RARITAN VALLEY COMMUNITY COLLEGE
ACADEMIC COURSE OUTLINE

CSIT 210 Data Management

I. Basic Course Information

A. Course Number and Title: **CSIT 210 Data Management**

B. New or Modified Course: **New**

C. Date of Proposal: Semester: **Fall** Year: **2024**

D. Effective Term: Fall 2025

E. Sponsoring Department: **Mathematics and Computer Science**

F. Semester Credit Hours**: 4**

G. Weekly Contact Hours**:** 5 Lecture**: 3**
Laboratory**: 2**
Out of class student work per week: **7**

H. ☒ Prerequisite (s): **CSIT 105 Foundations of Computer Science or CSIT 107 Introduction to Data Science**
☐ Corequisite (s):

I. Additional Fees**:** None

II. Catalog Description

Prerequistes: CSIT 105 Foundations of Computer Science or CSIT 107 Introduction to Data Science.
This course will provide students with the knowledge and skills needed to acquire and curate real world data using Python and some of its libraries. This includes cleaning, modifying and transforming data to suitable formats for data analysis. Visualizations of clean data are used to lead to discovery of patterns to help make predictions. Students will also learn how to manage large datasets by creating relational databases and doing analysis using SQL and noSQL.

III. Statement of Course Need

A. Data science has been and still is an up and coming area in many disciplines. This course will teach students how to manage data and utilize data libraries using tools such as Python and SQL.

B. This course has a lab component. The lab is essential for providing students with applied hands-on experience with data management using a variety of tools. Students are required to use the software in the Computer labs to complete their assignments.

C. This course generally transfers as a Data Science program requirement dependent on transfer institution.

IV. Place of Course in College Curriculum

A. Free Elective
B. This course serves as a Programming Elective in the Computer and Programming Electives List.
C. This course meets a program requirement for the Data Science Option in the Computer Science AS degree.
D. To see course transferability: a) for New Jersey schools, go to the NJ Transfer website, www.njtransfer.org; b) for all other colleges and universities, go to the individual websites.

V. Outline of Course Content

A. Introduction to Data Science / Data Management
B. Python Basics
    1. Variables, strings and data structures such as lists, tuples, sets and dictionaries
    2. File I/O and Exception Handling
C. Python Libraries
    1. Numpy
    2. Pandas Series and Data Frames with GroupBy
    3. Getting Data from csv files
D. Cleaning, Transforming and Visualizing Data
E. Regular Expression Fundamentals and Applications of them
F. SQL Basics
    1. Queries and joins
    2. Advanced Queries
G. Relational databases, SQL and noSQL
H. Analysis and Prediction with Linear Regression

VI. A. Course Learning Outcomes:

At the completion of the course, students will be able to:

1. Acquire and curate real world data. This includes cleaning, modifying and reshaping data to suitable formats for data analysis. (GE.4)
2. Analyze, discover and use visualizations to describe patterns in large datasets. (GE.4 *)
3. Manage large datasets using databases. (GE.4)
4. Make predictions based on data analysis using Python. (GE.4 *)

* critical thinking

B. Assessment Instruments

1. Research Paper (suggested)
2. Projects / Group Projects
3. Labs / Coding Challenges
4. Exams
5. Quizzes

VII. Grade Determinants

A. Homework
B. Labs
C. Exams / Quizzes
D. Projects

Given the goals and outcomes described above, LIST the primary formats, modes, and methods for teaching and learning that may be used in the course:

A. lecture/discussion
B. small-group work
C. computer-assisted instruction
D. laboratory
E. student collaboration

VIII. Texts and Materials

A. Suggested textbook:

Python for Data Analysis, 3E open edition: https://wesmckinney.com/book/
Murach's Python for Data Analysis, Scott McCoy

B. Code in Jupyter Notebooks
C. Videos supplied or found by instructor
D. Power points supplied by instructor

Please Note: The course outline is intended only as a guide to course content and resources. Do not purchase textbooks based on this outline. The RVCC Bookstore is the sole resource for the most up-to-date information about textbooks.

IX. Resources

Jupyter Notebook with Python under the Anaconda umbrella will be installed for students to use during lecture and lab time.

X. Check One: ☐Honors Course ☐Honors Options ☒ N/A