Module 3 | Overview

OVERVIEW

This module, we move forward with our journey into data acquisition. Before you can work with your data, you have to collect it; and data comes from many sources and different formats. We will first learn how to obtain data sets from raw text files. We begin exploring preprocessing and using the scikit learn and pandas libraries.

LEARNING OUTCOMES

This module will support the following course level learning outcome(s):

* Acquire, clean, and organize data from a variety of sources, including raw data files, SQL databases, and online repositories using the Python programming language.
* Apply toolkits to preprocess large datasets for input to statistical and machine learning algorithms, including methods of feature extraction and dimensionality reduction.
* Apply statistical and machine learning toolkits to large datasets, including applications of regression, classification, and clustering.
* Perform simple visualizations of data.

ASSIGNMENTS

This is an approximation of the amount of time each activity will take the average student. If you find you are taking more time than suggested, please contact me for help.

| ASSIGNMENT ORGANIZER | |
| --- | --- |
| **Activity** | **Time (Hrs)** |
| Study/Reflect Time | 1:00 |
| Participate in [Reading & Discussion: Scikit Learn Chapter 5 (Pages 342-359)](https://elearning.mines.edu/courses/52392/discussion_topics/256385) | 1:00 |
| Watch & Study [Instructional Videos & Learning Materials: Scikit Learn](https://elearning.mines.edu/courses/52392/pages/instructional-videos-and-learning-materials-scikit-learn) | 1:00 |
| Participate in [Reading & Discussion: Pandas Basics](https://elearning.mines.edu/courses/52392/discussion_topics/256386) | 1:00 |
| Watch & Study [Instructional Videos & Learning Materials: Pandas Basics](https://elearning.mines.edu/courses/52392/pages/instructional-videos-and-learning-materials-pandas-basics) | 1:00 |
| Watch & Study [Instructional Videos & Learning Materials: Data Exploration and Data Acquisition](https://elearning.mines.edu/courses/52392/pages/instructional-videos-and-learning-materials-data-exploration-and-data-acquisition) | 1:30 |
| Submit [Activity: Coding Checkpoint - Dictionaries, Pandas, File IO Quiz](https://elearning.mines.edu/courses/52392/quizzes/74250) | 0:30 |
| Submit [Activity: Linear Regression with Boston Data](https://elearning.mines.edu/courses/52392/assignments/354735) | 2:00 |
| Review [Semester Project Introduction](https://elearning.mines.edu/courses/52392/pages/semester-project-introduction) | 1:00 |
| Submit [Activity: Semester Project - Getting Started and Team Formation](https://elearning.mines.edu/courses/52392/assignments/354721) | 2:00 |
| Submit [Project 4: Pandas and Data Acquisition](https://elearning.mines.edu/courses/52392/assignments/354744) | 3:00 |
| **Total Hours:** | **15:00** |