

COURSE NAME: DAT300 Data Collection and Curation

Credit Value: 4
Total Course Hours: 56
Prerequisite Course(s): None
Corequisite Course(s): None

COURSE DESCRIPTION

Through this course, students are introduced to data sources, and data validation. Students will appreciate the importance of collecting data. Students will analyze data sources and assess the data against a variety of criteria such as accuracy, validity, completeness, consistency and uniformity. Finally, students will have the opportunity to collect data from websites using prevalent web-scraping tools.

PLAR INFORMATION

This course is eligible for Prior Learning Assessment and Recognition. Students are advised to discuss options with their program coordinator.

COURSE LEARNING OUTCOMES

Upon completion of this course, the student will have reliably demonstrated the ability to:

- | | |
|--|--|
| <p>1.0 Discuss the components and considerations of a collection action plan that addresses a desired outcome.</p> <p>1.1 Explain the differences between the collection of qualitative and quantitative data for analytical purposes.</p> <p>1.2 Distinguish between primary and secondary sources of data.</p> <p>1.3 Identify policies, procedures, metrics, technology and resources for ensuring data quality.</p> <p>1.4 Recognize ethical procedures and best practices in data collection activities.</p> <p>1.5 Identify potential threats to data security and ways to prevent unauthorized access, unauthorized use or loss of data.</p> <p>1.6 Show how to secure a data set.</p> <p>1.7 Describe the importance of understanding the EU General Data Protection Regulation (GDPR).</p> <p>2.0 Validate data from several sources.</p> <p>2.1 Specify steps in validating a data set.</p> <p>2.2 Assess the credibility of the data source.</p> <p>2.3 Validate potential sources of data to determine the accuracy of the data.</p> <p>2.4 Validate potential sources of data to determine the reliability of the data.</p> | <p>2.5 Validate potential sources of data to determine the completeness of the data.</p> <p>2.6 Validate potential sources of data to determine the consistency of the data.</p> <p>2.7 Validate potential sources of data to determine the uniformity of the data.</p> <p>2.8 Explain the strengths and limitations to validating data from secondary sources.</p> <p>2.9 Discuss ways to ensure integrity of data collection, storage, analysis and presentation processes.</p> <p>3.0 Collect primary data using a survey tool.</p> <p>3.1 Create a data collection plan for the collection of primary data.</p> <p>3.2 Explain sampling theory and how it applies to the data collection activity.</p> <p>3.3 Design the survey tool.</p> <p>3.4 Collect data using a survey.</p> <p>3.5 Collate and store data in preparation for analysis.</p> <p>4.0 Collect secondary data from an online data source that addresses an identified goal.</p> <p>4.1 Define open data and explain its importance.</p> <p>4.2 Produce a data collection plan for a secondary data source from the web.</p> <p>4.3 Create data sets from secondary sources of</p> |
|--|--|

structured and unstructured data found on the web.

4.4 Collate, prepare and store data from open data sources for an analysis project.

5.0 Design and use a basic Python script.

5.1 Contrast Python and Visual Basic for Applications (VBA) as a tool for data collection activities.

5.2 Define the structures and components of a Python program.

5.3 Explain the concepts of file input and output.

5.4 Install and run the Python Interpreter.

5.5 Interpret provided Python scripts.

5.6 Design and run a basic Python script.

6.0 Use Python to collect data from websites.

6.1 Explain the web scraping process and list the steps involved.

6.2 Identify common data scraping tasks that can support business decision making.

6.3 Discuss the ethical and privacy issues concerning data scraping on the web.

6.4 Analyze a web page to identify elements and tags which could be useful to specific scraping tasks.

6.5 Troubleshoot potential problems when web scraping such as duplicate information, infinite scroll websites, and iframes.

GENERAL EDUCATION

This is not a General Education course.

PROGRAM OUTCOMES

This course contributes to the following Ministry of Colleges and Universities approved program learning outcomes (PLO):

Enterprise Analysis and Research

1. Identify, collect, organize, manipulate and analyze data to support problem solving, organizational decision-making and opportunity identification.

8. Apply project management principles to gather information, build solutions and sustain analytics projects.

ESSENTIAL EMPLOYABILITY SKILLS OUTCOMES

This course contributes to the following Ministry of Colleges and Universities approved essential employability skills (EES) outcomes:

6. Locate, select, organize, and document information using appropriate technology and information systems.

7. Analyse, evaluate, and apply relevant information from a variety of sources.

EXTERNAL COURSE ACCREDITATIONS AND CONDITIONS

COURSE EVALUATION

30% In Class Exercises

30% Tests

40% Projects

PROGRAM SPECIFIC GRADING

Per College Grading System

GRADING SYSTEM

A+:	90-100%	B+:	77-79%	C+:	65-69%	D:	50-54%	S - Satisfactory
A:	85-89%	B:	73-76%	C:	60-64%	F:	0-49%	I - Incomplete
A-:	80-84%	B-:	70-72%	D+:	55-59%			F- Repeat Course, included in GPA
								FS- Failure Supplemental
								FR- Repeat course, excluded from GPA

*For a complete chart of grades and descriptions, please see the Grading Policy.

LEARNING RESOURCES

Other Resources:

Required:

A practical Introduction to Python Programming

Author: Brian Heinold, 2012

Creative Commons Attribution-Noncommercial-Share Alike 3.0

https://www.brianheinold.net/python/A_Practical_Introduction_to_Python_Programming_Heinold.pdf

Suggested Resources:

Foundations of Python Programming

Authors: Brad Miller, Paul Resnick, Lauren Murphy, Jeffrey Elkner, Peter Wentworth, Allen B. Downey, Chris Meyers, and Dario Mitchell

GNU Free Documentation License, Version 1.3

<https://runestone.academy/runestone/books/published/fopp/index.html>

Think Python - How to Think Like a Computer Scientist - Version 2.0.17

Author: Allen Downey, published 2012,

Green Tea Press, Massachusetts Creative Commons Attribution-NonCommercial 3.0

<http://greenteapress.com/thinkpython/thinkpython.pdf>

Python Practice Book

Author: Anand Chitipothu

Creative Commons Attribution-NonCommercial 4.0 International License

<https://anandology.com/python-practice-book/>

Python for You and Me

Hands On Python 3 Tutorial

Author: Andrew Harrington, 2019

Creative Commons Attribution-Noncommercial-Share Alike 4.0 United States License

<http://anh.cs.luc.edu/handsonPythonTutorial/#>

Python Tutorial, Release 3.7.0

Authors: Guido van Rossum and the Python development team

September 02, 2018, Python Software Foundation

https://bugs.python.org/file47781/Tutorial_EDIT.pdf

Python Notes for Professionals

The Python Notes for Professionals book is compiled from Stack Overflow Documentation, the content is written by the beautiful people at Stack Overflow. Text content is released under Creative Commons BY-SA. 816 pages, published on June 2018

<https://books.goalkicker.com/PythonBook/PythonNotesForProfessionals.pdf>

A Beginner's Python Tutorial

https://en.wikibooks.org/wiki/A_Beginner%27s_Python_Tutorial

Introduction to Research Programming with Python

<http://rits.github-pages.ucl.ac.uk/intro-research-prog/index.html>

Also: Online sources, pdf's and other hand-outs will be provided.

Resources listed on the course outline support the achievement of learning outcomes, and may be used throughout the course to varying degrees depending on the instructor's teaching methodology and the nature of the resource.

LEARNING ACTIVITIES

Readings, Lectures, Discussions, Group Work, Research

DELIVERY MODE

This course may be delivered, in whole or in part, in a number of modalities, including in class, online, hybrid, in a synchronous or asynchronous manner or a combination thereof, as per accreditation and/or regulatory standards where appropriate.

ACADEMIC POLICIES

Canadore College is committed to the highest standards of academic integrity, and expects students to adhere to these standards as part of the learning process in all environments. The College's Academic Integrity policy seeks to ensure that all students understand their rights and responsibilities in upholding academic integrity and that students receive an accurate and fair assessment of their work. Please review the Academic Integrity policy (A-18) and other academic policies found on our website:

<https://www.canadorecollege.ca/about/policies>.

COLLEGE POLICIES

- Protecting human rights in support of a respectful college community

For college policies please see: <http://www.canadorecollege.ca/about-us/college-policies>.

STUDENT SUCCESS SERVICES - Your Success Matters!

Student Success Services provides student-focused services to facilitate students' success in their studies. Staff provide support by reducing and/or removing educational-related barriers through individualized accommodations and supports to students with disabilities.

Please visit our webpage to learn more: <https://www.canadorecollege.ca/support/student-success-services> or look for our events on social media.

To connect with Student Success Services email studentsuccessnow@canadorecollege.ca or call 705.474.7600 ext 5205.

FIRST PEOPLES' CENTRE:

A culturally safe environment offering CONFIDENTIAL student focused services, drop in or make an appointment to access:

- One on one counselling
- Elder in residence program
- Peer tutoring
- Peer mentorship
- Lunch & learn workshops on study skills, self-care, life skills
- Learning Resource Centre

Drop by our offices at C254 College Drive, E101 Commerce Court or call 705 474 7600 Ext. 5961 College Drive / 5647 Commerce Court.

<https://www.canadorecollege.ca/experience/indigenous-student-experience>

WAIVER OF RESPONSIBILITY

Every attempt is made to ensure the accuracy of this information as of the date of publication. The college reserves the right to modify, change, add, or delete content.

HISTORICAL COURSE OUTLINES

Students use course outlines to support their learning. Students are responsible for retaining course outlines

for future use in applications for transfer of credit to other educational institutions.