ANL503 Data Wrangling

Level: 5

Credit Units: 5 Credit Units

Language: ENGLISH

Presentation Pattern: EVERY SEMESTER

Synopsis:

ANL503 Data Wrangling aims to equip students with advanced data acquisition and manipulation skills and techniques. Using MySQL, Python, and R, students learn the ins and outs of acquiring data from relational database systems (using SQL), web scraping, and web APIs (using Python) in a scalable and reproducible manner. Students also learn how to transform raw data into formats suitable for deeper analytics (using MySQL, Python, and R). The course rounds up with an introduction to visualisation (using R).

Topics:

- Introduction to the MySQL RDBMS and SQL as a glue language for analytics
- Essential concepts in probability and statistics
- Introduction to SQL and data manipulation with the SELECT statement
- Combining data from multiple sources with union and joins
- Understanding regular expressions
- Introduction to R
- Data manipulation with R
- Essential principles of data visualisation
- Introduction to Python programming
- Practical Python for data acquisition, mangling, and reporting
- Handling web APIs and web scraping
- Using Python for scalable spreadsheet data acquisition

Textbooks:

Zed A. Shaw: Learn Python 3 the Hard Way: A Very Simple Introduction to the Terrifyingly Beautiful World of Computers and Code (eText) Pearson
ISBN-13: 0134693903

Alan Beaulieu Released March 2020: Learning SQL, 3rd Edition 3rd edition O'Reilly Media, Inc
ISBN-13: 9781492057611

Learning Outcome:

- Assess appropriateness of database designs based on characteristics of data and analytics needs
- Critique data visualisations constructively
- Construct suitable SQL queries to acquire and reshape data
- Assemble effective data flows in MySQL, Python, and R as part of a reproducible workflow process
- Create Python scripts to automate the acquisition and processing of web and spreadsheet data
- Design and implement effective data visualisations in R

Assessment Strategies (Evening Class):

Components	**Description**	**Weightage Allocation (%)**
Overall Continuous Assessment	PRE-COURSE QUIZ 1	10
	TUTOR-MARKED ASSIGNMENT 1	30
	PARTICIPATION 1	10
Overall Examinable Components	ECA	50
Total		**100**