Harvard College Data Analytics

Harvard Undergraduate Data Analytics Group

PREPARED FOR

Compassion International

PREPARED DATE

Feb 5, 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.

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 Compassion International:

Project Description

Compassion International is one of the ten largest non-profit organizations in the United States, currently serving two million children in 25 developing nations living in the most extreme levels of poverty. Compassion's programs empower local leaders to invest deeply in children and youth in their communities over many years to tackle poverty holistically—recognizing poverty is not simply a lack of financial resources, but a generational cycle that must be broken by addressing its complex sources. Compassion offers a unique opportunity for sponsors and registered children to correspond through letters, exchanging encouragements.

Compassion call center transcripts could offer a wealth of insight around Compassion sponsors/donors' pain points, questions, emotions, and needs across their entire journey with our organization. Today, our call center tech stack supports very basic analytics (e.g. basic sentiment), but we would like to develop a Natural Language Processing Model (NLP) to automatically code quotes from the transcripts into more meaningful themes.

Datasets

The data for this project will be provided by Compassion and will consist of:

1. Compassion would like to test the NLP model with two of our marketing offices—USA and Australia. Each country will provide an export of supporter call center transcripts. (HDAG must be able to process Class I PII data, as supporter names will be included in these transcripts).

2. Compassion will also provide additional information about the sponsors whose transcripts are represented as potential filter categories—tenure, number of sponsored children, sponsor segment (as available), date of conversation, etc.

Preferred Coding Languages/Tools

Python—flexible and will allow our internal analysts to continue to maintain and update the NLP model even after the project dates.

Potential Roadblocks

1. Supporter information (including Class I data like names) will be included, so Caitlin from Compassion will work with HDAG to ensure appropriate privacy measures will be taken.
2. Sponsor demographic information is limited—age, ethnicity, gender, and segment are only captured for a subset of supporters, but could theoretically be extrapolated.

Harvard College Data Analytics

Desired Deliverables

- Among subset of supporters who call into the call center whose segment we know, identify key words more frequently used by each supporter segment.
- Develop a NLP model to identify key themes, based on a code frame the marketing team will provide—examples including:
    - Top themes by stage of the Compassion journey
    - Key pain points
    - Advanced sentiment analytics
    - Themes related to brand perceptions
- Visualize NLP results in a secure dashboard/app that can be accessed any time and automatically updates
    - Dashboard allows filtering: by segment, time frame, country, sentiment level, etc.
    - Dashboard can be accessed, modified, and maintained by Compassion analysts after project dates
- Enhance our understanding of drivers of +/- sentiment with regression analysis

Rough Engagement Timeline

Dates	Week	Tentative Schedule
2.6-2.19	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.
2.20-2.26	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).
2.27-3.5	2	Every member of each Client Case Team will follow the work plan, initially identifying key words frequently used by each supporter segment. 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.
3.6-3.12	3	
3.13-3.19	4	The team will develop an NLP model identifying key themes such as key pain points and top themes by stage of the Compassion journey. This model will be assessed and continually updated based on weekly client feedback.
3.20-3.26	5	
3.27-4.2	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, in particular the NLP model identifying key themes, based on a code frame the marketing team will have provided. Each HDAG case team will follow up with any questions the Client team might have during or after the presentation.
4.3-4.9	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 team will develop a secure dashboard application to visualize the results of the NLP model developed in the first half of the case. The dashboard will allow for filtering and can
4.10-4.16	8	
4.17-4.23	9	

		be accessible, modifiable, and maintainable by Compassion analysts after the project concludes.
4.24-4.30	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.
5.1 - 5.7	11	
5.8-5.22	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, February – May 2023
- Semester Case Fee: **$15,000**