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.

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.

Proposal for FemHealth Insights:

It is suggested that distributed ledger technology (DLT) can help increase the transparency and trustworthiness of the carbon markets. Broadly, the project HDAG will engage with South Pole would consist of two stages.

1) The HDAG team will first explore the potential for the application of DLT to the carbon market. In addition to providing recommendations relating to quality control and other measures that must be taken for proper implementation, the team will explore several DLT protocols and provide a recommendation for one which best fits the requirements outlined. Some protocols to first consider include: CHIA, TOUCAN / Polygon, NEAR, Ethereum, Binance chain, and EWC.

2) Using the chosen protocol, the HDAG team will develop a proof of concept for one or more project types (forestry, agriculture or renewable energy) focussing on the origination, verification, and potential fractionalisation of the asset. The HDAG team will leverage data previously collected for certifying historic projects to mimic the origination process.

At the end of the engagement, the HDAG team will deliver the following:

1) Slide decks detailing the value of DLT in this domain. South Pole would like to use these slide decks for internal educational purposes.
2) Ordered list of DLT protocols ranked on their fit with the needs of our various project types.
3) Proof of concept model that can be used to demonstrate increased transparency on the quality of carbon assets (forests, renewables, etc).

Rough Engagement Timeline

Dates	Week	Tentative Schedule
9.5-9.18	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.
9.19-9.25	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 JavaScript, D3.js, and other analytical tools as well as business analysts who have experience working in industry). Team members will begin by studying industry examples and client interviews. At the end of Week 2, a slide deck will be presented detailing the planned visualizations and analysis.
9.26-10.2	2	
10.3-10.9	3	Every member of each Client Case Team will follow the work plan, initially starting with familiarizing themselves with the client data and identifying any additional JS libraries to use. Team members will develop the code for the chosen visualization and insight tools. 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.
10.10-10.16	4	
10.17-10.23	5	

10.24-10.30	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. Each HDAG case team will follow up with any questions the Client team might have during or after the presentation.
10.31-11.6	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.

The HDAG team will work with the FemTech UX team to integrate the visualization code into the website. If time allows, the team will begin exploration into automated data scraping and/or applications of ML/AI. |
11.7-11.13	8	
11.14-11.20	9	
11.21-11.27	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 such as the code repository and sample visualizations. The HDAG team will follow up with any questions the Client business team might have during or after the presentation.
11.28 - 12.4	11	
12.5-12.11	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, September - December, 2022
- Semester Case Fee: $0 (pro bono)