Harvard
College
Data
Analytics

Harvard Undergraduate Data Analytics Group

PREPARED FOR

Octane OC

PREPARED DATE

Aug 28, 2023

ENGAGEMENT TIMEFRAME

Sept-Dec 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 on associates and analysts in industries where they have experience or interest, in producing 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 includes members of the Harvard faculty. Each team, composed of around seven to eight Harvard students, commits over 600 hours to a case over 10 weeks.

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 of COVID-19 outbreak to predict the rate of transmission and epidemic curves; the product was 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 employees’ progress and KPIs and to help employees better manage student feedback.

Harvard College Data Analytics

Proposal for Octane OC:

Summary:
HDAG will work on a case for the Octane OC Haystax investment intelligence team. Octane is building the SoCal of tomorrow by creating 55k jobs in the area by 2030. So far, their LaunchPad incubator has assisted 1,500 companies in raising over 5 billion dollars. HDAG will work to improve an ML algorithm that predicts startup fundability (currently with 90% accuracy) and build a variety of other data visualizations, LLM, and sentiment analysis models.

Dataset:
OctaneOC has collected data in the form of a 40-point scorecard/rubric for each company that has passed through the incubator. This includes a specific quantitative assessment of the company in a variety of categories by several judges (including a Chat GPT 13th judge). They also have data about the success of the startups and the funding they have received. Currently, data is accessible in a company dashboard with an array of related visualizations and ML-based predictions presented.

Tasks:
HDAG could work in a number of areas:

1. HDAG will work to improve upon the “Moneyball” inspired startup evaluation algorithm. The algorithm currently predicts a startup’s probability of receiving capital with 90% accuracy. HDAG will leverage new model types, enhanced data cleaning methods, and a larger dataset to improve upon the model.
2. HDAG will work on creating interactive visualizations, dashboards, or other predictive algorithms related to the prescreening tool, word clouds, coachability index, and AI PitchDeck evaluation platform.
3. Finally, HDAG will work with NLP and LLMs, improving the GPT-powered 13th judge or implementing a judge sentiment analysis algorithm.

Main Contacts:
Dave Beck, Creator and General Manager of Haystax

Engagement Timeline

Dates	Week	Tentative Schedule

9/18-9/22	0	Mohan and Dave will be in contact to finalize work expectations, and project details, and obtain the dataset
9/25-9/29	1	The HDAG Case Team Leader (CTL) will have a call with the Octane team to meet the team, 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 map out the weekly work plan for the semester: from both the perspective of technical execution and business analysis.
10/2-10/6	2	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 the industry).
10/9-10/13	3	Every member of the case team will follow the work plan, and start the data analytics, which includes every aspect of the data pipeline: data transferring, cleaning, exploration, modeling, visualization, etc. 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.
10/16-10/20	4	
10/23-10/27	5	
10/30-11/3	6	Wrap up the work for the first half of the semester, and prepare for the midway presentation to Octane which can include a technical product (algorithm, statistical model, web app) and a business presentation (slides).

11/6-11/10	7	Midway presentations with Octane: Each 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 Octane team might have during or after the presentation.
11/13-11/17	8	After the midway presentations, the CTL will integrate comments or suggestions from the Client team into 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 on the case team.
11/20-11/24	9	
11/27-12/1	10	
12/4-12/8	11	The case team will summarize their work for the entire semester and give a final presentation to the 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.
12/11+	Post-Project	HDAG team will follow up with Octane on the implementation of suggestions and deployment of analytical tools. We will ask for feedback on their work for the Spring of 2024.

Pricing

- Engagement Timeline: 12 weeks, September - December, 2023
- Semester Case Fee: $0, Pro bono