



Harvard Undergraduate Data Analytics Group

PREPARED FOR

Roche

PREPARED DATE

Feb 13, 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 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 Roche:

Background:

Roche conducts a plethora of studies and administers several treatments to patients across the world. They would like the HDAG team to develop a Patient and Society Benefit Metrics (PBSM), leveraging in particular the initial work conducted by the PBSI team (Patient Benefit and Societal Impact). The PBSI team has created an informative framework, defining key “domains” where benefits can be understood from a “Patient” and/or “Society” perspective. However, these are not quantified or measurable, and cannot be used for insights or comparison amongst Roche’s portfolio of studies.

Goal:

The team will work to develop a concise set of the most informative metrics on Patient and Society benefit and validate it both internally and externally. This will ultimately inform study design decision making through insights about the benefits that specific study designs have upon patients and society.

This work will contribute towards the development of trustworthy scientific evidence that will extend the benefit of therapies, services, and integrated solutions to broaden access to more patients and accelerate development. Aligned to a 1–3-year strategic horizon, Roche would like to accelerate the development and validation of PBSM for ultimate adoption and operationalization at the Disease Area (DA) level.

Datasets:

- PBSI Framework with “domains”
- *Based on the metrics designed by the team*, dummy data can be generated for any data that is unavailable, and real data can be sourced for any data that is available.

Deliverables:

Phase 1: Literature Review

- Conduct a literature review of the PBSI “domains” to determine valid metrics for each domain
- Where no valid metric exists, conduct literature research to determine/design potential new metrics that might cover the domain
- **Core deliverable:** Literature review covering metrics and theoretical approach to data model(s)

Phase 2: Conceptual Data Model

- Design a series of metrics based on the literature review
- Design a “composite” metric to aggregate value across all PBSI domains

- Develop a data model concept for data collection, processing, and analysis - to enable the analysis and comparison of Patient and Society benefits across studies/projects
- Validate the data model and metrics internally
- **Core deliverable:** Mockup of data model(s) and output(s), within initial validation

Phase 3: Physical Data Model

- Validate the physical data model internally
- Validate the physical model and insights externally
- **Core deliverable:** Implement an app-like prototype (R Shiny, React, or similar), with output validation

Rough Engagement Timeline

Dates	Week	Tentative Schedule
2.6-2.19	0	<p>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.)</p> <p>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.</p>
2.20-2.26	1	<p>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). The HDAG team will begin their literature review of the PSBI “domains”, exploring valid metrics for each domain.</p>
2.27-3.5	2	<p>Every member of each Client Case Team will follow the work</p>

3.6-3.12	3	<p>plan, continuing with the PSBI domain literature review.</p> <p>In particular, where no valid metric exists, the team will conduct literature research to determine or design potential new metrics that might cover the domain.</p> <p>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.</p>
3.13-3.19	4	<p>Once literature review insights have been collected, a literature summary report and an initial overview of theoretical approached to data model(s) will be delivered to the client. The team will subsequently transition to the Conceptual Model; the team will design a series of metrics based on the literature review and provide a “composite” metric to aggregate value across all PBSI domains.</p> <p>The team will develop an initial data model concept for data collection, processing, and analysis to enable the analysis and comparison of Patient and Society benefits across studies or projects.</p>
3.20-3.26	5	
3.27-4.2	6	<p>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 literature review report on PSBI metric domains and the initial data model concept to enable the analysis and comparison of Patient and Society benefit across all PBSI domains. Each HDAG case team will follow up with any questions the Client team might have during or after the presentation.</p>
4.3-4.9	7	<p>After the midway presentations, each CTL will integrate</p>

4.10-4.16	8	<p>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.</p> <p>Based on the feedback from the midpoint presentation and results of initial exploratory data analysis, the team will finalize the data model and validate it and its metrics internally. The team will then transition to the Physical model, where they will implement an app-like prototype (R Shiny, React, or similar), with output validation.</p>
4.17-4.23	9	
4.24-4.30	10	<p>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.</p>
5.1 - 5.7	11	
5.8-5.22	Post-Project	<p>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.</p>

Pricing

- Engagement Timeline: 12 weeks, February – May 2023
- Semester Case Fee: **\$20,000**