Harvard
College
Data
Analytics

Harvard Undergraduate Data Analytics Group

PREPARED FOR

Bluebonnet Data

PREPARED DATE

Feb 20, 2023

ENGAGEMENT TIMEFRAME

Feb - May 2023

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 seven to eight Harvard students, commits over 600 hours to a case over the course of a 10 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 Bluebonnet:

The goal of the project is to expand and optimize Bluebonnet's current algorithm tasked with matching teams of data and technology volunteers with clients. Specifically, the project will aim to incorporate team composition-based matching with the existing ranked preference matching. This will eliminate or reduce the time spent for manually optimizing matches generated by the current algorithm.

Exploratory data analysis will be performed on historical data of fellows' rankings, Bluebonnet's generated client preference scores, and the current matching algorithm to understand the structure of the data and guide algorithm development. In conjunction with consultation on Bluebonnet's initial algorithm attempts and ideas, research will be conducted on matching techniques applicable to the project. Algorithm improvements will be built and tested, guided by prioritized traits to match on and feedback from Bluebonnet.

Datasets & Code

- **Historical volunteer profile data**, including client rankings, background, and client preference scores
- **Algorithm:** Existing code implementation of current matching algorithm

Deliverables

- Optimized algorithm candidate(s) to match volunteers to teams and clients based on additional data. Algorithm and matching behavior honed according to provided criteria for optimal matching.
    - Algorithm code development base
    - Written explanation of algorithm to allow for replication
    - Initial candidates hand-off by 4/20/2023 for client testing
- Presentation summarizing results of EDA, algorithm development and function, and relevant visualizations to support key results

Harvard College Data Analytics

Rough Engagement Timeline

Dates	Week	Tentative Schedule
2.6-2.19	0	The 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 1-2 associates of the HDAG case team and map out the weekly work plan for the semester.
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. (On 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).
2.24-3.1	2	Every member of the Client Case Team will follow the work plan. This encompasses familiarization with the data, background research on matching algorithms, and exploratory algorithm development. Every week, the 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.
3.2-3.8	3	
3.9-3.15	4	
3.16-3.22	5	Wrap up the work for the first half of semester, and prepare for the midway presentation to Client which can include both a technical product (algorithm) and a business

		presentation (slides).
3.23-3.29	6	Midway presentations with Client: the whole team will present their findings and recommendations from the first half of the semester to the Client team. The HDAG case team will follow up with any questions the Client team might have during or after the presentation.
3.30-4.5	7	After the midway presentations, the CTL will integrate comments or suggestions from the Client team to the work plan. The 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. The CTL will hand off the optimized algorithm candidate(s) by April 20 for client testing.
4.6-4.12	8	
4.13-4.19	9	
4.20-4.26	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 and the business presentation. HDAG team will follow up with any questions the Client business team might have during or after the presentation.
4.26-6.1	Post-Project	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: Waived - Pro bono