

## DATA 301 - Introduction to Data Science

Fall 2018

### I. Catalog Description

#### DATA 301 Introduction to Data Science (4)

Introduction to the field of data science and the workflow of a data scientist. Types of data (tabular, textual, sparse, structured, temporal, geospatial), basic data management and manipulation, simple summaries, and visualization. 3 lectures, 1 laboratory.

Prerequisite: CPE/CSC 202; and one of the following: STAT 302, STAT 312, or STAT 313.

### II. Required Background and/or Experience

Prerequisite: CPE/CSC 202; and one of the following: STAT 302, STAT 312, or STAT 313.

Basic knowledge of statistics and the use of computers.

### III. Expected Outcomes

The student should:

- A. Identify different types of data used in data science and know their properties
- B. Write programs implementing key data manipulation, management, and analytic tasks
- C. Choose techniques to solve common data analytic tasks, and perform appropriate (simple) data analyses with given data
- D. Visualize, demonstrate and explain the results of data analyses to customers/data owners

### IV. Text and References

Recommended Texts:

- 1) McKinney, Wes. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. "O'Reilly Media, Inc.", 2012.
- 2) Liu, Bing. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. "Springer", 2011.
- 3) Mamoulis, Nikos. *Spatial Data Management*. "Morgan & Claypool Publishers", 2012.

### V. Minimum Student Materials

USB flash drive.

### VI. Minimum University Facilities

Availability of computing facilities.

**VII. Expanded Description of Content and Method**

<b>Content</b>	<b>Number of lectures</b>
A. <b>Introduction</b> .....	<b>4</b>
1. The Data Science workflow process	
2. Introduction to Python	
B. <b>Work with Tabular Data</b> .....	<b>8</b>
C. <b>Work with Time Series Data</b> .....	<b>8</b>
D. <b>Work with Textual Data</b> .....	<b>4</b>
E. <b>Work with Structured Data</b> .....	<b>4</b>
F. <b>Work with Market Basket Data</b> .....	<b>4</b>
G. <b>Work with Geospatial Data</b> .....	<b>8</b>
	<b>Total           40</b>

**Method**

Largely lecture with computer demonstrations of methods and problems, class discussion, supervised computer lab work and in-class exercises. Material from references and additional problems supplement the text.

**VIII. Method of Evaluating Outcome**

Problem and programming homework assignments, examinations, and projects.