ZU/WI/7/11

Unit Code:	BDM 221
Unit Title:	DATA SCIENCE PROGRAMMING
Pre-requisites	BDM 121, STA 113,, BSD 122.
Program(s):	BSE
Lecturer Name:	Mr. George Wainaina
Lecturer Contacts:	Email: george.wainaina@zetech.ac.ke, Phone No: 0718313173 / 0731919231
Consultation time:	

UNIT DESCRIPTION/ OBJECTIVES OF THE UNIT

The purpose of this course is to make students use their skills in math, statistics, programming, and other related subjects to organize large data sets. Students will learn how to apply their knowledge to uncover solutions hidden in the data to take on business challenges. Students will learn about business issues and how to devise smart and purposeful ways through use of analytical solutions by making sense of big sets of data.

EXPECTED LEARNING OUTCOMES

By the end of this course the students should be able to:

i). Apply theoretical concepts to build and assess data-based models to assist in carrying out research-based activities.

ii). Write fully-functional Python programs using commonly used data structures, custom functions, and reading and writing to files

iii). Apply computing theory, languages, and algorithms, as well as mathematical and statistical models, to appropriately formulate and use data analyses to solve real life problems.

iv). Interpret data using an ethically responsible method.

PRE-REQUISITE COURSE:

- **BDM 121:** Database Management Systems.
- **STA 113:** Probability and Statistics.
- **BSD 122**: Object Oriented Programming II.

COURSE SYLLABUS AND SCHEDULE

Week	Topic	Credit hours
Week 1.	**Python Basics:** IDE, Python Comments, Data types, variables, modules, libraries, arguments I/O. Functions: definitions, variables, parameters, importing modules, returning values	3
Week 2.	**Flow Control:** Conditional Statements, operators, loops. Virtual Environment: Creating virtual environment, manipulation of virtual environment.	3
Week 3	Introduction to data science process and the value of Learning data science Data science Lifecycle/ methodology Python Libraries for data science & visualization Using Statistics Case study	3
Week 4	Data Wrangling	3
Week 5	Feature Engineering **ASSIGNMENT 1**	3
Week 6	Exploratory Data analysis	3
Week 7	**CAT1**	3
Week 8	Machine Learning models Supervised methods	3
Week 9	**ASSIGNMENT 2**	3
Week 10	Machine Learning models Un Supervised methods	3
Week 11	Mini Data Science projects	3
Week 12	**CAT2**	3
Week 13	Exam Revision	3
Week 14	EXAMS	3

TEACHING/LEARNING

Mode of Delivery
The course unit will be delivered through face to face and blended learning.

Teaching Methodology
Lecture method, Group activities, Class discussions, Demonstrations, Illustrations and Lab practical sessions.

Instructional materials/Equipment:
Course texts, Handouts, presentation slides, computer software and Hardware, overhead projectors, Virtual Labs, LMS, Video Conference facility, GitHub.

Course Assessment

Continuous Assessment Test (CATS)	20%
Assignments	10%
Examination	70%
Total	**100%**

Core Reading Materials for the Course:

1. Shah, C. (2020). A Hands-On Introduction to Data Science (1st ed.). Cambridge University Press.
2. Grus, J. (2019). Data science from scratch: first principles with python. O'Reilly Media.
3. Saiz, Z. A., González, Q. C., Gil, H. L., & Ruiz, M. D. (2020). An Introduction to Data Analysis in R: Hands-on Coding, Data Mining, Visualization and Statistics from Scratch (Use R!) (1st ed. 2020 ed.). Springer.

Recommended Reference Materials:

1. Wickham, H., & Grolemund, G. (2017). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data (1st ed.). O'Reilly Media.
2. Spiegelhalter, D. (2019). The Art of Statistics: How to Learn from Data (Illustrated ed.). Basic Books.
3. Emmert-Streib, F., Moutari, S., & Dehmer, M. (2020). Mathematical Foundations of Data Science Using R (De Gruyter STEM) (1st ed.). De Gruyter Oldenbourg.
4. Knaflic, N. C. (2015). Storytelling with Data: A Data Visualization Guide for Business Professionals (1st ed.). Wiley.
5. Martin, O. (2018). Bayesian analysis with Python: introduction to statistical modeling and probabilistic programming using PyMC3 and ArviZ. Packt Publishing Ltd.

E-Resources:

1. Gökalp, M.O., Gökalp, E., Kayabay, K., Koçyiğit, A. & Eren, P.E. (2021). The development of the data science capability maturity model: a survey-based research. Online Information Review, Vol. ahead-of-print No. ahead-ofprint. https://doi.org/10.1108/OIR-10-2020-0469
2. Roman, D., Reeves, N., Gonzalez, E., Celino, I., Abd El Kader, S., Turk, P., Soylu, A., Corcho, O., Cedazo, R., Re Calegari, G., Scandolari, D. & Simperl, E. (2021). An analysis of pollution Citizen Science projects from the perspective of Data Science and Open Science. Data Technologies and Applications, Vol. 55 No. 5, pp. 622-642. https://doi.org/10.1108/DTA10-2020-0253