Harvard College Data Analytics

Harvard Undergraduate Data Analytics Group

PREPARED FOR

Thermo Fisher

PREPARED DATE

Aug 18, 2022

ENGAGEMENT TIMEFRAME

Sep - Dec 2022

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 Thermo Fisher:

Problem: It's currently difficult to discern our competitor's strategy by analyzing their traditional communication outlets (e.g., financial reports, statements, public announcements).

Goal: Get a better understanding of our competitor's strategy by using data analytics to investigate their digital footprint.

Key Questions

The following questions should be answered in the context of understanding how they advance the competitor's strategy:

Question	Example types of data
What roles are our competitors hiring?	LinkedIn, job boards, company career page
How has our competitor's product portfolio changed over time?	Looking at the product catalog numbers listed on their site over time using Wayback machine. Potential starting point: https://github.com/sangaline/wayback-machine-scraper
What topics and language are our competitors using and how has it changed over time? (i.e., leveraging NLP, topic modeling, and Wayback machine)?	Looking at the full text of the competitor and looking for topics like "automation", "sustainability", "biotech customers", "supply chain" etc.
How are our competitors spending their advertising dollars? What channels are they using (e.g., Facebook or, Twitter, Google, TikTok, etc.)?	To be determined pending feasibility analysis
What other data signals can we pull from competitor data footprints to help understand their strategy?	To be determined pending feasibility analysis

Key deliverables

- Final presentation answering for each of the key questions, quantitatively where possible
- Synthesis of key findings demonstrating how competitor actions relate to their strategy
- Annotated / well-documented source code

Harvard
College
Data
Analytics

Additional Details

The goal of this case is to establish a connection between digital footprint and competitor strategy. Web scraping will be leveraged to collect company "KPI metrics", as defined in the Key Questions above, and Thermo Fisher ultimately seeks to understand how each of the indicators considered correlate with financial performance or overall company strategy.

As such, this engagement has three phases:

1) **Scraping for feasibility.** While the Key Questions outlined serve as a starting point, once web scraping begins, it may become apparent that certain types of data are more feasible to collect than others. Thus, data will first be scraped from a large pool of competitor companies with the goal of understanding the types of data that are available to consider.

2) **Specification and deeper data collection.** Once it is understood which data types are available to be collected and which companies' digital footprints are more apparent, a collection of key competitor companies will be selected for a deeper dive. Any further data (e.g., longitudinal data) will be scraped and organized for deeper analysis. Factor analyses can subsequently be conducted to determine which features are correlated with company success or give indication for the introduction of new company products.

3) **Translating digital footprint to strategy.** Using the data collected and insights generated, the HDAG team will produce a quantitative and qualitative report on the digital footprint and perceived strategy of the key competitor companies. For example, what are the distinguishing features of each competitor's strategy/mission, and what steps are they taking to realize this end.

Harvard College Data Analytics

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 Python, R, SQL, and other analytical tools as well as business analysts who have experience working in industry).
9.26-10.2	2	Every member of each Client Case Team will follow the work plan, initially starting with understanding existing company KPIs, including but not limited to product launches, talent acquisition from career pages, and ad spending trends. Starting with the outlined Key Questions, web scraping for feasibility will be carried out to understand the types of data that are available to consider related to competitor strategy. 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.
10.3-10.9	3	
10.10-10.16	4	Once it is understood which data types are available to be collected and which companies' digital footprints are more apparent, a collection of key competitor companies will be
10.17-10.23	5	

		selected for a deeper dive. The HDAG team will perform a deeper web scrape of any additional suggested features (e.g., longitudinal data) based on weekly client feedback.
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. Using the data collected, the HDAG team will produce a quantitative and qualitative report on the digital footprint and perceived strategy of the key competitor companies. For example, what are the distinguishing features of each competitor's strategy/mission, and what steps are they taking to realize this end. Factor analyses can be conducted to determine which features are correlated with company success or overall strategy.
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 (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.
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 Fall of 2022.

Harvard
College
Data
Analytics

Pricing

- Engagement Timeline: 12 weeks, September - December 2022
- Semester Case Fee: **$30,000**