**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 210– Foundations of Data Analytics**

**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:** DSA 120

**Co-requisites:** MAT 285

**Course Description:**

Focusing on the programming of data science and common techniques of analytics, students will learn to use data mining software to develop data analysis projects including problem formulation, collecting relevant data, wrangling and formulating data, applying data analysis techniques, aggregating and visualizing data, and presenting results. Students will write programs to analyze various sized structured and unstructured data sets.

**General Education Status:** N/A

**Credits: 4 Lecture Hours: 2 Lab Hours: 4**

**Learning Outcomes:**

1. Identify the problem
2. Identify necessary data sets
3. Validate data sets
4. Perform data wrangling
5. Transform data to necessary formats for analysis
6. Extract data sets to retrieve good sample sets
7. Summarize data sets
8. Perform exploratory data analysis
9. Build functions to perform common tasks
10. Build a model based on a dataset
11. Visualize data to communicate results

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

1. Identify the problem
2. Identify necessary data sets
3. Validate data sets
4. Perform data wrangling
5. Transform data to necessary formats for analysis
6. Extract data sets to retrieve good sample sets
7. Summarize data sets
8. Perform exploratory data analysis
9. Build functions to perform common tasks
10. Build a model based on a dataset
11. Visualize data to communicate results

**Course Content Areas:**

1. Problem identification and preliminary analysis
2. Identifying data needed
3. Programming language introduction and syntax
4. Programming language tools and functions
5. Formulating data questions
6. Determining the data model wanted

* Simple models
* Visualizing models
* Formulas and model families

1. Determining the necessary data wrangling techniques

* Data frames
* Tibbles
* Tidy data
* Relational data

1. Extracting the data through programming
2. Writing data functions specific to the programming language
3. Modeling the results generated through programming
4. Performing the data analysis
5. Formatting the results based on programming language formats
6. Building graphics based on programming language tools available
7. Visual representations of the results and recommendations