



Harvard Undergraduate Data Analytics Group

PREPARED FOR

Compassion International

PREPARED DATE

Jul 31, 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.

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 Compassion International:

Phase 1: Automatic CIV categorization (NLP)

In addition to sponsorship activities, Compassion supports children and communities through Complementary Interventions (CIVs). These are generally one-time projects, intended to support a specific Facilitating Church Partner (FCP) or community. While CIVs are well-documented in Compassion's instance of Salesforce, there is a gap in the categorization available for analytics. A CIV may be labeled too vaguely with the specific activity associated with that intervention buried in free-text fields which are difficult to leverage for analysis. For example, a more detailed and actionable categorization for a "Beneficiary Income Generation" CIV may be "Livestock Rearing" (for example, it may be determined that CIVs containing free-text fields indicating "poultry", "pigs", or similar keywords may be grouped into such a category). The grouping would be much more easily leveraged for impact analysis. Note that if it is found that CIVs are best represented in multiple groups, the categorization may be best represented as an array (i.e. a CIV may be correctly labeled as both "Livestock Rearing" and "Business Skill Development").

Compassion is requesting assistance in creating a machine-learning approach to this type of categorization. Further, Compassion is requesting that scripts utilized in this process be developed in a way that can be integrated into existing systems for future use with scale.

Phase 2: Correlating CIVs with program outcomes

Compassion has recent "outcomes" data, collected via on-site surveys, for a significant number of participants in Tanzania and Indonesia. As a secondary activity, categorized CIVs could be correlated to program outcomes measured by this data. This analysis framework, and any initial findings, would be valuable in better understanding direct impact of specific CIV types.

For example, Compassion would be eager to understand whether FCPs implementing "Livestock Rearing" CIVs see improved outcomes in food security compared to FCPs implementing "Food Kit" CIVs.

Desired Deliverables

At the end of the engagement, the HDAG team will present their findings and deliver the following:

1. Data extract which categorizes all existing CIVs into useful groups based on the specific activities described in Salesforce.
2. Python scripts containing NLP or other methods of categorization, which can be integrated into Compassion's systems to group future CIVs.
3. An analysis framework for correlating completed CIVs with outcomes survey results for Tanzania and/or Indonesia.

Rough Engagement Timeline

Dates	Week	Tentative Schedule
9.5-9.18	0	<p>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.).</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>
9.19-9.25	1	<p>CTL will introduce the project and the work plan to the 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).</p> <p>The team will begin exploring the CIVs in Compassion's Salesforce database and generate initial CIV categorizations. At the end of Week 2, the HDAG team will present these initial CIV categorizations and conceptual frameworks (NLP) for automatic CIV classification based on free text descriptions.</p>
9.26-10.2	2	
10.3-10.9	3	<p>Every member of each Client Case Team will follow the work plan. Machine learning models for automated CIV classification will be implemented. Clustering and other unsupervised analyses can be presented to show method utility.</p>
10.10-10.16	4	
10.17-10.23	5	<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>

		<p>By the end of week 5, initial NLP/ML models for CIV classification should be implemented for presentation to Client.</p> <p>A data extract which categorizes all existing CIVs into useful groups based on the specific activities described in Salesforce, along with corresponding Python scripts (containing classification models) should be prepared for midpoint delivery.</p>
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	<p>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.</p> <p>If new ideas for CIV classification models are proposed, they can be implemented after the midway presentation before moving onto Phase 2 of this engagement.</p> <p>HDAG team will begin analyzing “outcomes” data (on-site surveys) in Tanzania and Indonesia. Categorized CIVs will be correlated with program outcomes to characterize the direct impact of specific CIV types.</p>
11.7-11.13	8	
11.14-11.20	9	
11.21-11.27	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>
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 2022.
------------	--------------	--

Pricing

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