NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

FACULTY OF APPLIED SCIENCES

DEPARTMENT OF INFORMATICS AND ANALYTICS

SIDS5101 BIG DATA ANALYTICS

Examination Paper

2024

This examination paper consists of 4 pages

Time Allowed:	3 hours
Total Marks:	100
Examiner's Name:	Prof N Gasela
External Examiner:	

INSTRUCTIONS

This is a closed book examination. Please attempt all questions. Questions have equal weight.

MARK ALLOCATION

QUESTION	MARKS
1.	20
2.	20
3.	20
4.	20
5.	20
TOTAL	100

QUESTION ONE

a. Discuss the big data analytics value chain, giving examples [14]

b. A car manufacturer has decided to install a global positioning system (GPS) in its new cars to collect data about customers' driving patterns. The customers' data will be stored in a centralized database in the car manufacturer's server room at its head office in Bulawayo.

Identify the key legal requirement that the car manufacturer needs to meet with regard to this data and describe what the manufacturer will need to do to meet this requirement.

[6]

QUESTION TWO

a. Describe any three (3) preprocessing steps and provide relevant examples. [6]

b. A Zimbabwean athletics official has collected thousands of track and field event results for athletes all around Australia in the last 10 years. This data has been collected from a range of data sets and then imported into a spreadsheet. Before the official can present findings using a data visualization, they will have to first cleanse the data.

   i. Explain why cleansing data is a necessary process during data manipulation and analysis. [2]

   ii. Describe one technique the official can use to cleanse data in the spreadsheet, and provide an example. [2]

c. Given a dataset named 'marks' with a numeric column called 'grade,' write Python code to eliminate rows with missing values and remove outliers using either the interquartile range method or the standard deviation method. [10]

QUESTION THREE

a. Discuss five (5) critical success factors of big data analytics. [10]

b. An organization would like to collect medical data from a range of primary and secondary sources around Zimbabwe. The organization wants to query a range of data sets, identify relationships and patterns, and manipulate and cleanse data. It would then like to sell this data to health insurance companies.

i. At some point in the next few years, the organization will have to consider archiving the files and disposing of the files on its servers. Explain the difference between archiving files and disposing of files. [6]

ii. Consideration is being given to enabling health insurers to access this data via cloud storage. State one advantage and one disadvantage of using cloud storage for storing data. [4]

QUESTION FOUR

a. Provide five arguments for visualizing data to aid a business in decision-making. [10]

b. Mandla has completed his dynamic data visualization on greenhouse gas emissions. He needs to test his visualization for navigation and error tolerance of the interface.

    i. Explain the difference between navigation and error tolerance. [2]

    ii. Provide one suitable technique for testing navigation. [1]

    iii. Provide one suitable technique for testing error tolerance. [1]

c. For each of the following plot types, describe at least one situation where you would choose this plot over the others and describe the type of data it would be appropriate for using the visual
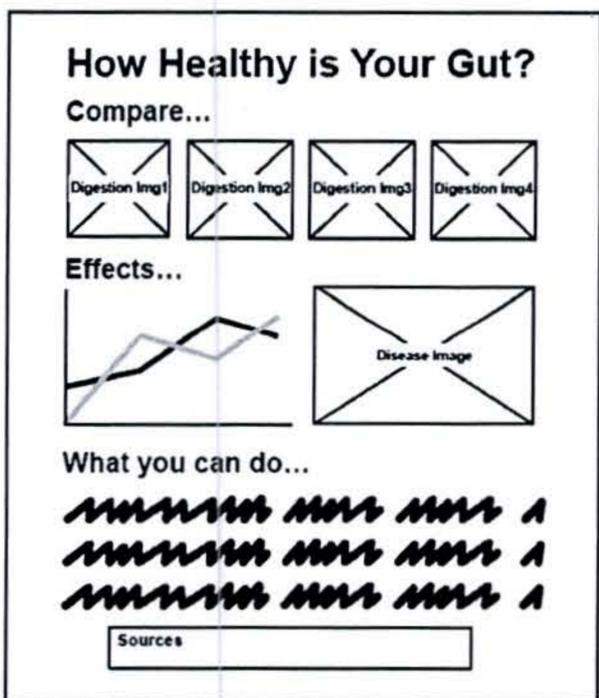
    i. Line plot. [3]

    ii. Box plot. [3]

QUESTION FIVE

a. Write notes on how data and image models are used in creating effective data visualizations. [8]

b. Explain why text is added to a data visualization presentation. [4]

c. An organization that promotes healthy eating feels that there is not enough knowledge about the health issues of elderly people related to poor diet and digestion. After collecting and analyzing data, the organization has decided to develop an infographic aimed at educating the elderly about digestive health.
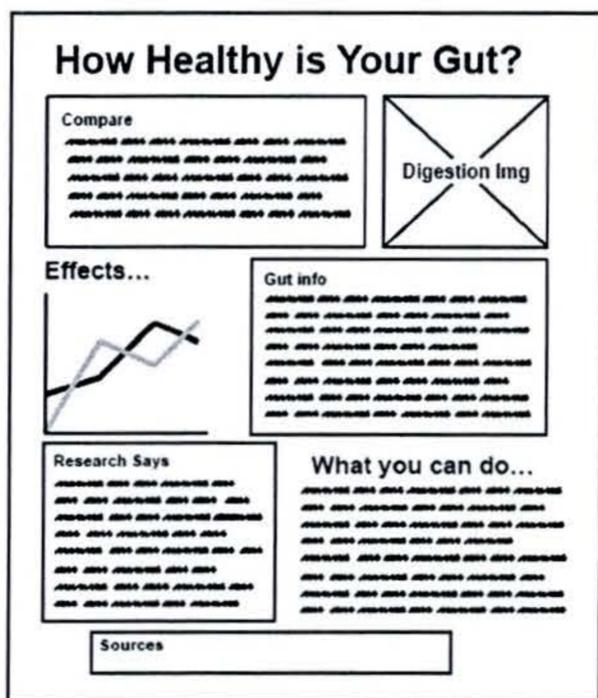
i.   Outline a technique the organization could use to generate design ideas for the infographic. [1]

ii. The organization will need to consider design principles when designing the infographic. Identify and describe one design principle related to appearance that is appropriate for the target audience. [2]

iii. One criterion for evaluating the alternative designs is stated below:

The infographic should be able to communicate the intended message to the target audience.

Based on the criterion above, justify which idea, A or B, would be more appropriate [5]

A

B

END OF QUESTION PAPER