STQD6014: DATA SCIENCE
1st SEMESTER: ACADEMIC YEAR 2024/2025

Name and course code:	STQD 6014, Data Science
Lectures time:	Tuesday (9am – 1pm)
Lecturers:	Dr. Nor Hamizah Miswan Dr. Nurul Afiqah Burhanuddin Dr. Bernard Lee Kok Bang
Tel. No.:	03-8921 3703 03-8921 3717 03-8921 5784
Email:	norhamizah@ukm.edu.my nurul.afiqah@ukm.edu.my bernardlkb@ukm.edu.my

Course Evaluation:

- Test 20%
- Projects 40%
- Final Exam 40%

Course objectives:

Upon the completion of this course, the students should be able to:

1. mastering the fundamentals of data science comprehensively and understanding the latest data technologies
2. develop programs for data-related analysis using Python.
3. apply concepts of data exploration, visualization, and cleaning.
4. communicate the results and statistical findings derived from real-world problem analysis.

Course Synopsis:

This course aims to expose students to the basic principles of data science and Python programming. Concepts and types of data related to it are also introduced to students. This course also covers algorithms, processes, methods and analyses used in the field of data science with examples and discussions using Python. Other topics discussed include the latest data technologies for data storage and archiving.

References:

1. McKinney, W. 2023. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter. 3rd Ed. Sebastopol: O'Reilly Media.
2. Matthes, E. 2023. Python Crash Course, 3rd Edition: A Hands-On, Project-Based Introduction to Programming. 3rd Ed. San Francisco: No Starch Press.

3. Agresti, A. & Kateri, M. 2021. Foundations of Statistics for Data Scientists: With R and Python. 1st Ed. United Kingdom: Chapman and Hall/CRC.
4. Klosterman, S. 2021. Data Science Projects with Python: A case study approach to gaining valuable insights from real data with machine learning. 2nd Ed. Birmingham: Packt Publishing.
5. Bruce, P., Bruce, A. & Gedeck, P. 2020. Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python. 2nd Ed. Sebastopol: O'Reilly Media.

Course contents

Week	Contents	Lecturer	Notes
Week 1 7-13 Oct 2024	**Introduction to Data Science**	Dr. Nor Hamizah	
Week 2 14-20 Oct 2024	**Introduction to Computing and Python:** • Variables & Objects • Lists	Dr. Nor Hamizah	
Week 3 21-27 Oct 2024	**Python**: • If Statements • Dictionaries • User Input and While Loops	Dr. Nor Hamizah	
Week 4 28 Oct-3 Nov 2024	**Python**: • Functions • Classes	Dr. Nor Hamizah	
Week 5 4-10 Nov 2024	**Python**: • Files and Exceptions • Testing Your Code	Dr. Nor Hamizah	**Test (20%)**
Week 6 11-17 Nov 2024	**NumPy:** Arrays and Vectorized Computation	Dr. Nurul Afiqah	
Week 7 18-24 Nov 2024	**NumPy:** Arrays and Vectorized Computation Data Manipulation with **Pandas**	Dr. Nurul Afiqah	
Week 8 25 Nov-1 Dec 2024	Data Manipulation with **Pandas**	Dr. Nurul Afiqah	**Project 1 (15%)**
Week 9 2-8 Dec 2024	**MID SEMESTER BREAK**		
Week 10 9-15 Dec 2024	Data Loading and Storage	Dr. Nurul Afiqah	

Week 11 16-22 Dec 2024	Data Wrangling: Clean, Transform, Merge, Reshape	Dr. Nurul Afiqah	
Week 12 23-29 Dec 2024	Plotting and Visualization **Matplotlib**	Dr. Bernard Lee	
Week 13 30 Dec 2024 – 5 Jan 2025	Data Aggregation and Group Operation	Dr. Bernard Lee	
Week 14 6-12 Jan 2025	Data Analysis	Dr. Bernard Lee	**Project 2 (25%)**
Week 15 13-19 Jan 2025	Time Series	Dr. Bernard Lee	
Week 16 20-26 Jan 2025	Introduction to Modelling	Dr. Bernard Lee	
27 Jan- 2 Feb 2025	**REVISION WEEK**		
3-23 Feb 2025	**FINAL EXAMINATION WEEKS**		