



Course outline

This programme first introduces HTML as a format for structuring and embedding semantic metadata into text documents, which can yield additional data insights when analysing web-pages. Participants then learn to perform web scraping to collect raw text data and perform textual analysis through one of the best text processing facilities of any modern programming language - the Python NLTK package.

Course objectives

Web scraping is an essential component of any data analyst's toolkit as it grants access to the world's largest repository of raw data - the internet. Used to trawl data for competitive advantages, web scraping has gained demand among firms as a skill to ultimately make sense of the raw findings.

At the end of the course, participants will be able to:

- Understand HTML as a semantic markup language for structuring text in a web page
- Complete an end-to-end web scraping and text processing as a practice

- Extract data from a webpage using BeautifulSoup 4 Library
- Learn to pre-process text data with NLTK package

Course details

1 week

Certificated by Singapore Management University (SMU)

Who should attend

Aspiring data science professionals seeking to apply Python to real world data problems
Anyone with an interest in learning about the advanced data analysis techniques and putting it in practice
Managers looking into costs and performance hurdles through predictive modelling

Pre-requisites

Preferably for participants who have completed the course

Tools

NLTK package

Model of training

Classroom,
Field trip

