**Omaha Data Science Academy (OSDA)**

**Data Manipulation and Management**

**Cohort 7 (July 8th – July 31st, 2019)**

Instructor(s): Jeremy Bergmann, BS/MS

Phone: 402-321-7530; Email: [jeremy@jbanalyticsconsulting.com](mailto:jeremy@jbanalyticsconsulting.com); Office: By Appointment

Course Materials: [Google Drive - Data Manipulation and Management](about:blank)

GitHub: <https://github.com/nathan-booth/data-manipulation-and-management>

**COURSE DESCRIPTION:**

This class will teach the student how to design, store, clean, query, and access data, specifically to be used within data science projects. It will also teach database design, SQL programming, Relational Database modeling & schemas, Python & NoSQL, data extract, cleaning/transformation and load (ETL) techniques, and cloud-based platform offerings.   
  
The class will include a “class project”, where students create an end-product, using a MySQL database and BI Visualization tool (such as Tableau), from the knowledge gained in the course.

.

**Overall Approach:**

The overall approach to the course is to provide the background needed for data manipulation within data science projects, using the SQL programming language and relational & non-relational databases. The topics include: Data Manipulation basics, writing SQL Queries, Manipulating/Managing data and utilizing tools (Dataiku) to assist and automate the build and deployment of data-science projects.

**Class Schedule:**

**Class 1 - Data Manipulation - Basics**

Hour 1 - Introductions, Class Overview, Data Science Philosophy, Tool Setup

Hour 2 - Data Manipulation Background, Definitions & Database Basics - Lecture

Hour 3 - Class Project Intro & Data Selection

**Class 2: Writing Queries**

Hour 1 - Class 1 Quiz, SQL Queries, Data Types, Functions & Sub-queries – Lecture

Hour 2 - Utilizing MySQL Workbench - Working Examples

Hour 3 - Exercises/Working Session

**Class 3 – Manipulating Data - Part 1**

Hour 1 - Create Databases, Create/Alter/Delete Tables, Constraints – Lecture

Hour 2 - Importing Project Data into MySQL - Import/Export Wizard

Hour 3 - Exercises/Working Session

**Class 4 – Manipulating Data - Part 2**

Hour 1 - Class 2&3 Quiz, SQL Operators, Views, Case Statements & Logical Functions, Data Cleansing

Hour 2 - SQL Joins – Lecture & Examples

Hour 3 - Exercises/Working Session

**Class 5 – Storing & Managing Data**

Hour 1 - ETL & Data Storage, Data Mgmt/Database Admin, Data Gov. & Quality - Lecture

Hour 2 - Working Example - Python: ETL & Measuring Data Quality

Hour 3 - Exercises/Working Session

**Class 6 – Structuring & Modeling Data**

Hour 1 - Class 4 & 5 Quiz, ER & Relational Diagrams, Data Models, Normal Forms – Lecture

Hour 2/3 - Class Project – Data Modeling - ER & Relational Diagrams (draw.io)

**Class 7 - Tools for Data Manipulation and Management**

Hour 1-2 – Cloud Platforms (AWS, Azure & Google), NoSql, ETL & Workflow Tools - Lecture

Hour 3 – PowerBI Cloud/Desktop - Exercises/Working Session

**Class 8 – Class Project**

Hour 1-3 – Working Session – Finish Class Project