**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).   
  
This class will include a COVID-19 dataset, where students can practice and apply their SQL knowledge via class assignments and test their knowledge through online quizzes (link).

.

**Overall Approach:**

The 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 relational database technologies. The course topics include: Data Manipulation basics, writing SQL Queries, Manipulating, Managing, and cleaning data, and utilizing tools to automate the 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 – Examples & Exercises - COVID Data Import

**Class 2: Writing Queries**

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

Hour 2 - Functions & Sub-queries – Lecture

Hour 3 - Examples & Exercises - Utilizing MySQL Workbench, Querying Data

**Class 3 – Extracting and Storing Data**

Hour 1 - Class 1&2 Quiz, Views, Create Databases, Create/Alter/Delete Tables, Constraints – Lecture Hour 2 - Importing Data into MySQL - Import/Export Wizard – Demographic Data

Hour 3 - Examples & Exercises - Storing Data for Analysis

**Class 4 – Joining & Merging Data**

Hour 1 - SQL Joins & Unions – Lecture & Examples

Hour 2/3 - Examples & Exercises - Joining Data for Analysis

**Class 5 – Cleaning Data Using SQL**

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

Hour 2 – Identifying & Removing Duplicates, Data Cleansing – String, Numeric & Date Functions

Hour 3 - Examples & Exercises - Cleaning Data for Analysis

**Class 6 – Structuring & Modeling Data**

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

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

**Class 7 – Storing & 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 - Examples & Exercises - 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 2/3 – Working Session – ETL Using Python, Import to Google Cloud MySQL, COVID-19 Dashboard