Series ϵFGHE

Set-4

Q.P. Code **368**

Roll No.

Candidates must write the Q.P. Code on the title page of the answer-book.

DATA SCIENCE

*Time allowed : **2** hours* *Maximum Marks : **50***

- Please check that this question paper contains **11** printed pages.
- Q.P. Code given on the right hand side of the question paper should be written on the title page of the answer-book by the candidate.
- Please check that this question paper contains **21** questions.
- **Please write down the serial number of the question in the answer-book before attempting it.**
- 15 minute time has been allotted to read this question paper. The question paper will be distributed at 10.15 a.m. From 10.15 a.m. to 10.30 a.m., the students will read the question paper only and will not write any answer on the answer-book during this period.

General Instructions :

(i) Please read the instructions carefully.

*(ii) This question paper consists of **21** questions in **two** sections : **Section A** and **Section B**.*

*(iii) **Section A** has Objective Type Questions whereas **Section B** contains Subjective Type Questions.*

(iv) Out of the given (5 + 16 =) 21 questions, a candidate has to answer (5 + 10 =) 15 questions in the allotted (maximum) time of 2 hours.

(v) All questions of a particular section must be attempted in the correct order.

(vi) ***Section A*** *: Objective Type Questions (24 marks) :*

*(a) This section has **5** questions.*

(b) There is no negative marking.

(c) Do as per the instructions given.

(d) Marks allotted are mentioned against each question/part.

(vii) ***Section B*** *: Subjective Type Questions (26 marks) :*

*(a) This section has **16** questions.*

*(b) A candidate has to do **10** questions.*

(c) Do as per the instructions given.

(d) Marks allotted are mentioned against each question/part.

Section A

(Objective Type Questions) 24 Marks

1. Answer any ***4*** out of the given ***6*** questions on Employability Skills. *4×1=4*

(i) A/An _____ may be defined as underlying characteristic of a person which results in effective and/or superior performance in a job. *1*

(a) uncertainty

(b) competence

(c) obstacles

(d) fear

(ii) Which of the following is ***not*** true for formulas in a spreadsheet ? *1*

(a) A formula always starts with an equal to (=) sign.

(b) A formula is displayed in the formula bar.

(c) Formulae are used to calculate results through arithmetic operations.

(d) In numeric formulae, you cannot make use of operators.

(iii) Given below are two statements, one labelled as Assertion (A) and the other labelled as Reason (R). Select the correct answer from the code given below : *1*

Assertion (A) : Physiological motivation can be guided by the need for achievement and the need for affiliation.

Reason (R) : The need for achievement is a social form of motivation involving a competitive drive to meet the standards of excellence.

Code :

(a) Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A).

(b) Both Assertion (A) and Reason (R) are true, but Reason (R) is not the correct explanation of Assertion (A).

(c) Assertion (A) is true, but Reason (R) is false.

(d) Assertion (A) is false, but Reason (R) is true.

(iv) If the column is not wide enough to display the value, which of the following errors is displayed ? *1*

(a) #VALUE! (b) #DIV/0!

(c) ##### (d) #COLUMN!

(v) Which term is used to describe a person's ability to recognize what results are important and the steps needed to be taken to achieve them ? *1*

(a) Motivation

(b) Result Orientation

(c) Stress Management

(d) Goal Setting

(vi) Given below are two statements, one labelled as Assertion (A) and the other labelled as Reason (R). Select the correct answer from the code given below : *1*

Assertion (A) : Many entrepreneurs are not considered experts in their line of business, but still they make important decisions and solve issues everyday within their industry.

Reason (R) : By preparing oneself to take on the challenge and by taking smaller steps to work towards it, fearful situations can also start to feel comfortable.

Code :

(a) Both Assertion (A) and Reason (R) are true and Reason (R) is the correct explanation of Assertion (A).

(b) Both Assertion (A) and Reason (R) are true, but Reason (R) is not the correct explanation of Assertion (A).

(c) Assertion (A) is true, but Reason (R) is false.

(d) Assertion (A) is false, but Reason (R) is true.

2. Answer any ***5*** out of the given ***6*** questions. *5×1=5*

(i) _____ is the right of an individual to have control over how his or her personal information is collected and used. *1*

(a) Data Privacy (b) Data Governance

(c) Governing (d) Data Quality

(ii) Which of the following refers to the process of identifying incorrect, incomplete and inaccurate data ? *1*

(a) Exploratory Data Analysis

(b) Data Cleaning

(c) Univariate Analysis

(d) Data Ethics

(iii) Every _____ represents the outcome of the test in a decision tree. *1*

(a) Internal node

(b) External node

(c) Branch

(d) Leaf node

(iv) The _____ algorithm is one of the most basic and easy-to-implement supervised machine learning algorithms. *1*

(a) Decision tree

(b) K-nearest neighbors (K-NN)

(c) Classification

(d) Scoping

(v) _____ is the square root of the variance of the residuals. *1*

(a) Root Mean Square Deviation

(b) Required Mean Similar Deviation

(c) Root Median Similar Deviation

(d) Regular Mean Square Deviation

(vi) In which type of learning do algorithms act on data without human intervention ? *1*

(a) Unsupervised

(b) Supervised

(c) Limited

(d) Conditional

3. Answer any ***5*** out of the given ***6*** questions. *5×1=5*

(i) PDP stands for *1*

(a) Personal Data Protection Bill

(b) Private Data Protection Bill

(c) Personal Device Protection Bill

(d) Personal Device Prevention Bill

(ii) _____ is a more complex form of statistical analysis technique and is used to analyze more than two variables in the data set. *1*

(a) Graphical method

(b) Multivariate analysis

(c) Univariate analysis

(d) Unsupervised learning technique

(iii) Decision trees are used to solve _____ problems. *1*

(a) Only classification

(b) Only regression

(c) Both classification and regression

(d) All universal

(iv) Can we use K-NN algorithms for data mining problems ? *1*

(v) Y = m * X + b is the formula for _____. *1*

(a) Mean Absolute Error

(b) Root Mean Square Deviation

(c) Simple linear regression

(d) Simple linear classification

(vi) State whether the following statement is *true* or *false* : *1*

The formula for non-linear and linear regression is same.

4. Answer any ***5*** out of the given ***6*** questions. *5×1=5*

(i) The Children's Online Privacy and Protection Act is a law that deals with *1*

(a) privacy policy for children who are less than the age of 13 years.

(b) privacy policy for adults.

(c) public policy for all children.

(d) protection policy for mothers of newborns.

(ii) The process of Exploratory Data Analysis is done with the help of summary statistics and _____ representations. *1*

(a) Analytical

(b) Graphical

(c) Logical

(d) Sequential

(iii) Regression trees are used when the dependent variable is non-continuous. Is the given statement *true* or *false* ? *1*

(iv) Which of the following is true for K-NN algorithm ? *1*

(a) Interpretability of the K-NN algorithm is very low.

(b) The K-NN algorithm explicitly has a training step.

(c) The K-NN algorithm is not sensitive to outliers.

(d) K-NN works well with a small number of input variables, but as the number of variables grow, the K-NN algorithm struggles to predict the output of a new data point.

(v) The actual value of the _____ depends on the data and accuracy required. *1*

(a) RMSE (b) MAE

(c) NLR (d) model

(vi) Trigonometric functions are examples of linear functions.

Is the given statement *true* or *false* ? *1*

5. Answer any ***5*** out of the given ***6*** questions. *5×1=5*

(i) Ethics govern the behaviour or actions of an individual. *1*

(a) True

(b) False

(ii) Bivariate analysis is also a good way to measure the _____ between the two variables. *1*

(a) Difference (b) Mean

(c) Correlation (d) Median

(iii) For every possible decision in a decision tree, stemming from the root makes a _____. *1*

(a) Branch (b) Tree

(c) Root (d) Node

(iv) _____ is a non-parametric algorithm, as it does not assume anything about the distribution of the data. *1*

(a) Cross Validation (b) Regression

(c) Dataset (d) K-NN

(v) What is the basic objective of linear regression ? *1*

(a) To reduce the distance between the line and data points to make it minimum.

(b) To increase the distance between the line and data points to make it maximum.

(c) To find the sum of the distance between the line and data points.

(d) To find the average of the distance between the line and data points.

(vi) A _____ regression equation has an intercept on the right-hand side and an explanatory variable with a coefficient. *1*

(a) circular (b) multiple

(c) complex (d) simple

SECTION B

(Subjective Type Questions) **26 Marks**

*Answer any **3** out of the given **5** questions on Employability Skills. Answer each question in 20 – 30 words.* *3×2=6*

6. List any two benefits of Entrepreneurial Competencies. *2*

7. Answer the following questions from the spreadsheet given below : *2*

My Store				
	A	B	C	D
1	Product	Cost Price	Selling Price	Profit/Loss
2	Ruler	9	15	
3	Crayons	23	25	
4	Pens	14	12	
5	Cardboard	35	45	
6	Highest			

(a) Write a formula in cell D2, to calculate the Profit or Loss for the item "Ruler" as the difference of "Cost Price" and "Selling Price".

(b) Write a function in B6 to display the highest cost price of the product.

8. List any two sources of motivation and inspiration. *2*

9. Differentiate between internal and external motivation. Give a suitable example. *2*

10. Match the given attitudes of an entrepreneur with their characteristics : *2*

Attitudes		*Characteristics*
(a) Interpersonal skills	(i)	Ability to take charge and act in a situation before others
(b) Taking initiative	(ii)	Ability to continue to do something even when it is difficult
(c) Decisiveness	(iii)	Ability to work with others
(d) Perseverance	(iv)	Ability to make quick and profitable decisions

*Answer any **4** out of the given **6** questions in 20 – 30 words each.* *4×2=8*

11. Name the areas of focus of data governance. *2*

12. Differentiate between Univariate analysis and Bivariate analysis. *2*

13. Why are decision trees considered to be versatile ? *2*

14. What is the principle on which K-NN algorithm works ? *2*

15. When do we use linear regression ? *2*

16. The formula for non-linear regression is

$y \sim f(x, \beta)$.

What do x and y denote in the given formula ? *2*

*Answer any **3** out of the given **5** questions in 50 – 80 words each.* *3×4=12*

17. Write a short note on General Data Protection Regulation (GDPR). *4*

18. List any four tools and methods used to perform Exploratory Data Analysis. *4*

19. List any four important features of decision trees. *4*

20. List any four disadvantages of K-NN compared to other algorithms. *4*

21. Explain in points the working of 'k-means clustering' algorithm. *4*