**Omaha Data Science Academy (OSDA)**

**Data Foundations – Business Intelligence**

**Data Manipulation and Management**

**ODSA Cohort 10 (October 18rd – November 10th, 2021)**

Instructor(s): Jeremy Bergmann, BS/MS/MBA

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. The topics in this class include: SQL programming, data cleansing/ transformation, data extract/load (ETL) techniques, advanced database functions (stored procedures, column functions, database triggers, temporary tables), Relational Database modeling and design, and an overview of cloud-based data storage offerings (NoSQL, Hadoop, Cloud SQL, etc.).   
  
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 (Github, MySql Workbench)

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

Hour 3 – Examples & Exercises – COVID Data, WorldDB

**Class 2: Writing Queries**

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

Hour 2 - Column and Scalar Functions – Lecture

Hour 3 - Examples & Exercises - Querying Data, Class Project – Overview, Queries

**Class 3 – Extracting & Storing Data**

Hour 1 - Class 1&2 Quiz, Column & Scalar Functions: Class 2 Deep-Dive

Hour 2 - DDL - Create Create/Alter/Delete Databases & Tables, Views, Constraints – Lecture

Hour 3 - DMM – Inserting, Deleting, and Updating Data - Lecture, Exercises, Project - Importing Data

**Class 4 – Joining & Merging Data**

Hour 1 - SQL Joins & Unions – Lecture

Hour 2 - Subqueries - Lecture, Examples & Exercises - Joining & Merging data

Hour 3 - Class Project – Joining & Merging Data for Analysis

**Class 5 – Cleaning Data Using SQL**

Hour 1 - Class 3&4 Quiz, Cleaning Data – Overview, SQL Wildcards, Identifying & Removing Duplicates

Hour 2 – Data Cleansing – String, Numeric & Date Functions

Hour 3 – Examples & Class Project - Cleaning Data for Analysis

**Class 6 – Database Design - Structuring & Modeling Data**

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

Hour 2 – Examples & Exercise – OLTP Data Modeling - ER & Relational Diagrams

Hour 3 – Class Project – ERD Diagram, OLAP Dimensional Modeling

**Class 7 – Storing & Managing Data**

Hour 1 – Class 5&6 Quiz, Enterprise Data Overview, ETL & Data Storage, Data Mgmt/DB Admin - Lecture

Hour 2 - Data Gov. & Stewardship, Data Quality - Lecture

Hour 3 - Class Project – Data Quality, Data Dictionary, Extract/Transform/Load (ETL) Pipeline

**Class 8 – Advanced Database Functionality**

Hour 1 – Temporary Tables, SQL Procedural Language - Stored Procs, Column Functions, and Triggers

Hour 2-3 - Finish Class Project - Connect MySQL to PowerBI Cloud, Answer Business Questions

**Bonus - Tools for Data Manipulation and Management**

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

Hour 2 – Final Exam

Hour 3 – Working Session – Using Google Cloud MySQL/Data Studio, MS Azure, COVID-

19 Dashboard