Omaha Data Science Academy

Around 2016, the Contemporary Analysis (CAN) data science team saw a need in the local community. There were job openings for data scientists, but no one to fill them. There were a few college students graduating the three local programs, but they lacked the experience necessary for the roles.

The "aha!" moment came when we learned multiple people already employed by the companies advertising the job openings had most of the skills needed to fill those roles, they just lacked one or two of the skills needed to function as a "data scientist".

CAN set out to create an academy that trains the individuals in the missing skill sets. We created the Omaha Data Science Academy (Oma-DSA).

The DSA is not meant to replace a degree in mathematics or economics. In fact, you need a degree to apply to the academy. Instead, it is meant to give a person who already has a degree, the skills necessary to become a fully functional data scientists--one needed by most of the companies in Omaha.

Graduates of the DSA will gain skills in 4 key areas:

Programming--Whether it be machine learning, web scrapers, applications, or computational modeling, a data scientist has to know how to think programmatically as well as write code to connect disparate systems.

Data Manipulations and Management--This class will teach the student how to design, store, clean, query, and access data which is the key foundation to any projects success.

Machine Learning and Data Science Modeling--Machine Learning is one of the most in-demand skills in job descriptions today. This skill is broken into 3 areas of knowledge-Basic Modeling, Model Evaluation, and Advanced Data Science Modeling.

Data Visualization Using Tableau--Expressing data is key to implementation, scale, and corporate buy-in. We teach the fundamentals of building and expressing data visually so non-technical users can make decisions from the data.

These skills will give a graduate the necessary knowledge to lead their companies down the road of data discovery and give those companies a leg up in both the local and the global economy.

Welcome.

Omaha Data Science Academy

Mission and Vision: To Provide a(n):

- Base knowledge of how to do:
    - Data Science Programming
    - Data Manipulation and Management
    - Data Science Modeling including Machine Learning
    - Data Visualization
- Understanding that Data Science is both an evolving and continual learning process including:
    - How to continue to learn
    - How do figure out how to do data science when traditional knowledge doesn't yield good results
- Mentor Network
    - CAN
    - Professors
    - Presenters
    - Omaha Data Scientists Slack Channel/Users Group
- Peer Network
    - Classmates
    - Former Classmates
    - Omaha Data Scientists Slack Channel/Users Group
- Career Guidance
    - Continuing Education Opportunities
    - Job Board
    - Internship Opportunities
    - Placement Assistance

Omaha Data Science Academy 2018/2019 Calendar

2018
Cohort 5

Introduction to Data Science   4 weeks--July 11th-August 6th
Data Manipulation and Management   4 weeks--August 15th-September 12th
Basic Model Building   4 weeks--September 26th-October 22nd
Mathematics of Model Evaluation   4 weeks--November 5th-December 3rd

2019

Advanced Data Science Modeling in Python   4 weeks--January 7th-February 4th
(2019)
Data Visualization Using Tableau   4 weeks--February 13th-March 11th
(2019)
Cohort 6
Introduction to Data Science   4 weeks--January 14th-February 6th
Data Manipulation and Management   4 weeks--February 18th-March 13th
Basic Model Building   4 weeks--March 25th-April 22nd
Mathematics of Model Evaluation   4 weeks--May 6th-June 3rd
Advanced Data Science Modeling in Python   4 weeks--June 10th-July 3rd
Data Visualization Using Tableau   4 weeks--July 15th-August 7th

Cohort 7
Introduction to Data Science   4 weeks--May 29th-June 24th
Data Manipulation and Management   4 weeks--July 1st-July 24th
Basic Model Building   4 weeks--August 5th-August 28th
Mathematics of Model Evaluation   4 weeks--September 9th-October 2nd
Advanced Data Science Modeling in Python   4 weeks--October 14th-November 6th
Data Visualization Using Tableau   4 weeks--Nov. 18th-December 16th

Fundamentals of Data Science Certificate

1. Introduction to Data Science   **4 weeks**

In this class, students will be introduced to some of the major concepts of Data Science (Python Programming, Database Management, Modeling, and Data Visualization) and some of the tools used in profession. The tools include a crash course in the basics of programming, data structures and object oriented design, basic web development, how to use Jupyter Notebooks, GitHub, and web scrapers as well as functional programming concepts and key Python libraries: (Numpy and Pandas).

2. Data Manipulation and Management   **4 weeks**

This class teaches a student how to store and transform data specifically to be used in modeling. It will also teach database design, SQL queries, different schemas, data cleaning techniques, and data appending. The class will also introduce a tool called Dataiku--a data platforming tool used for easier data engineering and visual/drag and drop data science**.

3. Machine Learning and Data Science Modeling: **12 weeks**
Machine Learning is one of the most in-demand skills in job descriptions today. This module is broken into 3 classes (Basic Modeling, Mathematics of Model Evaluation, and Python and Advanced Data Science Modeling). By taking these classes together, a student will have a firm grasp of Machine Learning and Data Science Modeling.

3a. Basic Model Building   **4 weeks**

Fundamentally Data Science is using statistics and economic modeling to predict what is likely to happen next. This class will teach the student the fundamentals of how to build models. This will include the basics of model evaluation, choosing target variables and characteristics, and basic machine learning. Students will learn both how to model on a data science platform, a standard moving forward in data science, as well as how to search for and write common algorithms in Python.

3b. Mathematics of Model Evaluation   **4 weeks**

This class will dive into the metrics behind evaluating an analytics model's performance using F1, Accuracy, Precision, Recall, AUC, Cost matrix, and Cumulative Lift. We will additionally show the steps to: building, testing, evaluating,

---

adjusting/rebuilding, re-testing, and re-evaluating a model. We will also teach how to choose which model to use the pitfalls to just using accuracy as an indicator.

3c. Python and Advanced Data Science Modeling   **4 weeks**

Data Science has gone from needing to know how to code to most modeling techniques having standardized libraries that can be pasted into a program. This means that one may do data science without understanding what the models mean or actually do. This class will drill into how to program the models the traditional way. We will use Word2Vec to scrape, debug, and enhance data science models. We will also show how to use Python to solve other gaps such as calculations, other data manipulation, and random number population.

4. Data Visualization Using Tableau   **4 weeks**

With an increase in demand of 1,581% since 2011, Forbes showed this one skill alone is immensely valuable in helping key non-technical business users understand data, create corporate buy-in, and make decisions from the data. In this class, we will teach the fundamentals of expressing data visually using Tableau, an industry-wide benchmark for quality visualization tools. Students will learn the necessary skills to build visualizations and best practices needed to make them implementable.

Notes:

*While Introduction to Data Science is a fundamental course with no prior knowledge of programming needed, a basic understanding of technology is expected. Introduction to Data Science is also the prerequisite for all other data science courses unless the student has previous IT experience or knowledge of other object oriented languages. Course skipping is approved on a case by case basis.

**While not there yet, there will be less and less actual programming in data science in the future (including database management). It is for this reason we teach Data Science both in programming and on a platform.

Learning Objectives
Fundamentals of Data Science Certificate --Cohort 5

Introduction to Data Science   4 weeks

- Data Science Concepts
  - What is it?
    - Programming
    - Databases
    - Statistics and Computational Modeling
    - Data Visualization
- Tools of Data Science:
  - Intro to Python:
    - Jupyter Notebooks
    - Numpy
    - Pandas
    - Scrapers

Data Manipulation and Management   4 weeks

- Database Joins
- Aggregation
- Target Variables
- Characteristic Selection

Basic Model Building   4 weeks

- How to Program ML
- Choosing Target Variables
- Choose Characteristics
- Public Libraries

Mathematics of Model Evaluation   4 weeks

- Base parameters of base algorithms:
- Regression
- Random Forest
- Gradient
- Accuracy of machine learning
- Precisions v F1 v Accuracy
- How to evaluate models using matplotlib and/or dataiku

Advanced Data Science Modeling in Python   4 weeks
- Numpy
- Pandas
- Advanced Modeling Techniques

Data Visualization Using Tableau   4 weeks
- Understanding Tableau Interface
- Connecting Tableau to data sources
- Chart Styles
- Geospatial Data
- Calculations
- Data Visualization Best Practices