**Omaha Data Science Academy (OSDA)**

**Data Manipulation and Management**

**Callers To Coders (June 3rd – June 25th, 2020)**

Instructor(s): Jeremy Bergmann, BS/MS

Phone: 402-321-7530; E-Mail: [jeremy@jbanalyticsconsulting.com](mailto:jeremy@jbanalyticsconsulting.com);

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, data cleaning/transformation, data extract/load (ETL) techniques and a variety of cloud-based storage offerings (MySQL & NoSQL).   
  
Additionally, this class will include a “class project”, where students create an end-product, using raw data, a MySQL database and BI Visualization tool (Google Data Studio). The class project will build the knowledge gained in the course lectures and it 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 Business Intelligence (BI) and data science projects, via SQL programming language and databases technologies. The course topics include: Data Manipulation basics, writing SQL Queries, Manipulating, Managing, and cleaning data and utilizing tools 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 Introduction - Data Import

**Class 2: Writing Queries**

Hour 1 - SQL Queries, Clauses & Data Types – Lecture

Hour 2 - Functions & Sub-queries – Lecture

Hour 3 - Exercises & Class Project - Utilizing MySQL Workbench, Querying Data

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

Hour 1 - Class 1&2 Quiz, Views, Create Databases, Create/Alter/Delete Tables, Constraints – Lecture

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

Hour 3 - Class Project - Storing Data for Analysis

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

Hour 1 - SQL Joins – Lecture & Examples

Hour 2/3 - Exercises & Class Project - Joining Data for Analysis

**Class 5 – Cleaning Data Using SQL**

Hour 1 - Class 3&4 Quiz, Case Statements, SQL Operators & Logical Functions, Data Cleansing

Hour 2/3 - Exercises & Class Project - Cleaning Data for Analysis

**Class 6 – Structuring & Modeling Data**

Hour 1-2 - ER & Relational Diagrams, Data Models, Normal Forms – Lecture

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

**Class 7 – Managing Data**

Hour 1 - Class 5&6 Quiz, 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 8 - Tools for Data Manipulation and Management**

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

Hour 2 – Class Project – Import to Google Cloud MySQL, Answer Business Questions

Hour 3 – Working Session – Finish Class Project