

UPDATED TIPS DATABASE QUERIES

#	QUERY DESCRIPTION	RELEVANT DATA
1	<p>CLUSTER ANALYSIS (WITH TIME-SERIES) OF INTERNSHIP DATA</p> <p>Group interns in each region (or ZIP code) and record metrics (average performance, gender ratio, age ratio, proportion of interns vs. non-profits in a particular area) for each region. This can inform outreach efforts and help TIPS decide where to allocate resources for acquiring more non-profits or recruiting more intern applicants.</p> <p>Results of this query could be visualized as a heat map showing how various metrics (e.g concentration of interns/non-profits) change over years.</p>	<p>Intern addresses</p> <p>Non-profit addresses</p> <p>Intern evaluations</p> <p>Intern demographics</p> <p>^ track year-to-year</p>
1a	Find the proportion of persons who submitted intern applications grouped by ZIP code and the proportion of nonprofit partners located in each ZIP code.	<p>Applicant addresses</p> <p>Non-profit addresses</p>
1b	Find all non-profit partners with longitude east of (less than) -122.25° and calculate the geographic center of the points as a reference location for a potential Eastside office for TIPS.	<p>Non-profit addresses</p> <p>(figure out how to convert to lat/long)</p>
2	<p>WEIGHTED CRITERIA FOR CANDIDATE SELECTION</p> <p>PRE-SCREENING</p> <p>Query should be able to automatically generate shortlist of best-fit matches for all candidate interns based on a multivariate regression model (or least squares) with vectors weighted for the importance or level of various skills required by the internship and possessed by a candidate, along with a separate set of weights for travel distance (or use k-nearest neighbors).</p>	<p>Skills listed in intern job descriptions and in applications, levels</p> <p>Relative weights / importance of skills</p> <p>Applicant addresses</p> <p>Non-profit addresses</p>
2a	For every applicant / candidate considered by the selection committee, return the list of non-profit partners that are included in both 1) the 15 non-profits located closest to that intern's address and 2) the 15 non-profits whose skills sought have the most matches with the candidate's skills.	<p>Skills listed in intern job descriptions and in applications.</p> <p>Applicant addresses</p> <p>Non-profit addresses</p>
3	<p>PREDICTIVE MODELS FOR DONOR AND EVENT ACTIVITY</p> <p>Apply probabilistic modeling techniques based on past donor history by type with types of donors identified by machine learning clustering. Use groups to develop targeted marketing based on donor characteristics, track donors' activity at events in response to different interactions, forecast donations.</p>	<p>Donor names, info</p> <p>Donation history, amounts, methods</p> <p>Interaction history</p> <p>Event information</p>

3a	Return count and sum of all donation amounts grouped by organization membership (table number), as well as number of donors and number of persons at a table for a fundraiser.	Donor names, info Event information Donation amounts
4	SUGGEST/PREDICT FIT OF ALUMNI MENTORSHIP Based on intern profile (school, interests, internship sector, skills), search for alumni with the closest matches and suggest alumni mentors based on those matches for each intern. Group students by various similarities or students with the same mentor and analyze any potential commonalities. Analyze past mentorships with evaluation surveys and use results to predict effectiveness of future mentorships, updating mentor suggestion model.	Alumni names, skills, education, sector. Intern names, info, skills, school, sector. Mentor evaluations.
4a	Return the first name, last name, email, phone, internship year, and skills of all alumni who possess a certain skill (or interest).	Alumni names, skills Alumni contact info
5	RANK & PROFILE OF OPTIMAL PERFORMANCE FROM EVALUATIONS Rank all internships based on normalized performance data and extract key features that contribute to optimal performance or the profile of an optimal intern relative to non-profit's needs, creating a dual recommender system.	Intern skills/abilities Non-profit needs Evaluation data Internship info
5a	For all internship evaluations where signed difference between mean of ratings received and mean of self-ratings is negative, return the Internship#, intern's name, and non-profit partner name. <i>See next page for notes.</i>	Evaluation data, both received & self-given Internship info
5i	(potential query; high difficulty level and more abstract) Run text-based natural language processing on all intern testimonials and compare with actions in response to text samples sent out to supporters to identify characteristics (or keywords, structures) of most effective / provocative testimonials and select for those characteristics or suggest more effective writing techniques to future interns.	Intern testimonials and qualitative evals

NOTES ON DATA FROM EVALUATIONS:

Suppose that TIPS has an evaluation system that works in the following way so that at least part of the evaluations are quantifiable. (This would not apply to binary yes/no questions.)

For evaluating interns, both interns and nonprofits will be asked the same series of questions where their answers are given on a 1-5 scale, where the intern evaluates themselves and the nonprofit evaluates the intern on the same question/characteristic. The nonprofit evaluation would work the same way (questions about how well the nonprofit was able to facilitate the internship and how well they were able to support and meet the intern's needs or expectations from both perspectives, etc.).

In order for evaluations to be useful actionable, TIPS would distribute the same set of questions during both the midterm and final evaluations (more time-sensitive questions could be added to the final) to be able to compare evaluations during the course of the internship.

With this evaluation system, we can interpret various statistics on evaluation data as follows:

Cumulative sum of ratings from all questions: the higher the sum, the better the ratings given by all parties for all four types of evaluations (intern self, nonprofit-intern, nonprofit self, intern-nonprofit). The "most successful" internship is the one with the highest cumulative sum.

Mean ratings received by intern (or non-profit): a normalized performance index between 1 and 5 for the party receiving the rating, which we assume to be less egocentrically biased. If the 1 to 5 scale is preserved between years, performance indices would be comparable even if cumulative sums are not because the content or number of questions changes.

Total variation distance between nonprofit's and intern's evaluations for intern (or nonprofit) performance: the lower the [total variation distance](#), the more similar the intern's and the nonprofit's ratings of intern (or non profit) performance.

Signed difference between mean of ratings received and mean of self-ratings for intern (or nonprofit):

If the difference is **positive**, the receiver underrated themselves on average relative to the other party's ratings.

If the difference is **negative**, the receiver overrated themselves on average relative to the other party's ratings.

The latter case is more concerning (the former case is relatively harmless and could be a sign of modesty), and the preservation of sign allows us to account for this directionality in identifying problematic or divergent evaluations to flag and follow up on.