Harvard
College
Data
Analytics

Harvard Undergraduate Data Analytics Group

PREPARED FOR

Galaxy

PREPARED DATE

Jan 20, 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 seven to eight 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 Galaxy:

The goal of this project is to create a dashboard that will consolidate data from major Layer 1/Layer 2 chains on 10+ key metrics identified by Galaxy. Broadly, the project HDAG will engage with Galaxy would consist of two main stages.

1) After agreeing upon the key networks and metrics with Galaxy, the HDAG team will first identify the best online data sources and consolidate the relevant data into tabular form. This will be done using API queries on major crypto data aggregation websites such as CoinGecko, DefiLlama, and CryptoSlam!, and in an automated way that will allow continual tracking of the metrics of interest.
2) The HDAG team will then proceed to create a web-app based on this tabular data, with potential functionalities including visualizations through graphing, labeling unusual fluctuations, and a ranking system based

At the end of the engagement, the HDAG team will deliver the following:

1) A codebase that will allow Galaxy to replicate and continue tracking methods created by the HDAG team, including scripts, requisite dependencies (e.g. Python libraries), and other software-related deliverables. All code will be provided to and owned by Galaxy upon termination of the project.
2) A slide deck that lays out the methodologies used for collecting and displaying the data.

Harvard College Data Analytics

Rough Engagement Timeline

Dates	Week	Tentative Schedule
2.13 - 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 a more detailed 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). By the end of Week 3, we plan to have identified the relevant data sources for the networks and metrics of interest identified by Galaxy and begin to use APIs to aggregate the data into one consolidated table structure.
2.27 - 3.5	2	
3.6 - 3.12	3	
3.13 - 3.19	4	Every member of each Client Case Team will continue to follow the work plan, mainly focusing on collecting data from their assigned API data source. 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. By the end of Week 5, a preliminary version of the dashboard will be finished, such that it can present a “snapshot” view of the different networks at a given point in
3.20 - 3.26	5	

		time. i
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. 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 initial table will move beyond the "snapshot" view into a real-time web-app dashboard that will allow for continual data collection and the identification of significant fluctuations in different metrics through visualization methods.
4.10 - 4.16	8	
4.17 - 4.23	9	
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. sentiment analysis model-based findings, overall trends in the dataset). 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 - 5.14	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. This extra week is a time that is mainly kept in the case of any delays in finishing deliverables when final

		presentations can still occur.

Pricing

- Engagement Timeline: 12 weeks, February - May, 2023
- Semester Case Fee: $40,000