

ONE PAGE COURSE OVERVIEW

Course Title: D206 – Data Cleaning

Course Description:

Data Cleaning is a data analytics course that continues to build proficiency in the data analytics life cycle with data preparation skills and knowledge. This course addresses exploring, transforming, and imputing data as well as handling outliers. Simply put, you will be getting a data set to clean by removing incomplete, irrelevant, corrupt, faulty, or dirty data.

Competencies:

- a. Predicting Obstacles in Data Analysis
- b. Preparing Data for Analysis
- c. Manipulating Data for Analysis

Performance Assessment (PA): Task covering the above competencies and involving:

- a. Explore various graphs and statistics to identify outliers
- b. Consider various methods to handle missing data, such as imputation
- c. Explore a basic use of principal component analysis (PCA) for data reduction of a set of variables.

Tools and Techniques: Python or R. (Students can choose).

Resources:

- a. Courseware content lessons 1- 40 primarily came from the text.
- b. **Text:** Larose, C. D., & Larose, D. T. (2019). Data science using Python and R. John Wiley & Sons. ISBN: 978-1-119-52684-1.
- c. LinkedIn Learning Videos.
- d. PluralSight Videos.
- e. Knowledge Check: Interactive labs, multiple choice questions, etc