UNIVERSITY EXAMINATIONS: 2022/2023

EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE IN DATA SCIENCE

BDS 1101: PRINCIPLES OF DATA SCIENCE

FULLTIME/ PART TIME/DISTANCE LEARNING

ORDINARY EXAMINATION

DATE: DECEMBER, 2023   **TIME: 2 HOURS**

INSTRUCTIONS: Question One Is Compulsory, Choose Two Other Questions

QUESTION ONE (20 MARKS) Compulsory

a) You have been hired by Safari Telcom limited as an expert, to give a lecture to its staff on data science and the tasks in data science. You are required to explain the following.

   i.   Give any two tools for data wrangling.   **(2 Marks)**

   ii. Explain the data wrangling process with respect to data science.   **(2 Marks)**

   iii. Describe any two known tools used for data cleaning.   **(2 Marks)**

   iv. Describe two known ways for data blending using Python.   **(2 Marks)**

   v.   Differentiate between structured data and unstructured data. Provide an example in each case.   **(4 Marks)**

b) Describe four most commonly used algorithm by data scientist.   **(4 Marks)**

c) Explain the relationship between data science and machine learning.

                                                                                 (4 Marks)

QUESTION TWO (15 MARKS)

a) Explain *Big data, data science* and *data analytics*, and hence suggest the relationship that exist amongst them, if any. **(5 Marks)**

b) Justify why data scientists should comply to privacy and confidentiality of data. **(5 Marks)**

c) "To ensure that data privacy is protected, data scientists must have a solid understanding of privacy laws and regulations, in the country they serve, as well as best practices for data handling and security". Justify this statement. **(5 Marks)**

QUESTION THREE (15 MARKS)

a) Discuss any three effective communication techniques for data analysts. **(6 Marks)**

b) Discuss the six Steps in the data preparation process. **(9 Marks)**

QUESTION FOUR (15 MARKS)

a) Discuss the following with respect to data science and hence explain the main goal, in each case.

    i. Classification. **(3 Marks)**

    ii. Regression. **(3 Marks)**

    iii. Time series analysis. **(3 Marks)**

b) Identify any four major data cleaning activities and suggest how you would resolve issues that may arise, as a data science expert. **(6 Marks)**