Miami Dade College

Course Description

CAP3321C | Data Wrangling | 4.00 Credits

This course offers a broad introduction to data wrangling, data retrieval, and the use of Python in data analytics. Students will use Python and other business intelligence tools to retrieve data from various sources, clean the dataset, and prepare it for data analysis tasks. Prerequisites: CAP1788 and CAP2761C.

Course Competencies:

Competency 1: The student will demonstrate an understanding of the importance of data wrangling by:

1. Explaining the need for high-quality and accurate data in analytics
2. Describing the typical data wrangling steps
3. Identifying various tools and languages that can perform data wrangling
4. Explaining why Python and various associated packages are ideal for data wrangling

Competency 2: The student will demonstrate an understanding of Python data structures and file management by:

1. Identifying the four built-in data types in Python that are used for data collection (lists, sets, tuples, dictionaries) and writing code using these data types
2. Describing the function of advanced data types such as stacks and queues
3. Opening, reading, writing, and closing a file

Competency 3: The student will explore various Python packages for data analytics by:

1. Creating an array and performing mathematical operations on it using NumPy
2. Creating series and data frames using Pandas
3. Performing basic column and row operations on data frames
4. Creating various charts using Matplotlib
5. Applying various statistical measures such as min, max, and median

Competency 4: The student will demonstrate an understanding of data cleaning and preparation using Python by:

1. Describing the importance of extracting only the subset of a dataset required for a specific business question
2. Sorting, filtering, indexing, and grouping a dataset
3. Finding missing values, duplicate values, and outliers, and performing various operations to handle them. Applying functions to entire rows and columns
4. Identifying the data types in each column and converting one to another when necessary
5. Concatenating and joining various datasets
6. Using Pandas methods to create pivot tables and perform additional data manipulation/ transformation tasks

Learning Outcomes:

- Use quantitative analytical skills to evaluate and process numerical data
- Solve problems using critical and creative thinking and scientific reasoning
- Use computer and emerging technologies effectively