Harvard College Data Analytics

Harvard Undergraduate Data Analytics Group

PREPARED FOR

CincyTech

PREPARED DATE

Feb 6, 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 CincyTech:

Goal:
The project goal is to analyze the VC round, deal type, deal amount, total invested capital, pre-money valuation, and post valuation for each company in multiple sectors. We will group the companies by industry and analyze general trends as well as casual relationships between variables provided by Pitchbook.

We will also develop a model to detect any causal relationships between a specific variable and the variance between pre-money and post valuation. This will give insight into what factors make an exit backed by a VC firm successful.

Exploratory data analysis will be conducted to quantify the information based on each category bucket.

Datasets:
- **Pitchbook**
  - Industry/vertical
  - VC round
  - Deal type
  - Total invested capital
  - Pre and post money valuation (difference)
- **National venture capital association**
  - PitchBook-NVCA Venture Monitor

Deliverables:
1. **Exploratory Data Analysis:** presentation (based on exploratory analysis) capturing the correlative factors and quantitative attributes analyzing exits by VC backed companies by industry

2. **Predictive modeling:** model to predict the likelihood of future VC activities based on historic VC activity and other related datasets
   a. Model code base (python) including inputs and output results
   b. Presentation summarizing predictive capabilities and model outputs with any relevant visualizations to support key findings based on VC exits based on industry

Harvard College Data Analytics

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 starting with exploratory data analysis (EDA) to identify and quantify the likelihood of predictive characteristics, attributes, and model focus areas. The team will explore each dataset to extract relevant EDA insights, namely features associated with service activity. 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	Once EDA insights have been collected, unsupervised clustering will be performed to identifying additional patterns

3.20-3.26	5	in the data and feature correlations with service activity need. Results will be presented weekly to the client to begin specifying the level of predictive model granularity.
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 EDA report/presentation capturing the correlative factors and quantitative attributes supporting this predictive service activity use-case. 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. Based on the feedback from the midpoint presentation and results of initial exploratory data analysis, correlative analysis, and/or unsupervised ML outcomes, the specific level of model granularity will be determined. The team will then focus on developing a model to predict the likelihood of future service activities based on historic service activity and other provided datasets. Predictions could be categorized by geographic region, specific customer, a specific diagnostics instrument/machine, depending on the level of granularity that the correlative factors and attributes support.
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. protocol ordering and recommendations). The HDAG team will follow
5.1 - 5.7	11	

		up with any questions the Client business team might have during or after the presentation.
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: **$3,000**