Booz | Allen | Hamilton

WOMEN IN DATA SCIENCE

Hack the Bay Final Report

November 2020

Executive Summary

Hack the Bay was a 7-week, virtual hackathon designed to help solve some of the toughest challenges facing the Chesapeake Bay watershed. The hackathon took place Aug. 3-Sep. 20, 2020 and was an entirely volunteer-led collaboration between Booz Allen Hamilton, the Chesapeake Monitoring Cooperative (CMC), and nearly thirty other partnering organizations. Hack the Bay was a year in the making, and ultimately drew 430 participants from around the world.

Highlights include:

- **A multi-office data scoping event in February 2020 with over 50 participants**
- **A winning submission to Databricks' Hackathon for Social Good during the challenge scoping process**
- **A series of five virtual hackathon events designed to keep participants engaged and provide them learning opportunities, which drew 390 attendees**
- **A conference published paper using CMC's data**
- **20 projects addressing four key questions related to CMC's data and watershed restoration**
- **A Booz Allen recruitment hiring event tailored and promoted to hackathon participants, which attracted applications from 4 out of the 12 winning team members who attended**

The hackathon's results were equally exciting. A survey of other online hackathons hosted on Devpost shows that Hack the Bay's participation far exceeded participation rates of other events with non-monetary prizes (and in many cases, also exceeded hackathons offering prizes in the tens of thousands of dollars). With 35 total submitters, Hack the Bay saw a **conversion rate of 8.1%.** Organizers received overwhelmingly positive feedback through the participant survey and from CMC.

Hack the Bay's data scientists helped shape CMC's effort in leading the first federally supported citizen science water quality data initiative. In doing so, participants helped prove the value of CMC's data in informing watershed restoration policy, helped CMC assess their own program for future expansion, and demonstrated the possibilities of introducing new machine learning methods into Bay pollution models. The hackathon also sourced key input on the challenges of working with this data and scoped a variety of techniques for extracting valuable insights for watershed restoration.

The team is seeking opportunities to leverage the products from the hackathon to further develop solutions that could address a broader market need for clients such as Environmental Protection Agency (EPA), United States Geological Survey (USGS), and National Oceanic and Atmospheric Administration (NOAA).

Testimonials

"I am impressed and inspired by the BAH commitment to positive social impact through data, especially the Women in Data Science. It was a privilege and an honor to help out the CMC and CBP teams advance their citizen scientist programs."
— Hackathon Participant

"[Hack the Bay] gave me an appreciation for the complexities of this subject and how blank policy recommendations are extremely difficult. It really excited me - I would love to continue on the subject matter given the opportunity."
— Hackathon Participant

"This work is so useful to us as far as demonstrating the success of the CMC project, particularly to our funders. The overall message from this submission is 'CMC fills the data gaps left by the Chesapeake Bay Program dataset' - which is exactly what we want to be able to demonstrate."
- CMC Judge

"You all are amazing, and I am grateful for the opportunity to combine efforts, wisdom and creativity here."
— CMC Judge

"I was very impressed with the presentations tonight by the Chesapeake Bay Hackathon winners. ...Thank you for sponsoring this."
— Booz Allen Partner

Table of Contents

Background

Social Good Opportunity

In 2015, the U.S. EPA's Chesapeake Bay Program (CBP) funded the formation of the CMC to expand their data collection efforts to capture watershed monitoring conducted by volunteer, academic, NGO, and local government groups. As part of this effort, CMC has brought together previously siloed water monitoring groups, developed a pipeline and platform for collecting, vetting, sharing, and scaling data on the Chesapeake Bay, and engaged residents in the conservation process.

The CMC is at its next steppingstone, seeking assistance in demonstrating and leveraging the value of the data collected and housed in its Chesapeake Data Explorer. Ultimately, the CMC aims to illustrate the value their data provide to the CBP partnership, state agencies, and community partners by: demonstrating enhanced insight and understanding of Chesapeake Bay and watershed health, connecting their data back to the impact of Bay restoration practices supporting adaptive management strategies, and hosting fully developed visualization/analytical tools on their website.

To support these goals, the CMC partnered with our team of Booz Allen volunteers to solicit crowd-sourced solutions through a public hackathon.

Event Goals & Objectives

A hackathon is an event where people with diverse skillsets collaborate to solve a problem. Hack the Bay set out with several goals:

Innovate for Social Good: Working with our nonprofit partner, design and execute an event that brings attention to real data-centered challenges and their impact on the world, while also sourcing creative and potentially innovative ideas for addressing them

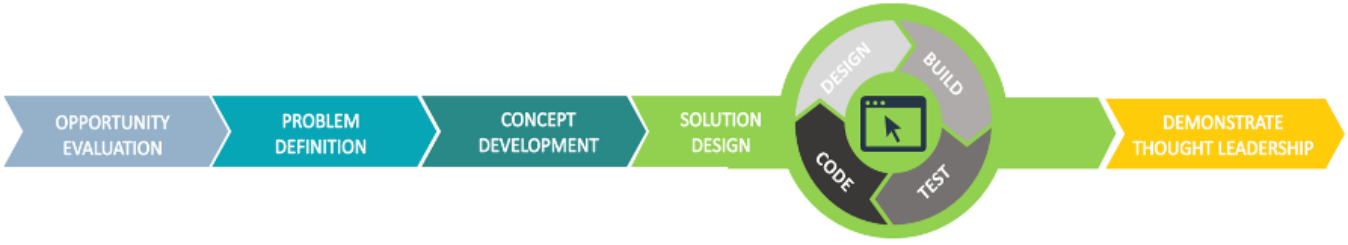
Support Technical Development: Create opportunities for hackathon participants to enhance and showcase their skills in data science, web development, design, communication, and project management

Unite Problem-Solvers: Partner with external organizations to bring together consultants, students, and professionals in the field to collaborate and network

In alignment with the goals of Women in Data Science, Hack the Bay strove to be inclusive of all genders, races, ethnicities, geographies, and skill levels.

Social Good Pipeline

Hack the Bay was designed to follow Women in Data Science's Social Good Pipeline: a process for scoping, developing, and scaling solutions that support real-world challenges using Booz Allen's resources and talent pool.



In anticipation of conceptualizing a scalable prototype (Stage 4: Solution Design), Hack the Bay sought to bring together various market subject experts to design challenges that were not only feasible from a capability perspective, but also interesting to the broader environmental and energy market.

Event Plan

Partners

Hack the Bay was a highly collaborative event. The event was led by Booz Allen with advisory support from members of the **Water Community of Practice** and **Infrastructure, Energy, and Environmental (IEE) Functional Community.** The nonprofit partner, **Chesapeake Monitoring Cooperative,** represented and engaged experts from the **Chesapeake Bay Program** (within EPA), **Alliance for the Chesapeake Bay, Izaak Walton League of America, University of Maryland Center for Environmental Sciences,** and **Alliance for Aquatic Resources Monitoring.**

Through Booz Allen market leads and connections within CMC, the Hack the Bay team also reached out to connections at the **Environmental Protection Agency (EPA), United States Geological Survey (USGS), National Oceanic and Atmospheric Administration (NOAA),** and **Northern Virginia Regional Commission** to invite them to participate in the event. Through recruiting virtual event panelists and judges, Hack the Bay also engaged environmental science experts from **Blue Water Baltimore, Clean Energy Leadership Institute, Choose Clean Water Coalition, South River Keepers, D.C. Department of Energy and Environment, Eco.Logic, Climate Justice Alliance, Latino Outdoors,** and **Chesapeake Conservancy.** A concerted effort was made to recruit college students to participate in the hackathon – in marketing efforts, the Hack the Bay team reached out to the **College of William & Mary, Georgia Tech, University of Maryland College Park, Carnegie Mellon University, Virginia Tech, Stanford University, Old Dominion University, Johns Hopkins University, Howard University,** and **General Assembly**.

Timeline & Milestones

Hack the Bay was a year in the making. Significant milestones in the process included:

AUGUST-DECEMBER 2019

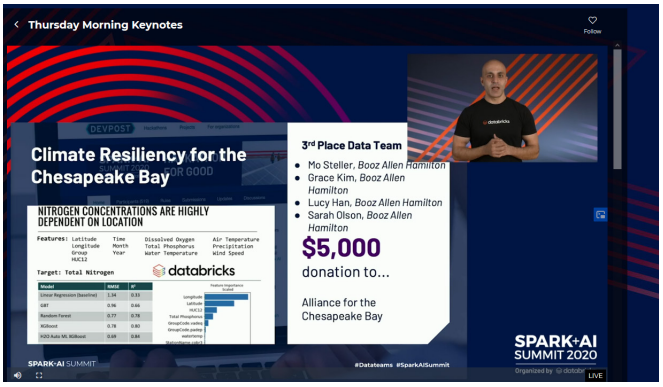
Booz Allen developed interest in supporting a hackathon centered on CMC's data and water quality issues. The initial scoping team solicited partners and leaders within Booz Allen to advise the scope of the event and potential impact on market initiatives. A plan was developed to orchestrate a hackathon during April 2020 to coincide with the 50th Anniversary of Earth Day.

FEBRUARY 2020

Hack the Bay kicked off with a Design-o-thon: a 2-hour ideation session designed to crowd-test ideas for the hackathon challenge tracks. The event was at Booz Allen's DC office (901 15th Street NW) and attended by over 50 people in-person and remotely (joined by a team from Booz Allen's San Diego office).

MARCH-JUNE 2020

Due to Covid-19, the hackathon was postponed, and the team made plans to pivot to a virtual format. Volunteers further explored CMC's data and collaborated with the nonprofit on scoping the challenges. A team of Booz Allen data and climate scientists entered a separate virtual hackathon (Databricks Hackathon for Social Good) using CMC's data to address the impact of climate change on water quality. This team won third place and $5,000 for CMC, validating the opportunity for innovation with this data and novel machine learning/AI tools. The team published a paper on their findings via the 2020 Climate Informatics Conference.



AUGUST-SEPTEMBER 2020

The hackathon was held virtually over a 7-week period, culminating in a virtual winner presentation session on Sep. 29.

Challenge Tracks

Hack the Bay offered participants the option to focus on one of four tracks designed to leverage CMC's data to address different problems and appeal to participants with skills across data science, web development, design, and communication:

1. **Develop a Restoration Case Study:** Tell a story about how water quality has changed over time in the Chesapeake Bay watershed. (Time Series / Visualization)
2. **Identify Data Gaps:** Demonstrate how and where CMC's data fills the gaps in the Chesapeake Bay Program's database, and where data gaps in the watershed still exist. (GIS / Mapping)
3. **Model Water Pollution:** Analyze potential causes for/ build a predictive model for pollution in a section of the Bay. (Machine Learning)
4. **Design a Water Quality Report Card:** Design a local or regional version of the Chesapeake Bay report card that ties water quality to the values of the communities living in the watershed. (Web Development / Design)

For each challenge, supplementary datasets were recommended as well as background reading and recommendations from the nonprofit partner on how to approach each problem.

Virtual Engagement

The Covid-19 pandemic and subsequent concerns around social gathering created new pressures for the Hack the Bay team. With the hackathon originally timed to occur in person during Earth Week (April 2020), the team postponed the event and made the strategic decision to move forward with a virtual, 7-week hackathon in the summer to ensure Hack the Bay would continue in a safe, smart, and socially responsible manner. While a virtual format presented its own unique challenges, there were some benefits: (1) we were able to engage a broader audience; (2) the extended timeline afforded teams more time and flexibility to develop their projects; and (3) the event served as a pilot for learning how to virtually engage participants in future socially distanced events.

We partnered with Devpost to host an online platform for the hackathon through which participants could sign up, create teams, submit projects, access resources and hackathon guidelines, and receive notifications on the event. In addition to Devpost, we leveraged a Slack workspace for communicating with participants and Webex for hosting virtual events.

Due to the extended timeline and virtual format, we coordinated several virtual events designed to keep participants engaged and provide them opportunities to learn about technical and domain areas related to the hackathon:

Hackathon Kick-Off (Aug. 3)
This virtual launch of the hackathon introduced participants to the CMC, Booz Allen (and Women in Data Science), the data and the challenges. A half hour of the event was dedicated to breakout sessions for each hackathon track, where participants could discuss the challenges in greater detail and ask questions.

Wrangling Geospatial Data (Aug. 5)
In partnership with Women in Data Science's Education Pillar, this coding tutorial was led by Booz Allen data scientists and provided an overview of useful open source packages and tools in Python, R, and QGIS for newcomers to geospatial data.

Bay Issues Panel (Aug. 11)
Organized by the CMC, this virtual panel invited experts from various watershed monitoring groups and the DC government to discuss current issues facing the Chesapeake Bay, including the status of the total maximum daily load, the emerging issue of microplastics, environmental policy priorities under Covid-19, and success stories of public-private partnerships and citizen science monitoring.

Environmental Justice Expert Panel (Aug. 18)
This event invited leaders from organizations focused on environmental justice to discuss the intersection of social inequality, economics, and the environment, covering topics from climate migration, to public health, to green infrastructure, to equity for clean public spaces.

Winner Presentations (Sep. 29)
In the culminating event of the hackathon, each winning team was invited to demo their solutions to a live audience. The event included remarks from Booz Allen Strategic Innovation Group (SIG) Vice President John Larson and an analysis of future application of the solutions at CMC from Liz Chudoba.

Collectively, these events were attended by over 390 participants.

Results

Participation & Exposure

430
Unique participants

37
Countries

172
Cities

103
Universities/ organizations

390+
Virtual event attendees

A survey of other online hackathons on Devpost shows that Hack the Bay's participation far exceeded participation rates of other events with non-monetary prizes (and in many cases, also exceeded hackathons offering prizes in the tens of thousands of dollars.) The Hack the Bay Team engaged in marketing efforts through multiple channels from July 20 (the date the Hack the Bay platform went live) through September 25, 2020.

6,659
Visits to Hack the Bay's online platform by 3,846 unique visitors

37.14%
Hack the Bay's bounce rate - the percentage of visitors who enter the site and leave rather than continuing to further explore the site (the average bounce rate for other hackathons on Devpost is 51.76%)

00:08:07
140% higher than the average for other Devpost hackathons

The following statistics show progress data of the 430 participants who registered for Hack the Bay. With 35 total submitters, Hack the Bay saw a conversion rate of 8.1%.

126
Participants joined the Hack the Bay Slack channel

202
Working solo

116
Have a team

99
Looking for teammates

1-2
Average number of people per team

Submissions

After seven weeks of hacking, we received a total of 20 project submissions from 35 participants hailing from all over the world – internationally, from Israel, Nigeria and Singapore, and within the U.S. from California, Washington, New Jersey, Oklahoma, Florida, Tennessee, and the DMV region. Several finalists were current data science students from various universities, including the University of Pennsylvania, Tufts University, George Washington University, Nigeria Teacher's Institute, and University of Maryland – College Park.

Our judges evaluated each submission based on scalability, robustness, and creativity. After a very challenging judging process (with one tie, and two categories where the winner was decided by a narrow margin), we recognized four winning solutions:

Develop a Restoration Case Study: Hack the Bay

A team of three recent graduates of the data analytics course at George Washington University.

Judge Comments: "Hack the Bay" evaluated multiple water quality parameters in CMC's data over time using a methodology that could be scaled to other watersheds and visualizations that told a clear story.

Identify Data Gaps: Mind the Gap

Four data science professionals from Maryland, Washington, D.C., Florida, and New York.

Judge Comments: "Mind the Gap" developed an online data visualization tool and novel data collection prioritization system that can be used as-is to help CMC plan for future monitoring.

Identify Data Gaps: Richard and Lu, Dynamic Duo

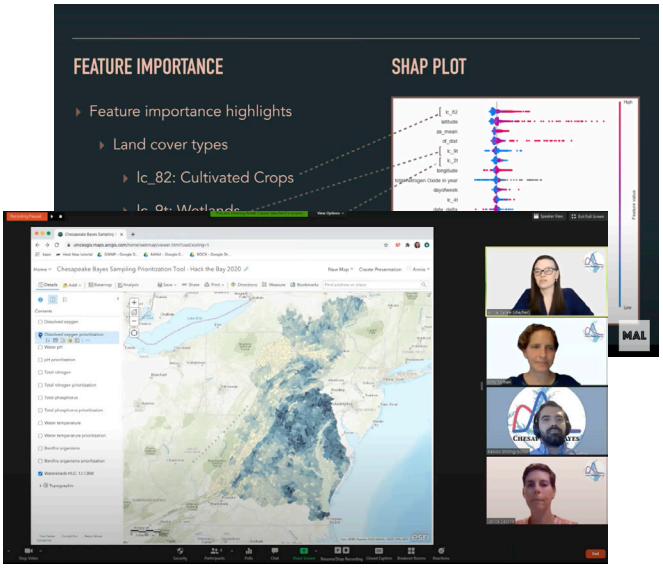
A team of two Booz Allen data scientists and current Washington, D.C. residents.

Judge Comments: "Dynamic Duo" developed an analysis aligned with CMC priorities that clearly demonstrated, interpreted, and tracked CMC's progress in filling the gaps in the Chesapeake Bay Program's database.

Model Water Pollution: Shore is Fun

Five data science professionals working together from Seattle, WA, San Francisco, CA, Tulsa, OK, Boston, MA, and Mill Valley, CA.

Judge Comments: "Shore is Fun" modeled total nitrogen water pollution using land and air data, and delivered a professional, robust analysis that was both a great showcase of machine learning tools and a novel solution for understanding water quality predictors.



Hackathon Feedback

Following the end of the hackathon, we solicited feedback from participants in a post-event survey. We received 22 responses that were overwhelmingly positive:

100% of respondents felt the hackathon supported their personal growth, knowledge or professional development, with 90% of respondents feeling at least very strongly about Hack the Bay's impact on their personal and professional growth.

All respondents indicated that the event was inclusive of all genders and races. Inclusion in a virtual and external-facing format was an important aspect for Hack the Bay.

77% of respondents felt confident that they would be using new skills developed during the hackathon.

A majority of respondents enjoyed the virtual events and panels, specifically the virtual panels on issues impacting the Bay and environmental justice.

Next Steps

Hack the Bay was a volunteer-led initiative to advance our understanding of watershed science for the Chesapeake Bay. Through the process, hundreds of data scientists, developers, and designers expanded their understanding of environmental science, surfaced unique challenges and opportunities of analyzing CMC's nontraditional water quality data, and scoped a variety of techniques for extracting valuable insights for watershed restoration.

Impressed with the solutions developed, our partners have offered further opportunities for hackathon participants to present their work to audiences in the environmental space. Furthermore, the hackathon has created an opportunity to source talent for Booz Allen and resulted in actionable insights for future virtual events. Our team is currently seeking opportunities to leverage the products from the hackathon to further develop solutions that could address a broader market need.

Hack the Bay helped shape CMC's effort in leading the first federally supported citizen science water quality data initiative. Participants helped prove the value of CMC's data in informing watershed restoration policy, empowered CMC to assess their program for future expansion, and demonstrated the possibilities of introducing new machine learning methods into Bay pollution models. By shedding new light on one of the most studied watersheds in the world, our team hopes this work inspires others to join the global citizen science movement and support local environmental and conservation efforts.