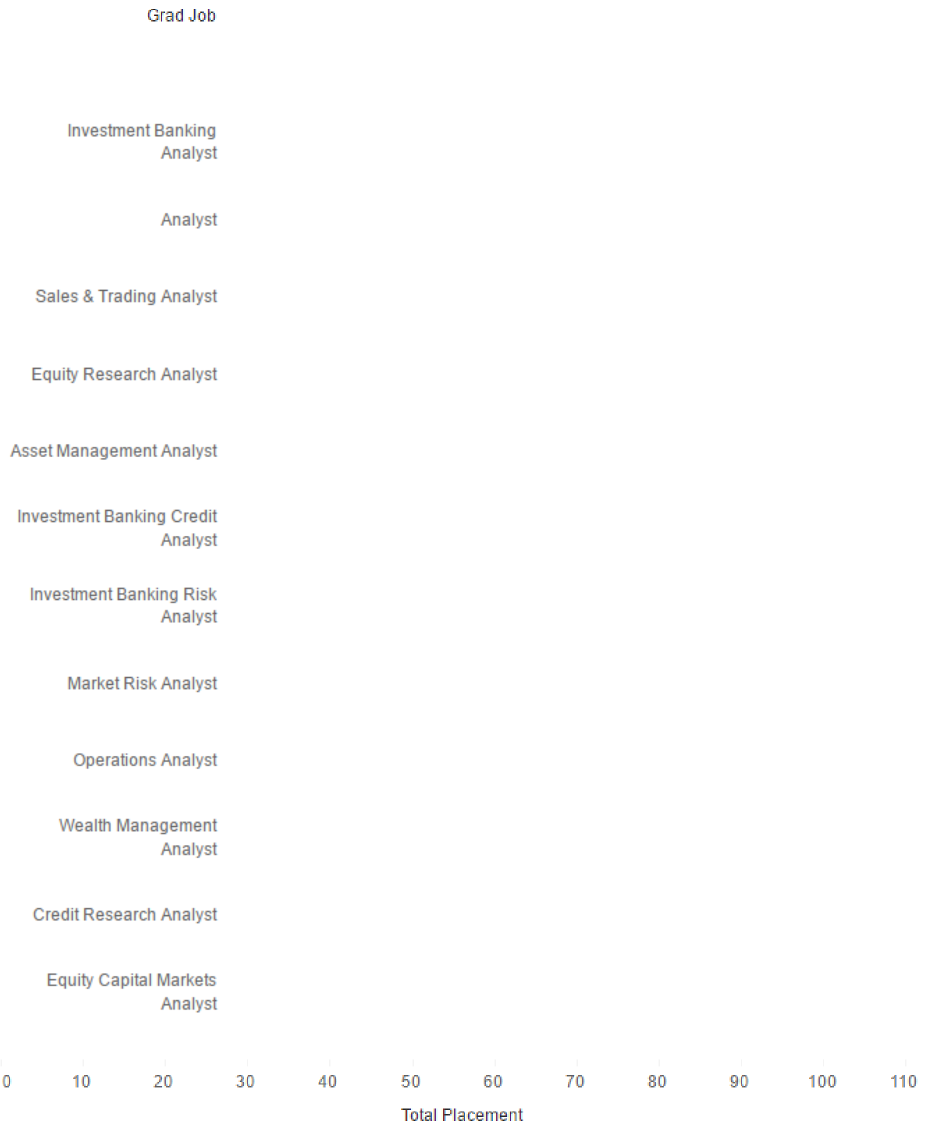
# Visualizations and Tools

## Overview of Charts and Tools

To visualize all this data, a Tableau tool was originally built and presented with the purpose of giving students the power to analyze the data themselves. Unfortunately, conversations with university administration deemed this level of insight potentially harmful to the Stern community. As the creator of such a tool, I would have to agree as well. The negative externalities could be a student thinking “Oh wow…so in order to get a job at ABC firm I must be part of XYZ club.” Therefore, the full tool will only be available to the Office of Student

Engagement and my thesis advisors. I apologize for the inconvenience, but hope you understand that this is for the greater good. Below are some harmless graphs I can speak more to.

Club vs. No Club by Job from 2014-2016



This tool here allows us to see, of all the graduates from 2014-2016 who worked in these jobs at the Top 11 Investment Banks, what percentage were part of clubs? Comparing each value with the fact that from 2014-2016, about 50% of all Sternies attended clubs, we would expect to see 50/50 splits in placement for each job for club vs. no-club. But for jobs like investment banking analyst, there is clearly a higher placement for those who are in clubs. Yet, as the total placement numbers get smaller, the sample becomes less reliable.

This leads me into one of the key shortcomings of the analysis.

## Confounding Variables and Shortcomings

One of the key realizations is that it is impossible to measure the inherent motivation and drive of a particular student. This negates any definitive causality from placement data because it may very well be that those who go to clubs were more motivation and driven to get involved and learn. Thus, the club itself didn't help them obtain a certain job - as they might have gotten this job anyway.

Another shortcoming is that the proper experimental design for this analysis is also unlikely to happen. We would need a completely controlled environment in which the club's executive leadership would stay for 4 years, all incoming freshmen are randomly assigned to a club including one group that is a control group - no club, and after 4 years, we measure these student outcomes. Thus, the best I can say is that my analysis tries to serve more color into this otherwise nebulous and vague territory.

Further complications arise when you consider that data on LinkedIn is not actually verified. People could have been claiming involvement in clubs they were not part of, or even left off clubs and job information for the sake of privacy. The incomplete data set here is worth keeping in mind.

On the club data side, executive leadership of various clubs told me that not every person who attends an event is always accounted for since people arrive late and those with the check-in swipers have either sat down or put away their gear. Thus, the club data should also be viewed in relative proportions on a club to club level.

## Conclusion

Broadly speaking, club involvement at Stern is quite prolific and engaging - around 50% of students get involved every year. Not shown in this analysis, but one of the unlisted graphs (due to data sensitivity issues mentioned above) showed certain clubs having 90%+ placement at certain smaller firms like Evercore and Lazard. Through three years of data, the likelihood of that happening due to chance is low. Thus, there is a big possibility that there is some backroom relationships going on, such as certain clubs having alumni connections at certain firms. This relates to how this analysis may affect some of the policy decisions made by the Office of Student Engagement.

In seeing the analysis of club life and post-grad placement, I can see OSE and club leaders working to strengthen event programming around the most popular type of events. In addition, clubs may be given different budgets according to their level of engagement and type of programming - especially if there proves to be a relationship in the type of programming a club holds (workshop vs. social events) and that club's job placement performance. In addition, OSE can better see which clubs may have special relationships with which certain firms, and work with them to perhaps open up the recruiting process to the broader Stern community so everyone can get a fair shot.

Lastly, I wanted to highlight the fact that the largest group represented by placement into companies was the non-club group. This is most likely due to the fact that there are just more people who are not involved in clubs. However, it does show that there are “non-club success stories” - meaning that it's possible to not be part of a club and still do well, anyway. From my experience and conversations, a lot of underclassmen stress about getting involved in club leadership or selective portfolio teams, because they see it as a critical path to take in order to do well, professionaly. And while the ambition is great, the data is inconclusive in being able to say definitely that joining a club will lead to a better job. Rather, this should serve as an encouraging word to undergrads to optimize on 1) their passion for what the club does and 2) the social nature of clubs. A fulfilling collegiate experience can hopefully come from treating one's extracurricular time as time not only to learn new skills, but for the pursuit of passions, to make new friends, and to have fun.

Thanks for reading!