

Dataforward - Data Science Upskilling Program For FDA Staff

Tomas Drgon PhD &
The Applied Learning Track Network
ODT/ODAR



Abstract

Dataforward is a program that upskills and reskills FDA employees' data science skillset. We support FDA's Data Science Community, offering employees' professional growth, experiential learning opportunities, and a cross-Center network. Through bringing data science to day-to-day work, Dataforward enhances and improves the FDA's ability to achieve our mission.

Introduction

Applied Learning Track Basecamp is an experiential, remote learning program, a great starting place for FDA employees interested in exploring data science.

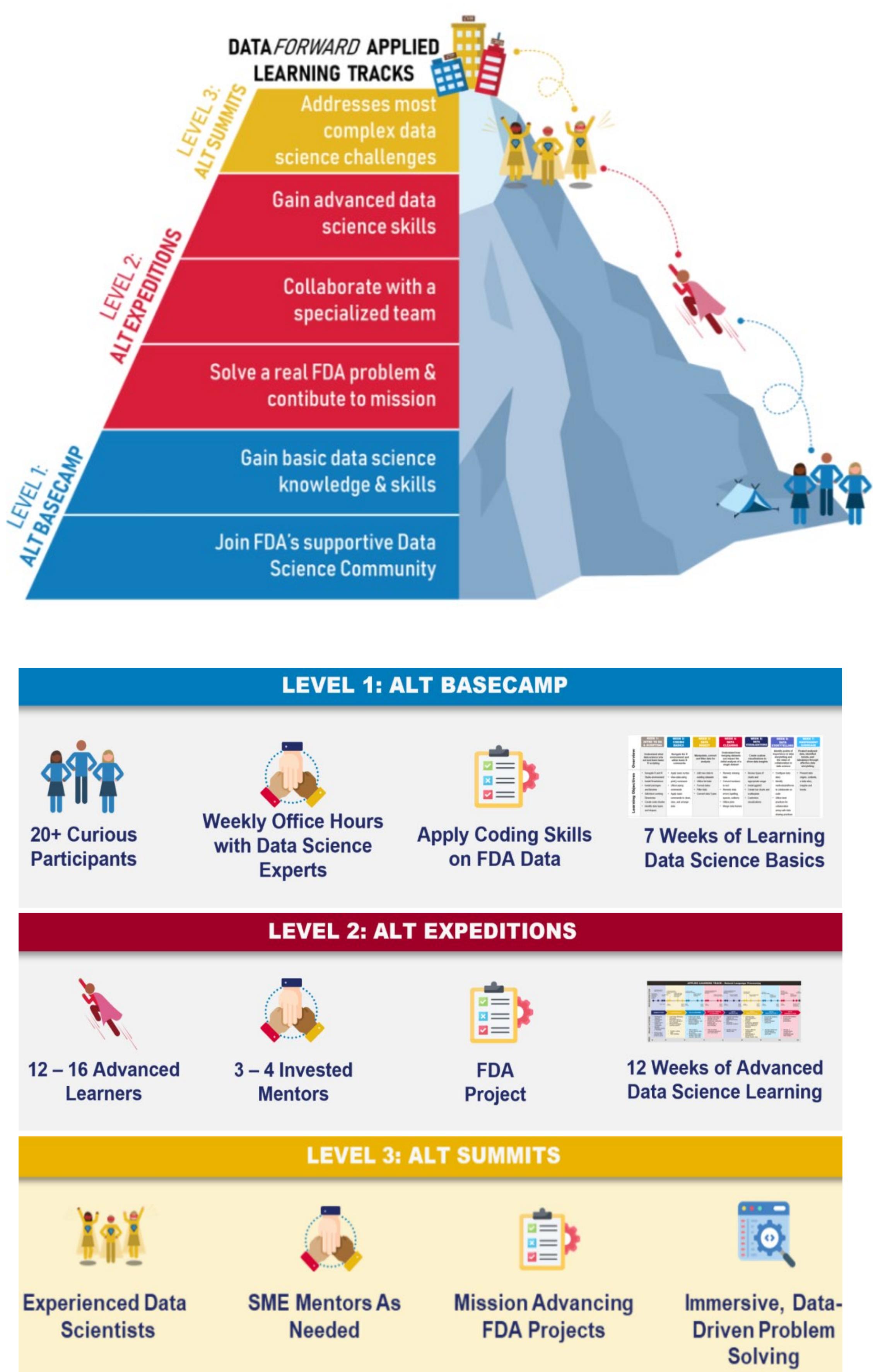
Participants will 1) gain basic data science skills to confidently approach problems at FDA, and 2) join FDA's robust Data Science Community with access to its resources. As a prerequisite for ALT Expeditions, ALT Basecamp allows employees to explore their interest in data science with support from FDA's Data Science Community. Empowered by new cross-agency connections, graduates return to their Centers and/or Offices ready to tackle data problems with poise.

Applied Learning Track Expedition is an advanced, experiential learning opportunity in data science that further develops FDA's data science talent network. This remote course helps highly motivated go-getters expand their data science skills in advanced topic areas to solve health crises and make a real impact at FDA. ALT Expeditions builds on the data science fundamentals taught in ALT Basecamp. Participants collaborate with a project team to solve a real problem at FDA using data from a Center and/or Office.

Applied Learning Track Summit is a continuation of ALT Expedition, where data scientists are deployed to advanced projects to solve complex data problems for Centers and/or Offices. Participants further develop their data science careers and skills while making a difference at the FDA. ALT Summits allow participants to apply and practice the problem-solving data science skills they acquired in previous ALT programs. This ALT program does not include a curriculum for guided learning. Instead, participants tackle a project independently with minimal guidance from a mentor.

Materials and methods

Dataforward is a community based facilitated self-learning educational program. The program fosters communication, teamwork, and employment of open-source tools (R, python). Budding data scientists of varying skills are matched with data science "champions" in Applied Learning Track Basecamp and led through a basic data science curriculum that combines in-house content with LinkedIn Learning and other sources. They work on weekly assignments for 6 weeks. After graduation, selected participants continue in Applied Learning Track Expeditions, where they are assigned to one of 3 or 4 teams and matched with data science mentors. Here they work on a specific real world FDA data problem presented to the ALT by FDA staff (project owners). The teams communicate and present a solution to the data science problem to the project owners. Graduates of the Expeditions track may further develop successful and implementable data solutions into real FDA implementations in ALT Summit. If successful, the implementations may inform regulatory decisions.



Results and discussion

The Applied Learning Track Basecamp has a capacity of 25 trainees, ALT Expeditions 12 trainees, and the ALT Summit 6-10 trainees in the Summit. The Dataforward graduated cohort now counts over 100 data scientists from across the FDA.

The Dataforward ALT participant selection process assures that the candidates represent:

1. All FDA Centers and Offices (CBER, CDER, CDRH, CFSAN, CTP, CVM, NCTR, OC, ORA)
2. The GS levels where most of the hands-on data science work is likely to be performed - GS 7 through GS15, with most participants at GS 12, 13 and 14
3. Skill levels - beginner to intermediate data scientists with focus on self reported "Basic experience", "Some experience", and "Intermediate experience"

The ROI analysis documents Dataforward ALT's substantial Return On Investment (2022 to date)

Value:

- Cost to replace the current ATL crisis response reserve through contract support - **\$35.9M**
- Value of the completed Dataforward ALT projects - **\$1.7M**
- Value of decreased turnover rate of participants - **~\$2.5M**
- Other non-monetary benefits: Increased job satisfaction, Community connection, Institutional knowledge retention, Cross-Center solutions

Investment:

FTE support - **~\$500K**

Contract support - **~\$1.75M**

Conclusion

The strength of Dataforward ALT program is in its community approach. The open-source software skills increase skill portability, help de-silo the data science community, foster cross-center collaboration, and collaboration with academia and industry. The program saves resources by relying on in-house development and in-house staff.

The Dataforward Applied Learning track accomplishments to date:

- Training offered in R and Python
- 111 graduates overall
- 61 Basecamp graduates
- 50 Expedition graduates
- 10 Summit participants
- 12 graduates Dataforward ALT became mentors or champions
- The Dataforward ALT cohort tackled and delivered **solutions to 5 FDA data problems** (CFSAN, ORA, CVM, CTP)
- The Dataforward has created a **Data Science Leadership program**: 3 sessions focused on training of non-scientist managers that manage data scientists and data science teams (first instance delivered in October/November 2024)
- With the overall investment of about \$2M to date the Dataforward ALT program has created a **tangible value to the FDA of over \$36M**.
- The Dataforward ALT has also created a data science community that **presents a non-tangible value to the FDA** with increased employee satisfaction, increased retention rates, retention of institutional knowledge, and creating an FDA-wide culture of data sharing and collaboration that exists outside of the regular FDA reporting structure, thus contributing to the robustness and readiness of the agency to respond to crises and future data efforts in data science, machine learning, and artificial intelligence.