

SOUTHEAST COMMUNITY COLLEGE
CONSTRUCTION MANUFACTURING AND TECHNOLOGY DIVISION
Computer Information Technology Program
Date: August 2024

I. CATALOG DESCRIPTION

Course Number: INFO2578
Course Title: Data Analysis with Python
Prerequisite(s): MATH1020 or higher, and either INFO1414 or INFO1434 or CSCI2311
Catalog Description: This course covers analyzing and visualizing data using the Python programming language. Topics include preparing data sets, creating visualizations, predictive analysis and an introduction to machine learning.
Credit Hours: 3
Class Hours: 45
Lab Hours: 0
Total Contact Hours: 45

II. COURSE OBJECTIVES: *Course will:*

- A. Review the Python programming language.
- B. Introduce Jupyter Notebooks.
- C. Introduce the Pandas and Seaborn Python libraries.
- D. Demonstrate data visualization techniques.
- E. Demonstrate retrieving, importing and cleaning data.
- F. Introduce data analysis methods.
- G. Demonstrate predictive analysis using regression models.
- H. Introduce machine learning concepts.

III. STUDENT LEARNING OUTCOMES AND GENERAL EDUCATION LEARNING OUTCOMES:

- A. Student Learning Outcomes: *Student will be able to:*
 - 1. Develop Jupyter Notebooks with Python.
 - 2. Use Pandas data objects.
 - 3. Retrieve data from files, databases and web APIs.
 - 4. Import, clean and process data with Python.
 - 5. Create data visualizations with Pandas and Seaborn.
 - 6. Develop and verify hypotheses about datasets.
 - 7. Develop linear and multiple regression models for predictive analysis.
 - 8. Describe and demonstrate basic machine learning techniques.
- B. General Education Learning Outcomes (GELOs)
 - 1. GELO #3: Critical Thinking & Problem Solving
Outcome 2: Synthesize information to arrive at reasoned solutions to problems

IV. CONTENT/TOPICAL OUTLINE

- A. Python review and tooling
 - 1. Python fundamentals
 - 2. Jupyter Notebooks
- B. Using Pandas and Seaborn
 - 1. Pandas DataFrames
 - 2. Plotting with Pandas
 - 3. Plotting with Seaborn

- C. Retrieving data
 - 1. Importing data into DataFrames
 - 2. Flat files
 - 3. Databases
 - 4. JSON web APIs
- D. Cleaning data
 - 1. Identifying problems in datasets
 - 2. Removing or replacing invalid data
 - 3. Dealing with outliers
- E. Preparing data
 - 1. Reshaping data
 - 2. Applying functions to datasets
- F. Analyzing data
 - 1. Melting columns
 - 2. Grouping and aggregating data
 - 3. Pivot tables
 - 4. Binning data
 - 5. Forming and testing hypothesis
- G. Predictive Analysis
 - 1. Linear regression
 - 2. Multiple regression
- H. Machine Learning
 - 1. Supervised Learning
 - 2. Classification

V. **INSTRUCTIONAL MATERIALS**

- A. Required Text(s): *Murach's Python for Data Analysis*, (Refer to CID and/or instructor for current edition)
- B. Other Resources: Required software available as a free download
- C. Computer and Internet access

VI. **METHODS OF PRESENTATION/INSTRUCTION**

- A. Methods of presentation typically include a combination of the following:
 - 1. Material is presented in a web-based training format
 - 2. Engaged learning activities

VII. **METHODS OF EVALUATION**

- A. Methods of evaluation, although determined by the individual instructor, traditionally includes a combination of the following:
 - 1. Programming assignments
 - 2. Data Analysis Projects
 - 3. Quizzes
 - 4. Tests