NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

FACULTY OF APPLIED SCIENCES

DEPARTMENT OF INFORMATICS AND ANALYTICS

SIIS5106

DATA ANALYTICS AND VISUALISATION

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

1. **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

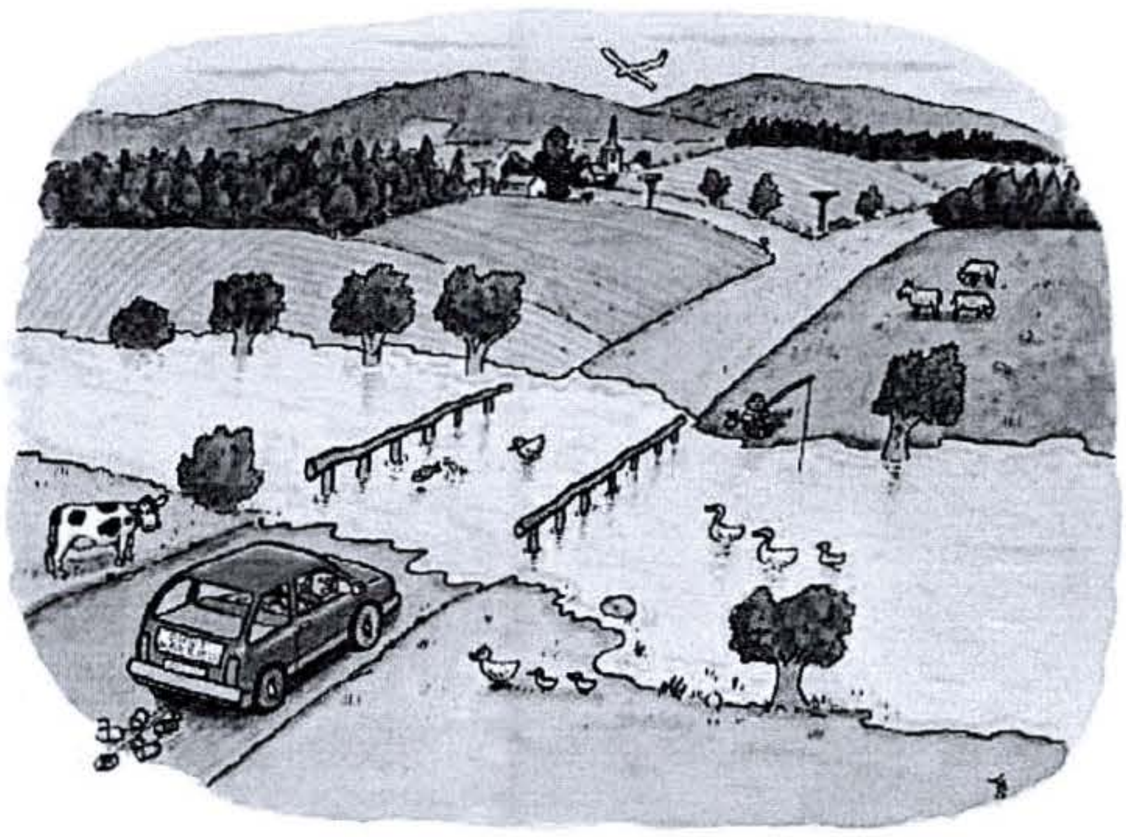
QUESTION ONE

a. Data analytics can be classified into four main types. Give a brief description of each of the types. [8]

b. Name and describe two sub types of categorical vs numerical data, **and** include an example for each. [4]

c. List four (4) differences between qualitative and quantitative data. [4]

d. Given the picture below, make two qualitative and two quantitative observations about it and explain why these are qualitative or quantitative observations. [4]



QUESTION TWO

a. Give a detailed discussion of the steps of the data analysis lifecycle [6]

b. What purpose does the availability of Python code libraries such as NumPy and Scikit-learn serve? [4]

c. Write Python code that uses the matplotlib library to create a line graph representing the marks of a student for 4 tests. The x-axis should be labelled 'Mark', and the y-axis should be labelled 'Tests'. The graph title should be 'Student Marks'. Use an asterisk (*) as the marker with a marker size 10.

The line style should be dashed, and the line width should be 2. The data for the graph is represented in Table 1.1 [10]

Table 1.1

Tests	Test1	Test2	Test3	Test4
Mark	58	45	75	80

QUESTION THREE

a. . What are the main data pre-processing steps? Briefly describe any three (3) of the steps and provide relevant examples. [9]

b. Sipho and Chipo run an online business that allows people to buy and sell old sneakers and clothes. The data from their business is stored in a database that is hosted on a server in their office. A full backup of their business data is performed at the start of each month, and a differential backup is performed each week on Friday night. All backups are stored on an external hard disk drive (HDD) located in the office. The backups are regularly tested, and all backups are created successfully.

   i. If the server had a hard drive failure, what is the maximum number of days of data that will be lost if the database needs to be restored from backup? Explain your answer. [4]

   ii. Suggest and justify **two** changes to the backing up strategy to improve data protection and the ability to recover from a disaster. [4]

c. A data analyst conducting a study stops collecting more data once the evidence starts to support the hypothesis. What type of bias has the analyst introduced into the findings? Explain. [3]

QUESTION FOUR

4.1. What is data visualization? Using examples explain why data visualization is important in data analytics. [10]

4.2. Meals by Delivery is a meal delivery service. Customers can use a mobile phone app to order meals, for dinner only, from fine-dining restaurants located in districts around the town hall of any city or town that Meals by Delivery services. The app connects to Meals by Delivery's central database. Currently, Meals by Delivery is restricting deliveries to customers who live within a 3 km radius of a town hall.

Meals by Delivery's sales manager has suggested that the business expand its operations to include deliveries within a 10 km radius, as he believes there are many more families that would enjoy fine-dining meals and would pay for the delivery service.

The sales manager has commissioned a marketing company to gather data from residents about their eating habits when it comes to meal deliveries in order to determine if expanding the company's operations would be worthwhile. This information will be presented as an infographic with data visualizations representing customer and family eating habits.

a. Describe a method that the marketing company could use to gather data to identify constraints for the data visualizations. Suggest why this method would be suitable for this task. [4]

b. Identify one example of a constraint for the data visualizations. [1]

c. Suggest how the marketing company could generate a list of the most important non-functional requirements for the data visualizations based on the data it gathers from both residents and Meals by Delivery. [5]

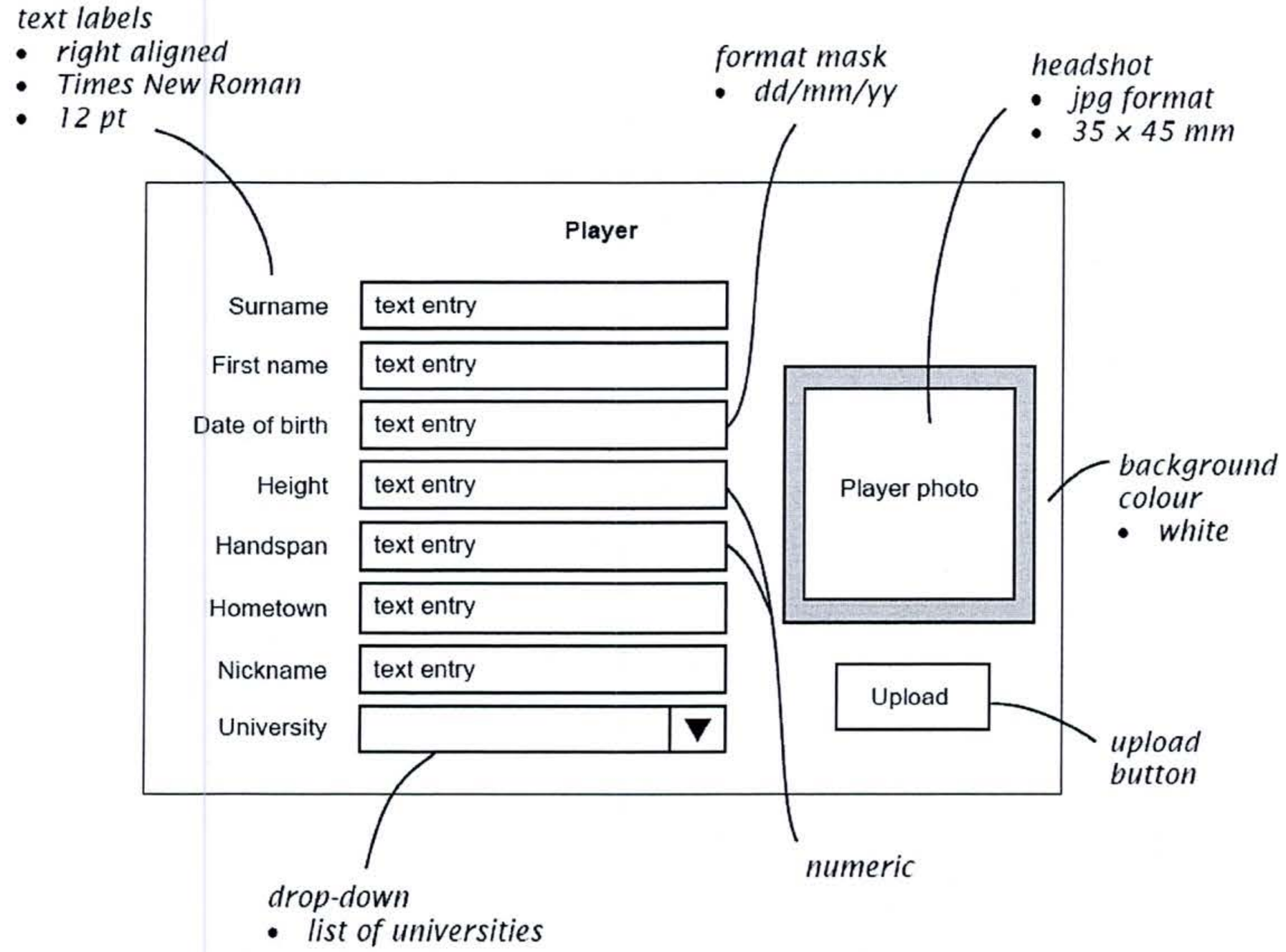
QUESTION FIVE

a. Discuss on how data visualization is used for achieving the following objectives:

   i. To determine the relative importance of variables. [5]

   ii. To find the relationships among data variables. [5]

b. The Zimbabwe Basketball League has developed the design below for the input form of its player database.



i. Identify the design tool that has been used to create this input form. [2]

ii. Identify one design principle relating to the appearance of this input form. [3]

Data from the input form has been exported into a spreadsheet, part of which is shown below. One of the team managers would like to know the average height and handspan of the players on their team.

Surname	**First name**	**Date of birth**	**Height**	**Handspan**	**Town**	**Nickname**	**Campus**
Banda	Joe	15/05/00	203 cm	26 cm	Gutu	Mude	NUST
Phiri	Ushe	27/12/01	2.01 m	24 cm	Gwanda	Shorty	LSU

c. Describe two techniques for cleansing data, which can be used before the data is manipulated and any calculations made. [5]

END OF QUESTION PAPER