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

Name and course code:	STQD 6014, Data Science
Lectures time:	Please see the Course Content below
Lecturer:	Dr. Bernard Lee Kok Bang Dr. Nor Hamizah Miswan
Tel. No.:	03-8921 3712
Email:	bernardlkb@ukm.edu.my norhamizah@ukm.edu.my

Course Evaluation:

- Projects 40%
- Mid-sem Test 20%
- 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

Lectures	Contents	Lecturer	Notes
Week 1 **(2024-10-19)** **9.00 am – 1.00 pm**	**Introduction to Data Science**	Dr. Nor Hamizah	
Week 2 **(2024-10-26)** **9.00 am – 1.00 pm**	**Introduction to Computing and Python:** • Variables & Objects • Lists	Dr. Nor Hamizah	
Week 3 **(2024-10-27)** **2.00 pm – 6.00 pm**	**Python**: • If Statements • Dictionaries • User Input and While Loops	Dr. Nor Hamizah	
Week 4 **(2024-11-09)** **2.00 pm – 6.00 pm**	**Python**: • Functions • Classes	Dr. Nor Hamizah	
Week 5 **(2024-11-16)** **9.00 am – 1.00 pm**	**Python**: • Files and Exceptions • Testing Your Code	Dr. Nor Hamizah	
Week 6 **(2024-11-23)** **2.00 pm – 6.00 pm**	**NumPy:** Arrays and Vectorized Computation	Dr. Nor Hamizah	**Mid-Sem Test (20%)**
Week 7 **(2024-11-30)** **9.00 am – 1.00 pm**	Data Manipulation with **Pandas**	Dr. Nor Hamizah	**Project 1 (15%)**
Week 8 **(2024-12-07 - 2024-12-08)**	**MID SEMESTER BREAK**		
Week 9 **(2024-12-14)** **9.00 am – 1.00 pm**	Data Loading and Storage	Dr. Bernard Lee	
Week 10 **(2024-12-14)** **2.00 pm – 6.00 pm**	Data Wrangling: Clean, Transform, Merge, Reshape	Dr. Bernard Lee	
Week 11 **(2024-12-28)** **2.00 pm – 6.00 pm**	Plotting and Visualization **Matplotlib**	Dr. Bernard Lee	**Project 2 (25%)**

Week 12 **(2025-01-04)** **2.00 pm – 6.00 pm**	Data Aggregation and Group Operation	Dr. Bernard Lee	
Week 13 **(2025-01-11)** **2.00 pm – 6.00 pm**	Data Analysis	Dr. Bernard Lee	
Week 14 **(2025-01-18)** **2.00 pm – 6.00 pm**	Time Series	Dr. Bernard Lee	
Week 15 **(2025-01-25)** **2.00 pm – 6.00 pm**	Introduction to Modelling	Dr. Bernard Lee	
2025-02-03 **-** **2025-02-23**	**FINAL EXAMINATION WEEKS**		**Dr. Nor Hamizah (15%)** **Dr. Bernard Lee (25%)**