**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; E-Mail: [jeremy@jbanalyticsconsulting.com](mailto:jeremy@jbanalyticsconsulting.com); Office: By Appointment

GitHub: <https://github.com/jbergmann56/OSDA-DataManipulationAndManagement>

**COURSE DESCRIPTION:**

This class will teach the student how to design, store, clean, query, and access data, with a focus on use within data science projects. It will also teach database design, SQL programming, Relational Database modeling, introduction to Python & NoSQL, data cleaning/transformation and load (ETL) techniques and a variety of cloud-based storage offerings.   
  
Additionally, this class will include a “class project”, where students create an end-product, using raw data, a MySQL database and BI Visualization tool (such as Tableau). The class project will build from knowledge gained in the course and provides the student a working example, similar to what one could expect within a business environment.

.

**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 building 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 - Exercises & Class Project - Data Import

**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 & Class Project - Querying Data

**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 & Class Project - Storing Data for Analysis

**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 & Class Project - Cleaning Data for Analysis

**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 & Class Project - Analyze Data Quality & Create Data Dictionary

**Class 6 – Structuring & Modeling Data**

Hour 1/2 - Class 4 & 5 Quiz, ER & Relational Diagrams, Data Models, Normal Forms – Lecture

Hour 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 – Class Project - Data Automation, Connect to PowerBI Cloud/Desktop

**Class 8 – Class Project**

Hour 1-3 – Working Session – Finish Class Project