The Data Incubator

Academic Catalog

1 January 2019 - 31 December 2019

41 E 11th St, 11th Floor
New York, NY 10003
-
1601 Connecticut Ave NW, Suite 400
Washington, DC 20009
-
2323 Broadway
Oakland, CA 94612
-
Telephone: (202) 905-6115
https://www.thedataincubator.com
admissions@thedataincubator.com

Table of Contents

Mission & Objectives

Mission
The Data Incubator

is a leading educator for individuals with advanced degrees looking to transition into the data science industry, committed to increasing data science literacy and adoption.

offers students a comprehensive and innovative education to achieve success as data scientists, delivered in a hands-on, practical learning environment, by an organization that values exploration, collaboration, integrity, and excellence.

fosters a community of data scientists among it students, alumni, staff, and industry partners, committed to the sharing and expansion of data science knowledge, continuing life-long education, and training.

Objectives
The Data Incubator is committed to offering practical education in a student-first environment where administrators and faculty work diligently to deliver services and instruction that meets individual student needs and prepares them for future success in the data science field. To achieve this, we have identified the following objectives:

1. Thoroughly qualify program candidates, through a technically rigorous application process, to ensure admitted students possess the necessary prerequisites.
2. Provide an educational setting that encourages engagement and fosters collaboration and community through in-classroom peer-to-peer learning exercises, and extracurricular group activities.
3. Embrace continuous improvement in our work to ensure an improved experience for each graduating cohort of students.
4. Enforce and maintain policies that ensure academic comprehension and accountability.
5. Attract a diverse and experienced group of administrators and instructors dedicated to the success of the students both during and after the program.

Courses Offered

The Data Incubator offers a full-time immersive Fellowship program as well as a part-time online Foundations course. The Data Science Fellowship Program is designed to prepare students for data science careers. The part-time Data Science Foundations program is designed to help students gain foundational skills in Python and data analysis. The part-time course is not geared toward career transitioning and may be designated as “avocational.” In some states, avocational, or non-occupational, courses are not intended to provide instruction that will result in the student’s acquisition of occupational skills for a particular job. The Data Incubator’s educational programs are not designed to lead to positions in a profession, occupation, trade, or career field requiring state licensure.

The Data Incubator offers the following courses:

Course	Length	Type of Course	
		Immersive	Part-time
Data Science Fellowship Program	320 hours / 8 weeks	✔	
Data Science Foundations Program	32 hours / 8 weeks		✔

Facilities & Equipment

Equipment:
Students are required to bring their own laptops. Laptops must be capable of connecting to the internet and running a modern Chrome browser. The Data Incubator provides each student with a cloud server (hosted by DigitalOcean) with a 4-core processor, 8GB of RAM, and 80GB of disk space for coursework completion.

Students participating remotely must also have a means of connecting to GoToMeeting for lecture (a webcam is not required, but the ability to view a video conference with audio is necessary). Students interested in joining the Foundations course, or the Fellowship Program online, are encouraged to test their computer’s ability to connect to GoToMeeting prior to signing up / applying. GoToMeeting provides the ability to join a test session here: https://support.citrixonline.com/en_US/Meeting/all_files/G2M050001

California (Bay Area) Facilities:
In California, The Data Incubator operates out of the Oakland Impact Hub, located at:

2323 Broadway
Oakland, CA 94612

The Hub provides classrooms for lectures and desks for out-of-lecture work, as well as wifi. A kitchen and restrooms are available onsite for student use as well.

The Data Incubator owns the equipment necessary for videoconferencing for lecture:

- Logitech ConferenceCam CC3000e integrated full duplex speakerphone with echo and noise cancellation
- Logitech 1080p HD video conferencing camera
- Asus Flip Tablet (Windows 10; 4GB RAM; 60GB hard drive; 64-bit OS; Intel Atom x5-Z8350 1.44 GHz)

The Impact Hub provides a 55” flat-screen HD TV that serves as a screen for lecture. Students are also able to follow along in the lecture notebooks on their laptops.

New York City Facilities:
In New York City, The Data Incubator has an office located within TechSpace at:

41 E 11th St, 11th Floor
New York, NY 10003

The office serves as classroom and student workspace and is outfitted with desks, chairs, and videoconferencing equipment. TechSpace provides a kitchen area, call rooms, wifi, and restrooms for all members.

Washington, DC Facilities:
In DC, The Data Incubator has an office suite located at:

1601 Connecticut Ave NW, Suite 400
Washington, DC 20009

The office contains a classroom for lectures and student work (outfitted with desks and chairs for all students as well as videoconferencing equipment), call rooms, a kitchen, and restrooms for student use.

Hours of Operation:
The Data Incubator’s offices are open from 9AM - 6PM, Monday-Friday, with exceptions made for Federal Holidays. After-hours and weekend access is typically available, but is subject to change. Students will be provided with details regarding after-hours access upon program start.

Data Science Fellowship Program

(320 hours / 8 weeks)

The Data Science Fellowship Program is an intensive 8-week course that aims to prepare master's students, PhDs, and postdocs in STEM and social science fields for careers in industry as data scientists (SOC Code 15-2051).

The course is divided up into eight modules, each with an associated assignment ("miniproject"). There is also a job search curriculum, comprised of both soft skills and technical interview practice, that is taught throughout the 8 weeks (there is one soft skills lecture per week along with an associated assignment, and each Friday technical interview practice takes the place of normal technical lecture).

Course Level: Intermediate

Topics Covered: *

- Data manipulation and cleaning in Python
- Machine learning (introductory and advanced)
- MapReduce
- Data visualization
- Spark
- TensorFlow
- Data science for business
- Soft skills & technical interview practice

**See the following pages for a detailed breakdown of each module and the topics covered therein.*

Students Will Learn How To:

- Effectively preprocess, clean, and join data from disparate sources
- Build predictive models to answer industry-relevant questions
- Write and deploy large-scale distributed applications using Spark and MapReduce
- Analyze and visualize data as part of decision making processes
- Understand the skills necessary to navigate the data science job market and secure employment in the field

Prerequisites:

- Master's or PhD (earned or to be completed within 1 year from admission)
- Some prior programming experience
- Foundational knowledge of statistics
- Proficiency in the English language

Pre-Course Preparation:
Students admitted to the Data Science Fellowship Program are provided access to preparatory material contained within a “12-Day Program.” They are asked to work through the material on their own and complete an associated Milestone Project. The Data Incubator recognizes that all of its students are starting out at slightly different levels, with different strengths and weaknesses, and exposure to all concepts necessary for program success is not uniform across the board. The purpose of the 12-Day Program and its associated Milestone Project is to ensure that all students entering the Fellowship have the foundational skills necessary for success.

Student Assessment:
The Data Incubator assesses student performance in the following ways:

- Each module has an associated miniproject that students must complete. These miniprojects require students to write code and submit their answers to an automated grader. While there are weekly miniproject deadlines, students who have not met the 90% passing grade on each question are encouraged to continue working on the project until they achieve 90%. Students must achieve a 90% on all miniproject questions in order to graduate from the program.
- Students will work on an individual Capstone Project throughout the course. This project will be of their own choosing and proposed during the admissions process. The goal of the Capstone is to give students a project for their portfolio that they can showcase in conversation with potential employers. Students will submit a brief (2-minute max) video each week that will be used by instructors to judge Capstone progress. Students will review one another’s videos each week and provide feedback; each Friday, a few videos will be selected for group discussion during the interview practice lecture. Submission of a final Capstone video that meets the standards in the rubric provided to students in the course syllabus is necessary for program graduation.
- Students complete daily coding challenges (via HackerRank) to help them become familiar with the types of questions they may be asked in a coding interview. While these are not graded, participation in the coding challenges is mandatory.

Graduation Requirements:
In order to graduate from the program, students must achieve a 90% on all miniproject questions and submit a final Capstone video.

Data Science Fellowship Program Admissions Policies

Entrance Requirements
Students must have (or be within one year of completing) either a Master's or PhD from an institution of higher education accredited by an accrediting association recognized by the US Department of Education. The Data Incubator does not admit ability-to-benefit students.

Prospective students must be proficient in instruction in the English language and possess the skills necessary to operate a computer. Some prior programming experience is necessary for successful admission, as well as a foundational knowledge of statistics.

Admittance Standards
The Data Science Fellowship Program is a competitive program for admittance. Our admissions process involves a written application, a technical coding challenge, resume review, and a videoconference interview.

Admissions Deadline
Admissions decisions are sent out at least 4 weeks prior to the first day of the course. The Data Incubator reserves the right to make exceptions and admit students to a given course after the admissions deadline has passed under special circumstances.

Data Science Fellowship Program Curriculum Overview

For a further breakdown of topics covered, see https://www.thedataincubator.com/curriculum.html

A week is spent on each of the following eight technical modules:

Data Manipulation and Cleaning in Python

Module Description:

The first step of data science is mastering the computational foundations on which data science is built. We cover the fundamental topics of programming relevant for data science - including pandas, numpy, scipy, matplotlib, regular expressions, sql, json, xml, checkpointing, and web scraping - that form the core libraries around handling structured and unstructured data in Python. Students gain practical experience manipulating messy, real-world data using these libraries. They also walk away with a firm understanding of tools like pip, git, ipython, jupyter notebooks, pdb, and unit testing that leverage existing open source packages to accelerate data exploration, development, debugging, and collaboration.

Topic Outline:

- Consuming APIs (and JSON)
- SQL - Standard Query Language
- Iterators, Generators, and Coroutines
- Overview of Scraping and Munging Technologies
- How to (Software) Engineer Real Good
- Pandas
- Scraping
- Dealing with Strings in Python
- NumPy, SciPy, and Matplotlib

Associated Project Work:

Students will scrape picture captions off of a website that tracks the goings-on of New York's socially well-to-do. By extracting names from these captions, they will assemble a graph of friendships amongst this crowd. Analysis of this graph will produce insights about the most connected New Yorkers.

Students will assemble a SQL database of 4 years' worth of NYC restaurant inspection data. They will write and execute queries against this database to understand the variations in scores and violations across the city and between different types of restaurants.

Introduction to Machine Learning

Module Description:

In a world with abundant data, leveraging machines to learn valuable patterns from structured data can be extremely powerful. We explore the basics of machine learning, discussing concepts like regression, classification, model evaluation metrics, overfitting, variance versus bias, linear regression, ensemble methods, model selection, and hyperparameter optimization. Through powerful packages such as scikit-learn, students come away with a strong understanding of core concepts in machine learning as well as the ability to efficiently train and benchmark accurate predictive models. They gain hands-on experience building complex ETL pipelines to handle data in a variety of formats, developing models with tools like feature unions and pipelines to reduce duplicate work, and practicing tricks like parallelization to speed up prototyping and development.

Topic Outline:

- K Nearest Neighbors
- Unsupervised Learning
- Bias, Variance, and Overfitting
- Scikit-learn Workflow
- Learning and Metrics
- Linear and Logistic Regression

Associated Project Work:

Students will develop a series of models to predict a venue's star rating from various features. Working from 100MB of real-world data, they will start with location-based models before building models based on other attributes of the venues. Finally, an ensemble model will blend the individual models into a final prediction of the venue's popularity.

Distributed Computation, Hadoop, and MapReduce

Module Description:

With the advent of big data, the disk-space, memory, and computational resources of a single computer are no longer sufficient. We introduce the basic concepts around distributed computing - spreading out workloads over multiple computers. Topics covered include hadoop, HIVE, partitioning, fault-tolerance, hadoop streaming, mrjob, and distributed machine learning. Students will walk away with a firm understanding of distributed computing paradigms, how to efficiently break up a workload across multiple nodes, and how to select between competing distributed computing paradigms. Students gain direct hands-on experience building, debugging, and deploying MapReduce jobs to run on large, real-world datasets in the cloud (AWS and Google Cloud Platform).

Topic Outline:

- mrjob in the Cloud
- Hadoop and MapReduce
- Naive Bayes
- Out-of-core and online learning
- Python MapReduce

Associated Project Work:

Students will process 10GB of XML data from the Wikipedia project using MapReduce. Taking a series of improved processing algorithms, they will answer questions about the distribution of words and letters on Wikipedia pages. Then they will examine the structure of Wikipedia as revealed by the links between pages.

Data Visualization

Module Description:

Data science is about helping humans understand the story behind the data, and visualizations provide a powerful tool for helping the analyst understand and communicate that story. We discuss the biases and limitations of both visual and statistical analysis to promote a more holistic approach.

Topic Outline:

- Explanatory Visualization with D3
- Advanced D3
- Visualizing Large / High Dimensional Datasets
- Exploratory Visualization
- Overview of Data Visualization
- JS Visualization in Python
- An HTML Primer
- A Javascript Primer
- Visualization Theory

Associated Project Work:

Students will build and deploy two interactive visualizations for their Capstone projects.

Additional optional practice:
Students will build an interactive website giving information on NYC's bus system. They will process historical data and develop plots to illustrate trends. Using a live feed of bus information, they will compare the current state to this historical average. All of the visualization will be deployed as a Flask app running on Heroku.

Advanced Machine Learning

Module Description:

While machine learning on structured data lays an important foundation, a larger world of analytical opportunities becomes available through understanding advanced machine learning techniques and how to handle unstructured data. We explore techniques such as support vector machines, decision trees, random forests, neural nets, clustering, KMeans, expectation maximization, time series, and signal processing. Students come away with intuition about the suitability of different techniques for different problems. In addition to handling structured data, students directly apply these techniques to large volumes of real-world unstructured data, solving problems in natural language processing using Word2Vec, bag of words, feature hashing, and topic modelling.

Topic Outline:

- Support Vector Machines (SVM)
- Deep Learning
- Recommendation Engine
- Decision Trees and Random Forests
- Time Series
- Comparing ML Algorithms
- Outlier Detection
- Unbalanced Classes
- Word Disambiguation Using NLP
- Digital Signals

Associated Project Work:

Students will use NLP techniques to extract sentiment from English text. Working with 300MB of venue reviews, they will build a series of models to predict the star rating associated with a given review. They will also examine statistically improbable phrases that appear in the text corpus.

Students will examine methods of dealing with seasonality, as they build models to predict temperatures in several cities. The training data come from National Weather Service observations and must be cleaned before use.

Students will build a machine learning model to classify musical snippets by genre. Working from raw waveforms, they will extract a number of features to use as input to a model. In processing 200MB worth of data, students must consider efficiency and performance issues.

Data Science for Business

Module Description:

Sometimes the most important question to ask in data science comes from thinking beyond the data itself. We explore a myriad of topics that affect data science decision making as a whole, and affect the implementation of data-driven business policies. Important topics include data fidelity, relevance, and the value of additional data. Bias is a major theme, and students think about how their conclusions are influenced by data collection, external factors, internal structuring, procedural artifacts, and more. Students gain a broader understanding of how to balance trade-offs to suit the business problem, such as when to favor accuracy over interpretability and vice versa. We also discuss more practical engineering considerations like building for prediction speed or robustness, and deploying to different environments. Students apply this knowledge to case studies that simulate what they would be expected to contribute as part of a real-world team faced with a business problem.

Topic Outline:

- Hypothesis Testing
- Personal Interview Questions
- Algorithms and Data Structures
- The purpose of case studies
- What the data really says
- Statistics

Distributed Computing with Spark

Module Description:

Spark is a technology at the forefront of distributed computing that offers a more abstract - but more powerful - API. This module includes the basics of Scala like map, flatmap, for comprehension, and data structures. We cover core concepts of Spark like resilient distributed datasets, memory caching, actions, transformations, tuning, and optimization. Students get to build functioning applications from end to end and learn critical tooling around Spark (sbt, jvm) to make them more productive. They apply that knowledge to directly developing, building, and deploying Spark jobs to run on large, real-world datasets in the cloud (AWS and Google Cloud Platform).

Topic Outline:

- Tweet mini case study
- SparkR
- Streaming Technologies
- PySpark Intro
- Spark MLlib
- DataFrames
- Creating Spark Applications
- Spark Intro
- Scalding (Scala on Cascading)
- Scala Primer
- PySpark MLlib
- Advanced Topics in Spark
- DataFrames

Associated Project Work:

Students will use Spark to parse and process 10GB of data on posts and users of a popular Q&A website. They will extract insights on the posting habits of users and develop predictors of user behavior from their posts. Spark's machine learning capabilities will be used to discover meaning in unstructured text data.

Machine Learning with TensorFlow

Module Description:

TensorFlow is taking the world of deep learning by storm, but it is actually a fairly general platform for large numeric computation. We demonstrate its capabilities through its Python interface and build some simple machine learning models. We give a brief overview of the theory of neural networks, including convolutional and recurrent layers. Students will practice building and testing these networks in TensorFlow, using real-world data. They will come away with both theoretical and practical understanding of the algorithms behind deep-learning algorithms.

Topic Outline:

- Introduction to TensorFlow
- Iterative Algorithms
- Machine Learning in TensorFlow
- Basic Neural Networks
- Deep Neural Networks
- Adversarial Noise
- Convolutional Neural Networks
- DeepDream and the Inception Model
- Variational Autoencoders
- Recurrent Neural Networks
- Other Resources

Associated Project Work:

Students will build a series of models to classify images from the Cifar-10 dataset. These models will include basic image analysis, convolutional neural networks, and transfer learned deep neural networks.

In addition, we have a Job Search curriculum that is taught throughout the program:

Job Search Curriculum

Our Job Search curriculum is designed to teach both the technical aspects of job searching in data science - coding and programming interviews - as well as the soft skills aspects. We assist Fellows in designing a data science resume, crafting cover letters, and developing behavioral interviewing techniques as well as strategies for professional salary and benefit negotiation.

We hold two job search sessions a week: a soft skills lecture, and an interview practice session focusing on coding and statistics interview questions.

As part of the Job Search curriculum, Fellows complete resumes, cover letters, and interview practice sessions which are graded and required for completion of the program.

Data Science Foundations Program

(32 hours / 8 weeks)

This part-time online course is geared towards giving working professionals with no previous data science experience an immersive, hands-on experience with foundational data science techniques. Classes are held Tuesdays and Thursdays from 7 - 9 PM Eastern over the course of 8 weeks.

Course Level: Beginner / Introductory

Topics Covered: *

- Introduction to data wrangling in Python
- Introduction to predictive analytics

**See the following pages for a detailed breakdown of each module and the topics covered therein.*

Students Will Learn How To:

- Manipulate and analyze structured data
- Programmatically work with differently-formatted data and messy unstructured data
- Code efficiently leveraging advanced programming concepts
- Leverage probability and statistics fundamentals in order to understand noisy data
- Use the fundamentals of modelling to predict future trends
- Utilize accurate modelling to understand non-normal distributions

Prerequisites:

- High school diploma or GED

We also recommend that students have:

- Previous experience manipulating data in spreadsheets (experience with macro programming a plus)
- Completed some college-level mathematics
- A basic familiarity with statistics
- Proficiency in the English language

Pre-Course Preparation:

Students are asked to attempt to complete the Python lesson provided by Codecademy (https://www.codecademy.com/learn/python) prior to the first class, so that they come in with at least an initial familiarity with Python. Students are also asked to review several probability concepts covered in the publicly available textbook, A First Course in Probability, 8th Edition, by Sheldon Ross (ISBN 013603313X). A study guide is provided to students in the form of a PDF.

Student Assessment:
Students in the Data Science Foundations Program complete two miniproject assignments. These miniprojects require students to write code and submit their answers to our automated grader. Students receive immediate scoring and are able to continue to work on their code and submit until they achieve a passing score. Passing scores are defined as scores greater than or equal to 90% on each miniproject question. Students must achieve a 90% on all miniproject questions for successful program completion.

Completion Requirements:
Students must achieve a 90% on all miniproject questions for successful program completion.

Employment Preparation:
No job placement assistance is provided for the Foundations program. The Foundations program is not geared toward career transitioning and may be designated as “avocational.” In some states, avocational, or non-occupational, courses are not intended to provide instruction that will result in the student’s acquisition of occupational skills for a particular job.

Data Science Foundations Program Admissions Policies

Entrance Requirements
This course is open to the public. Anyone with a high school diploma or GED may register. We recommend that students have prior experience manipulating data in spreadsheets, some coding experience, and a basic familiarity with statistics. Students must be capable of operating a computer and receiving instruction in the English language. The Data Incubator does not admit ability-to-benefit students.

Admittance Standards
This course is open to the public. Anyone with a high school diploma or GED may register.

Admissions Deadline
The deadline for registration is 2 weeks prior to the first meeting of the course. The Data Incubator reserves the right to make exceptions and admit students to a given course after the admissions deadline has passed under special circumstances.

Data Science Foundations Program Curriculum Overview

For a further breakdown of topics covered, see https://www.thedataincubator.com/curriculum.html

The following two modules are taught in the Data Science Foundations Program:

Introduction to Data Wrangling

Module Description:

Much of the world's data cannot be easily processed in Excel or other spreadsheets. It is either too inaccessible, too messy, too unstructured, too varied, or too large for elementary handling. This course is designed to equip students with core tools to start down the path towards becoming data scientists. We cover the basic data structures and file formats, and then move into classes and SQL. Basics of Pandas, Numpy, and matplotlib follow from that. We introduce the fundamental building blocks of data manipulation. We also demonstrate how to translate simple Excel commands into more powerful languages like SQL and Python, as well as how to build on the existing powerful open-source libraries.

Topic Outline:

- Numpy
- Pandas
- SQL - Standard Query Language
- Python Built-in Data Structures
- Python Data Formats
- Object-oriented programming
- Functions
- Exceptions

Associated Project Work:

Students will gain experience with Python-based data wrangling technologies to extract insights from a structured, web-API-based dataset. Students will learn the fundamental building blocks of data extraction, manipulation, and aggregation via Pandas DataFrames and good Python programming practice.

Introduction to Predictive Analytics

Module Description:

Machine learning requires strong statistical foundations. In this module, we solidify that groundwork by reviewing probability concepts - such as important distributions, Bayes' Rule, and conditional expectation - and move on to rigorous statistical analysis: parameter estimation, hypothesis testing, p-values, z-scores, and other core concepts in statistical inference. We extend this basis of statistical knowledge into machine learning basics: regression, regularization, overfitting, and important learning metrics.

Topic Outline:

- Bias, Variance, and Overfitting
- Scikit-learn Workflow
- Probability
- Linear and Logistic Regression: What this is all about
- Learning and Metrics
- Introduction to Statistics

Associated Project Work:

Students will gain experience applying statistical concepts (parameter estimation, hypothesis testing, p-values, z-scores) on real-world datasets, using Python’s stats-oriented libraries to ask interesting, relevant questions and draw concrete inferences from population data.

Transfer of Credit

Data Incubator courses are not credit-bearing. The Data Incubator does not accept hours or credit from other institutions through transfer of credit, challenge examinations, achievement tests, or experiential learning. Courses taken at The Data Incubator are unlikely to count as transfer credit at another institution. The Data Incubator does not give or evaluate experiential learning credit.

Academic Policies

Hours:
Academic credit is measured in clock hours. One (1) hour of instructional time is defined as a sixty (60) minute period.

Probation:
The Data Incubator does not provide a probation option. If a student is not making satisfactory progress as determined by the Student Assessment procedures described in the Course Description, they may be dismissed from the program.

Attendance:
This attendance policy exists to ensure that students get the most out of the program. In order to successfully complete the program, students are expected to attend every session. Attendance is not required in the case of job interviews and sickness, though students must communicate their absences to a Data Incubator instructor. Likewise, if an unexpected conflict arises, contact a Data Incubator instructor.

Completion:
For the Fellowship, a LinkedIn Certificate and Certificate of Completion are issued within 10 days of the end of the two-week grace period for assignment completion following the end of the course (in other words, 10 days from the end of “Week 10”) to each student who has successfully completed the Data Science Fellowship Program. A Letter of Completion will also be available upon request.

For the Foundations Program, a Certificate of Completion is issued within 10 days of the end of the two-week grace period for assignment completion following the end of the course to each student who has successfully completed the Data Science Foundations Program.

Student Rights:

1. Students have the right to equal opportunity education and non-discrimination based on sex, gender, race, color, religion, ancestry, national origin, disability, medical condition, genetic information, marital status, sexual orientation, or other categories protected by law of the states in which we operate.
2. Students have the right to view their own academic records.
3. Students have the right to cancel or withdraw from their course, per The Data Incubator's Cancellation, Withdrawal, and Refund Policy.
4. Students have the right to file a grievance, per The Data Incubator's Grievance Procedure.

Student Services

Housing:
The Data Incubator does not provide student housing.

Library and Learning Resources:
Fellowship students at The Data Incubator have access to over 650 pages of comprehensive notes for over 125 hours' worth of lectures contained in interactive Jupyter notebooks to allow for hands-on learning. Students are given access to the notebooks at the start of the program and are able to maintain access after program completion. Materials are available 24 hours a day and accessible with an internet connection, which is provided at each of The Data Incubator's locations.

In addition to the course notebooks, resources available to Fellowship students include:

- Full-time instructors on site in each city (who are also available via Slack / email throughout the day for questions from online and in-person students)
- Substantial references to both print and readily available online resources contained within curriculum materials
- A dedicated Slack team, to serve as a forum for questions, discussion, and collaboration
- Technical lectures twice a day (one of which is focused on discussing the day's coding challenge)
- Dedicated office hours (hosted on GoToMeeting), held three times per week

Fellowship instructors have dedicated office hours three times a week, but are also available throughout the day (both physically and via Slack) to answer questions and offer assistance. Availability often continues after-hours via Slack as well.

Additionally, Fellowship students retain access to the pre-course preparation materials (contained within the “12-Day Program”) provided prior to program start. These can be used as a reference throughout the course as well.

Foundations students at The Data Incubator have access to approximately 150 pages’ worth of comprehensive lecture notes contained in interactive Jupyter notebooks to allow for hands-on learning. Students are given access to the notebooks at the start of the program and are able to maintain access after program completion. Materials are available 24 hours a day and accessible with an internet connection.

The Foundations course is taught in real time by a live instructor. Lectures are recorded and shared with students via YouTube for later reference. The instructor holds office hours twice a week, and is available for questions via Slack and email outside of normal class hours.

Foundations students are added to a dedicated Slack team which serves as a forum for questions, collaboration, and discussion.

Career and Job Placement Assistance:
The Data Incubator provides job placement assistance for the Fellowship program only. No placement assistance is provided for the Foundations program.

For the Fellowship, The Data Incubator has a network of 300+ hiring partner companies and its goal is to help its students find employment in data science roles with these companies. Students who maintain passing scores in the program (minimum of 90% on all miniproject assignments) and complete the soft skills assignments are given access to the contacts and open roles with the Incubator’s hiring partners.

Our Job Search curriculum is designed to teach both the technical aspects of job searching in data science - coding and programming interviews - as well as the soft skills aspects. We assist Fellows in designing a data science resume, crafting cover letters, and developing behavioral interviewing techniques as well as strategies for professional salary and benefit negotiation.

We hold two job search sessions a week: a soft skills lecture, and an interview practice session focusing on coding and statistics interview questions.

As part of the Job Search curriculum, Fellows complete resumes, cover letters, and interview practice sessions which are graded and required for completion of the program.

The Incubator assists its Fellowship students in placements as often as possible, but does not guarantee job placement to any student.

Distance Education:
Students participating in one of The Data Incubator's courses online ("distance education") will receive instruction in real time and feedback on their assignments at the same time as their in-person peers. Miniprojects are graded via an automated grader, so students receive their scores within minutes.

Student Records and Files:
Student final grades and descriptions of courses offered are maintained permanently. All other school and student records will be maintained electronically for 50 years.

Students may view their own academic records. Students who seek to view their own records should contact the Program Experience Director.

The Data Incubator will take reasonable steps to protect the privacy of personal information contained in student records.

Grievance Procedures

Internal Grievance Procedure:
Persons seeking to resolve problems or complaints should contact the Program Experience Director for any ethics issues or grievances about classes, students, instructors, staff members, or situations. If a resolution does not occur, students who have concerns may contact the COO or CAO. Any contact/discussion with the COO / CAO is completely confidential.

External Grievance Procedure:
A student or any member of the public may file a complaint about this institution with the Bureau for Private Postsecondary Education by calling the toll-free phone number at 888-370-7589, or by completing a complaint form, which can be obtained on the Bureau's Internet Web Site: www.bppe.ca.gov.

Cancellation, Withdrawal, Dismissal, and Refund Policies

Cancellation:
The Student has the right to cancel without any penalty or obligation through attendance at the first class session, or seven days after enrollment, whichever is later. Cancellation is effective when student provides written notice of cancellation to Admissions Lead via email:

Fellowship: Alyssa Thomas | alyssa@thedataincubator.com
Foundations: Sarah Fugate | sarah@thedataincubator.com

The notice must clearly express the student no longer wishes to be bound by the Enrollment Agreement. If the Agreement is cancelled in the manner described within Section 4 of the Agreement, the Institution will refund the student any net institutional fees paid, less the Registration Fee, within 30 days after notice of cancellation is received

Withdrawal:
If a student withdraws after the first day of the program, they will receive a pro-rata refund if they have completed sixty percent (60%) or less of the course through the last day of attendance. Student shall be responsible for 100% of the tuition if they completed more than sixty percent (60%) of the course. For the purposes of determining the amount of the refund, the date of the student's withdrawal shall be deemed the later of the last day of attendance or notification of withdrawal. All refunds will be paid within 30 days of the withdrawal. Student may withdraw from the program if they provide written notice of withdrawal to Admissions Lead via email:

Fellowship: Alyssa Thomas | alyssa@thedataincubator.com
Foundations: Sarah Fugate | sarah@thedataincubator.com

The refund shall be based on net institutional fees paid, less the Registration Fee.

Dismissal:
The Data Incubator reserves the right to terminate a student's training at any point.

Refund Policy:
All refunds will be paid within 30 days of cancellation. Students who cancel their enrollment in accordance with the Cancellation Policy will receive a refund of all moneys paid, less the Registration Fee and any Non-Institutional Fees. Students who withdraw after the cancellation period should reference their signed Enrollment Agreement for applicable refund provisions.

Tuition and Fees

Unless otherwise agreed to in a private lending agreement and as approved by The Data Incubator, students must issue payment for 100% of the total tuition by the first day of instruction. The Data Incubator may allow students to delay payment of tuition under extraordinary circumstances. If you believe you have an extraordinary circumstance, please reach out to admissions@thedataincubator.com.

Fellowship Program:
Note: A period of attendance comprises the entire program. Hence, the fees for a period of attendance are the fees for the entire educational program.

	Fellow	Scholar: on-site, direct payment	Scholar: on-site, financed payment	Scholar: online, direct payment	Scholar: online, financed payment
Registration Fee (Institutional)	$250.00	$250.00	$250.00	$250.00	$250.00
Supplies & Equipment Fee (Non-institutional)	$1,750.00	$1,750.00	$1,750.00	$1,750.00	$1,750.00
Student Tuition Recovery Fund (STRF) Fee * (Non-institutional, *non-refundable*)	$0.00	$0.00	$0.00	$0.00	$0.00
Tuition Fee (Institutional)	$15,000.00	$15,000.00	$15,000.00	$8,000.00	$8,000.00
Direct Payment Discount (Institutional)	- $1,000	- $1,000	N/A	- $1,000	N/A
Fellowship Discount (Institutional)	- $15,750	N/A	N/A	N/A	N/A
Total Fees for Period of Attendance **	**$250.00**	**$16,000.00**	**$17,000.00**	**$8,000.00**	**$9,000.00**
Total Program Fees for Entire Educational Program **	**$250.00**	**$16,000.00**	**$17,000.00**	**$8,000.00**	**$9,000.00**

*STRF: $0.00 for every $1,000 of tuition rounded to the nearest $1,000. Applies only to students in California.
**A period of attendance comprises the entire program. Hence, the fees for a period of attendance are the fees for the entire educational program.

Please see Appendix C for information regarding the Student Tuition Recovery Fund.

Foundations Program:
Note: A period of attendance comprises the entire program. Hence, the fees for a period of attendance are the fees for the entire educational program.

Registration Fee (Institutional)	$250.00
Supplies & Equipment Fee (Non-institutional)	$250.00
Student Tuition Recovery Fund (STRF) Fee * (Non-institutional, *non-refundable*)	$0.00
Tuition Fee (Institutional)	$2,995.00
Total Fees for Period of Attendance **	**$3,495.00**
Total Program Fees for Entire Educational Program **	**$3,495.00**

*STRF: $0.00 for every $1,000 of tuition rounded to the nearest $1,000. Applies only to students in California.
**A period of attendance comprises the entire program. Hence, the fees for a period of attendance are the fees for the entire educational program.

Please see Appendix C for information regarding the Student Tuition Recovery Fund.

Financial Assistance

The Data Incubator does not participate in federal or state financial aid programs and we do not provide institutional financing.

Loans:

If a student obtains a loan to pay for an educational program, the student will have to repay the full amount of the loan plus interest, less the amount of any refund. If the student receives federal student financial aid funds, the student is entitled to a refund of the moneys not paid from federal financial aid funds.

If a student defaults on a federal or state loan, both of the following may occur: (1) The federal or state government or a loan guarantee agency may take action against the student, including garnishing an income tax refund; and (2) The student may not be eligible for any other government financial assistance at another institution until the loan is repaid.

General Information for California Students

As a prospective student, you are encouraged to review this catalog prior to signing an enrollment agreement. You are also encouraged to review the School Performance Fact Sheet, which must be provided to you prior signing an enrollment agreement. Students will be provided with a PDF version of the catalog before receiving an enrollment agreement. The catalog will also be made available on The Data Incubator's website at https://www.thedataincubator.com/ca/catalog.html.

The school catalog is reviewed at the end of each year for regulatory and compliance requirements. All course information, school policies, and student information are updated where necessary. The Data Incubator reserves the right to change any provision of the catalog at any time. Students are expected to read and be familiar with the information contained in the catalog and with all school policies. By enrolling in The Data Incubator, the Student agrees to abide by the terms stated in the catalog and all school policies.

The Data Incubator does not participate in federal or state financial aid programs.

The Data Incubator is not accredited by an accrediting agency recognized by the United States Department of Education at this time, and therefore does not qualify to offer federal (Title IV), state, or any other government financial aid.

The Data Incubator is not operating as a debtor in possession, has not filed a petition within the preceding five years, nor has this institution had a petition in bankruptcy filed against it within the preceding five years that resulted in reorganization under Chapter 11 of the United States Bankruptcy Code (11 .S.C. Sec.1101 et seq.) 94909(a)(12).

Appendix A: Management

Michael Li, Chief Executive Officer and Chief Academic Officer
Patrick Smith, Chief Operating Officer
Mike Anderson, Chief Strategy Officer
Megan Nissel, Director of Operations
Alyssa Thomas, Director, Program Experience and Career Advising
Sarah Fugate, Director, Online Learning
Erika Underwood, Director, Partnerships

Appendix B: Faculty

The Data Incubator employs the following full-time instructors (Data Scientists in Residence). Biographies for instructors can be found on The Data Incubator's website at https://www.thedataincubator.com/curriculum.html.

Instructor	Course	Degree	Institution	Years of Experience
Robert Schroll	Fellowship	PhD, Physics	University of Chicago	12+ years data analysis; 10+ years Python programing; 8+ years software development
Richard Ott	Fellowship	PhD, Physics	Massachusetts Institute of Technology	10+ years data analysis and software development
Dylan Bargteil	Fellowship, Foundations	PhD, Physics	New York University	6+ years data analysis and research; 5+ years teaching
Zachary Glassman	Fellowship, Foundations	MS, Chemical Physics	University of Maryland, College Park	6+ years Python programming, research, and data analysis
Timothy Pollio	Fellowship	PhD, Mathematics	University of Virginia	6+ years research and teaching
Tim Schwuchow	Fellowship	MA, Economics	Duke University	7+ years teaching and course design, data analysis
Don Fox	Foundations	PhD, Chemical Engineering	Cornell University	7+ years data analysis and research
Ana Hocevar	Fellowship	PhD, Physics	University of Ljubljana	10+ years data analysis, research, and programming
Kai Zhang	Fellowship	PhD, Physical Chemistry	Duke University	10+ years programming, data analysis, research
Alyssa Thomas	Fellowship (Soft Skills)	MA, Education: Curriculum and Instruction	University of South Florida	7+ years teaching, corporate training, and curriculum design

Appendix C: Specific Disclosures Required by the California Bureau for Private Postsecondary Education

The Data Incubator is a private education institution approved to operate by the California Bureau for Private Postsecondary Education (BPPE). Approval to operate means compliance with state standards set forth in the CEC and 5, CCR.

INTERNATIONAL STUDENTS AND ENGLISH LANGUAGE SERVICES
The Data Incubator does not provide international student visa services nor vouch for a student's status, and/or any related or associated charges. The Data Incubator does not offer English as a second language instruction. All instruction occurs in English. Language proficiency is judged in the Fellowship admissions interview.

NOTICE CONCERNING TRANSFERABILITY OF CREDITS AND CREDENTIALS EARNED AT OUR INSTITUTION
The transferability of credits you earn at The Data Incubator is at the complete discretion of an institution to which you may seek to transfer. Acceptance of the certificate you earn in the educational program is also at the complete discretion of the institution to which you may seek to transfer. If the credits or certificate that you earn at this institution are not accepted at the institution to which you seek to transfer, you may be required to repeat some or all of your coursework at that institution. For this reason you should make certain that your attendance at this institution will meet your educational goals. This may include contacting an institution to which you may seek to transfer after attending The Data Incubator to determine if your certificate will transfer.

ARTICULATION AGREEMENTS
The Data Incubator has not entered into transfer or articulation agreements with any other college or university.

LEAVE OF ABSENCE POLICY
The Data Incubator does not grant leaves of absence.

HOUSING
The Data Incubator does not assume responsibility for student housing, does not have dormitory facilities under its control, and does not offer student housing assistance. Short term housing for approximately $1,500 - $2,000 per month is available within 15 miles of the office. Housing can be sourced from short term rentals through services such as AirBnB.

STUDENT TUITION RECOVERY FUND (STRF)

The State of California established the Student Tuition Recovery Fund (STRF) to relieve or mitigate economic loss suffered by a student in an educational program at a qualifying institution, who is or was a California resident while enrolled, or was enrolled in a residency program, if the student enrolled in the institution, prepaid tuition, and suffered an economic loss. Unless relieved of the obligation to do so, you must pay the state-imposed assessment for the STRF, or it must be paid on your behalf, if you are a student in an educational program, who is a California resident, or are enrolled in a residency program, and prepay all or part of your tuition.

You are not eligible for protection from the STRF and you are not required to pay the STRF assessment, if you are not a California resident, or are not enrolled in a residency program.

It is important that you keep copies of your enrollment agreement, financial aid documents, receipts, or any other information that documents the amount paid to the school. Questions regarding the STRF may be directed to the Bureau for Private Postsecondary Education, 2535 Capitol Oaks Drive, Suite 400, Sacramento, CA 95833, (916) 431-6959 or (888) 370-7589.

To be eligible for STRF, you must be a California resident or are enrolled in a residency program, prepaid tuition, paid or deemed to have paid the STRF assessment, and suffered an economic loss as a result of any of the following:

1. The institution, a location of the institution, or an educational program offered by the institution was closed or discontinued, and you did not choose to participate in a teach-out plan approved by the Bureau or did not complete a chosen teach-out plan approved by the Bureau.
2. You were enrolled at an institution or a location of the institution within the 120 day period before the closure of the institution or location of the institution, or were enrolled in an educational program within the 120 day period before the program was discontinued.
3. You were enrolled at an institution or a location of the institution more than 120 days before the closure of the institution or location of the institution, in an educational program offered by the institution as to which the Bureau determined there was a significant decline in the quality or value of the program more than 120 days before closure.
4. The institution has been ordered to pay a refund by the Bureau but has failed to do so.
5. The institution has failed to pay or reimburse loan proceeds under a federal student loan program as required by law, or has failed to pay or reimburse proceeds received by the institution in excess of tuition and other costs.
6. You have been awarded restitution, a refund, or other monetary award by an arbitrator or court, based on a violation of this chapter by an institution or representative of an institution, but have been unable to collect the award from the institution.
7. You sought legal counsel that resulted in the cancellation of one or more of your student loans and have an invoice for services rendered and evidence of the cancellation of the student loan or loans.

To qualify for STRF reimbursement, the application must be received within four (4) years from the date of the action or event that made the student eligible for recovery from STRF.

A student whose loan is revived by a loan holder or debt collector after a period of noncollection may, at any time, file a written application for recovery from STRF for the debt that would have otherwise been eligible for recovery. If it has been more than four (4) years since the action or event that made the student eligible, the student must have filed a written application for recovery within the original four (4) year period, unless the period has been extended by another act of law.

However, no claim can be paid to any student without a social security number or a taxpayer identification number.

GENERAL INFORMATION

As a prospective student, you are encouraged to review this catalog prior to signing an enrollment agreement. You are also encouraged to review the School Performance Fact Sheet, which must be provided to you prior to signing an enrollment agreement.

Any questions a student may have regarding this catalog that have not been satisfactorily answered by the school may be directed to the Bureau for Private Postsecondary Education at 2535 Capitol Oaks Drive, Suite 400, Sacramento, CA 95833, www.bppe.ca.gov, toll-free telephone number (888) 370-7589 or by fax (916) 263-1897.