

**Part 2**

# Department Master Syllabus

**Camden County College**

**Blackwood, New Jersey**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Course Number:**  DSC-102 | | **Course Title:**  Data Science II | | | |
| **Department/Program:** Math/Data Science | | | | | |
| **Date of Review:** March | | 2020 | | | |
| (This Department Master Syllabus has been examined by the program/department faculty members and it is decided that no revision is necessary at this time.) | | | | | |
| **Date of Revision:** New Course, March | | | | 2020 | |
| (This Department Master Syllabus has been examined by the program/department faculty members and it is decided a change requiring a revision is necessary at this time.) | | | | | |
| N.B. A change to the course materials alone (textbooks and/or supplementary materials) may not constitute a revision. Any other change to the items listed below on this form is considered a revision and requires approval by the department/program faculty at a department/program meeting and by the division at a Chairs and Coordinator meeting. | | | | | |
| **Credits:**3 | | | | | |
| **Contact Hours** | **Lecture:** 2 | | **Lab:** 2 | | **Other:** 0 |
| Prerequisites: DSC-101, CSC-106 | | | | | |
| Co-requisites: CSC-171 | | | | | |
| Course Description: This course is the second course in a three-course sequence that examines the core activities that a professional data scientist performs throughout the data science workflow.  Building on the knowledge attained in DSC-101, this course will focus on the data preparation phase by applying standard data wrangling practices to collections of raw data gathered from a variety of sources and mapping it into a coherent form for analysis.  Topics include ‘dirty’ raw data issues of inconsistent naming, and formatting, missing data and outliers, as well as reshaping and pivoting data as appropriate. The student will examine real-world case studies and use modern data preparation tools and programming techniques to complete data wrangling.  Topics in this course are well-aligned with those of CSC-171. | | | | | |
| **Student Learning Outcomes (SLOs)**  Course specific student learning outcomes  Upon completion of this course the student will be able to:   * Discover and extract the significant data from large sets of raw data from a variety of sources as assessed by homework, quizzes, tests, and projects * Practice using modern software tools and programming techniques to clean and prepare data for analysis as assessed by homework, quizzes, tests, and projects * Create new structured datasets from raw sources by cleaning the data through combining, renaming, reformatting, omitting and merging data fields as needed for analysis as assessed by homework, quizzes, tests, and projects * Analyze and verify the consistency, accuracy, quality and security of the clean data as assessed by homework, quizzes, tests, and projects   As assessed by:  Students will be evaluated on the degree to which student learning outcomes are achieved. A variety of methods may be used such as tests, class participation, programming projects, homework assignments, online learning tools, etc. | | | | | |
| **General Education Student Learning Outcomes**  If this course has applied for General Education Elective Status the general education student learning outcomes listed below must exactly match those the sponsor has identified on the General Education Request form.  General Education SLOs:  N/A  As assessed by:  N/A | | | | | |
| **Program Learning Outcomes**  List all course level student learning outcomes that interconnect to a particular program learning outcome.  All CSLOs target the following DSC.AAS PSLOs  Develop solid analytical reasoning, critical thinking and technical skills in order to extract, mangle, analyze and present data for multiple disciplines to broad audiences that follow professional standards to enhance understanding and decision-making.  Demonstrate the ability to work independently and as a member of a team with modern technical tools to accomplish data life cycle project goals and meet deadlines.  Describe the assessment of the interconnected program learning outcome(s).  Various course level assessment instruments will be used to target specific program learning outcomes. | | | | | |
| **Course Outline:**  I. Overview of Exploring Data   * 1. Diagnose data for cleaning      1. Loading and viewing data      2. Reading and writing data      3. Interacting with Database   2. Visual exploratory data analysis      1. Plotting with pandas and seaborn: line plots, bar plots, histograms and density plots, scatter or point plots, facet grids and categorical data      2. Other Python Visualization Tools   II. Data Cleaning and Preparation   * 1. Handling missing data      1. Filtering out missing data      2. Filling in missing data   2. Data transformation      1. Removing duplicates      2. Transforming data using a function or mapping      3. Replacing values      4. Renaming axis indexes      5. Detecting and filtering outliers      6. Permutation and random sampling      7. Computing indicator/Dummy variables   3. String manipulation      1. String object methods      2. Regular expressions      3. Vectorized string functions in pandas   III. Data Wrangling: Join, Combine, and Reshape   * 1. Hierarchical indexing      1. Reordering and sorting levels      2. Summary statistics by level      3. Indexing with a DataFrame’s columns   2. Combining and merging Datasets      1. Database-stype DataFrame joins      2. Merging on index      3. Concatenating along an axis      4. Combining data with overlap   3. Reshaping and Pivoting      1. Reshaping with Hierarchical indexing      2. Pivoting “long” to “wide” format      3. Pivoting “wide” to “long” format   IV. Overview of Modeling in Python   * 1. Creating model descriptions with Patsy   2. Introduction to statsmodels   3. Introduction to scikit-learn   V. Data Analysis Case studies | | | | | |
| **Course Activities:**    The classroom activities will include formal and informal lectures and structured, supervised active learning laboratory sessions. During lectures, new material and assigned problems will be explained. Students are encouraged to contribute to the discussion and to ask questions about the material. Active laboratory learning sessions will include individual and team projects that use individual and collaborative modern software tools to examine, develop and report on data science case studies. | | | | | |
| **Course Materials:**  Textbook(s): TBD  Supplemental Materials: **:**  Software Licenses: Free software tools  Computers: Students will need access to a computer for assigned projects and homework | | | | | |
| **Course Assessment Plan**  How often and by what means will the effectiveness of this course as part of the curriculum be assessed?    Assessment cycle to be determined by the members of the department. Students will be evaluated on the degree to which student learning outcomes are achieved. Assessment instruments may be in the form of tests and/or programming projects. | | | | | |