**Course Abstract**

***If you need accommodations due to a disability, contact Disability Services in***

***Edison Hall Room 100, 732.906.2546.***

***To foster a productive learning environment, the College requires that all students adhere to the Code of Student Conduct which is published in the college catalog and website.***

**Course ID and Name: DSA 120 – Foundations of Data Collection and Cleansing**

**Department: Business and Computer Science**

Chairperson or Course Coordinator: Dr. Aslihan Cakmak

Office Location: ED123

E-mail Address: [ACakmak@Middlesexcc.edu](mailto:ACakmak@Middlesexcc.edu)

Telephone: 732-906-2526

**Prerequisites:** CSC 106 and DSA 110

**Co-requisites:** None

**Course Description:**

Students will be introduced to the data processing method. Students will learn how to collect raw data, cleanse, extract, sort, process, analyze, store and present the data in a readable format. Students will use various techniques to collect and cleanse data from different sources available. The use of common data collection and cleansing software will be integrated throughout the course.

**General Education Status:** N/A

**Credits: 3 Lecture Hours: 2 Lab Hours: 2**

**Learning Outcomes:**

1. Identify locations where data is collected
2. Identify the types of data commonly collected
3. Explain how to locate data files stored at different locations.
4. Explain the similarities and differences in data based on its’ file format
5. Demonstrate how to collect, extract, and store data
6. Demonstrate how to parse and filter data
7. Explain and demonstrate common data cleansing techniques
8. Demonstrate how to extract and organize cleaned data
9. Analyze data through basic statistical functions
10. Evaluate the level of data quality
11. Present outcomes determined from cleaned and organized data

**Upon successful completion of this course, a student will be able to:**

1. Identify locations where data is collected.
2. Identify the types of data commonly collected.
3. Explain how to locate data files stored at different locations
4. Explain the similarities and differences in data based on its’ file format
5. Demonstrate how to extract and organize cleaned data
6. Analyze data through basic statistical functions
7. Evaluate the level of data quality
8. Present outcomes determined from cleaned and organized data

**Course Content Areas:**

1. Identifying the problem and the data needed
2. Locating the data needed including local files, server files, cloud-based files
3. Understanding data and file-formats including spreadsheets, database tables, PDFs, flat files, text files
4. Collecting the data based on file formats
5. Extracting and storing the data needed
6. Local files
7. Server files
8. Cloud files
9. Parsing and filtering data
10. Data cleansing techniques
    1. Identifying values for cleanup
    2. Determining and cleaning bad data
    3. Identifying outlier data
    4. Formatting data
    5. Identifying duplicate data
11. Standardizing and scripting cleaned data
12. Exploring and extracting cleaned data
13. Searching for content
14. Sorting and subtotaling
15. Creating subsets
16. Analyzing cleaned and organized data
    1. Basic statistical functions
    2. Interpreting results
17. Evaluating data quality
18. Determining cleaning conclusions
19. Documenting cleansing conclusions
20. Presenting data
    1. Identifying what is to be presented
    2. Formatting
    3. Charting
    4. Uploading data