Harvard
College
Data
Analytics

Harvard Undergraduate Data Analytics Group

Harvard College Data Analytics Group (HDAG) is a non-profit student organization at Harvard dedicated to helping organizations make smarter and more data-driven decisions. We assist companies in achieving their strategic goals by translating their data into meaningful and actionable information. We aim to pair teams of well-trained, highly motivated Harvard students with our partners, specifically focusing associates and analysts in industries where they have experience or interest, in order to produce the highest quality of work possible. From data collection to strategy implementation, we want to be there every step of the way to help organizations make data their new superpower.

We competitively recruit undergraduate students at Harvard with demonstrated competence, dedication, and problem-solving skills, many of whom have prior experience working in top management consulting or data science teams. All our team leaders have experience working in or leading data science teams at Fortune 500 companies, and our board of technical advisors include members of the Harvard faculty. Each team, composed of around six Harvard students, commits over 600 hours to a case over the course of a 10-12 week span.

We enjoy different challenges and work with a diverse set of organizations and problems. Our clients range from local businesses to Fortune 500 companies to international non-profits. Using our capabilities in visualization, machine learning, and predictive analytics, among others, we help organizations diagnose problems and identify strategies across their sales, marketing, financial or operational functions. Client confidentiality is our utmost priority.

Harvard College Data Analytics

Team Capabilities

1. Data Analytics Consulting: deriving valuable insights from data

 a. Case study 1 - Providing IT resource management analytics for a multinational Fortune 500 company in energy and automation: Through statistical analysis of over 100k anonymized employees, we identified help desk call volume and demographic trends to help inform executive decisions on employee satisfaction and IT resource allocation.

 b. Case study 2 - Providing data processing service for a Wall Street fintech company: Through scraping the Securities and Exchange Commission (SEC) website and extracting relevant data en masse, we created well-formatted databases to advance the client's core digital offerings.

2. Machine Learning Algorithms: training and deploying predictive models

 a. Case study 1 - Providing IT security service for a multinational Fortune 500 company in energy and automation: By building ML models, we enabled predictive analytics for the company's future spending on Indirect Procurements and introduced data integrity improvement design to the purchase request process.

 b. Case study 2 - Providing AI algorithm advancements for a leading sports analytics company: Using "Big 5" European club leagues' pre-game and in-game data, we created models that predict win, loss, and draw probability and provided an evaluation of the accuracy and probability calibration of the models.

3. Business Intelligence Visualizations: creating interactive visual dashboards

 a. Case study: Providing visualization services for the World Health Organization Region for the Americas: We developed a web app to visualize models on COVID-19 outbreak to predict rate of transmission and epidemic curves; product delivered to WHO country offices in Latin America for projections of varying health intervention measures.

4. Whole-Set Solutions: providing comprehensive digitalization systems

a. Case study: Creating an HR and user management system for an educational foundation in China: We developed a system from scratch to help the management team keep track of employee's progress and KPI and to help employees better manage student feedback.

Harvard College Data Analytics

Proposal for Cambridge Volunteer Clearinghouse:

Background
The COVID pandemic and subsequent fallout have resulted in a dramatic reduction in the number of individuals volunteering, especially for long-term, in-person work. This crisis in volunteer recruitment and retention is essentially a crisis in service delivery for nonprofit agencies that rely on volunteers to execute programs. However, most information about this crisis is only anecdotal.

Project Description
The goal of this project will be to measure and quantify the decline in volunteerism in the Cambridge community, and more broadly in other communities. The work will help nonprofits take informed action to counter the trend by adjusting program structure or offering volunteer roles to accommodate permanent changes in volunteer expectations and needs. The HDAG team will begin by designing surveys to investigate volunteerism decline in the local Cambridge community, leveraging statistical inference techniques for proper survey design and implementing qualitative analysis on responses. The websites of other volunteer agencies will be scraped to capture a more granular view of the volunteerism landscape in the Boston community.

The team will then consider more broadly the trends in volunteerism across the US by analyzing large-scale data from AmeriCorps, an independent government agency supporting national service and volunteerism. Longitudinal modeling and trend analysis will be techniques particularly relevant to this endeavor. The team will in particular seek to characterize and quantify the overall reduction in volunteerism due to the pandemic.

Desired Deliverables
1. Survey to explore volunteerism decline in the local Cambridge community. If data are collected in time, analysis of the raw survey data can be delivered. A final written report will be delivered summarizing insights collected.
2. Datasets compiled from the scraping of other volunteer agencies.
3. Results from longitudinal modeling and trend analysis of the landscape of volunteerism across the US (AmeriCorps data).
4. The code (Python or R) that was used to develop the analysis and produce the results.

Rough Engagement Timeline

Dates	Week	Tentative Schedule
2.6-2.19	0	Each HDAG Case Team Leader (CTL) will have a call with the respective Client liaison to better understand work expectations and align goals for this semester (in terms of research questions, final format of deliverables, etc.) After the meeting, CTL will consult with the 1-2 associates of the HDAG case team and map out the weekly work plan for the semester: from both the perspective of technical execution and business analysis.
2.20-2.26	1	CTL will introduce the project and the work plan to the rest of the case team and start delegating tasks to each individual. (In each team we have data scientists who are proficient in Python, R, SQL, and other analytical tools as well as business analysts who have experience working in industry).
2.27-3.5	2	Every member of each Client Case Team will follow the work plan, initially beginning with survey design and distribution. A code framework will be laid out for brief analysis of prospective data collected. Every week, each CTL will update the Client liaison on the progress that the case team has made over the past week. There is also a weekly meeting between the case team where each member will discuss their work with the others, and the CTL will delegate work for next week.
3.6-3.12	3	

3.13-3.19	4	The team will begin explore the viability of scraping other volunteer agencies' websites to capture a more granular view of the volunteerism landscape in the Boston community. This plan will be assessed and continually updated based on weekly client feedback.
3.20-3.26	5	
3.27-4.2	6	Midway presentations with Client: each whole team will present their findings and recommendations from the first half of the semester to the Client team, in particular the local survey and results from web scraping. Each HDAG case team will follow up with any questions the Client team might have during or after the presentation.
4.3-4.9	7	After the midway presentations, each CTL will integrate comments or suggestions from the Client team to the work plan. Each CTL will list out the remaining questions or technical tasks for the latter half of the semester and delegate them to each individual of the case team.
4.10-4.16	8	
4.17-4.23	9	
		Based on the feedback from the midpoint presentation, the team will address any gaps in survey collection. The team will also consider more broadly the trends in volunteerism across the US by analyzing large-scale data from AmeriCorps. Longitudinal modeling will be implemented as the team characterizes and quantifies the overall reduction in volunteerism.
4.24-4.30	10	The case team will summarize their work for the entire semester and give a final presentation to Client. This will include both technical deliverables (e.g. code repository, curated data sets) and the business presentation (e.g. protocol ordering and recommendations). The HDAG team will follow up with any questions the Client business team might have during or after the presentation.
5.1 - 5.7	11	

5.8-5.22	Post-Project	The HDAG team will follow up with Client on the implementation of suggestions and deployment of analytical tools. We will ask for feedback on their work for the Spring of 2023.

Pricing

- Engagement Timeline: 12 weeks, February – May 2023
- Semester Case Fee: **$0**