CMPT 1520: Data Science I with Python

EFFECTIVE DATE

March 2021

DEPARTMENT

Computers - City Centre

DESCRIPTION

In this course, you will be introduced to fundamental data science concepts using the Python programming language. The course will review the major data science knowledge areas, including data visualization, linear algebra, statistics, and probability. The course ends with practical how-to material on both retrieving and preparing data for subsequent levels of analysis. Students will interact with a subject matter expert for 10 synchronous one-hour sessions and are expected to engage in self-directed learning for approximately two hours per week. This training assumes that the student already has at least basic programming skills. Please note: one of the course modules provides a crash introduction to Python for those that need to review Python or are switching to Python from another language. Required: A computer running Windows, MacOS, or Linux; stable high-speed internet connectivity; a microphone and webcam (ideal).

CREDITS

0.0

YEAR OF STUDY

Continuing Professional Development

PREREQUISITES

Either: CMPT 1510 (Introduction of Python Programming); or equivalent prior programming knowledge/experience.

COREQUISITES

None

COURSE LEARNING OUTCOMES

Upon successful completion of this course, students will be able to:

- List the major recent technical advances that led to the identification of Data Science as a distinct body of knowledge.
- Use Matplotlib for a range of data visualizations.
- Demonstrate basic familiarity of linear algebra (vectors and matrices), calculate the dot product of two

vectors, adding matrices, and multiply a vector by a matrix.

- Calculate the mean and standard deviation of a data sample and determine the correlation between two random variables.
- Explain the concept of a PDF (probability distribution function) and define some of the most frequently referred to PDFs, such as the Binomial Distribution and Normal Distribution.
- Extract data from various data sources (databases, web-scraping, Application Programming Interfaces (APIs) and integrate such data into forms that are usable for further analysis.
- Describe the concept of Gradient Descent for optimization

PRIOR LEARNING ASSESSMENT & RECOGNITION (PLAR)

None

HOURS

Lecture: 10
Self-paced: 20
Other: 20

INSTRUCTIONAL STRATEGIES

GRADING SYSTEM

Satisfactory/Unsatisfactory

PASSING GRADE

S, based on minimum 80% attendance

EVALUATION PLAN

None

COURSE TOPICS

- Introduction to Data Science
  A short intensive review of Python
  Data Visualization
  Linear Algebra
  Statistics
  Probability
  Hypothesis and Inference
  Retrieving data from databases, web scraping, APIs
  Working with Data
  Gradient Descent

LEARNING RESOURCES

Data Science from Scratch by Joel Grus

Notes:

- Course contents and descriptions, offerings and schedules are subject to change without notice.
- Students are required to follow all College policies including ones that govern their educational experience at VCC. Policies are available on the VCC website at:

  https://www.vcc.ca/about/governance--policies/policies/ .
- To find out how this course transfers, visit the BC Transfer Guide at https://www.bctransferguide.ca.

Broadway campus
1155 East Broadway
Vancouver, B.C. Canada
V5T 4V5

Downtown campus
250 West Pender Street
Vancouver, B.C. Canada
V6B 1S9

Annacis Island campus
1608 Cliveden Avenue
Delta, B.C. Canada
V3M 6P1

604.871.7000

vcc.ca

Generated at: 6:59 pm on Apr. 10, 2021