



OMA–DSA

Data Science Academy

2020
Cohorts 8 & 9

Omaha Data Science Academy

In late 2015, 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 (ODSA).

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 scientist--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 (SQL 101)--This class will teach the student how to design, store, clean, query, and access data which is the key foundation to any projects success.

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.

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.

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.

How this works:

The ODSA is broken into two tracks, designed to fill two different roles at a company:

Data Analyst and Data Scientist

Tracks	Data Analyst			Data Scientist		
Class number	1	2	3	4	5	6
				Machine Learning/AI		
Class Name	Python 101	Data Manipulation and Management (SQL 101)	Data Visualization	Basic Model Building	Model Evaluation	Advanced Model Building

Tracks:

Data Analyst (or Business Intelligence) consists of the first 3 classes--teaching you to find data, save data, manipulate data, and visualize data. This will allow a person to better understand how to make data-driven decisions off of the data and be better prepared for the future with their recommendations.

Data Scientist builds off of Data Analyst and is the Machine Learning and AI part of the academy. Data Scientists are tasked with predicting the future and learning what to prescribe in an effort to affect the outcomes positively. These 3 classes teach you how to take data you have accumulated and visualized in the data analyst track and understand what is going to happen before it happens.

You are welcome to take all the classes (and receive a Certificate of the Fundamentals of Data Science), one track, a group of classes, or just one class to better your knowledge.

We are happy to help as you better your career.

Mission and Vision:

Our goal is to make Business Intelligence and Data Science sustainable in Omaha. That takes more than just knowledge, it takes a(n):

- Base knowledge of how to do:
 - Data Science Programming
 - Data Manipulation and Management
 - Data Visualization
 - Data Science Modeling including Machine Learning and AI
- 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

Our job is to make you successful in 6 months *and* in 6 years.

Omaha Data Science Academy 2019 and 2020 Calendar

Cohort 8 (approximately. Dates Subject to Change)

<u>Introduction to Data Science</u>	4 weeks--January 27th-February 19th
<u>Data Manipulation and Management (SQL 101)</u>	4 weeks--March 2nd-March 25th
<u>Data Visualization Using Tableau</u>	4 weeks--April 6th-April 29th
-----	-----
<u>Basic Model Building</u>	4 weeks--May 11th-June 8th
<u>Mathematics of Model Evaluation</u>	4 weeks--June 15th-July 8th
<u>Advanced Data Science Modeling in Python</u>	4 weeks--July 20th-August 12th

Cohort 9 (approximately. Dates Subject to Change)

<u>Introduction to Data Science</u>	4 weeks--June 1st-June 24th
<u>Data Manipulation and Management (SQL101)</u>	4 weeks--July 6th-July 29th
<u>Data Visualization Using Tableau</u>	4 weeks--August 10th-September 2nd
-----	-----
<u>Basic Model Building</u>	4 weeks--September 14th-October 7th
<u>Mathematics of Model Evaluation</u>	4 weeks--October 19th-November 16th
<u>Advanced Data Science Modeling in Python</u>	4 weeks--Nov. 23rd-December 16th

*There are no classes: New Year's Day, Memorial Day, July 4th, Labor Day, Veterans Day, Thanksgiving Wednesday and Thursday, or Christmas.

**If in the event of weather cancelation (snow, etc.), we will make the decision by noon on those days based on weather forecasted and predicted road conditions. If class is canceled, that class will be made up during the period between classes or at a different date based on a class vote.

Fundamentals of Data Science Certificate

Half 1: Data Analyst

1. Introduction to Data Science Programming (Python 101) **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 (SQL 101) **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. 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.

Half 2: Data Scientist

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.

5. 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.

6. 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.

7. 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.

Fundamentals Course Syllabus-Cohort 8

Introduction to Python and Data Science

Class 1 - Class Overview / 'Zero to One'

- Class Structure
- Learning Effectively
- Technology Review/Quiz
- Computer Anatomy
- Programming Concepts
- Digital Storage
- Structure of the Internet
- Careers in Tech
- Common Job Titles
- Team Structures
- Data Science Concepts
- Base Software / Installations (Anaconda, Editors, CLIs, etc)
- Bash / Shell Basics

Class 2 - Tooling Overview / Python Intro

- Bash / Shell Review
- Installation Checks
- Version Control (GitHub: Git, CLI, and Desktop)
- Data Management and Manipulation
- SQL / NoSQL
- Excel, Access, Airtable, BigQuery, etc.
- Modeling Languages (Python, R, Scala, Julia)
- Visualization (Tableau, Matplotlib, Plotly, Seaborn, D3)
- Platforms (Dataiku, DataRobot, C3 Oracle)
- Python Basics
- Structure and Syntax
- Full Language Dive
- Libraries / Adjunct code

Class 3 - Python / Help / First Project

- Python Review
- Code Dive
- Using/Building Libraries
- Mini-Projects
- Getting Help

- Web Resources / Using StackOverflow
- Pros/cons of video tutorials
- Creative Use of Python
- Build Your Own Project
- Guided / Exploratory Building

Class 4 - Python / Getting Data / Visualization

- Projects Presentation
- Importing Data with Python
- Working with CSV, JSON, etc
- Regex / re Library
- Basic Web Scrapers
- Web Page
- Weather Data
- Stock Data
- Visualizing Data with Matplotlib

Class 5 - Statistics / Data Processing in Python

- Intro to Statistics
- Processing Data in Python
- NumPy: Arrays / Uses
- Pandas: DataFrames / Functions
- Practice / Experimentation

Class 6 - Statistics / Pandas / Databases

- Deeper Dive into Statistics
- Pandas: DataFrames / Functions
- DBMS DataBase Management Systems
- SQL Structured Query Language
- SQLAlchemy

Class 7 - Intelligent Code

- NLP Natural Language Processing (NLTK, Genism)
- Machine Learning
- Algorithms Exploration
- Overview and Understanding
- Scipy and Scikit-learn
- Deep Learning (Keras, TensorFlow, Theano)
- Tensorflow Mini-Project

Class 8 - Bringing it All Together

- Concepts Test / Review
- Written / Multiple Choice
- Verbal
- Final Test Project:
- Web Scraping
- Data Processing

- Visualization
- Presentation / Communication

Data Manipulation and Management (SQL 101)

Class 1 - Basics

- The need for Data Manipulation
- The target (table with characterizations)
- Joins

Class 2 - SQL Code

- SQL code format
- SQL as a declarative language
- Table Structure and Field Type

Class 3- Joins in SQL

- Inner
- Left Outer

Class 4 - Using Joins

- Examples (interpret results)

Class 5 - Aggregation

- Aggregation of query results
- Combined with Joins

Class 6 - Simpson's Paradox

- Simpson's Paradox

Class 7 - SQL on a Platform

- Dataiku

Class 8 - Data Management in Toolsets other than SQL

- Other Tools - Cloud
 - Microsoft Azure
 - Google Big Query
 - AWS

Data Visualization Using Tableau

Class 1 - Intro to Tableau

Tableau Fundamentals

- Understanding the Tableau authoring interface, drag & drop analytics
- Key Tableau terminology and vocabulary

Connecting data to Tableau:

- Creating data sources
- How to work within data sources
- How to join/blend disparate data sources

Class 2 - Visualizing Your Data

How to build the top 3 essential data visualizations:

- Bar charts
- Line graphs
- Scatter Plots

Class 3 - Other ways to Visualize your data

How to build other useful charts in Tableau

- Heat maps
- Dual axis charts
- Highlight table
- Crosstabs
- Tree maps
- Gantt charts
- Pie charts

Class 4 - Visualizing Your Geospatial Data

- Geospatial data overview
- Creating maps in Tableau
- Custom geocoding
- Custom shapefiles in Tableau
- Using map layers
- Mapbox Integration

Class 5 - Analyzing data in Tableau

Calculated fields

- Arithmetic calculations
- Quick table calculations
- Aggregations
- Working with dates & times

Class 6 - Advanced Analysis in Tableau

Analytics Pane

- Lines - Constant / Average / Reference
- Median with Quartiles
- Box Plots
- K-means Clustering
- Trend Lines (Linear Regression)
- Forecast

Level of Detail Calculations

Advanced Data Charts

Class 7 - Key Tableau Technical Features

Parameters

Groups vs Sets

Hierarchies

View Data

Tooltips

Class 8 - Building data driven dashboards

Using the dashboarding interface

- Float vs tiled dashboards
- Adding images & webpages to dashboards
- Dashboard filters
- Dashboard actions

Dashboard design best practices

- How to build a 3 step drill down dashboards
- How to create a stories in Tableau

Basic Model Building

Class 1 - What is a Model

- What is a Model
- How is it built
- How is it consumed
 - API
 - The Resultant Table
- The differences between AI and ML

Class 2 - Review Model examples

- Review an existing example

Class 3- Building some models (Dataiku)

- Target variables

Class 4 - How to process a model

- Options for how to process model

Class 5 - Basics of the model evaluation

- Business
- Model scores

Class 6 - Choosing different algorithms

- Choosing different algorithms

Class 7 - Common Algorithms

- From scikit-learn

Class 8 - Putting it all together

Mathematics of Model Evaluation

Class 1 -- The metrics (part 1)

- Accuracy
- Precision

Class 2 -- The metrics (part 2)

- F1

- Recall
- Class 3 -- The metrics (part 3)**
 - AUC
 - Cost Matrix
- Class 4 -- The metrics (part 4)**
 - Cumulative Lift
- Class 5 -- Using the metrics**
 - Metrics and the business problem
 - How to discuss results
 - When is a ML model finished?
- Class 7 -- Putting it all together**
 - Build
 - Test
 - Evaluate
 - Adjust build
 - Re-test
 - Re-evaluate
- Class 8 -- The Results**
 - Present Results

Python and Advanced Modeling

- Class 1 - Applying Model Building to Python Programming**
 - Review Libraries
 - Review Structures
 - Review Data
- Class 2 - Building Word2Vec model**
 - Revisit Scraping data
- Class 3 - Model Building**
 - Code, Enhance, Review, Repeat
- Class 4 - Model Building on a different data source**
 - Refactoring the code
- Class 5 - Expand the Scikit Learn Library**
 - Review standard Library Codes
 - Run Dataiku models as Python codes
- Class 6 - Add Python to Dataiku flow**
 - Understand Results
- Class 7 - Python code to solve other gaps**
 - Calculations
- Class 8 - Other Data Manipulation**
 - Random Number Population (Benford's Law)