



CONTINUING EDUCATION

Course Outline

Data Mining

With Data Mining, businesses can extract hidden information and knowledge from large datasets and build models from historical data to predict future behaviors. In this course, you will learn to manipulate, process and clean the dataset using Python data analysis libraries; and perform statistical analysis on this mined and clean data to solve data analysis problems using real world scenarios.

Who Should Take This Course?

This hands-on course is designed for those who are interested in practical data analytics to solve real business problems. This includes identifying existing data, cleaning messy data, and identifying opportunities to collect additional data to solve the problem at hand, popularly referred to as “data munging.” For this course you will use Python. For data mining in Python, the Pandas library will be the main subject. It is a common library that is frequently used for data manipulation. Data Mining is the sixth required course in the Data Analyst Certificate.

Course Objectives

- Describe the exploratory data analysis (EDA) method
- Manipulate, merge, join, melt and slice the data in the data frame similar to how you would interact with a database
- Use time series data
- Create visualizations to explore the data using Python with matplotlib
- Describe the data using conventional statistical methods such as the mean, median, variance and standard deviation

Course Info

- Length: 18 hours
- Format: Classroom instruction, demonstration and labs
- Prerequisite: Python for Data Analysis

Course Content

Describe the exploratory data analysis (EDA) method

- Detail the investigative phase of EDA
- Communicate how EDA enables the discovery of patterns and anomalies
- Explain how EDA is used to test hypotheses
- Describe how to check assumptions using summary statistics and visualizations

Manipulate, merge, join, melt and slice the data in the data frame similar to how you would interact with a database

- Understand how Online Analytical Processing (OLAP) works
- Convert tabular data into multi-dimensional arrays
- Slice the data frame by column
- Slice the data frame by location
- Slice the data frame using filters
- Use multi-index slicing
- Make long data wide
- Make wide data long
- Create a new data frame with the transformed variables
- Impute missing values
- Manipulate data using groupby
- Join and combine the data frames
- Manipulate data using Regex

Use time series data

- Identify date time types
- Use up-sampling and interpolation
- Use resampling

Create visualizations to explore the data using Python with Matplotlib

- Understand matplotlib, a special library for visualization in Python
- Identify nominal/categorical, ordinal, interval and ratio data in the dataset
- Use visualizations to describe nominal data
- Use visualizations to describe ordinal data
- Use visualizations to describe interval and ratio data
- Use visualizations to describe correlations in the data
- Select appropriate visualization techniques for categorical data
- Select appropriate visualization techniques for interval and ratio data
- Analyze the data visually to identify if any of the variables are correlated

Describe the data using conventional statistical methods such as the mean, median, variance and standard deviation

- Calculate descriptive statistics for the variables in the data frame
- Review the descriptive statistics

Assessment /Evaluation

This course is not assigned a letter or numerical grade. However, in some cases, skill assessments may be administered during the course to gauge progress and comprehension.

Course Completion / Continuing Education Unit (CEU) Letter

If your company requires proof of course completion, or if you would like to have proof for your own records, you may request a Course Completion/Continuing Education Unit (CEU) letter **after the final session of your course**. You must attend* 80 percent of the course (or 100 percent of a single-day course) to qualify for the letter. You can [request a Course Completion/CEU letter online](#).

Certificates of Completion for Certificate Programs

To receive a Certificate of Completion for the Data Analyst Certificate, you must take all eight required courses; attend 80 percent of each course (calculated by number of sessions in each course), except the Capstone courses, where 100 percent attendance is required; participate in class; and successfully complete the SQL Server Capstone project and the Data Analyst Certificate Capstone project by the deadlines given. (Please note: Saturday sessions lasting longer than four hours will count as two sessions.) If you complete a certificate program by taking courses individually (not through a series), you can [request a Certificate of Completion online](#).

*Attendance for CEU letters and certificate programs is verified via the sign-in sheets provided at each class session. It is your responsibility to sign-in.

Green River College is committed to providing access to all who visit, work and study on campus. The College will provide reasonable accommodations for individuals with disabilities, with advance notice of need. If you require accommodations, please contact Disability Support Services as soon as possible to determine eligibility and/or request accommodations. Accommodations are determined on a case-by-case basis. Please contact Disability Support by email at dss@greenriver.edu; by phone at 253-833-9111, ext. 2631; TTY 253-288-3359; or in person at the Student Affairs and Success Center, Room 210, to request accommodations. For additional information, please visit www.greenriver.edu/dss.

The accommodations authorized on your forms should be discussed with your instructor. All discussions will remain confidential. Accommodations are not provided retroactively, so it is essential to discuss your needs at the beginning of the quarter. Additionally, only accommodations approved by Disability Support Services will be provided. This syllabus is available in alternate formats upon request. Green River College is an equal opportunity educator and employer. Learn more at www.greenriver.edu/accessibility.