UNIVERSITY OF MALAYA

EXAMINATION FOR THE DEGREE OF MASTER OF DATA SCIENCE

ACADEMIC SESSION 2018/2019   : SEMESTER I

WQD7001   : Principles of Data Science

Jan 2019   Time : 2 hours

INSTRUCTIONS TO CANDIDATES :
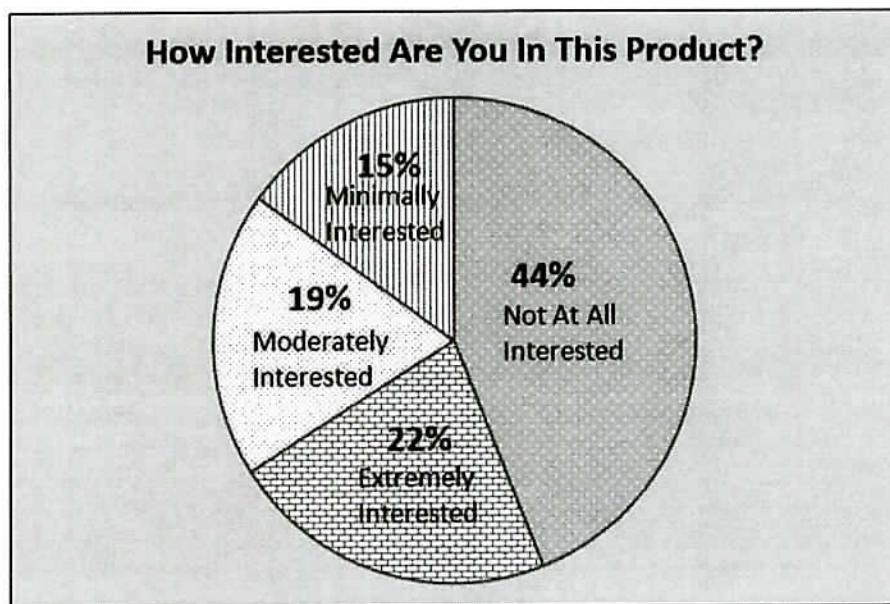
Answer **ALL** questions (50 marks).

(This question paper consists of 11 questions on 3 printed pages)

1. State **THREE (3)** types of data properties that you can investigate in exploratory data analysis (EDA).

(3 marks)

2. Why skipping EDA is a bad idea?

(4 marks)

3. Which language is more suitable for text analytics? R or Python? Defend your answer.

(3 marks)

4. Data scientist requires skills in statistics.
   Differentiate between descriptive statistics and inferential statistics.

(5 marks)

5. Illustrate the five-number summary using a boxplot.

(5 marks)

6. Linear Regression Analysis consists of more than just fitting a linear line through a cloud of data points.
   Structure and explain the **THREE (3)** steps for linear regression.

(6 marks)

7. Discuss the **FIVE (5)** questions that machine learning can help answer, and for each question indicate the related algorithm.

(5 marks)

8. In enhancing reproducibility, what is your advice or principles in terms of facilitating reproducibility and enabling discoverability?

(4 marks)

9. DJ Patil defines a Data Product as "... a product that facilitates an end goal through the use of data".
   Provide **TWO (2)** broad groups of data products function.

(5 marks)

10. Summarize using a diagram, potential data storytelling visuals by grouping them based on relationship, comparison, distribution, or composition.

(6 marks)

11. Evaluate the pie chart below and suggest a way that could improve the data storytelling. Depict your suggestion.

How Interested Are You In This Product?

15% Minimally Interested

19% Moderately Interested

44% Not At All Interested

22% Extremely Interested

(4 marks)

END